Research on Improving Recruitment Using Natural Language Processing and AI Technology

Lankun Chen

Amsterdam Business School, University of Amsterdam, Amsterdam, North-Holland, 1018 TV, Netherlands

lankunchen@outlook.com

Keywords: Natural Language Processing; AI Technology; Human Resource Management

Abstract: This thesis examines the use of large language models to enhance recruitment by automatically extracting and analyzing regional recruitment requirements from the job descriptions. This study addresses the key weaknesses in current assessment tools, such as ambiguity in job descriptions, excessive length, and bias. This research uses Natural Language Processing (NLP), specifically regular expressions, and the Bidirectional Encoder Representations from Transformers (BERT) algorithm to extract relevant elements from job descriptions. It employs a Topic Modeling approach to identify differences in regional recruitment preferences. The research demonstrates that automated NLP methods can be used to generate customized questionnaires, reducing ambiguity in job descriptions and unconscious bias while speeding up the hiring process. The research reveals significant differences in the recruitment requirements of different regions and platforms. It provides insights into developing more efficient and equitable strategies for recruitment. The thesis contributes to the field by providing a practical solution for improving the recruitment process using advanced AI and NLP technology while highlighting the need to be transparent, customized, and cost-efficient. Future research directions will include expanding data sources and addressing biases. They will also enhance model interpretability and assess the long-term effects of AI on recruiting.

1. Introduction

This paper will focus on using NLP technology to provide applied solutions to some of the shortcomings of current assessment tools, and conduct in-depth research on whether different recruitment sources will lead to different quality recruitment preferences.

Edwards denotes that Natural language descriptions of job requirements often suffer from ambiguity, excessive length, and complexity, leading to fewer candidate responses [1]. Additionally, the use of natural language in job descriptions by HR departments may inadvertently create biases that prevent qualified candidates from applying [2]. However, people are all working with nature language disregarding it is impossible. Additionally, Manual questionnaire compilation not only wastes time but also fails to meet customized needs. Therefore, it is necessary to find a cost-efficient way to identify hidden patterns in natural language job descriptions, such as skill requirements or personality traits, using AI technology [3]. These patterns can be used to generate questionnaires for job candidates, reducing ambiguity and unconscious bias in job descriptions while expediting the hiring process.

The thesis is structured as follows: We delve into the notion of e-HRM, establishing the foundation for comprehending its significance. Next, we examine present techniques in assessment development, emphasizing significant deficiencies. We will explore the extensive literature on identifying constructs in the field of Natural Language Processing (NLP), leading to a basic conceptual framework. The next section outlines our research methodology, which utilizes regular expressions and the BERT algorithm to extract elements from job descriptions while applying a topic modeling approach to identify regional differences in recruitment requirements. Next, we provide our findings from the data analysis, using a comparative approach to examine the current literature and gain valuable insights. In conclusion, we analyze the significance of our discoveries for theoretical understanding, practical applications, limitations, and future investigations, therefore opening up

  DOI: 10.25236/icemeet.2024.051

opportunities for further discovery in this area.

2. Literature review

This section reviews current trends in recruitment operations while outlining their theoretical basis and feasibility. In particular, artificial intelligence (AI) and large language models could play an instrumental role in recruitment processes. Subsequently, we examine the identified shortcomings and limitations in the current recruiting studies, underscoring the significance of extracting ideas from job descriptions. Next, we analyze the advancements in computer science regarding the extraction of structures and highlight the limitations of manual labeling. We also explore the feasibility of employing the BERT technique for semi-supervised learning to extract components from recruitment requirements.

2.1 Digital recruitment 3.0 and e-HRM

2.1.1 From HRM to e-HRM

Human Resource Management (HRM) refers to the systematic approach of enhancing the intellectual workforce to assist firms in attaining their goals, missions, and diverse objectives [4]. Currently, HR is perceived as a set of specialized skills, with the goal of HR optimization centered on using new technology to improve efficiency, provide investment returns, and foster organizational growth [5].

The widespread integration of information technology (IT) in HRM operations has led to the emergence of a new HRM paradigm known as e-HRM. Bondarouk defined e-HRM as integrating HRM with information technologies [6]. It encompasses many procedures and contents that aim to provide value for employees and management within and between businesses. e-HRM includes operations, such as recruitment, training, and managing compensation and benefits [7]. This thesis largely focuses on the recruitment function within HR. Next, we examine the continuous development of recruitment alongside the progress of electronic Human Resource Management (e-HRM).

2.1.2 The Development of the recruitment process

Laumer and Eckhardt described recruitment as the process of locating and attracting new employees, whereas selection involves making decisions from a pool of candidates to find those who can bring value to the firm [8]. Recruitment is a key human resources strategy since it provides the initial chance to anticipate the future behavior of potential workers and plan accordingly.

Historically, organizations have relied on conventional methods such as newspaper advertisements and employee referrals to attract suitable candidates [9]. However, these traditional recruitment techniques are increasingly seen as inefficient, requiring considerable time and often failing to produce the best outcomes.

The age of internet hiring began in the mid to late 1990s with the process of converting employment and candidate records into digital format, significantly increasing the range and depth of available data [10]. HR departments could now disseminate detailed job descriptions to a large number of candidates without incurring the costs of printing and distributing newspapers. This digital method also reduced the time and effort required by candidates, eliminating the need to physically visit different locations or complete numerous forms.

Digital Recruitment 2.0 emerged ten years after the introduction of Digital Recruitment 1.0, consolidating job postings from several independent employment websites. This technological breakthrough enabled recruiters to identify outstanding candidates across multiple platforms, while job seekers enjoyed the convenience of not having to manually search each network for job openings.

By 2015, the Digital Recruitment 2.0 environment had reached its peak, paving the way for the digital recruitment 3.0 era. The current period is characterized by the rise of AI technology, including advanced AI software that can comprehend spoken language, assess emotions, identify photos, and make judgments using multiple criteria [11]. AI software utilizes sophisticated algorithms to streamline various recruitment processes, fundamentally revolutionizing how organizations identify and hire skilled individuals. This technology enhances the benefits and possibilities of HRM by minimizing and almost eliminating obstacles in job search and talent acquisition [12].

