Preprints.org

Article

Optimizing Curriculum Vitae Concordance: A Comparative Examination of Classical Machine Learning Algorithms and Large Language Model Architectures

Mohammed Maree [*] and Wala'a Shehada

Article

Optimizing Curriculum Vitae Concordance: A Comparative Examination of Classical Machine Learning Algorithms and Large Language Model Architectures

Mohammed Maree [1],* and Wala'a Shehada [2]

1 Department of Information Technology, Faculty of Information Technology, Arab American University, 13 Zababdeh, P.O Box 240, Jenin, Palestine; mohammed.maree@aaup.edu

2 Department of Natural, Engineering and Technology Sciences, Arab American University, Palestine; w.shehada1@student.aaup.edu

* Correspondence: mohammed.maree@aaup.edu; Tel.: +00970-4-2418888 ext. 1123

Abstract: Digital recruitment systems have revolutionized the hiring paradigm, imparting exceptional efficiencies and extending the reach for both employers and job seekers. This investigation scrutinizes the efficacy of classical machine learning methodologies alongside advanced large language models (LLMs) in aligning resumes with job categories. Traditional matching techniques, such as Logistic Regression, Decision Trees, Naïve Bayes, and Support Vector Machines, are constrained by the necessity of manual feature extraction, limited feature representation, and performance degradation, particularly as dataset size escalates, rendering them less suitable for large-scale applications. Conversely, LLMs like GPT-4, GPT-3, and LLAMA, adeptly process unstructured textual content, capturing nuanced language and context with greater precision. We evaluate these methodologies utilizing two datasets comprising resumes and job descriptions to ascertain accuracy, efficiency, and scalability. Our results reveal that while conventional models excel with structured data, LLMs significantly enhance the interpretation and matching of intricate textual information. This study highlights the transformative potential of LLMs in recruitment, offering insights into their application and future research avenues.

Keywords: digital recruitment systems; classical machine learning; large language models (LLMs); performance degradation; methodology comparison; recruitment transformation

1. Introduction

The proliferation of digital recruitment methodologies has catalyzed an increased demand for more efficacious automated solutions. By harnessing the capabilities of the internet and cutting-edge technologies, these systems have revolutionized the employment process, conferring numerous advantages over traditional recruitment practices. Historically, the recruitment landscape was dominated by newspaper advertisements and job offers, with the processing of employment applications being executed manually—a process characterized by extensive time consumption and substantial resource expenditure, alongside inherent geographical limitations. The advent of electronic recruitment systems has ameliorated these inefficiencies, rendering access to requisite competencies both efficient and effective [20]. The incorporation of artificial intelligence (AI), machine learning (ML), and natural language processing (NLP) technologies has further engendered transformative changes in online hiring systems. These sophisticated models possess the capability to perform comprehensive analyses of curricula vitae (CVs) and execute optimal candidate-job matching algorithms, augmented by predictive analytics. Consequently, the integration of these

technologies has substantially enhanced the efficacy and efficiency of the recruitment process, providing a robust framework for future advancements in the field [1].

Despite the recent improvements in online recruitment systems, they still encounter formidable challenges, particularly concerning data privacy and AI biases. The sensitive personal data collected must be securely stored to avert breaches, which can result in identity theft and significant reputational damage. Implementing robust cybersecurity measures and clear data retention policies is imperative. Moreover, cross-border data transfers necessitate compliance with diverse regulatory frameworks. AI algorithms employed for resume screening and candidate matching can perpetuate historical biases if trained on biased datasets, leading to inequitable hiring practices. The utilization of diverse and representative training data, meticulous algorithm design, and stringent human oversight are essential to mitigate these biases. Additionally, Blockchain technology can bolster data security and accuracy within recruitment processes.

The primary objective of this article is to furnish a comprehensive review of the historical evolution and current state of online recruitment platforms. This study examines the technological advancements driving innovation in this sector, including AI, machine learning, and natural language processing. Furthermore, the article aims to explore the various models and frameworks employed in online recruitment systems, with a particular emphasis on recent developments such as large language models. By synthesizing extant data and identifying gaps, this article seeks to provide valuable insights to developers, researchers, and professionals engaged in the development and implementation of future recruitment systems. Ultimately, this study aspires to guide future research and development endeavors by proposing strategies to address current challenges and capitalize on emerging trends, thereby enhancing the efficacy and equity of electronic recruitment methodologies.

