

AI and Semantic Modeling for Profile–Opportunity Matching: A Systematic Review with Quantitative Synthesis

Jáder Felipe Silva Duarte
jader_duarte@live.com
d2026100837@unifei.edu.br
Federal University of Itajubá (UNIFEI)
Itabira, Minas Gerais, Brazil

Adler Diniz de Souza
adlerdiniz@unifei.edu.br
Federal University of Itajubá (UNIFEI)
Itajubá, Minas Gerais, Brazil

ABSTRACT

Digital recruitment platforms and web-based recommender systems have intensified the need for methods that connect academic and professional profiles with job, internship, and career opportunities. However, evidence on profile–opportunity matching remains fragmented across task formulations, datasets, ground-truth sources, validation protocols, and metric families. This paper reports a systematic review with subgroup-based quantitative synthesis of AI-based and semantic modeling approaches for profile–opportunity matching. The review protocol was defined in Parsifal and followed PRISMA reporting recommendations. Searches covered studies published between 2020 and 2026 in ACM Digital Library, IEEE Xplore, ScienceDirect, and Scopus. In total, 425 records were identified; after removing 69 duplicates, 356 records were screened, and 64 studies were selected for full-text assessment. After in-depth reading and methodological quality assessment, 43 studies were retained in the main synthesis. Among these studies, 23 were classified as *Strong Include*, 18 as *Include with limitations*, and 2 as *Narrative only*. Quantitative synthesis eligibility classified 13 studies as *Eligible*, 28 as *Partially eligible*, and 2 as *Not eligible*. The evidence is concentrated in resume–job classification, candidate ranking, job recommendation, fit prediction, and student–job recommendation, using NLP, machine learning, deep learning, BERT/SBERT, Transformers, LLMs, recommender systems, knowledge graphs, ontologies, and graph neural networks. Results also show frequent reliance on private or industrial datasets, which limits reproducibility and direct comparability. Because a single statistical meta-analysis was not methodologically appropriate, the quantitative synthesis separates classification-based fit prediction from Top-K recommendation/ranking, providing a more replicable organization of comparable evidence.

KEYWORDS

Artificial Intelligence, Semantic Modeling, Profile–Opportunity Matching, Job Recommendation, Internship Recommendation, Resume–Job Matching, Candidate Ranking, Natural Language Processing, Recommender Systems, Knowledge

Graphs, Systematic Literature Review, Quantitative Synthesis

1 INTRODUCTION

Digital recruitment platforms, career platforms, and internship portals have increased the availability of resumes, job descriptions, academic profiles, and professional records. However, the growth in data volume does not, by itself, eliminate the difficulty of identifying appropriate matches. The problem requires modeling compatibility among profiles, requirements, experiences, skills, and interaction histories rather than merely identifying matching terms [9, 16, 33, 47].

A central limitation of traditional approaches is their reliance on keywords. Resumes, academic profiles, and job descriptions contain semistructured or unstructured textual information in which equivalent skills may be expressed using different vocabularies. To address this limitation, the literature explores machine learning, deep learning, natural language processing, embeddings, Transformer-based models such as BERT and SBERT, ontologies, *knowledge graphs*, Graph Neural Networks (GNNs), Large Language Models (LLMs), hybrid approaches, and recommender systems [1, 7, 11, 14, 15, 18, 26, 37, 40, 44, 45]. These techniques support resume–job classification, candidate ranking, job and internship recommendation, *person–job fit* prediction, Top-K recommendation, semantic search, and skill–job alignment [9, 16, 18, 36, 40, 45, 49].

Despite these advances, the literature remains fragmented in terms of task formulations, compared entities, datasets, *ground truth* construction, validation protocols, baselines, and metrics. This heterogeneity hinders direct comparisons and does not allow the highest values reported under different protocols to be interpreted, in isolation, as evidence of overall superiority. Moreover, the frequent use of private, industrial, or insufficiently documented data reduces reproducibility and limits external validation. Therefore, the evidence must be organized by tasks, metric families, and methodologically comparable subgroups.

This challenge is particularly relevant to matching academic profiles with internship opportunities. While general recruitment platforms can leverage employment histories, applications, and recruiter decisions, student-oriented scenarios must represent courses, projects, developing skills, interests, and early-career goals. Studies addressing *freshers*, higher

education, and early career indicate that this subdomain remains less consolidated than matching in general employment platforms [23, 34, 37, 39, 46].

This article aims to systematically identify, analyze, and synthesize Artificial Intelligence (AI) and semantic modeling approaches used to match academic/professional profiles with employment, internship, and career opportunities. The scope is restricted to studies that explicitly address matching, recommendation, ranking, alignment, or fit prediction tasks, distinguishing this focus from broader reviews of AI applied to recruitment automation.

Based on this scope, the main contributions of this review are: (i) to systematically map AI and semantic modeling approaches applied to profile–opportunity matching and classify the studies by task, methodological family, dataset, strength of the *ground truth*, and metric family; (ii) to assess the methodological quality of the studies and their eligibility for quantitative synthesis; (iii) to synthesize the quantitative evidence within comparable subgroups, separating classification-based fit prediction studies from Top-K recommendation/ranking studies, and to complement this analysis with a narrative synthesis of the main methodological families; and (iv) to identify limitations, research gaps, and future opportunities, with an emphasis on matching academic profiles with internship opportunities.

2 POSITIONING RELATIVE TO PREVIOUS REVIEWS

Related reviews organize the *e-recruitment* literature around challenges, classify job recommender systems (JRSs) by applications, methods, and metrics, analyze traditional and emerging techniques across hiring stages, or synthesize the use of AI in broader human resource decision-making [3, 4, 10, 28]. Table 1 summarizes the differences in scope. This review differs by requiring explicit matching between a profile and an opportunity and by assessing, for each study, the task, data, strength of the *ground truth*, metrics, quality, and eligibility for quantitative synthesis. Results are compared only within compatible subgroups, without inferring superiority across distinct protocols. This topic also aligns with WebMedia: a recent study combines semantic content, user–item interactions, and graph-based recommendation, showing that strong predictive performance does not necessarily imply equivalent ranking performance [12].

3 METHODOLOGY

This systematic review was conducted based on a protocol defined in Parsifal [31] and guided by the PRISMA recommendations [30]. The methodological objective was to identify, select, assess, and synthesize primary studies that apply Artificial Intelligence (AI), Natural Language Processing (NLP), ontologies, *knowledge graphs*, embeddings, *recommender systems*, and other semantic modeling approaches to the problem of matching academic/professional profiles with employment, internship, or career opportunities.

The methodological process comprised the following stages: definition of the scope and research questions, formulation of the PICOC framework, development of the search strategies, study selection, methodological quality assessment, data extraction and normalization, and definition of the synthesis strategy. The counts and percentages reported in the results sections were calculated based on the final set of 43 studies retained in the main synthesis, excluding studies removed after the quality assessment. Therefore, summary tables generated during data extraction were not used directly without first filtering the final corpus.

3.1 Review Protocol and Scope

The scope of the review was limited to the problem of *profile–opportunity matching*. We considered studies that propose or evaluate methods for *resume–job matching*, *candidate–job matching*, *candidate ranking*, *person–job fit*, *student–job recommendation*, *internship recommendation*, *skill–job alignment*, and *fit prediction*.

The review focused on methods that perform matching, recommendation, ranking, alignment, or fit prediction between profiles and opportunities. Therefore, studies focused exclusively on recruitment automation, such as *resume parsing*, *resume screening*, applicant tracking systems (ATSs), interviews, or job-description generation, were excluded when they did not present explicit profile–opportunity matching mechanisms. Recommender systems in unrelated domains, such as courses, movies, products, news, or e-commerce, were also excluded, except when they were directly associated with academic or professional opportunities.

3.2 Research Questions and PICOC

The review was guided by six research questions formulated to connect the study scope, data extraction, subgroup-based quantitative synthesis, and discussion of methodological limitations:

- **RQ1 – Techniques:** Which AI, NLP, and semantic modeling techniques have been used to match academic/professional profiles with opportunities?
- **RQ2 – Matching tasks:** Which matching tasks are addressed, including recommendation, ranking, resume–job classification, *person–job fit*, *fit prediction*, and student–job recommendation?
- **RQ3 – Datasets and ground truth:** Which datasets, matching entities, and sources of *ground truth* are used?
- **RQ4 – Evaluation metrics:** Which metrics are used to evaluate classification, ranking, recommendation, and *fit prediction*?
- **RQ5 – Quantitative evidence:** Which methodological families report the best quantitative results within comparable subgroups of *profile–opportunity matching*?
- **RQ6 – Comparability limitations:** Which methodological limitations affect study comparability and the feasibility of quantitative synthesis?

Table 1: Positioning of this review relative to related reviews.

Review	Scope	Emphasis	Difference from this study
Mashayekhi et al. [28]	Since 2012 (123 studies)	Eight specific <i>e-recruitment</i> challenges	Challenge-oriented synthesis; does not classify the strength of the <i>ground truth</i> or eligibility for quantitative synthesis.
Çelik Ertuğrul and Bitirim [10]	2010–Apr. 2023 (57 studies)	Taxonomy of JRSs, applications, methods, and metrics	Includes applications beyond explicit matching and does not stratify results according to protocol comparability.
Aminou et al. [4]	2017–2024 (40 studies)	Traditional and emerging techniques across hiring stages	Includes screening, assessment, selection, and final decision-making; does not require profile–opportunity matching.
Amanullah et al. [3]	2023–2025 (25 studies)	AI for decision-making in recruitment, performance, and retention	Adopts a broad human resource scope; does not detail entities, <i>ground truth</i> , and metrics by matching task.
This study	2020–Jun. 2026 (43 studies)	AI and semantic modeling for explicit matching	Study-level quality and eligibility; synthesis by task and metric; focus on academic profile–internship matching.