2.2 Current practices in AI-enabled recruitment

2.2.1 Benefits and risks of deploying AI-enabled recruitment tool

The advancement of e-recruitment has shifted talent evaluation methodologies from conventional human-based approaches to AI-powered systems. Recent studies emphasize the advantages of AI-driven evaluation technologies, such as time efficiency, heightened precision [13], reduced prejudice [14], improved competitiveness for skilled individuals, better adaptability [15], and the enhancement of HR experts' analytical abilities [16].

Nevertheless, recent research has also revealed that AI-powered recruitment methods have disadvantages. Artificial intelligence (AI) systems have the potential to unintentionally reinforce pre-existing biases that are present in the data used to train them, resulting in discriminatory results. If the historical data exhibits prejudiced recruitment practices, the AI system may reproduce similar trends, hence strengthening discrimination [17]. In addition, there are worries over the clarity and comprehensibility of AI choices in the recruiting process. Several AI algorithms function as "black boxes," posing challenges for HR professionals and applicants in comprehending decision-making procedures [18]. Furthermore, while AI can enhance efficiency, it may lead to a lack of personalization in candidate experiences, potentially impacting the organization's employer brand. Lastly, although AI-driven tools show promise in improving recruitment processes in the short term, their long-term impact on organizational culture and employee satisfaction remains unclear.

These gaps emphasize the necessity for more investigation to create AI recruiting tools that are more transparent, impartial, and favorable to candidates and comprehend their enduring impacts on businesses. The objective is to fill these knowledge deficiencies by utilizing Natural Language Processing (NLP) techniques and sophisticated AI models to improve recruiting procedures. More precisely, we employ regular expressions and the BERT algorithm to reliably and exhaustively extract elements from job descriptions. Through the use of a topic modeling methodology, our objective is to discern geographical disparities in recruiting prerequisites, therefore establishing a recruitment strategy that is both customized and equitable.

2.2.2 Evolution of screening tools and their advantages and disadvantages

Traditional Psychological Assessments: The first step in analyzing candidates, of course, is to use standard psychological tests. These include the Myers-Briggs Type Indicator (MBTI) and the Minnesota Multiphasic Personality Inventory (MMPI), which are often used to measure aspects of people's personality traits or psychological states. These tests are primarily intended to measure more complex information about a candidate's personality traits, which can be used in anticipation of job performance and fit with the organization. The techniques described have been adequately standardized and are known to predict aspects of individual job performance or behavior with reasonable accuracy in a variety of population [19], which generally has made them useful for many purposes that experience no requirements for generalizability beyond the parameters measured from these methods' use [20]. However, the administration and scoring of these tests remain a cumbersome process, along with social and cultural biases [21].

Questionnaire-Based Surveys Using Job Descriptions: Questionnaires and surveys specially designed for job descriptions: Many organizations use such assessments in addition to direct psychological assessments. The content of these surveys is derived from a job description and asks respondents to rate candidates on the specific skills, credentials, and personal attributes deemed necessary for the role [22]. This methodology also means candidate quality is assessed based on what matters for the job. In addition, these questionnaires can be administered and completed online, enabling the collection of large-scale data [23]. Such measures can be susceptible to impression management and are typically limited in their ability to capture complex, real-time interactions with contextual drivers of behavior [24-25]. Additionally, it would be inefficient and cost-consuming for HR professionals to compile specific questionnaires [26].

Situational Judgment Tests (SJTs): Situational judgment tests (SJTs) gained popularity in the 1980s and 1990s. These tests show candidates different situations related to employment and present a few possible options for how the candidate should deal with those situations. Therefore, SJTs, which are job specific, can accurately measure judgment as they apply to the specified job and are oriented concerning specific features of a particular role [27]. While they are less discriminatory toward minorities than cognitive ability tests [28], candidates may fake good responses rather than provide an insight into true behavior, and developing JSIs can be a costly process in terms of both time and resources spent on reliability validation.

Modern AI-Enabled Tools: In recent times, tools such as HireVue and Pymetrics have become popular. These two AI-enabled tools use artificial intelligence (AI) to analyze video interviews and neuroscience-based games for cognitive and emotional trait assessment, respectively. As a result of their ability to handle big data quickly, such AI-driven solutions reduce the time spent on screening. Moreover, the complicated datasets can be understood by them, thereby giving more profound insights into candidate capabilities. Also, it facilitates the standardization of evaluation criteria by eliminating some types of human bias [29]. Nonetheless, there is a danger that AI systems could perpetuate existing biases in the training data, leading to discriminatory outcomes [30]. Many "black box" algorithms used in AI make it impossible for users to know how decisions have been made. Additionally, a less personalized candidate experience may result from a lack of human interaction.

2.2.3 Research gap in assessment tool

There have been several developments, but there are still enormous gaps in current systems. It is pertinent to achieve a delicate balance between user friendliness, cost efficiency, automation, and data security, specifically. Specifically, it is crucial for designing effective screening systems that involve identifying elements from job descriptions. Job descriptions commonly contain detailed information about skills, qualifications, and attributes necessary for particular positions. Employing Natural Language Processing (NLP) techniques like the BERT algorithm may enhance the accuracy and coverage of this extraction method. To make sure that applicants match job requirements better, companies could adapt their selection procedures based on certain criteria that are integral to a specific job opening, thus allowing them to improve recruiting efficiency as well as equitable treatment for all the applicants in question by adopting recruiting strategies that are tailored towards addressing components critical to any given role within the organization.

Our Solution is to develop a kind of automated NLP methods be used to extract the constructs of a job description or a more general oral description of the talent given by hiring manager in order to generate a specific questionnaire for the job. This could largely cut down on labor costs while improving productivity. There are numerous benefits, including less risk of data leaking, customizable and time-saving, cost-saving and user-friendly.

2.2.4 Solution to this gap

Our Solution is to develop a kind of automated NLP methods be used to extract the constructs of a job description or a more general oral description of the talent given by hiring manager in order to generate a specific questionnaire for the job. This could largely cut down on labor costs while improving productivity. There are numerous benefits below.

Less risk of data leaking: Being a publicly available job description or the hiring manager's verbal description of the need means the privacy of candidates is not an issue here.

Customizable and time-saving: Traditional selection questionnaires need to be written by HR experts. When a company has many positions and inconsistent requirements for candidates, writing questionnaires will be a time-consuming and laborious task. When this process is automated, a customized questionnaire about this occupation can be generated quickly and efficiently, even without the need for specialized HR knowledge.

