International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Named Entity Recognition Driven Synthesis of IT Job Descriptions in Morocco: A Comparative Analysis of BERT and BiLSTM Models

Zineb Elkaimbillah[1], Zineb Mcharfi[1], Mohamed Khoual[1] and Bouchra El Asri[1]

Submitted:12/03/2024 **Revised**: 27/04/2024 **Accepted**: 04/05/2024

Abstract*:* The information technology (IT) sector, characterized by its dynamism and diversity, represents a major challenge for jobseekers and recruiters alike, who have to navigate through massive lists of job offers to extract relevant information. This article proposes a new approach to meeting this challenge by integrating Named Entity Recognition (NER) into the synthesis of job descriptions in the IT domain. This exploration in the IT sector offers a significant contribution to the optimization of job search processes and recruitment strategies specific to this sector. Our approach, which includes the conceptualization, data preparation and training of BERT (Bidirectional Encoder Representations from Transformers) and BiLSTM (Bi-directional Long Short-Term Memory) models, enables us to compare the performance of two NER models through in-depth evaluation. The originality of our approach lies in the use of Named Entity Recognition (NER) as the cornerstone of automatic synthesis. By harnessing the power of NER, we simplify and streamline the process of efficiently extracting crucial information such as organizations, locations and job titles. The results underline the transformative potential of NER in improving the accessibility and comprehensibility of complex information contained in job advertisements in the IT sector. By automating the extraction of relevant entities such as job titles, skills required, company names, work locations, responsibilities requested, technical and non-technical skills, diplomas and years of experience required, we facilitate the job search process. Our evaluations show that BERT models outperform BiILSTM models in terms of accuracy and performance in named entity recognition, demonstrating their superiority for this specific task.

Keywords: BERT, BiLSTM, Information Technology, Job descriptions, Named Entity Recognition, Summarization.

1. Introduction

In the ever-changing landscape of information technology (IT), the process of sifting through a multitude of job offers to identify key information can be time-consuming and difficult. IT jobs are unstructured texts that contain important information for candidates, such as job title, skills required, responsibilities requested, company location, etc. This information can be considered as entities in their own right. This information can be thought of as named entities, like words or phrases that designate specific elements of the real world.

To alleviate this problem, researchers have developed innovative approaches to make the job search process more efficient and accurate. One such approach is the automatic synthesis of job offer using Named Entity Recognition (NER). NER is a sub-field of natural language processing [1] that focuses on extracting relevant textual information and identifying named entities in a particular piece of text. By leveraging NER, researchers can automatically extract relevant entities from job advertisements and use this information to synthesize job descriptions that accurately reflect the requirements of a given IT role.

The application of NER technology for text summarization is a popular application in various fields such as business intelligence, automatic writing of news articles, analysis of online comments and product recommendation. The field of IT task analysis is no exception to this trend, and the use of NER for the analysis and synthesis of IT tasks is becoming increasingly common.

The main objective of this article is to explore the field of automated synthesis of IT job offers, by examining how NER, a subset of artificial intelligence, can streamline the extraction of crucial information from job descriptions, helping jobseekers to navigate the complex landscape of IT job offers.

[1]*dept. IMS Team, ADMIR Laboratory Rabat IT Center ENSIAS, Mohammed V University in Rabat, Morocco.*
ORCID ID: 0009-0008-4061-7714
** Corresponding Author Email: zineb_elkaimbillah@um5.ac.ma*

The structure of the rest of this article is as follows: Section 2 presents technical background, then Section 3 reviews related works. Section 3 describes the methodology. Section 4 presents the findings and discussions. Finally, Section 5 concludes the paper and suggests directions for future research.

2. Background

2.1. Named entity recognition

The goal of named entity recognition (NER), a subtask of information extraction, is to locate and categorize named entities in text into predetermined groups, including names of people, places, dates, and other particular entities. It is a crucial component in various natural language processing (NLP) applications including information retrieval, question answering, machine translation, and more.

NER systems can be rule-based, machine learning-based, or deep learning-based. Rule-based systems use handcrafted rules and patterns, while machine learning-based systems rely on statistical models trained on annotated corpora. Deep learning-based systems, particularly those using neural networks, have recently achieved state-of-the-art performance due to their ability to learn complex features from large amounts of data.

2.2. BERT

BERT (Bidirectional Encoder Representations from Transformers) [2] is a model introduced by Google in 2018 that has significantly improved various NLP tasks, including Named Entity Recognition (NER). Its key features include:

Bidirectional Context: BERT processes text in both directions simultaneously, capturing context more effectively than traditional models.

Pre-training and Fine-tuning: BERT is pre-trained on large text corpora using tasks like Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). It can then be fine-tuned on specific tasks with smaller amounts of data.

Transformer Architecture: Utilizing self-attention mechanisms, BERT efficiently processes and links different parts of the input text.

