

Article

Not peer-reviewed version

---

# Operationalizing Fit in Recruitment: A Multi-KPI, Evidence-Based Matching Architecture

---

[Angelo Leogrande](#)<sup>\*</sup>, Mauro di Molfetta, [Nicola Magaletti](#), Valeria Notarnicola, Stefano Mariani

Posted Date: 3 March 2026

doi: 10.20944/preprints202603.0023.v1

Keywords: job-candidate matching; semantic embeddings; multi-KPI framework; skill coverage; decision support systems



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Operationalizing Fit in Recruitment: A Multi-KPI, Evidence-Based Matching Architecture

Angelo Legrande <sup>1,\*</sup>, Mauro di Molfetta <sup>2</sup>, Nicola Magaletti <sup>2</sup>, Valeria Notarnicola <sup>2</sup> and Stefano Mariani <sup>2</sup>

<sup>1</sup> LUM University Giuseppe Degennaro, Casamassima, Italy

<sup>2</sup> LUM Enterprise s.r.l., Casamassima, Italy

\* Correspondence: legrande.cultore@lum.it

## Abstract

The current study proposes a multi-KPI approach for the alignment of job offers with candidate profiles that combines semantic signals, skill coverage indicators, and behavioral/contextual dimensions into a single decision support architecture. The approach is based on a semantic-first paradigm that represents job offers and candidate profiles as text embeddings that can be compared using the cosine similarity measure to produce a scalable baseline ranking that remains robust across different languages and styles. The baseline approach can then be extended using additional KPIs that capture different dimensions of the candidate-job alignment: the Hard Skill Coverage Ratio (HSCR) and the Skill Gap Index (SGI) measure the satisfaction of hard skill requirements and the remaining skill gaps; the Hard Skill Proficiency Similarity (HSPS) measures semantic similarity in the hard skill domain; the Soft Skill Semantic Alignment (SSSA) and the Soft Skill Evidence Density (SSED) measure the semantic similarity of the candidate behavior and the quality of the corroborating evidence; the Cultural & Team Fit Score (CTFS) addresses the organizational fit; and other indicators cover the operational feasibility of the candidate-job match. The approach is implemented using Python and tested using a realistic testbed of 300 EURES job postings and 300 heterogeneous candidate profiles across different formats and languages. The results show that the semantic approach yields a stable baseline that can be used to produce coherent and well-ordered shortlists of candidate-job pairs while the additional KPIs improve the interpretability of the approach and the feasibility of constraint satisfaction. The analysis of the candidate-job scores and the separations of the rankings also reveals that there are clusters of technically equivalent candidates and that there are clear cases of dominance. Overall, the approach can support an incremental move toward multi-criteria decision making while balancing scalability, transparency, and governance-by-design requirements.

**Keywords:** job–candidate matching; semantic embeddings; multi-KPI framework; skill coverage; decision support systems

## 1. Introduction

Digital recruitment platforms and labor markets are increasingly employing automated or semi-automated systems to support job-c candidate matching. The increase in unstructured data contained in job advertisements and curricula vitae (CVs), which are composed of heterogeneous writing styles and multiple languages, has further fueled research into decision support systems. Therefore, research has been carried out on a broad spectrum of techniques ranging from rule-based techniques to machine learning-based techniques and embedding-based semantic similarity models (Kurek et al., 2024; Frazzetto et al., 2025). Although these techniques are significantly better at handling large sets of documents, they also pose a trade-off in semantic flexibility, interpretability, and usability. For example, embedding-based techniques provide powerful tools to measure semantic-level similarity between job advertisements and candidate CVs to address issues with keyword-based matching. The

potential of embedding-based techniques has also been enhanced by employing graph-based techniques and inductive learning techniques, among others (Frazzetto et al., 2025). The recruitment process involves a complex multi-criteria decision problem where technical needs, skill gaps, soft skills, organization culture, and logistics must be considered simultaneously. However, research carried out by experts has shown that hiring involves complex trade-offs. The decision process cannot be reduced to a single measure of semantic similarity. In fact, most research has concentrated on a single best measure of semantic similarity or skills match. The ranking produced by most decision support systems is difficult to interpret or audit. Moreover, it does not match real-world decision-making constraints. Although novel techniques such as knowledge graph-based semantic relatedness in job title matching have been proposed (Zadykian et al., 2025), a gap exists between high-performance black-box-based techniques and decision support systems where recruitment decision-making needs to be explained, governed, and controlled in addition to being scalable and having high performance (Aleisa et al., 2023). The research question that is being addressed in the paper is: "How can job-candidate matching be operationalized as a multi-dimensional, interpretable, and scalable decision-support process that combines semantic similarity with explicit skill-based, behavioral, and contextual constraints?" To achieve this, we propose a novel multi-KPI algorithmic framework that is intended to help decision-makers make better choices with the help of a number of structured "fit" indicators, including technical requirement coverage, skills gaps, semantic similarity of technical profiles, soft skills compatibility, evidence density, cultural team fit, and contractual/contextual compatibility. The novelty of the work resides not in the introduction of an additional score, but in the integration of heterogeneous "fit" indicators within a single unifying framework. Although zero-shot recommendation models like Kurek et al. (2024) and graph neural networks like Frazzetto et al. (2025) achieve outstanding results in automated matching, they are mostly optimized towards a single modeling paradigm. The novelty of our work resides in the combination of discrete rule-based KPIs with continuous embedding-based semantic measures and evidence-based "fit" indicators, allowing the system to operate both as a high-recall semantic retrieval system and as a constraint-based multi-criteria evaluation system, with incremental refinement capabilities without compromising interpretability. In an empirical context, the study positions itself within the context of an industrial project that is funded using public resources and aims at developing data support systems for labor market matching and workforce policymaking. The relevance of the study is further highlighted by the evaluation of the proposed system using various job postings and CVs that are more likely to represent real-world operational contexts than artificially generated data sets. The use of further AI recruitment tools within the context of national labor markets further highlights the relevance of the study, as mentioned in Aleisa et al. (2023). The relevance of the study can be highlighted from the context of the existing literature on matching systems along three different dimensions. First, the study proposes a structured methodological framework that can integrate semantic matching with constraint handling using a single architecture. Second, the study proposes a set of KPIs that can be used for decision-making purposes within the context of matching systems, while also highlighting the relevance of interpretability. Third, the study proposes an empirical analysis that is based on an industrial project funded using public resources, thus highlighting the relevance of the study.

The article continues as follows: in Section 2, the literature review is provided, while Section 3 defines the overall methodological framework, from scores to the multi-KPI approach for recruitment matching. Section 4 defines the embedding-based semantic approach for job-CV matching, while Section 5 evaluates the Hard Skill Coverage Ratio (HSCR) approach, especially the requirement-based ranking approach. Section 6 defines the Hard Skill Proficiency Similarity (HSPS) approach, especially the semantic approach for high-recall matching. Section 7 defines the Skill Gap Index (SGI) approach for providing additional information for hard skills-based matching. Section 8 defines the Soft Skill Semantic Alignment (SSSA) approach for continuous soft skills-based matching, while Section 9 defines the Soft Skill Evidence Density (SSED) approach for providing evidence-based support for soft skills-based matching. Section 10 defines the Cultural and Team Fit Score (CTFS)

approach for providing additional information for organizational compatibility-based matching, while Section 11 defines the Contract Compatibility Score (CCS) approach for providing evidence-based support for multi-stage recruitment pipeline-based matching. Section 12 defines the Location and Mobility Fit Index (LMFI) approach for constraint-based matching, while Section 13 evaluates the Seniority and Compensation Alignment (SCA) approach for providing a high-resolution measure of structural compatibility-based matching. Section 14 evaluates the overall performance of the multi-KPI approach for providing evidence-based support for multi-stage recruitment pipeline-based matching, while Section 15 evaluates the overall approach of integrating hard skills, soft skills, and other factors for recruitment matching purposes. Finally, Section 16 defines the overall limitations of the article, while Section 17 defines the conclusion of the article.

## 2. Literature Review

Recent studies on AI-based recruitment processes indicate that there is a strong trend toward semantic similarity, deep learning, and end-to-end automation. Ajjam and Al-Raweshidy (2026) suggest an embedding-based semantic matching framework similar to our work but still centered around one dominant matching criterion. Kim (2026) changes the focus by examining the implications and risks of AI-based recruitment processes at the macro-level of society. Although this aligns with our emphasis on the importance of awareness and understanding at the macro-level, it does not offer any framework. The majority of the recent contributions focus on resume screening and scoring from an engineering and workflow optimization point of view. Liu et al. (2026), Dangeti et al. (2026), Sawant et al. (2026), Chihab et al. (2025), Barath et al. (2025), Hepzibah et al. (2025), Yadav et al. (2025), Ammupriya et al. (2025), Nabila et al. (2025), Singla et al. (2025), Dhobale et al. (2025), Tilve et al. (2025), Gangoda et al. (2024), and Waghmare et al. (2024) are some of these papers. Although these systems improve the efficiency and effectiveness of the recruitment process, they still use the monolithic concept of matching as an optimization problem. In these systems, different dimensions of matching are fused into one score. From the recommender systems and methodology point of view, Çelik Ertuğrul and Bitirim (2025) indicate the problems of explainability, hybridization, and evaluation in matching. They suggest the concept of multi-sided fairness in algorithmic hiring by Kaya and Bogers (2025). Likewise, hyper-personalization-based approaches (Alqudah et al., 2025) or personality-based approaches (Khan et al., 2025) enhance the notion of "fit" but integrate these dimensions into unifying, largely uninterpretable scoring systems. A second group of papers enhances the technical scope of matching using multimodal data or advanced learning techniques (Wu et al., 2025; Yazici et al., 2024; Dilli Ganesh et al., 2025; Kurek et al., 2024; Zhang, 2024), or predictive/zero-shot techniques (Waghmare et al., 2024; Kurek et al., 2024). Again, while technically sophisticated, these systems continue to view matching as an end-to-end prediction or ranking problem, rather than a decomposable multi-criteria decision problem. Other papers take a broader approach, proposing infrastructure, platform, or strategic-level AI-driven recruitment systems or models (Badouch & Boutaounte, 2025; Wahyuningrum et al., 2025; Es-Said et al., 2025; Saouabe et al., 2025; Kumar et al., 2025; Pandit et al., 2024; Dilusha et al., 2024; Jamil et al., 2024; Bhalke et al., 2024; Choudhuri et al., 2024). Other related papers address adjacent components in the recruitment pipeline, including resume generation, job portals, or user experience (Jha et al., 2025; Kulkarni et al., 2025; Haneef et al., 2025; Babalola et al., 2024), or organizational/performance outcomes in particular contexts, including social media recruitment (Al-Dmour et al., 2025) or career development (Sathish et al., 2024; Noel & Sharma, 2024). While these papers provide further evidence of the increasing ubiquity of AI in HR systems, they do not address the question of how to formally address matching as a controllable, multi-dimensional problem. From methodological and historical perspectives, Mat Saad et al. The recent works by (2022) and Rojas-Galeano et al. (2022) provide reviews and bibliometric analyses that reinforce the prevalence of model-centric, performance-oriented approaches. Previous systems based on classical machine learning and deep learning paradigms, as in Najjar et al. (2021), Mridha et al. (2021), Vasilescu et al. (2019), demonstrate the progress of automated screening and ranking systems, while Martínez and Fernández (2019) introduce rule-

based and ontology-driven approaches specifically targeting ethical and legal auditing. Sharma and Garg (2024) further explore optimization-based HR analytics, while Mohamed et al. (2024) propose a precision CV matching system, again based on unified optimization/scoring paradigms. Together, they attest to the continuous expansion of technical, data, and application domains, while also exemplifying the conceptual limitation of the unified approaches: job-candidate matching is still treated as a prediction, ranking, or process-automation problem. In contrast, the current article proposes a new methodological direction by defining matching as a multi-criteria, interpretable, and auditable decision support process, based on a unified set of complementary KPIs covering semantic similarity, skill sets, behavioral evidence, as well as contextual/contractual feasibility, which may be seen as filling the gap left unaddressed in the literature, which is largely implicit or unstructured. See Table 1.

**Table 1.** Positioning the Proposed Multi-KPI Framework within the Recruitment and Matching Literature.

Macro-Theme	Representative Works	Main Focus & Results	Methods / Algorithms	Critical Limit w.r.t. Our Approach
Semantic & ML-Based Matching	Ajjam & Al-Raweshidy (2026); Wu et al. (2025); Yazici et al. (2024); Dilli Ganesh et al. (2025); Kurek et al. (2024); Zhang (2024); Waghmare et al. (2024)	Improve matching accuracy via embeddings, deep learning, multimodal data, zero-shot or hybrid models	Sentence embeddings, Transformers, CNNs, multimodal models, knowledge graphs, zero-shot learning	Matching is framed as a single end-to-end prediction or ranking problem; lacks decomposition into interpretable decision dimensions (e.g., coverage, gaps, constraints, evidence)
Resume Screening, Parsing & Ranking Pipelines	Liu et al. (2026); Dangeti et al. (2026); Sawant et al. (2026); Chihab et al. (2025); Barath et al. (2025); Hepzibah et al. (2025); Yadav et al. (2025); Ammupriya et al. (2025); Nabila et al. (2025); Singla et al. (2025); Dhobale et al. (2025); Tilve et al. (2025); Gangoda et al. (2024); Mohamed et al. (2024); Sharma & Garg (2024); Najjar et al. (2021); Mridha et al. (2021); Vasilescu et al. (2019)	Automate screening, parsing, scoring, and ranking to improve efficiency and throughput	NLP pipelines, rule-based systems, ML/DL classifiers, GA+NN optimization, scoring functions	Typically rely on monolithic scores that fuse heterogeneous aspects of "fit" into a single opaque value, limiting explainability, audibility, and governance
Recommender Systems, Personalization & Fairness	Çelik Ertuğrul & Bitirim (2025); Kaya & Bogers (2025); Alqudah et al. (2025); Khan et al. (2025)	Highlight explainability, hybridization, fairness, and enriched notions of "fit" (e.g., personality, personalization)	Recommender systems, transfer learning, personality-based models, fairness-aware frameworks	Although conceptually rich, "fit" is still embedded in unified scoring models and not operationalized as explicit, auditable KPIs

Platforms, Infrastructure & Strategic HR AI	Badouch & Boutaounte (2025); Wahyuningrum et al. (2025); Es-Said et al. (2025); Saouabe et al. (2025); Kumar et al. (2025); Pandit et al. (2024); Dilusha et al. (2024); Jamil et al. (2024); Bhalke et al. (2024); Choudhuri et al. (2024); Jha et al. (2025); Kulkarni et al. (2025); Haneef et al. (2025); Babalola et al. (2024); Al-Dmour et al. (2025); Sathish et al. (2024); Noel & Sharma (2024); Lamikanra & Obafemi-Ajayi (2023)	Focus on end-to-end platforms, UX, Web platforms, LLM infrastructure, tools, RAG, blockchain, workflow systems, organizational strategic frameworks impact, and HR transformation	Address ecosystem and process rather than the core decision model; matching remains a black-box or secondary component
5. Reviews, Bibliometrics & Governance- Oriented Foundations	Yeju et al. (2026); Mat Saad et al. (2022); Rojas-Galeano et al. (2022); Martínez & Fernández (2019)	Map the field, highlight trends, and introduce early governance/ethics perspectives	Systematic reviews, bibliometrics, ontology-based and rule-based auditing approaches Identify the need for transparency and accountability, but do not provide a concrete multi-KPI operational decision framework

Note. This table synthesizes major research streams in AI-based recruitment, highlighting their methods and limitations, and situates our multi-KPI framework as an interpretable, modular, and governance-oriented alternative to monolithic matching approaches.

### 2.1. From Matching Pipelines to Decision Support: Structural Patterns in the Recruitment AI Literature

The network analysis offers a quantitative and structural understanding of the conceptual organization of the literature on AI-driven recruitment and job/candidate matching processes (Potočník et al., 2024; Van Esch et al., 2019). Table X displays the most important centrality measures of the core KPIs, including degree centrality, weighted degree centrality, and betweenness centrality. The structural configuration of the network reveals that the overall structure is highly polarized around the core nodes. The small number of concepts controls most of the network's connectivity and intermediation. As shown in Table X, job/candidate matching is found to be the most important node in the network. It bears the highest degree centrality score (9), the highest weighted degree centrality score (55), and a very high betweenness centrality score (0.4167). The last score reveals that job/candidate matching not only tends to co-occur most often with other concepts but also acts as an important intermediary by appearing on a considerable percentage of the shortest paths connecting different parts of the network. From the structural configuration of the network, it becomes evident that the literature centers around the general issue of matching labor demand and supply. Job/candidate matching acts as the most important concept connecting different research streams. These streams comprise resume/job matching, automated resume screening, semantic similarity, AI-driven recruitment, and issues of explainable AI in recruitment and AI hiring in government (Sajjadiani et al., 2019; Liem et al., 2018; Raghavan et al., 2020; Bogen & Rieke, 2018). This finding quantitatively supports the interpretation that, regardless of their technical or methodological specializations, the majority of the contributions frame their objectives with regard to the more general problem of job-candidate matching. A structural finding of particular interest is the role of semantic embeddings. While the degree of this node is lower than that of job-candidate matching, at 6, and the weighted degree, at 15, is also lower, the highest value of betweenness centrality, at 0.4676, is found for this node. This finding can be interpreted to mean that, more than with regard to any

other pair of weakly connected subdomains, this concept serves as a bridge. It can be concluded that embedding-based representations and text embeddings in the context of HR constitute the primary methodological infrastructure that unifies different approaches to the problem of recruitment matching (Ajjam & Al-Raweshidy, 2025). The table thus provides quantitative evidence that, far from being a technical option among many, semantic embeddings constitute the primary connective mechanism for linking different approaches, such as semantic similarity models, resume-job matching, AI-based approaches to recruitment matching, automated resume screening, work on explainability, and work on interpretability (Liem et al., 2018). The application-oriented nature of this field is seen in the quantitative data provided by the metrics for resume-job matching and automated resume screening, which, at 7 each, have relatively high values for the degree metric, while their very high values for the weighted degree metric, at 44 each, indicate a strong co-occurrence with many other concepts in the corpus. At the same time, their betweenness centrality remains relatively low at 0.0625 for both authors, which implies that they are not important conceptual bridges between various research areas. On the contrary, these are groups of densely interconnected nodes, each of which is focused on operational pipelines for filtering, ranking, and screening CVs (Sajjadiani et al., 2019; Van Esch et al., 2019). This is also in line with the prevalence of engineering-related contributions, which focus on optimization, efficiency, and scalability, often viewing the matching process as an optimization of workflow automation, as opposed to multi-criteria decision making. The same can be said of the use of semantic similarity and AI-based recruitment, both of which are indicated by a moderate to high degree value of 6 and 7, respectively, with corresponding weighted degree scores of 22 and 33, indicating strong usage and diffusion of these concepts within the literature (Ajjam & Al-Raweshidy, 2025; Potočník et al., 2024). However, both of these concepts are accompanied by very low betweenness centrality scores of 0.0093 and 0.0069, suggesting that these are not acting as structuring agents within the overall network, but are instead being used as part of existing approaches, perhaps as technical tools or as part of the broader context of existing approaches. In terms of structural interpretation, this again supports the view that these concepts are not being used as part of an explicit multi-KPI evaluation framework, but are being incorporated into monolithic models in which different aspects of fit, such as hard skill coverage, skill gap analysis, and soft skill assessment, are being implicitly combined within a single ranking or scoring function (Sajjadiani et al., 2019). The table also helps to clarify the position of explainable AI within recruitment and fairness, as well as fairness in AI-based recruitment approaches. These nodes also have non-negligible degree and weighted degree values (5 and 7, respectively; 11 and 38, respectively), which indicate their presence and discussion in conjunction with core technical issues (Raghavan et al., 2020; Mehrabi et al., 2021). Nevertheless, their betweenness centrality remains low at 0.0787 and 0.0069, respectively, which indicates their peripheral nature. Quantitatively, this also indicates that issues related to transparency, accountability, and governance tend to be related to existing AI-based recruitment and automated resume screening approaches rather than the core architecture of the matching system itself (Bogen & Rieke, 2018; Liem et al., 2018). Lastly, the least centrally located nodes in the network relate to recruitment decision support systems and skill extraction. The degree, weighted degree, and betweenness centrality values for this node are minimal at 1, 2, respectively, and 0, respectively. This provides strong quantitative evidence that the literature rarely discusses the issue of matching in the context of a decision support system, nor does it discuss the issue of skill extraction as a separate, relevant dimension for decision-making. This provides evidence that this issue tends to be treated as a component of a broader, end-to-end prediction system rather than being treated as a component of a multi-criteria decision-making system (Potočník et al., 2024). Overall, the quantitative metrics reported in the table, along with their structural implications, provide a coherent quantitative portrait of the literature. The literature tends to be dominated by approaches related to semantic matching, with semantic embeddings being the primary integrative methodology (Ajjam & Al-Raweshidy, 2025). The literature also tends to be dominated by application-focused research on resume-job matching and automated resume screening approaches (Sajjadiani et al., 2019; Van Esch et al., 2019). At the same time, the notion of the multi-KPI, interpretable, and governance-aware

recruitment process as a decision support process is structurally weak and underdeveloped (Raghavan et al., 2020; Mehrabi et al., 2022). This quantitative structure is consistent with the qualitative findings of the literature review, supporting the claim of the proposed multi-KPI framework as an attempt to place itself in an underexplored area of the field, shifting the focus away from monolithic matching scores toward a more modular, interpretable, and auditable decision support architecture. See Table 2.

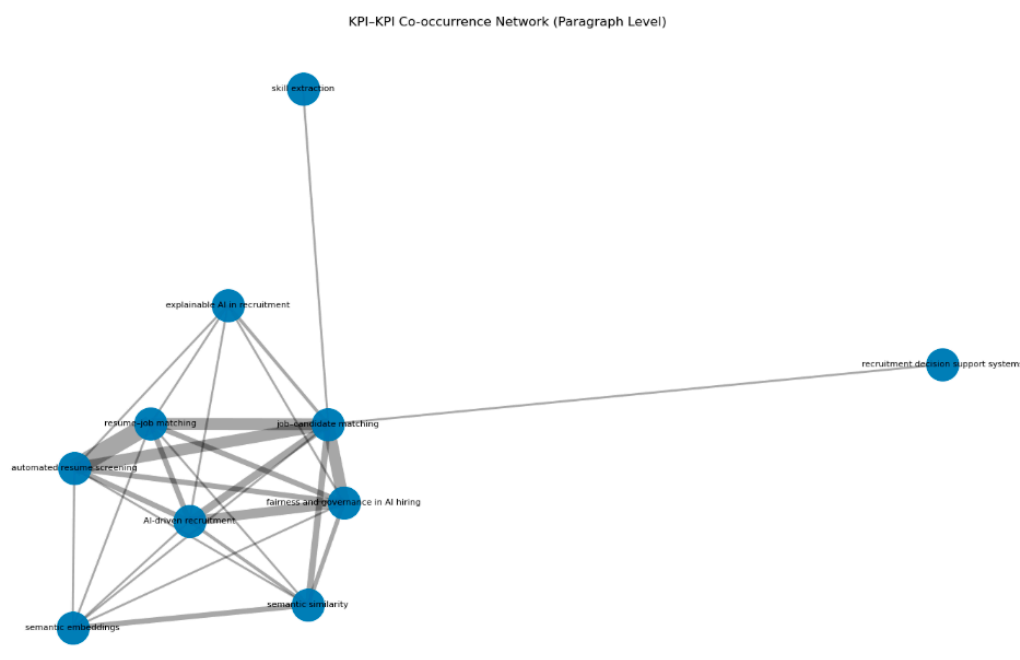
**Table 2.** Network Centrality Analysis of Core Concepts in AI-Based Recruitment Research.