The remainder of this manuscript is structured as follows: Section 2 presents a comprehensive literature review that examines recent advancements in AI and ML, with a focus on the influence of large language models (LLMs). It also outlines the methodology employed in this study. Section 3 presents the findings obtained from the experiments conducted during for evaluating the various ML and LLM models employed in this work. In Section 4, future directions are explored, including the integration of Blockchain for secure data management, algorithm optimization strategies, and the potential role of Virtual Reality in candidate assessment. Finally, Section 5 provides concluding remarks for this article.

2. Materials and Methods

The integration of AI, ML, and LLMs into online recruitment platforms has revolutionized the employment landscape. AI and ML algorithms enhance candidate screening and matching capabilities by analyzing extensive datasets to identify optimal candidates based on skill sets, professional experience, and cultural alignment [1]. LLMs such as GPT, LLAMA, and BERT contribute to streamlining recruitment workflows by automating communication, generating interview queries, and providing immediate feedback during candidate interactions [8]. Collaboratively, these technologies enhance operational efficiency, mitigate bias, and enhance overall candidate experience throughout the recruitment process.

2.1. Enhancements in Online Recruitment Systems through AI and ML

Recent research has extensively examined the integration of AI and machine learning in online recruitment platforms. Several models and methodologies have been examined in the literature, emphasizing AI and machine learning's potential to improve various parts of the recruiting procedure. For instance, in [1], the authors focus on the usage of machine learning and artificial intelligence to improve the staff selection process. The work uses latent semantic analysis along with bidirectional encoder representations from transformers (BERT) to detect and comprehend hidden patterns in textual resume data. Support vector machines (SVM) are then used to create and improve the screening model. The study shows that LSA and BERT can effectively retrieve essential themes, but SVM enhances the model's efficiency through cross-validations and variable selection procedures. This approach offers HR practitioners useful insights for building and enhancing

recruitment processes and offers more interpretable outcomes than current machine learning-based resume screening systems. In [2], the researchers investigate the creation and implementation of an interview bot intended to improve the recruitment process. The project utilizes NLP and AI technology to develop a system for conducting initial interviews with job prospects. The bot evaluates candidate responses using complex NLP algorithms to determine their eligibility for the role. This automated technique intends to improve the recruitment process, decrease human bias, and increase productivity by swiftly filtering out unsuitable prospects, freeing up human recruiters to focus on the most promising individuals. The research illustrates the potential of AI-driven technologies to change traditional human resource procedures, delivering a more objective and scalable solution for talent acquisition. The authors of [3] explore the advancements in recruitment technology for developing smart recruitment systems (SRS), which use deep learning and natural language processing (NLP) to classify and rank resumes. These systems standardize and parse resume information, transforming diverse forms into English and extracting relevant data. By including executive requirements in the scoring process, the algorithms ensure that resume rankings meet the unique objectives of recruiters. This strategy strives to supply high-quality applicants by evaluating resumes based on technical capabilities and other important information, increasing the hiring process's efficacy and efficiency.