BERT's architecture shown in figure 1 illustrates its input and internal processing structure. The input tokens, represented by the yellow boxes at the bottom, include special tokens such as [CLS] (classification token) and [SEP] (separation token) as well as the actual tokens of the input sequences. These tokens are first converted into embeddings (blue boxes) that combine position, segment and token records to maintain context. These embeddings are then fed into the BERT model, which processes them through several layers of transformers, capturing the bidirectional context. The resulting hidden states (pink boxes) at the top represent contextualized embeddings for each word, which are used for downstream tasks such as Named Entity Recognition (NER). This architecture enables BERT to understand and represent the nuanced context of each token in relation to its surrounding tokens, making it highly effective for a variety of NLP applications.

2.3. BiLSTM

BiLSTM (Bidirectional Long Short-Term Memory) networks improve upon traditional LSTM models by analyzing input sequences in both forward and reverse directions. allowing them to capture context from both past and future states simultaneously. The architecture consists of two LSTM layers: one processes the sequence from the beginning to the end (forward layer), and the other processes it from the end to the beginning (backward layer). The outputs from both layers are concatenated at each time step, providing a comprehensive understanding of the input sequence. This bidirectional approach is particularly effective for tasks like Named Entity Recognition (NER), where understanding the context from both directions can significantly improve the accuracy of entity classification. (Figure 2) [3].

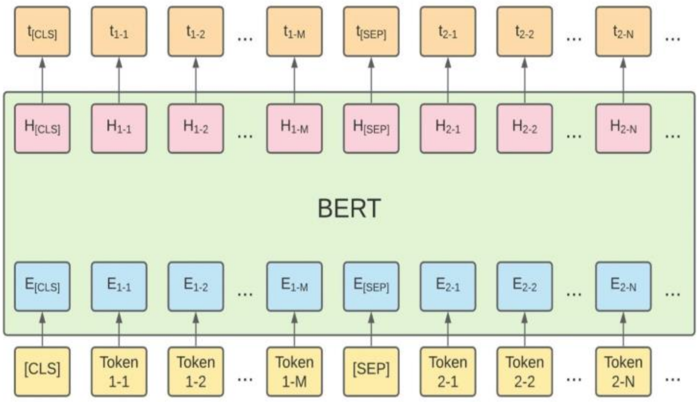

Fig1. BERT model architecture

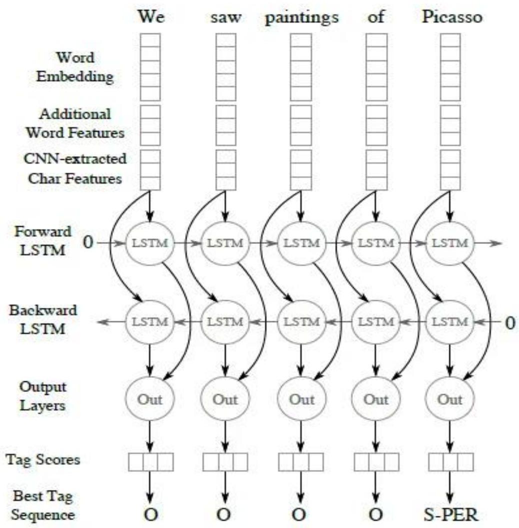

Fig2. BiLSTM model architecture

3. Related works

Extensive research has leveraged deep learning methodologies to automate the process of named entity recognition (NER), and this body of work can be categorized based on the employed model, the nature of the data utilized, and the specific domains under investigation. Thus, this section is structured around three key dimensions: the initial dimension, outlined in Table ,1 provides an overview of recent NER endeavors across various domains; the second dimension, detailed in Table 2, delves into the implementation of NER within the education sector; and the third dimension encapsulates studies showcasing the application of NER in the context of text summarization.

3.1. NER model applied in different domains

[4] developed a BERT-based model for Named Entity Recognition in the German legal domain using a newly created annotated dataset. The model achieved high accuracy, making it a useful tool for legal practitioners in analyzing legal documents.

[5] proposed a novel FinBERT-MRC model for financial named entity recognition using BERT and machine reading comprehension. The model shows impressive results in recognizing different financial named entities in texts, making it an essential tool for financial analysts and investment professionals.

[6] created a manually annotated dataset by extracting cyber security related texts from different sources, including news articles, blogs, and research papers. They proposed a word embeddings-based NER model for cyber security domain that outperforms traditional NER models. The ability to learn context-specific entities enhances the model's effectiveness for detecting vulnerabilities and threats in cyber systems.

[7] proposed a hybrid ALBERT-BiLSTM model for Chinese NER in football to improve the accuracy of identifying football-specific named entities. The model shows high performance in recognizing relevant entities in their collected dataset and benchmark dataset, making it a valuable tool for the football industry.

[8] applied an unsupervised approach using a weighted distributional semantic model for NER in the agricultural domain. The method shows high accuracy in categorizing and identifying relevant entities in agricultural texts, making it a valuable tool for agricultural professionals.

[9] proposed a method for explaining the BERT model's decisions for near-duplicate news article detection based on NER. The generated attention maps provide valuable insights into the model's decision-making process, enabling more accurate and trustworthy decision-making for news professionals.

