

인공지능 기반 맞춤형 장애인 구인·구직 추천 연구*

임윤정** / 권지우*** / 송규원****

본 연구의 목적은 장애인 노동시장의 수급 불균형을 해소하기 위해 인공지능을 활용하여 장애인 노동시장의 수요자 맞춤형 구인·구직 정보를 추천하는 알고리즘을 개발하는 것이다. 이를 위하여 본 연구에서는 장애인 전문 채용 포털에서 다양한 채용공고 데이터를 수집하고, 데이터의 품질 측면에서 장애인 구직자의 접근성을 저해하는 요소를 분석하였다. 또한, 구직자의 장애 특성과 직무 특성 간의 정밀한 매칭을 위해, 거대 언어 모델(Large Language Model)을 활용하여 채용공고 내의 비정형 텍스트를 구조화하고, 명시되지 않은 은닉 정보(예: 적합 장애유형 및 정도)를 식별 및 추출하여 추천에 활용하였다. 구인·구직을 위한 추천 알고리즘으로 기계학습 및 딥러닝 기반의 단어 임베딩 모델인 TF-IDF(Term Frequency-Inverse Document Frequency)와 Sentence-BERT를 사용하여 구인·구직 데이터를 벡터화하고, 이들 간의 유사도 점수를 산출하여 최적의 추천 리스트를 도출하였다.

본 연구에서 수행한 실험의 평가를 위해, 상위 k개의 추천 리스트 중에서 채용공고의 직종과 구직자의 희망 직종 일치 여부를 비교하여 Top-k Accuracy와 MAP(Mean Average Precision)를 산출하였다. 또한, 실제 추천 결과를 직접 비교 및 분석하여 정성적으로 평가를 진행하였다. 그 결과, 전반적으로 Sentence-BERT의 결과가 TF-IDF의 결과보다 우수함을 확인하였고, 구직 추천의 경우 0.842의 Top-5 Accuracy를, 0.716의 MAP@5를 달성하였으며, 구인 추천의 경우 0.695의 Top-5 Accuracy를, 0.4의 MAP@5를 달성하였다. 정성 평가 결과에서도, 수요자의 특성에 맞는 적절한 구인·구직 추천이 이루어진 것을 확인할 수 있었다.

본 연구를 통해, 장애인 노동시장의 수요자 맞춤형 구인·구직 추천 알고리즘에 LLM의 적용 가능성을 확인하였다. 다양한 데이터의 확보 및 LLM의 한계 극복 등이 이루어진다면, 더 좋은 추천 결과를 기대해 볼 수 있을 것이다. 본 연구를 기반으로, 장애인 인력 고용기회가 확대되고, 체계적이고 능동적인 장애인 일자리의 제공이 이루어지기를 기대한다.

주제어 _ 거대 언어 모델, 자연어 처리, 장애인, 일자리 추천

* 본 연구는 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. RS-2024-00397607, 데이터 기반 장애인 데이터 탐색·활용 해결 기술 개발)

** 차세대융합기술연구원 연구원(공동 1저자)

*** 차세대융합기술연구원 연구원(공동 1저자)

**** 차세대융합기술연구원 책임연구원(교신저자)

AI-Based Personalized Job Recommendation System for People with Disabilities*

Im Yun-jeong** / Kwon Ji-woo*** / Song Gyu-won****

This study aims to develop an AI-driven algorithm that provides personalized job recommendations tailored to the labor market for people with disabilities, addressing existing imbalances in job supply and demand. The authors collected job postings from a specialized employment portal for individuals with disabilities and analyzed accessibility barriers. A large language model (LLM) was used to structure unstructured job posting text and extract implicit information, such as the type and degree of disability, to improve alignment between job seekers' needs and job requirements.

The recommendation algorithm employs TF-IDF and Sentence-BERT to vectorize job and applicant data, using similarity scores to generate optimized recommendations. The system's performance was evaluated using Top-k Accuracy and Mean Average Precision (MAP) by comparing the alignment between recommended job categories and job seekers' desired positions. Results show that Sentence-BERT outperformed TF-IDF, achieving a Top-5 Accuracy of 0.842 and MAP@5 of 0.716 for job search recommendations, and a Top-5 Accuracy of 0.695 and MAP@5 of 0.400 for job offers. Qualitative analysis further confirmed the relevance of recommendations to user needs.

This research demonstrates the potential of LLMs in creating tailored job recommendation systems for people with disabilities. With more data and continued improvements in LLMs, this approach could significantly expand employment opportunities for individuals with disabilities.

Key words _ Large Language Model, Natural Language Processing, Disabilities, Job Recommendation

* This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.RS-2024-00397607, Development of data-based technology to explore and utilize data for people with disabilities)

** Researcher, Advanced Institute of Convergence Technology(Co-First Author)

*** Researcher, Advanced Institute of Convergence Technology(Co-First Author)

**** Principal Researcher, Advanced Institute of Convergence Technology(Corresponding Author)

I. 서론

장애인 노동시장은 여전히 여러 가지 구조적 장애물과 사회적 편견으로 인해 불균형한 상황에 직면해 있다(이수련·양성원, 2023). 우리나라 통계에 따르면 전체 노동 인구 중 장애인의 고용률은 비장애인에 비해 현저히 낮으며, 이는 사회적 배제와 경제적 기회의 불균등을 반영한다(한국장애인고용공단, 2023). 장애인은 신체적 또는 인지적 제한으로 인해 다양한 직무에 접근하는 데 어려움을 겪으며, 이로 인해 적합한 일자리를 찾기 어렵다는 문제가 지속적으로 제기되고 있다. 또한, 고용 과정에서의 차별과 편견, 접근성 부족, 직장 내 지원 체계의 미비 등은 장애인의 고용 안정성을 저해하는 주요 요인으로 작용하고 있다(이수련·양성원 2023). 장애인 노동시장의 수급 불균형 문제에 관한 연구는 지속적인 관심을 받고 있으나, 그 해결을 위한 체계적인 접근은 부족한 실정이다.

인공지능을 활용한 일자리 추천 연구는 최근 몇 년간 급격한 발전을 이루며 다양한 접근법과 알고리즘이 제안되고 있다(Qin et al., 2023). 초기 연구들은 기계학습(Machine Learning)과 자연어 처리(Natural Language Processing) 등의 기술을 활용하여 구직자의 이력서, 기술, 선호도 등을 분석하고, 이를 바탕으로 최적의 일자리를 추천하는 시스템을 개발하였다(Guo et al., 2014; Alsaif et al., 2022; Mishra and Rathi, 2020; Patel and Vishwakarma 2020). 특히, 협업 필터링은 사용자 간의 유사성을 기반으로 추천을 수행하여 구현이 비교적 간단하고 실시간 추천에 유리한 장점을 보였으나, 신규 사용자나 아이템에 대한 콜드 스타트 문제를 겪는 한계가 있었다. 이러한 한계를 보완하기 위해 콘텐츠 기반 필터링(Content-Based Filtering)과의 결합을 시도한 하이브리드 모델이 제안되었으며, 이는 추천 정확도와 사용자 만족도를 동시에 향상시키는 데 기여하였다. 최근에는 딥러닝과 거대 언어 모델(Large Language Model, 이하 LLM)의 발전이 일자리 추천 시스템에 새로운 가능성을 열어주고 있다(Wu et al., 2024). BERT(Bidirectional Encoder Representations from Transformers)와 같은 사전 학습된 언어 모델을 활용하여 이력서와 직무 설명서 간의 의미적 유사성을 분석하는 연구가 활발히 진행되고 있으며, 이는 기존의 기계학습 알고리즘 보다 높은 정확도를 제공한다(Fang et al., 2023; Guan et al., 2024).