Previous research has emphasized the challenges organizations face when manually sifting, analyzing, recruiting, and alerting job applicants. For example, the authors of [5] proposed a system to automate the recruitment procedure utilizing the waterfall model, which is simple and sequential in nature. This intelligent job application screening and recruitment system, developed with PHP and MySQL, makes use of artificial intelligence techniques to improve Human Resource Management. The approach makes the recruitment process faster, more efficient, and inexpensive. According to evaluations, this strategy improves the accuracy with which individuals are matched to appropriate occupations. In a similar line of research [6], the researchers proposed an analytics-based method to improve the competitiveness of the HR recruitment process. The authors used existing tools to extract features from job postings and match them with potential candidates' resumes in a database. The finest suitable candidates were chosen using similarity analysis, taking into account the desired traits and their resumes. In one case, a structure learned on the profiles of 1029 job candidates at an IT company reduced screening manual efforts by 80%. This decrease is predicted to save significantly on time and operational costs. Although the study was limited to the situation of a technology company in India, the suggested artificial intelligence-based approach can be used across many industries. The study in [11] provides a complete review of current e-recruitment systems, classifying them based on a variety of evaluation criteria. The suggested approach in this work differs from typical systems in that it uses several semantic resources, such as WordNet and YAGO3, along with NLP, extracting features, and skill-relatedness algorithms, were used to discover the latent semantic dimensions in resumes and job postings. Unlike prior algorithms that consider the entire text of documents, this approach matches specific sections of applications to relevant parts of job postings. To address missing background knowledge, the HS dataset is used to augment job postings with semantically relevant ideas. Initial studies with a real-world dataset yielded high precision results, demonstrating the approaches' usefulness in giving relevance scores to candidate resumes and job offers. Also, in [18], the research presents an automated online recruitment method that combines numerous semantic resources and statistical concept-relatedness measurements. The system starts by using NLP techniques to discover and extract applicant concepts from job postings and resumes. The retrieved concepts are then refined using statistical concept-relatedness m9easures. As a result, various semantic resources are used to determine the semantic elements of resumes and job postings. To compensate for the restricted domain coverage of these semantic resources, the HS dataset is used to supplement job postings with new semantically relevant ideas. Initial testing with different resumes and job postings showed encouraging precision results, verifying the effectiveness of the suggested approach.

In [7], a decision-making model is offered as a critical requirement for the use of AI technology in corporate employment interviews. The Analytic Hierarchy Process (AHP) approach was used to

create the model, which included eighteen primary criteria defined by prior research using the Unification Model of Usage and Adoption of Technology. The factors were grouped into four major categories: expectations of performance, expectations of effort, societal impact, and facilitating environment. A survey of 40 professionals who were either users or producers of AI job interview solutions discovered the simplicity of use, job fitness, perceived utility, and perceived consistency were the most important criteria influencing the adoption of an AI-based job interview system. The results indicate that greater attention should be focused on integrating AI job interview systems into an organization's internal recruitment procedures, rather than focusing on external surroundings or situations.

In the current job market, organizations must hire and keep people who are the best match for the position. Research shows that employees who feel their jobs are important and enjoyable are more effective and less inclined to quit. Throughout the recruitment procedure in order to select the most qualified candidates. However, due to the large number of candidates, rigorous screening and interviews with each candidate are frequently impractical. To address this, researchers created systems such as JobFit, which uses recommender systems, machine learning algorithms, and historical data to forecast the best candidates for a job [4]. JobFit examines job requirements and profiles of applicants to generate a JobFit score that rates candidate based on their compatibility. This approach assists HR professionals by selecting a small group of top candidates for additional screening and interviews, making sure the best applicants are not ignored

2.2. Enhancements in Online Recruitment Systems through LLMs

Previous research has shown how Large Language Models (LLMs) have transformed the processing of natural language jobs across numerous domains due to their exceptional capabilities. However, the possibility of graph semantic processing in employment suggestions has not been thoroughly examined. [8] This work aims to demonstrate the potential of large language models to understand behavioral graphs and utilize their knowledge to improve recommendations for online recruitment, specifically advertising of (OOD) applications. They give a novel technique for studying behavioral graphs and detecting underlying patterns and correlations by utilizing the extensive contextual knowledge and semantic descriptions supplied by large language models. They suggest a meta-path prompting constructor to help LLM recommenders grasp the syntax for behavioral graphs for the initial time, as well as a path supplementation component to reduce prompt bias resulting from path-based sequencing input. This design allows for individualized and accurate employment recommendations for specific individuals. Their method is evaluated on big, actual-world data sets, demonstrating its capability to increase the importance and value of proposed outcomes. This study not only demonstrates the hidden strength of large language models but also provides useful insights for constructing advanced recommendation systems for the field of employment. The findings improve the area of processing natural languages while having clear implications for improving job search processes. With the growing popularity of large language models (LLMs), utilizing their extensive knowledge and sophisticated reasoning capabilities appears to be a potential strategy for improving resume completeness for more precise job suggestions. However, directly employing LLMs can lead to concerns like fabricated generation and the few-shot issue, which can reduce resume quality. To overcome these problems, [9] this work provides a revolutionary LLM-based GANs Interactive Recommendation (LGIR) technique. They improve resume completeness by extracting explicit user features (such as skills and interests) from self-descriptions and inferring implicit qualities from behaviors. To address the few-shot problem, which occurs when limited interaction records prevent high-quality resume production, they used Generative Adversarial Networks (GANs) to align low-quality resumes with high-quality created equivalents. Extensive trials using three large real-world recruiting datasets demonstrate the effectiveness of their proposed strategy. Current job recommendation systems primarily rely on shared filtering or person-job matching algorithms. Such models frequently function as "black-box" structures, without the transparency required to deliver clear information to job seekers. Furthermore, traditional matching-based systems can only rank and retrieve current vacancies from a database, limiting their utility as