The PICOC framework guided the construction of the search strings and the definition of the review scope [22], as detailed below:

- **Population:** students, candidates, professionals, resumes, CVs, and academic/professional profiles.
- **Intervention:** AI, ML, DL, NLP, embeddings, ontologies, *knowledge graphs*, Transformers, GNNs, LLMs, *semantic similarity*, and *recommender systems*.
- **Comparison:** traditional recruitment methods, manual screening, *keyword-based matching*, TF-IDF, rule-based methods, and classical baselines.
- **Outcome:** *profile–opportunity matching*, *resume–job matching*, recommendation, ranking, *skill–job alignment*, *person–job fit*, and *fit prediction*.
- **Context:** online recruitment, *e-recruitment*, career platforms, internship platforms, higher education, the labor market, and *talent acquisition*.

3.3 Search Strategy and Study Selection

The searches were conducted in the ACM Digital Library, IEEE Xplore, ScienceDirect, and Scopus. The main search string combined three conceptual blocks: terms related to job matching and recommendation; terms associated with AI, ML, DL, NLP, and recommender systems; and terms related to semantic modeling, including ontologies, embeddings, *knowledge graphs*, semantic similarity, and text mining. Table 2 presents the general structure of the search string used as the basis for adaptation to each digital library.

The queries were adapted to the syntax and available resources of each database. Eight queries were executed, with combinations varying among matching terms, semantic representation or textual extraction, NLP techniques, and evaluation terms, while preserving the profile–opportunity matching scope. We considered primary studies published from 2020 onward, written in English, and available in full text. Study selection followed the PRISMA recommendations [30], including duplicate removal, title and abstract screening, full-text assessment, and the final inclusion decision. Numerical details of the selection flow are presented in the results section. The final search update was conducted in June 2026.

Title and abstract screening, full-text reading of the studies, methodological quality assessment, and data extraction were performed by the first author. Because these stages were not conducted independently in duplicate, no inter-rater agreement measure was calculated, and no formal disagreement-resolution procedure was applied.

3.4 Inclusion and Exclusion Criteria

The inclusion and exclusion criteria were defined to ensure transparency, traceability, and reproducibility in the study-selection process. These criteria were applied during both title and abstract screening and full-text assessment to retain only studies directly related to the *profile–opportunity matching* problem. Table 3 presents the criteria used.

These criteria standardized study selection, reduced ambiguities during screening and full-text assessment, and made the methodological process more reproducible. However, the final inclusion decision was not determined exclusively by the quality score. It also considered adherence to the review scope, the centrality of the profile–opportunity matching task, and the presence of sufficient methodological evaluation to support the synthesis.

3.5 Methodological Quality Assessment

Methodological quality was assessed after full-text reading using five criteria: centrality of matching; clarity of the method; description of the data and *ground truth*; quality of the experimental evaluation; and appropriateness of the metrics. For each criterion, *Yes* (2.0) indicated complete and explicit fulfillment; *Mostly Yes* (1.5), fulfillment with a minor omission; *Partial* (1.0), incomplete but interpretable evidence; *Weak Partial* (0.5), only indirect or insufficient evidence; and *No* (0.0), absence or incompatibility. The total score (0–10) initially defined the categories *Strong Include* (8.5–10), *Include with limitations* (7.0–8.4), *Narrative only* (6.0–6.9), *Weak / Consider exclusion* (5.0–5.9), or *Exclude* (0–4.9).

The score supported but did not replace the decision. A critical scope mismatch, absence of explicit matching, uninterpretable data or labels, or an incompatible evaluation could lead to a study being downgraded or excluded regardless of

Table 2: Main search string used as the conceptual basis for the database-specific queries.

Conceptual block	Search terms
Matching and recommendation	"job recommendation" OR "job matching" OR "career recommendation" OR "internship recommendation" OR "resume matching" OR "candidate matching" OR "skill matching"
AI, ML, NLP, and recommender systems	"artificial intelligence" OR "machine learning" OR "deep learning" OR "recommender system" OR "natural language processing" OR NLP
Semantic modeling	"ontology" OR "knowledge graph" OR "semantic similarity" OR "semantic matching" OR embedding OR "text mining"

Boolean structure: (Matching and recommendation) AND (AI, ML, NLP, and recommender systems) AND (Semantic modeling).

Table 3: Inclusion and exclusion criteria used in the study-selection process.

ID	Type	Criterion
IC1	Inclusion	The study addresses jobs, careers, internships, recruitment, candidates, resumes, skills, or academic/professional profiles.
IC2	Inclusion	The study presents or supports a method for matching, recommendation, ranking, alignment, or fit assessment between profiles, candidates, resumes, or skills and jobs, internships, or careers.
IC3	Inclusion	The study uses AI, ML, NLP, recommender systems, semantic similarity, ontologies, <i>knowledge graphs</i> , embeddings, Transformers, fuzzy logic, or related techniques.
IC4	Inclusion	The study is a primary research contribution published as a journal article, conference paper, proceedings paper, or book chapter.
IC5	Inclusion	The study is written in English and its full text is available.
EC1	Exclusion	The study is not related to jobs, careers, internships, recruitment, candidates, resumes, skills, or academic/professional profiles.
EC2	Exclusion	The study does not present, evaluate, or support matching, recommendation, ranking, alignment, or fit assessment between profiles and opportunities.
EC3	Exclusion	The study focuses on unrelated domains, such as courses, movies, music, products, news, research papers, learning objects, or e-commerce.
EC4	Exclusion	The study uses AI, NLP, ontologies, <i>knowledge graphs</i> , embeddings, or graph-based methods without applying them to profile–opportunity matching.
EC5	Exclusion	The study focuses exclusively on resume screening, candidate scoring, resume classification, or resume ranking without matching a profile with an opportunity.
EC6	Exclusion	The study only describes datasets, recruitment platforms, ATSs, hiring systems, or filtering systems without computational matching.
EC7	Exclusion	The study focuses on bias, fairness, ethics, explainability, visualization, or auditing without evaluating matching methods.
EC8	Exclusion	The study focuses exclusively on interviews, job-description generation, engagement, scheduling, communication, or offer management.
EC9	Exclusion	The study is a survey, mapping study, systematic review, or literature review that does not present a primary matching method.
EC10	Exclusion	The document is not a primary research study, such as an editorial, presentation, poster, abstract, table of contents, or non-scientific document.
EC11	Exclusion	The study is not written in English or its full text is unavailable.
EC12	Exclusion	The study was published before 2020 and is therefore outside the temporal scope of this review, which considers studies published from 2020 onward.

its total score. Overall quality and eligibility for quantitative synthesis were recorded separately, with a documented justification at the study level during data extraction.

3.6 Data Extraction and Normalization

Data extraction was conducted in a structured manner and covered study identification, application context, matching entities, computational method, semantic technique or representation, dataset, source of *ground truth*, experimental protocol, baselines, metrics, reported quantitative results, contributions, and limitations.

Derived fields were also defined to support the comparative synthesis, including task type, methodological family, strength of the *ground truth*, metric family, quantitative synthesis group, and eligibility for quantitative synthesis. These fields were used to distinguish different levels of analysis, avoiding the combination of methodological family, specific technique, and proposed model or architecture within the same category. For example, *Transformer*, BERT, GNN, ontology, and recommender system were treated as techniques or methodological families, whereas the names of the specific models proposed in the articles were recorded separately.

The extracted data were normalized by standardizing metrics, grouping techniques into methodological families, and

classifying datasets according to their origin, availability, and documentation. To preserve the numerical consistency of the review, statistics concerning the included studies were calculated only after filtering the final corpus of 43 studies retained in the main synthesis, without including studies excluded after the quality assessment.

3.7 Data Synthesis and Eligibility for Quantitative Synthesis

Data synthesis combined descriptive analysis, narrative synthesis, and subgroup-based quantitative synthesis. Descriptive analysis was used to characterize the corpus in terms of publication year, task type, methodological family, datasets, sources of *ground truth*, and reported metrics. Narrative synthesis was used to interpret methodological patterns and research gaps. Quantitative synthesis, in turn, was restricted to studies with greater methodological and metric comparability.

To avoid ambiguity regarding the term “eligible study”, eligibility for quantitative synthesis was operationalized according to five dimensions: quality score, clarity of the *ground truth*, dataset description, reported metrics, and task compatibility with the analytical subgroups. In this study, the *ground truth* was considered clear when the source of labels, relevance

judgments, or interaction signals was explicitly described, including, for example, application histories, clicks, interviews, recruiter decisions, hires, human annotations, or benchmark labels. Private datasets could also be considered eligible, provided that they offered sufficient documentation regarding entities, size, label construction, experimental protocol, and metrics.

Eligibility for quantitative synthesis was not determined solely by the quality score. Studies with good methodological scores could be classified as *Partially eligible* or *Not eligible* when they presented critical limitations related to the *ground truth*, dataset documentation, metric extraction, or task compatibility with the quantitative subgroups. Table 4 presents the criteria used to classify studies as *Eligible*, *Partially eligible*, or *Not eligible* for quantitative synthesis.