KPI	Degree	Weighted Degree	Betweenness Centrality
semantic embeddings	6	15	0.467593
job-candidate matching	9	55	0.416667
explainable AI in recruitment	5	11	0.078704
resume-job matching	7	44	0.062500
automated resume screening	7	44	0.062500
semantic similarity	6	22	0.009259
fairness and governance in AI hiring	7	38	0.006944
AI-driven recruitment	7	33	0.006944
recruitment decision support systems	1	2	0.000000
skill extraction	1	2	0.000000

Note. This table reports degree, weighted degree, and betweenness centrality for key concepts in the literature network, highlighting which topics act as hubs and bridges in the research landscape and motivating the integrative, multi-KPI perspective adopted in this study.

The above figure 1 depicts the KPI-KPI co-occurrence network at the paragraph level in the literature review section. As can be observed, it provides a visual representation of the conceptual structure of the field. At the center of this structure, it can be noted that the process of job-candidate matching has become the main hub. This suggests that the literature has been centered on addressing the overarching problem of matching labor demand with labor supply. Moreover, it has been observed that most concepts are discussed in direct relation to this overarching objective. This suggests that most concepts are discussed in direct relation to this overarching objective (Potočnik et al., 2024). The cluster of concepts such as semantic embeddings, semantic similarity, resume-job matching, automated resume screening, AI-based recruitment, fairness and governance in AI hiring, and explainable AI in recruitment surrounds the main hub. The thickness of the lines connecting these concepts suggests that they are discussed in close proximity to one another in the paragraphs. This suggests that this field has been following a research direction where semantic concepts are combined with automated resume screening processes, along with concerns related to fairness and explainability (Ajjam & Al-Raweshidy, 2025; Raghavan et al., 2020). Practically, the figure corroborates that the dominant literature is organized around semantic matching and automated screening pipelines, with ethical and transparency issues being integrated within the same technical framework, rather than changing it (Mehrabi et al., 2021; Raghavan et al., 2020). Semantic embeddings and semantic similarity can be found as well-integrated nodes within the dominant cluster, which highlights the importance of distributed text representations as a unifying methodological framework across many of the literature works. These concepts are not independent of one another, as they are strongly intertwined with application aspects such as screening/matching, which reflects the recent trend of incorporating embedding-based methodologies and semantic similarity approaches in the literature (Ajjam & Al-Raweshidy, 2025). In contrast, two nodes can be easily recognized as being on the periphery of the network, namely the one related to decision support systems in recruitment and the one related to skill extraction. The former node is strongly connected to the rest of the network only through job-candidate matching, which highlights that the decision support perspective is still on the periphery of the literature, failing to function as an important structuring dimension (Potočnik et al., 2024). The same can be said about the node of skill

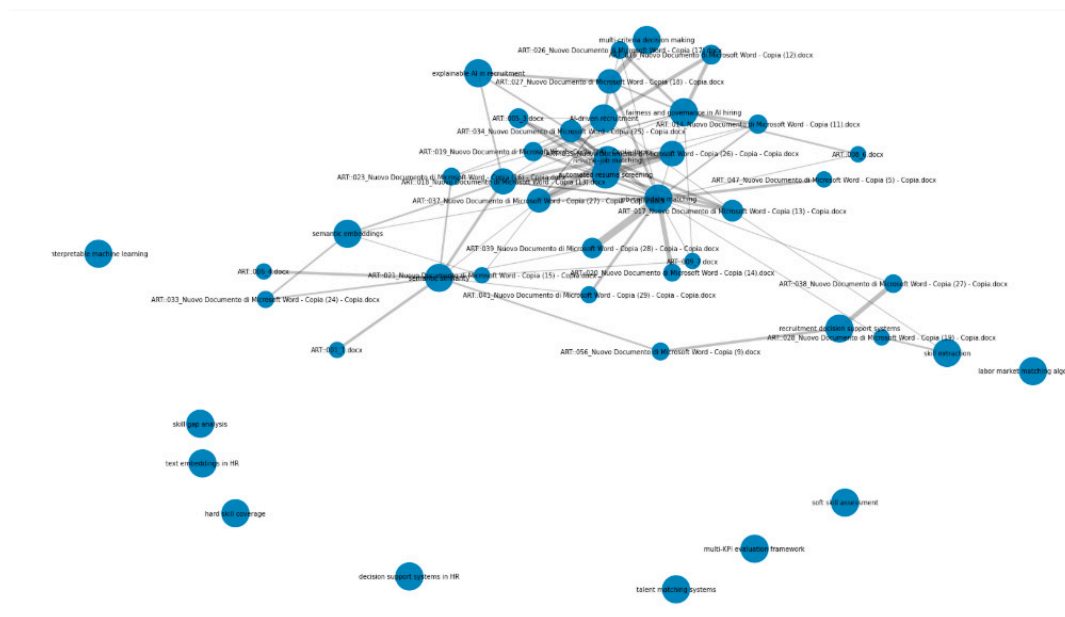
extraction, which is found to be isolated, reflecting its treatment as an auxiliary technical step, rather than an independent, conceptually central dimension of the decision process. The figure can be recognized as providing a visual synthesis that is in line with the theoretical analysis. The dominant literature is centered around a technical, application-oriented core, which revolves around semantic matching and automated screening, while the decision support perspective on the multi-criteria, interpretable, and governance-aware nature of the recruitment process is on the periphery of the conceptual network, as highlighted by the analysis of the nodes (Raghavan et al., 2020; Mehrabi et al., 2021). From this vantage point, the multi-KPI framework can be recognized as targeting a structurally underexplored domain, aiming to transform the matching process from an exclusive ranking/prediction task to a structured, decomposed, and controllable decision-making process. See Figure 1.



**Figure 1.** KPI-KPI Co-Occurrence Network in Recruitment Research (Paragraph Level). Note. Nodes represent recurring concepts and edges paragraph-level co-occurrence, revealing job-candidate matching as the central hub, dense links among semantic and screening themes, and peripheral roles for decision support and skill extraction within the literature.

The above Figure 2 represents the co-occurrence network of articles, keywords, and key performance indicators (KPIs), providing a more comprehensive and detailed overview of the conceptual space of the literature under examination. In contrast to the KPI network, it can clearly distinguish between document nodes and conceptual nodes, making it easier to monitor the variety of different contributions to specific topics, which are grouped around specific keywords (Potočník et al., 2024). In the center of the network, there is a dense group of nodes, which are highly interconnected, covering a significant number of keywords, such as job candidate matching, resume-job matching, automated resume screening, semantic embeddings, semantic similarity, AI-based recruitment, explainable AI in recruitment, fairness, and governance in AI hiring. The importance of the above group of keywords, which are highly connected, can be identified by the significant number of related articles, implying that the literature under examination is highly focused on the application of semantic matching and automated screening, with a strong emphasis on the technical-operational aspects of the respective tasks (Ajjam & Al-Raweshidy, 2025). A small portion of more recent literature is focused on explainability and fairness, too, in the context of semantic matching and

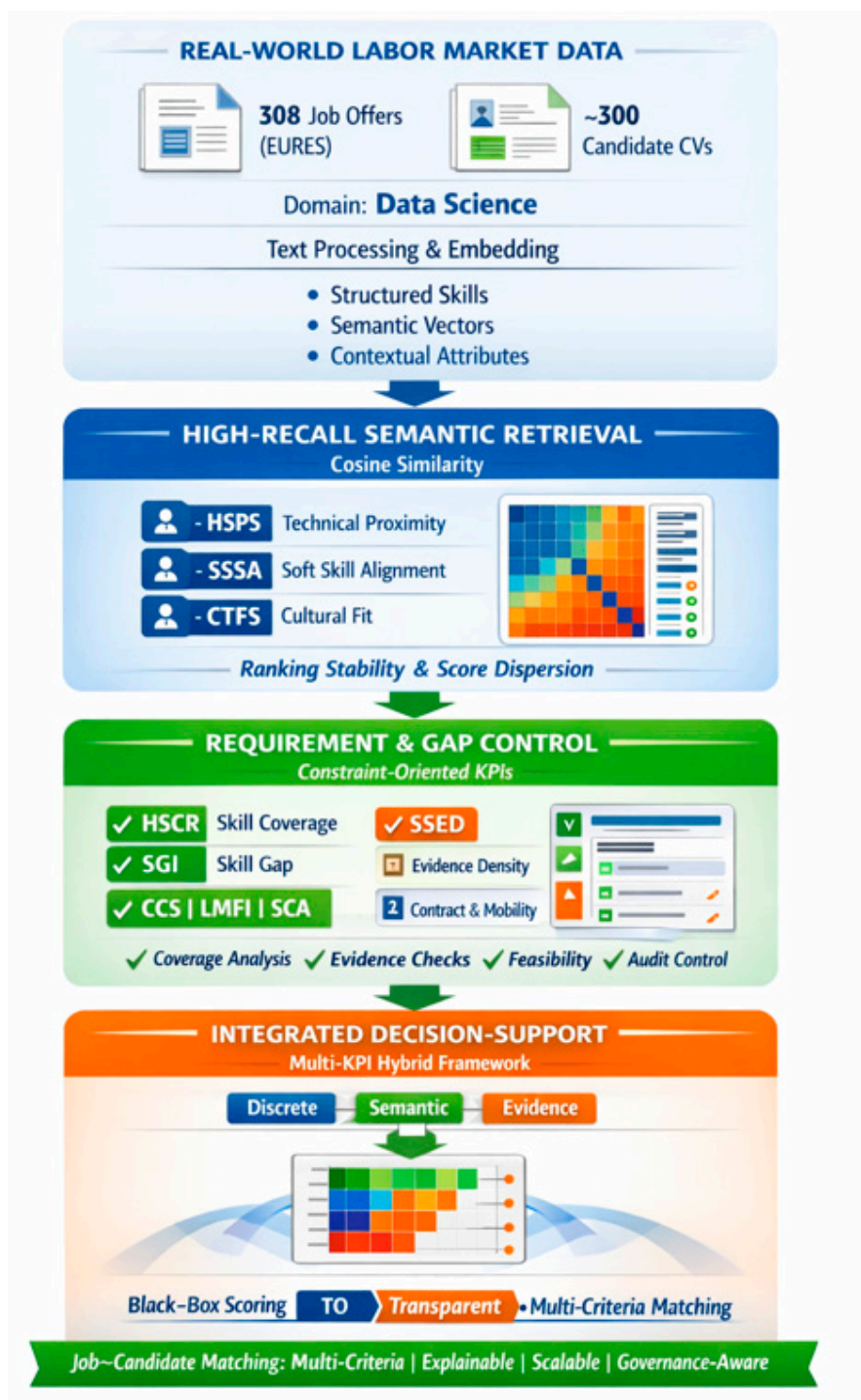
automated screening, as suggested by Raghavan et al. (2020), Mehrabi et al. (2021), still within the technical-operational context. The significant presence of the keyword “semantic embeddings” as one of the highly connected nodes of the network further supports the above finding that it provides a unifying, shared infrastructure for a significant portion of the literature under examination (Ajjam & Al-Raweshidy, 2025). A significant number of articles are related to more than one keyword of the group, implying the existence of different approaches incorporating aspects of semantic similarity, automated screening, and, to some extent, explainability. Apart from the above-mentioned core, some of the peripheral concepts that are weakly interconnected include multi-KPI evaluation framework, talent matching, soft skills evaluation, decision support system in HR management, hard skills coverage, skills gaps evaluation, and text embeddings in HR management. The weak connectivity of the above-mentioned concepts with the core concepts as well as with each other might be due to the lower representation of these concepts as unique entities in the literature. The concepts that are associated with an explicitly multi-criteria decision-support approach, along with an emphasis on multi-KPI evaluation, are located at the periphery of the network, where connectivity with other concepts is minimal, i.e., with only a few articles. The figure indicates that the literature of this domain is mostly focused on the technical application of the core of semantic matching/screening. The dimensions of semantic matching, which specify the criteria, hard skills evaluation, soft skills evaluation, and decision support, are located at the periphery of the network. The connectivity of the above-mentioned concepts with each other as well as with the core of the network supports the idea that the field of semantic matching/screening is mostly focused on model-centric as well as pipeline-centric approaches (Ajjam & Al-Raweshidy, 2025). The above figure also emphasizes the significance of the current manuscript, as it covers an under-explored part of the network with an alternative multi-KPI-based decision-support framework, as opposed to monolithic approaches of matching, as emphasized in the literature (Raghavan et al., 2020; Mehrabi et al., 2021). See Figure 2.



**Figure 2.** Article-Keyword-KPI Co-Occurrence Network of the Recruitment Matching Literature. Note. This network maps articles, keywords, and KPIs, revealing a dense core around semantic matching and screening, with explainability and multi-KPI decision support remaining peripheral, highlighting an underexplored space addressed by the proposed framework.

### 3. From Monolithic Scores to Modular Evidence: A Multi-KPI Methodology for Recruitment Matching

The methodological choice underlying this framework is to move beyond monolithic matching strategies—whether purely semantic or purely rule-based—and to decompose the notion of “fit” into a structured family of complementary, interpretable, and modular KPIs. See Figure 4.



**Figure 4.** Architecture of the Multi-KPI, Evidence-Based Job–Candidate Matching Framework. Note. The figure illustrates a layered pipeline integrating discrete constraints, semantic similarity, and evidence density KPIs, transforming black-box scoring into transparent, multi-criteria decision support for scalable, interpretable, and governance-aware recruitment matching across real-world job and CV data.

Each of these measures is associated with a particular, theoretically grounded dimension of compatibility, including explicit satisfaction of technical requirements (HSCR, SGI), semantic proximity of technical profiles (HSPS), behavioral and interpersonal compatibility (SSSA, SSED, CTFS), and feasibility constraints of operations (CCS, LMFI, SCA). What is original in our proposal is not any particular measure taken in isolation, but their systematic integration into a single framework that integrates different types of evidence, including discrete measures defined on sets (HSCR, SGI, CCS), continuous measures of semantic similarity between embeddings (HSPS, SSSA, CTFS), and density-based evidence measures (SSED). This mixed approach is consistent with current research efforts to transcend traditional, monolithic matching scores in favor of structured, evidence-based job matching models (Martínez-Manzanares et al., 2024) and with research that calls for integrating semantic textual relatedness with explicit knowledge representations to improve their interpretability (Zadykian et al., 2025). In particular, our approach, which integrates different types of evidence, avoids the limitations of embedding-based recruitment pipelines, as recently recognized in related research (Aleisa et al., 2023). From a methodological point of view, it should be noted that the framework explicitly differentiates between a semantic core layer (HSPS, SSSA, CTFS), which allows for high recall matching with robustness to natural languages, and other layers focused on constraint, evidence, and decision support (HSCR, SGI, SSED, CCS, LMFI, SCA), which reintroduce auditability, feasibility, and decision support. This is in line with recent trends in explainable matching, which combine semantic similarity models with structured knowledge to ensure interpretability, controllability, and explainability of matching outcomes (Zadykian et al., 2025), as well as empirically validated models of recruitment that combine expert criteria with automated scoring approaches (Martínez-Manzanares et al., 2024). In addition, recent trends in system-level implementations of AI-powered recruitment approaches highlight the importance of modularity, which allows for the implementation of constraints without requiring the full redesign of the matching pipeline (Aleisa et al., 2023). In this regard, the differentiation between layers supports the development of systems that can support the implementation of different organizational priorities without requiring modifications to the core matching process. The set of KPIs outlined in Table X can therefore be seen to support the development of job-candidate matching as a multi-criteria, explainable, and governance-aware decision support process, which represents a significant methodological step-change relative to more simplistic matching approaches, which may only support single-criterion matching outcomes, or approaches that are entirely opaque to the matching process (Martínez-Manzanares et al., 2024; Zadykian et al., 2025; Aleisa et al., 2023). See Table 3.

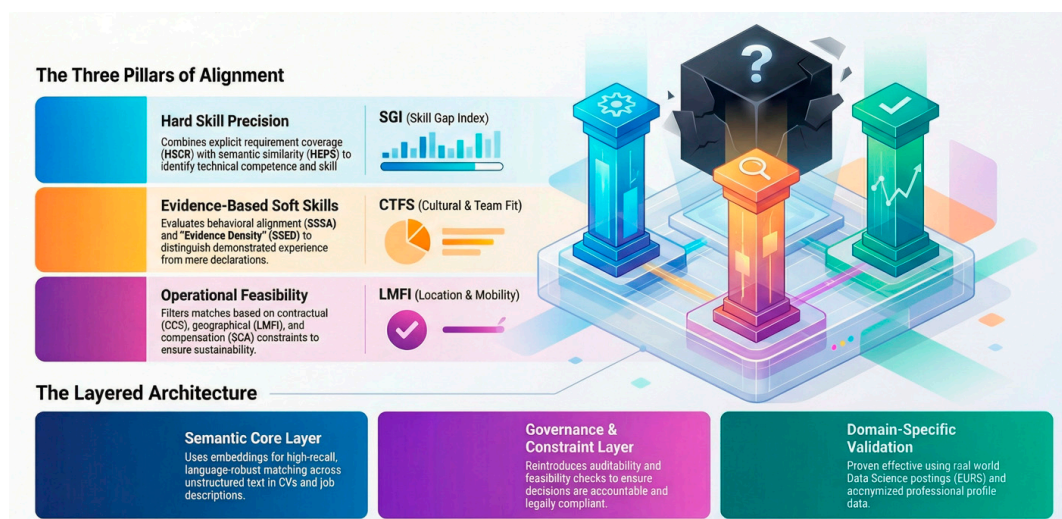
**Table 3.** Overview of the Proposed Multi-KPI Framework for Job–Candidate Matching.

Category	KPI	Description	Utility for the Article
	Hard Skill Coverage Ratio (HSCR)	Measures the proportion of required technical skills in a job offer that are explicitly covered by a candidate profile. Formally, $HSCR =  Skills_{job} \cap Skills_{cv}  /  Skills_{job} $ .	Provides an interpretable and auditable measure of technical completeness with respect to explicit job requirements; supports requirement-driven ranking and constraint-based filtering.
Hard Skills	Hard Skill Proficiency Similarity (HSPS)	Measures semantic similarity between the technical content of job descriptions and CVs using embedding-based cosine similarity.	Serves as the semantic core of the system, capturing technical domain proximity beyond keyword overlap and enabling scalable, language-robust matching.
	Skill Gap Index (SGI)	Measures the proportion of missing required skills, defined as $SGI = 1 - HSCR$ .	Reframes coverage into a gap-oriented perspective, supporting reskilling, training planning, and developmental matching strategies.

	Soft Skill Semantic Alignment (SSSA)	Measures semantic alignment between soft-skill-related content in job descriptions and candidate profiles using embeddings.	Introduces a behavioral and interpersonal dimension of fit, enabling differentiation among technically similar candidates based on soft skill compatibility.
Soft Skills	Soft Skill Evidence Density (SSED)	Measures the density of sentences or segments in a profile that provide concrete evidence of soft skills relative to document length.	Distinguishes between declared and demonstrated soft skills, improving robustness, explainability, and evidence-based ranking.
	Cultural & Team Fit Score (CTFS)	Measures semantic and contextual compatibility between organizational culture/team context and candidate profile.	Supports assessment of organizational and team-level integration, contributing to retention-oriented and governance-aware decision making.
	Contract Compatibility Score (CCS)	Measures the proportion of compatible contractual conditions between job offer and candidate constraints or preferences.	Filters out technically suitable but contractually infeasible matches, increasing operational feasibility of recommendations.
Contractual / Contextual	Location & Mobility Fit Index (LMFI)	Measures geographical and mobility compatibility considering distance, remote work policies, relocation, and travel constraints.	Addresses logistical frictions, especially in cross-country and international labor markets.
	Seniority & Compensation Alignment (SCA)	Measures alignment between required seniority/compensation levels and the candidate's experience and expectations.	Reduces risks of overqualification, underqualification, and early turnover by promoting sustainable long-term matches.

Note. This table summarizes the nine KPIs grouped by hard skills, soft skills, and contractual/contextual dimensions, outlining their definitions and roles within the framework to support interpretable, multi-criteria, and governance-aware recruitment decisions.

Figure 5 shows a visual representation of the proposed methodological framework, with both the three pillars of alignment and the multi-layered structure of the multi-KPI method being clearly outlined. See Figure 5.



**Figure 5.** The Three Pillars and Layered Architecture of the Multi-KPI Recruitment Matching Framework. Note. The figure summarizes hard-skill precision, evidence-based soft skills, and operational feasibility within a

layered architecture, showing how semantic cores, governance constraints, and domain validation jointly support transparent, multi-criteria, and accountable job–candidate matching decisions.

The three pillars represent the three dimensions of compatibility that are explicitly addressed within the framework, namely hard skill precision, evidence-based soft skills, and operational feasibility, thereby reinforcing the view that “fit” is not just some latent construct, but an emergent property of the interplay between different, albeit complementary, dimensions of compatibility, in keeping with the latest advancements in recruitment research that advocate multi-dimensional models of assessment (Potočnik et al., 2024). With regard to the hard skill dimension, Figure X clearly illustrates the interplay between explicit requirement coverage and semantic similarity, thereby reinforcing the view that HSCR, SGI, and HSPS are part of an integrated framework of technical competence assessment, in keeping with the latest advancements in AI-based job matching methodologies, which combine both semantic similarity-based approaches with structured skill alignment methodologies (Ajjam & Al-Raweshidy, 2025). With regard to the behavioral dimension, the emphasis placed upon the interplay between declared and demonstrated soft skills, in keeping with the SSSA, SSED, and CTFS, is clearly illustrated, thereby reinforcing the view that evidence density and contextual alignment are critical dimensions of compatibility, with the latter being an emergent property of the interplay between different dimensions of compatibility, in keeping with the latest advancements in algorithmic decision-making, wherein both explainability and traceability of evidence are critical design dimensions of algorithmic decision-making processes (Mehrabani et al., 2021; Raghavan et al., 2020). The pillar of operational feasibility explicitly communicates that matching is not only subject to competence constraints but also to real-world constraints, as expressed by CCS, LMFI, and SCA, which filter and qualify plausible matches on the basis of contractual, geographical, and seniority-related constraints. The inclusion of contextual and governance-aware constraints reflects the increasing awareness of the importance of incorporating accountability and decision justification systems in recruitment systems, going beyond predictive accuracy (Raghavan et al., 2020). The lower section of the figure completes the conceptual breakdown of the system by depicting the multi-layered architecture of the system. The semantic core layer refers to the embedding-based components HSPS, SSSA, CTFS, which are intended to offer high recall rates and language robustness in matching unstructured text in CVs and job descriptions. This layer reflects the “semantic first” principle of the framework, which explains why embedding-based similarity is used as the primary matching function in the system (Ajjam & Al-Raweshidy, 2025). Above it, the governance and constraint layer reinstates the importance of auditability, feasibility, and accountability by means of discrete, rule-like KPIs such as HSCR, SGI, SSED, CCS, LMFI, and SCA, which ensure that matching recommendations are not only semantically plausible but also operationally feasible. The distinction between the predictive core and the governance layer reflects recent calls for fairness-aware and explainable AI systems in high-stakes applications such as hiring (Mehrabani et al., 2021; Raghavan et al., 2020). The domain validation layer finally emphasizes that the entire system is grounded in real-world data and applications, facilitating validation without affecting the core decision-making process, which reflects recent views on evidence-based recruitment systems and selection systems (Potočnik et al., 2024). The cumulative evidence substantiates the methodological argument that the framework represents an advancement over monolithic matching approaches by assembling diverse evidence types into a modular and understandable decision architecture. The image illustrates the manner in which ongoing semantic similarity signals, coverage and constraint checks, and evidence-based indicators collectively function as constituent elements of a monolithic system rather than mutually exclusive approaches. The image thus illustrates the set of KPIs presented in Table X as the functional core of a multi-criteria decision support process that is both explainable and governance-conscious and illustrates the manner in which semantic robustness, scalability, and auditability are achieved simultaneously through architectural separation and controlled integration of the constituent layers (Ajjam & Al-Raweshidy, 2025; Mehrabani et al., 2021; Raghavan et al., 2020).