그러나 이러한 연구들은 주로 비장애인을 대상으로 한 일반 채용시장에 집중되어 있으며, 장애인 채용시장을 고려한 연구는 많지 않다. 장애인 노동시장은 신체적, 정신적 조건, 환경 적응력 등 다양한 요소가 복합적으로 작용하는 특수한 시장으로, 기존의 추천 알고리즘으로는 이를 충분히 반영하기 어렵다. 따라서, 장애인의 개인적 특성, 능력, 선호도 등을 분석하여 최적의 맞춤형 일자리를 제공하는 추천 알고리즘의 개발이 필요하다.

이에 본 연구는 직무수행 능력과 업무별 요구되는 신체 능력을 중심으로 맞춤형 구인·구직 추천 알고리즘을 구현하고자 한다. 본 연구에서 제안하는 시스템은 장애인을 대상으로 개인의 희망 직무와 장애 유형 및 정도를 고려하여 개인 맞춤형 일자리를 추천한다. 또한, 장애인을 채용하기 희망하는 기업에 대해 채용 직무에 적합한 맞춤형 인재를 추천하여 장애인 채용시장에 존재하는 수급 불균형 현상을 해소하고자 한다. 이 과정에서, 최신 인공지능 기술인 LLM을 기반으로 비정형 데이터의 구조화를 수행하고, 고도화된 추천을 위해 직접적으로 명시되지 않은 은닉 정보(Hidden Value)를 도출하여 활용하였다.

본 연구가 추구하는 목표는 크게 두 가지다. 첫째, LLM과 프롬프트 엔지니어링을 통한 채용공고에서의 비정형 텍스트 정형화 및 명시되지 않은 은닉 정보 추출 작업의 적용 가능성을 검토하고자 한다. 둘째, 인공지능을 활용하여 장애인 노동시장을 위한 수요자 맞춤형 구인·구직 정보를 자동으로 추천해주는 기술을 개발하고자 한다. 이를 통해 장애인들이 자신의 능력에 적합한 일자리를 효과적으로 찾을 수 있도록 지원한다. 이를 위하여, 장애인 노동시장의 현황과 문제점을 분석하고, 기존의 일반 노동시장을 대상으로 한 인공지능 기반 일자리 추천 연구들을 검토하였다. 그동안 장애인에 맞춘 일자리 추천 연구가 많지 않았다는 점에서 본 연구의 차별성을 찾을 수 있겠다. 본 연구를 통해 장애인 취업 시장에 존재하는 불균형을 해소하고, 장애인 인력 고용 활성화에 이바지할 수 있으리라 기대한다.

본 논문은 다음과 같이 구성된다. 2장에서는 장애인 고용시장에 대한 기존 문헌 연구를 검토하고, 인공지능 기술 중 특히 자연어 처리, 단어 임베딩, LLM, 구인·구직 추천 알고리즘에 관한 연구를 소개한다. 3장에서는 구인·구직 추천 알고리즘을 개발하기 위한 데이터 수집 방법, 데이터 분석 절차, 그리고 알고리즘 구현 방법 및 절차를 상세히 기술한다. 4장에서는 제안된 추천 알고리즘의 신뢰성과 유용성을 평가한다. 5장에서는 연구의 주요 결과와 한계점을 논의하며, 이를 바탕으로 향후 연구 방향에 대한 제언을 제시한다. 특히, 추천 알고리즘의 개선 가능성과 장애인 고용 증진을 위한 추가적인 기술적, 정책적 접근을 모색한다.

II. 관련 연구

1. 장애인 고용시장 문헌연구

한국장애인고용공단 자료에 따르면, 2023년 장애인 고용률은 전체 평균 3.17%, 공공부문 장애인 고용률은 3.86%로 의무고용률 3.6%를 상회했지만, 민간부문은 2.99%로 의무고용률 3.1%에 미치지 못

한다. 현재 장애인 구인·구직 서비스는 장애인 구직자와 구인 기업 간의 정보 비대칭으로 인해 장애인 고용시장의 수급 불균형 현상이 심화되고 있다. 2023년 장애인경제활동실태조사에 따르면, 장애인들은 취업 과정에서 ‘취업 알선(13.4%)’, ‘일자리 정보 제공(12.5%)’, ‘장애인 구분모집/특별채용(5.8%)’ 등과 같은 직접적인 지원을 필요로 한다. 유효한 취업 알선을 위해서는 구직자의 정보 뿐 아니라 기업의 다양한 정보 역시 필요하다. 그러나, 장애인들이 공공 취업 알선 기관을 통해 구직을 진행하는 비율은 상대적으로 낮으며, 주로 주변 지인을 통해 일자리 정보를 얻는 경향이 있다. 이는 공공기관의 취업 지원 서비스를 통해 구인 기업의 정보를 획득하는 데 어려움이 있음을 나타낸다. 장애 유형별 장애인 실업자의 희망 직업을 묻는 질문에서는, 전반적으로 단순노무 종사자라는 응답이 가장 많이 나왔으며 뒤를 이어 서비스, 기능, 관리자 직종에 대한 선호도가 높게 나타났다. 하지만 장애 유형별로 선호하는 직종의 순위는 상이했다. 장애인 중에서도 다양한 분야의 업무기술을 보유한 구직자들도 많다는 점을 고려한다면 기업들의 구인 정보가 적절한 시기에 장애인 구직자에게 충분히 제공되는 것이 필요하다.

한편, 2023 기업체장애인고용실태조사에서 장애인 고용기업체가 장애인 근로자 채용 시 가장 중요하게 고려하는 요소는 ‘장애 유형(23.4%)’으로 나타났으며, 그 다음으로는 ‘장애 정도(중증/경증)’, ‘의사소통 능력’, ‘이동 및 동작 수행 능력’이었다. 이는 기업들이 장애인의 개별적 특성과 역량을 중요하게 생각하고 있음을 시사한다. 그러나 현재의 구인·구직 플랫폼은 이러한 세부적인 정보를 효과적으로 반영하지 못하고 있어, 적합한 인재 매칭에 어려움을 겪고 있다.

실제로, 한국장애인고용공단 내부 시스템에서 관리되고 있는 취업 알선을 위한 구직자, 구인 기업의 정보의 지속적 확보 및 활용에 대한 애로사항이 있으며, 축적된 데이터 연계가 되지 않아 활용도가 적은 상황이다(전희주 외, 2021). 축적된 정보는 단순히 열람, 검색만 가능하며 이를 체계적으로 활용하고 분석하는 데는 어려움이 있다. 게다가 기업의 채용 정보는 이미지나 비정형 텍스트 형태로 제공되는 경우가 많아, 플랫폼에서 이러한 정보를 효과적으로 처리하기 어렵다. 기업마다 채용 공고의 형식과 표현이 다르며, 이는 정보 기술을 활용한 데이터 분석의 필요성을 부각시킨다.