The decision rule was applied as follows: studies that met all essential criteria were classified as *Eligible*; studies with one or two partial limitations that were still relevant to quantitative interpretation were classified as *Partially eligible*; and studies with critical failures in methodological quality, *ground truth*, metrics, or task compatibility were classified as *Not eligible*. Thus, a study was not required to report both Accuracy and F1-score to be eligible, provided that it presented a minimum set of comparable metrics within the subgroup. Conversely, studies reporting only one isolated metric, without complementary metrics or sufficient protocol clarity, were treated as partially eligible or not eligible, depending on the severity of the limitation.

Because of the heterogeneity among tasks, datasets, experimental protocols, baselines, and metrics, a single aggregate statistical meta-analysis was not conducted. Instead, we adopted a *subgroup-based quantitative synthesis* strategy, primarily separating studies on *classification-based fit prediction* from studies on *Top-K recommendation/ranking*. This decision preserves the internal comparability of each subgroup and avoids combining metrics that measure different phenomena, such as Accuracy, F1-score, and AUC in classification tasks, or MAP, MRR, NDCG@K, Precision@K, and Recall@K in ranking and recommendation tasks.

To support transparency and consistency, the extracted data were organized in a structured extraction spreadsheet during the review process, including study identification, inclusion decision, quality criteria, final classification, task type, methodological family, dataset, source of *ground truth*, reported metrics, quantitative synthesis group, eligibility status, and justification for the final classification. The aggregate results derived from this extraction process are presented in the following sections.

4 RESULTS

This section presents the descriptive findings obtained after applying the review protocol. To avoid redundancy with the methodology, the results focus on characterizing the included corpus, matching tasks, methodological families, datasets, sources of *ground truth*, reported metrics, and eligibility for

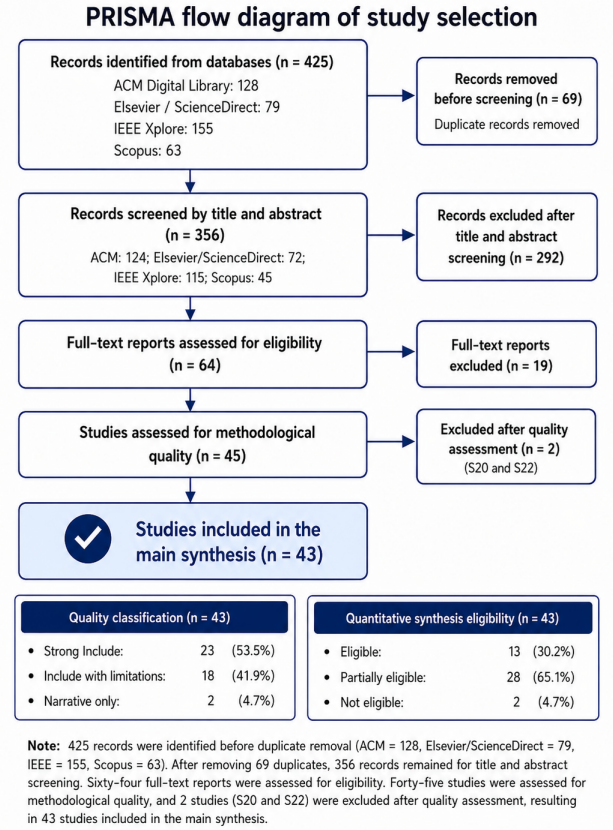


Figure 1: PRISMA flow diagram of the study-selection process.

quantitative synthesis. All statistics presented in this section consider only the 43 studies retained in the main synthesis.

4.1 PRISMA Flow and Study Selection

Figure 1 presents the PRISMA flow of the study-selection process. Initially, 425 records were identified across the four databases. After removing 69 duplicates, 356 records remained for title and abstract screening. At this stage, 292 records were excluded. After screening, 64 studies were selected for full-text assessment. Following full-text reading, application of the eligibility criteria, and methodological quality assessment, 43 studies were retained in the main synthesis.

4.2 Overview of the Included Studies

The temporal distribution of the included studies shows a strong concentration of publications in recent years, as presented in Table 5. Considering only the 43 studies retained in the main synthesis, 30 were published between 2025 and 2026, representing approximately 69.8% of the included corpus. This concentration suggests recent growth in the topic, possibly associated with advances in Transformers, LLMs, GNNs, embeddings, and recommender systems applied to digital recruitment.

Table 4: Eligibility criteria for quantitative synthesis.

Criterion	Eligible	Partially eligible	Not eligible
Quality score	Score ≥ 8.5 and no critical failures in the remaining criteria.	Score between 7.0 and 8.4, or score ≥ 8.5 with partial limitations concerning the dataset, <i>ground truth</i> , or metrics.	Score < 7.0 , exclusion due to methodological quality, or a critical failure in an essential criterion.
<i>Ground truth</i>	The source of labels, relevance judgments, or interaction signals is explicitly described and consistent with the matching task.	The source of labels is partially described, based on a proxy, or affected by transparency limitations.	The source of labels is absent, vague, or cannot be directly related to profile-opportunity matching.
Dataset	A public or private dataset sufficiently documented in terms of entities, size, labels, experimental protocol, and data split.	A private, restricted, or domain-specific dataset with a partial description that is sufficient for interpretation but limited for replication.	The dataset is absent, insufficiently described, or has no clear relationship with the matching evaluation.
Metrics	Metrics are extractable and comparable within the subgroup. For classification, at least two metrics among Accuracy, Precision, Recall, F1-score, or AUC. For ranking/recommendation, at least one position-sensitive metric, such as MAP, MRR, NDCG@K, Hit Rate@K, Precision@K, or Recall@K, with the value of K defined when applicable.	Metrics are only partially extractable, the metric set is incomplete, only one isolated metric is reported, or details regarding K, the protocol, or the calculation are absent.	Metrics are absent, not extractable, incompatible with matching, or restricted to non-comparable operational measures, such as latency or execution time.
Task and protocol	The task is compatible with one of the main subgroups: <i>classification-based fit prediction</i> or <i>Top-K recommendation/ranking</i> , with a sufficiently described experimental protocol.	The task is related to the review scope but has mixed objectives, an auxiliary evaluation, an incomplete protocol, or low direct comparability.	The task is incompatible with the quantitative subgroups or lacks an effective evaluation of matching, ranking, recommendation, or <i>fit prediction</i> .

Table 5: Included studies by publication year.

Year	Studies	Percentage
2020	4	9.3%
2021	1	2.3%
2022	2	4.7%
2023	3	7.0%
2024	3	7.0%
2025	23	53.5%
2026	7	16.3%
Total	43	100.0%

Regarding publication type, among the 43 studies included in the main synthesis, 34 were published in conference proceedings and 9 in journals, corresponding to 79.1% and 20.9% of the corpus, respectively. This predominance of conference publications indicates that the topic is still evolving rapidly and remains strongly associated with research communities in AI, Web, NLP, recommender systems, and intelligent recruitment.

4.3 Matching Tasks and Methodological Families

The included studies addressed different formulations of *profile-opportunity matching*. As shown in Table 6, the predominant tasks are resume-job classification (17; 39.5%), candidate ranking (9; 20.9%), and job recommendation (7; 16.3%). Student-job recommendation comprises 3 studies (7.0%), highlighting a gap in matching academic profiles with internship opportunities.

Regarding computational approaches, a distinction was made between *methodological family* and *technique*. In this study, a methodological family represents a broad category of approaches, such as Transformer-based methods, graph-based methods, structured-knowledge approaches, recommender systems, or classical machine learning. A technique, in turn, corresponds to a specific model, algorithm, resource, or representation used within a family, such as BERT, SBERT,

GCN, RGCN, ontologies, *knowledge graphs*, *dual-tower models*, *learning-to-rank*, SVM, or Random Forest.

Because many studies combine more than one approach, the classification was performed as a multi-label categorization. Therefore, the same study may appear in more than one methodological family, and the percentages presented in Table 7 do not sum to 100%. This organization makes it possible to identify the most recurrent families and techniques in the corpus without combining distinct conceptual levels.

Table 7 shows that the most recurrent families are Transformer-based methods, recommender systems, knowledge-based approaches, and graph-based methods. This pattern suggests that recent studies tend to combine textual semantics, structured knowledge, graph relationships, and recommendation logic, moving away from purely lexical approaches or isolated classification models. Classical ML/NLP and fuzzy or multicriteria-scoring approaches also appear as complementary components, particularly in hybrid pipelines that combine semantic representations with traditional supervised models or rule-based scoring.

4.4 Quality Assessment and Eligibility for Quantitative Synthesis

The quality assessment indicated that most of the included studies had substantial adherence to the scope of this review. Considering only the 43 studies retained in the main synthesis, the mean *Quality Score* was 8.5, with values ranging from 7.0 to 10.0. In total, 31 studies achieved a score of at least 8.0, and 25 studies achieved a score of at least 8.5. However, as defined in the methodology, the final quality classification was determined not only by the numerical score but also by methodological clarity, the centrality of the matching task, the quality of the evaluation, and the comparability of the results.

Table 8 presents the distribution of studies by quality classification and eligibility for quantitative synthesis. Among the

Table 6: Distribution and traceability of studies by matching task.

