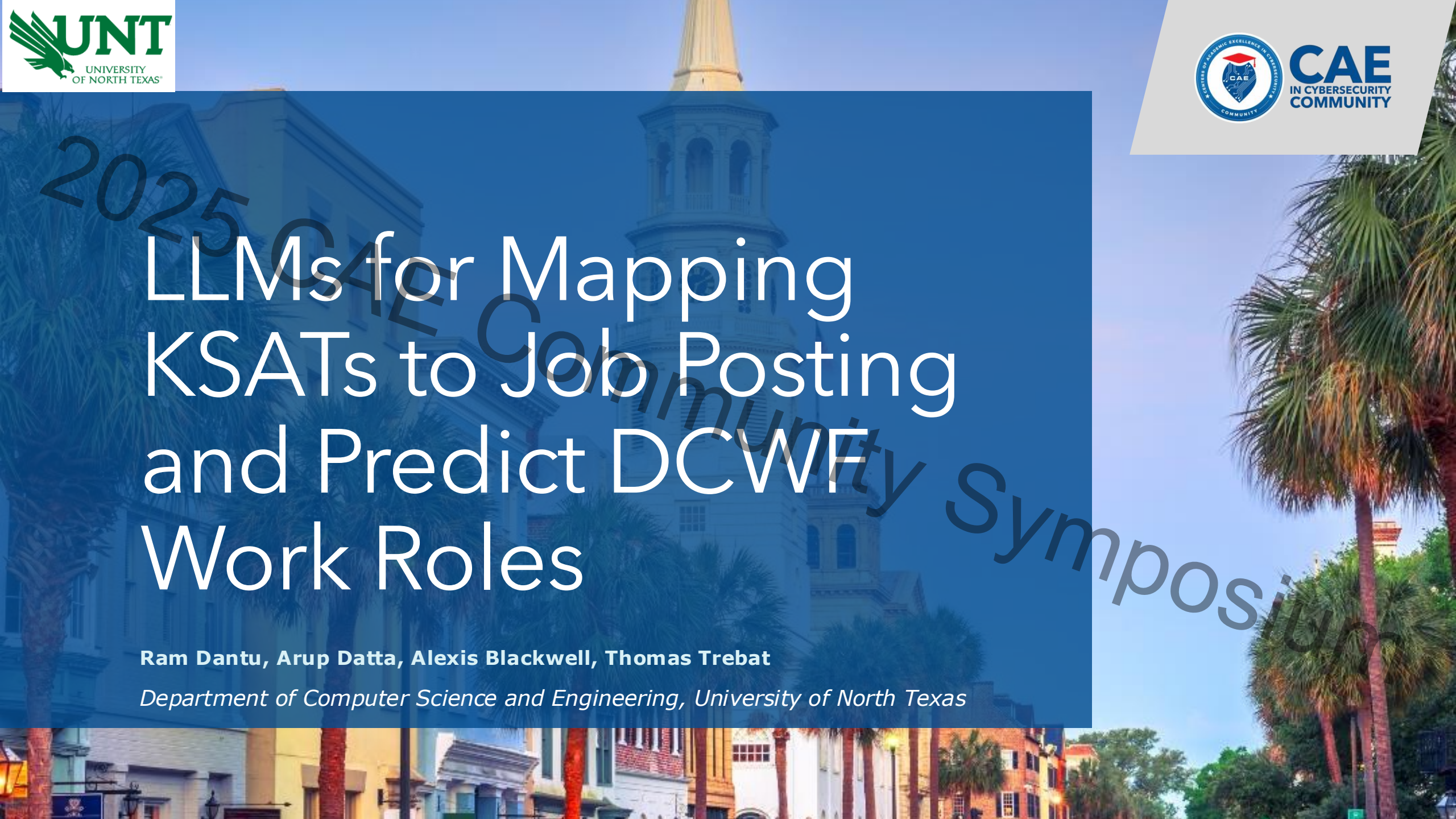


LLMs for Mapping KSATs to Job Posting and Predict DCWF Work Roles

Ram Dantu, Arup Datta, Alexis Blackwell, Thomas Trebat

Department of Computer Science and Engineering, University of North Texas

Abstract

Abstract: A crucial part in today's rapidly changing labor market for both companies and job seekers is accurately determining skills needed for a particular job. Job descriptions include a wealth of information about the duties, abilities, and credentials required to carry out a particular job. This information can be extremely difficult to extract. Additionally, the extracted data is frequently unstructured leading to problems in the usability of the information.