더불어 적극적인 기업의 채용을 유도하기 위해 구인 직무에 최적화된 구직자를 추천하는 등의 맞춤형 채용 알선 서비스 구축 역시 필요하다고 할 수 있다. 구인 기업 측면에서 장애인 지원자가 없거나 업무능력을 갖춘 장애인 인력이 부족하다는 것은 구인 기업들의 구인 정보가 능력 있는 장애인들에게 적절하게 제공되지 않을 뿐 아니라 구인 기업들에게도 장애인 구직자 및 장애인 노동시장의 공급자 정보가 제대로 제공되지 않고 있음을 시사한다. 이는 장애인 일자리 추천 시스템의 구축 운영의 필요성을 보여준다.

2. 인공지능 기반 일자리 추천 기술

인공지능 기술은 크게 세 가지 범주, 즉 인공지능(Artificial Intelligence), 기계학습(Machine Learning), 딥러닝(Deep Learning)으로 분류될 수 있다. 첫 번째 범주인 인공지능은 인간의 학습, 추론, 지각 능력을 모방하여 컴퓨터 프로그램으로 구현한 기술로, 기계학습과 딥러닝을 포함하는 포괄적인 개념이다. 두 번째 범주인 기계학습은 컴퓨터에 명시적인 지침을 주지 않고도 데이터에서 패턴을 학습하고 예측할 수 있도록 하는 기술이다. 마지막으로, 딥러닝은 인간 두뇌의 작동 방식을 모델로 한 ‘인공신경망’을 사용하는 기계학습의 하위 집합이다. 기계학습의 다양한 작업을 수행할 수 있으며, 이미지 인식, 문서 번역, 음성 인식 등 다양한 고차원적인 작업을 수행할 수 있다.

1) 자연어 처리와 단어 임베딩 모델

자연어 처리는 인간의 언어를 컴퓨터가 처리할 수 있도록 해주는 총체적인 기술을 의미하며, 정보 검색 추천 시스템, 질의응답 시스템, 문서 요약 등의 다양한 분야에서 활용되고 있다(Young et al., 2018). 자연어를 처리하기 위해서는 사람의 언어를 컴퓨터가 이해하고 처리할 수 있도록 단어를 0과 1로 이루어진 형태로 바꾸어야 하는데, 이를 벡터화 혹은 단어 임베딩(Word Embedding)이라 한다. 비정형 텍스트나 문장 같은 비구조적 데이터의 경우, 단어를 벡터화하기 전에 먼저 단어 단위로 분할해야 한다. 이를 위해 형태소 분석기(Tokenizer)가 사용된다. 특히, 한국어 처리에서는 한국어 형태소 분석기가 필수적이다. 한국어는 형태소 기반 언어로, 어미와 조사가 어근에 결합해 단어가 형성되므로, 이를 분리하지 않으면 단어의 의미와 문맥을 정확하게 파악하기 어렵다. 예를 들어, “갔다”는 “가다”라는 동사에 시제 어미 “-ㅂ니다”가 붙은 형태로, 이를 분석하지 않고 단순하게 처리하면 동일한 의미의 다양한 형태를 구분할 수 없다. 이는 한국어의 복잡한 문법적 특성 때문에 필수적인 절차다.

단어 임베딩 방법은 크게 빈도 기반(Count-based), 예측 기반(Predictive-based), 그리고 사전 학습된 임베딩(Pretrained Embedding)로 구분된다(Almeida & Xexéo, 2019; Wang et al., 2020). 첫 번째, 빈도 기반 방법은 텍스트에서 단어가 얼마나 자주 등장하는지를 수치화하여 벡터로 표현하는 방식이다. 대표적으로 BoW(Bag of Words)와 TF-IDF(Term Frequency-Inverse Document Frequency)가 있다. BoW는 단순히 단어의 등장 횟수를 카운트하여 벡터를 생성하는 방법으로, 문맥을 고려하지 않는다는 한계가 있지만 구현이 간단하고 직관적으로 해석할 수 있다. TF-IDF는 BoW의 한계를 보완한 방식으로, 단어의 빈도 외에도 그 단어가 특정 문서에서 얼마나 중요한지를 반영한다. 두 번째, 예측 기반 방법은 딥러닝을 기반 단어가 주변 단어들과의 연관 관계를 학습하여 벡터를 생성하는 방식으로, Word2Vec, GloVe, FastText 등이 대표적이다. Word2Vec은 단어가 주어졌을 때 주

변 단어를 예측하는 방식(Continuous Bag of Words, CBOW) 또는 주변 단어가 주어졌을 때 중심 단어를 예측하는 방식(Skip-gram)을 통해 단어의 벡터를 학습한다. 세 번째, 사전 학습된 임베딩은 대규모 데이터에서 사전 학습된 언어 모델을 활용해 단어와 문장의 벡터를 생성하는 방법이다(Wang et al., 2020). 대표적인 모델로는 BERT, GPT, RoBERTa 등이 있다. 이러한 모델들은 언어의 문맥을 더 잘 이해할 수 있도록 설계되어 있으며, 문장의 복잡한 의미 관계나 다양한 언어적 특성을 벡터에 반영할 수 있다. BERT는 양방향으로 문맥을 고려해 단어를 벡터화하며, GPT는 순차적으로 단어를 예측하는 방식으로 문장을 처리한다. 사전 학습된 언어 모델은 텍스트 분류, 질의응답, 번역 등 다양한 자연어 처리 작업에서 우수한 성능을 보여주고 있다.

본 연구에서 주로 활용한 자연어 처리 기술은 구인·구직 텍스트 데이터를 형태소 단위로 나누기 위한 한국어 형태소 분석기와 단어 수준 또는 문장 간의 유사도를 비교하기 위해 TF-IDF, Sentence-BERT 기반 임베딩 모델을 사용하였다.

(1) TF-IDF 모델

TF-IDF(Term Frequency-Inverse Document Frequency)는 문서에서 단어의 중요도를 평가하는 통계적 모델로, 주로 키워드 추출, 검색 결과의 순위 결정, 문서 간 유사도 측정 등에 활용된다(Tata & Patel, 2007). TF-IDF 값이 클수록 해당 단어가 문서 내에서 더 중요한 의미를 갖는다고 간주된다. TF(Term Frequency)는 특정 문서에서 특정 단어 t 의 출현 빈도를 나타낸다. 예를 들어, 문서 1에서 “사무”라는 단어가 5번 출현하면, 그 문서에서 “사무”의 TF는 5가 된다. 특정 문서에서 TF 값이 클수록 해당 단어가 상대적으로 중요한 단어로 간주된다. 식 (1)에서 $f_{t,d}$ 는 단어 t 가 문서 d 에서 출현한 횟수, $|d|$ 는 문서 d 의 전체 단어 수를 의미한다.