comprehensive career advisors. To address these limitations, the authors of this study [10] introduce GIRL (Generative Job Recommendation Based on Large Language Models), a new approach that leverages recent breakthroughs in LLM. GIRL employs the Supervising Fine-Tuning (SFT) approach to train an LLM-based generation to generate appropriate Job Descriptions (JDs) depending on the job seeker's curriculum vitae. To improve the generator's performance, they train and fine-tune a model that assesses the match between CVs and JDs utilizing a Proximal Policies Optimizing (PPO)-based reinforced learning (RL) technique. This connection with recruiter feedback ensures that the resulting outputs better reflect employer preferences. GIRL is a job seeker-centric generative algorithm that provides employment ideas without the necessity for a pre-defined candidate list, hence boosting the capabilities of existing job recommendation models via produced content. Extensive experiments on an extensive real-world dataset confirm their technique's great success, confirming GIRL as a paradigm-shifting solution in job system recommendation that promotes a more personal and detailed job-seeking experience.

Table 1. A summary of the reviewed systems.

Reference	Methodology	Summary
[1]	Latent Semantic Analysis (LSA) and BERT with Support Vector Machines (SVM)	Improves resume screening by detecting patterns in textual data and enhancing model efficiency through cross-validation
[2]	Natural Language Processing (NLP) and AI	Development of an interview bot that evaluates candidate responses to improve recruitment efficiency and reduce bias.
[3]	Deep Learning and NLP (SRS)	Smart Recruitment Systems (SRS) classify and rank resumes by standardizing and parsing information, improving candidate selection.
[4]	(JobFit) Recommender Systems, Machine Learning, and Historical Data	JobFit predicts the best candidates for a job by generating a compatibility score, aiding HR professionals in candidate selection.
[5]	Waterfall Model, PHP, and MySQL	An automated recruitment system that uses AI to improve efficiency and accuracy in matching candidates to jobs.
[6]	Analytics-Based Recruitment	An analytics-based method to match job postings with resumes, significantly reducing manual screening efforts.
[7]	Analytic Hierarchy Process (AHP)	Decision-making model for AI recruitment systems focusing on criteria like ease of use, job fit, and perceived utility.
[8]	Behavioral Graph Analysis (BGA) with LLMs	Enhances job recommendations by understanding and utilizing behavioral graphs to improve recruitment outcomes.
[9]	LLM-Based GANs Interactive Recommendation (LGIR)	Improves resume completeness by extracting user features and using GANs to align low-quality resumes with high-quality generated equivalents
[10]	GIRL Generative Job Recommendation with LLMs and Reinforced Learning	GIRL generates job descriptions based on CVs, providing personalized job suggestions and enhancing the job-seeking experience.
[11]	Semantic Resources (WordNet and YAGO3) and NLP	E-recruitment system that match resumes with job postings by extracting features and using skill-relatedness algorithms to improve candidates' relevance.

In our proposed research work, we use a dataset of resumes from various industries to extract features and train multiple machine learning models (Naive Bayes, SVM, Logistic Regression, Decision Tree, XGBoost), and large language models (ChatGPT, LLaMA) for job classification. The performance of the models in categorizing resumes was tested using accuracy measures. Figure 1 depicts the outline of the methodology, followed by a detailed explanation:
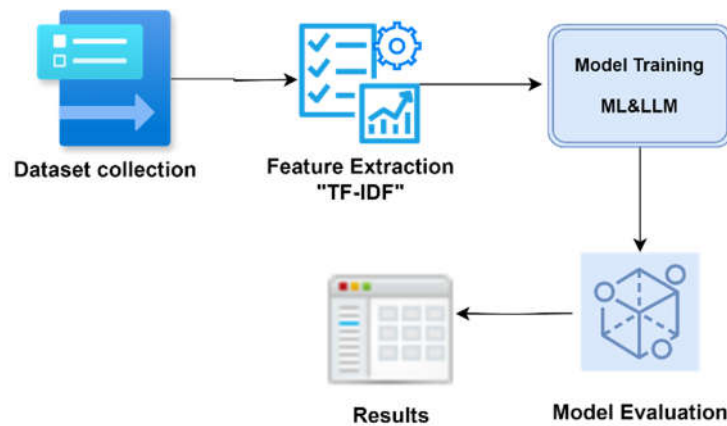
Figure 1. Online Recruitment Architecture for comparing the exploitation of ML against LLMs.

