

Recommendation System of Client-Requested Projects in the Swedish Consultancy Market with LLMs

Diogo Buarque Franzosi^{1,2,*}, Kristijan Capovski¹, Maycel Isaac¹ and Stefan Byttner³

¹Synteda AB
Gothenburg, Sweden

²Blekinge Institute of Technology
Karlskrona, Sweden

³Halmstad University
Halmstad, Sweden

Abstract

This article presents the recommendation system of Personas, a microservice-based platform designed to assist Human Resources (HR) teams in streamlining the recommendation and presentation of candidates to clients based on posted project descriptions. Personas offers functionalities for recommendation, automatic generation of tailored curricula and motivation letters, and conversational support through client- and consultant-facing chatbots.

At its core, the recommendation system suggests relevant projects posted by clients to each candidate on a daily basis. It leverages both structured and unstructured textual data, including web-scraped content, user-uploaded documents, curated profiles, and texts generated by Large Language Models (LLMs). All documents are embedded into a shared semantic vector space, enabling fast similarity computations and facilitating Retrieval-Augmented Generation (RAG) workflows.

The recommendation pipeline consists of a two-stage process. First, lightweight pre-selection models apply filters and semantic similarity metrics to narrow down the pool of potential assignments. Then, in-depth analyses using LLMs provide refined compatibility assessments. Notably, the LLM-based evaluations serve not only to improve ranking precision, but also as high-quality proxy labels for evaluating and improving pre-selection models.

This paper describes each stage of the pipeline—document collection, structuring, curation, pre-selection, and LLM-based analysis—and presents quantitative results demonstrating the system's effectiveness on a large-scale dataset.

Keywords

Recommendation System, Semantic Similarity, Large Language Models, Job Matching

1. Introduction

Recruitment processes are becoming increasingly complex, involving large volumes of candidate profiles, job postings, and client-specific requirements. Traditional systems struggle to scale

SAIS2025: Swedish AI Society Workshop 2025, 16-17 June 2025, Halmstad, Sweden.

*Corresponding author.

✉ diogo@synteda.com (D. Buarque Franzosi); kristijan.capovski@synteda.com (K. Capovski); mi@synteda.com (M. Isaac); stefan.byttner@hh.se (S. Byttner)

🌐 <https://www.synteda.se/> (M. Isaac)

🆔 0000-0003-1432-5255 (D. Buarque Franzosi)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

efficiently with this complexity while providing high-quality matches. In recent years, artificial intelligence (AI) and machine learning (ML) have been increasingly adopted to support and automate parts of the recruitment pipeline, such as resume screening, job recommendation, and candidate ranking. In particular, the rise of Large Language Models (LLMs) has opened new possibilities for semantic matching and deep content analysis in Human Resources (HR) tools [1].

This paper presents the recommendation system of *Personas*, a microservice-based platform that supports HR teams in streamlining the recommendation and presentation of candidates to different clients according to posted project descriptions. *Personas* includes tools for candidate recommendation, automatic generation of personalized curricula and motivation letters, and conversational support via consultant and client-facing chatbots. At the core of this platform lies a recommendation system that automatically suggests relevant client-requested projects to each candidate, typically on a daily basis.

The system leverages both structured and unstructured data sources, including web-collected projects posted by clients, user-uploaded resumés, and documents curated or generated by LLMs. All documents are embedded into a vector database using pre-trained sentence embedding models, allowing efficient semantic comparisons through cosine similarity and enabling Retrieval-Augmented Generation (RAG) techniques.

To handle the large volume of incoming client requests, *Personas* uses a two-stage recommendation pipeline. First, a pre-selection stage quickly filters and ranks client requests using lightweight similarity-based models. Then, an in-depth analysis stage applies LLM-based scoring methods to assess candidate-assignment compatibility with greater nuance. Interestingly, these in-depth LLM analyses also serve a dual role as soft labels for evaluating the quality of pre-selection models—supporting a continuous improvement cycle for the recommendation system.

This paper provides a detailed description of each component of the recommendation pipeline, including document collection and curation, embedding and structuring, pre-selection, in-depth LLM analysis, and evaluation procedures. We support this discussion with experimental results based on a large corpus of over 35,000 assignment projects requested by clients and a curated sample of candidate CVs. The results highlight the effectiveness of in-depth LLM models and their utility as proxies for human evaluation.