$$TF(t, d) = \frac{f_{t,d}}{|d|} \quad (1)$$

DF(Document Frequency)는 특정 단어가 출현한 문서의 수를 의미한다. 예를 들어, “미용”이라는 단어가 10개의 문서 중 3개의 문서에서 언급되었다면, “미용”의 DF는 3이다. IDF(Inverse Document Frequency)는 $DF(t)$ 에 반비례하는 수를 의미한다. 특정 단어가 여러 문서 집합 내에서 자주 사용되는 경우, 이는 그 단어가 흔하게 등장한다는 것을 의미한다. 흔히 사용되는 단어일수록 낮은 IDF 값을 가지고, 드물게 사용되는 단어일수록 높은 IDF 값을 가진다. 예를 들면, 여러 문서에서 “은”, “는”, “이”, “가”와 같은 조사가 많이 등장하지만, 그 문서를 대표하는 중요한 단어는 아니다. 반면, “기계학습”이라

는 용어는 문서 집합에서 자주 등장하지 않고, 그 문서의 핵심인 경우가 많다. 식 (2)에서 N 은 문서 집합 수, $|d \in D: t \in d|$ 는 단어 t 가 출현한 문서 수를 의미한다. TF-IDF는 각 단어의 TF와 IDF를 곱하여 계산하며, 식 (3)과 같다.

$$IDF(t, D) = \log \left(\frac{N}{|d \in D: t \in d|} \right) = \log \left(\frac{N}{1 + df(t)} \right) \quad (2)$$

$$TF-IDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (3)$$

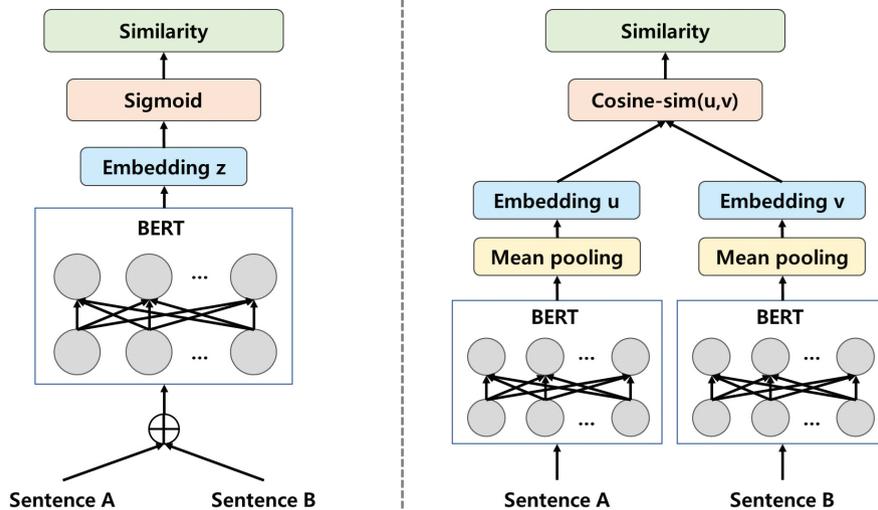
(2) Sentence-BERT 모델

BERT(Bidirectional encoder representations from transformers)는 Google에서 개발한 사전 학습된 언어 모델이다(Devlin et al., 2018). Transformer 모델의 Encoder 부분만 사용하여 구성되었으며, 양방향에서 문맥을 이해할 수 있도록 설계되었다. 이전의 단방향 언어 모델과 달리 문장의 양쪽에서 오는 정보를 모두 활용하여 단어의 의미를 더 정확하게 포착할 수 있다. 최근에는 다양한 변형된 BERT 구조들이 많이 등장하여 특정 작업(task)에 맞게 개선되고 있다(Liu, 2019; Sun et al., 2022).

Sentence-BERT(이하 SBERT)는 이러한 BERT 모델 위에 Pooling layer를 쌓아 올린 구조이며, 두 문장의 의미적 유사성을 측정하는 Semantic Textual Similarity(STS) 및 Natural Language Inference(NLI) 데이터로 학습된 문장 임베딩 모델이다(Reimers et al., 2019). SBERT는 두 문장을 각각 임베딩한 후, 이들의 코사인 유사도 점수를 계산하여 의미적 유사성을 측정하는 데 사용된다. 문장의 의미를 보다 효과적으로 임베딩에 반영하며, 문장 간의 유사도를 비교하는 작업에서 탁월한 성능을 보이고 있다.

Sentence Transformers는 SBERT의 다양한 모델에 쉽게 접근하고 사용할 수 있도록 UKPLab에서 개발한 Python 라이브러리로, Hugging Face에서 관리되고 있다. 현재 Hugging Face에는 5,000개 이상의 사전 훈련된 Sentence Transformers 모델들이 배포되어 있으며, 이 중에는 Massive Text Embeddings Benchmark(MTEB) 리더보드에서 최첨단(State-of-the-art) 성능을 기록한 모델이 다수 포함되어 있다. 이러한 모델을 기반으로 응용 시스템 사례에 맞는 맞춤형 모델을 손쉽게 훈련하거나 미세 조정할 수 있어, 특정 애플리케이션에 최적화된 임베딩 모델을 만들 수 있다.

〈그림 1〉 BERT(좌측) 및 Sentence-BERT(우측) 구조 (Han et al., 2023)



2) LLM

LLM은 대규모의 텍스트 데이터를 기반으로 사전 학습된 딥러닝 모델로, 수십억 개 이상의 파라미터(parameter)를 포함하고 있다. LLM은 자연어 이해와 생성에 뛰어난 성능을 발휘하며, 텍스트 생성, 질의응답, 감정 분석, 번역 등 다양한 자연어 처리 작업에 광범위하게 활용되고 있다.

LLM에서 프롬프트(prompt)는 사용자가 제공한 입력 또는 질의로, 모델이 응답을 생성하는 시작점 역할을 한다. 프롬프트는 단순한 단어나 구절, 간단한 질문에서부터 복잡하고 상세한 지침에 이르기까지 다양한 형태와 수준의 복잡성을 가질 수 있다(Wu et al., 2024). 프롬프트 엔지니어링(Prompt Engineering)은 LLM의 출력 품질을 향상시키기 위해 입력 방식을 전략적으로 설계하고 최적화하는 과정을 의미한다. 이 과정은 모델이 특정 작업을 수행하도록 유도하기 위해 입력 문구를 구성하고 조정하는 것을 포함한다. 특히, 작업에 대한 응답 예시를 여러 개 제공하거나 복잡한 작업에 대해 단계적으로 추천하도록 지시하는 명령문을 활용함으로써 작업 성능을 크게 향상시킬 수 있다(Wu et al., 2024).

본 연구의 목적은 LLM을 사용하여 장애인을 대상으로 한 채용공고에 일반적으로 포함되지 않는 은닉 정보(예: 장애 유형, 장애 정도)를 추출하여, 보다 효율적인 구인·구직 매칭을 실현하는 데 있다. 이에 본 연구에서는 GPT-4o-mini의 강력한 자연어 처리 능력을 이용하여 채용공고 텍스트에서 적합한 장애 특성을 효과적으로 식별하고 추출할 수 있는 텍스트 프롬프트 구조를 설계한다. 또한, 추출한 은닉 정보를 실제 구인·구직 추천에 적용함으로써 설계된 프롬프트의 실효성을 검증하고, 이를 통해 장애인에게 적합한 직무 매칭을 지원하는 실질적인 방안을 제시하고자 한다.