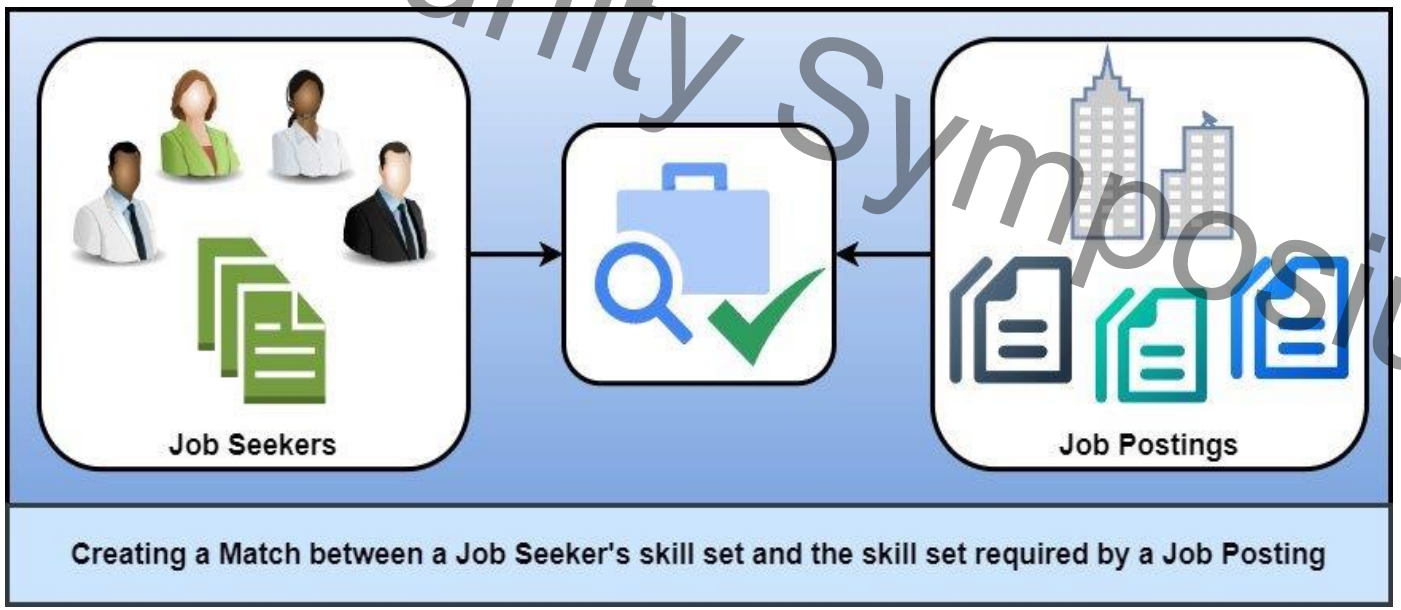
The standardized DoD Cyber Workforce Framework (DCWF) can assist job searchers in locating condensed and organized data on the knowledge, abilities, and skills needed as well as the task units (also known as KSAT) that must be completed to be successful in that position. However, there are no current currently no effective approaches available that can truly extract information from job postings advertising and map those with the KSATs specified in the DCWF.

Large Language Models (LLMs) with their vast knowledge base and ability to be expanded to handle nearly any task without any training. LLMs can learn from examples given in the input context and use that knowledge to perform similar tasks.

This research work aims to fill the skill extraction and mapping gap by leveraging the reasoning capabilities of LLMs and few-shot learning and providing an effective method that identifies explicit and implicit skills and aligns them with the DCWF framework.