[10] applied a BERT-based approach for Chinese Semantic NER in marine engine room systems. The approach outperforms other state-of-the-art models, and the study provides insights into the BERT model's performance based on different factors. This

Table 1. Related work for NER applied in different domain

REF	*NER Usage*	*Data Source*	*Models*	*Evaluation Measures*
[4]	Facilitating more accurate and context-aware information extraction from legal documents written in German.	Legal documents in German	BERT	F1=96%
[5]	Integrate a large amount of financial knowledge existing in unstructured texts into structured formats	ChFinAnn	FinBERT-MRC	F1=92.78%
[6]	Have a machine-assisted analysis of such information	cyber security information (Google News, Wikipedia)	BERT	F1=0,974
[7]	Analyze the developmental tendencies of football	Chinese corpus in the field of football	ALBERT-BiLSTM	F1=84.37%
[8]	AGRONER system (Agricultural)	Corpus agricoles	BERT	F1=80.43%
[9]	Detect near-duplicates	newspaper articles	BERT	F1=97%
[10]	Knowledge storage for marine engine room systems	Marine engine room semantic datasets	BERT	F1=90.30%
[11]	Food safety Domain	Food Texts on the web, Weibo NER	ERNIE-BiLSTM-CRF	F1=70%
[12]	Legal text processing	Turkish legal texts	BiLSTM	F1=92,27%
[13]	Problem of multiple meanings, long lengths, and close connections with context information in the recognition of named entities in the field of Chinese tourism	tourism dataset: (Go.com and Baidu Encyclopedia)	ERNIE-CNN-BiLSTM-CRF	F1= 98.60

research has practical applications in the field of marine. engineering, facilitating more accurate and efficient data analysis and decision-making processes.

[11] proposed a novel Chinese NER approach based on ERNIE-BiLSTM-CRF for the food safety domain. The approach outperforms other state-of-the-art models, and the study provides insights into the impact of different components in the proposed approach. This research has practical applications in the field of food safety, facilitating more accurate and efficient data analysis and decision-making processes.

This study [12] has important practical applications, as it could facilitate the development of Turkish legal text processing applications, such as information extraction, summarization, and translation. By accurately identifying and classifying named-entities, the system could help automate legal document analysis and improve the efficiency of legal professionals and researchers.

[13] proposed a tourism NER method based on knowledge enhancement. The method outperforms other state-of-the-art models and shows the importance of incorporating external knowledge sources for improving entity representation and recognition. Maintaining the Integrity of the Specifications.

3.2. NER applied in field of Education

[14] proposed a SkillNER approach that can automatically mine and map soft skills from any text. The approach outperforms other state-of-the-art methods and has practical applications in talent management.

[15] proposed a hybrid approach for NER in resumes that combines rule-based and machine learning-based methods. The approach shows promising results and has important practical applications in talent acquisition and management where accurate NER is crucial for identifying relevant candidates and assessing their suitability for a particular job. By combining rule-based and machine learning-based methods, the proposed approach could facilitate more efficient and accurate resume screening and analysis processes.

[16] explore a Title2Vec approach for ONER and other applications. The approach shows promising results and has the potential to facilitate more efficient talent acquisition and career management processes.

Table 2. Related work for NER applied in education domain

REF	*NER Usage*	*Data Source*	*Entities*	*Models*	*Evaluation Measures*
[14]	detection of communities of job profiles based on their shared soft skills	Scientific papers ESCO (European Skill Qualification and Occupation)	Skills	SVM	72.6%
[15]	Extract education and work experience Information from resumes.	Resumes in the field of education	City -Country History Diploma Diploma Section Job Title Oral Talent	RoBERT	89.63%
[16]	Occupational data mining for industry and job.	IPOD datasets	Responsibility Function Location	LSTM-CRF	99.8%

3.3. Summarization text Using NER

The integration of Named Entity Recognition (NER) in text summarization has garnered attention in various research works. Here are a few related works on the topic:

The proposed method in [17] has important practical applications drug discovery and development, regulatory compliance, and in the pharmaceutical industry, where accurate NER is crucial for adverse event reporting. By applying text summarization techniques, the proposed method could facilitate more efficient and accurate NER for pharmaceutical articles.

[18] proposed a method to enhance sequence-to-sequence models for abstractive text summarization using word sense disambiguation and semantic content generalization. The method shows promising results and has important practical applications in various fields.

[19] demonstrates the effectiveness of using named entities in text summarization and proposes a graph-based method for summarizing Czech news articles. The proposed method could have important practical applications in fields such as journalism, media monitoring, and information retrieval.