Matching task	Studies (ID and reference)	<i>n</i> (%)
Resume–job classification	S06 [14]; S07 [9]; S08 [7]; S14 [5]; S16 [8]; S17 [13]; S21 [43]; S23 [24]; S26 [2]; S27 [29]; S28 [6]; S29 [51]; S32 [38]; S40 [35]; S41 [1]; S44 [47]; S45 [27]	17 (39.5%)
Candidate ranking	S02 [42]; S03 [16]; S04 [33]; S05 [45]; S11 [19]; S15 [40]; S18 [32]; S31 [21]; S34 [44]	9 (20.9%)
Job recommendation	S10 [15]; S12 [36]; S24 [41]; S33 [49]; S35 [39]; S39 [37]; S42 [18]	7 (16.3%)
Fit prediction	S01 [11]; S09 [25]; S13 [26]; S19 [20]	4 (9.3%)
Student–job recommendation	S25 [34]; S36 [23]; S37 [46]	3 (7.0%)
Employee–position matching	S30 [17]; S38 [50]	2 (4.7%)
Semantic expert search	S43 [48]	1 (2.3%)
Total	–	43 (100.0%)

Note. The study IDs retain the original extraction identifiers and are therefore nonconsecutive.

Table 7: Methodological families and techniques identified in the included studies.

Methodological family	Identified techniques or resources	Studies	%	Study IDs
Transformer-based	BERT, SBERT, RoBERTa, Sentence-BERT, sentence-pair BERT, <i>dual encoders</i> , <i>cross-encoders</i> , contextual embeddings	27	62.8%	S01, S02, S03, S05, S06, S07, S09, S10, S11, S14, S15, S16, S18, S21, S24, S27, S28, S30, S32, S33, S34, S35, S36, S38, S40, S41, S45
Recommender systems	<i>dual-tower models</i> , <i>learning-to-rank</i> , matrix/factorization-based models, collaborative filtering, Top-K recommendation, reranking	24	55.8%	S02, S03, S04, S07, S10, S11, S12, S13, S15, S19, S21, S24, S29, S30, S32, S34, S37, S39, S40, S41, S42, S43, S44, S45
Knowledge-based	<i>knowledge graphs</i> , ontologies, skill taxonomies, ESCO, O*NET, TransE, structured knowledge	22	51.2%	S01, S02, S05, S06, S09, S10, S11, S13, S14, S17, S18, S26, S29, S30, S33, S35, S36, S37, S38, S39, S41, S43
Graph-based	GNN, GCN, RGCN, heterogeneous graphs, recruitment graphs, candidate–job interaction graphs	14	32.6%	S01, S05, S06, S07, S08, S09, S13, S29, S33, S35, S36, S37, S38, S44
LLM-based	LLMs, RAG, text generation or enrichment, explanatory feedback, conversational or recommendation-support components	5	11.6%	S08, S15, S21, S28, S35
Fuzzy / Multicriteria scoring	Fuzzy logic, AHP, hybrid rules, multidimensional scoring, semantic weights, and ensemble-based supervised scoring	3	7.0%	S14, S28, S31
Reinforcement learning	Reinforcement learning applied to recommendation, ranking, or dynamic adaptation of results	2	4.7%	S30, S33
Classical ML/NLP	SVM, Random Forest, XGBoost, TF-IDF, cosine similarity, classical supervised models, and traditional NLP pipelines	7	16.3%	S17, S23, S24, S25, S26, S31, S39
Semantic search	Semantic search, similarity-based retrieval, and <i>expert search</i>	1	2.3%	S43

Note. The study IDs retain the original extraction identifiers and are therefore nonconsecutive. The categories are multi-label and reflect relevant methodological components identified in each study; therefore, the percentages do not sum to 100%.

Table 8: Quality classification and eligibility for quantitative synthesis among the included studies.

Dimension	Category	Studies	%
Quality classification	<i>Strong Include</i>	23	53.5%
Quality classification	<i>Include with limitations</i>	18	41.9%
Quality classification	<i>Narrative only</i>	2	4.7%
Quantitative eligibility	<i>Eligible</i>	13	30.2%
Quantitative eligibility	<i>Partially eligible</i>	28	65.1%
Quantitative eligibility	<i>Not eligible</i>	2	4.7%

43 included studies, 23 were classified as *Strong Include*, 18 as *Include with limitations*, and 2 as *Narrative only*. Regarding quantitative eligibility, 13 studies were classified as *Eligible*, 28 as *Partially eligible*, and 2 as *Not eligible*.

The main quantitative groups identified were *classification-based fit prediction* and *Top-K recommendation/ranking*, each comprising 19 studies. In addition to these two main groups,

5 studies were treated separately because they involved semantic-search evaluation, human relevance assessment, or low quantitative comparability. These findings reinforce the methodological decision to adopt subgroup-based quantitative synthesis rather than a single statistical meta-analysis.

5 QUANTITATIVE SYNTHESIS

The quantitative synthesis was conducted as a subgroup-based comparison restricted to the 13 studies classified as *Eligible*. These studies presented greater methodological clarity, extractable metrics, and greater compatibility among the task, experimental protocol, and metric family. A single aggregate meta-analysis was not conducted because the studies differ substantially in task formulation, datasets, *ground truth* construction, baselines, validation protocols, and reported metrics. Therefore, classification metrics, such as Accuracy, Precision, Recall, F1-score, and AUC, were analyzed separately from ranking and recommendation metrics, such as

MAP, MRR, NDCG@K, Precision@K, Recall@K, and Hit Rate@K.

5.1 Comparison of Eligible Studies by Metric Family

The comparison of the 13 eligible studies was organized by metric family. Classification metrics, such as AUC, Accuracy, Precision, Recall, and F1-score, assess the model's ability to predict classes or binary fit between a profile and an opportunity. Ranking and recommendation metrics, such as MAP, MRR, NDCG@K, Precision@K, Recall@K, and Hit Rate@K, assess the quality of the ordering of recommended candidates or opportunities.

Because some studies report more than one metric family, the same study may appear in more than one table. Therefore, Tables 9 and 10 should be interpreted as syntheses by metric family rather than as mutually exclusive groups of studies. This separation avoids combining classification and ranking metrics within a single textual column and directly connects the results to the research questions concerning datasets, *ground truth*, methodological families, and quantitative performance.

Tables 9 and 10 show that the eligible studies are not directly comparable through a single aggregate metric. Within the classification metric family, S42 [18] reported the highest AUC, at 0.9718, followed by S03 [16], S11 [19], and S04 [33]. In terms of Accuracy, S03 [16] reported the highest value, at 0.9290, whereas S42 [18] reported the highest F1-score, at 0.9184. Among the studies most directly aligned with *classification-based fit prediction*, S44 [47] reported an F1-score of 0.8103, while S07 [9] was the study with the most complete set of classical classification metrics, reporting AUC, Accuracy, Precision, Recall, and F1-score.

Within the ranking and recommendation metric family, S15 [40] reported the highest NDCG@10, at 0.8800, and the highest MRR@10, at 0.8400. S05 [45] also presented strong ranking performance, with $MRR = 0.8310$, $NDCG@10 = 0.7820$, $P@5 = 0.8560$, and $R@10 = 0.7410$. S33 [49] reported a more complete set of Top-K metrics on a public dataset, including MAP, P@10, R@10, and NDCG@10, whereas S42 [18] combined classification and ranking metrics on a private platform dataset.

Despite these results, their interpretation must account for differences in datasets, sources of *ground truth*, experimental protocols, values of K, and relevance definitions. Therefore, the reported values were not combined into a single statistical estimate. Instead, they were used to support a quantitative synthesis by metric family and task subgroup, preserving internal comparability between classification metrics and ranking/recommendation metrics.

5.2 Partially Eligible Studies and Separate Evidence

In addition to the 13 eligible studies, 28 were classified as *Partially eligible*. They remained in the review because they provided useful quantitative or methodological evidence but

were analyzed separately because of one or more limitations: unclear *ground truth* or labels, incomplete validation protocols, private, specific, or poorly documented data, non-standardized metrics, poorly comparable baselines, platform-centered evaluations, limited samples, or low reproducibility. These limitations were not mutually exclusive.

These studies contributed to the narrative synthesis of trends, architectures, limitations, and gaps but were not combined with the 13 eligible studies in the main quantitative comparison. This separation preserves the validity of the synthesis and prevents values obtained using insufficiently documented data, labels, or protocols from being treated as directly comparable to the results of the eligible studies.

6 NARRATIVE SYNTHESIS

The narrative synthesis was organized by methodological family because the reviewed corpus is strongly hybrid. Many studies combine embeddings, Transformers, graphs, ontologies, *knowledge graphs*, taxonomies, and recommendation mechanisms within the same architecture. Therefore, the categories discussed in this section should be interpreted as overlapping analytical lenses rather than mutually exclusive classes.

6.1 Semantic, Transformer-Based, and Embedding-Based Matching

A clear trend in the corpus is the transition from lexical overlap to semantic representation. TF-IDF and *cosine similarity* still appear as baselines, initial filters, or interpretable components of hybrid systems [2, 21, 24, 34, 37]. However, recent studies increasingly rely on contextual embeddings, BERT, SBERT, RoBERTa, Siamese architectures, *dual-tower encoders*, and *cross-encoders* to represent resumes, job descriptions, academic profiles, and career trajectories [5, 8, 15, 19, 27, 29, 32, 35, 38, 40, 42, 44, 45].

This shift is important because profile-opportunity matching cannot be reduced to keyword matching. Equivalent skills and experiences may be expressed using different vocabularies in resumes, academic records, and job descriptions. Studies such as S15 and S34 illustrate this transition by using sentence-level representations, contrastive fine-tuning, vector retrieval, and Siamese architectures to capture semantic equivalence between profiles and opportunities [40, 44]. LLM-based approaches also appear in the corpus, but mainly as complementary components for feedback, explanation, text enrichment, or recommendation support rather than as the primary model evaluated quantitatively [6, 7, 39, 40, 43].