Cost-saving and user-friendly: Compared to the need of HR experts equipped with more advanced analytical knowledge in digital recruitment process, our solution does not require professional skills to generate questionnaires. It only requires a job description, whether written or oral, formal or informal, from which keywords can be identified to generate questionnaires. It is very important that by using some existing open-source large language models or deep learning models to fine-tune the domain required by the profession, an ideal recognition model for this field can be obtained, which is much more cost-effective than building a separate artificial intelligence system, as shown in table1.

Table 1. Summary of Assessment Tools for their Advantages and Disadvantages

Practices	Advantages	Disadvantages
Traditional Psychological Assessments (e.g., MBTI, MMPI)	• Standardization • Predictive validity	• Time-consuming • Potential bias • Lack of context
Questionnaire-Based Surveys Using Job Descriptions	• Job relevance • Ease of administration	• Self-report bias • Time-consuming
Situational Judgment Tests (SJTs)	• Job relevance • Reduced adverse impact	• Response bias • Development cost
Modern AI-Enabled Tools (e.g., HireVue, Pymetrics)	•Efficiency • Enhanced precision • Bias mitigation	• Bias reinforcement • Lack of transparency • Personalization issues

3. Diversity in Recruitment

As mentioned above, the advent of e-recruitment 3.0 has not only triggered the development of the assessment methods we discussed in section 3.2, but also brought about globalization of employers and employees.

3.1 Influence of Culture

Cultural values play a pivotal role in recruitment processes. According to Hofstede cultural dimension theory [31], countries with various cultural backgrounds place different emphasis on recruitment criteria; according to Birkelund et al.'s findings, cultures with high uncertainty avoidance like Japan and South Korea tend to have stringent qualifications requirements and rigorous recruitment procedures while cultures with lower uncertainty avoidance such as Australia are generally more flexible and emphasize practical skills over qualifications [32].

3.2 Benefits and Drawbacks of a Diverse Workplace

Cultural diversity can foster creativity and innovation, increase customer acquisition, and strengthen problem-solving abilities for companies [33]. However, cultural diversity also presents unique challenges such as negative stereotypes that inhibit multicultural team integration; miscommunication between cultures; conflicts across work styles.

3.3 Research Gap

Previous studies have hinted at differences in recruitment requirements across countries, so it is worth exploring whether there are differences in job requirements from different sources, with the advent of e-recruitment 3.0 bringing more recruitment sources. We will use Latent Dirichlet Allocation (LDA) algorithm in Topic Modeling Algorithm to see whether there is any preference among data sets from China, United Kingdom, United States and a recruitment software - LinkedIn.

3.4 Construct Extraction Technique in Natural Language

3.4.1 Construct Mining Pipeline

Herderich et al. proposed an innovative text mining methodology called Con- struct Mining Pipeline that can locate structures within natural language text. The nine-step procedure begins with data collection and ends with cluster naming. Each step in these nine processes includes Data Collection, Sentence Embedding, Evaluation, Dimension Reduction, Clustering, Robustness Checking, Verifying Clusters, Interpreting Results, and Naming. Psychological identification involves the recognition and understanding of psychological constructs or unfamiliar concepts. However, this research shows that current approaches rely on manual annotation of clustering outcomes for training datasets, limiting their overall use. Our goal in this research is therefore to create a system that removes manually describe constructs while filling any current gaps within this technology.

3.4.2 Topic Modeling

Topic modeling is a text mining technique for discovering latent topics within a corpus of documents using Latent Dirichlet Allocation (LDA). LDA represents documents as probabilistic distributions over multiple topics [34]. Recently, researchers have enhanced topic models by including word embeddings and neural networks into them - leading to improved accuracy and interpret ability for these models. Dynamic Topic Models (DTM) and Hierarchical Topic Models (HTM) have also been proposed, in order to account for temporal variations and complex data structures. Topic models find use in information retrieval, text classification and recommendation systems- making great strides forward in natural language processing. We plan to use topic modeling to analyze whether there are indeed differences in recruitment needs from different sources.

3.5 Keyword Extraction Algorithm in NLP

Keyword extraction is an integral component of natural language processing (NLP), consisting of extracting key words and phrases from text for applications like summarization, retrieval and content analysis. Over the years many algorithms and strategies have been devised in order to maximize keyword extraction's precision and effectiveness.

3.5.1 Early Approaches

Initial keyword extraction algorithms relied heavily on statistical techniques for extraction; one such text mining approach described by Dillon was called Term Frequency-Inverse Document Frequency (TF-IDF), which calculates word importance by comparing their frequency in documents against that in larger collections - thus revealing any potentially important keywords within specific contexts.

3.5.2 Statistical Methods

RAKE (Rapid Automatic Keyword Extraction), developed by Rose et al. and first widely adopted in 2010, has also proven highly popular [35]. RAKE utilizes co-occurrence and frequency patterns within text to detect keywords - making it highly effective at extracting them without needing access to a huge corpus of documents.

3.5.3 Linguistic Approaches

Linguistic techniques employ syntactic and semantic analysis to enhance keyword extraction. A research project conducted by (mihalcea-2004)) using TextRank - an unsupervised algorithm using graphs to rank text - brought a significant change to keyword extraction by taking advantage of text's structure. TextRank creates a word network dis- playing co-occurrence associations; then employs PageRank's scoring system to evaluate and prioritize keywords according to their scores.

3.5.4 Machine Learning Approaches

Machine learning has revolutionized keyword extraction. Successful keyword extraction problems have been achieved using supervised learning algorithms like Conditional Random Fields (CRFs) and Support Vector Machines (SVMs), both requiring datasets with annotations to train models which predict keyword relevance using variables like word location, frequency and part-of-speech tags.

3.5.5 Deep learning approaches

Recent advances in deep learning approaches have dramatically advanced keyword ex- traction skills. Neural network models that utilize word embeddings like Word2Vec and GloVe, combined with transformer like BERT have achieved outstanding levels of performance - particularly models based on BERT are capable of understanding contextual words found within documents, providing more precise, contextually relevant keyword extraction [36].