*Data Sources and Dataset Construction.* The empirical analysis relies on two different data sources: one consisting of job offers retrieved from the EURES platform and the other consisting of candidate profiles retrieved from a prominent professional social networking platform that are publicly available on the web. The rationale for choosing these data sources is that it allows the analysis to be more ecologically valid since it relies on real-world data while at the same time adhering to the requirements of the current data recruitment pipeline that relies on authentic data sources such as job offers or candidate profiles (Frazzetto et al., 2025; Mashayekhi et al., 2022). Both the job offers and candidate profiles are retrieved using a keyword-based search query using the keyword “Data Science.” This approach allows the analysis to be more focused on a well-defined professional domain while at the same time allowing for the heterogeneity of job offers or candidate profiles that are available in the data science job market (Khaouja et al., 2021). The job offers are retrieved from the EURES platform, which offers structured as well as semi-structured job offers that include the skills required for the job, the responsibilities of the job, the contractual conditions of the job, and other relevant information pertaining to the job. This data is used to create the job representation layer of the proposed system. Pre-processing of the data is performed using various techniques such as text normalization, tokenization, and the removal of boilerplate content from the job offers (Khaouja et al., 2021; Mashayekhi et al., 2022). The candidates' profiles were obtained from publicly accessible professional profiles. The emphasis here was on information that is relevant to professional matching. The information obtained from the profiles was then encoded using information extraction and embedding-based encoding. The use of these techniques falls under the recent pipelines that integrate information extraction and semantic analysis for recruitment optimization. Importantly, the dataset was created strictly for research purposes and under the guidelines of relevant ethical and legal standards. The encoded representation does not enable any form of re-identification of the individuals and is strictly for the evaluation of the proposed matching framework. From the methodological point of view, the dataset supports the multi-dimensional character of the proposed framework. The dataset does not treat job postings and candidates' profiles as monolithic entities. Rather, both job postings and candidates' profiles are represented as composed of several semantic and functional elements. This enables the computation of different KPIs related to semantic similarity, skill coverage and gaps, behavioral evidence, and contextual and contractual compatibility. The use of these elements aligns with challenge-based evaluations of e-recruitment recommendation systems. The use of these elements aligns with challenge-based evaluations of e-recruitment recommendation systems (Mashayekhi et al., 2022). In conclusion, this dataset offers a realistic and diverse testbed to evaluate the proposed multi-KPI decision support architecture while ensuring alignment with reproducibility, transparency, and governance requirements in data-driven recruitment research (Frazzetto et al., 2025; Khaouja et al., 2021; Ifakir et al., 2025).

#### 4. Evaluating an Embedding-Based Semantic Pipeline for Job–CV Matching

The study aims to design and test a semantic pipeline in matching job descriptions and CVs. The study seeks to bridge the gaps in traditional recruitment systems that often use exact keyword matching. The study is in line with the latest advancements in embedding-based and transformer-based recruitment systems (Kurek et al., 2024; Li et al., 2025). The study focuses on evaluating the similarity between job descriptions and CVs at the level of meaning by using distributional representations of texts. The study follows the latest advancements in semantic textual relatedness in recruitment systems (Zadykian et al., 2025). The study uses an experimental dataset of approximately 300 job descriptions from EURES and 300 CVs in different formats (PDF, DOCX, and TXT). The study follows the latest advancements in using end-to-end pipelines in extracting valuable insights from candidates' profiles (Frazzetto et al., 2025). The study follows the latest advancements in using dense vector representations of texts by using a pretrained multilingual sentence embedding model. The study follows the latest advancements in computing semantic similarity between job descriptions and CVs using cosine similarity. The study follows the latest advancements in using established e-recruitment recommendation paradigms that use latent semantic representations (Mashayekhi et al.,

2022; Sukri et al., 2024). For each job posting, a ranked list of candidates is generated, with the top 10 candidates retained in the output. This transforms a complex, unstructured search space into a compact, semantically coherent output, making it easier for human decision-makers to process, thus improving outputs (Kurek et al., 2024). Quantitatively, the results offer strong evidence of the approach's efficacy and stability. As can be seen in Table X, the experiment covers 308 job offers, resulting in 3,080 job-candidate associations from the top-10 rankings, thus making it applicable to real-world scenarios, albeit for medium-sized scenarios, rather than small-scale scenarios (Mashayekhi et al., 2022). Furthermore, there are 149 unique candidates at least once in the top-10 rankings, thus making it unlikely that it is in danger of settling into a trivial solution with a limited number of candidate profiles, but rather offering a relatively diverse set of recommendations. With regard to the distribution of the similarity scores, it can be seen that the global mean is at 0.446, with a standard deviation of 0.051, thus making it a relatively moderate, yet significant, level of semantic alignment between recommended candidates, with a certain level of variability to discriminate between candidates. Furthermore, it can be seen that the minimum and maximum scores are at 0.294 and 0.715, respectively, thus making it a relatively comprehensive range of similarity, capturing both strong and weak semantic alignments, a phenomenon that is typical in transformer-based recruitment systems (Li et al., 2025). Furthermore, it can be seen that there is a certain level of quantitative validation in the proposed approach, with the average score between candidates in the rankings, from position one to position ten, diminishing from 0.491 to approximately 0.420, thus making it a coherent, well-ordered structure, rather than a flat distribution, a phenomenon that is typical in zero-shot, deep-learning-based job-candidate matching systems (Kurek et al., 2024). The above analysis of ranking separation also supports this finding. The average value of the gap between the first and tenth positions is 0.071, and the standard deviation of these gaps is 0.027. Thus, in most cases, the top-ranked profile will be considerably more relevant than the least relevant one among the top 10. The existence of both dense and less dense ranking distributions with dominant top-ranked profiles is also supported by the range of the gaps between the top and tenth positions. Such scenarios are quite common and align with real-world recruitment scenarios discussed in challenge-based evaluations of various recommendation systems (Mashayekhi et al., 2022). The statistics of the number of candidates' recurrences also support the above findings. The median value of these statistics is 7, and the maximum value is 234. Thus, there are some semantically central candidates that match a considerable number of job postings. The quantitative evidence supports the above qualitative assumption that semantic similarity may serve as an effective high-recall retrieval mechanism but may not be used as a standalone decision criterion. The latter issue may also be resolved by using the proposed system as part of the decision support system that combines various indicators and constraints (Frazzetto et al., 2025; Ifakir et al., 2025; Zadykian et al., 2025). Thus, the above numerical analysis supports the assumption that the proposed embedded semantic matching may serve as a strong quantitative baseline for the job-candidate matching problem and may serve as a good starting point for the development of the multidimensional decision support system that incorporates various indicators and constraints. Such decision support systems may also be developed using deep learning and graph-based recruitment systems (Li et al., 2025). See Table 4.

**Table 4.** Aggregated Statistics of the Semantic Job-CV Matching Experiment.

Category	Metric	Value	Description
Dataset Size	Number of job offers	308	Total number of job postings in the experiment
	Total job-CV matches (top-10)	3,08	10 candidates per job offer
	Unique candidates involved	149	Number of distinct CVs appearing in top-10 lists
Semantic Score (Global)	Mean semantic similarity	0.446	Average similarity over all job-CV pairs in top-10

Ranking Structure	Standard deviation	0.051	Variability of semantic similarity scores
	Minimum similarity	0.294	Lowest observed semantic similarity
	Maximum similarity	0.715	Highest observed semantic similarity
	Mean similarity at rank 1	0.491	Average score of the top-1 candidate per job
	Mean similarity at rank 2	0.470	Average score of the second-ranked candidate
	Mean similarity at rank 3	0.458	Average score of the third-ranked candidate
	Mean similarity at rank 5	0.445	Average score of the fifth-ranked candidate
Ranking Separation	Mean similarity at rank 10	$\approx 0.420$	Average score of the tenth-ranked candidate
	Mean gap (rank 1 – rank 10)	0.071	Average difference between top-1 and top-10 scores per job
	Std. dev. of gap	0.027	Variability of the top-1 vs top-10 separation
	Minimum gap	0.023	Smallest observed separation between top and bottom of top-10
Candidate Recurrence	Maximum gap	0.230	Largest observed separation between top and bottom of top-10
	Mean appearances per CV	20.7	Average number of times a CV appears in top-10 lists
	Median appearances per CV	7	Median recurrence of a CV in top-10 lists
	25th percentile	2	25% of CVs appear at most 2 times
	75th percentile	22	25% of CVs appear more than 22 times
	Maximum appearances	234	Most frequent CV in the top-10 rankings

Note. This table reports dataset size, global similarity statistics, ranking structure, separation between top-ranked candidates, and candidate recurrence patterns, summarizing the quantitative behavior of the embedding-based semantic job–CV matching pipeline.

## 5. Interpretable, Requirement-First Ranking in Recruitment: The Role of HSCR

The research suggests an approach to job-candidate matching based on KPIs. The HSCR plays a key role in this approach as a requirement-based KPI. The HSCR aims to offer a direct measure of how well the information given in a CV matches the explicit technical needs set forth in a job offer. This approach directly addresses the call for more transparent and manageable recruitment technologies (Nikolaou, 2021; Potočnik et al., 2024). The HSCR differs from semantic-based approaches that calculate an estimate of textual proximity. The HSCR measures “technical fitness” as a simple numerical value. This differs from embedding-based recruitment pipelines that heavily rely on distributional similarity (Ajjam & Al-Raweshidy, 2025). The HSCR can be defined as follows: Let  $(S_R)$  be a set of skills required by a job offer and  $(S_C)$  be a set of skills contained in a CV. The HSCR can be defined as  $(HSCR = |S_R \cap S_C| / |S_R|)$ , where  $(|S_R|)$  denotes the number of skills required by a job offer. The HSCR has a range of  $[0, 1]$ , where 0 means no coverage at all and 1 means full coverage. The HSCR is a ratio; therefore, it remains comparable across various job offers despite differences in the number of skills required by a job offer. In the context of the broader multi-KPI framework described in the article, which incorporates semantic signals, coverage indicators, and contextual dimensions, HSCR provides a key methodological component by introducing a system of explicit requirement control, which can counterbalance a system that could be dominated

by distributional similarity measures (Ajjam & Al-Raweshidy, 2025). While embedding-based approaches provide a robust system for achieving high recall, HSCR provides a system for ranking that is tied to verifiable technical constraints, which can be seen as a key component of transparency, governance, and explainability, all of which have become key components of hiring research in the context of AI hiring (Raghavan et al., 2020; Mehrabi et al., 2021). In terms of data, the article uses a dataset of real-world collections of job offers and CVs that exist in heterogeneous formats, such as PDF, DOCX, and TXT, which can be seen as a realistic system that aligns with the application of modern AI hiring systems (Van Esch et al., 2019). A key component of the system is a controlled vocabulary of hard skills, which includes programming languages, frameworks, platforms, and methodological terms, which serves as a lightweight ontology for the system, aligning with a structured approach to translating work histories into a measurable space, which has become a key component of translating applicant work histories into a measurable space, as described in Sajjadi et al. (2019). By using a combination of direct substring matching and lightweight fuzzy matching, the system is able to identify a range of skills for each document, which can be seen as a key component of translating unstructured text into a measurable space, which can be contrasted with the use of dense vectors, which can be seen as a key component of embedding-based approaches (Ajjam & Al-Raweshidy, 2025). Once the skill sets are extracted, HSCR is computed for each job-candidate pair. Candidates are ranked based on their ability to cover the required skills for a given job, and a top-10 list is created for each job offer, thus fulfilling the requirement-first selection criteria. The output is recorded in a structured table with job IDs, candidate IDs, ranking, and HSCR score. A score close to 1 signifies that all required skills for a particular job have been met, while lower scores indicate that some skills have been met, along with specific technical gaps. Since the HSCR score for each job-candidate pair can be broken down into the skills met and those that have not been met, this measure is not only quantitative but also explainable, thus fulfilling the need for human-in-the-loop decision-making while also aligning with the need for interdisciplinary research that calls for explainable and accountable screening mechanisms (Raghavan et al., 2020; Liem et al., 2018). Table X presents the aggregate statistics for the HSCR-based matching experiment. The data consists of 308 job offers, each with a top-10 list, thus creating 3,080 job-CV matches. Out of the rankings, 133 unique candidates exist, thus demonstrating that the system does not reduce to a trivial solution with a small number of profiles dominating the recommendation space. In fact, globally, HSCR values have a mean of 0.81 with a standard deviation of 0.36, covering the entire range from 0 to 1, thus demonstrating that this measure successfully captures the different levels of technical coverage, from those who do not meet even the explicit requirements to those who meet all the required skills. This relatively high mean value is consistent with the ranking strategy, which gives priority to higher coverage ratios within the top-10 lists. The ranking structure also provides additional evidence of the consistency of the HSCR-based ordering, with the mean HSCR value of rank 1 being close to 0.84, gradually decreasing to 0.79 at rank 10. This continuous decline in HSCR value with rank position confirms that the ranking system generates a meaningful ordering of the candidates, as opposed to generating flat scores. The small differences between consecutive rank positions suggest that, technically, most of the candidates are equivalent with regard to skill coverage, which is consistent with the labor market reality, where several profiles are expected to satisfy most of the essential demands of a given job position (Potočnik et al., 2024). Finally, additional interesting results are provided by the analysis of ranking separation, with the average gap between the first and tenth candidate being close to 0.042, accompanied by a standard deviation of 0.09. The minimum value of the gap is 0, indicating that, for some job offers, the ranking system identifies the top-10 lists where the coverage ratio of the top-ranked candidate is equal to that of the last-ranked candidate, suggesting the existence of technically equivalent profiles. The maximum value of the gap, equal to 0.4, refers to job offers where the coverage ratio of the top-ranked candidate is significantly higher than that of the other shortlisted candidates, thereby confirming the system's ability to identify clear winners, as well as situations where several technically equivalent profiles need to be evaluated using additional decision-making criteria, thereby supporting structured decision-making processes in an

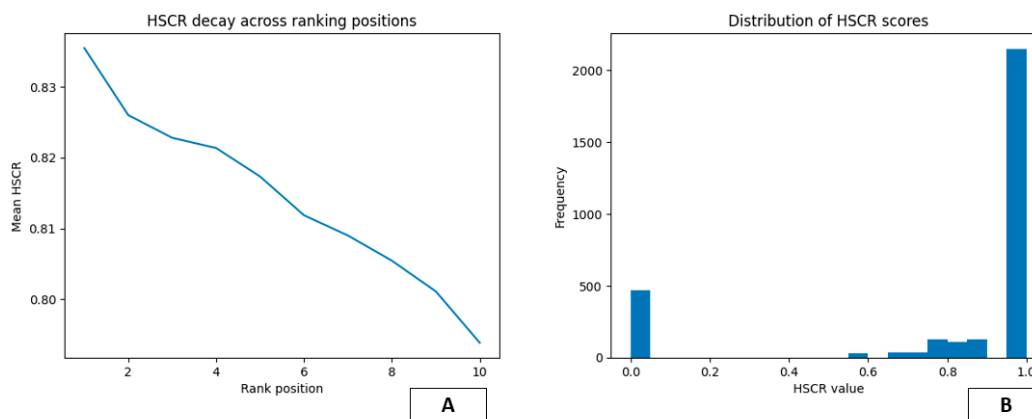
organizational context (Nikolaou, 2021; Raghavan et al., 2020). Lastly, the statistics on candidate recurrence are heavily skewed with a long tail. The average candidate occurs on the top-10 lists approximately 23 times, with a median of 8. Additionally, 25% of all candidates occur no more than 3 times. However, a small set of them occur extremely frequently with a maximum of 188. This suggests a set of a few generally compatible, technically “central” candidates and a set of more specialized candidates. The quantitative evidence supports HSCR-based matching as yielding structured, interpretable, and technically meaningful rankings. Thus, it supports HSCR-based matching as a solid, requirement-oriented component of a transparent multi-KPI recruitment decision support process (Ajjam & Al-Raweshidy, 2025; Mehrabi et al., 2021). See Table 5.

**Table 5.** Aggregate Statistics of the HSCR-Based Requirement-First Matching Experiment.

Category	Metric	Value
Dataset Size	Number of job offers	308
	Total job–CV matches (top-10)	3080
	Unique candidates involved	133
HSCR (Global)	mean	0.814413652
	std	0.35555628
	min	0
	max	1
Ranking Structure	Mean HSCR at rank 1	0.835456607
	Mean HSCR at rank 2	0.82598691
	Mean HSCR at rank 3	0.822786539
	Mean HSCR at rank 4	0.821343537
	Mean HSCR at rank 5	0.817322459
	Mean HSCR at rank 6	0.811848073
	Mean HSCR at rank 7	0.80899428
	Mean HSCR at rank 8	0.805438312
	Mean HSCR at rank 9	0.801109307
	Mean HSCR at rank 10	0.793850495
Ranking Separation	Mean gap (rank 1 – rank 10)	0.041606112
	Std. dev. of gap	0.090059743
	Minimum gap	0
	Maximum gap	0.4
Candidate Recurrence	Mean appearances per CV	23.15789474
	Median appearances per CV	8
	25th percentile	3
	75th percentile	25
	Maximum appearances	188

Note. This table summarizes dataset scale, global HSCR distribution, ranking behavior, separation, and candidate recurrence, illustrating how explicit skill extraction enables interpretable, requirement-driven ranking and supports transparent, auditable, human-in-the-loop recruitment decisions.

Hence, Figures 6 A and B present a different look at the behavior of the HSCR-based ranking system, describing the characteristics of the developed ranking list as well as the overall distribution of skill-coverage scores for all possible job-candidate pairs within the top-10 list. See Figure 6.



**Figure 6.** HSCR Ranking Decay and Score Distribution Across Top-10 Job–Candidate Matches. Panel A shows monotonic HSCR decay across ranks, confirming coherent requirement-based ordering; Panel B shows a skewed score distribution with peaks at extremes, illustrating coverage behavior and motivating complementary criteria within multi-KPI recruitment decision support.

Figures A and B confirm the numerical data presented in Table 5, helping to better understand the HSCR measure from a practical perspective, matching modern criteria for evaluating e-recruitment recommendation system rankings (Mashayekhi et al., 2024). Figure A shows the average behavior of HSCR as a function of ranking position. It can be seen that the curve has a monotonically decaying nature from rank 1 to rank 10, with the maximum average HSCR at rank 1 gradually decreasing for lower ranks. This behavior also supports the correctness of the implementation of the ranking model, with a larger fraction of skills being covered by candidates with higher ranks. This behavior is also expected for a structured ranking model of job candidates (Kurek et al., 2024). The smooth nature of the curve also indicates that the average HSCR values for different ranks vary only slightly, suggesting a large number of technically equivalent top-10 candidates. This is a plausible phenomenon, especially for a professional recruitment context, where a large number of candidates may meet basic requirements (Potočnik et al., 2024). This behavior is also desirable from a decision-support perspective, as it indicates that there is a single best candidate, with a small number of near-best alternatives. Figure B displays a histogram showing the distribution of HSCR score values for all pairs of jobs and candidates included in the top-10 lists. It is noticeable that the distribution is highly skewed, with two peaks at score values 1.0 and 0.0, respectively. The first indicates that top candidates within the top-10 lists have full or near-full coverage of the required skills, which further validates the effectiveness of the proposed HSCR-based ranking system. On the contrary, candidates with score values near 0.0 indicate that some bottom candidates within the top-10 lists fail to cover at least one explicitly defined required skill, which often happens with jobs that have very sparse skill sets defined in their descriptions. This is a known limitation with coverage- or extraction-based systems, which often depend on entity recognition and preprocessing steps, which, in turn, depend on the quality of entity recognition itself, as indicated by Ifakir et al. (2025). Figures A and B, therefore, indicate that the proposed coverage-based system, while being effective, also presents some limitations that make further selection criteria important, especially for cases with candidates having similar levels of coverage or with jobs having poorly defined descriptions, as indicated by Mashayekhi et al. (2024).

## 6. HSPS as a High-Recall Semantic Signal for Job–Candidate Matching

The present study proposes Hard Skill Proficiency Similarity (HSPS) as a supplementary measure to Hard Skill Coverage Ratio (HSCR). The proposed HSPS aims at capturing a specific aspect of technical fit that cannot be measured by the presence or absence of a list of skills, which is what HSCR does. While HSCR offers a specific, requirement-based measure of how well a set of required

skills is represented within a CV, HSPS offers a continuous, quantitative perspective on technical fit, which is consistent with recent embedding-based models of job-candidate matching (Kurek et al., 2024; Li et al., 2025). More specifically, HSPS is proposed as a function of cosine similarity between dense representations of technical information contained within a set of job offers and a CV, which is derived by a multilingual pretrained transformer model, thereby allowing for semantically meaningful comparison across languages and writing styles, as recent semantic textual similarity models for recruitment propose (Zadykian et al., 2025; Sukri et al., 2024). From a computational perspective, we normalize the dense representations by L2 norm, which means that cosine similarity is computationally equivalent to a dot product, allowing for efficient computation of the entire job-CV similarity matrix by means of a matrix multiplication, which is essential for scalability to datasets containing hundreds of job postings and CVs, as recent e-recruitment models propose (Mashayekhi et al., 2024). In the experiment that is reported, the system processes 308 job offers and produces 3,080 job-candidate associations by choosing the top 10 candidates for each job. This is the scale at which the HSPS can be applied realistically, as shown in the experiment, aligning with the application of the deep learning recruitment pipeline (Li et al., 2025; Frazzetto et al., 2025). The quantitative distribution of the HSPS scores offers valuable insights into the performance of the HSPS metric. The scores range globally from a minimum of 0.164 to a maximum of 0.734, with an average of 0.528 and a standard deviation of 0.075. This reflects a moderate-to-high level of semantic similarity between the shortlisted candidates. The range of the scores reflects the ability of the HSPS to distinguish between poor and good technical alignments of the candidates, a characteristic that is common in semantic ranking using embeddings (Kurek et al., 2024). Lower scores indicate that the candidate's technical narrative is only loosely related to the technical aspect of the job posting, while higher scores indicate substantial overlap between the domains, tools, and conceptual stacks used. The lack of scores closer to 1.0 is also expected because the job description and the candidate's CV are unlikely to match entirely (Sukri et al., 2024). The ranking of the HSPS scores also supports the internal consistency of the HSPS-based candidate matching process. The mean HSPS at the first rank is 0.578, while the scores decrease monotonically to 0.497 at the 10th rank. This indicates that the HSPS consistently ranks the semantically more aligned candidates at the top of the ranked candidate pool for each job posting, a characteristic that is common in the transformer-based recruitment pipeline (Li et al., 2025). Although the slope of the HSPS scores is declining as the ranks decrease from 1 to 10, the slope is not steeply declining. This indicates that semantically similar candidates are more common at the top of the ranked pool of candidates for the job posting, aligning with the evaluation of the recruitment recommendation system (Mashayekhi et al., 2024). Additional evidence is provided from the analysis of ranking separation, in which it is found that on average, there is a gap of approximately 0.081 between the first-ranked candidate and the tenth-ranked candidate, with a standard deviation of 0.039. Furthermore, it is found that there is a minimum gap of approximately 0.020, in which all the top-10 candidates are semantically similar, forming tight clusters of candidate profiles. On the other hand, it is found that there is a maximum gap of approximately 0.293, in which the top-ranked candidate is significantly more aligned with the job requirements compared to all the remaining candidates in the top-10 list, thus indicating that HSPS can both identify strong leaders and reveal scenarios in which all candidates in the top-10 list are almost equivalent, thus warranting further evaluation with additional criteria, including components related to extraction or explainability (Ifakir et al., 2025; Zadykian et al., 2025). The recurrence statistics of the candidates provide further insight into how the semantic model distributes the matches among the candidates. The average recurrence count among the 208 distinct candidates appearing in the top-10 lists is 14.8, with a median recurrence count of 5. The recurrence count at the 25th percentile is only 2. This again suggests a highly skewed distribution where a small set of broadly relevant candidates are seen to recur frequently. The maximum recurrence count of 122 further supports the presence of a small set of "hub" profiles semantically close to a large range of technical roles. This long-tail distribution is expected in embedding-based retrieval systems. The above statistics support our view of HSPS as a high-recall meaning-based filtering system rather than a constraint-based one (Kurek et al., 2024;

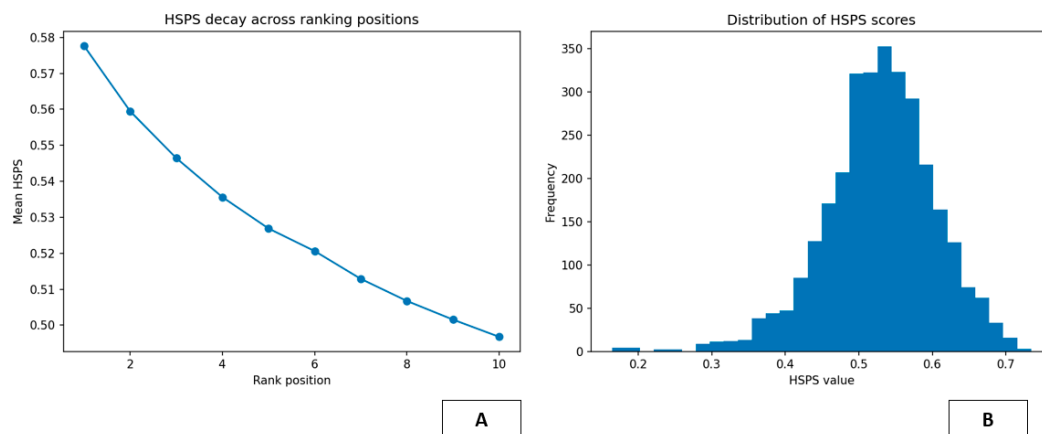
Frazzetto et al., 2025). The quantitative statistics support our view of HSPS as an effective ranking criterion. The numerical properties of the HSPS metric are well-structured score distributions with coherent rankings showing smooth decay across ranks. The HSPS also provides meaningful differentiation among candidates in both clustered and polarized scenarios. The numerical properties of HSPS further support our view of HSPS as an effective measure of technical domain proximity with sufficient variability among jobs and candidates. The HSPS acts as a robust semantic component of a multi-KPI-based data-driven recruitment decision support system (Mashayekhi et al., 2024; Li et al., 2025). See Table 6.