Despite overlapping with broader recruitment practices, the consultancy market—particularly within regional environments—poses distinct challenges. These include high turnover rates, rapid project-based hiring cycles, and the need for precise skill-client alignment under tight deadlines. Furthermore, assignment descriptions and candidate resumes are often maintained in multiple languages depending on client and candidate backgrounds. In the case of our study, we focus on a Swedish-English dual-language context, which introduces additional semantic and syntactic complexity. Standard online job recommendation systems often overlook these localized and multilingual nuances. This work aims to address these gaps by adapting LLM-based scoring to the consultancy domain and by demonstrating a robust performance across linguistic variations.

The rest of the paper is organized as follows: Section 2 describes related work. Section 3 introduces the methodology and technical design of the *Personas* recommendation system. Section 4 presents the experimental setup and quantitative results. Section 5 discusses the

implications of our findings and outlines limitations. Finally, Section 6 concludes with future directions.

2. Related Work

Recent work has explored the application of Large Language Models (LLMs) to job recommendation systems from various perspectives. For instance, Kavas *et al.* [2] illustrates a multilingual, hybrid system that combines LLMs and recruiter input for better CV-job matching. [1] experiments on extracting, matching and ranking skills between CV and job profiles. Zheng *et al.* [3] and Wu *et al.* [4] examine generative and graph-based approaches using LLMs for candidate-job alignment. Other studies investigate multilingual and zero-shot matching techniques [5, 6], as well as hybrid recommendation pipelines [7, 8]. A broader review of LLM applications in recommendation tasks is provided in surveys by Wu *et al.* [9] and Hou *et al.* [10], highlighting the rapid convergence of recommendation systems and large-scale language modeling. A summarization of the how the concept presented in this paper relates to previous presented work is shown in Table 1.

Table 1

Comparison of Related Work and the Personas Recommendation System.

Feature / Method	Prior Work	This Paper (Personas System)
Multilingual Matching	✓	✓ (focus on Swedish-English context)
Human-in-the-Loop Curation	✗ or limited	✓ (frontend for review and editing)
LLMs for Match Scoring	✓	✓ (with structured, explainable prompts)
Embedding-Based Pre-selection	Partial ✓	✓ (three models evaluated)
LLM Scores as Soft Labels for Training	✗	✓ (used to supervise pre-selection)
Real-World Consultancy Deployment	✗	✓ (live system, 35,000+ assignments)
Integration with Chatbots and RAG Tools	✗	✓ (used for summarization and QA)

3. Methodology

The goal of this article is to present the design and evaluation of the Personas recommendation system, with a particular focus on how lightweight pre-selection models, supported by LLM-based in-depth analyses, can effectively streamline candidate–assignment matching at scale.

This section describes the architecture, data processing steps, and algorithms used in the Personas recommendation system. The system is composed of several modular components responsible for document ingestion, structuring, pre-selection of client assignments for each candidate, and in-depth analysis of candidate-assignment pair using LLMs. Figure 1 provides an overview of the pipeline. We use the terms client request, project, and assignment interchangeably to refer to job descriptions posted by clients, which outline specific tasks to be completed within a defined time frame.

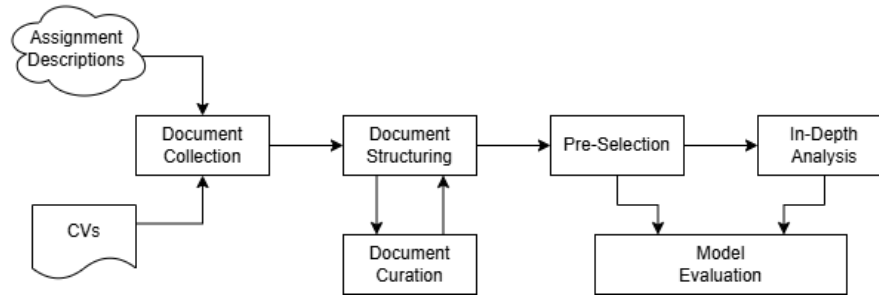


Figure 1: Personas' recommendation system flow chart.

3.1. Personas Recommendation System

Figure 1 shows the data flow of Personas' recommendation system. First, data is collected, both assignments posted by clients from the web and resumes are uploaded by HR and consultants. These documents are further processed using LLMs or curated via human-in-the-loop pipelines. Third, the pre-selection of assignments for each candidate takes place as a first step of the recommendation system. Finally, the pre-selected assignments are further analyzed in detail by LLM-based chains. Both pre-selection and in-depth models are used for evaluation. Each component is further described in the following sections.