3) 구인·구직 추천 알고리즘

일자리 추천 시스템은 구직자의 특성과 선호도를 분석하여 적합한 일자리를 제안함으로써 구인·구직 과정을 효율화하는데 도움을 주는 시스템을 의미한다(De et al., 2021; Patil et al., 2023; Mashayekhi et al., 2024). 일자리 추천 시스템은 온라인에 남겨진 구직자의 명시적(explicit) 또는 암묵적(implicit) 피드백을 기반으로 구직자의 선호도를 분석 및 예측하며, 이를 통해 가장 적합한 직무, 기업, 채용공고 등을 제공할 수 있다(Patil et al., 2023). 명시적 피드백은 구직자가 직접 입력한 이력서 정보나 자기소개서, 만족도 등을 의미하며, 암묵적 피드백은 구직자의 검색 기록, 클릭 행위, 지원 이력과 같은 행동 분석을 통해 시스템에서 추론 가능한 정보를 의미한다.

일자리 추천 알고리즘은 콘텐츠 기반 방식(Content-Based Filtering), 협업 필터링 방식(Collaborative Filtering), 하이브리드(hybrid) 방식의 세 가지가 있다(손지은 외, 2015). 콘텐츠 기반 접근 방식은 직무의 속성을 기반으로 분류하며, 사용자의 과거 선택 또는 지원 이력에 포함된 직무와 추천 후보 직무 간의 유사성을 계산하여 유사도가 높은 직무를 추천한다(Guo et al., 2014; Alsaif et al., 2022). 협업 필터링 방식은 비슷한 경력이나 관심사를 가진 다른 구직자들의 행동과 의견을 기반으로 추천하는 기술이다(Mishra and Rathi, 2020; Patel and Vishwakarma, 2020). 이는 특정 직무에 대한 선호도가 유사한 구직자 집단이 존재하며, 이들이 선호하는 다른 직무도 비슷한 선호를 보일 것이라는 가정에 기반한다. 예를 들어, A 구직자가 선호한 직무와 유사한 직무를 선호한 다른 구직자들이 높게 평가한 직무를 A에게 추천하는 방식이다. 마지막으로 하이브리드 방식은 다양한 추천 알고리즘의 단점을 보완하고 성능을 향상시키기 위해 두 가지 이상의 추천 기법을 결합하는 방식이다(Saha, 2023). 예를 들어, 콘텐츠 기반 방식과 협업 필터링 방식을 통합하여 구직자의 개별 선호도와 사용자 간의 유사성을 동시에 반영한 추천을 제공할 수 있다. 하이브리드 방식은 추천의 정확성과 다양성을 동시에 높일 수 있으며, 특정 상황이나 도메인에 최적화된 맞춤형 추천 시스템을 구현할 수 있다.

일자리 추천 연구는 주로 기계학습 및 딥러닝 기술을 중심으로 발전해 왔다. Gupta와 Garg(2014)에서는 구직자의 직무 선호도와 과거의 직업 행동 패턴을 바탕으로, 구직자의 프로필과 일치하는 일자리를 추천하는 규칙 기반 데이터 마이닝 기법을 제안하였다. Martinez-Gil et al.(2018)은 기계학습 알고리즘을 사용하여 구직자와 구인 공고 간의 매칭 정확도를 높이는 방법을 탐구하였다. Chen et al.(2018)에서는 구인 공고와 구직자 프로필을 바탕으로, 빅 데이터를 지원하는 트리 기반 학습 모델을 사용해 개인화된 구인 또는 구직자 추천 시스템을 제안한다. 온라인에서 구직자의 행동을 실시간으로 추적하고, 이를 바탕으로 맞춤형 추천을 제공하는 방법을 제시하였다. Saha et al.(2023)에서는 콜드 스타트(Cold Start)와 데이터 희소성 문제를 해결하기 위해, 구직자와 구인자의 정보를 활용한 콘텐츠 기반과 협업 필터링 데이터를 결합한 하이브리드 추천 시스템을 제안하였다. Wang et al.(2016)에서는

양방향 선호도를 기반으로 한 상호 추천 모델을 제안하였다. 이 모델은 구직자와 구인자 간의 상호 선호도를 고려하여 추천을 수행하며, 이를 통해 보다 정교한 매칭을 가능하게 한다.

또한, 기술 추출(skill extraction)에 대한 연구도 중요한 역할을 하고 있다. 기술 추출은 이력서나 구인 공고와 같은 비정형 텍스트에서 구직자나 직무의 요구되는 핵심 기술을 자동으로 식별하는 것으로, 구직자와 일자리 간의 매칭 정확도를 향상시킬 수 있다. Gugnani와 Misra(2020), Khaouja et al.(2021), Fang et al.(2023), 그리고 Guan et al.(2024) 등의 연구에서는 구직자의 이력서나 채용 공고에 나타난 키워드와 스킬을 자동으로 추출하는 방법을 제안하여, 직무와 구직자 간의 스킬 매칭을 개선하는 방안을 모색하였다. Gugnani와 Mishra(2020)는 구직자의 스킬을 자동으로 추출하는 기술을 발전시키기 위해 자연어 처리(NLP) 기법을 도입하여, 이력서 내 명시된 스킬과 채용 공고의 요구 스킬을 비교하는 시스템을 제안했다. 이 연구는 키워드 기반의 스킬 매칭을 통해 일자리와 구직자 간의 스킬 불일치를 최소화하는 데 중점을 둔다. Khaouja et al.(2021)은 채용 과정에서 기술 격차를 줄이기 위해 Word2Vec과 같은 임베딩 기법을 활용한 추천 모델을 제안했다. 이 모델은 구직자의 이력서를 벡터화하여, 비슷한 스킬을 가진 다른 지원자들과 비교하고, 이 정보를 바탕으로 적합한 일자리를 추천한다. 이러한 연구는 정확한 스킬 정보가 직무 추천 시스템의 성능을 크게 향상시킬 수 있음을 보여준다.

최근 연구에서는 LLM을 활용한 접근도 활발하다. Guan et al.(2024)은 GPT-3와 같은 대규모 언어 모델을 사용해, 구인·구직 문서의 의미를 심층적으로 이해하고 분석하는 방법을 제안했다. 이러한 연구는 단순한 키워드 매칭을 넘어, 구직자와 일자리 간의 문맥적 유사성을 고려하여 보다 정교한 추천을 가능하게 했다.

본 연구에서 장애 유형 맞춤형 구인·구직 추천은 콘텐츠 기반 필터링 방식을 채택한다. 장애인을 대상으로 한 일반적인 채용공고에는 특정 장애 유형이나 장애 정도와 같은 세부 정보가 명시되지 않는 경우가 많아, 구직자의 장애특성과 직무특성 간의 적합한 매칭이 어려운 실정이다. 이러한 문제를 해결하고자, LLM을 활용하여 채용공고에서 중요한 은닉 정보(예: 적합/부적합 장애 유형, 장애 정도)를 효과적으로 식별하고 추출할 수 있는 프롬프트 구조를 설계하였다. 추출된 은닉 정보(Hidden Value)는 실제 구직자와 구인자 간의 보다 정밀한 매칭을 가능하게 한다. 추천 알고리즘은 TF-IDF와 SBERT를 사용하여 구직자와 구인자의 텍스트 데이터를 벡터화하고, 이들 간의 유사도 점수를 산출하여 최적의 추천을 제공한다.