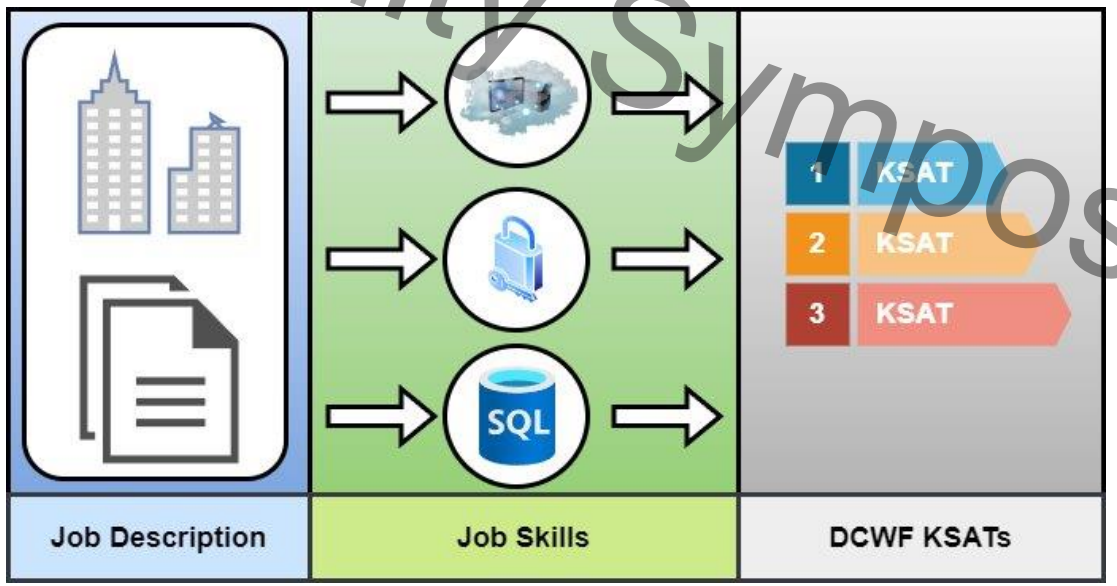
DCWF Skill Extraction

- Large Language Models (LLMs) can be used to identify explicit and implicit skills in the skill extraction process.
- Extracted Job skills can be mapped to the DCWF KSAT's as a holistic approach to determine if an applicant could be successful at a particular role.



Problem Motivation

- DCWF offers KSATs as an option to determine which knowledge, abilities, skills, and task units must be completed to be successful in a work role.
- It can be difficult to extract skills from job description.
  - Often extracted data is in an unstructured format.
- Most methods to investigate matching utilizes keywords.
  - Using keywords loses the latent information when mapping it with the DCWF KSAT's



Dataset Preparation

Role Prediction Dataset

1

100 Job postings for different role such as Cyber Defense Analyst, Data Scientist, Database Administrator. For each job descriptions, the top 5 suitable roles were listed in descending order of best match.

KSAT Mapping Dataset

2

The first 20 job postings of the 100 annotated Job Postings were mapped to the KSATs depending on the relevancy of the KSAT with the job description.

LLM Solution

LLM Model	LLama-3.1-8B
Job Database	MongoDB
Python Libraries	Pandas, Numpy, transformers, matplotlib, bitsandbytes

1

Prompt Template

Prompt template for role prediction and KSAT mapping task.

2

Top 5 Role Prediction

Top 5 role prediction with the selected template for the 100 annotated job descriptions and compare their results with the ground truth and generate necessary matrices