3.1.1. Document Collection

Every day, the system collects hundreds of new client request descriptions from various sources, including public job boards, company websites, and internal client submissions. Data collected are cleaned, parsed, and stored in a standardized text format. Additionally, candidate CVs are uploaded directly by users or HR consultants through the Personas platform. Each document, whether a CV or an assignment description, enters the same downstream pipeline for semantic processing.

3.1.2. Document Structuring and Summarization

All collected documents are transformed into high-dimensional semantic vectors using OpenAI's text-embedding-ada-002 model. This model encodes natural language into 1536-dimensional dense embeddings that preserve contextual semantics, enabling fine-grained comparison between documents and document components.

To enhance interpretability and facilitate targeted similarity matching, raw documents are restructured into standardized JSON schemas using prompting techniques with LLMs. The structuring prompt processes heterogeneous file formats (PDFs, Word documents, plain text) and outputs a normalized representation. For candidate CVs, we extract fields such as:

- Personal Information
- Short Description
- Experiences

- Skills
- Education
- Publications
- Achievements
- Courses
- Certificates

We use two additional LLM prompts to separately extract:

- Candidate Keywords – terms related to competencies, tools, and domains of expertise.
- Candidate Roles – role titles or job functions expressed in the document.

For client request descriptions, a parallel extraction process retrieves:

- Required skills and technologies
- Language
- Required skills
- Good to have skills
- Soft skills
- Location
- Company
- Industry sector

These summarizations and document structuring are achieved using services provided by Personas, which are connected to an Agents RESTful API, introduced in [11]. It contains simple prompt chains as well as RAG chains and ReAct agents.

These structured fields are used in subsequent matching stages. Prompt templates and formatting instructions are available in supplemental material.

3.1.3. Document Curation

The Personas system incorporates a human-in-the-loop approach for document curation, involving both candidates and sales consultants. Once documents—such as candidate summaries, generated CVs, or motivation letters—are automatically produced, they can be reviewed, edited, or enriched through a dedicated user interface before being presented to clients. This curation process ensures higher-quality, context-aware content, better aligned with client expectations and market standards.

Beyond presentation, curated documents play an important role in improving the overall recommendation system. By refining or correcting the automatically extracted or generated information, the system benefits from more accurate and relevant data in subsequent matching steps. Personas includes a frontend platform specifically designed to support this interactive workflow, allowing users to annotate, validate, or update content with minimal friction.

3.1.4. Pre-Selection

The pre-selection stage aims to efficiently reduce the candidate search space from thousands of potential assignments to a manageable shortlist suitable for more intensive analysis. This is achieved through lightweight semantic similarity models that rank assignments for each candidate. In our daily pipeline, these methods are run over the assignments collected in the same morning, but the endpoints provided by the tool allow also a more flexible range of assignment poolk based on the collected date.

In this study, we evaluate three pre-selection models:

PS1: Skill-to-Skill Matching

This model evaluates how well a candidate’s skills match the requirements of a given assignment by comparing their respective skill sets using semantic embeddings. Specifically, for each assignment-candidate pair, we represent the required skills of the assignment as a set of text embeddings \mathbf{a}_i , and the candidate’s skills as another set \mathbf{c}_j , where both sets are derived from the summarization process described in Section 3.1.2. The embeddings are generated using the text-embedding-ada-002 model.

To measure similarity, for each assignment skill embedding \mathbf{a}_i , we find the closest matching candidate skill embedding \mathbf{c}_j based on cosine distance:

$$d_i = \min_j (\text{distance}(\mathbf{a}_i, \mathbf{c}_j)), \quad (1)$$

where cosine distance is defined as:

$$\text{distance}(\mathbf{a}, \mathbf{b}) = 1 - \cos(\mathbf{a}, \mathbf{b}), \quad (2)$$

with values ranging from 0 (identical vectors) to 2 (completely dissimilar).

The final matching score is computed by averaging the minimal distances across all required skills, then scaling the result to a 0–100 range:

$$s = \frac{2 - \text{mean}(d_i)}{2} \times 100. \quad (3)$$

This approach generalizes simple keyword matching by capturing semantic similarity between skills. For example, a perfect match between all assignment and candidate skills yields $d_i = 0$ for all i , resulting in a score of 100.

One problem with this algorithm is that it grows with $\mathcal{O}(N \times M \times K)$ the number of candidates (N), number of assignments (M) and number of required skills in assignment (K).

PS2: Keyword-to-Assignment Matching

This model assesses the overall relevance of a candidate’s competency profile to a given assignment by leveraging a vector similarity search engine—specifically, the built-in nearest neighbor (NN) search provided by Chroma. Unlike the previous model (PS1), which computes

pairwise distances between individual skill embeddings, this approach compares aggregated representations of candidate and assignment data.