3.5.6 Evaluation and Challenges

Evaluating keyword extraction algorithms typically involves metrics such as accuracy, recall and F1-score to measure their performance. Unfortunately, keyword extraction models still face challenges such as managing multiple meanings within certain fields and annotated datasets for supervised techniques, and current studies are being done to ensure their applicability across various text genres and languages.

In conclusion, Natural Language Processing (NLP) keyword extraction technology has rapidly advanced over time, from basic statistical techniques to intricate deep learning models. Each strategy offers distinct advantages that may suit various applications or situations more successfully; in this thesis study we investigate extracting HR constructs specifically job requirements via extraction algorithms.

4. Research Questions

After analyzing the above literature review, this thesis mainly addresses the following two research questions:

-Identification and Extraction of Constructs: How effectively can constructs such as 'Approachability,' 'Empathy,' and 'Kindness' be identified and extracted from natural language job descriptions using extraction algorithms?

-Qualitative Differences in Constructs Across Sources: Do the psychological constructs identified from job descriptions differ qualitatively among the three sources? If differences exist, what are the nature and extent of these differences?

5. Methodology

The entire project will be based on CRISP-DM because it is a complete scientific data mining process. Based on CRISP-DM, model prototypes can be improved and fine- tuned in a timely and sensitive manner based on real HR feedback meetings organized subsequently. The benefit of choosing the CRISP-DM method is that its standardized and systematic process helps ensure that each stage of the data science project can be evaluated and optimized through quantitative methods, thereby improving the scientific nature of the project [37].

5.1 Business Understanding

5.1.1 Motivation

Recruitment processes are essential to any organization's success; yet traditional approaches often rely on subjective assessments and intuition - often leading to inefficiency and bias. Due to an increase in unstructured candidate descriptions, more objective, scientific methods have become increasingly important for evaluating candidates for specific jobs; general questionnaire design techniques do not accommodate customization or scientific rigor. Digital recruitment has become more prominent due to globalization advances; existing assessment tools contain flaws; while cultural and job resources diversity also drive an increase in recruitment needs as discussed in literature review [38].

5.1.2 Project Purposes

This project will use NLP algorithms to help HR professionals and hiring mangers create scientifically sound questions for each position. We hope to improve the recruitment process by automating the structure recognition, filtering and weighing processes, as well as the questionnaire creation using AI. This will result in better candidate selection and better organizational outcomes, with greater efficiency, savings, user-friendliness, and cost savings compared to the current tools discussed at Section 3.2. Topic modeling will determine if the recruitment needs of different sources are significantly different. The topic modeling will give recommendations on how recruitment can be differentiated and bias eliminated in the recruitment sources.

5.2 Data Understanding

The dataset for this project was provided by our supervisor with 75 entities following features:

-Source (BNLD for English respondents from UK and US, CNLD for Chinese respondents and JD for job descriptions)

-Native Language Description for Step 1 Constructs Identification is an initial de- scription of ideal candidate.

Example: Our project team is looking for an energetic team player with superior project and program management abilities who can navigate ambiguity while aligning on project goals, as well as deploy globally-relevant solutions in a fast-paced environment.

-Constructs that would be the input for Step 2 Constructs Weighting and Filtering are the key characteristics identified based on descriptions of expected candidates for specific job [39].

Example: Creativity, Teamwork, Project Management Skills, Adaptability, Ambiguity Tolerance, Goal Alignment, Self-Management, Global Perspective

-Constructs and Weights that would be the input for Step 3 Questionnaire Development represent the key characteristics from Step 2 but with assigned weights that express the importance of characteristics for each case.

Example: Creativity - 15%, Teamwork- 15%, Project, Management Skills - 15%, Adapt- ability - 10%, Ambiguity Tolerance - 10%, Goal Alignment - 10%, Self-Management - 15%, Global Perspective - 10%.

The work in this thesis will use Native Language Description and its corresponding Constructs to build a NLP extraction algorithm because data-point is text. With using Source to group the data, the Latent Dirichlet Allocation (LDA) algorithm would be trained to study what the main keywords of job descriptions from different sources are, so as to determine whether they are different.

5.3 Data Preparation

To prepare the data, we tokenized Job description data with TweetTokenizer. We then filtered out words that were shorter than three letters and lemmatized all remaining words. The stop words list was extended with domain-specific terminology. The final corpus of words was transformed into a Bag of Words by converting a Gensim dictionary that was filtered for common and rare terms.

5.4 Modelling

5.4.1 BERT Model

Google introduced Bidirectional Encoder representations from Transformers in 2018 as a pre-trained, groundbreaking model. BERT is a bidirectional model that reads text in both directions, simultaneously considering the left and right contexts. The bidirectional approach allows BERT better understand language and context. This leads to improved performance in various NLP tasks, as shown in figure 1.

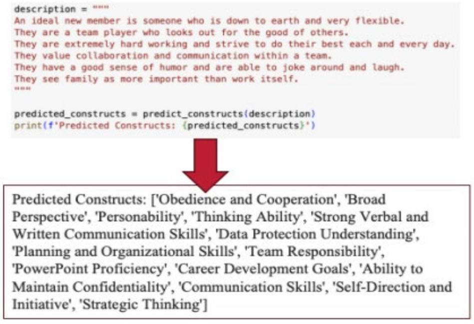

Figure 1. Practical use of BERT Model in Constructs Extraction

BERT addressed this limitation by taking advantage of the Transformer architecture, which provides for bidirectional context understanding. A key innovation was using bidirectional training of Transformer encoders in order to deepen word context both ways - an approach which marked a marked improvement over previous models for many NLP tasks.

-Mathematical Formulas in BERT

$$\text{Self-Attention Mechanism: } \mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

In this formula: - Q: Query matrix - K: Key matrix - V: Value matrix - dk: Dimension of the key vectors.

This self-attention mechanism allows the model to weigh the importance of different words in a sentence dynamically.

Multi-Head Attention:

$$\mathrm{MultiHead}(Q, K, V) = \mathrm{Concat}(\mathrm{head}_1, ..., \mathrm{head}_h)W^O \quad (2)$$

Each head in the multi-head attention mechanism independently computes attention, and the results are concatenated and linearly transformed.

Application in this dataset: Since the BERT Model is a pre-trained model, it only needs to be fine-tuned on our dataset to start working as a classifier. However, the work cannot be started directly. The job description in the dataset is not labeled. After manually labeling some datasets and fine-tuning BERT, it is found that it can only perform well in some general categories, but it is still far from being able to extract exactly the same dataset as in the dataset. The Figure 1 shows an example of the effect of the BERT algorithm on construct extraction after several training sessions [40].