[20] presented the effectiveness of using document clustering and named entity recognition in automatic text summarization. The proposed method could have important practical applications in fields such as journalism, information retrieval, and data analytics

4. Methodology

4.1. General architecture

The methodology presented in figure 3, aims to collect a diverse dataset of job offers in the IT field, pre-process and evaluate the performance of the models (Bert and Bilstm) for the NER task by comparing them to select the best model used to extract the entities, that are used to generate the synthesized job descriptions

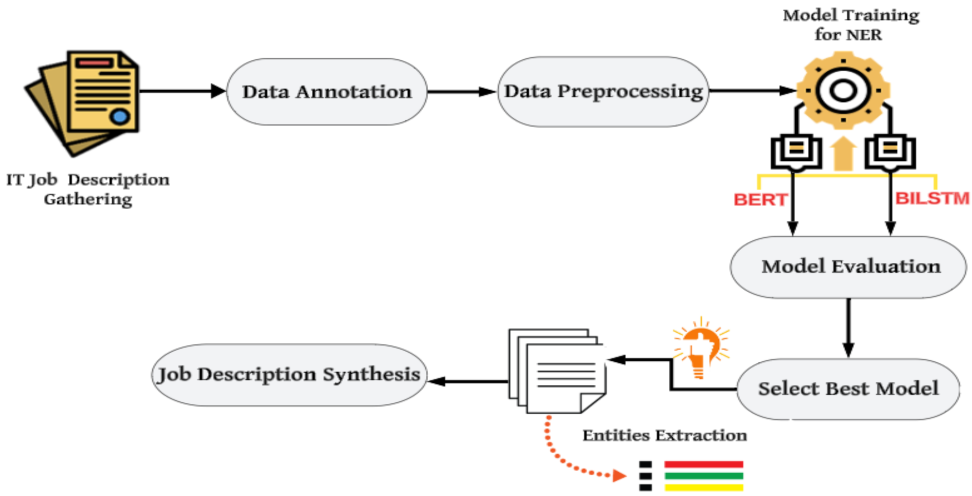

Fig 3. The process of NER-based job description synthesis

4.2. Description of Dataset

This article analyzes the descriptions of job offers related to the IT domain in Morocco. The dataset collected online, comprising 9 entity types (Company, Location, Profile, Diploma, Experience, Formation, ItSkill, SoftSkill, and Responsibility), are tagged using the BIO (Inside, Outside, Beginning) tagging method [21], which comprises 19 tags with over 6900 words in total as shown in figure 4 and Figure 5. Further details on the method and the annotation tool used can be found in the article [22].

4.3. Model implementation

4.3.1. Training the NER model with BERT

Integrating BERT into my corpus of IT job offers, annotated in BIO for named entity recognition, was a strategic decision to improve the accuracy of text analysis. BERT (Bidirectional Encoder Representations from Transformers) is renowned for its ability to understand the bidirectional context of words in a sentence, making it particularly suited to the complex task of recognizing entities in often nuanced job descriptions.

This division into three parts was chosen to enable effective cross-validation and evaluation of our model, which is part of an approach to guaranteeing robust modeling and balanced evaluation. The training phase, representing the majority of the data, enables the model to acquire a thorough understanding of the structures and patterns in the corpus. The development partition is used to fine-tune the model's hyperparameters, while the test partition objectively evaluates the model's performance on unknown data.

The tokenization stage is critical when employing BERT for Named Entity Recognition (NER) in IT job descriptions.

During this phase, the input text is broken down into smaller units typically words or subword fragments (subtokens) which are then fed into the BERT model. This is essential to align the text with the requirements of BERT, which operates at a subtoken level of granularity. In this way, tokenization ensures an accurate, tailored representation of the language specific to IT job offers, optimizing model performance during the training stage and named entity recognition.

4.3.2. Training the NER model with BiLSTM

Bi-directional Long Short-Term Memory (BiLSTM) is a popular neural network model that is used for sequence labeling tasks, including NER, due to its ability to capture relevant context information. Several studies have been conducted on the effectiveness of BiLSTM in NER tasks [23].

One such study was conducted by [24], the method has been trained and evaluated on various scientific literature datasets, achieving state-of-the-art results in many cases. The model has been shown to effectively handle the complexity of scientific language and the diversity of named entities present in this domain.

	Tag	count
0	B-COMPANY	130
1	B-DIPLOMA	86
2	B-EXPERIENCE	180
3	B-FORMATION	105
4	B-ITSKILL	1171
5	B-LOCATION	151
6	B-PROFILE	173
7	B-RESPONSABILITY	540
8	B-SOFTSKILL	348
9	I-COMPANY	43
10	I-DIPLOMA	93
11	I-EXPERIENCE	357
12	I-FORMATION	84
13	I-ITSKILL	484
14	I-LOCATION	9
15	I-PROFILE	359
16	I-RESPONSABILITY	2172
17	I-SOFTSKILL	449