The candidate’s profile is represented as a single embedding vector derived from the embeddings of their extracted keywords. The assignment is similarly represented by a single embedding computed from its full textual description. These embeddings are generated using the same text-embedding-ada-002 model described earlier.

To compute the similarity, the model uses Chroma’s internal similarity search mechanism, which indexes the assignment embeddings and allows for efficient retrieval of the most relevant assignments for a given candidate query embedding. The matching score is determined by the cosine similarity between the candidate’s keyword embedding \mathbf{c} and the assignment embedding \mathbf{a} :

$$s = \frac{(2 - \text{distance}(\mathbf{c}, \mathbf{a}))}{2} \times 100, \quad (4)$$

where distance is given in Eq. 2. This results in a score between 0 and 100, with higher scores indicating stronger overall alignment between the candidate’s profile and the assignment content.

By comparing entire profiles rather than individual skills, this model captures broader semantic alignment, making it suitable for assessing general fit or potential suitability across loosely defined tasks.

This model solves the $\mathcal{O}(N \times M \times K)$ grow using NN search in Chroma, reducing it to $\mathcal{O}(N)$

PS3: Role-to-Title Matching

This model follows the same similarity search approach as PS2 but operates on different inputs. Instead of using candidate keywords and full assignment descriptions, it compares the embedding of the candidate’s identified roles with the embedding of the assignment title. This captures how closely the candidate’s professional identity or career path aligns with the nature of the role being offered.

3.1.5. In-Depth Analysis with Large Language Models

The client requests that pass the pre-selection phase are further evaluated using more computationally expensive but semantically rich LLM-based models. These models read both the candidate and client request documents in detail and produce a matching score from 0 to 100 based on nuanced semantic understanding.

We consider two in-depth LLM-based models:

M1: Generic Fit Scorer

This prompt reads the full CV and client request description and outputs a score indicating the candidate’s suitability for the role. The score is based on inferred relevance, skill overlap, and contextual cues. No training is involved—this is a prompt-based model using zero-shot capabilities of the LLM. Zero-shot LLM capabilities have shown good results in job matching [5].

The prompt template receives the full text of candidates’s resume and the full description of client request to output a score from 0-100 and a textual analysis.

M2: Role-Contextual Fit Scorer

This model also leverages the zero-shot capabilities of the LLM, but applies prompt engineering to decompose the evaluation into multiple targeted dimensions. The prompt instructs the model to assess the candidate’s fit by considering several key factors individually, resulting in a more structured and explainable score. The evaluation is broken down as follows: Skill Matching (0–40 points), Role Alignment (0–30 points), Strengths and Weaknesses Analysis (0–20 points), and Additional Considerations (0–10 points). This formulation emphasizes recent experience and contextual relevance, encouraging the model to focus not only on content overlap but also on transferable experience and strategic fit.

These prompts are designed to be interpretable, meaning they output not only a score but also a natural language justification, which can be logged for future audits or included in client reports. Both M1 and M2 prompt templates are available in the supplemental material.

3.1.6. Model Evaluation

The outputs of the in-depth models are used both as final recommendations and as proxy labels for evaluating the performance of pre-selection models.

We also include a subset of assignments that were manually tagged by HR experts, allowing for comparison between automated evaluations and human judgment. This triangulation enables us to quantify how well each automated stage replicates expert decisions.

4. Results

In this section, we present a first evaluation of the Personas recommendation system. We begin with a description of the dataset used in our experiments, followed by a qualitative assessments of in-depth LLM analyses compared to pre-selection models. We then analyze the performance of the pre-selection models and conclude by a quantitative analysis comparing automated scores to human evaluation labels.

4.1. Data Description

Our dataset consists of 35,679 client requested project descriptions collected from May 2024 to February 2025. These client requests span a wide variety of industries, roles, and technical domains. Among them, 4,328 assignments have received human annotations, indicating whether they were deemed relevant for a particular candidate by HR professionals.

To evaluate the system’s ability to model candidate relevance, we selected 20 candidate CVs representative of diverse experience levels and domains, ranging from junior software developers to senior project managers.

Each candidate was evaluated against a pool of assignments using both pre-selection and in-depth models, with correlations computed between the various scoring mechanisms and available human judgments.

4.2. Qualitative Assessment of LLM Analyses and Pre-Selection Models

The in-depth LLM analyses (M1 and M2) demonstrated the ability to produce nuanced evaluations, often surfacing insights that keyword-based models overlooked. For instance, the models could infer transferable skills—such as familiarity with Agile methodologies in project management—even when explicit technologies or terms were not mentioned.