3

Predicting the KSAT's

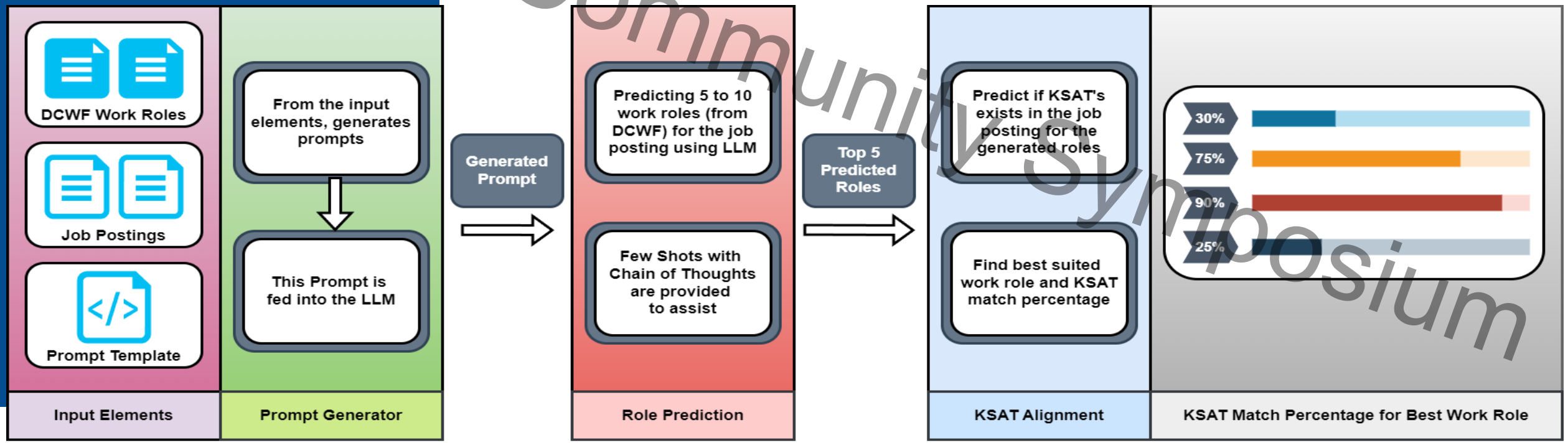
Predicting which KSAT's are relevant for the predicted roles and generating required matrices

4

Refining the Job Description

Running the 2nd and 3rd stage with refining the job description by removing irrelevant information from the beginning and the end of the job description.