**Table 6.** Aggregate Statistics of the HSPS-Based Semantic Job–Candidate Matching Experiment.

Category	Metric	Value
Dataset Size	Number of job offers	308
	Total job–CV matches (top-10)	3080
	Unique candidates involved	208
HSPS (Global)	Mean	0.528452
	Standard deviation	0.075312
	Minimum	0.164417
	Maximum	0.734151
Ranking Structure	Mean HSPS at rank 1	0.577612
	Mean HSPS at rank 2	0.559451
	Mean HSPS at rank 3	0.546433
	Mean HSPS at rank 4	0.535588
	Mean HSPS at rank 5	0.526878
	Mean HSPS at rank 6	0.520601
	Mean HSPS at rank 7	0.512881
	Mean HSPS at rank 8	0.506745
	Mean HSPS at rank 9	0.501556
	Mean HSPS at rank 10	0.496773
Ranking Separation	Mean gap (rank 1 – rank 10)	0.080838
	Std. dev. of gap	0.038557
	Minimum gap	0.019954
	Maximum gap	0.292521
Candidate Recurrence	Mean appearances per CV	14.80769
	Median appearances per CV	5
	25th percentile	2
	75th percentile	15
	Maximum appearances	122

Note. This table reports dataset scale, global HSPS distribution, ranking behavior, separation, and candidate recurrence, illustrating scalable embedding-based similarity computation and confirming HSPS as a high-recall semantic signal for ranking technical domain proximity.

Figure 7 propose an alternative point of view on the behavior of the Hard Skill Proficiency Similarity (HSPS) key performance indicator used in the candidate ranking process.



**Figure 7.** HSPS Ranking Decay and Score Distribution Across Top-10 Job–Candidate Matches. Note. Panel A shows monotonic HSPS decay across ranks, indicating coherent semantic ordering; Panel B shows a near-normal score distribution, confirming HSPS as a continuous, high-recall measure of technical similarity supporting stable, interpretable ranking in recruitment decision support systems.

Figure A shows the plot of the mean HSPS value as a function of the candidate’s position in the ranking process. The HSPS value monotonically decreases from the highest position down to the tenth position. This means that the candidates placed at the top of the ranking tend to have a higher semantic similarity to the content of the technical offer, while the candidates placed at the bottom of the ranking tend to have a reduced semantic similarity. This behavior is consistent with the internal consistency of the ranking process. Moreover, the behavior of the HSPS value is consistent with the evaluation behavior of the state-of-the-art e-recruitment recommendation systems (Mashayekhi et al., 2024). The behavior of the HSPS value is not random; instead, the algorithm defines the ranking along a continuous range of technical proximity. This behavior has also been observed in zero-shot or embedding-based job matching algorithms (Kurek et al., 2024). The smooth slope of the curve also implies that the top candidates tend to be technically comparable. This behavior has also been observed in the recruitment algorithms based on the transformer architecture, where semantic similarity plays the role of a graded measure of the ranking (Li et al., 2025). The behavior of the HSPS value from the point of view of the decision-making process is also satisfactory because the algorithm does not propose a single winner; instead, it proposes a set of top candidates that are technically proximal, thus allowing the decision-making process to play its role in the final evaluation of the candidates. This behavior is also consistent with the paradigm of the explainable semantic matching (Zadykian et al., 2025). Figure B illustrates the range of HSPS score value distribution among all job-candidate pairs contained within the top-10 lists. The histogram illustrates a normal distribution curve with most score values concentrated in the mid-to-high similarity range and tapering off as the score value increases. This supports the assertion that HSPS measures a range of technical similarity and does not fall into the binary ‘good’ and ‘bad’ categories as might be expected from legacy recruitment systems. The clustering of score values around a specific region of the y-axis supports the assumption that most profiles will likely have similar domain knowledge and tooling and/or project descriptions, even if not exactly the same. This is similar to large-scale semantic ranking systems (Kurek et al., 2024). The left-hand side of the histogram illustrates that although contained within the top-10 lists, some candidates have lower levels of similarity with the job’s technical focus. This phenomenon is more common in job descriptions that are less descriptive. Figures A and B collectively illustrate that HSPS supplies a soft but useful semantic signal that supplies a stable ordering of candidates and clusters similar candidates while also offering a natural and continuous score value distribution. The latter phenomenon is particularly important when considering decision support systems and the requirement to support human decision-making with a logical and

understandable listing of candidates based on their semantic technical similarity (Zadykian et al., 2025; Mashayekhi et al., 2024).

## 7. From Skill Coverage to Skill Gaps: Introducing SGI as a Complement to HSCR

The study also introduces a new concept, Skill Gap Index (SGI), which is a natural and necessary extension of HSCR. The shift in focus changes from "how well an applicant already meets the defined requirements" to "how much still needs to be addressed with respect to the clearly defined criteria outlined in a job offer." The quantification of what is still missing is a direct result of HSCR, and SGI is defined as one minus HSCR. Therefore, SGI also inherits the positive numerical properties of HSCR, including that its range will be between 0 and 1, that comparisons across different job offers with different numbers of requirements will be possible, and that it will still be easy to interpret. An SGI of 0 means that an applicant is fully meeting the requirements outlined, while an SGI close to 1 means that there is a growing gap between the applicant's skills and the required skills. This reformulation of HSCR into SGI also has important implications because, with this approach, the skills mismatch is quantified and can be related to training initiatives that aim at reducing this mismatch. This approach is also related to other numerical approaches to quantify skills shortages and mismatch in the labor market (Bachmann et al., 2020; Chinn et al., 2020). Instead of applying a binary suitable/unsuitable test criterion on an applicant's profile, recruiters might now apply SGI to quantify how dissimilar an applicant's profile is with respect to a desired profile. The quantification approach is based on a dataset containing 308 job offers, with a corresponding top-10 shortlist, resulting in a total number of 3,080 job-CV pairs and 157 distinct candidates. This setting represents a realistic medium-scale recruitment scenario and is comparable with datasets that have been used for evaluations of state-of-the-art e-recruitment recommendation systems (Mashayekhi et al., 2024; Li et al., 2025). From a global statistical point of view, the results indicate that the measure has an average of 0.675 and a standard deviation of 0.408. The extremal values indicate that the measure does not collapse to a small range of scores. The identification of skills used in the evaluation of HSCR follows contemporary methods of skill identification from job offers and CVs (Khaouja et al., 2021; Ifakir et al., 2025; Frazzetto et al., 2025). Similarly, SGI, which represents the complement of HSCR, follows the same distribution but with an average of 0.325 and the same standard deviation of 0.408. The results cover the entire range from 0 to 1. From an operational point of view, these results indicate that on average, candidates from the top-10 shortlisted matches do not meet one-third of the explicitly required skills. The extremal values indicate that the measure does not collapse to a small range of scores. The range of the results indicates that some matches have SGI close to 0 while others have SGI close to 1. The range of SGI indicates that not only does SGI indicate the suitability of candidates but also estimates the required effort to align the candidate with the job by means of training or re-skilling. The results resonate with the importance of skill-gap measurement at the organizational and policy levels (Bachmann et al., 2020; Chinn et al., 2020). The ranking structure further supports the validity of the HSCR ordering, with the mean HSCR for rank 1 being approximately 0.724, which gradually decreases to approximately 0.675 for rank 5. This gradual decrease supports the premise that the system places candidates with higher HSCR at the top of the list, which is in line with the evaluation framework of the recruitment recommendation system (Mashayekhi et al., 2024). The relatively small slope also supports the premise that there are many technically equivalent candidates at the top of the list, which is in line with real-world labor markets where many candidates meet the majority of the core requirements of the job (Li et al., 2025). Further insight can be gained by examining the ranking separation, with the average distance between the first-ranked candidate and the tenth-ranked candidate being approximately 0.094, with a standard deviation of approximately 0.144. The minimum distance of 0 signifies cases where the top 10 candidates are technically equivalent, with no difference in terms of skill sets, while the maximum distance of approximately 0.667 signifies cases where one of the candidates is significantly better than the rest of the candidates in terms of required skills. The numerical evidence supports the premise that HSCR, in combination

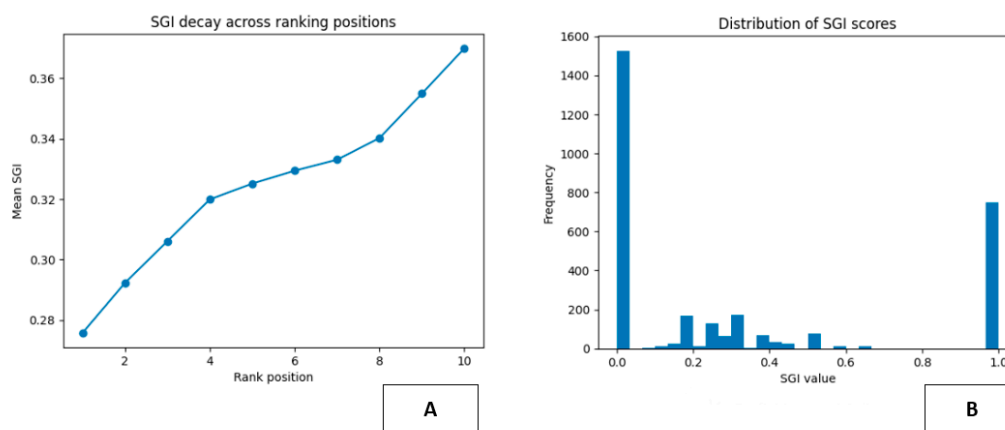
with SGI, provides a clear and quantitatively actionable representation of technical fit, with the numerical evidence supporting the premise that SGI has the potential to support both strict selection strategies as well as training-oriented strategies, which complements the coverage-based and semantic representation in providing support for the recruitment process, as outlined by Mashayekhi et al. (2024) and Khaouja et al. (2021). See Table 7.

**Table 7.** Aggregate Statistics of the SGI-Based Skill Gap Analysis in Job–Candidate Matching.

Category	Metric	Value
Dataset Size	Number of job offers	308
	Total job–CV matches (top-10)	3080
	Unique candidates involved	157
	Mean	0.675351
	Standard deviation	0.408481
	Minimum	0.000000
	Maximum	1.000.000
SGI (Global)	Mean	0.324649
	Standard deviation	0.408481
	Minimum	0.000000
	Maximum	1.000.000
Ranking Structure	Mean HSCR at rank 1	0.724298
	Mean HSCR at rank 2	0.707747
	Mean HSCR at rank 3	0.693948
	Mean HSCR at rank 4	0.680104
	Mean HSCR at rank 5	0.674884
Ranking Separation	Mean gap (rank 1 – rank 10)	0.094203
	Std. dev. of gap	0.143568
	Minimum gap	0.000000
	Maximum gap	0.666667

Note. This table reports dataset scale, global SGI distribution, ranking structure, and separation, showing how SGI complements HSCR by quantifying residual skill gaps, supporting both selection decisions and training-oriented strategies in recruitment pipelines.

Figure 8 provides complementary insights into the behavior of the Skill Gap Index (SGI) within the ranking produced by the HSCR-based matching pipeline.



**Figure 8.** GI Ranking Trend and Skill-Gap Score Distribution Across Top-10 Job–Candidate Matches. Panel A shows monotonic SGI increase across ranks, indicating growing skill gaps; Panel B displays a polarized

distribution with peaks near zero and one, evidencing heterogeneous mismatch patterns and supporting SGI as a graded, decision-support signal.

Figure A reports the average SGI value as a function of the ranking position, while Figure B shows the distribution of SGI scores across all job–candidate pairs included in the top-10 shortlists. Together, these visualizations illustrate both the internal coherence of the ranking mechanism and the overall structure of skill gaps observed in the recommended matches, in line with evaluation practices discussed in contemporary e-recruitment recommendation systems (Mashayekhi et al., 2024). The curve in Figure A exhibits a clear and monotonic increase of mean SGI values from rank 1 to rank 10. This pattern is the expected mirror image of the HSCR decay: candidates placed at the top of the ranking are characterized by smaller skill gaps, whereas lower-ranked candidates progressively exhibit larger gaps with respect to the explicit technical requirements of the job offers. The smoothness of the curve suggests that the ranking is not driven by abrupt threshold effects, but rather reflects a gradual degradation of technical completeness as one moves down the shortlist. Similar gradient-based ranking behaviors have been observed in deep learning–based resume–job matching systems (Li et al., 2025). This behavior is particularly desirable in decision-support contexts, as it indicates that the system orders candidates along a meaningful continuum of residual training needs rather than producing unstable or noisy rankings. Figure B complements this view by showing the empirical distribution of SGI scores. The histogram reveals a strongly polarized structure, with a large concentration of values near zero and another prominent mass close to one, together with a smaller but non-negligible density of intermediate values. The peak near zero corresponds to candidates who fully, or almost fully, satisfy the explicitly detected skill requirements, and therefore exhibit negligible or very small gaps. Conversely, the peak near one represents cases in which candidates lack most of the required skills, despite being included in the top-10 lists due to the relative scarcity or ambiguity of requirements in some job descriptions. The identification and quantification of such gaps depend on structured skill extraction from job ads and CVs, as highlighted in recent surveys on skill identification (Khaouja et al., 2021) and NER-based recruitment optimization pipelines (Ifakir et al., 2025). The intermediate region reflects partial matches, where candidates cover a substantial subset of the required competencies but still require non-trivial upskilling. Taken together, these figures highlight the dual role of SGI in the proposed framework. On the one hand, its monotonic increase across ranks confirms that it provides a consistent ordering signal aligned with the notion of technical distance from the target role, consistent with structured evaluation approaches in e-recruitment systems (Mashayekhi et al., 2024). On the other hand, its bimodal and dispersed distribution underscores the heterogeneity of real-world matching scenarios, where both fully aligned and strongly misaligned profiles can coexist, reinforcing the need for combining SGI with semantic and contextual KPIs in a multi-criteria decision process (Li et al., 2025).

## 8. SSSA as a Continuous Measure of Behavioral Compatibility in Job Matching

The study proposes the Soft Skill Semantic Alignment KPI, which extends the matching framework to cover more than just technical competencies, specifically incorporating the behavioral and interpersonal aspects, which are crucial in real-world recruitment scenarios. While technical skills are essential to determine one’s technical qualification for the position, soft skills affect how well the person will fit into the team, communicate, lead, etc., as has been extensively proven by organizational psychology literature (Dalal et al., 2020). Since soft skills are difficult to communicate in any standard way, SSSA proposes a semantic, metric-based way of modeling the soft skills dimension, which is in line with the new digital talent signals being used in personnel evaluation (Chamorro-Premuzic et al., 2016), as well as the increased reliance on technology in the hiring process (Nikolaou, 2021). In practical terms, SSSA is given by the cosine similarity between the dense vector representation of soft skill-related content in the job description and the candidate profile. The vector representation of the content is generated using a sentence embedding model, with the resulting score being a continuous number measuring the degree of behavioral and interpersonal matching. This

kind of vector representation is in line with recent advances in transformer-based recruitment systems (Li et al., 2025), as well as other models of semantic textual relatedness, which are being developed to support explainable matching (Zadykian et al., 2025). In its nature, SSSA is normalizable, comparable, and suitable for ranking-based retrieval, which is in line with the evaluation metrics of e-recruitment recommendation systems (Mashayekhi et al., 2024). This way, the soft skill dimension can be treated as measurable and gradable, rather than being based on subjective, qualitative evaluation. From the point of view of the dataset, the experiment covers 308 job offers, each with its own top 10, resulting in 3,080 job/CV matches. The total number of unique individuals within the ranking is 203, suggesting that the metric does not focus its recommendations on an extremely narrow subset of profiles but, rather, covers them over a fairly broad range. Numerical diversity is particularly relevant in soft skill analysis, as behavioral characteristics are often more varied than technical skills, especially in different environments, as argued in Dalal et al. (2020). The global distribution of the SSSA metric scores provides the first quantitative characterization of the metric. The scores are distributed with an average of 0.56, standard deviation of 0.065, suggesting that the metric captures a fairly high degree of semantic alignment between the soft skill narratives of job descriptions and candidate profiles. The minimum scores are close to 0.35, and the maximum scores are close to 0.75, suggesting that the metric covers a fairly good range of scores, both low and high, in terms of behavioral alignment between the two types of documents. The absence of scores close to the extremes of 0 or 1 makes sense, as soft skills are rarely described in perfectly overlapping or perfectly non-overlapping ways in actual documents. Instead, the distribution of scores reflects a continuous concept of alignment, which is precisely what the metric is intended to capture, echoing the behavior of other semantic recruitment pipelines using deep learning approaches, as in Li et al. (2025) and Mashayekhi et al. (2024). The ranking structure further supports the internal consistency of the SSSA-based ranking. The average SSSA score of the first-ranked candidates is approximately 0.61, which gradually decreases to about 0.56 for the fifth-ranked candidates. This shows that the score consistently ranks more behaviorally aligned candidates higher, which is consistent with the overall ranking dynamics of embedding-based recruitment systems, as suggested by Mashayekhi et al. (2024). At the same time, the gradual nature of the score drop also suggests that many of the candidates in the higher ranks of the ranking are relatively equally aligned in their behavioral responses, which is similar to real-world recruitment scenarios where several candidates may have similar behavioral narratives on teamwork, communication, or leadership, even though they may differ in the exact wording of their answers, as suggested by Nikolaou (2021). Further insights can be gained by examining the ranking separation analysis. The average gap between the first-ranked and the tenth-ranked candidates is approximately 0.078, with a standard deviation of approximately 0.027. This shows that, on average, the first-ranked candidate is more aligned with regard to soft skills compared to the lower-ranked candidates, although not by an extremely large margin. The minimum gap of about 0.02 shows that, in some cases, the top 10 candidates may be almost equally aligned with regard to soft skills, while the maximum gap of about 0.15 shows that, in other cases, there may be a clear winner with regard to behavioral matching. In terms of quantifying the overall ranking, it can be said that SSSA can identify clear leaders as well as clusters of equally good candidates, which can be useful for more explainable matching approaches, as suggested by Zadykian et al. (2025). Lastly, the statistics on candidate recurrence provide insight into how SSSA's metric spreads out across the CV set. The median here is at 5 occurrences, with an interquartile range spanning from 2 to 15. While we see a degree of skewing here, we also note that certain candidates recur frequently. The highest recurrence rate we observed was at 122. This indicates the existence of "behaviorally versatile" individuals whose soft-skill stories are more universally applicable to various work scenarios. The long tail of this curve further supports SSSA's interpretation as a high-recall ranking system rather than a filtering one (Li et al., 2025; Mashayekhi et al., 2024). Thus, we find from our numerical data that SSSA offers a stable measure of behavioral and interpersonal compatibility. The distributions of SSSA's score, ranking properties, and separation statistics all support SSSA's interpretation as a continuous measure of soft-skill compatibility. As such, SSSA remains an essential

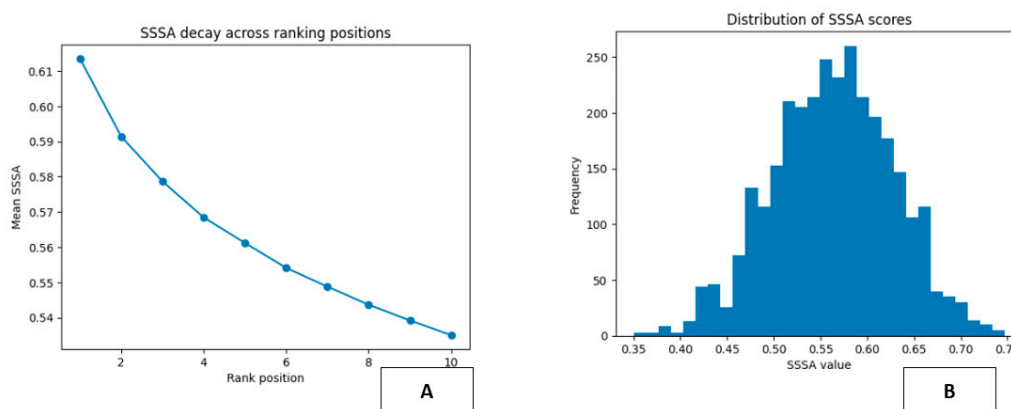
quantitative component of a multi-KPI-based human-centric recruitment decision support system (Nikolaou, 2021; Mashayekhi et al., 2024). See Table 8.

**Table 8.** Aggregate Statistics of the SSSA-Based Soft Skill Semantic Alignment Experiment.

Category	Metric	Value
Dataset Size	Number of job offers	308
	Total job–CV matches (top-10)	3080
	Unique candidates involved	203
SSSA (Global)	Mean	0.563406
	Standard deviation	0.064695
	Minimum	0.350591
	Maximum	0.747142
Ranking Structure	Mean SSSA at rank 1	0.613531
	Mean SSSA at rank 2	0.591410
	Mean SSSA at rank 3	0.578668
	Mean SSSA at rank 4	0.568449
	Mean SSSA at rank 5	0.561198
Ranking Separation	Mean gap (rank 1 – rank 10)	0.078486
	Std. dev. of gap	0.027387
	Minimum gap	0.020400
	Maximum gap	0.151449
Candidate Recurrence	Mean appearances per CV	1.480.769
	Median appearances per CV	5
	25th percentile	2
	75th percentile	15
	Maximum appearances	122

Note. This table reports dataset scale, global SSSA distribution, ranking behavior, separation, and candidate recurrence, demonstrating SSSA as a continuous, high-recall semantic measure of behavioral and interpersonal compatibility within a multi-KPI recruitment decision support framework.

Figure 9 provides different insights into the performance of the Soft Skill Semantic Alignment (SSSA) key performance indicator within the candidate ranking process.



**Figure 9.** SSSA Ranking Decay and Score Distribution Across Top-10 Job–Candidate Matches. Panel A shows monotonic SSSA decline across ranks, indicating decreasing behavioral alignment; Panel B shows an unimodal,

mid-range distribution, confirming SSSA as a continuous, explainable soft-skill similarity signal supporting nuanced ranking among technically comparable candidates.