Ⅲ. 연구 방법

1. 연구 목표 및 대상

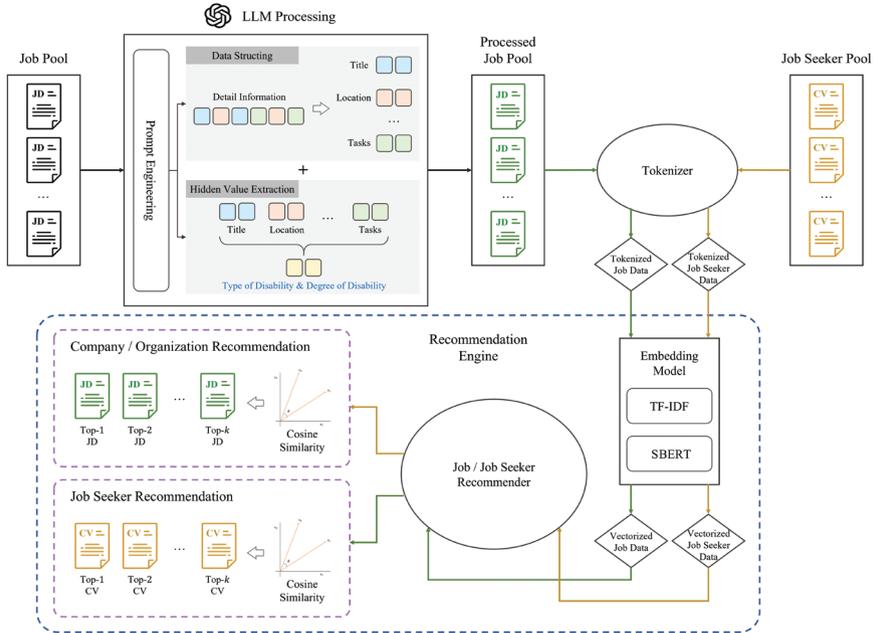
본 연구에서는 직무수행 능력과 업무별 요구되는 신체 능력을 중심으로, 장애인 고용 활성화를 위한 맞춤형 구인·구직 추천 알고리즘을 구현하는 것을 목표로 한다. 본 연구의 추천 알고리즘은 장애인의 개인적 특성과 직무 요구사항을 정밀하게 매칭하여, 개인의 신체적·인지적 특성에 맞는 직무를 추천하는 동시에, 기업에게도 직무에 적합한 인재를 추천하는 기능을 포함한다.

본 연구에서는 장애인 구직 희망자의 실제 프로필 정보와 장애인을 대상으로 하는 채용공고를 직접 수집하여 활용한다. 또한, LLM을 기반으로 수집된 비정형 데이터를 구조화하고, 채용공고에 직접적으로 명시되지 않은 은닉 정보를 추출하여 추천에 사용한다. 여기서 은닉 정보란, 일반적으로 채용공고에 명시되지 않지만, 해당 직무를 수행하기에 적합 또는 부적합한 장애 유형 및 정도와 같은 중요한 정보를 의미한다. 이러한 요소들은 장애인의 직업 만족도와 생산성에 영향을 미칠 수 있으므로, 이러한 정보들이 정확하게 고려된 맞춤형 추천이 필요하다.

이를 위하여, LLM을 기반으로 채용공고에 명시된 업무 내용 및 작업환경을 분석하기 위한 프롬프트 엔지니어링을 수행하였다. 이 과정에서 채용공고의 비정형 데이터를 효과적으로 구조화하기 위해, 직무 요구사항과 작업환경에 대한 핵심 정보를 추출할 수 있도록 설계된 맞춤형 프롬프트를 사용하였다. 최종적으로, 정형화된 구인·구직 데이터 간의 유사도 점수를 산출하여 상위 k개의 추천 리스트를 제공한다.

〈그림 2〉는 본 연구에서 제시하는 시스템의 구조를 도식화하여 나타낸 것이다.

〈그림 2〉 장애인 맞춤형 구인·구직 추천 알고리즘 구조도



2. 데이터 수집 및 분석

본 연구에서 활용한 데이터는 장애인 전문 채용 포털에 등록된 구직자의 프로필과 채용공고이다. 구직자 프로필은 해당 포털로부터 제공받았으며, 채용공고는 직접 수집하였다.

개인정보보호법(법률 제19234호)에 따라 구직자 프로필은 성명, 나이, 성별 등의 민감한 개인정보를 모두 제외한 상태로 제공받았다. 세부 항목은 '나이대', '장애 유형', '장애 정도', '최종 학력', '전공', '자격증/어학', '희망 직종', '희망 직무', '희망 근무지' 이다. 41개의 구직자 프로필 중 '장애 유형', '장애 정도', '희망 직무' 항목 결측치가 있는 데이터를 처리하여, 최종적으로 총 38개의 구직자 프로필을 활용하였다. <표 1>는 장애 유형과 정도에 따른 구직자의 분포를 나타낸다. 분석 결과, 지체장애의 비율이 가장 높고, 뇌병변장애와 청각장애 순으로 나타났다. 장애 정도가 있는 장애 유형만 봤을 때, 중증은 16명, 경증은 15명으로 비슷한 비율로 분포되어 있었다.

〈표 1〉 ‘장애 유형’ 및 ‘장애 정도’에 따른 사용자 데이터 분포

지체		청각		시각		뇌병변		뇌전증	내부	지적	정신	계
경증	중증	경증	중증	경증	중증	경증	중증	경증	중증	-	-	
5	5	3	4	2	1	4	5	1	1	3	5	39*

* 전체 38명 중 1명은 지적장애와 뇌병변장애(중증)를 동시에 수반하고 있어 중복 집계함

‘희망 직종’에 따른 구직자의 분포는 〈표 2〉에서 확인할 수 있다. ‘희망 직종’은 장애인 전문 채용 포털에서 제시하는 직종 분류 기준에 따라, 사용자가 직접 선택한 항목을 기반으로 분류되었다.

〈표 2〉 ‘희망 직종’에 따른 구직자 데이터 분포

희망 직종	인원
경영·사무·금융·보험	22
연구 및 공학기술	2
교육·법률·사회복지·경찰·소방 및 군인	2
보건·의료	1
예술·디자인·방송·스포츠	4
미용·여행·숙박·음식·경비·돌봄·청소	2
영업·판매·운전·운송	1
건설·채굴	0
설치·정비·생산·기계·금속·재료	0
설치·정비·생산·전기·전자·정보통신	2
설치·정비·생산·화학·환경·섬유·의복·식품·가공	0
설치·정비·생산·인쇄·목재·공예 및 제조 단순	2
농림어업직	0
총 계	38

데이터의 분포를 살펴보면, ‘경영·사무·금융·보험’ 관련 직종에 종사하기를 희망하는 구직자들이 과반 수 이상을 차지한다. 반면, ‘건설·채굴’, ‘설치·정비·생산·기계·금속·재료’, ‘설치·정비·생산·화학·환경·섬유·의복·식품·가공’, ‘농림어업직’의 경우는 구직 수요가 낮은 것으로 분석된다. 이는 타 직종에 비해 비교적 높은 신체 능력을 요구하거나, 비교적 높은 위험도를 가진 직종이기 때문인 것으로 보인다. 그러나, 적은 표본으로 산출된 결과이므로 특정 직종에 대한 구직 수요가 과소평가 되었을 가능성도 존재한다.