6.2 Knowledge-Based, Graph-Based, and History-Based Matching

Knowledge-based and graph-based approaches extend textual matching by representing relationships among skills, occupations, experiences, courses, tools, job requirements, candidates, and platform interactions. Ontologies, *knowledge graphs*, skill taxonomies, O*NET, ESCO, and graph embeddings are used to structure semantic relationships and reduce

Table 9: Classification metrics reported by the eligible studies.

ID / Ref.	Fam.	Specific model	Dataset	Ground truth	AUC	Acc.	Prec.	Rec.	F1	
S01 [11]	T+G	CSAGNN / WHIN + BERT	Private dataset	platform	Platform interactions + behavioral signals	0.7227	0.7237	–	–	0.6958
S03 [16]	RS	PJFFF	Ping An HR-X		Recruitment records + fit labels	0.9530	0.9290	–	–	0.7710
S04 [33]	RS	TAPJFNN / TAPJFNN-K	Baidu recruitment dataset		Recruitment history + fit assessment	0.9300	0.8508	–	–	0.8454
S07 [9]	T+G	MV-CoN	BOSS Zhipin		Resume-job interaction data	0.7370	0.6780	0.6830	0.6490	0.6680
S11 [19]	T	SPBERT	Jobindex dataset		Candidate-job pairs + offline/online evaluation	0.9460	–	–	–	–
S15 [40]	T+RS	SBERT + contrastive fine-tuning + reranking	University placement dataset		Academic relevance + matching assessment	–	0.8600	–	–	0.8000
S33 [49]	G+RS	GCN + RL + DCF	XING Challenge Dataset		Public recommendation interactions	–	–	–	–	0.7210
S34 [44]	T	SimCSE RoBERTa + Siamese	HF resume-job pairs		Labeled resume-job pairs	–	0.7200	0.7000	0.7400	0.7200
S42 [18]	RS	Segment-Aware LorentzFM	Kariyer.net		Real-world job-platform interactions	0.9718	–	0.9045	0.9327	0.9184
S44 [47]	G+RS	PJFCANN	Online recruitment dataset		Recruitment history + person-job fit labels	–	0.8076	0.7994	0.8215	0.8103

Note. Fam. = methodological family; T = Transformer-based; G = graph-based; RS = recommender system; HF = HuggingFace. Dashes indicate metrics that were not reported or could not be extracted for the selected model.

Table 10: Ranking and recommendation metrics reported by the eligible studies.

ID	Ref.	Specific model	Dataset	Ground truth	MAP	MRR	NDCG	P@K	R@K	HR@10	
S05	[45]	KG-enhanced BERT	Dual-Tower	Manufacturing companies	Candidate-job relevance derived from company data	–	0.8310	0.7820@10	0.8560@5	0.7410@10	–
S10	[15]	JobFormer	Private data + benchmark		Applications, expert skill assessments, and clicks	–	0.2982	0.3430@20	–	0.6998@20	–
S11	[19]	SPBERT	Jobindex dataset		Candidate-job pairs and offline/online evaluation	–	–	0.1860@20	–	–	–
S12	[36]	MBJ-DA	GLIT job-search dataset		Multiple search and application behaviors	–	0.4014	0.4208@5	–	–	–
						–	–	0.4570@10	–	–	–
						–	–	0.4866@20	–	–	–
S15	[40]	SBERT + contrastive fine-tuning + reranking	University placement dataset		Academic relevance and matching assessment	–	0.8400@10	0.8800@10	–	–	–
S33	[49]	GCN + RL + DCF	XING Challenge Dataset		Public recommendation interactions	0.4360	–	0.3530@10	0.3870@10	0.2890@10	–
S34	[44]	SimCSE RoBERTa + Siamese	HuggingFace resume-job pairs		Labeled resume-job pairs	–	–	–	0.7500@10	–	–
S42	[18]	Segment-Aware LorentzFM	Kariyer.net		Real-world job-platform interactions	–	0.5000	0.5010@10	–	–	0.5087

reliance on superficial textual similarity [1, 23, 37, 39, 45, 46, 48, 50]. These methods are particularly relevant to matching academic profiles with internships because they can connect courses, projects, developing skills, and job requirements in a more explainable manner.

Graph-based and history-based models also exploit behavioral and relational signals that cannot be captured through text alone. These signals include candidate–job interactions, applications, clicks, interviews, recruiter decisions, and platform histories [9, 11, 17, 26, 47, 49, 51]. Among the eligible studies, S01, S07, and S44 show how graph neural networks, attention mechanisms, *co-teaching*, and recruitment histories can support *person–job fit* or resume–job classification [9, 11, 47]. This family has strong practical relevance but also reinforces a central limitation of the field: many high-performing approaches depend on private or industrial datasets, reducing reproducibility and external validation.

6.3 Recommender Systems and Student-Oriented Matching

Recommender systems are central to tasks in which the objective is to rank candidates or opportunities rather than merely classify profile–job pairs. Within this family, metrics such as MAP, MRR, NDCG@K, Precision@K, Recall@K, and Hit Rate@K are more appropriate than binary classification metrics [14, 15, 18, 19, 36, 40, 45, 49]. The reviewed studies include *dual-tower models*, collaborative filtering, matrix- or factorization-based models, *learning-to-rank*, reranking, graph-enhanced recommendation, and reinforcement learning [15, 18, 20, 36, 40, 45, 49].

The subgroup focusing on students, *freshers*, internships, and higher education is smaller but central to the motivation of this review [13, 23, 34, 37, 39–41, 46]. These studies suggest that academic profile–internship matching requires models capable of representing incomplete career trajectories, academic signals, developing skills, projects, courses, and early-career interests. S15 is particularly transferable to academic profile–job matching because it combines SBERT,

vector retrieval, and reranking in an academic-profile context [40]. Other studies indicate the potential of integrating *knowledge graphs*, ontologies, embeddings, skill-gap analysis, and personalized recommendation for higher education and early-career scenarios [13, 23, 37, 39, 41, 46].

Overall, the narrative synthesis shows that the field has moved beyond purely lexical matching toward hybrid architectures that combine semantic representation, structured knowledge, graph relationships, behavioral signals, and recommendation mechanisms. However, the evidence is not equally mature across all subdomains. Recommendation/ranking and graph/history-based approaches show greater quantitative maturity, whereas student-oriented and internship-focused matching remains less consolidated and offers an important opportunity for future research.

7 DISCUSSION

This section interprets the review findings in light of the six research questions, connecting the descriptive results, subgroup-based quantitative synthesis, and narrative synthesis. Rather than repeating the data presented in the previous sections, the discussion emphasizes what the observed patterns indicate about the maturity of the field, the strength of the available evidence, recurrent methodological limitations, and implications for matching academic profiles with internship opportunities.

7.1 Answers to the Research Questions

Regarding RQ1 and RQ2, the reviewed studies show that *profile–opportunity matching* has been addressed mainly through NLP, ML, DL, BERT/SBERT, Transformers, *recommender systems*, *knowledge graphs*, ontologies, GNNs, LLMs, and hybrid methods [1, 15, 19, 23, 37, 39, 40, 44, 45, 48]. The most recurrent tasks were *resume–job classification*, *candidate ranking*, *job recommendation*, and *fit prediction*, whereas studies focusing on students, internships, *freshers*, and higher education appeared less frequently [9, 11, 16, 18, 23, 34, 36, 39, 40, 45–47]. This indicates that the literature is still more mature in general corporate recruitment than in matching academic profiles with internships.

Regarding RQ3 and RQ4, most studies rely on private or industrial datasets, frequently based on applications, clicks, recruiter decisions, interviews, or platform interactions [9, 11, 15, 16, 18, 19, 33, 36, 45, 47]. These sources increase ecological validity but limit reproducibility and external comparison. The reported metrics also differ according to task formulation: classification studies use Accuracy, Precision, Recall, F1-score, and AUC, whereas ranking and recommendation studies use MAP, MRR, NDCG@K, Precision@K, Recall@K, and Hit Rate@K [9, 11, 18, 19, 36, 40, 44, 45, 47, 49]. Therefore, these metric families must be interpreted separately.

Regarding RQ5 and RQ6, the most comparable quantitative evidence is concentrated in the *classification-based fit prediction* and *Top-K recommendation/ranking* subgroups. Among the eligible studies, the recommender-system family, represented by S42 [18], reported the highest AUC (0.9718)

and F1-score (0.9184) values in the classification subgroup. In the recommendation/ranking subgroup, the combination of Transformers and recommender systems, represented by S15 [40], reported the highest NDCG@10 (0.8800) and MRR@10 (0.8400) values. However, these results do not establish the overall superiority of any methodological family because their comparability is affected by task heterogeneity, private datasets, unclear *ground truth*, different baselines and validation protocols, incompatible metrics, and variation in the values of K [1, 9, 11, 18, 26, 37, 40, 41, 45, 47, 49, 51]. These factors justify the decision not to conduct a single aggregate statistical meta-analysis and support the use of subgroup-based quantitative synthesis.

7.2 Methodological Interpretation of the Findings

The results indicate a transition from lexical similarity toward hybrid approaches that combine textual representation, structured knowledge, graphs, behavioral signals, and recommendation [9, 11, 15, 18, 40, 45, 47, 49]. However, differences in reported values do not result solely from the architecture; they also reflect the task, dataset, construction of the *ground truth*, baselines, validation protocol, metrics, and, for ranking tasks, the value of K. Therefore, a higher value under a specific protocol demonstrates performance in that context but does not establish overall superiority.