Fig 4. Count of IOB tags assigned to each entity

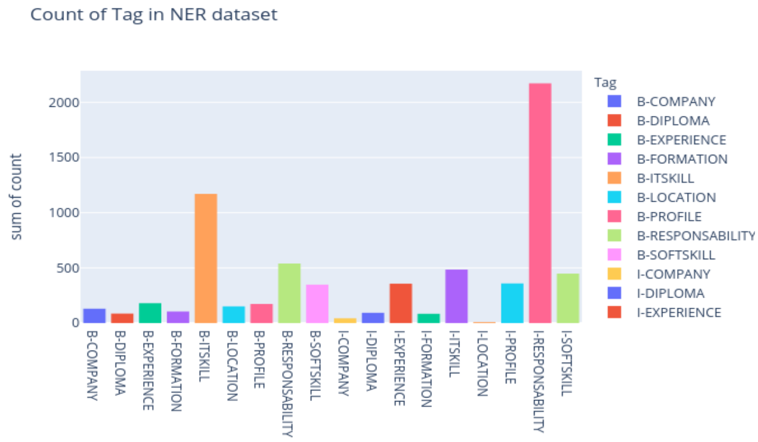

Fig 5. Count of Tag in NER dataset

Likewise, [25] introduced a method for Chinese clinical named entity recognition using a deep learning framework that combines multi-head self-attention with BiLSTM-CRF. This approach shows promise in effectively identifying and extracting named entities from Chinese clinical texts.

Data manipulation involves symbolizing each word in sentences and assigning corresponding labels. Vocabularies are created for words and labels based on the tokenized data. The word-index dictionary assigns a unique index to each word, facilitating subsequent numerical representation.

The data is divided into training and test sets, and the model architecture comprises an input layer, an integration layer, a spatial stall and a bidirectional LSTM layer. The model is compiled with an Adam optimizer, a sparse categorical cross-entropy loss and an accuracy metric.

An early stop is implemented, interrupting training if val_accuracy does not improve after 15 epochs. Model fitting is then used for training, with test data used for validation

4.4. Job Description Synthesis Process

Having applied named entity recognition (NER) to IT job descriptions using the BERT and BiLSTM models, we compared their performance in order to select the most effective model. Synthesizing an IT job offer from NER results can be done systematically by following a defined process.

First, the relevant entities extracted by the NER model, such as job titles (IT profile), company, location, IT skills, soft skills, responsibilities, degree, years of experience and training, are identified, as explained in detail in the article [19]. Next, these entities are prioritized according to their importance and relevance to the candidate. Once the entities have been ranked, a summary of each job offer is generated, focusing on the most frequent and relevant entities. This synthesis includes a summary of the key information extracted, presented in a concise and structured manner. The generated summaries are evaluated to check their relevance and accuracy against the original job descriptions. This stage includes feedback from users to improve the quality of the summaries.

This process enables job offers to be summarized using the key entities identified by the NER, generating a result for each offer based on the most frequent and relevant information. Summaries produced in this way can greatly facilitate the job search and decision-making process for candidates, by presenting essential information in a clear and structured way.

5. Results & Discussion

5.1. Results

The BiLSTM model shows interesting results when trained on the IT job vacancy dataset. According to Figure 6, the training and

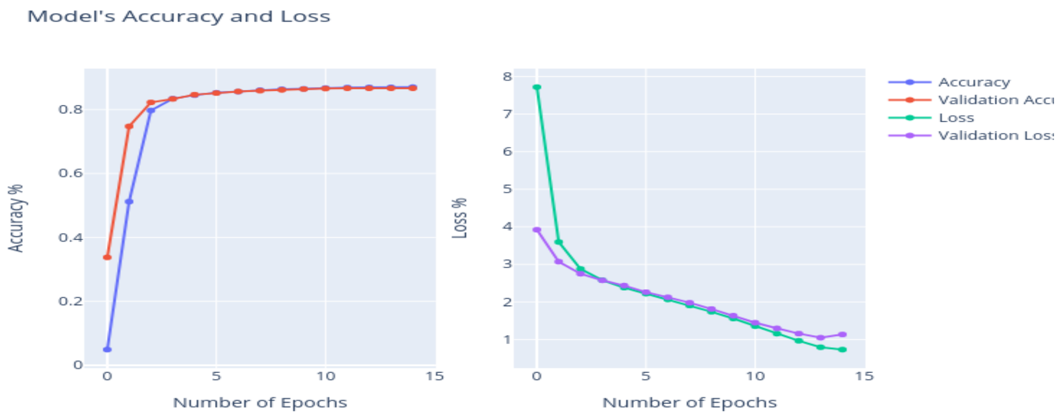

Fig 6. Count of IOB tags assigned to each entity

validation curves, the model achieves an accuracy of 0.86 after several epochs, indicating that the model can generalize well on the training and validation data. However, the training loss curve stabilizes at an average value of 0.73, and the validation loss reaches 1.13. This suggests a slight over-fitting problem, where the model learns from specific patterns in the training data but fails to generalize perfectly to the validation data. The difference between training loss and validation loss is significant here: a higher validation loss than training loss indicates that the model could benefit from further regularization or hyperparameter tuning to improve its generalization ability.accuracy and loss of the BiLSTM model.