- **Dataset Description:**

This study's dataset consists of resumes and job categories, in multiple key industries like Finance, Technology, and Healthcare. Our dataset includes 962 resumes that were obtained from a publicly available resource and meant to give a wide representation of job categories to enable the training and evaluation of classification algorithms [23].

- **TF-IDF Vectorization:**

Term Frequency-Inverse Document Frequency (TF-IDF) was used to translate textual information contained in resumes into numerical features [24]. This model emphasizes the value of a word in a document to the total corpus. The Scikit-Learn library's TF-IDF vectorizer was used with parameters set to eliminate English stop words and a maximum document frequency threshold of 85% to filter out extremely common phrases. As a result, each resume was converted into a feature vector that may be used as input by machine learning models.

- **Machine Learning Models:**

Multiple machine learning models were trained using the TF-IDF features extracted from the resumes. The models included in this experiment were:

Multinomial Naïve Bayes:

A probabilistic framework using Bayes' theorem and a multinomial distribution. It works particularly well for text categorization tasks. The model computes the probability of each job category using the features retrieved from the resumes and assigns the category with the highest probability [24].

Support Vector Machine (SVM):

A classifier that locates the optimal hyperplane in a high-dimensional space that most effectively separates the classes. A linear kernel was utilized in this study to efficiently handle text data [25].

Logistic Regression:

A linear model that has been adapted to multi-class classification using the one-vs-rest method. The model calculates the probabilities of each job category and chooses the one with the greatest probability [26].

Decision Tree:

A non-linear model that divides data into subsets according to feature values. The tree is built by recursively separating the data at the feature that provides the most information gain [26].

XGBoost:

A gradient-boosted decision tree implementation that has been tuned for performance and speed. XGBoost develops an ensemble of trees sequentially, with each tree correcting the mistakes of the preceding ones.

- **Large Language Models:**

A large language model (LLM) is an artificial intelligence system that understands, generates, and manipulates human language with high efficiency. These models are trained on massive volumes of text data and employ sophisticated algorithms to understand language patterns, context, and nuances. The models included in this experiment were ChatGPT [28] and LLaMA [29].

ChatGPT:

Powered by OpenAI's GPT-3.5-turbo, this model was used to map resumes to their corresponding job categories. The resumes were submitted as input to the model, along with a system prompt that presented a list of potential job categories. The program identified the most relevant job category for each resume.

Meta's LLaMA:

The LLaMA model was also applied for resume classification. Similar to ChatGPT, the model accepted resumes as input, along with a system prompt showing job categories. The model's responses were processed to retrieve the expected job categories.

- **Evaluation Metrics**:

To assess the models' performance, we utilized the Accuracy metric, which is the ratio of accurately predicted job categories to the total number of predictions. It's computed as:

$$Accuracy = \frac{TP+TN}{TP+TN+TN+FN} \tag{1}$$

3. Results

In this experiment, we used a dataset of resumes and their corresponding job categories, which we divided into training and testing sets to simplify model training and evaluation. The dataset was divided so that 80% of the data was allocated to the training set and 20% for the testing. To ensure that the data split could be reproduced, a random seed of 42 was employed. So, our code reads the input dataset, randomly selects 80% of the records for the training set and allocats the remaining 20% to the testing set. The split data is recorded in two different .csv files named "_dataset.csv" and "_predict.csv".

To better support our results, we created a new dataset (named UPDATED DATASET) from the original dataset with a slight change in semantics and reduced the size of the resumes to compare the results between both machine learning models and large language models.