The methodological choice should therefore be aligned with the task type. For resume–job classification, supervised models evaluated using AUC, Accuracy, and F1-score may be appropriate. For ranking and recommendation, position-sensitive metrics such as NDCG@K, MRR, MAP, Precision@K, and Recall@K are more suitable. In academic and internship contexts, where professional experience may be limited, semantic and knowledge-based approaches are particularly relevant because they can represent courses, projects, developing skills, and training gaps.

7.3 Limitations of Datasets, Ground Truth, and Metrics

The main methodological limitations of the field concern private datasets, heterogeneous sources of *ground truth*, and non-comparable metric families. Many studies use application histories, clicks, recruiter decisions, interview records, or platform behavior as labels [9, 11, 15, 16, 18, 19, 33, 36, 47]. Although these signals are useful, they represent different notions of relevance: a click may indicate interest, an application may indicate preference, and a recruiter decision may reflect expert judgment but also organizational bias. Therefore, reported performance must be interpreted in relation to how each study defines and constructs its *ground truth*.

Beyond comparability, adoption involves *trade-offs*: lexical or rule-based methods are simpler and more interpretable but represent semantic equivalences with less flexibility; Transformers, graphs, and hybrid architectures expand this capacity but may increase computational cost, reduce interpretability, and reproduce biases present in the data and labels [1, 21, 25, 26, 37, 43]. Future evaluations should report

the task, data, *ground truth*, data split, baselines, metrics, and value of K, as well as efficiency, explainability, and fairness when relevant. This minimum set of information supports transparent and reproducible comparisons.

7.4 Implications for Academic Profile–Internship Matching

The findings have direct implications for matching academic profiles with internship opportunities. First, the small number of studies directly focusing on students, *freshers*, internships, and higher education indicates that this domain remains underexplored [13, 23, 34, 37, 39–41, 46]. Most of the literature has been developed around professional resumes, corporate platforms, and large-scale recruitment histories.

Second, the results suggest that a robust solution for academic profile–internship matching should not depend on a single technique. A promising model could combine NLP and embeddings to represent resumes, academic records, and job descriptions; ontologies or *knowledge graphs* to structure skills, fields of study, tools, and requirements; and fuzzy or hybrid mechanisms to handle partial degrees of fit. This combination is particularly appropriate for internships because compatibility is rarely binary: a student may partially meet the requirements of a position and still be a promising candidate.

Third, explainability should be treated as a central requirement. In academic contexts, a matching system should not only rank students or opportunities but also explain why a particular opportunity is recommended, which skills justify the fit, which gaps need to be addressed, and which courses, projects, or experiences support the compatibility score. Models based on ontologies, *knowledge graphs*, fuzzy logic, and structured semantic representations offer promising paths toward this objective [1, 21, 23, 37, 39, 48].

These implications indicate that academic profile–internship matching requires a specific methodological design. Unlike traditional professional recruitment, this domain must account for developing skills, incomplete academic trajectories, indirect evidence of competence, and early-career goals. Therefore, future models should be evaluated not only by aggregate performance but also by their explanatory capacity, usefulness to students, alignment with company needs, and support for decision-making in educational institutions.

7.5 Research Gaps

The review reveals six main gaps. The first is the scarcity of public datasets and shared benchmarks, which limits reproducibility and prevents fair comparisons among models. The second is the limited focus on internships and higher education, despite the social and economic relevance of this domain. The third is the low transparency of labels, particularly when clicks, applications, or recruiter decisions are used as proxies for compatibility.

The fourth gap is limited external validation. Few studies test their models across multiple datasets, sectors, institutions, or countries, which restricts generalization to universities, small businesses, internship programs, and regional labor markets. The fifth gap is the absence of systematic comparisons among semantic components. Many studies propose hybrid architectures but do not always isolate the contributions of embeddings, ontologies, graphs, Transformers, or recommendation modules through *ablation studies*.

The sixth gap is explainability. Although some studies address fairness, interpretability, or feedback, most still prioritize aggregate performance metrics [5, 6, 43]. In academic–professional matching systems, explanations are essential to support human decisions, guide skill development, and prevent recommendations from becoming black boxes.

Overall, the results suggest that AI and semantic modeling approaches are promising for *profile–opportunity matching*, but the field still needs to advance in reproducibility, *ground truth* transparency, external validation, metric comparability, and explainability. For academic profile–internship matching, these gaps represent not only limitations of the literature but also concrete opportunities for scientific and applied contributions.

8 THREATS TO VALIDITY

This review is subject to threats related to the search process, study selection, construct interpretation, data extraction, synthesis strategy, external validity, and publication bias. To reduce these risks, the protocol defined explicit inclusion and exclusion criteria, standardized extraction fields, methodological quality assessment, and a subgroup-based quantitative synthesis strategy. Descriptive statistics were calculated only for the final corpus of 43 studies retained in the main synthesis.

Search and selection validity. The search strategy may not have retrieved all relevant studies because similar problems are described using different terms, such as *job matching*, *person–job fit*, *candidate ranking*, *career recommendation*, *resume–job matching*, and *internship recommendation*. This risk was mitigated by using multiple digital libraries, adapting the search string to each database, employing broad synonyms, and applying explicit selection criteria. Selection bias may also arise in borderline studies that address recruitment or resume analysis without clearly presenting profile–opportunity matching. Title and abstract screening and full-text reading of the studies were performed by the first author. At both stages, the decision considered whether each study included an explicit matching, recommendation, ranking, alignment, or fit prediction task. Conducting the selection process through a single researcher, without an independent duplicate assessment, constitutes a limitation of the study.

Construct and data extraction validity. A specific threat concerns the definition of *profile–opportunity matching*, as the literature uses related terms such as *resume–job classification*,

candidate ranking, job recommendation, person–job fit, fit prediction, and skill–job alignment. To reduce this ambiguity, the review distinguished task type, methodological family, specific technique, dataset, source of *ground truth*, metric family, and quantitative eligibility. Data extraction may also be affected by incomplete reporting of datasets, labels, baselines, or metrics. The quality assessment and data extraction were performed by the first author using explicit criteria and a standardized form, with missing or ambiguous information recorded as *not reported, not clear, or not described*. Because these activities were not conducted independently in duplicate, no formal disagreement-resolution procedure or inter-rater agreement calculation was performed.

Synthesis and conclusion validity. The main threat to synthesis validity is the heterogeneity among tasks, datasets, sources of *ground truth*, baselines, validation protocols, and metric families. Combining classification metrics with ranking metrics could lead to artificial conclusions. For this reason, a single aggregate statistical meta-analysis was not conducted. Instead, the quantitative synthesis separated evidence from *classification-based fit prediction* and *Top-K recommendation/ranking*, while partially eligible studies were treated as complementary narrative evidence.

External validity and publication bias. The generalizability of the findings is limited by the frequent use of private, industrial, or platform-specific datasets. Although real-world interaction data increase ecological validity, they reduce reproducibility and hinder external validation. Therefore, results obtained on specific platforms should not be automatically generalized to universities, internship programs, small businesses, or regional labor markets. The review may also be affected by publication bias because studies reporting positive results, novel architectures, or substantial improvements tend to be published more frequently than studies reporting negative or inconclusive results.

9 CONCLUSION

This systematic review analyzed 43 studies on Artificial Intelligence and semantic modeling approaches applied to *profile–opportunity matching*. The results show that the field is driven primarily by NLP, ML, deep learning, Transformers, BERT/SBERT, LLMs, *recommender systems*, *knowledge graphs*, ontologies, GNNs, and hybrid approaches. The most frequent tasks were *resume–job classification*, *candidate ranking*, and *job recommendation*, whereas studies directly focusing on students, internships, *freshers*, and higher education still appear less frequently.

Because of the heterogeneity among task formulations, datasets, sources of *ground truth*, baselines, validation protocols, and metric families, conducting a single aggregate statistical meta-analysis was not methodologically appropriate. Instead, the review adopted a *subgroup-based quantitative synthesis*, separating evidence from *classification-based fit prediction* and *Top-K recommendation/ranking*, while treating partially eligible studies as complementary narrative evidence.

This strategy made it possible to organize the quantitative results while preserving internal comparability among studies with similar tasks and metrics.

The evidence indicates that hybrid semantic approaches are particularly promising for matching profiles with opportunities, especially when they combine textual representation, structured knowledge, graph-based relationships, interaction histories, and recommendation mechanisms. However, the field still faces relevant limitations related to the limited availability of public datasets, low transparency in *ground truth* construction, frequent use of private datasets, heterogeneous evaluation protocols, and metrics that are not always comparable.

For the domain of academic profile–internship matching, the findings suggest a relevant opportunity for research and application. Unlike traditional professional recruitment, this context requires models capable of representing developing skills, academic trajectories, completed courses, projects, early-career interests, and partial degrees of fit between students and opportunities. Therefore, future research should prioritize reproducible datasets, transparent protocols, external validation, explainability, and the integration of NLP, ontologies, *knowledge graphs*, fuzzy logic, and recommender systems.

Overall, this review contributes by organizing the recent literature on *profile–opportunity matching*, distinguishing methodological families and matching tasks, assessing study quality and comparability, and synthesizing quantitative evidence by subgroup. The results reinforce that Artificial Intelligence and semantic modeling approaches have the potential to improve matching between academic/professional profiles and opportunities, but advances in reproducibility, data transparency, and evaluation standardization are still needed to consolidate the field.

10 ETHICAL CONSIDERATIONS

This study is a systematic literature review based exclusively on previously published scientific articles and extracted bibliographic and methodological metadata. No human participants were recruited, no primary personal data were collected, and no personally identifiable information was processed by the authors.