For the BERT model, the results are very encouraging, as shown in figure 7. The learning loss curve stabilizes at an average value

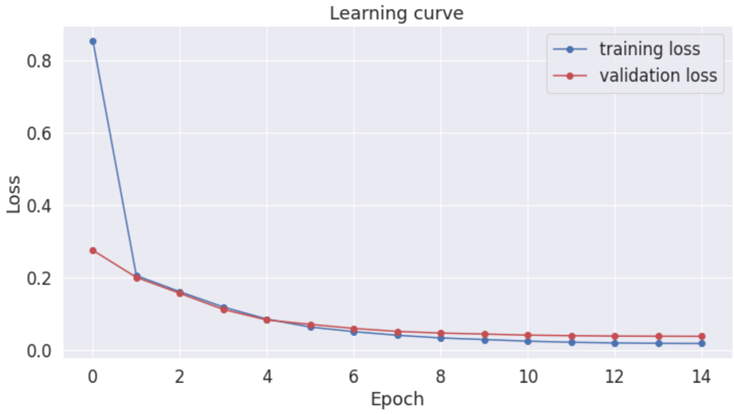

Fig 7. Count of Tag in NER dataset

of 0.88 after several epochs, indicating the model's excellent learning and generalization capability. Moreover, this accuracy of 0.88 shows that the BERT model outperforms the BiLSTM model in terms of accuracy on this dataset. The small difference between training loss and validation loss indicates that the model does not suffer from overfitting and is well calibrated for this specific task. BERT's superior performance can be attributed to its ability to better understand context and complex relationships in texts, which is crucial for analyzing job offers in the IT sector. BERT

uses bidirectional attention, which enables it to capture long-term dependencies and obtain richer, more contextual word representations than BiLSTM. This division into three parts was chosen to enable effective cross-validation and evaluation of our model, which is part of an approach to guaranteeing robust modeling and balanced.

The results obtained when applying the BERT and BiLSTM models to a dataset of computer job descriptions for named entity recognition provide significant insights as summarized in Table 3. The BERT model shows superior overall performance, these results underline the power of pre-trained language models such as BERT in solving natural language processing tasks, providing improved performance over more traditional architectures such as BiLSTM, although both models perform well, BERT stands out for its better generalization ability and slightly higher accuracy, making it the preferred model for analyzing IT job postings in this case. An example of a BI Engineer profile job offer obtained by prediction using our best model BERT is shown below in Figure 8.

5.2. Evaluation of NER Job Syntheses

To evaluate the quality of job, offer syntheses in IT generated by Named Entity Recognition (NER), we formed a panel of evaluators comprising IT job candidates. These evaluators were trained to understand the criteria for evaluating the syntheses, including relevance, accuracy, clarity, and usefulness.

Next, we presented the evaluators with the original job descriptions and the generated syntheses. Each synthesis was evaluated based on the following criteria:

- **Relevance**: Does the synthesis cover all essential information?
- **Accuracy**: Are the extracted details accurate and correctly represented?
- **Clarity**: Is the synthesis clear and easy to understand?
- **Usefulness**: Is the synthesis useful for candidates?

Evaluators rated each criterion on a scale of 1 to 5, with 1 indicating very poor performance and 5 indicating excellent performance. Feedback was collected through detailed questionnaires and feedback sessions, and this feedback was analysed to identify the strengths and weaknesses of the generated syntheses.

The results of this evaluation were modelled in a histogram (see figure 9), providing a clear visualization of the average scores for each criterion. The data shows that the relevance of the syntheses received an average score of 4.3, indicating that the syntheses effectively covered the essential information from the original job descriptions. Accuracy received an average score of 4.1, showing that the extracted information was generally accurate and correctly represented. In terms of clarity, the syntheses scored particularly well, with an average of 4.5, indicating that they were clear and easy to understand. Finally, the usefulness of the syntheses, rated at 4.2, suggests that these syntheses are perceived as helpful for candidates by facilitating a quick and comprehensive

Table 3. Comparisons of the proposed Models Performances using IT job offer dataset

Deep Learning Model	*Validation Loss*	*Average train loss*	*Accuracy*
BERT	0,039	0,02	0,88
Bidirectional LSTM	1,13	0,73	0,86