Methodology and Architecture



Accuracy Results

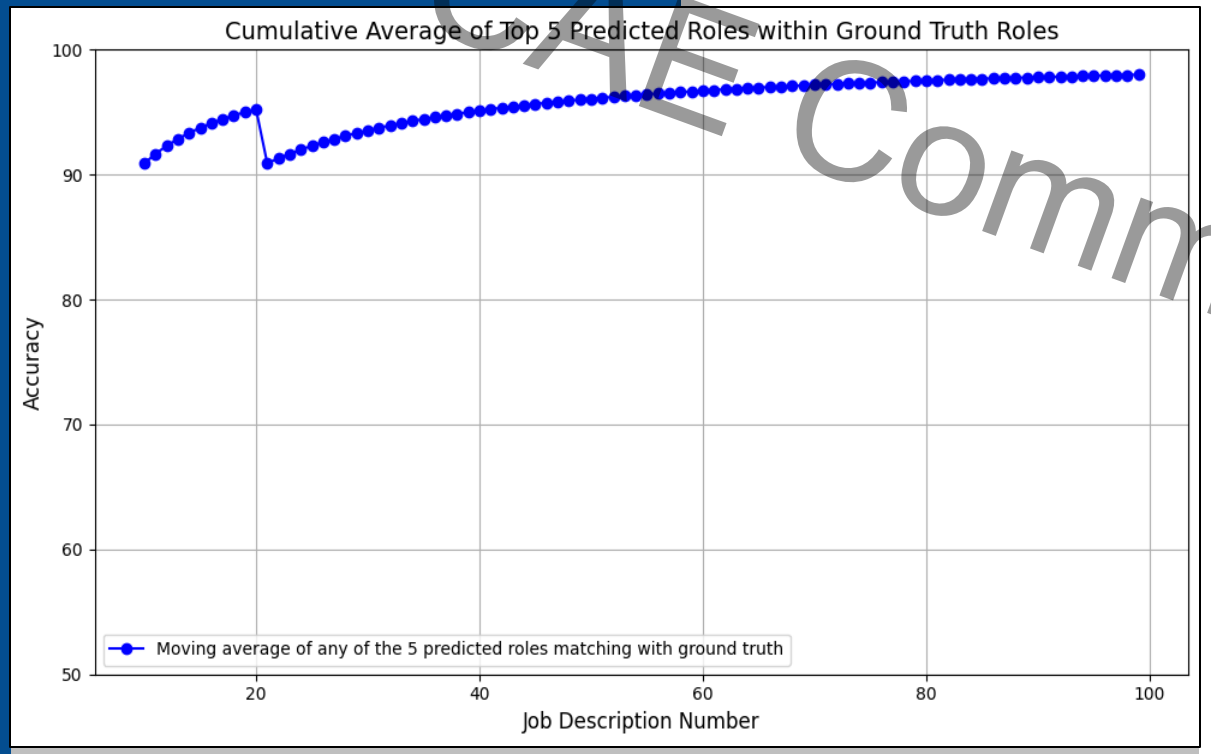


Fig. 5. The cumulative average of the Top 5 Predicted Role compared to the Ground Truth Roles. The final accuracy is approximately 96%.

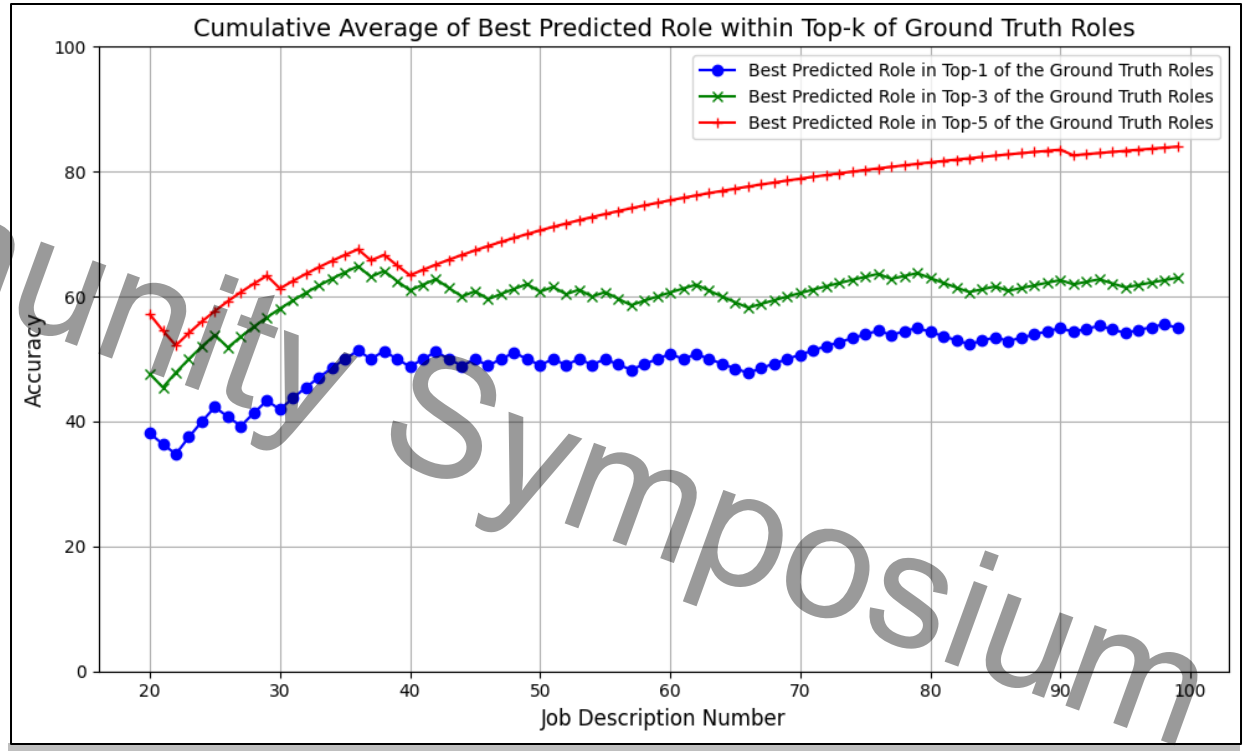


Fig. 6. The cumulative average of the Best Predicted Role within the Top-k Ground Truth Roles. The final accuracy for the Top 1, Top 3, and Top 5 predictions are 53%, 62%, and 83% respectively.