5.4.2 Latent Dirichlet Allocation algorithm

Figure 2. Word Cloud in initial analysis

As shown in figure 2, this thesis implements LDA using Gensim. Gensim is a Python library that is used for document indexing, topic modeling and retrieving similarity with large corpora. It is presumed that there is a certain quantity of themes *k* inside the collection of papers. At first, we assign a certain subject to each word *W* in the collection of documents. The topic assignment is determined by θ_i, which follows a Dirichlet distribution Dir(α). Here, *i* ranges from 1 to *M*. This code snippet initializes a topic model. Next, we make the assumption that the assigned topic for a word *W* is incorrect, while assuming that all other topics are correct. This involves calculating the conditional probabilities p (topic *t* | document *d*), which represents the probability that the document *d* is as-signed to topic *t*, and *p* (word *w* | topic *t*), which represents the probability that topic *t* is assigned to word *w*. Next, we modify the topic of the document to be the subject with the greatest likelihood of being assigned to this document, which is calculated as the product of the probability of the topic given the document and the probability of the word given the topic (*p* (topic *t* | document *d*) · *p* (word *w* | topic *t*)).

When the number of topics is preset to 8, we get the word cloud diagram as shown in the figure 3. Each Topic is composed of keywords. Each keyword in a Topic has a weight. The greater the weight, the more representative the word is in this topic, and vice versa.

For example, in Topic 4, the weight of 'data' is the largest, so Topic 4 largely represents the topic of the data class. This thesis will introduce in detail the optimal parameter of LDA -the evaluation method of the number of topics, and use the trained optimal model to see if the three different sources have statistically different topic distributions [41-42]. Figure 3 shows the LDA model overview.

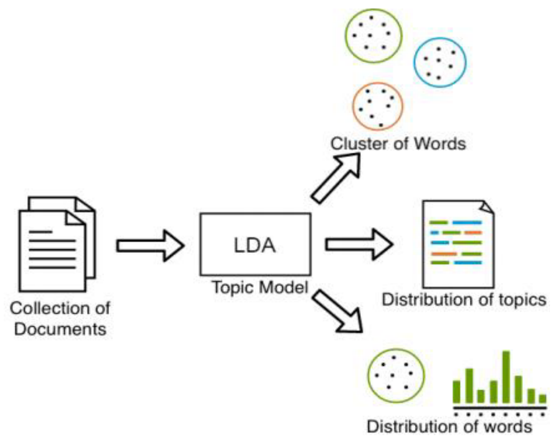

Figure 3. LDA model overview

5.5 Evaluation

5.5.1 Evaluation of Fine-tuned BERT Model

A classification report provides an overview of the performance of a model, including metrics like precision and recall rates. It also includes F1 scores, and support for specific classes. You can then assess the model's accuracy at predicting different categories. From Table 2, we can see that the fine-tuned BERT Model performs averagely, especially in predicting requirements other than personality and skills. However, this does not affect its use. Because in the complete process, users, such as HR, will select or delete unwanted results on the final interactive interface. Therefore, the performance of the model is sufficient [43] as shown in table 2.

Table 2. Classification Report of BERT Model

Class	Precision	Recall	F1Score	Support
Personality	0.67	0.64	0.65	653
Skills and Ability	0.34	0.51	0.41	138
Other requirements	0.15	0.10	0.12	216
Accuracy			0.65	
Macro avg	0.23	0.25	0.24	1024
Weighted avg	0.50	0.50	0.50	1024

5.5.2 Determine the Optimal Number of Topics for LDA Model

Topic models produce topics with coherence scores that help to evaluate their quality. This score is used to measure how similar the words are in a topic. Higher coherence measures indicate more meaningful topics. Below are some coherence measurements.

-UMass Coherence

•Definition: UMass coherence measures the coherence between topics by evaluating the words that appear together in the documents.

• Calculation: It is calculated using the conditional probabilities of co-occurrence of words, which are derived from documents that were used to train the models.

-UCI Coherence

•Definition: UCI coherence measures topic coherence by evaluating point-wise mutual information between word pairs using an external reference corpus.

• Calculation: It measures how often words in a topic co-occur more frequently than expected by chance in a large reference corpus.

-C_V Coherence

• Definition: C_V Coherence is a hybrid coherence measure that combines UMass and UCI measures. It uses sliding windows and word vectors to capture context and co-occurrence [45].

•Calculation: The coherence is calculated by aggregating word similarity pairwise within topics. This includes both direct co-occurrences and word embeddings.

As shown in the figure below, LDA performs best in determining the topics of job descriptions when the number of topics is 4, because at this time the values of the three coherence scores are the highest, as shown in figure 4-5. Therefore, the number of topics for the final deployed model is 4.