3.1. Category Mapping:

When applying large language models, we noticed that the models can answer the job classification using semantically-related terms that carry the same meaning. For example, the "data science" category was classified according to large language models into "Data Scientist". To obtain more precise results, the code refers to a JSON file named "data_map. json". This file contains concept mappings from the raw job title labels contained in the "Category" column to a more standardized semantically-overt form. This mapping can assist in enhancing the accuracy of machine learning models by decreasing ambiguity and inconsistency in the labels. Figure 2 illustrates an example of how the concept mapping could be organized in the "data_map. json" file.

```json
{
    "Data Science": [
        "Data Scientist",
        "Data Science Consultant"
    ],
    "HR": [
        "Human Resources Manager",
        "HR Manager"
```

Figure 2. Concept mapping captured in the "data_map. json" file.

3.2. API Integration:

To use large language models (LLMs) such as ChatGPT and LLaMA, the code communicates with the application programming interfaces (APIs) of ChatGPT and LLaMA to incorporate them into the job title prediction process. These APIs serve as gates to the LLMs' functionality. The code first generates a message, or prompt, instructing the LLM to behave as a recruitment expert.

For the used ChatGPT OpenAI's GPT-3.5-turbo model, the request includes the resume's content and a system prompt to select the most suitable job title from a predefined list, and the response was parsed to extract the job title suggested by ChatGPT.

For Meta's LLaMA model, similar to ChatGPT, the request comprised the resume text and a system prompt. The answer was retrieved in two steps: first, fetching the prediction URL, and then retrieving the final job title from the prediction URL.

Once the prompt is created, the code uses libraries such as requests to send it to the ChatGPT and LLaMA APIs. Finally, the code extracts the anticipated job title from each LLM's response, using their insights alongside the predictions from the machine learning models.

Machine learning models were employed to predict job categories for resumes in the prediction data. For ChatGPT and LLaMA, the functions getGPT and getLAMA were used to get predictions for the resumes. The results were saved into a CSV file. Overall, the evaluation and comparative analysis process was critical for determining the effectiveness of the various methodologies utilized for resume-to-job classification tasks.

3.2. Comparative Analysis:

In this section, we compare the performance of several models on the original and updated datasets. The models tested were GPT, LLAMA, SVM, Multinomial Naive Bayes, Logistic Regression, Decision Tree, and XGBoost. The datasets were divided into 80% for training and 20% for testing. See Table 2.

Table 2. Accuracy performance of machine learning algorithms and large language models.

Model	Original Dataset	Updated Dataset
GPT	90.6%	84.9%
LLAMA	72.5%	81.3%
SVM	100%	100%
Multinomial Naive Bayes	91.7%	83.4%
Logistic Regression	97.9%	97.9%
Decision Tree	98.9%	91.7%
XGBoost	99.0%	96.8%

As demonstrated in Table 2, the findings reveal that changes in data can have an impact on model accuracy, with various models performing differently on both the original and updated datasets. GPT's accuracy dropped from 0. 906 to 0.849, demonstrating susceptibility to dataset changes, whereas LLAMA's accuracy increased substantially from 0.725 to 0.813, indicating improved feature distinctions in the new dataset. SVM achieved perfect accuracy (100%) on both datasets, demonstrating robustness. Multinomial Naive Bayes decreased from 0.917 to 0.834, probably due to changes in word distributions. Logistic Regression stayed at 0.979, indicating stability. The accuracy of the Decision Tree reduced from 0.989 to 0.917, indicating possible overfitting on the original dataset. XGBoost's accuracy dropped from 0.99 to 0.968, yet it still performed well despite being slightly sensitive to data changes. This emphasizes the need to assess

model resilience and flexibility to ensure consistent performance across a variety of data circumstances.

4. Discussion

Large language models are often trained on vast amounts of text data and are capable of generating human-like text. However, there can be a mismatch between the general capabilities of these models and the specific requirements of tasks or applications. For example, while LLMs excel in generating coherent text, they may struggle with tasks requiring deep domain knowledge or specific contextual understanding. In the context of our work, domain knowledge about occupational categories was not precisely captured by the exploited LLMs. For instance, for the a resume of "HR Manager", the two following result ("Human Resources Coordinator") was retrieved when consulting ChatGPT about the most appropriate job title for the given resume. As such, mismatches and ambiguity can undermine the accuracy and reliability of LLM outputs. In applications requiring precise information or sensitive content (e.g., medical advice, legal documents), errors or misinterpretations due to mismatched capabilities or ambiguous outputs can have serious consequences.