Moreover, the LLMs exhibited sensitivity to temporal factors (e.g., recency of experience), job seniority, and domain-specific terminology. The generated justifications were coherent and often aligned closely with human reasoning, making them well-suited for explainable AI applications.

Crucially, the LLM analyses provided detailed, interpretable justifications for individual candidate-assignment matches. This level of granularity enables qualitative assessments of model behavior, which is not possible with pre-selection models that output only a numerical score. By analyzing these justifications, we gain confidence in the LLM-generated evaluations and can therefore use their scores as a reliable benchmark for assessing the performance of pre-selection models.

This analysis was conducted in collaboration with HR experts, who reviewed various reports daily over several months.

4.3. Evaluation of Pre-Selection Models

To evaluate the quality of the pre-selection stage, we compared the scores produced by each pre-selection model (PS1–PS3) against those generated by the in-depth LLM-based scoring models M1 and M2. The models were applied to assignment-candidate pairs collected during March 2025 and 13 candidates, each one presenting one or two CVs in either English or Swedish (total of 20 CVs). During this period, 2,547 assignments were collected. To manage computational complexity—particularly due to the pairwise comparisons required by PS1—we limited our evaluation to this subset rather than using the full database of assignments. Each pre-selection model suggests 15 assignments from the pool for each CV, culminating in 15 or 30 assignments per candidate. Each of these assignments are then analyzed by the LLM-based models M1 and M2. Figure 2 summarizes the results using two boxplots—one for M1 and one for M2—showing the distribution of scores across candidates selected by each pre-selection method.

We found no clear preference for any specific pre-selection model indicating that lighter models based on semantic searches can perform as well as skill-to-skill comparisons.

4.4. Comparison with Human Label

We also assessed the alignment between in-depth LLM scores and human expert labels across seven candidate profiles, denoted as C1, C2, C3, C4, C5, C6 and C7. The profiles extend expertise from IT programming industry to project manager. The human labels were tags (Good, Maybe or Bad) indicating whether each project was relevant for the candidate.