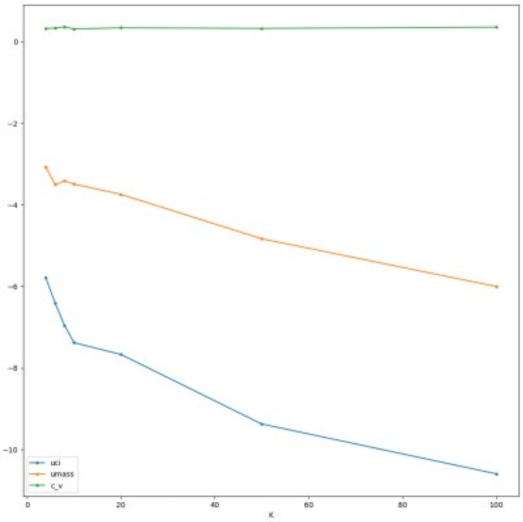

Figure 4. Coherence Scores v.s Number of Topics

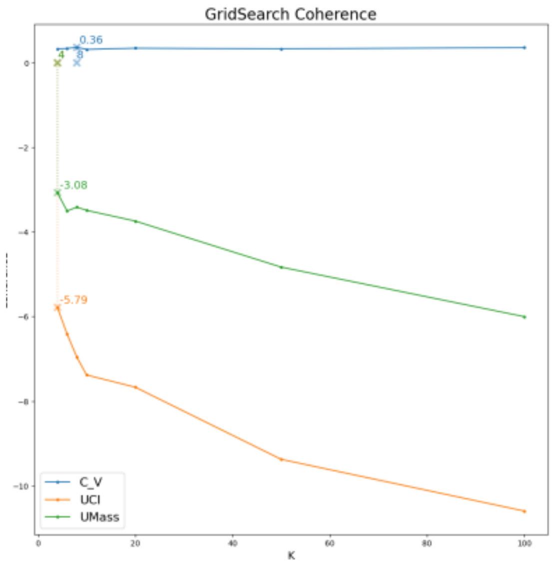

Figure 5. GridSearch Coherence Scores v.s Number of Topics

5.6 Deployment

After the model training is completed, all our work will be packaged into an app with a user-friendly interface. It will be similar to the Chatgpt interface. HR only needs to enter a job description similar to that in the dataset in the dialog box to get an automatically generated questionnaire corresponding to the description. HR can also feedback questions and add descriptions in the dialog box to get more satisfactory results. Our final product will be highly interactive and user-friendly. The MvP would be in deliverables [44].

6. Important Results and Findings

6.1 Results

• BNLD (UK and United States)
– Dictionary size: 17
– Corpus size: 25

– Topic Proportions:
* Topic 1: 32%
* Topic 3: 48%
* Topic 4: 20%
• CNLD (China)
– Dictionary size: 29
– Corpus size: 25
– Topic Proportions:
* Topic 1: 16%
* Topic 2: 4%
* Topic 3: 60%
* Topic 4: 20%
• JD (Linkedin)
– Dictionary size: 91
– Corpus size: 25
– Topic Proportions:
* Topic 1: 8%
* Topic 3: 92%, as shown in figure 6-7.

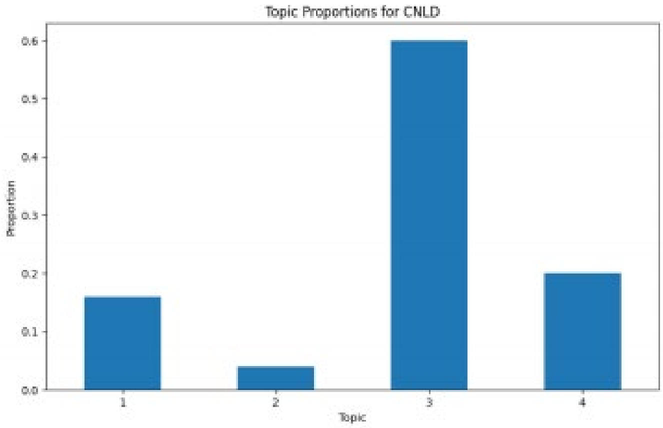

Figure 6. CNLD

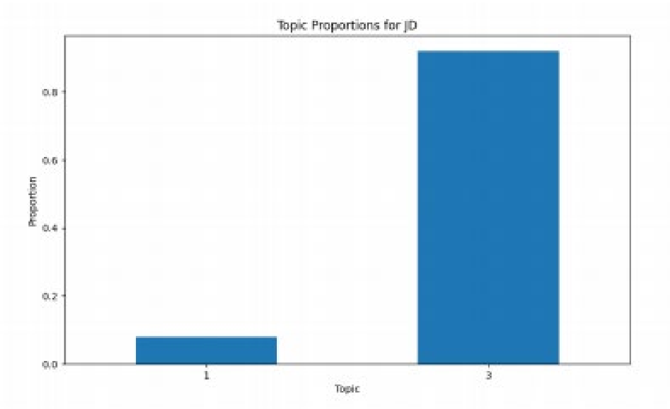

Figure 7. JD

6.2 Interpretation

• Topic Dominance

– For BNLD, Topic 3 is the most dominant topic, accounting for 48% of the corpus. Topics 1 and 4 are represented at 32% and 20% respectively.

– For CNLD, Topic 3 is significantly dominant, covering 60% of the corpus. Topics 1, 2, and 4 have lesser contributions, with Topic 2 being the least represented at 4%.

– For JD, Topic 3 overwhelmingly dominates, representing 92% of the corpus, following Topic 1 has a minimal presence at 8%, but other topics have no weights in this model.

• Topic Distribution and Focus

– The dominance of Topic 3 in CNLD and JD, and its substantial presence in BNLD, suggests a common theme or focus area that is highly relevant across team-work, strong and experienced candidates.

– The variance in the representation of other topics (1, 2, and 4) across sources indicates differing secondary focus areas. For instance, CNLD shows a broader spread across four topics, which may reflect a more diverse job requirements for candidates compared to BNLD and JD.

• Dictionary Size and Corpus Size

– The increasing dictionary size from BNLD (17) to CNLD (29) to JD (91)

suggests that the vocabulary used in Linkedin (JD) is much richer and more varied compared to the more specialized texts in BNLD and CNLD.

–Despite the variations in dictionary size, the corpus size remains consistent across all sources (25), indicating an equal number of documents analyzed for each source, as shown in table 3.

Table 3. Chi-Square Test Results for Topic Distributions

Comparison	**Chi2 Statistic**	**p-value**
BNLD vs CNLD	2.6667	0.4459
BNLD vs JD	12.0571	0.0024**
CNLD vs JD	8.3509	0.0393*

* Significant at $p \leq 0.05$ ** Significant at $p \leq 0.01$

• Chi-Square Test Interpretation

– Table 3 shows the Chi-Square test results indicate that there are significant differences in topic distributions between BNLD and JD ($p = 0.0024$), and between CNLD and JD ($p = 0.0393$). However, the difference between BNLD and CNLD is not statistically significant ($p = 0.4459$). This suggests that while there are notable differences in how topics are distributed between JD and the other two sources, the distribution between CNLD and BNLD sources is relatively similar [45].

6.3 Word Cloud Analysis

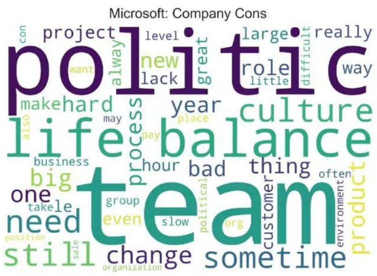

Figure 8. Word Cloud when number of topics optimal

Figure 8 shows the word cloud provides a visual representation of the most frequent terms across the topics. Key terms like team, data, strong, experience, and management appear prominently, reflecting great weight in the correlated topic. The prominence of these terms suggests their central role in the discussions and descriptions within the corpus.