Figure 3. Domain concept mismatch between the actual job title and the LLM's suggestion.

In subsequent sections, we aim to delineate significant avenues pertinent to resume screening and classification. These directions draw inspiration from recent advancements in Blockchain and Virtual Reality (VR) technologies, which, to our understanding, have yet to be explored within the scope of this research endeavor.

4.1. Future Direction in Online Recruitment Systems:

Integrating blockchain technology in online recruiting platforms improves data security by keeping information about candidates in a decentralized and secure manner. This ensures strong protection against unauthorized access and fraud. Furthermore, blockchain simplifies the verification of credentials, making background checks more efficient and reliable.

4.2. The Role of Virtual Reality in Candidate Assessment:

Virtual Reality (VR) is changing candidate assessments by giving immersive job simulations that enable realistic and objective evaluations of candidates' abilities and performance, removing biases and providing a more accurate assessment than traditional approaches.

5. Conclusions

The technological advancements have significantly transformed the landscape of online recruitment, enhancing efficiency in hiring processes and expanding opportunities for both employers and job seekers. This study undertook a comparative analysis between traditional

machine learning models and large language models (LLMs) in the context of resume-to-job category matching across two distinct datasets. Traditional models such as Logistic Regression, Decision Trees, Naive Bayes, and SVM demonstrated robust performance when applied to structured data. Conversely, LLMs such as GPT-3.5-turbo and LLAMA exhibited exceptional proficiency in interpreting and matching complex, unstructured textual information. Our findings underscore the substantial potential of LLMs in revolutionizing recruitment practices by effectively capturing nuanced language nuances and contexts, thereby enhancing the accuracy and scalability of online recruitment systems.

However, we also identified a critical issue that warrants further attention in this domain—specifically, the challenge of concept mismatch and ambiguity. This issue becomes particularly pronounced when LLMs are tasked with comprehending domain-specific knowledge or assuming an expert role within specific domains, such as occupational categories in our current study. Addressing this challenge necessitates additional research efforts aimed at aligning LLM outputs with expert domain knowledge encoded in ontological repositories. Such alignment would enable LLMs to more accurately discern semantic orientations and domain concepts, thereby yielding more precise and contextually relevant outcomes.

In summary, while LLMs offer significant advancements in recruitment technology, mitigating the effects of concept mismatch and ambiguity through enhanced semantic alignment remains a critical area for further investigation and development.

References

1. Tian, X., Pavur, R., Han, H., & Zhang, L. (2023). A machine learning-based human resources recruitment system for business process management: using LSA, BERT and SVM. Business Process Management Journal, 29(1), 202-222.
2. Josu, A., Brinner, J., & Athira, B. Improving job recruitment with an interview bot: a study on NLP techniques and AI technologies. In Advances in Networks, Intelligence and Computing (pp. 579-586). CRC Press.
3. Ponnaboyina, R., Makala, R., & Venkateswara Reddy, E. (2022). Smart recruitment system using deep learning with natural language processing. In Intelligent Systems and Sustainable Computing: Proceedings of ICISSC 2021 (pp. 647-655). Singapore: Springer Nature Singapore
4. Appadoo, K., Soonnoo, M. B., & Mungloo-Dilmohamud, Z. (2020, December). Job recommendation system, machine learning, regression, classification, natural language processing. In 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) (pp. 1-6). IEEE.
5. Ele, B. I., Ele, A. A., & Agaba, F. (2024). INTELLIGENT-BASED JOB APPLICANTS' ASSESSMENT AND RECRUITMENT SYSTEM. Technology, 6(1), 25-46
6. Ujlayan, A., Bhattacharya, S. & Sonakshi A Machine Learning-Based AI Framework to Optimize the Recruitment Screening Process. JGBC 18 (Suppl 1), 38–53 (2023). https://doi.org/10.1007/s42943-023-00086-y
7. Lee, B. C., & Kim, B. Y. (2021). A DECISION-MAKING MODEL FOR ADOPTING AN AI-GENERATED RECRUITMENT INTERVIEW SYSTEM. Management (IJM), 12(4), 548-560.
8. Wu, L., Qiu, Z., Zheng, Z., Zhu, H., & Chen, E. (2024, March). Exploring large language model for graph data understanding in online job recommendations. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 38, No. 8, pp. 9178-9186).