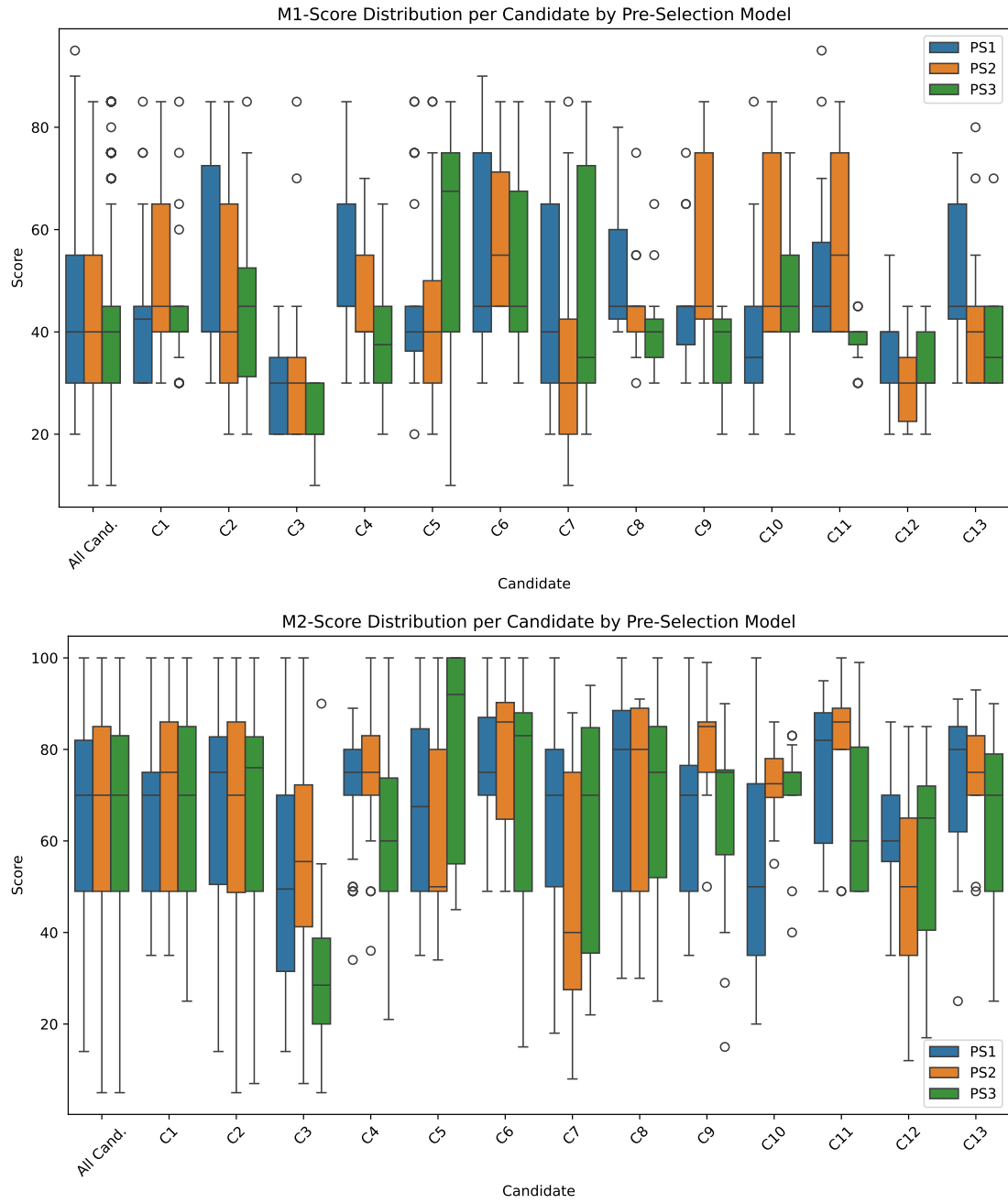


Figure 2: Distribution of M1-Scores (top) and M2-Scores (bottom) for candidates C1-C13 for assignments collected in March.

Figure 3 shows the Spearman's correlation between LLM scores and human labels for all candidates and for each candidate separately. On the left only assignments in Gothenburg, illustrating the possible bias of personal preference, since the company is based in Gothenburg

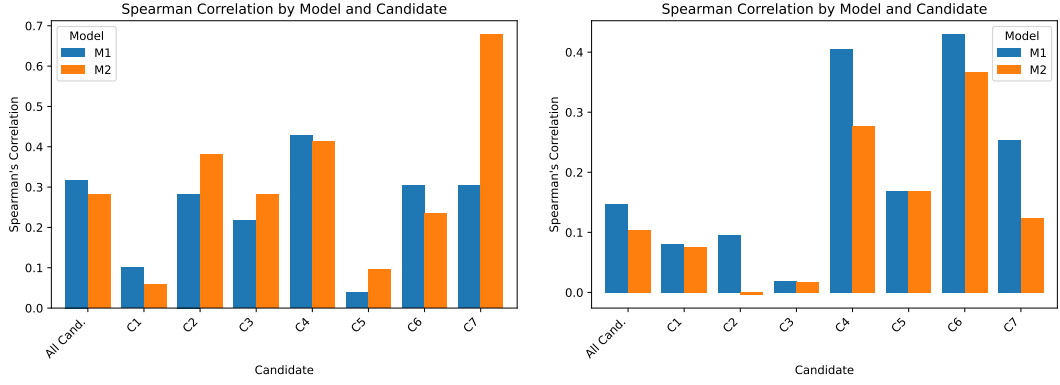


Figure 3: Spearman's correlation between human labeled and in-depth model matches for assignments in Gothenburg (left) and all over Sweden (right).

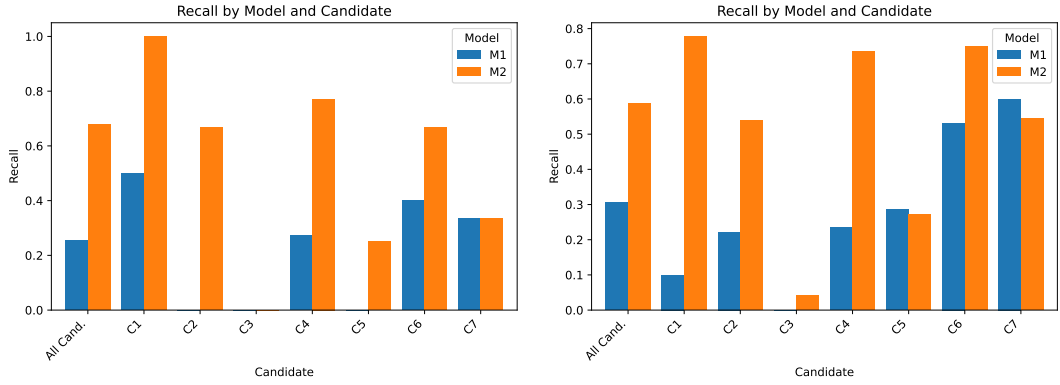


Figure 4: Recall of binary classification using human labeled and in-depth model matches for assignments in Gothenburg (left) and all over Sweden (right).

many candidates consider location as a strong component, but this is not evaluated by the models.