Figure A shows how the average value of SSSA changes with the rank of the candidate. The curve shows that, on average, the SSSA value decreases monotonically from rank 1 to rank 10, implying that higher-ranked candidates tend to have higher levels of semantic alignment with the behavioral and interpersonal expectations outlined within the job description, while lower-ranked candidates tend to have lower levels of semantic alignment with the job description. This shows consistency within the ranking process, an essential attribute of technology-enabled ranking processes, as outlined within the framework of candidate ranking in technology-enabled settings (Nikolaou, 2021). Furthermore, the process is not random, with the ranking of the candidates based upon their continuum of soft skills, an attribute of the principles of transformer-based semantic matching, as outlined within the study of Li et al. (2025). It is, of course, important to note that although the curve is smooth, implying continuous reduction of SSSA with increasing rank, the rate of decline is shallow, implying that although the SSSA value changes with increasing rank, the differences are not significant. For example, some of the top-ranked candidates have almost identical behavioral characteristics, an attribute of the real world, where several individuals may have almost identical communication skills, soft skills, or leadership attributes, as outlined within their personal narratives or résumés, respectively. This, therefore, shows that within the real world, several almost equivalent individuals are expected to be produced, an attribute of ranking processes, as outlined within several recent studies of semantic textual relatedness within the candidate ranking process (Zadykian et al., 2025). In terms of decision-making, therefore, an individual top-ranked candidate may be produced, with several almost equivalent individuals available for consideration within the recruitment process, an attribute of ranking processes, as outlined within several recent studies of ranking within technology-enabled settings (Nikolaou, 2021). The overall distribution of all SSSA scores for all job-candidate pairs within the top-10 shortlists is given in Figure B. From the histogram, it can be seen that there is an unimodal distribution of SSSA scores with most of the scores concentrated in the middle to upper range of the semantic alignment spectrum, with fewer scores in the extremes. This demonstrates that SSSA indeed measures a continuous construct of soft skill compatibility, as opposed to a hard distinction between “good” and “poor” soft skill compatibility, which is different from other embedding-based approaches to recruitment (Li et al., 2025; Mashayekhi et al., 2024). Moreover, it can be noted that, on average, most candidates will have moderate to good soft skill compatibility with the job description, with most of the scores being clustered closely together in the middle of the spectrum. However, it can also be noted that there are some lower-end SSSA scores in the top-10, implying that some candidates will have lower behavioral or interpersonal similarity with the job description, particularly if one includes broad definitions of soft skills. The figures have substantiated that SSSA indeed behaves as a soft, ranking metric with a monotonic relationship to rank, as can be expected of an explainable semantic matching model (Zadykian et al., 2025). Indeed, the SSSA score behaves in a way that is compatible with its usage as a multi-criteria decision support tool, which can make distinctions between technically equivalent candidates based on soft skill compatibility (Mashayekhi et al., 2024; Nikolaou, 2021).

## 9. Measuring What Supports Soft Skills: The Soft Skill Evidence Density Metric

The Soft Skill Evidence Density metric was proposed with the aim of moving away from a keyword analysis approach towards an evidence-based approach for soft skills evaluation, with a quantification of the degree of support available for the soft skills listed on an individual's curriculum vitae. The proposed metric does this by quantifying the degree of evidence available substantiating the skills listed on an individual's curriculum vitae, such as “teamwork,” “leadership,” and “communication.” Each element of evidence that substantiates the skills listed on an individual's curriculum vitae, for example, descriptions of cross-functional projects that an individual was a part of, descriptions of conflicts that were resolved by an individual, mentoring experiences, scenarios

that involve stakeholder management, and feedback received, contributes towards an individual's Soft Skill Evidence Density score. Thus, an individual's Soft Skill Evidence Density score is continuous and takes into account not only the availability of soft skills but also the degree of support available for the skills listed on an individual's curriculum vitae. With regard to the benefits that SSED offers over the state of the art in keyword analysis for curriculum vitae-based evaluation, the main advantage that SSED offers over existing approaches is that SSED's signal quality is better than that which is available with keyword analysis or semantic analysis-based approaches for soft skills evaluation. The problem with keyword analysis or semantic analysis for soft skills evaluation is that an individual's score can be maximized by using templates, which is a problem that has been extensively discussed with regard to recruitment technology. SSED, therefore, can be seen as a step up from keyword analysis, given that it provides a measurable assessment of soft skills that is based on the level of evidence provided for the skills referred to in the curriculum vitae presented by the candidates. This makes it easier to discriminate among candidates, thus providing a framework for using SSED as a tool for ranking the shortlisted candidates (Ifakir et al., 2025; Li et al., 2025). SSED also provides a framework for comparing the candidates more adequately. Although both profiles may indicate that "communication skills" is a key skill, the profile of one candidate may indicate three different instances of such skills, while the profile of another candidate may indicate only a single instance of the skills. Based on SSED, the former will be ranked higher than the latter since more evidence is provided for the skills referred to in the profile. This ability of SSED to create quantitative distinctions from qualitative ones also makes it easier to be used for ranking. Moreover, the quantitative scale used for evaluating the level of evidence associated with the skills on the curriculum vitae makes it easier to carry out comparisons among candidates within the e-recruitment context (Mashayekhi et al., 2024). The quantitative nature of SSED is also confirmed by the fact that experiments carried out on a large dataset consisting of 308 job offers, each with a corresponding top 10 ranking list, resulted in a total of 3,080 matches between the job offers and the candidates' curricula vitae, with 136 different candidates being involved. This also points to a concentration effect, where a small number of candidates is associated with a high SSED soft skills evaluation (Li et al., 2025). In all instances across the global set, the SSED score's numerical distribution has an average of 0.8775 with a standard deviation of 0.2813. In numerical form, it has been observed that the shortlisted candidates' set has a high density of relevant soft-skill information. However, it must be noted that there is considerable variability. Importantly, it has also been observed that the range from 0.0 to 1.0 has been covered by the model. This is an important aspect to ensure that the model retains expressiveness to maintain discriminative ability in ranking (Mashayekhi et al., 2024). The ranking structure further supports the internal consistency of the metric, with an average SSED score for rank 1 being 0.9708, monotonically decreasing to 0.8876 for rank 5. This shows that higher-ranked individuals have denser and more relevant evidence to support their soft skills, hence providing a meaningful ranking with quantitatively significant differences, similar to state-of-the-art deep learning-based ranking models (Li et al., 2025). The ranking separation statistics further provide more insights into the overall behavior of the SSED metric. From the analysis, it is clear that the average separation between ranks 1 and 10 is 0.1606, which is on a scale of 0 to 1. This clearly differentiates between higher-ranked individuals. Moreover, the standard deviation of the separation between ranks is 0.3145, with values ranging between 0.0 and 1.0, showing significant variations across different job postings, with some cases being almost impossible to differentiate based on the SSED, while in other cases, the metric clearly differentiates between candidates based on the explicitness of the soft skill information. Lastly, the recurrence analysis clearly shows that there is a significant long-tail effect, with the average number of times the CV recurred being 22.65, with the most recurrent CV being 248 times. This clearly shows that there is indeed a small number of highly evidenced profiles with versatile, high-level soft skills that can be applied to different situations. In other words, the SSED tends to identify candidates with highly developed, well-evidenced, and versatile soft skills, which can be applied to multiple situations, similar to other large-scale AI recruitment systems (Li et al., 2025; Mashayekhi et al., 2024). From the numerical analysis, it is clear that the SSED is indeed a

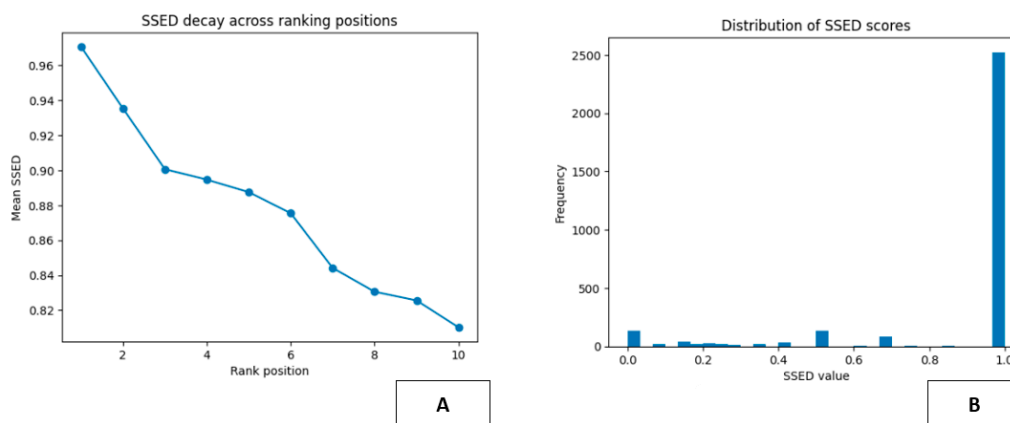
selective, well-behaved, and expressive metric. In other words, the numerical analysis supports the claim that it is possible to transform soft skills from unsubstantiated claims to evidence, hence providing a robust metric layer for talent matching, ranking, and shortlisting (Nikolaou, 2021; Dalal et al., 2020). See Table 9.

**Table 9.** Aggregate Statistics of the Soft Skill Evidence Density (SSED)–Based Ranking Experiment.

Category	Metric	Value
Dataset Size	Number of job offers	308
	Total job–CV matches (top-10)	3080
	Unique candidates involved	136
SSED (Global)	Mean	0.8775
	Standard deviation	0.2813
	Minimum	0.0000
	Maximum	10.000
Ranking Structure	Mean SSED at rank 1	0.9708
	Mean SSED at rank 2	0.9355
	Mean SSED at rank 3	0.9007
	Mean SSED at rank 4	0.8947
	Mean SSED at rank 5	0.8876
Ranking Separation	Mean gap (rank 1 – rank 10)	0.1606
	Std. dev. of gap	0.3145
	Minimum gap	0.0000
	Maximum gap	10.000
Candidate Recurrence	Mean appearances per CV	22.65
	Median appearances per CV	6
	25th percentile	2
	75th percentile	11
	Maximum appearances	248

Note. This table reports dataset scale, global SSED distribution, ranking structure, separation, and recurrence, showing how evidence-based soft-skill measurement improves discriminative power beyond keyword claims and supports interpretable, robust ranking in recruitment decision support systems.

Figure 10 offers a unique perspective on the behavior of the SSED matching model, not only on the ranking but also on how the score is distributed across all possible pairs of jobs and candidates.



**Figure 10.** SSED Ranking Decay and Score Distribution Across Top-10 Job–Candidate Matches. Panel A shows a sharp, near-monotonic SSED decline across ranks, indicating strong separation of evidence-rich profiles; Panel B shows a right-skewed distribution concentrated near one, confirming SSED’s discriminative, evidence-based ranking behavior.

This two-fold perspective is similar to what is seen in contemporary survey results on e-recruitment recommendation systems. For instance, ranking coherence and score distribution are considered some of the most important factors that determine the quality of an e-recruitment system (Mashayekhi et al., 2024). Figure A indicates that, with every step away from the top-ranked candidates towards the tenth-ranked candidates, there is a corresponding decrease in the mean SSED score, with a sharp, near-monotonic curve indicating a sharp decline. It is also indicated that the top SSED scores are reserved for the top-ranked candidates, with a corresponding decrease in score with a decline in ranking. This could indicate that the SSED matching model produces a coherent ranking, with the top candidates having the highest density of evidence supporting their soft skills. This is similar to what is seen with contemporary ranking-based e-recruitment systems that make use of deep learning algorithms for ranking candidates (Li et al., 2025). The sharp decline between the top-ranked candidates and those who appear further down the ranking suggests that the SSED model produces a ranking that highlights a few candidates with strong soft skills and a corresponding amount of corroborating evidence, with a corresponding decline in SSED score with a decline in ranking. This is a characteristic that is seen with ranking systems that have the capacity to identify the “best matches” and separate them from the rest, making the SSED matching model useful for e-recruitment systems (Nikolaou, 2021). Figure B shows the distribution of SSED scores over all matches. The distribution is significantly right-skewed, with an increased tendency towards scores closer to 1.0, gradually decreasing towards lower scores. This shows that there are a significant number of shortlisted matches with high SSED scores, implying higher levels of alignment between evidence statements and soft skills, as well as lower numbers of matches with lower levels of alignment. The tendency towards 1.0 also shows that, for a significant number of job–CV pairs, the evidence statements within the CV are significantly aligned with the soft skills obtained from job descriptions, as indicated by the results of applying both semantic-based and entity-based extraction techniques to recruitment data (Ifakir et al., 2025). At the same time, the presence of scores closer to the lower end of the range shows that SSED scores are dynamic, an essential attribute of scores in modern AI-based recruitment and recommendation systems (Mashayekhi et al., 2024). Both figures demonstrate that, in addition to being locally interpretable, SSED scores can be part of an overall ranking system with high scores concentrated towards the higher end of the ranking, with the distribution of scores over the entire candidate pool also being provided, an essential attribute of evidence-based recruitment approaches that use sophisticated semantic-based and deep learning-based architectures, as indicated in Li et al. (2025) and Nikolaou (2021).

## 10. CTFS as a Fine-Grained Signal for Cultural and Team Compatibility in Recruitment

The Cultural & Team Fit Score (CTFS) is a measure that aims to measure the level of similarity between the personal values, working style, and personal characteristics of an individual and the culture of a particular organization or the working environment of a specific team within an organization. This measure can be regarded as an enhancement of the current technology-enhanced recruitment/selection methodologies that take into consideration a multi-dimensional criterion of fit between the applicant and the organization or the working environment of the team that the applicant is expected to join (Nikolaou, 2021; Potočnik et al., 2024). Contrary to the simplistic semantic analysis of the applicant’s profile that might result in the misidentification of the applicant as culturally fit due to textual similarities that might be drawn between the applicant’s personal profile and the organizational culture or working environment of the team that the applicant is expected to join—such as the use of phrases like “fast-paced environment” or “team-oriented mindset” that are

common in applicant profiles that are similar to the organizational/working environment of the team that the applicant is expected to join—the applicant's behavior is evaluated under similar conditions, such as decision-making styles, conflict resolution styles, receptiveness to feedback, leadership/followership tendencies, adaptability to change, or other personal characteristics that can be elicited from the applicant using other information sources that are increasingly used in technology-enhanced recruitment/selection methodologies, such as surveys on the sophistication of the recruitment methodology used (Nikolaou, 2021; Mashayekhi et al., 2024). The primary advantage of using CTFS over semantic analysis is its sensitivity to the context of culture. Culture is not just defined by the words an organization uses; culture is defined by expectations, norms, and trade-offs. For example, two organizations may share the same culture defined by words like “innovation.” However, one may prioritize speed, while the other may prioritize validation. The differentiation of expectations is achieved by mapping an individual's potential behavior against the organizational culture signature and team culture signature. This method moves beyond the boundaries of text-based semantic relatedness or embedding-based matching mechanisms (Zadykian et al., 2025; Li et al., 2025; Kurek et al., 2024). It also makes possible precise predictions at the team level, not the organizational level, allowing for granular evaluation, e.g., between a research team and a sales team within an organization. This framework helps to mitigate micro-level misalignments, ensuring alignment with modern recruitment processes using artificial intelligence technology (Ifakir et al., 2025). Moreover, the use of CTFS shows higher predictive validity in employee retention and employee performance. Research in organizational psychology shows that alignment between individual characteristics and organizational contexts is strongly, directly related to employee performance and turnover outcomes (Van Iddekinge et al., 2011). It can also predict possible integration problems during the employee onboarding process. The method addresses fairness, transparency, and accountability issues with algorithm-based recruitment processes (Raghavan et al., 2020; Mehrabi et al., 2021). It provides explanations to the hiring teams about the reasons behind the predictions, e.g., employee tolerance of ambiguity, employee adherence to structured processes, employee collaboration propensity, etc. The method answers the call for transparent recruitment processes using artificial intelligence technology (Mashayekhi et al., 2024; Zadykian et al., 2025). Conceptually, the Cultural & Team Fit Score can be seen as a step beyond similarity at the word level, focusing on behavioral and contextual alignment. The table provides a quantitative summary of CTFS behavior and discriminative power, which can be seen as a quantitative summary of a large-scale matching experiment with 308 job offers, each with a top-10 shortlist, resulting in a total of 3,080 job/CV matches. This data can be used to create a strong evaluation system, similar to medium-scale evaluations presented in recent research on AI-assisted recruitment (Li et al., 2025; Kurek et al., 2024). The data consists of 308 unique job roles, with 75 unique candidates in the top-10 shortlists, i.e., candidates culturally compatible with many different job roles. The global CTFS statistics show that, on average, a score of 0.9352 with a standard deviation of 0.0594 is obtained, indicating that most pairs of jobs/CVs have considerable cultural and team alignment. Moreover, the lowest score is 0.7333, which indicates that even at the lowest possible score, there is some degree of alignment, while the highest score is 1.000, indicating highly aligned pairs of jobs/CVs. These results indicate that CTFS is a fine-grained ranking signal for plausible candidates, which is consistent with a natural fit with layered evaluation architectures, as proposed by recent intelligent recruitment systems (Li et al., 2025; Mashayekhi et al., 2024). This ranking structure is further corroborated by a gradual decrease in mean CTFS values for candidates at rank 1, which is 0.9574, and a small difference between consecutive ranks, which indicates that CTFS is effective at differentiating between highly similar candidates, a characteristic that is unique to highly precise fit measures. Furthermore, ranking separation metrics also indicate that CTFS differentiates between candidates, with a mean difference between ranks 1 and 10 being 0.0316, which indicates that while there is a limited range of possible scores, there is still a meaningful score range, which indicates that CTFS differentiates between candidates with different levels of cultural compatibility, which is a concern that is often cited with regard to algorithmic hiring systems (Raghavan et al., 2020; Mehrabi et al., 2021). Finally, recurrence

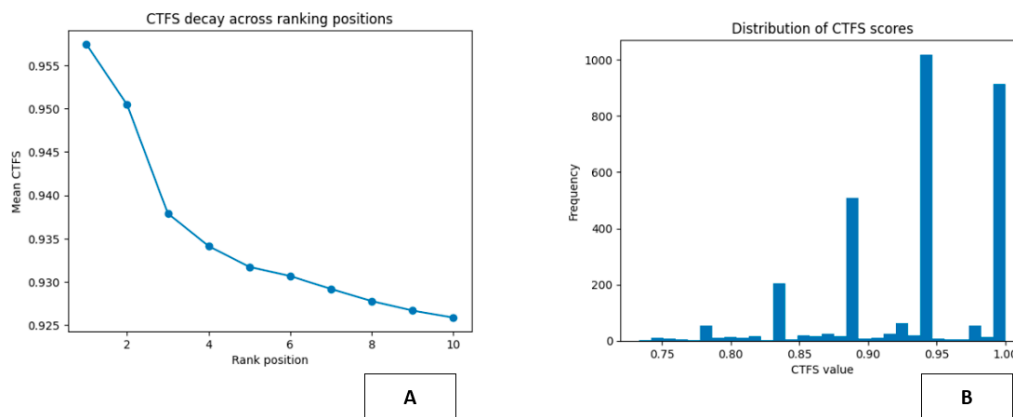
metrics indicate that there is a concentration effect, with a number of candidates showing cultural compatibility with multiple positions, which further indicates that CTFS is effective at identifying a group of highly suitable candidates with a broad range of applicability, which is relevant for a range of diverse team settings. The data characteristics, which include high scores with minimal variance, a smooth ranking structure, minimal inter-rank separation with meaningful differentiation at higher ranks, and significant recurrence, further indicate that the proposed measure is effective at conducting a nuanced evaluation of cultural and team compatibility between qualified candidates, which is consistent with semantic and embedding-based compatibility metrics, which have been proposed by recent AI-based hiring systems (Mashayekhi et al., 2024; Li et al., 2025; Kurek et al., 2024). See Table 10.

**Table 10.** Aggregate Statistics of the Cultural & Team Fit Score (CTFS)-Based Ranking Experiment.

Category	Metric	Value
Dataset Size	Number of job offers	308
	Total job-CV matches (top-10)	3080
	Unique candidates involved	75
CTFS (Global)	Mean	0.9352
	Standard deviation	0.0594
	Minimum	0.7333
	Maximum	10.000
Ranking Structure	Mean CTFS at rank 1	0.9574
	Mean CTFS at rank 2	0.9505
	Mean CTFS at rank 3	0.9379
	Mean CTFS at rank 4	0.9341
	Mean CTFS at rank 5	0.9317
Ranking Separation	Mean gap (rank 1 – rank 10)	0.0316
	Std. dev. of gap	0.0336
	Minimum gap	0.0000
	Maximum gap	0.1111
Candidate Recurrence	Mean appearances per CV	41.07
	Median appearances per CV	14
	25th percentile	5.5
	75th percentile	51
	Maximum appearances	214

Note. This table summarizes CTFS distribution, ranking structure, separation, and recurrence, showing consistently high cultural alignment, fine-grained discrimination among similar candidates, and concentration effects, supporting CTFS as a precise signal for team- and culture-level compatibility in multi-stage recruitment pipelines.

Figure 11 offers a clear and complementary picture of the behavior of the Cultural & Team Fit Score (CTFS) within the employed ranking framework for the job candidate matching process, as supported by the most recent advances in the application of AI for the development of recruitment and recommendation systems (Mashayekhi et al., 2024; Nikolaou, 2021). See Figure 11.



**Figure 11.** CTFS Ranking Decay and Score Distribution Across Top-10 Job–Candidate Matches. Note. Panel A shows a smooth, monotonic CTFS decline across ranks, indicating stable ordering among culturally compatible candidates; Panel B shows a high-score–skewed distribution, confirming CTFS’s fine-grained discrimination within a plausible candidate pool.

Panel A of the graph shows the decay of the mean CTFS across the different positions from 1 to 10, while Panel B of the graph shows the overall distribution of the CTFS across the different matches that are evaluated. This allows the evaluation of the internal consistency of the employed ranking approach as well as the global statistical properties of the CTFS, which are the primary concerns of the most contemporary intelligent recruitment architectures (Li et al., 2025). Regarding the first of these panels, the smooth curve showing the monotonic decay of the mean CTFS across the different positions from 1 to 10 is noteworthy. The highest ranked candidates are the ones that achieve the highest levels of the CTFS, almost reaching 0.96 at the first position. Between the first and second positions, there is a notable decrease, although this is followed by a steeper decline between the second and third positions. After this initial separation of the candidates from the others, the decay of the CTFS becomes progressively flatter across the remaining positions until the 10th position is reached, at which the CTFS remains at 0.926. This pattern is very informative because it suggests that the ordering is not arbitrary; that the top positions are clearly distinguished from the rest of the candidates, especially at the top of the ordering, while the second half of the top-10 list contains candidates whose cultural fit is very similar. In practice, CTFS is thus very successful at identifying the best-fitting candidates while also indicating that there are several alternatives among the shortlist that are almost equivalent from a cultural and team fit perspective. The consistency of this pattern of decay is also consistent with the requirements of transparent and ordered ranking that are important for the design of algorithmic hiring systems (Raghavan et al., 2020). The picture provided in Panel B supports this perspective because it shows the distribution of CTFS scores for all job-CV matches. The histogram of the distribution is heavily skewed towards high scores, indicating that the great majority of the scores are between 0.90 and 1.00, while there is a clear peak at the upper end of the scale. Only a small number of matches are found at the lower end of the scale, between 0.75 and 0.85. This confirms that the CTFS operates in the high-score regime because the analysis is focused on the top 10 candidates for each job, not the overall applicant pool. Thus, the CTFS is a fine-grained discriminative signal among the pool of plausible and relevant candidates, rather than a coarse filter that distinguishes between good and poor candidates. The interpretability of such fine-grained similarity measures is also consistent with the requirements of the newly developing area of explainable matching systems (Zadykian et al., 2025). This juxtaposition of the two figures points to a key methodological concern. First, as a result of the prevalence of both high and concentrated score values, small numerical value differentials become relevant from a ranking perspective. Thus, Panel A demonstrates that even small declines in mean CTFS exhibit a systematic pattern with regard to their rank, suggesting that this metric is a stable measure for ordering. At the same time, Panel B

demonstrates that overall, CTFS has a constrained range, which explains why the distance between lower ranks is not large because, in fact, many of the candidates are very similar in cultural and team fit. In terms of fairness, this is a key concern because, from a fairness and robustness perspective, it is critical to understand the distribution of score values to avoid ranking being driven by spurious fluctuations (Mehrabi et al., 2021). From a practical point of view, these patterns indicate that CTFS is well-adapted to support shortlist prioritization and selection discussions as part of technology-enabled recruitment processes (Nikolaou, 2021). The top-ranked individual is, on average, the best cultural fit, but several strong alternatives are close behind. On the other hand, the strong clustering of scores at the top also reveals that cultural fit is not unusual but rather a common trait among the core group of candidates who consistently score well as cultural fits—an established phenomenon when AI-based recommendation systems are applied to recruitment processes (Mashayekhi et al., 2024; Li et al., 2025). To summarize, the above two figures clearly indicate that CTFS acts as a cohesive and high-resolution ranking signal that identifies subtle but systematic differences in cultural and team alignment among candidates while at the same time satisfying the requirements of explainability, robustness, and fairness as part of modern algorithmic recruitment systems (Raghavan et al., 2020; Mehrabi et al., 2021; Zadykian et al., 2025).