Predicted Roles Results

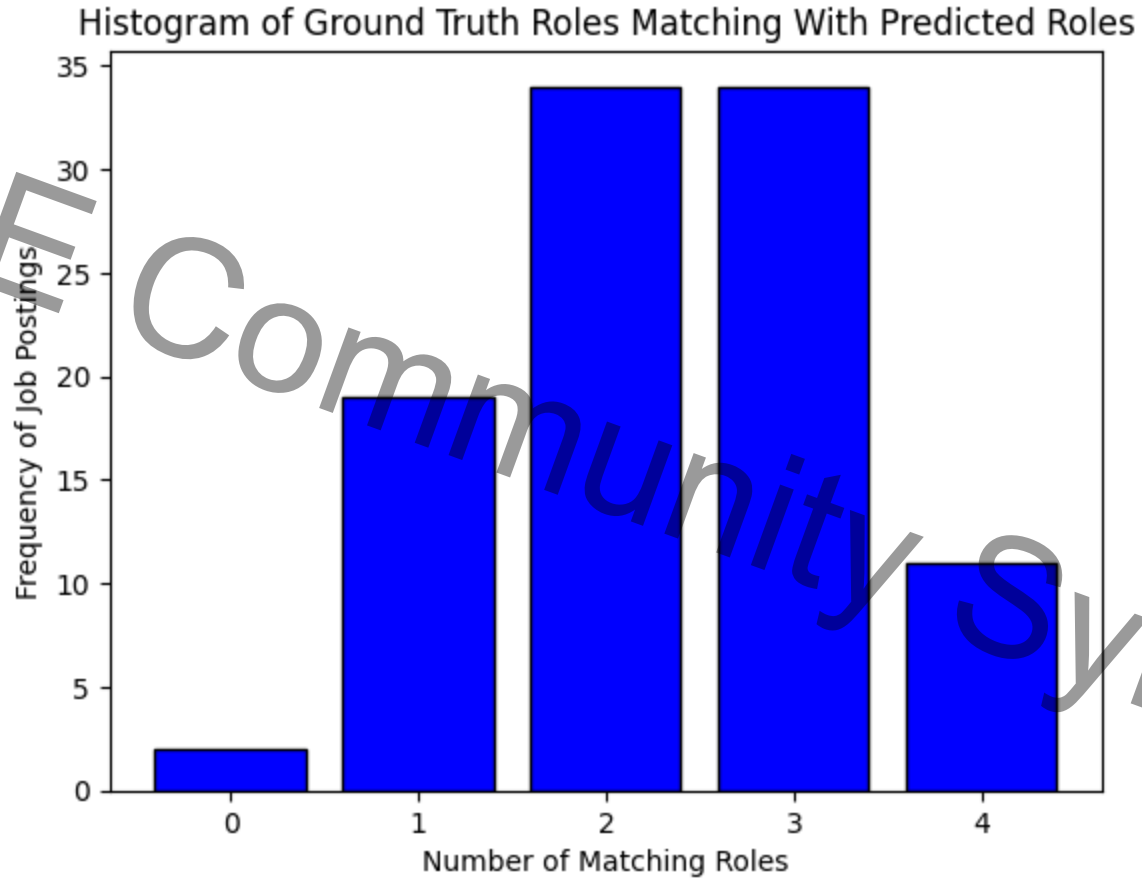


Fig. 7. Frequency of number of predicted role matching with the ground truth labels. As can be seen in the Figure, 2 to 3 Work Roles is most common number of matches due to it being unlikely that a job description would match to less than 2 or greater than 4.

Conclusion