Therefore, the study does not involve direct research with human participants and follows the ethical and data-protection requirements applicable to the secondary analysis of scientific literature, including Brazilian National Health Council Resolution No. 510/2016 and the Brazilian General Data Protection Law (LGPD).

ACKNOWLEDGMENTS

Generative AI assistance was used to support manuscript organization, textual clarity, and the synthesis of the extracted evidence. All methodological decisions, study selection, data extraction, interpretation of the results, and final responsibility remained with the authors.

REFERENCES

- [1] Rubén Alonso, Danilo Dessì, Antonello Meloni, and Diego Reforgiato Recupero. 2025. A novel approach for job matching and skill recommendation using transformers and the O*NET database. *Big Data Research* 39 (2025), 100509. doi:10.1016/j.bdr.2025.100509
- [2] Ahmad Alsharef, Sonia, Hasan Nassour, and Jitender Sharma. 2023. Exploring the Efficiency of Text-Similarity Measures in Automated Resume Screening for Recruitment. In *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE, 36–42. <https://ieeexplore.ieee.org/document/10112537>
- [3] Md Amanullah, Mirza Fahim Ahmed, Khadiza Khatun Tumpa, Malina Khatun, Ripon Kumar Paul, and Prosenjit Tarafdar. 2026. Artificial Intelligence for Data-Driven Human Resource Decision-Making: A Systematic Review. *Social Sciences & Humanities Open* 13, Article 102909 (2026). doi:10.1016/j.ssaho.2026.102909
- [4] Loubna Aminou, Abdelaziz Daaif, Maha Soulam, Abderrahim Chalfauat, and Mohamed Youssef. 2025. Are Cutting-Edge Technologies Transforming Classical Approaches in Automated Hiring Systems? A Systematic Literature Review. *Systems and Soft Computing* 7, Article 200421 (2025). doi:10.1016/j.sasc.2025.200421
- [5] P. Felix Antony, P. Eshanth, and S. Julie Sharine. 2026. An Intelligent Resume–Job Matching System: Bridging Candidates and Opportunities through Advanced Semantic Understanding. In *2026 9th International Conference on Computational Intelligence in Data Science (ICCIDS)*. 1–6. doi:10.1109/iccids69108.2026.11407881
- [6] Durga B. J. B. Simha, and Shinu Abhi. 2025. Hybrid Resume Fitment Scoring with LLMs and AHP. In *2025 4th International Conference on Applied Artificial Intelligence and Computing (ICAATIC)*. 45–50. doi:10.1109/icaaic64647.2025.11330300
- [7] Amin Baghbanzadeh and Dan Wu. 2025. Resume–Job Compatibility Scoring Using Graph Neural Networks and Large Language Models. In *Proceedings of the 13th International Conference on Information Technology: IoT and Smart City (ICIT '25)*. Association for Computing Machinery, New York, NY, USA, 182–187. doi:10.1145/3787330.3787359
- [8] Irene L. Balmes and Melvin Ballera. 2025. Assessing the performance of an Improved Recruitment System in Resume Extraction and Candidate Matching. In *2025 International Conference on Decision Aid Sciences and Applications (DASA)*. 747–752. doi:10.1109/dasa68193.2025.11498743
- [9] Shuqing Bian, Xu Chen, Wayne Xin Zhao, Kun Zhou, Yupeng Hou, Yang Song, Tao Zhang, and Ji-Rong Wen. 2020. Learning to Match Jobs with Resumes from Sparse Interaction Data using Multi-View Co-Teaching Network. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management (Virtual Event, Ireland) (CIKM '20)*. Association for Computing Machinery, New York, NY, USA, 65–74. doi:10.1145/3340531.3411929
- [10] Duygu Çelik Ertuğrul and Selin Bitirim. 2025. Job Recommender Systems: A Systematic Literature Review, Applications, Open Issues, and Challenges. *Journal of Big Data* 12, 1, Article 140 (2025). doi:10.1186/s40537-025-01173-y
- [11] Hao Chen, Lun Du, Yuxuan Lu, Qiang Fu, Xu Chen, Shi Han, Yanbin Kang, Guangming Lu, and Zi Li. 2024. Professional Network Matters: Connections Empower Person–Job Fit. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining (Merida, Mexico) (WSDM '24)*. Association for Computing Machinery, New York, NY, USA, 96–105. doi:10.1145/3616855.3635852
- [12] Eduardo Ferreira da Silva, Joel Pires, Denis Robson Dantas Boaventura, Mayki dos Santos Oliveira, and Frederico Araujo Durão. 2025. A Gated Review Attention Framework for Topics in Graph-Based Recommenders. In *Proceedings of the 31st Brazilian Symposium on Multimedia and the Web. Sociedade Brasileira de Computação, Porto Alegre, RS, Brasil*, 19–27. doi:10.5753/webmedia.2025.15515
- [13] Dhruv Prakash Daberao, Gitesh Dalal, and R. Sreemathy. 2025. ResumeInsight: An AI-Driven Framework for Semantic Resume–Job Matching and Skill-Gap Analysis. In *2025 Global Conference on Information Technology and Communication Networks, GITCON 2025*. Institute of Electrical and Electronics Engineers Inc., 1–6. doi:10.1109/gitcon65266.2025.11377312
- [14] Xiao Gu, Ling Jian, Chongzhi Rao, Zhaoxue Bu, and Xianggang Cheng. 2025. Together Is Better: Knowledge-aware Model with Resume Fusion for Online Job Recommendation. *ACM Trans. Knowl. Discov. Data* 19, 3, Article 70 (mar 2025), 15 pages. doi:10.1145/3716503
- [15] Zhihao Guan, Jia-Qi Yang, Yang Yang, Hengshu Zhu, Wenjie Li, and Hui Xiong. 2024. JobFormer: Skill-Aware Job Recommendation with Semantic-Enhanced Transformer. *ACM Trans. Knowl. Discov. Data* 19, 1, Article 18 (dec 2024), 20 pages. doi:10.1145/3701735
- [16] Junshu Jiang, Songyun Ye, Wei Wang, Jingran Xu, and Xiaosheng Luo. 2020. Learning Effective Representations for Person–Job Fit by Feature Fusion. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management (Virtual Event, Ireland) (CIKM '20)*. Association for Computing Machinery, New York, NY, USA, 2549–2556. doi:10.1145/3340531.3412717
- [17] Zhang Jianguo, Shen Yancen, and Li Guimei. 2026. Research on optimization model of person-post matching and intelligent recommendation system based on multi-feature deep learning. *IET Conference Proceedings* 2025, 47 (mar 2026), 321–326. doi:10.1049/icp.2026.0202
- [18] Kübra Karacan Uyar and Yücel Batu Salman. 2026. Segment-Aware LorentzFM: Hyperbolic embedding approach for personalized job recommendations. *Array* 30 (2026), 100740. doi:10.1016/j.array.2026.100740
- [19] Mesut Kaya and Toine Bogers. 2026. Analyzing the Effects of a Human-in-the-Loop Candidate Recommendation Algorithm at Jobindex. *ACM Trans. Recomm. Syst.* (mar 2026). doi:10.1145/3801092
- [20] Aparup Khatua and Wolfgang Nejdl. 2020. Matching recruiters and jobseekers on Twitter. In *Proceedings of the 12th IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (Virtual Event, Netherlands) (ASONAM '20)*. IEEE Press, 266–269. doi:10.1109/asonam49781.2020.9381392
- [21] V. Kiruthiga, Kalpana Deshmukh, V.S. Narayana Tinnaluri, M. Priyadharsini, Manu Vasudevan Unni, and S. Ramya. 2025. Enhancing Recruitment Efficiency in HRM through Intelligent Resume Screening and Job Matching Using Fuzzy Logic and Ensemble Learning. In *2025 IEEE 3rd Global Conference on Wireless Computing and Networking, GCWCN 2025*. Institute of Electrical and Electronics Engineers Inc., 1–6. doi:10.1109/gwcwn66157.2025.11448508
- [22] Barbara Kitchenham and Stuart Charters. 2007. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*. Technical Report EBSE-2007-01. Keele University and Durham University Joint Report.
- [23] Yifeng Li, Cong Wang, Yongsheng Zhao, Fengling Sun, Tiancheng Wang, and Wenjing Li. 2026. Optimization Method for University Talent Skill–Job Matching Based on ARG Multi-Dimensional Data and Knowledge Graph Embedding. In *Proceedings of 2026 3rd International Conference on Informatics Education and Computer Technology Applications, IECA 2026*. Association for Computing Machinery, Inc, 143–148. doi:10.1145/3802133.3802158
- [24] Xiaohui Liang. 2025. Research on Resume Automatic Analysis and Matching Technology Based on Machine Learning and Natural Language Processing. In *2025 International Conference on Digital Analysis and Processing, Intelligent Computation (DAPIC)*. 261–265. doi:10.1109/dapic66097.2025.00054
- [25] Ping Liu, Rajat Arora, Xiao Shi, Benjamin Hoan Le, Qianqi Shen, Jianqiang Shen, Chengming Jiang, Nikita Zhiltsov, Priya Bannur, Yidan Zhu, Liming Dong, Haichao Wei, Qi Guo, Luke Simon, Liangjie Hong, and Wenjing Zhang. 2025. A Scalable and Efficient Signal Integration System for Job Matching. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (Toronto ON, Canada) (KDD '25)*. Association for Computing Machinery, New York, NY, USA, 4659–4669. doi:10.1145/3711896.3737185
- [26] Ping Liu, Haichao Wei, Xiaochen Hou, Jianqiang Shen, Shihai He, Qianqi Shen, Zhujun Chen, Fedor Borisov, Daniel Hewlett, Liang Wu, Srikant Veeraraghavan, Alex Tsun, Chengming Jiang, and Wenjing Zhang. 2025. LinkSAGE: Optimizing Job Matching Using Graph Neural Networks. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1 (Toronto ON, Canada) (KDD '25)*. Association for Computing Machinery, New York, NY, USA, 2448–2457. doi:10.1145/3690624.3709396