## 11. CCS as a Risk-Reduction Filter in Multi-Stage Recruitment Pipelines

The proposed metric, named Contract Compatibility Score (CCS), is meant to measure the level of compatibility between a candidate's expectations, restrictions, and preferences, on one side, and the terms of a particular contract, on the other, with a particular emphasis on structured recruiting practices, including those enabled with the help of modern technologies, in line with modern scholarship (Nikolaou, 2021; Potočník et al., 2024). Unlike semantic analysis, which can, at best, detect words like "full-time," "freelance," or "remote" in a resume or a job description, CCS is meant to measure multidimensional feasibility in terms of structured variables like contract type, duration, working hours, flexibility, benefits, risk, or legal restrictions, etc. From a conceptual point of view, CCS is meant to treat contract fit as an early-stage, quantified, rather than a late-stage, qualitative, recruitment decision variable, in line with modern, multi-layer intelligent architectures of recruiting systems, including structured constraints, like those proposed in Li et al. (2025) or Mashayekhi et al. (2024). Nevertheless, it is necessary to understand CCS in relation to its metric behavior in a broader, empirical context. In particular, the analysis is conducted on a pool of 308 job offers, each with a corresponding top-10 shortlist, resulting in a total of 3,080 job-CV pairs. In this pool, there are merely 53 unique candidates, with their CVs appearing in all rankings, indicating a strong concentration effect, with a relatively small group of candidates satisfying contract requirements in a multitude of job offers. From a quantitative point of view, it can be stated that CCS is capable of reliably detecting a relatively rare attribute in a candidate pool, with a particular group of structurally compatible candidate profiles, in a manner analogous to concentration phenomena in recommendation systems, like those proposed in Kurek et al. (2024) or Mashayekhi et al. (2024). In particular, with regard to the distribution of scores, CCS is found to have a mean of 0.6448 with a standard deviation of 0.0598. These figures position the metric on the lower end of the spectrum, clearly below the average for other cultural/semantic fit metrics. This is not to say that it is a failing, as it reflects the tight constraints associated with contractual compatibility. Clearly, factors such as work hours, contract duration, legal status, and compensation structures place real-world limits on achievable compatibility. The minimum CCS figure of 0.4250 clearly shows that even the lowest-ranked of the top 10 candidates still meet a certain minimum threshold of contractual feasibility, while the maximum CCS of 0.8750 clearly shows that nearly ideal contractual compatibility is possible, albeit relatively rare. Moreover, the absence of extremum clustering of the data suggests that CCS does not artificially inflate its figures, providing more real-world applicability to the algorithmic hiring process. In terms of governance, the inclusion of such constraints can only serve to increase transparency and accountability in algorithmic hiring systems, as suggested by Raghavan et al. (2020), Mehrabi et al. (2021). The ranking structure of CCS can further clarify its function as a metric. Clearly, the average

CCS of the top-ranked candidates is 0.6483, with minimal variation to 0.6445 by the fifth position. The minuscule differences between these figures clearly suggest that the top-ranked candidates for any given position will share similar contractual terms. In terms of the CCS as a metric, it is clearly not intended to be used as a fine-grained preference, as the primary contractual conditions will clearly be met by all candidates, with minimal practical application of the CCS figures to separate them. This two-stage process of constraint-based filtering, followed by more fine-grained ranking, clearly aligns with more modern approaches to the design of recruitment systems, as suggested by Li et al. (2025), Mashayekhi et al. (2024). This can be further evidenced by the calculation of ranking separation statistics. The average gap between rank 1 and rank 10 is 0.0063. The standard deviation is 0.0241, with a minimum gap of 0.0000. These values suggest that in many instances, it can be said that the best candidate and the tenth-best candidate are virtually indistinguishable with regard to CCS. At the same time, it must be noted that with a maximum gap of 0.2200, it is also evident that in a small minority of instances, there are significant gaps. This may be due to various strong constraints such as relocation needs, structured work schedules, or certain types of contracts. Quantitatively, this again supports the assertion that CCS is primarily a risk-reduction and feasibility tool rather than a discriminator. The statistics on candidate recurrence also offer certain structural characteristics. The average recurrence of each CV in the top-10 lists is 58.11. The median recurrence is 16, with a maximum recurrence of 271. This again points to a high degree of skewness in the data. The implication here is a strong "core-periphery" structure in which a small set of candidates are highly compatible with the conditions of most contracts, while the rest are much more intermittent. This has certain operational advantages. It also suggests that CCS can identify a stable set of "structurally compatible" profiles that can be given priority early on in the pipeline to avoid late-stage failure, such as offer rejection or early turnover—goals that are core to modern data-informed hiring practices (Nikolaou, 2021; Potočnik et al., 2024). All of this numerical data demonstrates that the CCS algorithm provides a moderate score with tight clusters, minimal average separation between ranks, and strong recurrence of a small set of candidates. This is the metric behavior expected from a KPI that seeks to model feasibility rather than similarity. CCS does not attempt to score the fine-grained preferences within a small range of viable options but rather seeks to stabilize the set of viable candidates by imposing hard constraints on the structural risk. This metric behavior aligns with technical, semantic, and cultural metrics, which may have a broader score range for viable candidates, while the CCS provides a firm foundation for the hiring decision, thus improving the predictability, efficiency, and effectiveness of the matching pipeline as a whole (Raghavan et al., 2020; Mehrabi et al., 2021). Table 12.

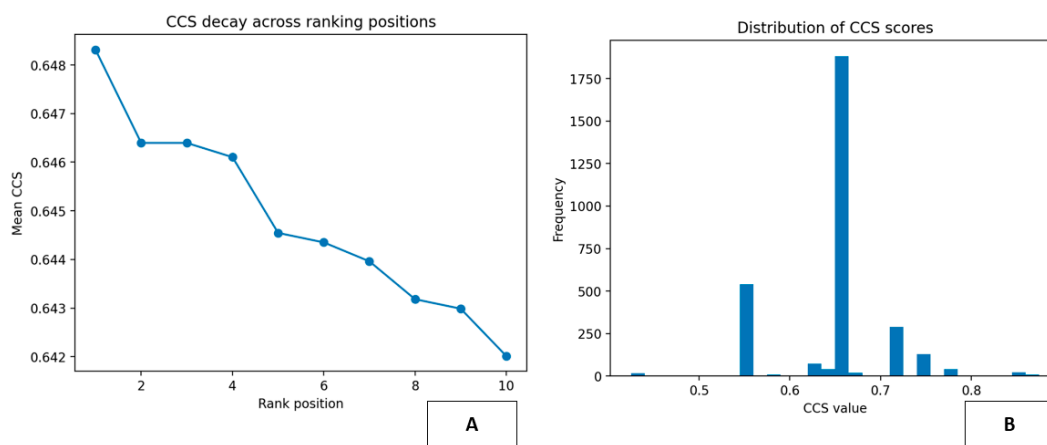
**Table 12.** Aggregate Statistics of the Contract Compatibility Score (CCS)-Based Feasibility Filtering Experiment.

Category	Metric	Value
Dataset Size	Number of job offers	308
	Total job-CV matches (top-10)	3080
	Unique candidates involved	53
CCS (Global)	Mean	0.6448
	Standard deviation	0.0598
	Minimum	0.4250
	Maximum	0.8750
Ranking Structure	Mean CCS at rank 1	0.6483
	Mean CCS at rank 2	0.6464
	Mean CCS at rank 3	0.6464
	Mean CCS at rank 4	0.6461
	Mean CCS at rank 5	0.6445
Ranking Separation	Mean gap (rank 1 – rank 10)	0.0063
	Std. dev. of gap	0.0241

Candidate Recurrence	Minimum gap	0.0000
	Maximum gap	0.2200
	Mean appearances per CV	58.11
	Median appearances per CV	16
	25th percentile	2
	75th percentile	31
	Maximum appearances	271

Note. This table summarizes CCS distribution, ranking structure, separation, and recurrence, showing tightly clustered scores, minimal inter-rank differences, and strong concentration effects, confirming CCS as a feasibility and risk-reduction filter rather than a fine-grained preference ranking signal.

Figure 12 provides an informative and concise visual representation of the behavior of the Contract Compatibility Score (CCS) in the broader framework of the ranking. See Figure 12.



**Figure 12.** CCS Ranking Decay and Score Distribution Across Top-10 Job–Candidate Matches. Note: Panel A shows a shallow, monotonic CCS decline, indicating feasibility-based ordering with minimal discrimination; Panel B shows concentrated mid-range scores, confirming CCS acts primarily as a constraint filter rather than a fine-grained preference optimizer.

Panel A represents the mean CCS over the different positions in the ranking framework, while Panel B represents the overall distribution of CCS over all the job-candidate pairs. With regard to Panel A, the smooth curve represents the gradual decline in the mean CCS from the first to the tenth position in the framework. The top-ranked candidates have the highest average CCS, slightly above 0.648. The smoothness of the curve gradually decreases and reaches 0.642 at the tenth position. Although the differences are small, it is noteworthy that the monotonicity of the mean CCS over the positions in the framework reveals that the ranking framework is appropriate and that the top-ranked candidates are more contractually compatible with the position than the lower-ranked candidates. This behavior follows from the established properties of structured recruitment technologies that take into account feasibility constraints in the ranking framework (Nikolaou, 2021; Mashayekhi et al., 2024). The smoothness of the curve also reveals that the top-ranked candidates per position in the framework are similar in terms of contractual compatibility. As discussed above, contractual compatibility is an overarching concept in which these types of factors often play the role of hard constraints. After satisfying the principal factors such as contract type, working hours, level of flexibility, and remuneration scheme, candidates often fall within a narrow range of feasibility. Therefore, the CCS does not intend to discriminate among candidates but rather to ensure that shortlisted candidates meet the overarching contractual requirements of the position. This constraint-based behavior follows from the algorithmic recruitment technologies that focus on the overall

minimization of risks (Raghavan et al., 2020; Mehrabi et al., 2021). Moreover, Panel B provides corroboration of this result with regard to the distribution of the CCS score, as shown in the histogram. The distribution of the data appears to be concentrated in the middle to higher end of the score range, with a strong peak at about 0.65-0.70, accompanied by smaller peaks at neighboring scores. The number of data points is small at the lower end of the scale and even smaller at the higher end, as expected given the bounded nature of contractual compatibility. Such score concentration phenomena are known to occur in other e-recruitment recommendation system problems with large-scale data, where structured feasibility constraints are incorporated, as reported in Mashayekhi et al. (2024). The concentration of scores at the dominant peak accounts for the smooth decline in the score distribution, as shown in Panel A, reflecting the ranking of candidates with close contractual compatibility. With regard to fairness, the stability of the score distribution within a narrow range of possible matches prevents arbitrary differentiation among equally qualified candidates, as discussed in Raghavan et al. (2020) and Mehrabi et al. (2021). Overall, the charts above demonstrate that the CCS is a stable and conservative ordering signal since there is a gradual decay while maintaining the same range of scores. This further supports the assertion that contract compatibility is primarily centered on the feasibility and risk-reducing aspects of the preference maximization problem rather than the maximization of the preference itself. Thus, the CCS can be used for initial filtering out of structurally incompatible candidates while also producing shortlists of candidates that are all contractually compatible along various other dimensions such as skills or cultural fit (Nikolaou, 2021; Mashayekhi et al., 2024).

## 12. Turning Location Into a Constraint: A Data-Driven Approach with LMFI

Location & Mobility Fit Index (LMFI) is presented as a tool for assessing the level of congruence that can be expected between a given candidate's geographical constraint factors, their mobility preferences, and their lifestyle needs, on the one hand, and the corresponding location needs and travel expectations for a particular given position, on the other. This measure distinguishes itself from other approaches to semantic analysis, which only allow for the detection of superficial expressions such as "remote," "hybrid," "willing to relocate," etc. This measure goes beyond the usual approaches to information extraction that have typically been discussed in the context of entity detection (Khaouja, Kassou, & Ghogho, 2021; Ifakir et al., 2025). Location, from a metrics perspective, is a multidimensional constraint that includes commuting distance, commuting time, visa/work permit, time zone overlap, travel tolerance, personal constraints, family constraints, and willingness to relocate. This framework allows for the LMFI to differentiate between the statements made by two different candidates who state "willing to relocate." In line with the latest developments in technology-mediated approaches to recruitment strategies and decision models (Nikolaou, 2021; Potočnik et al., 2024), geographical feasibility has also been made a key component of the overall multi-KPI framework (Li et al., 2025). In terms of evaluation, the dataset includes 308 job offers, each with a corresponding top 10 shortlist, which translates into a total of 3,080 job/CV pairs. A key quantitative finding from this research is that the diversity of unique candidates is very low, at only 34. This finding points to a strong concentration effect, where a small number of candidates meet the criteria for multiple job opportunities. Geographical feasibility is a hard constraint, which greatly reduces the solution space. From a fairness and structural bias perspective, this phenomenon is in line with the broader literature on constraint effects in algorithmic systems, especially in the context of machine learning (Mehrabi et al., 2021). With regard to the global score, the mean was 0.5409 with a standard deviation of 0.0740. This is a relatively compressed scale. In addition, unlike other factors such as culture and skill, which may be less constrained, geographical feasibility is subject to real-world limitations such as commuting bounds and legal issues. The minimum global score was 0.5000, which means that all pairs of users that meet the top 10 criteria will meet the minimum threshold of geographical feasibility, while the maximum global score was 0.7500, which represents a relatively low incidence of nearly perfect pairs. The ranking structure also provides further insight into the behavior of the LMFI as a signal. The average LMFI of ranks 1 through 4 is flat at 0.5422, then

gradually dips to 0.5402 at rank 5. This flatness of the curve further supports the notion that the top candidates for any given role are largely indistinguishable with regard to location and mobility fit. In other words, quantitatively, it further supports the notion that the LMFI is less of a fine-grained optimizer of user preferences and more of a gate to ensure that users meet the primary geographical/mobility constraints. This is further borne out by the ranking separation metrics, which showed that the average distance between rank 1 and rank 10 across all roles was 0.0023, with a standard deviation of 0.0116 and a minimum of 0.0000. In addition, the maximum distance was 0.0600, which means that for most roles, the top-ranked user and the user at rank 10 are nearly identical with regard to the LMFI. The recurrence statistics for candidates also shed further light on the operational implications of the LMFI's behavior. On average, a CV will recur 90.59 times within top-10 lists, with a median recurrence rate of 35 times, a 75th percentile rate of 229.5, and a maximum rate of 277 times. This indicates that a few candidates will recur frequently due to their location and mobility compatibility, while a majority will be systematically excluded by constraint effects. This concentration effect is also consistent with constraint-based filtering effects that have been reported by Li et al. (2025) for large-scale recruitment recommendation systems. Overall, these metrics appear to indicate a KPI that produces a range of moderate, tightly grouped values, a flat ranking profile, minimal average separation across positions, and a highly concentrated distribution of candidate recurrence. This suggests that the KPI is designed to operate within a model that takes account of geographical feasibility and mobility sustainability but does not account for preference or semantic similarity between candidates or locations. The LMFI framework is designed to filter candidates by hard constraints and identify a small, stable set of structurally compatible candidates, thereby eliminating risks related to burnout, relocation failure, and unsustainable travel. It thereby translates a vague textual representation of location preference into a clear, data-driven decision variable within a multi-KPI-based recruitment framework. See Table 13.

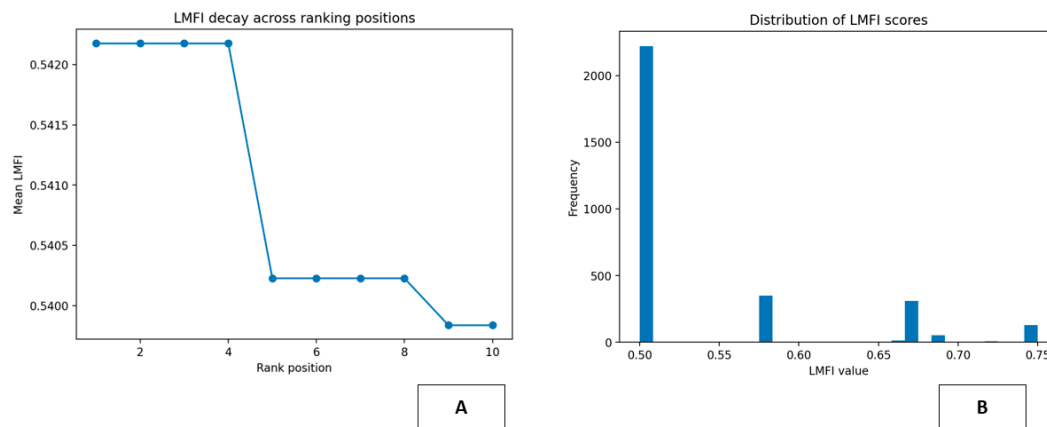
**Table 13.** Aggregate Statistics of the Location & Mobility Fit Index (LMFI)-Based Feasibility Filtering Experiment.

Category	Metric	Value
Dataset Size	Number of job offers	308
	Total job-CV matches (top-10)	3080
	Unique candidates involved	34
LMFI (Global)	Mean	0.5409
	Standard deviation	0.0740
	Minimum	0.5000
	Maximum	0.7500
Ranking Structure	Mean LMFI at rank 1	0.5422
	Mean LMFI at rank 2	0.5422
	Mean LMFI at rank 3	0.5422
	Mean LMFI at rank 4	0.5422
	Mean LMFI at rank 5	0.5402
Ranking Separation	Mean gap (rank 1 – rank 10)	0.0023
	Std. dev. of gap	0.0116
	Minimum gap	0.0000
	Maximum gap	0.0600
Candidate Recurrence	Mean appearances per CV	90.59
	Median appearances per CV	35
	25th percentile	21

	75th percentile	229.5
	Maximum appearances	277

Note. This table reports LMFI distribution, ranking structure, separation, and recurrence, showing flat rankings, minimal inter-rank differences, and strong concentration effects, confirming LMFI as a hard-constraint feasibility filter for geographical and mobility compatibility rather than a preference-based ranking signal.

Figure 13 offers a compact and understandable representation of the behavior of the Location & Mobility Fit Index (LMFI) within the ranking system. See Figure 13.



**Figure 13.** LMFI Ranking Profile and Score Distribution Across Top-10 Job-Candidate Matches. Note. Panel A shows a flat LMFI profile with minimal decline, indicating feasibility filtering rather than preference ranking; Panel B shows discrete, clustered scores, confirming LMFI operates as a constraint-based gate for geographic and mobility compatibility.

Panel A displays the decline of the mean LMFI across ranking positions, while Panel B displays the distribution of LMFI across all possible job-candidate pairs. For Panel A, we observe that the curve is flat across the first four ranking positions, with the mean LMFI retaining its value at approximately 0.542. This indicates that the top candidates are virtually indistinguishable with regard to location and mobility fit. The decline only manifests itself beyond the fourth ranking position, with a decline first to around 0.540, then to slightly below that value for the lower ranking positions. It is important to note that even at this latter range, the difference between a first-place and a tenth-place ranking is extremely minimal. The flatness of this curve indicates that LMFI primarily operates as a feasibility criterion, which is consistent with general trends in the broader labor market with regard to structural mobility constraints and hybrid work arrangements (Di Battista et al., 2023; Burbridge, 2025). Beyond a certain set of fundamental geographical and mobility constraints related to a specific role, such as acceptable commuting distances, relocation, or tolerance for traveling, candidates will tend to score closely together, with the ranking system not differentiating candidates on the basis of strong preference but instead grouping candidates who are uniformly compatible with regard to location fit for a specific role. This is, of course, an expected result given the increasingly regulated and contractual nature of both border-crossing work arrangements and remote work, in which feasibility is defined by both legal, tax, and organizational conditions, not merely personal preference (Escribano, 2024). Panel B of Figure 3 above corroborates the above result, as it shows the value distribution of LMFI to be heavily concentrated. The histogram shows an overwhelmingly dominant spike at 0.50, with smaller groups of scores clustered at 0.58, 0.66, and 0.75. The discrete nature of the distribution shows that LMFI is heavily influenced by a small number of underlying configurations of location and mobility, with the result that a number of applicants are classified within the same compatibility groups, thereby receiving the same scores or near-same scores. The

strong spike at the lower end of the scale, 0.50, shows that the majority of the matches are merely compatible within the minimum feasible conditions, with a smaller number of matches receiving higher scores, indicating strong relocation readiness or high levels of flexible mobility, among other factors. Such stepwise feasibility patterns are, of course, consistent with the nature of International Labor Governance, in which both compliance and mobility are governed by rules based upon threshold conditions, not continuous scales (Chau, 2025). Overall, the two panels of Figure 3 above show that LMFI functions within a limited score space with both limited variance and significant clustering. While this may be seen as limiting, it is, of course, an accurate reflection of the nature of both location and mobility, in which binary conditions are often at work. From a practical point of view, the above results show that LMFI is best suited to ensuring that the shortlisted pool of applicants contains merely those who are realistically compatible with the geographical configuration of the job, but is not suited to ranking within that feasible pool. In other words, LMFI turns location into a specific, practical constraint, which can be useful for risk reduction, retention opportunities, and more predictable hiring outcomes that align with the emergent global trends in labor mobility (Di Battista et al., 2023; Burbridge, 2025).

### 13. Seniority & Compensation Alignment (SCA): Measuring Structural Fit in Hiring

The Seniority & Compensation Alignment (SCA) Key Performance Indicator (KPI) has been developed to measure the level of correspondence between the experience level, scope of responsibility, and market positioning of a given candidate in comparison to the seniority and compensation framework inherent to a given function. While semantic-based measures may be based on labels such as “Senior Manager” or phrases such as “10+ years of experience,” SCA measures the structural level of correspondence with regard to the economic and organizational structure of a given function. In a quantitative sense, this involves aggregating information with regard to breadth of responsibility, decision impact, team size, benchmarking, and compensation into a single measure of structural correspondence. This approach is consistent with recent studies on changes to job structure and compensation in labor markets impacted by artificial intelligence (Di Battista et al., 2023; Mindell & Reynolds, 2023; Drozd & Tavares, 2024; Mer, 2023). From an empirical perspective, this involves an evaluation based on a database of 308 job offers with corresponding top-10 rankings, resulting in a total of 3,080 matches between jobs and CVs. Of the total rankings, 80 unique candidates are observed. While this represents a level of concentration effect with regard to structural correspondence to numerous roles with regard to seniority and compensation expectations, it also suggests a level of diversity with regard to market-positioning constraints that only a limited subset of candidates are able to meet across various job opportunities—a quantitative measure of diversity closely tied to issues of wage dispersion, benchmarking, and stratification dynamics in labor markets (Vazquez-Alvarez et al., 2022; Peña-Casas et al., 2025). Considering the global SCA score distribution, the mean value equals 0.7219, and the standard deviation equals 0.1123. Such an average level of alignment with considerable variability may be seen as a sign of both medium-to-high levels of alignment and considerable variability compared to stricter feasibility criteria. Such variability may be seen as conceptually appropriate because the alignment of seniority and compensation depends on various continuous criteria rather than binary constraints like the location or contract type. Such multidimensional structures are at the core of contemporary debates on compensation and remuneration (Turba, 2024; Aumayr-Pintar & Baggio, 2025). The minimum value of the SCA score equals 0.6200. Thus, even the weakest top-10 matches will always have at least some level of structural compatibility. The maximum SCA score equals 0.8750. Thus, even though there are some cases of almost optimal levels of alignment, they are not dominant. The SCA score does not restrict candidates to a narrow range but reflects various levels of alignment appropriate from the contemporary point of view of HR analytics and compensation alignment (Mer, 2023). The ranking structure also supports that SCA acts as an effective and consistent ordering signal. The SCA at rank 1 has an average of 0.7274. The SCA at rank 2 drops slightly but remains close to the previous one.