9.   Du, Y., Luo, D., Yan, R., Wang, X., Liu, H., Zhu, H., ... & Zhang, J. (2024, March). Enhancing job recommendation through llm-based generative adversarial networks. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 38, No. 8, pp. 8363-8371).

10. Zheng, Z., Qiu, Z., Hu, X., Wu, L., Zhu, H., & Xiong, H. (2023). Generative job recommendations with large language model. arXiv preprint arXiv:2307.02157

11. Maree, M., Kmail, A. B., & Belkhatir, M. (2019). Analysis and shortcomings of e-recruitment systems: Towards a semantics-based approach addressing knowledge incompleteness and limited domain coverage. Journal of Information Science, 45(6), 713-735.

12. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/the-consumer-data-opportunity-and-the-privacy-i

13. https://www.emptor.io/blog/data-security-in-hiring-best-practices-for-protecting-sensitive-information/

14. GDPR, G. D. P. R. (2018). General data protection regulation. URL: https://gdpr-info. eu/[accessed 2020-11-21].

15. Zhang, G. (2023, September). User-Centric Conversational Recommendation: Adapting the Need of User with Large Language Models. In Proceedings of the 17th ACM Conference on Recommender Systems (pp. 1349-1354).

16. Tuffaha, M. (2023). The Impact of Artificial Intelligence Bias on Human Resource Management Functions: Systematic Literature Review and Future Research Directions. European Journal of Business and Innovation Research, 11(4), 35-58.

17. Gao, R., & Shah, C. (2020, September). Counteracting bias and increasing fairness in search and recommender systems. In Proceedings of the 14th ACM Conference on Recommender Systems (pp. 745-747).

18. Kmail AB, Maree M, Belkhatir M, Alhashmi SM. An Automatic Online Recruitment System based on Multiple Semantic Resources and Concept-relatedness measures. In Proceedings of the 23rd IEEE International Conference on Tools with Artificial Intelligence (ICTAI'15). 2015.

19. Aseel B. Kmail, Mohammed Maree, Belkhatir M. MatchingSem: Online Recruitment System based on Multiple Semantic Resources. 12th IEEE International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'15)2015.

20. Arman, Md. (2023). The Advantages of Online Recruitment and Selection: A Systematic Review of Cost and Time Efficiency. Business Management and Strategy. 14. 220-240. 10.5296/bms.v14i1.21479.

21. N. S. Dhanala and R. D., "Implementation and Testing of a Blockchain based Recruitment Management System," 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020, pp. 583-588, doi: 10.1109/ICCES48766.2020.9138093.

22. https://www.finextra.com/blogposting/25678/smart-contracts-and-automation-unlocking-seamless-transactions

23. https://www.kaggle.com/datasets/jillanisofttech/updated-resume-dataset

24. Wendland, A., Zenere, M., & Niemann, J. (2021). Introduction to text classification: impact of stemming and comparing TF-IDF and count vectorization as feature extraction technique. In Systems, Software and Services Process Improvement: 28th European Conference, EuroSPI 2021, Krems, Austria, September 1–3, 2021, Proceedings 28 (pp. 289-300). Springer International Publishing.

25. Xu, S., Li, Y., & Wang, Z. (2017). Bayesian multinomial Naïve Bayes classifier to text classification. In Advanced Multimedia and Ubiquitous Engineering: MUE/FutureTech 2017 11 (pp. 347-352). Springer Singapore.

26. Pisner, Derek A., and David M. Schnyer. "Support vector machine." Machine learning. Academic Press, 2020. 101-121.

27. Alnuaimi, Amer FAH, and Tasnim HK Albaldawi. "An overview of machine learning classification techniques." BIO Web of Conferences. Vol. 97. EDP Sciences, 2024

28. Qi, Z. (2020, June). The text classification of theft crime based on TF-IDF and XGBoost model. In 2020 IEEE International conference on artificial intelligence and computer applications (ICAICA) (pp. 1241-1246). IEEE

29. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30

30. https://llama.meta.com/docs/model-cards-and-prompt-formats/meta-code-llama-70b.