6.4 Conclusion

Topic 3: The consistent dominance of Topic 3 across all sources highlights a key area of focus in team-work and strong experienced candidates.

Diverse Focus: Employers from CNLD (China) shows a more varied focus com- pared to BNLD (UK and United States) and JD (Linkedin), reflecting a broader range of topics covered in job requirements.

Vocabulary Richness: The source JD (Linkedin) have a much richer and varied vocabulary, indicating a more detailed and nuanced description of roles and responsibilities.

Statistical Significance: Significant differences in topic distributions between Linkedin (JD) and the other two sources suggest that Job descriptions from different sources will have preferences for recruitment requirements, which gives us inspiration to study how to eliminate or utilize bias in employment sources during the recruitment process [46].

These insights can help us study the preferences of recruitment requirements from different sources, but the demand for strong employees with team and experience is present in all three sources studied. In particular, the recruitment needs from LinkedIn use more diverse and rich terms than those from China, the United States, and the United King- dom. The recruitment requirement from China are broader than those from the other two sources, which maybe because China has a complete range of industries and a high volume of people, so the recruitment requirements are also richer [47].

7. Limitations and Future Research

7.1 Limitations

While this study offers valuable insights into the use of NLP and AI for recruitment processes, several limitations need to be acknowledged:

Data Source Diversity: The dataset used in this study consists primarily of job descriptions from a small number of sources (BNLD CNLD JD). This limitation may limit the generalizability across other recruitment platforms or geographic regions [48].

Bias in Job Descriptions: Inherently, job descriptions contain biases based on cultural, organizational and individual HR practices. AI models can perpetuate these biases and affect the fairness of the recruitment process.

Model Interpretability: Although advanced models such as BERT are highly ac- curate in text processing, their interpretability is still a challenge. Understanding and explaining these models' decisions is essential, especially when it comes to a domain with high stakes like recruitment.

Limited Focus on Long-term Impacts: This study is primarily focused on the immediate impacts of NLP and AI used in recruitment. The impact of NLP and AI on recruitment, organizational culture, and employee satisfaction was not fully addressed.

Dependency on Manual Annotations: The study relies on manual labeling of data for both training and validation, despite the sophisticated models used. This is a time-consuming process that may introduce human errors into the dataset [49].

7.2 Future Research Directions

In order to overcome these limitations and improve the use of NLP for recruitment, future research should focus on:

Expansion of Data Sources: A large database can lead to better model fitting and training. In addition, we need to find more data sources to verify whether there are differences in the sources of recruitment needs. Bias Mitigation Strategies: An extensive database can improve model fitting and training. We need to collect more data to determine if there are any differences in the recruitment sources.

Enhanced Model Interpret-ability: An extensive database can improve model fitting and training. We need to collect more data to determine if there are any differences in the recruitment sources.

Long-term Impact Studies: In order to assess the impact of AI and NLP in recruitment over time, longitudinal studies are required. It is crucial to assess their impact on employee retention and diversity, as well as organizational culture and performance.

Automation of Annotations: Future research may investigate the use semi-supervised or unsupervised methods to reduce reliance upon manual annotations. Then utilizing unlabeled big datasets can enhance the resilience and precision of AI models.

8. Conclusion

In conclusion, this paper has demonstrated the potential of NLP technology to address the ambiguities, biases, and inefficiencies inherent in current job assessment tools. By leveraging NLP to analyze natural language job descriptions, we have shown that hidden patterns, such as skill requirements and personality traits, can be identified cost-effectively. This approach not only reduces the ambiguities and unconscious biases often present in manual questionnaire compilation but also expedites the hiring process. Our research methodology, which combined regular expressions, the BERT algorithm, and topic modeling, provided a robust framework for extracting and analyzing job description elements, enabling us to uncover regional differences in recruitment requirements. The findings from our data analysis have contributed valuable insights to the existing literature and have implications for both theoretical understanding and practical applications in the field of e-HRM. Despite the promising results, our study is not without limitations, such as the potential for NLP algorithms to perpetuate biases present in the data they are trained on. Future investigations should continue to refine NLP techniques and explore additional variables that may influence recruitment preferences. Overall, this research opens up exciting opportunities for further discovery in the application of NLP technology to improve human resource management practices.

References

[1] Edwards D C. PEOPLE and TECHNOLOGY: A Winning Recruiting Combination[J]. Career Planning & Adult Development Journal, 2016, 32(3).

[2] Collins B W. Tackling unconscious bias in hiring practices: The plight of the Rooney rule[J]. NYUL Rev., 2007, 82: 870.

[3] Upadhyay A K, Khandelwal K. Applying artificial intelligence: implications for recruitment[J]. Strategic HR Review, 2018, 17(5): 255-258.

[4] Chilunjika A, Intauno K, Chilunjika S R. Artificial intelligence and public sector human resource management in South Africa: Opportunities, challenges and prospects[J]. SA Journal of Human Resource Management, 2022, 20: 12.

[5] Tsiskaridze R, Reinhold K, Jarvis M. Innovating HRM recruitment: a comprehensive review of AI deployment[J]. Marketing i menedžment innovacij, 2023, 14(4): 239-254.

[6] Bondarouk T V, Ruël H J M. Electronic Human Resource Management: challenges in the digital era[J]. The international journal of human resource management, 2009, 20(3): 505-514.

[7] Strohmeier S. Research in e-HRM: Review and implications[J]. Human resource management review, 2007, 17(1): 19-37.

[8] Laumer S, Eckhardt A. Help to find the needle in a haystack: integrating recommender systems in an IT supported staff recruitment system[C]//Proceedings of the special interest group on management information system's 47th annual conference on Computer personnel research. 2009: 7-12.

[9] Singh P, Finn D. The effects of Information Technology on recruitment[J]. Journal of Labor Research, 2003, 24(3).

[10] Black J S, van Esch P. AI-enabled recruiting: What is it and how should a manager use it? [J]. Business Horizons, 2020, 63(2): 215-226.

[11] Zahidi F, Imam Y, Hashmi A U, et al. Impact of artificial intelligence on HR management–A[J]. Research gate, 2020.