From then on, the SCA gradually decreases as the rank decreases. The SCA at rank 5 drops to 0.7223. Although the SCA changes slightly from one rank to another, the gradual decline from rank 1 to rank 5 indicates that candidates at higher ranks are more aligned with the seniority and compensation structure of the job. From the quantitative perspective, this supports the effectiveness of SCA as an ordering signal. The above phenomenon also holds when considering the ranking separation statistics. The average ranking separation from rank 1 to rank 10 is 0.0094. The standard deviation of ranking separation is 0.0314. The minimum ranking separation is 0.0000. The maximum ranking separation is 0.1500. The above phenomenon reflects the heterogeneity of the market and the structural wage segmentation (Vazquez-Alvarez et al., 2022; Peña-Casas et al., 2025). Other information can be derived from the candidate recurrence statistics. Each candidate on average appears 38.50 times across the top 10 lists, while the median is 8. The 75th percentile is 44.50, while the maximum is 236. This implies that there is a group of candidates who are best placed in terms of market-aligned seniority and salary expectations, making them suitable for a number of positions. However, the range of appearances in relation to other tighter feasibility KPIs (such as location or contract) implies that SCA allows for a more diverse candidate pool without focusing too heavily on a very small number of candidates. This is an important aspect of modern-day discussions on inclusive and resilient labor markets (Peña-Casas et al., 2025; Di Battista et al., 2023). Overall, the quantitative outcomes situate SCA at an intermediate position among KPIs, being at once more discriminative than feasibility KPIs because of their broader score distribution, while at the same time remaining closer to structural KPIs because of their grounding in economic and organizational reality. Finally, in terms of metric characteristics, with their relatively high mean scores, notable variance, consistent trend in rankings, and varying degrees of separation, it is evident that SCA is picking up economically relevant differences in alignment, rather than merely superficial textual similarity. From a practical standpoint, this makes SCA useful in risk reduction and negotiation, in that it can be used to assess whether or not there is a structural match between a candidate's level of seniority and their level of compensation, thus allowing for anticipation of potential issues that could lead to candidate disengagement, failed negotiations, or even issues of equity in compensation that have become increasingly prominent in both theoretical and practical discussions on compensation governance and transparency (Aumayr-Pintar & Baggio, 2025; Turba, 2024). Furthermore, it can also be useful in strategic workforce planning in addressing mismatches between role designs and available pools in an increasingly fluid labor market impacted by technological change (Mindell & Reynolds, 2023; Drozd & Tavares, 2024). Quantitatively, it is thus demonstrated that SCA is a useful KPI in providing a strong, meaningful, and economically relevant signal in assessing alignment, in addition to technical, cultural, and feasibility KPIs in a multi-dimensional recruitment decision-making process. See Table 14.

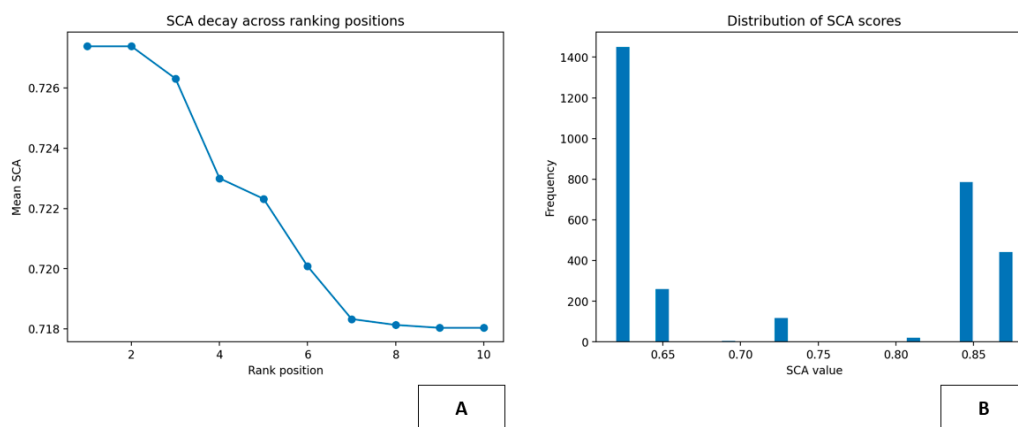
**Table 14.** Aggregate Statistics of the Seniority & Compensation Alignment (SCA)-Based Structural Fit Analysis.

Category	Metric	Value
Dataset Size	Number of job offers	308
	Total job-CV matches (top-10)	3080
	Unique candidates involved	80
SCA (Global)	Mean	0.7219
	Standard deviation	0.1123
	Minimum	0.6200
	Maximum	0.8750
Ranking Structure	Mean SCA at rank 1	0.7274
	Mean SCA at rank 2	0.7274
	Mean SCA at rank 3	0.7263

	Mean SCA at rank 4	0.7230
	Mean SCA at rank 5	0.7223
Ranking Separation	Mean gap (rank 1 – rank 10)	0.0094
	Std. dev. of gap	0.0314
	Minimum gap	0.0000
	Maximum gap	0.1500
Candidate Recurrence	Mean appearances per CV	38.50
	Median appearances per CV	8
	25th percentile	3.75
	75th percentile	44.50
	Maximum appearances	236

Note. This table summarizes SCA distribution, ranking structure, separation, and recurrence, showing moderate variability, coherent ordering, and structural differentiation, confirming SCA as an economically meaningful signal for assessing seniority–compensation alignment within multi-KPI recruitment decision-making pipelines.

Figure 14 provides an unambiguous view of what it means to be in the Seniority & Compensation Alignment (SCA) ranking framework. See Figure 14.



**Figure 14.** Structural Compatibility Alignment Across Ranks and Score Distribution. Note. Panel A shows monotonic SCA decay from rank one to ten with early separations and late plateau; Panel B reveals clustered score regimes, indicating discrete alignment tiers and limited discrimination.

Specifically, Panel A provides an overview of how the average SCA changes as we move across ranks from 1 to 10. Panel B provides an overview of the distribution of SCA across all job-candidate matches. The smooth curve in Panel A indicates a monotonic decline in average SCA as we move across ranks. The top-ranked candidates have the highest average SCA, which is just above 0.727. The SCA then gradually declines to about 0.718 by rank 10. Although the differences are small, it's important to note that this pattern is consistent across all ranks. The large decline in SCA from ranks 3 to 4 and ranks 5 to 6 suggests that the ranking system does an excellent job of distinguishing a small set of top candidates with higher structural alignment from the rest. Once we get past rank 7, we see a plateau. This suggests that lower-ranked candidates are all quite similar to one another in SCA. The only difference among them is in degree. Panel B reinforces this picture with a visualization of the distribution of SCA scores. From the histogram, it is clear that there is a strong presence of scores in the range of approximately 0.62 to 0.65, with a secondary, albeit significant, presence of scores in the range of approximately 0.84 to 0.86. There is a smaller presence in the mid-range, between 0.72 and

0.75. This reinforces the idea that SCA is primarily influenced by a few structural configurations or discrete alignment categories, likely related to combinations of seniority level, scope of responsibility, and compensation band matching. In addition, it reinforces the idea that SCA is not a continuously distributed metric, with scores clustering around a few typical alignment regimes, in line with the idea that roles and candidate profiles tend to be organized into relatively standardized market tiers. The two panels demonstrate that SCA is a relevant, albeit not extremely discriminative, signal in candidate rankings. In particular, Panel A shows that there is a strong, coherent signal in candidate rankings, with higher ranks associated with better structural alignment. Panel B shows that, in spite of this, there is a strong presence of a few principal alignment groups among candidates, in line with the relatively smooth decay in scores and with the limited differences between consecutive ranks. In practical terms, this means that SCA is well suited to identifying candidates with an experience level and compensation expectations that are structurally compatible with the role, while also highlighting that there is a strong presence of candidates within relatively similar market alignment tiers, thus offering viable alternatives from both economic and organizational points of view.

#### **14. Designing a Multi-Stage Matching Process: Insights from Cross-KPI Performance Patterns**

The following table represents a comprehensive comparison of a multi-dimensional matching framework, highlighting how various key performance indicators behave in terms of their score level, variability, their power in ranking, and candidate concentration. Instead of focusing on one particular definition of "fit," it shows how each metric measures a different aspect of compatibility between candidates and how these aspects interact with each other in a complex matching environment, in line with recent studies on structural changes in the labor market and skills restructuring (Di Battista et al., 2023; Andabayeva et al., 2024). First, with regard to Semantic Similarity, although it shows a relatively low mean score of 0.446, it also shows relatively low variability at 0.051, indicating that it is functioning at a conservative level, primarily at the lexical level. Furthermore, with a relatively high rank separation of 0.071, it is also capable of providing a certain level of candidate ranking, although with 149 unique candidates and a medium recurrence pattern, it is relatively permissive in nature, not strongly restricting the candidate pool, as expected from a text-based approach that measures surface-level similarity. The HSCR metric, on the other hand, is noteworthy for having a high mean at 0.814, together with extremely high variability at 0.356, indicating that it is relatively polar in nature, with candidates clustering at extremely high or low levels of fit. Furthermore, with a relatively moderate level of rank separation at 0.042, together with a medium recurrence pattern, it is relatively less capable of candidate ranking, functioning more like a binary signal that is capable of distinguishing between strong mismatches or strong matches, but not capable of distinguishing between strong or weak matches. On the other hand, HSPS, with its lower average value of 0.528, higher rank separation of 0.081, and highest number of unique candidates of 208, shows strong discriminative power in ranking-based ordering, with considerable diversity maintained among the profiles. Its lower recurrence value confirms that different candidates can satisfy different hard skill requirements, making HSPS more efficient in ordering than in focusing too much on a narrow subset of candidates. This behavior follows the overall employment market changes, including the diversification of skill portfolios, as seen in developing and advanced economies (Andabayeva et al., 2024). A very similar, albeit slightly more text-based, pattern is seen in the case of SSSA, with its average value of 0.563, rank separation of 0.078, good discriminative power, high number of unique candidates of 203, and lower-medium value of recurrence, suggesting that SSSA works within a more liberal space of soft skill semantics. However, the situation changes significantly with SSED, with its high average value of 0.878, high variability of 0.281, and the highest rank separation of 0.161, suggesting that soft skill evidence-based modeling has the highest discriminative power among the three models. The medium value of recurrence, along with the lower number of unique candidates of 136, confirms that requiring evidence of soft skill possession naturally leads to a lower number of candidates, with meaningful ordering differences between the remaining ones, as seen in modern

debates about valuing essential and relational work, as well as its documentation (Stevano, 2024). Finally, the case of CTFS, with its highest mean value of 0.935, minimal variance of 0.059, small rank separation of 0.032, and merely 75 unique candidates, clearly shows that the system is working within a high-fit setting, fine-tuning the ranking of already culturally compatible candidates. The high recurrence value shows that the focus is entirely on the small group of culturally compatible profiles, making the CTFS optimization signal very fine-grained. The two constraint-based KPIs, CCS (Contract Compatibility) and LMFI (Location & Mobility), show an even clearer picture of the two being primarily feasibility filters, with both having very small rank separation, 0.006 and 0.002, respectively, and both having very high recurrence rates, with LMFI having merely 34 unique candidates. This clearly shows that the two are primarily feasibility filters, with the candidates being either compatible with the constraints or not, with no room for ranking within the feasible space. In policy terms, such structural constraints are becoming increasingly relevant with the changing employment market, pay scales, and regulations (Aumayr-Pintar & Baggio, 2025). Finally, SCA (Seniority & Compensation) takes an intermediate position, with a mean value of 0.722, greater variability (0.112), a moderate rank separation (0.009), and a total of 80 unique candidates. This is consistent with recent macro-level studies on labor market restructuring and compensation realignment with technological and organizational change (Di Battista et al., 2023; Andabayeva et al., 2024). The table clearly illustrates that an efficient matching system requires a combination of highly discriminative ranking signals, such as SSED and HSPS, with strong feasibility constraints, including LMFI and CCS, and structural alignment, including SCA and CTFS. Each KPI contributes to a comprehensive process that replaces a semantic approach with a multi-dimensional, operational, and realistic approach, mirroring broader changes in skills, compensation, and employment structures (Di Battista et al., 2023; Stevano, 2024). See Table 15.

**Table 15.** Comparative Performance Profile of KPIs in the Multi-Dimensional Matching Framework.

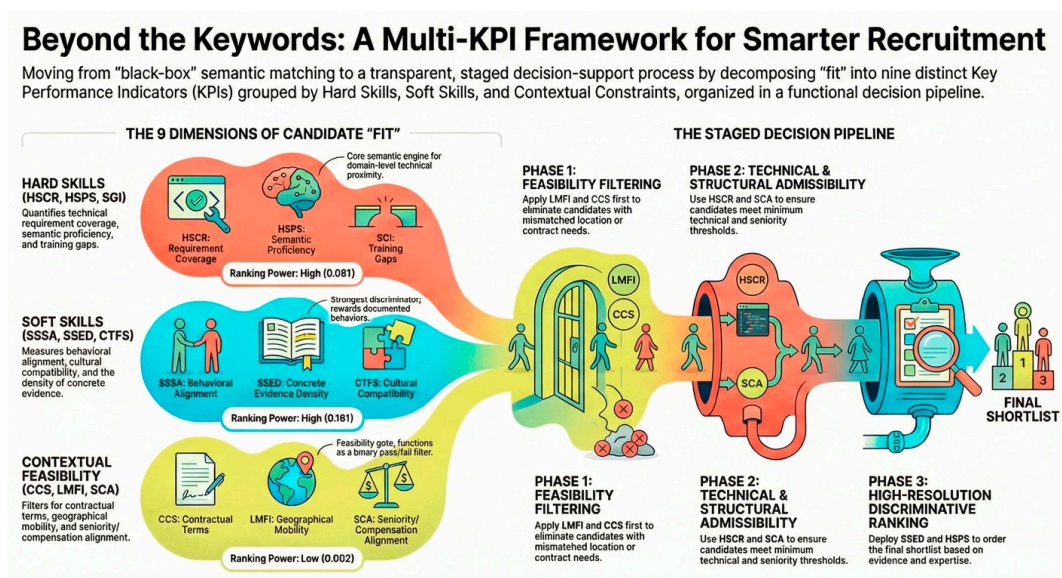
KPI	Mean Score	Variability (Std)	Rank Separation (Top1–Top10)	Unique Candidates	Recurrence Pattern	Main Insight
Semantic Similarity	0.446	0.051	0.071	149	Medium ( $\approx 21$ avg)	Good lexical discrimination, moderate concentration
HSCR	0.814	0.356	0.042	133	Medium ( $\approx 23$ avg)	Strong binary-like signal, high variance
HSPS (Hard Skills)	0.528	0.075	0.081	208	Low ( $\approx 15$ avg)	Strong ranking power, diverse candidates
SSSA (Soft Skills – Semantic)	0.563	0.065	0.078	203	Low–Medium	Good discrimination but still text-driven
SSED (Soft Skills – Evidence)	0.878	0.281	0.161	136	Medium	Strongest separation,

						evidence-based signal
CTFS (Cultural Fit)	0.935	0.059	0.032	75	High ( $\approx 41$ avg)	Very high fit, fine-grained ranking among similar profiles
CCS (Contract Compatibility)	0.645	0.060	0.006	53	High ( $\approx 58$ avg)	Feasibility filter, weak rank discrimination
LMFI (Location & Mobility)	0.541	0.074	0.002	34	Very high ( $\approx 91$ avg)	Strong constraint filter, near-binary feasibility
SCA (Seniority & Compensation)	0.722	0.112	0.009	80	Medium ( $\approx 39$ avg)	Structural alignment, moderate discrimination

Note: This table compares KPIs across mean scores, variability, ranking power, and candidate concentration, illustrating complementary roles of semantic, evidence-based, structural, and constraint-based signals within a staged recruitment decision process rather than a single monolithic optimization objective.

The comparative evaluation of all KPIs points to the fact that ranking algorithms according to their absolute “quality” is not only meaningless but also methodologically incorrect, since all KPIs measure a specific dimension of compatibility between jobs and candidates. Rather, our research points to the feasibility of applying a functional order of KPIs according to a multi-staged decision process. Thus, we shift the problem of matching from an optimization problem to a multi-staged decision process according to the logic of real recruitment processes. This approach complies with current research on human resource architectures with support by AI technologies (Raisch & Krakowski, 2021; Tambe, Cappelli, & Yakubovich, 2019; Van Esch, Black, & Ferolie, 2019). The first stage of this decision process will be dominated by those KPIs that are most constraint-oriented: The Location and Mobility Fit Index and the Contract Compatibility Score. These KPIs are characterized by extremely small rank differences and high recurrence rates. The function of these KPIs is not to evaluate candidates according to subtle differences but to check whether a certain profile can be considered for a given position at all. This approach complies with critical research on automated recruitment processes. The latter points to the feasibility of distinguishing clearly between feasibility checks and subsequent decision processes to reduce potential biases (Bogen & Rieke, 2018; Sánchez-Monedero, Dencik, & Edwards, 2020; Köchling & Wehner, 2020). Having determined this feasibility space, the subsequent phase considers various criteria related to structural and organizational compatibility, which are given by Seniority and Compensation Alignment and Cultural and Team Fit Score. These criteria are characterized by high average values, low variability, and high candidate concentration. These characteristics empirically validate their role as refinement criteria among already compatible candidates. These criteria guarantee that only candidates who are feasible are also structurally coherent with regard to the organization’s context, market positioning, and cultural framework in which the position operates. This approach to talent management and identification is in line with contemporary research on the development of AI-based talent management tools (Akhtar et al., 2019; Tambe et al., 2019). Having navigated this structural framework, the process now considers various criteria related to technical admissibility. The Hard Skill Coverage Ratio plays a pivotal role in this phase. The high average value of this indicator, combined with extremely high variability and high rank separation, suggests a bipolar distribution. This phenomenon can be best

explained by assuming a gatekeeping function. As such, this indicator is best suited to guaranteeing minimum standards are met rather than ranking candidates. This structure is in line with interdisciplinary research at the interface of psychology and machine learning. The latter requires a combination of admissibility criteria with subsequent ranking criteria to ensure an appropriate talent management process (Liem et al., 2018). It is only after establishing the feasibility, structural coherence, and technical admissibility that we should make use of the most discriminative ranking signals, which embedding-based and evidence-based KPIs offer. It is at this final stage that Soft Skill Evidence Density takes its position as the primary driver, with its considerably highest rank separation, indicating its capacity for differentiation on actual behavioral evidence. Hard Skill Proficiency Similarity and Soft Skill Semantic Alignment follow, providing continuous signals of proximity between candidates on a continuous scale, while Semantic Similarity, although less discriminative, provides a stabilizing effect on a lexical level. The patterns that emerge with regard to rank separation, variation, and recurrence of candidates across the data set itself justify this order and indicate that these KPIs function optimally when combined, rather than individually. The constraint-oriented KPIs guide the search space, structural alignment metrics stabilize and contextualize the search space, while requirement-based KPIs ensure technical viability, and finally, the highly discriminative semantic and evidence-based KPIs ensure a meaningful ranking. The proposed model operationalizes a concept of fit that combines soft skills, hard skills, and contextual and contractual conditions, moving beyond a semantic approach towards a more realistic, understandable, and governance-conscious approach to a recruitment-related decision-making process, which is in line with recent discussions on accountable and human-centered AI in hiring processes (Bogen & Rieke, 2018; Köchling & Wehner, 2020; Sánchez-Monedero et al., 2020). See Figure 15.



**Figure 15.** Beyond Keywords: A Multi-KPI, Staged Decision Framework for Candidate Ranking. Note. The figure decomposes candidate fit into nine KPIs across skills and context, organizing them into a three-phase pipeline that filters infeasible matches, enforces technical thresholds, and performs evidence-based, high-resolution ranking.

## 15. Orchestrating Hard Skills, Soft Skills, and Constraints in Job–Candidate Matching

The main value of the proposed framework does not lie in the identification of the "best" single metric but, rather, in the ability of the developed framework to demonstrate the orchestration of

multiple KPIs within a unifying decision process that reflects the realities of the recruitment practice. From the managerial point of view, the analysis supports the argument that the job-candidate matching process should not be viewed as an optimization task that relies on a single number or score. Rather, the process should be viewed as a multi-stage decision-making pipeline where different indicators play different albeit important roles, as supported by the recent studies on the use of AI-based recruitment systems and recommendation systems (Mashayekhi et al., 2024; Kurek et al., 2024). The first of the key implications of the developed framework is the clear differentiation between the KPIs that are primarily used as feasibility constraints and the KPIs that can be used for candidate ranking. For example, the Location and Mobility Fit KPI and the Contract Compatibility KPI are characterized by an extremely low separation of ranks and a very high candidate recurrence. This behavior of the respective indicators is similar to the behavior of binary filters. The most important application of such indicators is the identification of the pool of candidate profiles that can realistically be considered admissible for the respective job position. From the decision-maker's point of view, the application of such indicators should be performed at the initial stages of the decision-making process in order to ensure that the respective recommendation system does not produce technically optimal yet operationally impossible candidate recommendations. This logic of the initial filtering of the candidate pool aligns with the studies that emphasize the importance of the use of transparent constraint handling at the initial stages of the recruitment process (Raghavan et al., 2020; Mehrabi et al., 2021). The next layer of decision-making, which is represented by structural and contextual alignment metrics, includes Seniority and Compensation Alignment and Cultural and Team Fit. The patterns that have been noted, including high means, minimal variance, and a strong concentration of candidates, indicate that these KPIs primarily function as refinement signals within an already compatible group of candidates, rather than functioning as a broader form of selection. The managerial value that these metrics offer is primarily that they help mitigate risks related to potential issues with organization, economy, and culture, which recruiters must then navigate while seeking candidates that are not only feasible but also coherent with the context, team, and market position that a specific role represents, which is an increasingly important consideration within contemporary models of labor market analysis (World Economic Forum, 2023; WHO & ILO, 2022). The next form of decision-making that is represented is that which focuses on technical admissibility, with Hard Skill Coverage Ratio and gap-related metrics being highly relevant within this form of assessment. The extremely high variance that was noted with these metrics suggests a polarized form of behavior, with candidates being grouped into those that are clearly suitable and those that are clearly unsuitable, which indicates that their primary managerial value is likely that they function as a form of gatekeeping mechanism that enforces a minimum threshold, while also providing information that enables workforce planning, allowing an organization to identify candidates that are close to the ideal profile and for whom training investments could be strategically justified, a form of assessment that is highly relevant within contemporary models of AI-based matching and skill modeling (Li et al., 2025; Kurek et al., 2024). It is only after the feasibility, structural coherence, and technical adequacy of this evidence have been confirmed that it becomes possible to rely on the most discriminative ranking signals, namely those based on semantic similarity, with particular emphasis on evidence-based modeling of soft skills. The experimental results demonstrate that Soft Skill Evidence Density achieves the highest rank separation by far, thus being the most powerful discriminative signal for final-stage candidate selection. Hard Skill Proficiency Similarity and Soft Skill Semantic Alignment complement this ranking signal with continuous, semantically informed signals of proximity in technical and soft skills, respectively, while the baseline Semantic Similarity contributes a less discriminative but still useful lexical signal. This complex integration of embedding- and evidence-based signals follows a broader trend in intelligent recruitment systems, where deep learning and graph-based approaches can be used to boost fine-grained ranking performance (Li et al., 2025; Mashayekhi et al., 2024). This means that, for managers, the final stages of shortlisting can be aided by signals that capture subtle but crucial differences in professional narratives, behaviors, and practices. From the candidates' point of view, the implications of the multi-

KPI structure are as follows. The framework makes it evident that optimization along one dimension alone is not enough. Candidates will benefit from making their hard skills and soft skills explicit and, most importantly, from providing evidence of their behavioral skills. On the other hand, clarity regarding availability, mobility, contractual conditions, and seniority levels will have a direct impact on the candidates' visibility during the early stages of the selection pipeline. Profiles that are well-rounded and comprise both technical and soft skills, along with contextual conditions and constraints, will have a greater likelihood of passing through the successive stages of the pipeline and achieving high ranks during the final ordering. Thus, the system not only facilitates the selection but also, in a way, incentivizes candidates to provide more complete and realistic self-representation. All in all, the proposed ordering turns the matching problem from one-dimensional score-based matching into a decision process that considers constraints, requirements, context, and high-dimensional semantic signals. The ordering makes the system more explainable and controllable and better aligns with the needs of recruitment governance. The system will enable managers to better handle complexity and candidates to better understand how different dimensions of their profiles contribute to their employability (Raghavan et al., 2020; Mehrabi et al., 2021).