- [27] Xuezhao Liu. 2025. Deep Learning-Based Intelligent Resume-Position Matching System: Semantic Understanding and Recommendation of BERT Model in Massive Recruitment Data. In *Proceedings of the 2025 International Symposium on Machine Learning and Social Computing (MLSC '25)*. Association for Computing Machinery, New York, NY, USA, 8–13. doi:10.1145/3778450.3778452
- [28] Yoosof Mashayekhi, Nan Li, Bo Kang, Jefrey Lijffijt, and Tijl De Bie. 2024. A Challenge-Based Survey of e-Recruitment Recommendation Systems. *Comput. Surveys* 56, 10, Article 252 (2024), 33 pages. doi:10.1145/3659942
- [29] Vikas Mishra, Vineet Singh, Shikha Singh, Bramah Hazela, Pooja Khanna, and Pawan Singh. 2025. A Hybrid Semantic-Structural Approach for Fair and Accurate Automated Resume Screening. In *2025 IEEE 7th International Conference on Computing, Communication and Automation (ICCCA)*. 1–5. doi:10.1109/iccca66364.2025.11325535
- [30] Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, Roger Chou, Julie Glanville, Jeremy M. Grimshaw, Asbjørn Hróbjartsson, Manoj M. Lalu, Tianjing Li, Elizabeth W. Loder, Evan Mayo-Wilson, Steve McDonald, Luke A. McGuinness, Lesley A. Stewart, James Thomas, Andrea C. Tricco, Vivian A. Welch, Penny Whiting, and David Moher. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372 (2021), n71. doi:10.1136/bmj.n71
- [31] Parsifal. 2026. Parsifal: Perform Systematic Literature Reviews. <https://parsifal.ai/>
- [32] Aravinda Reddy Putluru, Priyanshu Arora, C. Ashokkumar, and C. Jothi Kumar. 2026. Semantic Resume Screening using Transformer Networks and Learning-to-Rank Models. In *Proceedings of the 2026 International Conference on AI-Driven Smart Systems and Ubiquitous Computing, ICAUC 2026*. Institute of Electrical and Electronics Engineers Inc., 1947–1955. doi:10.1109/icauc68182.2026.11441279
- [33] Chuan Qin, Hengshu Zhu, Tong Xu, Chen Zhu, Chao Ma, Enhong Chen, and Hui Xiong. 2020. An Enhanced Neural Network Approach to Person-Job Fit in Talent Recruitment. *ACM Trans. Inf. Syst.* 38, 2, Article 15 (feb 2020), 33 pages. doi:10.1145/3376927
- [34] Vishnu Pandian R, Yugendran S, M Revathi, and J Gold Beulah Patturose. 2025. Resume Ranker for Freshers using NLP. In *2025 7th International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*. 902–907. doi:10.1109/icidca66325.2025.11280571
- [35] Sima Rezaeipourfarsangi and Evangelos E. Milios. 2023. AI-powered Resume-Job matching: A document ranking approach using deep neural networks. In *Proceedings of the ACM Symposium on Document Engineering 2023 (Limerick, Ireland) (DocEng '23)*. Association for Computing Machinery, New York, NY, USA, Article 22, 4 pages. doi:10.1145/3573128.3609347
- [36] Yosuke Saito and Kazunari Sugiyama. 2023. Multi-Behavior Job Recommendation with Dynamic Availability. In *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region (Beijing, China) (SIGIR-AP '23)*. Association for Computing Machinery, New York, NY, USA, 264–271. doi:10.1145/3624918.3625314
- [37] Farah Saraswati and Z.K.A. Baizal. 2025. Hierarchical TF-IDF for Tech Job Recommendation: Comparing ESCO and MIND Ontologies. In *Beyond Technology Summit on Informatics International Conference, BTS-I2C 2025*. Institute of Electrical and Electronics Engineers Inc., 120–125. doi:10.1109/bts-i2c67944.2025.11399358
- [38] S. Senthilkumar, M. Sundares, J. Jeswin, and C. Guru Prasad. 2025. HireSight: AI and Blockchain-Driven Candidate Matching, Verification, and ATS Resume Generation. In *OCIT 2025 - Proceedings 23rd OITS International Conference on Information Technology*. Institute of Electrical and Electronics Engineers Inc., 1–6. doi:10.1109/ocit66168.2025.11400344
- [39] Adarsh Kumar Singh, Lakshay Khandelwal, Harshit Chaudhary, Ramneet Kaur, and Sushant Jhingran. 2025. AI-Powered Job Matching and Resume Enhancement for Freshers Using Skill Graph Embeddings and Language Models. In *2025 International Conference on Emerging Technologies and Innovation for Sustainability, EmergIN 2025*. Institute of Electrical and Electronics Engineers Inc., 460–465. doi:10.1109/emergin67762.2025.11450922
- [40] Jitendra Singh, Mekhala, Kamini Tanwar, Ashish Kumar, Aanchal Vij, and Ashish Jain. 2026. AI-Powered Resume and Job Description Matcher: A Semantic Similarity Approach Using SBERT and Contrastive Fine-Tuning. In *2026 International Conference on Smart Futuristic Technology*. 1–9. doi:10.1109/icsft66733.2026.11507545
- [41] B. Sribharathi, Sv Balamurugan, S Megavarmaraj, S Deepak, and S Kajendhiran. 2025. Scopira: An AI-Driven Career Guidance System Using Resume Parsing, Skill Gap Analysis, and Intelligent Job Matching. In *2025 10th International Conference on Smart Structures and Systems (ICSSS)*. 1–8. doi:10.1109/icsss66939.2025.11346170
- [42] Lun Su. 2025. Research and Practice on Resume-Job Matching Model Based on Multi-modal Fusion. In *Proceedings of the 2025 6th International Conference on Big Data Economy and Information Management (BDEIM '25)*. Association for Computing Machinery, New York, NY, USA, 1276–1284. doi:10.1145/3800000.3800198
- [43] Abhishek Tiwari, Adit Katiyar, Sneha Dhanuka, and Neeraj Kumar Verma. 2025. HireIndex: A RAG-Enhanced AI Recruitment Platform for Fair, Transparent, and Efficient Hiring. In *2025 8th International Conference on Emerging Technologies in Computer Engineering: Advances in Computing, Healthcare and Smart Systems (ICETCE)*. 198–201. doi:10.1109/icetce66848.2025.11387932
- [44] Diana Vyshotravka, Victoria Vysotska, Zhengbing Hu, Dmytro Uhryn, Yuriy Ushenko, and Kyrilo Smelyakov. 2025. Smart Application for Recruiting Based on Natural Language Processing Methods, Transformer Models and Siamese Neural Network Architecture. *International Journal of Intelligent Systems and Applications* 17, 5 (2025), 95–157. doi:10.5815/ijisa.2025.05.07
- [45] Liang Wang and Shuang He. 2025. Research on Enterprise Human Resource Allocation Optimization Method Based on Knowledge Graph-Enhanced Dual-Tower BERT Model. In *Proceedings of the 2025 2nd International Conference on Economic Data Analytics and Artificial Intelligence (EDAI '25)*. Association for Computing Machinery, New York, NY, USA, 470–475. doi:10.1145/3789297.3789372
- [46] Yiqian Wang. 2024. Research on Personalized Employment Recommendation System for Higher Vocational Students Based on Deep Learning Algorithms. In *Proceedings of 2024 4th International Conference on Big Data, Artificial Intelligence and Risk Management, ICBAR 2024*. Association for Computing Machinery, Inc, 479–483. doi:10.1145/3718751.3718828
- [47] Ziyang Wang, Wei Wei, Chenwei Xu, Jun Xu, and Xian-Ling Mao. 2022. Person-job fit estimation from candidate profile and related recruitment history with co-attention neural networks. *Neurocomputing* 501 (2022), 14–24. doi:10.1016/j.neucom.2022.06.012
- [48] Agnieszka Wosiak. 2021. Using semantic enrichment methods in expert search system for recruitment process in IT corporation. *Procedia Computer Science* 192 (2021), 2422–2431. doi:10.1016/j.procs.2021.09.011
- [49] Bowen Yang and Zhixuan Shen. 2025. Knowledge graph construction and talent competency prediction for human resource management. *Alexandria Engineering Journal* 121 (2025), 223–235. doi:10.1016/j.aej.2025.02.043
- [50] Juanmei Yang. 2025. Knowledge Graph-Based Intelligent Recommendation and Dynamic Matching System for Tobacco Industry Human Resources. In *Proceedings of 2025 International Conference on Economic Management and Big Data Application, ICEMBDA 2025*. Association for Computing Machinery, Inc, 478–482. doi:10.1145/3770177.3770254
- [51] Kaichun Yao, Jingshuai Zhang, Chuan Qin, Peng Wang, Hengshu Zhu, and Hui Xiong. 2022. Knowledge Enhanced Person-Job Fit for Talent Recruitment. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*. 3467–3480. doi:10.1109/icde53745.2022.00325