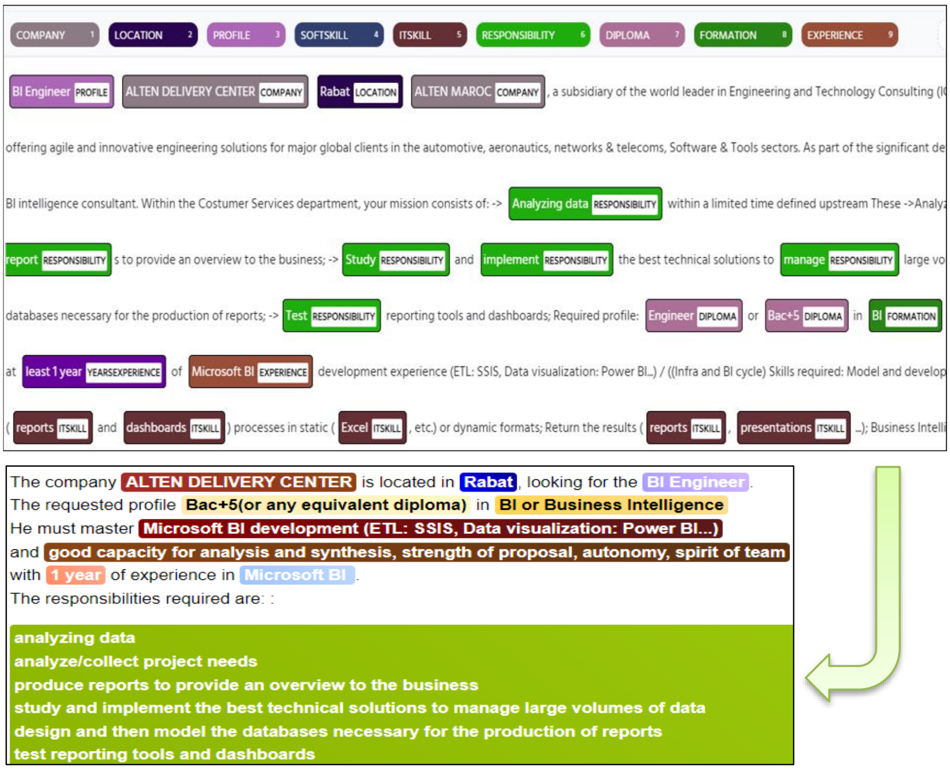

Fig 8 . IT Job Offer of a profile BI Engineer and its Summary as an output

understanding of the job offers.

The resulting histogram from these evaluations shows a balanced distribution of scores, with overall positive performance for each criterion. This feedback highlights the added value of the syntheses generated via NER while indicating areas for continuous improvement to achieve even higher levels of accuracy.

6. Conclusion

In summary, the integration of Named Entity Recognition (NER) into the automatic synthesis of Information Technology (IT) jobs signifies a significant advancement in optimizing the recruitment process. Leveraging the capabilities of NER, we have demonstrated the proficiency to extract crucial details, including organizational entities, locations, individuals, and job titles, from extensive job descriptions. This not only streamlines information retrieval for job seekers but also furnishes recruiters with a valuable tool for swiftly and comprehensively grasping the nuances of diverse IT job opportunities. The exploration of this automated synthesis approach emphasizes the transformative potential of NER within the realm of IT job listings. The discernment and categorization of entities within job descriptions contribute to the formulation of concise and informative job summaries. As the IT industry evolves, the imperative for efficient and intelligent recruitment tools becomes increasingly evident, with NER standing out as a valuable asset to meet these evolving demands.

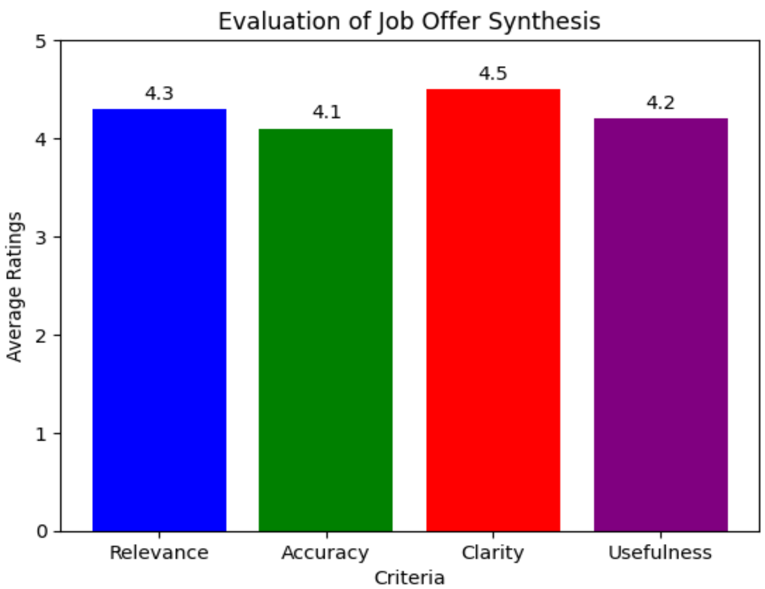

Fig 9 . Evaluation of Job Offer Synthesis

Looking ahead, further research and development in this domain can delve into optimizations, scalability, and the integration of advanced natural language processing techniques to enhance the synthesis process. Additionally, tailoring the approach to align with the dynamic landscape of IT job requirements and incorporating domain-specific knowledge could further enhance the precision and relevance of automated summaries.

REFERENCES

[1] Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. Journal of the American Medical Informatics Association, 18(5), 544-551.

[2] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[3] Chiu, J. P., & Nichols, E. (2016). Named entity recognition with bidirectional LSTM-CNNs. Transactions of the association for computational linguistics, 4, 357-370.