## 16. Limitations and Future Research Directions of the Multi-KPI Matching Framework

Despite the methodological and practical advantages of the proposed method of defining the multi-KPI framework, several limitations should be mentioned and may serve as guidelines for further study. First and foremost, the dataset's scope and composition may be seen as one of the limitations. Although using real-world job postings and CVs with considerable variability in terms of diversity and writing style may be seen as advantageous, the evaluation of the proposed method's effectiveness is still limited to one professional domain and one medium-sized dataset. Although this may be seen as advantageous in terms of ensuring greater internal consistency and control over the experiment's conditions, the evaluation of the proposed method's transferability to other domains remains questionable. As discussed in recent surveys of e-recruitment systems (Mashayekhi et al., 2024), one of the most important issues facing any e-recruitment system is its transferability from one domain to another. The patterns of system performance may not necessarily transfer from one domain to another. Domains such as the healthcare industry, manufacturing, and creative fields may have different patterns of skill representation, soft skills, and/or contractual conditions that may affect the system's relative effectiveness in terms of the proposed KPIs. Another important issue that may be seen as one of the limitations of the proposed method and framework is the quality and structure of the input data. Indeed, the proposed framework may assume that the job postings and CVs will have considerable textual content. However, as discussed in our recent study on semantic analysis and named entity recognition in recruitment (Ifakir et al., 2025), the quality of the extracted information may depend on the clarity and quality of the input. Although the proposed multi-KPI framework may alleviate some of these problems by using the weighted average of the indicators, the quality of the output still depends on the quality of the input. A third limitation pertains to the framework's reliance on existing vocabularies and heuristics for the derivation of some of the KPIs, such as HSCR and SGI. Although the use of vocabularies and heuristics is beneficial for transparency and auditability, there is a possibility that some of the vocabularies and heuristics might not be effective for new technologies and skills, which could lead to a lack of inclusion of some candidates' skills. This is a limitation that has also been discussed in systematic reviews on algorithmic HR decision-making systems (Köchling & Wehner, 2020). A fourth limitation pertains to the semantic models used for the derivation of embedding-based KPIs. Although multilingual transformers have the potential for providing a general-purpose solution, it is possible that such a solution might not be fine-tuned for a specific application, such as a recruitment application. This is a limitation that has also been discussed in recent deep learning research on resume-job matching (Li et al., 2025). Moreover, the use of embeddings does not inherently incorporate logical relationships, which necessitates the use of rule-based indicators. This also supports the argument that the framework's

use of semantic alignment is not absolute for determining suitability. This limitation also supports the argument that a more formalized strategy for optimizing the different indicators for a final decision is not provided by the framework. Although the framework has emphasized the use of a staged approach for orchestration, the final decision still relies on human judgment or policy with regard to the different indicators. This limitation has also been discussed in the literature on algorithmic hiring governance (Sánchez-Monedero et al., 2020). Lastly, the issue of fairness, biases, and ethics is another significant limitation of the framework. Although the multi-KPI framework improves the overall level of transparency and auditability, as opposed to the monolithic black-box approach, fairness, biases, and ethics are not guaranteed in the framework's design. Biases, if incorporated during the data, language models, or even the definition of the constraints and criteria, may still influence the final ranking results. It is an established fact that algorithmic hiring tools, if not adequately evaluated, may result in the perpetuation of biases and inequalities among different types of profiles, including non-traditional ones, as emphasized in the recent study of algorithmic hiring tools (Raghavan et al., 2020; Mehrabi et al., 2021). To summarize, although the framework successfully proves the viability of the overall benefits of the multidimensional, interpretable-based approach to job-candidate matching, its limited scope, dependency on the quality of the provided data, vocabularies, and language models, as well as the lack of formalization of the decision aggregation layer, are significant limitations that need to be addressed in future work.

## 17. Conclusions

In this study, the multi-KPI approach is proposed and empirically tested as a framework for job candidate matching, recasting the recruitment process as a multi-dimensional, interpretable, and governance-conscious decision support task instead of an optimization problem with a single performance measure. The experiments carried out on a realistic corpus of heterogeneous job offers and candidate CVs show that semantic matching using embedded representations offers a solid and scalable approach that can produce well-structured shortlists of relevant candidates even in multilingual and stylistically varied environments. At the same time, the study also reveals that semantic similarity is not sufficient on its own to fully capture the richness of recruitment decisions since it does not take into account mandatory requirements, skill gaps, behavioral indicators, organizational context, or operational requirements. The main contributions of this study are the empirical demonstrations of how different KPIs vary systematically in terms of score distributions, ranking separation, or candidate concentration, showing how these differences can be used to build a multi-stage and functionally differentiated decision support process. The coverage- or gap-based KPIs HSCR and SGI are shown to be useful for imposing hard technical requirements or highlighting skill gaps for the purposes of selective or developmental recruitment strategies. The embedding-based KPIs HSPS and SSSA are shown to be useful for providing continuous measures of technical or behavioral proximity between the candidate pool and the ideal candidate profile, thus facilitating fine-grained ranking of technically admissible candidates. The evidence-based KPI SSED is shown to be an important discriminator that can shift the recruitment process from declared competencies to observable behavior, thus greatly improving the robustness of soft skill evaluation. Finally, the contextual/contractual KPIs CTFS, CCS, LMFI, and SCA are shown to be useful for imposing hard or soft requirements that can filter the candidate pool while keeping the recommendations organizationally relevant. From a practical perspective, this framework provides a clear way to combine hard skills, soft skills, and context constraints within a single framework. Rather than forcing heterogeneous signals into a monolithic, uninterpretable score, this system maintains the semantic and decision-theoretic integrity of each signal while enabling their orchestration in a transparent, controllable, and auditable way. This directly addresses a major limitation of many existing AI-based recruitment frameworks, which often require a compromise on explainability, control, and compliance in favor of predictive power. In contrast, the multi-KPI framework presented in this work demonstrates that scalability, semantic robustness, explainability, constraint control, and human-in-the-loop decision-making can coexist. The originality of this work does not consist in the proposal of

a new matching algorithm, but rather in the decomposition of the concept of “fit” into a coherent set of complementary, interpretable KPIs, and their integration within a unified framework for analysis. This shift in methodology also shifts the perspective on job-candidate matching from a predictive/retrieval framework to a multi-criteria decision support framework, which is closer to real-world practices and compliance. The modularity of this architecture also makes it more relevant to practical needs, as it allows for the integration of new dimensions of fit, new constraints, or new indicators without requiring a complete redesign of the system. With regard to innovation, the framework shows how specific, rule-based measures, continuous semantic similarities, and evidence density indicators can all be integrated in one single pipeline, with each one offering a different, theoretically grounded perspective on the question of compatibility. Empirical analysis of the ranking structures and score distributions shows that these indicators are not redundant, but offer different, complementary views of the matching process. This, in turn, supports the overall claim of the paper: that to support recruitment decisions, we need not better scores, but better decision processes. Overall, the study contributes to the field of AI-based recruitment by providing conceptual and empirical evidence of how multi-dimensional, explainable, and governance-aware matching systems can be developed in practice. The proposed approach provides the foundation for further development toward more accountable, human-centric, and policy-aligned decision support systems for matching in the labor market.

**Acknowledgments:** This research was supported by the project “LUtech Campus Ecosystem – LUCE”, Project Code 22ROJB5, funded under a subsidized financing scheme of the Puglia Region within the framework of a Program Agreement, Contratto di Programma. The authors gratefully acknowledge this financial support, which made this study possible.

## References

- Ajjam, M.-H., & Al-Raweshidy, H. S. (2026). AI-driven semantic similarity-based job matching framework for recruitment systems. *Information Sciences*, 724, 122728.
- Akhtar, R., Winsborough, D., Lovric, D., & Chamorro-Premuzic, T. (2019). Identifying and managing talent in the age of artificial intelligence. In *Workforce readiness and the future of work* (pp. 169-185). Routledge.
- Al-Dmour, R., Al-Dmour, H., Al-Dmour, A., Basheer Amin, E., & Al-Dmour, Y. (2025). AI and big data-driven social media recruitment: The mediating role of talent acquisition and employee engagement in bank performance. *Digital Transformation and Society*, 1–19.
- Aleisa, M. A., Beloff, N., & White, M. (2023). Implementing AIRM: a new AI recruiting model for the Saudi Arabia labour market. *Journal of Innovation and Entrepreneurship*, 12(1), 59.
- Alqudah, N., Abuein, Q. Q., & Shatnawi, M. Q. (2025). AI-driven hyper-personalization and transfer learning for precision recruitment. *IAES International Journal of Artificial Intelligence*, 14(5), 4271–4278.
- Ammupriya, A., Deivanai, S., Niranchana, A., Naveenkumar, S., & Kumaran, S. (2025). Artificial intelligence in recruitment: Automating candidate screening and talent acquisition. In *Proceedings of the IEEE 4th World Conference on Applied Intelligence and Computing (AIC 2025)* (pp. 234–239).
- Andabayeva, G., Movchun, V., Dubovik, M., Kurpebayeva, G., & Cai, X. (2024). Labor market dynamics in developing countries: analysis of employment transformation at the macro-level. *Journal of Innovation and Entrepreneurship*, 13(1), 65.
- Aumayr-Pintar, C., & Baggio, M. (2025). Taking stock: Further experiences in gender pay transparency implementation and effectiveness.
- Bachmann, R., Frattini, F. F., Hauret, L., Kirov, V., Lewandowski, P., Martin, L., ... & Zierahn-Weilage, U. Skills gaps, skill and labour shortages, and mismatch—Existing evidence.
- Badouch, M., & Boutaounte, M. (2025). Strategies for AI and Big data in recruitment.
- Barath, G. V., Sanjev, R., Priya, V. S., Nabi, F. G., & Ravindran, S. (2025). AI-driven recruitment: Resume screening and skill matching with NLP. In *Proceedings of the 2025 IEEE International Conference on Computer, Electronics, Electrical Engineering and Their Applications (IC2E3 2025)*.

- Bogen, M., & Rieke, A. (2018). Help wanted: An examination of hiring algorithms, equity, and bias. *Upturn*, December, 7.
- Burbridge, C. (2025). Living and working in Europe 2024.
- Çelik Ertuğrul, D., & Bitirim, S. (2025). Job recommender systems: A systematic literature review, applications, open issues, and challenges. *Journal of Big Data*, 12(1), 140.
- Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., & Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world?. *Industrial and Organizational Psychology*, 9(3), 621-640.
- Chau, A. (2025). The Paradox of Visibility and the ILO's Decent Work for Domestic Workers Convention. In *Human Rights and the United Nations* (pp. 102-116). Routledge.
- Chihab, M., Boussetta, H., Chiny, M., Chihab, Y., & Hadi, M. Y. (2025). AI-driven professional profile categorization and recommendation system. *International Journal of Advanced Computer Science and Applications*, 16(11), 378-385.
- Chinn, D., Hieronimus, S., Kirzherr, J., & Klier, J. (2020). The future is now: Closing the skills gap in Europe's public sector. *McKinsey & company*, 27, 2020.
- Chung, D. J., Kim, B., & Park, B. G. (2019). How do sales efforts pay off? Dynamic panel data analysis in the Nerlove-Arrow framework. *Management Science*, 65(11), 5197-5218.
- Dalal, R. S., Alaybek, B., & Lievens, F. (2020). Within-person job performance variability over short timeframes: Theory, empirical research, and practice. *Annual Review of Organizational Psychology and Organizational Behavior*, 7(1), 421-449.
- Dangeti, S. R., Tejaswini, P., Dharani, S., Sravani, V., & Jaya Sri, R. (2026). AI-powered resume parsing and candidate scoring for efficient hiring workflows. In *Lecture Notes in Networks and Systems* (Vol. 1493 LNNS, pp. 73-80). Springer.
- Di Battista, A., Grayling, S., Hasselaar, E., Leopold, T., Li, R., Rayner, M., & Zahidi, S. (2023, November). Future of jobs report 2023. In *World Economic Forum* (pp. 978-2). Geneva, Switzerland: World Economic Forum.
- Drozdz, L. A., & Tavares, M. (2024). Generative AI: A turning point for labor's share. *Economic Insights*, 9(1), 2-11.
- Escribano, E. (2024). A New Model Tax Convention for a World of Increasing Remote Work and Mobility of Individuals. *World Tax Journal*, 16(2).
- Es-Said, B., Badouch, M., Mahmoud, H., & Boutaounte, M. (2025). Enhancing hiring processes with e-learning and recommendation systems. In *Emerging technologies for recruitment strategy and practice* (pp. 1-25).
- Frazzetto, P., Haq, M. U. U., Fabris, F., & Sperduti, A. (2025). From Text to Talent: A Pipeline for Extracting Insights from Candidate Profiles. *arXiv preprint arXiv:2503.17438*.
- Haneef, F., Varalakshmi, M., & Peer Mohamed, P. U. (2025). Leveraging RAG for effective prompt engineering in job portals. In *Proceedings of the 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN 2025)* (pp. 717-721).
- Hepzibah, A. R., Jeyasanthi, J., Devi, D. G., Senthil, S., & Ramnath, M. (2025). AI-powered resume screening system for smart hiring: Leveraging NLP and large language models for efficient and fair recruitment. *Journal of Theoretical and Applied Information Technology*, 103(23), 10146-10155.
- Hernández-Alvarez, M., & Torres-Hernández, E. (2025). Decentralized systems and AI: Exploring the intersection of Web3, blockchain, and predictive machine learning. In *Communications in Computer and Information Science* (Vol. 2529, pp. 304-311). Springer.
- Ifakir, I., Mohtaram, N., Nfaoui, E. H., Zannou, A., & El Hassouni, M. (2025). An Approach Based on Named Entity Recognition and Semantic Analysis for Recruitment Efficiency and Optimization. *International Journal of Advanced Computer Science & Applications*, 16(8).
- Jha, R., Paliwal, G., & Jha, B. K. (2025). AI-powered resume builder: Enhancing job applications with artificial intelligence. In *Proceedings of the 3rd International Conference on Disruptive Technologies (ICDT 2025)* (pp. 1655-1660).
- Kaya, M., & Bogers, T. (2025). Mapping stakeholder needs to multi-sided fairness in candidate recommendation for algorithmic hiring. In *Proceedings of the 19th ACM Conference on Recommender Systems (RecSys 2025)* (pp. 257-267).

- Khan, H. W., Sattar, M. U., Noor, S., & Alyousef, M. I. (2025). A personality-informed candidate recommendation framework for recruitment using MBTI typology. *Information*, 16(10), 863.
- 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>
- Kim, S. (2026). AI-driven hiring: A boon or a barrier to finding the right talent? *AI & Society*, 41(1), 557–564.
- Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, 13(3), 795-848.
- Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, 13(3), 795-848.
- Kulkarni, M. V., Khan, M. H. A., Raje, R. R., Wheeler, R., & Ganci, A. M. (2025). RealCV: An AI-powered resume generator. In *Proceedings of the IEEE Inter*
- Kumar, V., Sangeetha, G., Gopal, K., & Kumar, P. (2025). JobNexus: AI powered recruitment solution. In *Proceedings of the 2025 International Conference on Computing and Communication Technologies (ICCCCT 2025)*.
- Kurek, J., Latkowski, T., Bukowski, M., Świdorski, B., Łępicki, M., Baranik, G., ... & Dobrakowski, Ł. (2024). Zero-shot recommendation AI models for efficient job-candidate matching in recruitment process. *Applied Sciences*, 14(6), 2601.
- Li, H., Tang, X., Liu, Q., Liu, B., & Zhao, S. (2025). Enhancing intelligent recruitment with generative pretrained transformer and hierarchical graph neural networks: Optimizing resume-job matching with deep learning and graph-based modeling. *Journal of Organizational and End User Computing*, 37, 1–24. <https://doi.org/10.4018/JOEUC.377613>
- Liem, C. C., Langer, M., Demetriou, A., Hiemstra, A. M., Sukma Wicaksana, A., Born, M. P., & König, C. J. (2018). Psychology meets machine learning: Interdisciplinary perspectives on algorithmic job candidate screening. In *Explainable and interpretable models in computer vision and machine learning* (pp. 197-253). Cham: Springer International Publishing.
- Liu, J., Fu, Y., & Luo, Y. (2026). Smart resume filter for vocational qualification analysis. In *Communications in Computer and Information Science* (Vol. 2540 CCIS, pp. 118–134). Springer.
- Martínez-Manzanares, M. E., Urias-Paramo, J. J., Waissman-Vilanova, J., & Figueroa-Preciado, G. (2024). An Empirical Job Matching Model based on Expert Human Knowledge: A Mixed-Methods Approach. *Applied Artificial Intelligence*, 38(1), 2364158.
- Mashayekhi, Y., Li, N., Kang, B., Lijffijt, J., & De Bie, T. (2024). A challenge-based survey of e-recruitment recommendation systems. *ACM Computing Surveys*, 56(10), 1-33.
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6), 1-35.
- Mer, A. (2023). Artificial Intelligence in Human Resource Management: Recent Trends and Research Agenda. *Contemporary Studies in Economic and Financial Analysis*, 111, 31-56.
- Mindell, D. A., & Reynolds, E. (2023). *The work of the future: Building better jobs in an age of intelligent machines*. Mit Press.
- Nabila, C., Ayyoub, C., Abdelhak, S. E., & Fairouz, N. (2025). How to optimize recruitment with artificial intelligence: A case study on resume analysis and matching with Shneider AI Recruit. *Proceedings on Engineering Sciences*, 7(4), 2181–2188.
- Nandi, A., Das, A. K., Sur, H., Das, A., & Podder, S. (2025). AI-driven intelligent system for personalized job recommendations using real-time skill identification and industry trend analysis. In *Proceedings of the International Conference on Computing Intelligence and Application (CIACON 2025)*.
- Nikolaou, I. (2021). What is the Role of Technology in Recruitment and Selection?. *The Spanish journal of psychology*, 24, e2.
- Peña-Casas, R., Ghailani, D., & Sabato, S. (2025). Building resilient and inclusive labour markets in Europe: unpacking policy synergies and challenges. In *Deliverable. WeLar Project*.

- Perea-Trigo, M., Botella-López, C., Martínez-del-Amor, M. Á., Álvarez-García, J. A., Soria-Morillo, L. M., & Vegas-Olmos, J. J. (2024). Synthetic corpus generation for deep learning-based translation of spanish sign language. *Sensors*, 24(5), 1472.
- Potočnik, K., Born, M. P., & Kristina/Anderson Potočnik (Nei). (2024). Recent Developments in Recruitment and Selection.
- Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020, January). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 469-481).
- Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of management review*, 46(1), 192-210.
- Sajjadiani, S., Sojourner, A. J., Kammeyer-Mueller, J. D., & Mykerezzi, E. (2019). Using machine learning to translate applicant work history into predictors of performance and turnover. *Journal of Applied Psychology*, 104(10), 1207.
- Sánchez-Monedero, J., Dencik, L., & Edwards, L. (2020, January). What does it mean to 'solve' the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 458-468).
- Saouabe, A., Boulahoual, A., & Oualla, H. (2025). AI-driven recruitment: Enhancing profile recommendations in digital-age management. In *Lecture Notes in Networks and Systems* (Vol. 1548, pp. 433-440). Springer.
- Sawant, A., Yadav, S., Shah, A. H., Kolekar, U., & Balpande, S. (2026). Designing and developing comprehensive framework with improved insights for optimizing talent acquisition process of engineering graduates. In *Lecture Notes in Electrical Engineering* (Vol. 1460 LNEE, pp. 99-109). Springer.
- Singla, L., Parihar, B., Paliwal, N., Muwal, M., & Ahuja, V. (2025). A hybrid transformer-based human resource recruitment system for efficient business process management. In *Proceedings of the 5th International Conference on Intelligent Technologies (CONIT 2025)*.
- Stevano, S. (2024). The Devaluation of Essential Work: An Assessment of the 2023 ILO Report. *Development & Change*, 55(4).
- Sukri, S., Samsudin, N. A., Fadzrin, E., Ahmad Khalid, S. K., & Trisnawati, L. (2024). Word2vec-based Latent Semantic Indexing (Word2Vec-LSI) for Contextual Analysis in Job-Matching Application. *International Journal of Advanced Computer Science & Applications*, 15(3).
- 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.
- Turba, L. Marta. "TR Is Key to Successfully Integrating AI and Work." *The Journal of Total Rewards* 33.1 (2024): 1-1.
- Van Esch, P., Black, J. S., & Ferolie, J. (2019). Marketing AI recruitment: The next phase in job application and selection. *Computers in Human Behavior*, 90, 215-222.
- Van Iddekinge, C. H., Roth, P. L., Putka, D. J., & Lanivich, S. E. (2011). Are you interested? A meta-analysis of relations between vocational interests and employee performance and turnover. *Journal of Applied Psychology*, 96(6), 1167.
- Vazquez-Alvarez, R., Xu, D., & Belser, P. (2022). Global wage report 2022-23: the impact of inflation and COVID-19 on wages and purchasing power. *International Labour Organization*, 10.
- Wahyuningrum, S. R., Ghofur, A., Mufrihah, A., Hasanah, I., & Aisa, A. (2025). Self-adaptive systems: Redefining best practices in AI and big data in recruitment. In *Emerging technologies for recruitment strategy and practice* (pp. 27-58).
- World Health Organization (WHO). (2022). *International Labour Organization (ILO)*. (2022). *Mental health at work*.
- Wu, X., Liu, K., Wang, J., Lv, R., & Song, J. (2025). Candidate evaluation with multimodal data-driven for recruitment. In *Lecture Notes in Computer Science* (Vol. 15308, pp. 81-96). Springer.
- Yadav, K., Yuvaraj, S., Jugran, S., Sharma, G., & Koilraj, T. (2025). AI-powered recruitment: Enhancing hiring efficiency and candidate experience in modern HR. In *Proceedings of the 2025 International Conference on Technology Enabled Economic Changes (InTech 2025)* (pp. 141-147).

Yeju, V. V., Lakshman, S., & Prathima, T. (2026). A systematic review of application tracking systems with virtual assistance: Current trends and future directions. In *Lecture Notes in Networks and Systems* (Vol. 1505 LNNS, pp. 537–549). Springer.

Zadykian, V., Andrade, B., & Afli, H. (2025). Towards explainable job title matching: Leveraging semantic textual relatedness and knowledge graphs. arXiv. <https://doi.org/10.48550/arxiv.2509.09522>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.