Similarly, Figure 4 presents the recall of the binary classification. We consider positive when the Score is larger than 60. A true positive (TP) is therefore when the human labeled Good and the in-depth model score return larger than 60, while a false negative is when the human label a Good but the model returns below 60. We find recall

$$R = \frac{TP}{TP + FN} \quad (5)$$

to be the most important metric of the confusion matrix, since we want to be sure that for every Good assignment according to a human, that will also be considered good by the LLM model.

The results indicate moderate positive correlations between LLM-based scores and human judgments, especially when evaluations are constrained to relevant geographic or domain contexts. When including all client requests, including those outside the candidate's preferred

locations or industries, the correlation drops, reflecting the models' limitations in capturing implicit preferences not expressed in text.

5. Discussion

The experimental results presented in the previous section offer several important insights into the effectiveness and limitations of the Personas recommendation system.

First, while our results show that pre-selection models correlate to some extent with in-depth LLM scores, we did not observe a clear performance advantage for any specific pre-selection strategy. However, this does not necessarily imply that the models are equally effective at identifying the best matches overall. A more definitive evaluation would require running the in-depth LLM analyses exhaustively across all candidate–assignment pairs, which was beyond the scope of this study. Without such comprehensive scoring, it remains difficult to assess how well the pre-selection models truly prioritize the most suitable assignments from the entire pool.

Second, the qualitative and quantitative assessments of the LLM-based models show that large language models are capable of nuanced judgment in candidate-assignment matching. Their ability to consider contextual fit, infer latent skills, and synthesize complex job requirements makes them valuable tools for augmenting HR workflows. However, the reliance on textual descriptions means they are inherently limited by what is explicitly stated in the documents. This was particularly evident in the human comparison experiments, where factors such as geographic preference, salary expectations, or client-specific cultural fit played a key role in expert evaluations but were often absent from the CVs and client request descriptions.

These findings underscore an important trade-off: while LLMs bring rich semantic understanding, they cannot reason beyond the provided inputs. This suggests two avenues for improving future iterations of the system. First, incorporating structured preference data (e.g., preferred locations, target roles, availability) directly into the matching process may help bridge the gap between textual analysis and real-world candidate intent. Second, training domain-specific LLMs or fine-tuning existing models on annotated HR datasets could help models better internalize implicit selection criteria. It is also worth noting that the system can process much richer, multi-page CVs, whereas the human evaluation relied on shorter CVs due to practical limitations—humans need to read through multiple CVs quickly, making longer documents impractical.

Another noteworthy point is the use of LLM scoring as a source of soft labels. This enables a continuous learning pipeline, where lightweight pre-selection models can be evaluated and improved without relying on scarce human-labeled data. Over time, this setup has the potential to create a virtuous cycle of feedback, where pre-selection models improve in alignment with human-like preferences—even in the absence of direct human supervision.

One key limitation of this study is the absence of a traditional pre-selection analysis, which could serve as a baseline for comparison. Traditional pre-selection methods often rely on rule-based systems, keyword matching, or straightforward criteria such as years of experience or educational background. While these approaches are widely used in HR systems, they tend to be less flexible and context-sensitive than modern embedding-based models. Without this

baseline, it is difficult to assess whether the embedding-based pre-selection models outperform or simply offer a more nuanced approach to candidate-job matching. Future work should consider incorporating a traditional pre-selection model to directly compare the performance of the Personas system against these more established methods, providing a clearer understanding of the strengths and weaknesses of embedding-based pre-selection in a recruitment context.

Finally, the human evaluation itself is a potential source of bias. Tags such as “good match” or “not relevant” are subject to individual consultant preferences, which may vary widely. These judgments also frequently consider external factors not modeled in this study, such as location of the assignment, team composition, communication style, or organizational fit. Future work should aim to incorporate multi-dimensional human assessments, possibly through structured annotation schemes or post-recommendation feedback loops.

6. Conclusions

This study presents an in-depth exploration of the Personas recommendation system, a hybrid pipeline that combines lightweight semantic filtering with powerful LLM-based analysis to support HR teams in the task of candidate-to-assignment matching. Through a combination of structured document processing, semantic embedding, and prompt-driven evaluation, the system is able to generate daily recommendations at scale while maintaining relevance and interpretability.