[12] Rathore S P S. The Impact of AI on Recruitment and Selection Processes: Analysing the role of AI in automating and enhancing recruitment and selection procedures[J]. International Journal for Global Academic & Scientific Research, 2023, 2(2): 51-63.

[13] EL MNOUER O, KATFI A, KATFI H, et al. Artificial Intelligence and Recruitment in Morocco: Innovative Frontiers for Optimizing the Collaborator Experience[J]. International Journal of Accounting, Finance, Auditing, Management and Economics, 2023, 4(6-2): 223-238.

[14] Chen Z. Collaboration among recruiters and artificial intelligence: removing human prejudices in employment[J]. Cognition, Technology & Work, 2023, 25(1): 135-149.

[15] Horodyski P. Applicants' perception of artificial intelligence in the recruitment process[J]. Computers in Human Behavior Reports, 2023, 11: 100303.

[16] Albassam W A. The power of artificial intelligence in recruitment: An analytical review of current AI-based recruitment strategies[J]. International Journal of Professional Business Review, 2023.

[17] Caliskan A, Bryson J J, Narayanan A. Semantics derived automatically from language corpora contain human-like biases[J]. Science, 2017, 356(6334): 183-186.

[18] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead[J]. Nature machine intelligence, 2019, 1(5): 206-215.

[19] Hough L M, Connelly B S. Personality measurement and use in industrial and organizational psychology[J]. Psycnet, 2013.

[20] Barrick M R, Mount M K. The big five personality dimensions and job performance: a meta‐analysis[J]. Personnel psychology, 1991, 44(1): 1-26.

[21] Camara W J, Nathan J S, Puente A E. Psychological test usage: Implications in professional psychology[J]. Professional psychology: Research and practice, 2000, 31(2): 141.

[22] Goldberg L R. An alternative "description of personality": The Big-Five factor structure[M]//Personality and Personality Disorders. Routledge, 2013: 34-47.

[23] Goleman D. Emotional intelligence. Why it can matter more than IQ[J]. Learning, 1996, 24(6): 49-50.

[24] Paulhus D L. Measurement and control of response bias[J]. Measures of Personality and Social Psychological Attitudes/Academic Press, Inc, 1991.

[25] Richardson J D, Lounsbury J W, Bhaskar T, et al. Personality traits and career satisfaction of health care professionals[J]. The health care manager, 2009, 28(3): 218-226.

[26] Birkelund G E, Lancee B, Larsen E N, et al. Gender discrimination in hiring: Evidence from a cross-national harmonized field experiment[J]. European Sociological Review, 2022, 38(3): 337-354.

[27] Weekley J A, Ployhart R E. An introduction to situational judgment testing[J]. Situational judgment tests: Theory, measurement, and application, 2006: 1-10.

[28] Chan D, Schmitt N. Video-based versus paper-and-pencil method of assessment in situational judgment tests: subgroup differences in test performance and face validity perceptions[J]. Journal of applied psychology, 1997, 82(1): 143.

[29] Campos R, Mangaravite V, Pasquali A, et al. YAKE! Keyword extraction from single documents using multiple local features[J]. Information Sciences, 2020, 509: 257-289.

[30] Caliskan A, Bryson J J, Narayanan A. Semantics derived automatically from language corpora contain human-like biases[J]. Science, 2017, 356(6334): 183-186.

[31] Hofstede G. Dimensionalizing cultures: The Hofstede model in context[J]. Online readings in psychology and culture, 2011, 2(1): 8.

[32] Birkelund G E, Lancee B, Larsen E N, et al. Gender discrimination in hiring: Evidence from a cross-national harmonized field experiment[J]. European Sociological Review, 2022, 38(3): 337-354.

[33] Al-Jenaibi B. The scope and impact of workplace diversity in the United Arab Emirates–An initial study[J]. Acta Universitatis Danubius. Communicatio, 2011, 5(1).

[34] Jelodar H, Wang Y, Yuan C, et al. Latent Dirichlet allocation (LDA) and topic modeling: models,

applications, a survey[J]. Multimedia tools and applications, 2019, 78: 15169-15211.

[35] Rose S, Engel D, Cramer N, et al. Automatic keyword extraction from individual documents[J]. Text mining: applications and theory, 2010: 1-20.

[36] Devlin J. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.

[37] Christiansen N D, Burns G N, Montgomery G E. Reconsidering forced-choice item formats for applicant personality assessment[J]. Human Performance, 2005, 18(3): 267-307.

[38] Devlin J. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.

[39] Dieng A B, Ruiz F J R, Blei D M. Topic modeling in embedding spaces[J]. Transactions of the Association for Computational Linguistics, 2020, 8: 439-453.

[40] Goleman D. Emotional intelligence: Why it can matter more than IQ[M]. Bloomsbury Publishing, 2020.

[41] Griffiths T, Jordan M, Tenenbaum J, et al. Hierarchical topic models and the nested Chinese restaurant process[J]. Advances in neural information processing systems, 2003, 16.

[42] Herderich A, Freudenthaler H H, Garcia D. A computational method to reveal psychological constructs from text data[J]. Psychological Methods, 2024.

[43] Hulth A. Improved automatic keyword extraction given more linguistic knowledge[C]//Proceedings of the 2003 conference on Empirical methods in natural language processing. 2003: 216-223.

[44] Liu Y. Roberta: A robustly optimized bert pretraining approach[J]. arXiv preprint arXiv:1907.11692, 2019, 364.

[45] Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures[C]//Proceedings of the eighth ACM international conference on Web search and data mining. 2015: 399-408.

[46] Rosner F, Hinneburg A, Röder M, et al. Evaluating topic coherence measures[J]. arXiv preprint arXiv:1403.6397, 2014.

[47] The Adoption and Effect of Artificial Intelligence on Human Resources Management, Part A[M]. Emerald Publishing Limited, 2023.

[48] Vivek R. Enhancing diversity and reducing bias in recruitment through AI: a review of strategies and challenges[J]. Информатика. Экономика. Управление/Informatics. Economics. Management, 2023, 2(4): 0101-0118.

[49] Yilma B A, Aghenda N, Romero M, et al. Personalised visual art recommendation by learning latent semantic representations[C]//2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA. IEEE, 2020: 1-6.