채용공고 데이터는 구직자 프로필과 마찬가지로 장애인 전문 채용 포털에서 수집하였다. 2024년 7월 9일부터 9월 8일까지 약 두 달간 해당 채용 포털에 게시되는 모든 채용공고 정보를 Selenium을 기

반으로 크롤링하여 수집하였다. 수집한 채용공고 데이터는 텍스트 또는 이미지 형태로 존재로 존재한다. 이미지 형태의 채용공고에 많은 텍스트가 포함될 경우, 이미지에서 텍스트를 추출하는 광학 문자 인식(Optimal Character Recognition, OCR) 기술과 거대 멀티모달 모델(Large Multimodal Model, LMM)을 활용하여도 정확한 내용 추출에 한계가 있음을 확인하였다. 특히, 이러한 한계는 디지털 정보 접근이 제한적인 장애인들이 채용공고 내용을 확인하기 어려울 수 있기 때문에, 향후 디지털 정보 접근성을 높일 수 있는 방향으로 개선되어야 할 것이다.

본 연구에서는 텍스트 식별 및 추출이 어려운 이미지 데이터를 제외하고, 텍스트만으로 구성된 채용공고 630개를 활용하였다. 채용공고 데이터는 ‘직종’, ‘제목’, ‘경력 유무’, ‘학력’, ‘성별’, ‘연령’, ‘전공’, ‘장애인 채용구분’, ‘우대조건’, ‘고용형태’, ‘직무’, ‘담당업무’, ‘직급’, ‘직책’, ‘근무지역’, ‘근무시간/요일’, ‘근무형태’, ‘복리후생’, ‘근무기간’, ‘급여조건’, ‘계약기간’, ‘작업환경’, ‘장애인 편의시설’, ‘담당자명’, ‘연락처’, ‘이메일’, ‘접수기간’, ‘접수방법’, ‘전형일정’, ‘홈페이지’, ‘기업정보’, ‘세부 정보’를 포함하고 있다. ‘세부 정보’는 구인 기업에서 별도의 양식 없이 자유롭게 작성하는 항목이다. ‘직종’의 경우, 본 연구를 위해 장애인 전문 채용 포털에서 제시하는 직종 분류에 따라 ‘제목’, ‘직무’, ‘담당업무’를 기반으로 직접 분류한 항목이며, ‘개인 능력에 따라 추후 배치’와 같이 명확한 업무 내용의 명시가 없는 경우 ‘기타’로 분류하였다. ‘직종’에 따른 채용공고 데이터의 분포는 아래 <표 3>와 같다.

<표 3> ‘직종’에 따른 채용공고 데이터 분포

직종	개수
경영·사무금융보험	185
연구 및 공학기술	1
교육·법률·사회복지·경찰·소방 및 군인	66
보건·의료	23
예술·디자인·방송·스포츠	24
미용·여행·숙박·음식·경비·돌봄·청소	195
영업·판매·운전·운송	28
건설·채굴	1
설치·정비·생산·기계·금속·재료	6
설치·정비·생산·전기·전자·정보통신	25
설치·정비·생산·화학·환경·섬유·의복·식품·가공	21
설치·정비·생산·인쇄·목재·공예 및 제조 단순	40
농림어업직	0
기타	15
계	630

본 연구에서 수집된 채용공고 데이터에 따르면, 대부분 장애인 채용공고는 ‘경영·사무·금융·보험’ 또는 ‘미용·여행·숙박·음식·경비·돌봄·청소’ 직종에 속한다. ‘경영·사무·금융·보험’ 관련 직무들 역시 전문성을 요구하는 업무보다는, 단순 문서 정리, 데이터 입력, 유선 혹은 온라인 고객 상담 업무가 주를 이루고 있다. 가장 높은 비율을 차지하고 있는 ‘미용·여행·숙박·음식·경비·돌봄·청소’ 직종의 경우, 대부분의 업무는 청소 및 빨래를 포함한 환경 미화, 경비 등이다. ‘미용·여행·숙박·음식·경비·돌봄·청소’ 직종은 가장 높은 비율로 장애인 채용이 이루어지고 있으나, 실제 구직자 데이터 분포에서 나타난 희망 직종 통계와 비교했을 때, 해당 직종의 선호도 높지 않아 기업 입장에서 채용에 어려움을 겪을 것으로 보인다.

3. LLM 프롬프트 엔지니어링(LLM Prompt Engineering)

본 연구에서는 수집한 채용공고 데이터를 정형화하고 은닉 정보를 추출하기 위해 LLM의 프롬프트 엔지니어링 기법을 활용하여 가공하는 작업을 수행한다.

1) 비정형 데이터 구조화

채용공고 양식은 기업마다 서로 상이하다. 특히, 일부 공고에서는 모든 정보를 ‘세부 정보’ 항목에 몰아서 기재하는 경우도 있어, 직무 요구사항, 자격 조건, 근무 환경 등의 중요한 정보들이 명확하게 구분되지 않는다. 이러한 경우, 분석 과정에서 필요한 정보를 정확하게 추출하는 데 어려움이 있다.

따라서, 본 연구는 GPT-4o-mini를 기반으로 ‘세부 정보’에 나열된 텍스트를 사전에 선별한 데이터 항목에 따라 구조화하는 방법을 제시한다. 이는 GPT-4o-mini의 프롬프트 엔지니어링을 통해 ‘세부 정보’의 내용을 분석하고, 해당 내용을 ‘세부 정보’ 외의 항목들에 적절히 배치 및 보완하는 방식으로 진행하였다.

2) 공고 내 은닉 정보 추출

장애인 전문 채용 포털을 통해 전국의 장애인 채용 정보가 통합되어 제공되고 있으나, 직무별 특수한 채용 조건 등으로 인해 해당 채용공고의 내용이 장애인 본인에게 적합한 직무인지 판단하는 데에는 어려움이 있을 수 있다. 일반적으로 장애 정도에 대한 제한 및 우대사항은 있을 수 있으나, 채용 중인 직무에 적합한 장애 유형이나 업무상 요구되는 신체 조건 등은 구체적으로 명시되어 있지 않아 공고문에 대한 면밀한 분석 없이는 개인의 적합 여부를 파악하기 어렵다.

따라서, 본 연구에서는 LLM을 활용하여 채용공고에 직접적으로 명시되어 있지 않은 은닉 정보를 추출하고, 이를 추천 알고리즘에 활용하도록 하였다. GPT-4o-mini의 프롬프트 엔지니어링을 수행하여

채용공고에 명시된 업무 내용 및 환경을 분석하고, 이를 통해 명시되지 않은 ‘적합 장애 유형’ 정보를 얻을 수 있었다. 데이터 구조화 과정과 은닉 정보 추출 과정을 동시에 진행하기 위해, 프롬프트를 작성하여 데이터 가공을 수행하였고, 입력에 따른 결과는 아래 <그림 3>과 같다. 이때, ‘적합 장애 유형’ 정보를 추출하기 위해 장애 유형 목록을 제시하고 해당 목록에서만 선택하도록 제한했는데, 이는 장애인 복지법 시행규칙(보건복지부령 제1050호)에서 규정하는 장애 유형 및 장애 정도를 적용한 것이다. 적절한 프롬프팅으로 데이터의 구조화와 은닉 정보 추출을 동시에 수행 가능하였고, 원하는 형태의 데이터로 가공이 가능함을 확인하였다.