[4] Darji, H., Mitrović, J., & Granitzer, M. (2023). German BERT model for legal named entity recognition. arXiv preprint arXiv:2303.05388.

[5] Zhang, Y., & Zhang, H. (2023). FinBERT–MRC: Financial Named Entity Recognition Using BERT Under the Machine Reading Comprehension Paradigm. Neural Processing Letters, 1-21

[6] Srivastava, S., Paul, B., & Gupta, D. (2023). Study of Word Embeddings for Enhanced Cyber Security Named Entity Recognition. Procedia Computer Science, 218, 449-460

[7] An, Q., Pan, B., Liu, Z., Du, S., & Cui, Y. (2023). Chinese Named Entity Recognition in Football Based on ALBERT-Bilstm Model. Applied Sciences, 13(19), 10814.

[8] Veena, G., Kanjirangat, V., & Gupta, D. (2023). AGRONER: An unsupervised agriculture named entity recognition using weighted distributional semantic model. Expert Systems with Applications, 229, 120440.

[9] Novo, A. S., & Gedikli, F. (2023, February). Explaining BERT model decisions for near-duplicate news article detection based on named entity recognition. In 2023 IEEE 17th International Conference on Semantic Computing (ICSC) (pp. 278-281). IEEE.

[10] Shen, H., Cao, H., Sun, G., & Chen, D. (2023). Research on Chinese Semantic Named Entity Recognition in Marine Engine Room Systems Based on BERT. Journal of Marine Science and Engineering, 11(7), 1266.

[11] Yuan, T., Qin, X., & Wei, C. (2023). A Chinese Named Entity Recognition Method Based on ERNIE-Bilstm-CRF for Food Safety Domain. Applied Sciences, 13(5), 2849.

[12] Çetindağ, C., Yazıcıoğlu, B., & Koç, A. (2023). Named-entity recognition in Turkish legal texts. Natural Language Engineering, 29(3), 615-642.

[13] Leng, T., Altenbek, G., Ma, Y., & Haisa, G. (2023, October). Tourism named entity recognition method based on knowledge enhancement. In Fifth International Conference on Artificial Intelligence and Computer Science (AICS 2023) (Vol. 12803, pp. 782-789). SPIE.

[14] Fareri, Silvia, et al. "SkillNER: Mining and mapping soft skills from any text." Expert Systems with Applications 184 (2021): 115544.

[15] Kesim, E., & Deliahmetoglu, A. (2023). Named entity recognition in resumes. arXiv preprint arXiv:2306.13062.

[16] Liu, J., Ng, Y. C., Gui, Z., Singhal, T., Blessing, L. T., Wood, K. L., & Lim, K. H. (2022). Title2Vec: A contextual job title embedding for occupational named entity recognition and other applications. Journal of Big Data, 9(1), 99.

[17] Dobreva, J., Jofche, N., Jovanovik, M., & Trajanov, D. (2020). Improving NER performance by applying text summarization on pharmaceutical articles. In ICT Innovations 2020. Machine Learning and Applications: 12th International Conference, ICT Innovations 2020, Skopje, North Macedonia, September 24–26, 2020, Proceedings 12 (pp. 87-97). Springer International Publishing.

[18] Kouris, P., Alexandridis, G., & Stafylopatis, A. (2021). Abstractive text summarization: Enhancing sequence-to-sequence models using word sense disambiguation and semantic content generalization. Computational Linguistics, 47(4), 813-859.

[19] Marek, P., Müller, Š., Konrád, J., Lorenc, P., Pichl, J., & Šedivý, J. (2021). Text summarization of czech news articles using named entities. arXiv preprint arXiv:2104.10454.

[20] Senthamizh, S. R., & Arutchelvan, K. (2022). Automatic text summarization using document clustering named entity recognition. International Journal of Advanced Computer Science and Applications, 13(9).

[21] Diab, M. (2009, April). Second generation AMIRA tools for Arabic processing: Fast and robust tokenization, POS tagging, and base phrase chunking. In 2nd international conference on Arabic language resources and tools (Vol. 110, p. 198).

[22] Elkaimbillah, Z., El Asri, B., Mikram, M., & Rhanoui, M. (2023). Construction of an Ontology-based Document Collection for the IT Job Offer in Morocco. International Journal of Advanced Computer Science and Applications, 14(7).

[23] Roy, Bipraneel, and Hon Cheung. "A deep learning approach for intrusion detection in internet of things using bi-directional long short-term memory recurrent neural network." 2018 28th international telecommunication networks and applications conference (ITNAC). IEEE, 2018.

[24] Hou, Linlin, et al. "Method and dataset entity mining in scientific literature: a CNN+ Bilstm model with self-attention." Knowledge-Based Systems 235 (2022): 107621.

[25] An, Y., Xia, X., Chen, X., Wu, F. X., & Wang, J. (2022). Chinese clinical named entity recognition via multi-head self-attention based Bilstm-CRF. Artificial Intelligence in Medicine, 127, 102282.