Our findings show that simple embedding-based models, provide good performance as pre-selection filters and correlate well with more computationally expensive LLM-based evaluations. These in-depth analyses offer nuanced and context-aware assessments of fit, acting not only as scoring mechanisms but also as a source of soft supervision for continuous improvement of upstream components.

Importantly, we observed moderate alignment between LLM scores and human expert tags, especially in constrained settings. However, the divergence in broader contexts highlights the need to explicitly model candidate preferences and non-textual factors—such as geography, compensation expectations, and cultural fit—that are crucial in human decision-making.

This work contributes to the growing body of research on the application of large language models in HR and recommendation systems. It highlights both the promise and limitations of current AI technologies in replicating complex human judgment and suggests practical pathways for system refinement.

Future work will focus on expanding the candidate dataset, incorporating explicit preference modeling, expanding the knowledge base of each candidate, exploring ways to better access specific parts of the documents, and exploring fine-tuned LLMs trained on HR-specific tasks. We also plan to deepen our integration of feedback loops from real-world usage, enabling more adaptive and personalized recommendations over time.

Acknowledgments

This project is supported by Vinnova (T.A.R.G.E.T. (2024-00242)), Kunskapsstiftelsen (KKS) (SERT research profile (2018-01-22)), and our research partner Synteda.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] R. Alonso, D. Dessí, A. Meloni, D. Reforgiato Recupero, A novel approach for job matching and skill recommendation using transformers and the o*net database, *Big Data Research* 39 (2025) 100509. URL: <https://www.sciencedirect.com/science/article/pii/S2214579625000048>. doi:<https://doi.org/10.1016/j.bdr.2025.100509>.
- [2] H. Kavas, M. Serra-Vidal, L. Wanner, Using large language models and recruiter expertise for optimized multilingual job offer - applicant cv matching, in: *International Joint Conference on Artificial Intelligence*, 2024. URL: <https://api.semanticscholar.org/CorpusID:271494727>.
- [3] Z. Zheng, Z. Qiu, X. Hu, L. Wu, H. Zhu, H. Xiong, Generative job recommendations with large language model, *ArXiv abs/2307.02157* (2023). URL: <https://api.semanticscholar.org/CorpusID:259342592>.
- [4] L. Wu, Z. Qiu, Z. Zheng, H. Zhu, E. Chen, Exploring large language model for graph data understanding in online job recommendations, *ArXiv abs/2307.05722* (2023). URL: <https://api.semanticscholar.org/CorpusID:259836967>.
- [5] J. Kurek, T. Latkowski, M. Bukowski, B. Świdorski, M. Łępicki, G. Baranik, B. Nowak, R. Zakowicz, Łukasz Dobrakowski, Zero-shot recommendation ai models for efficient job-candidate matching in recruitment process, *Applied Sciences* (2024). URL: <https://api.semanticscholar.org/CorpusID:268564829>.
- [6] D. Sileo, W. Vossen, R. Raymaekers, Zero-shot recommendation as language modeling, in: *European Conference on Information Retrieval*, 2021. URL: <https://api.semanticscholar.org/CorpusID:244954768>.
- [7] P. Singla, V. Verma, A hybrid approach for job recommendation systems, 2024 15th *International Conference on Computing Communication and Networking Technologies (ICCCNT)* (2024) 1–5. URL: <https://api.semanticscholar.org/CorpusID:273840570>.
- [8] B. L. Prasad, K. Srividya, K. N. Kumar, L. K. Chandra, N. S. Dil, G. V. Krishna, An advanced real-time job recommendation system and resume analyser, 2023 *International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)* (2023) 1039–1045. URL: <https://api.semanticscholar.org/CorpusID:265935829>.
- [9] L. Wu, Z. Zheng, Z. Qiu, H. Wang, H. Gu, T. Shen, C. Qin, C. Zhu, H. Zhu, Q. Liu, H. Xiong, E. Chen, A survey on large language models for recommendation, *ArXiv abs/2305.19860* (2023). URL: <https://api.semanticscholar.org/CorpusID:258987581>.
- [10] Y. Hou, J. Zhang, Z. Lin, H. Lu, R. Xie, J. McAuley, W. X. Zhao, Large language models are zero-shot rankers for recommender systems, *ArXiv abs/2305.08845* (2023). URL: <https://api.semanticscholar.org/CorpusID:258686540>.

- [11] D. B. Franzosi, E. Alégroth, M. Isaac, Llm-based labelling of recorded automated gui-based test cases, in: Proc. ICST, IEEE, 2025, pp. 453–463.