<그림 3> 입력 프롬프트와 프롬프트 엔지니어링 결과 예시

<pre><system> 나는 지금부터 장애인 직업 추천 전문가야. 내가 보여주는 채용공고를 읽어보고 요구에 맞게 정리해줘. <user> 채용공고의 내용을 꼼꼼히 읽어보고 아래 포맷에 맞게 json 형태로 입력해줘. data_format = { "제목": "", "회사명": "", "기업정보": "", "직공": "", "장애인 채용구분": "", "우대조건": "", "직무": "", "담당업무": "", "직급": "", "직책": "", "근무지역": "", "근무시간/요일": "", "복지후생": "", "급여조건": "", "작업환경": "", "장애인 편의시설": "", "적합 장애유형": "", ... } "적합 장애유형" 항목의 경우, 담당업무, 우대조건, 장애인 편의시설 등 공고 내용을 바탕으로 해당 업무에 적합하거나 부적합한 장애 유형을 아래 목록에서 1개 이상씩 골라서 내용을 입력해줘. ["지체장애(경증)", "지체장애(중증)", "뇌병변장애(경증)", "뇌병변장애(중증)", "시각장애(경증)", "시각장애(중증)", "청각장애(경증)", "청각장애(중증)", "언어장애(경증)", "언어장애(중증)", "안면장애(경증)", "안면장애(중증)", "내부장애(경증)", "내부장애(중증)", "보청장애(경증)", "보청장애(중증)", "지적장애", "자폐성장애", "정신장애"]</pre>	<pre><input> { 'title': '[주] ○○○○ 공장 내 세탁물 수거 및 세탁, 정리 등 직원 채용', 'compName': '(주) ○○○○', 'compDetail': { 'industry': '', 'empCnt': '○○명', 'CEO': '', 'compType': '중소기업', 'address': '경기 ○○시 ○○○로 123' }, 'reqCareer': '경력무관', 'reqEduc': '학력무관', 'disabHireType': '장애정도 무관', 'prefCondition': '장애인, 취업보호대상자', 'empType': '정규직', 'jobGrade': '사원급', 'jobPosition': '팀원', 'location': '서울 ○○구', 'workDays': '08:30 ~ 16:30Wn주6일', 'benefits': '연차, 식사제공, 4대보험 가입', 'salary': '월급Wn(215만원~215만원)', 'contractPeriod': '', 'disabledFacilities': '', 'infoDetail': '[주] ○○○○ 공장 내 세탁물 수거 및 세탁, 정리 등 직원 채용Wn(주) ○○○○는 2015년에 설립된 회사로 자본금 ○억원, 사원수 ○○○명 규모의 중소기업입니다. 경기 ○○시 ○○○로 123에 위치하고 있습니다. Wn모집분야 및 자격요건Wn세탁원 (1명)Wn담당업무Wn· 세탁물 수거 및 세탁Wn· 정리 및 배송지원 등 업무 수행Wn 자격요건Wn· 학력: 학력무관Wn· 경력: 경력무관 Wn 우대사항Wn· 장애인 우선 채용Wn 전형 절차Wn· 서류전형 > 1차면접 > 최종합격Wn· 면접일정은 추후 통보됩니다.Wn유의사항Wn· 허위사실이 발견될 경우 채용이 취소될 수 있습니다.Wn', 'HRName': '김○○', 'HRNo': '', 'hireEmail': 'example@email.com', 'registDate': '2024.07.04 ~ 2024.09.04', 'registMethod': '장애인채용포털 접수, 이메일', 'homepage': 'https://www.example.co.kr/recruit/homepage.com' }</pre>	<pre><output> { "제목": "[주] ○○○○ 공장 내 세탁물 수거 및 세탁, 정리 등 직원 채용", "회사명": "(주) ○○○○", "기업정보": { "업종": "", "사원수": "○○명", "대표자": "", "기업 형태": "중소기업", "주소": "경기 ○○시 ○○○로 123" }, "경력 유무": "경력무관", "학력": "학력무관", "성별": "", "연령": "", "전공": "", "장애인 채용구분": "장애정도 무관", "우대조건": "장애인, 취업보호대상자", "고용형태": "정규직", "직무": "세탁원", "담당업무": "세탁물 수거, 세탁, 정리 및 배송지원 등", "직급": "사원급", "직책": "팀원", "근무지역": "서울 ○○구", "근무시간/요일": "08:30 ~ 16:30 주6일", "근로형태": "", "복지후생": "연차, 식사제공, 4대보험 가입", "근무기간": "", "급여조건": "월급 (215만원~215만원)", "계약기간": "", "작업환경": "", "장애인 편의시설": "", "담당자명": "김○○", "연락처": "", "이메일": "example@email.com", "접수기간": "2024.07.04 ~ 2024.09.04", "접수방법": "장애인채용포털 접수, 이메일", "전형일정": "서류전형 > 1차면접 > 최종합격", "홈페이지": "https://www.example.co.kr/recruit/homepage.com", "적합 장애유형": ["지체장애(경증)", "청각장애(경증)", "부적합 장애유형": ["언어장애(중증)", "정신장애"] }</pre>
--	--	--

4. 구인구직 추천 알고리즘

본 연구의 구인·구직 추천은 구직자 프로파일과 채용공고 텍스트를 기반으로 가장 유사한 직무와 인력을 추천한다. 자연어 처리 분야에서 일반적으로 사용되는 임베딩 방법으로는 TF-IDF, Word2Vec, SBERT 등이 있다. 이 중 Word2Vec의 경우 학습 데이터의 크기에 민감하게 영향을 받는 경향이 있으며(Mikolov et al., 2013; Mitra et al., 2017), 본 연구에서 활용하는 장애인 구인·구직 데이터는

그 특수성으로 인해 데이터의 수집이 용이하지 않고, 제한적이다. 따라서, 본 연구에서는 TF-IDF와 SBERT를 기반으로 한 추천 알고리즘을 활용하고자 한다. 각 구직자와 채용공고 내 선정된 변수들을 임베딩 벡터로 변환한 후, 각 벡터에 대해 코사인 유사도(Cosine Similarity)를 적용하여 최적의 일자리와 인재를 추천하도록 한다.

1) TF-IDF 기반 추천 알고리즘

TF-IDF 기반 추천 알고리즘은 Algorithm 1에 나타나 있으며, 자세한 구현 방법은 다음과 같다. 먼저, 구직자 프로파일과 채용공고 텍스트에 형태소 분석기를 적용해 문장을 형태소 단위로 분리한다. 형태소는 뜻을 가진 가장 작은 말의 단위를 의미한다. 예를 들어, “회계프로그램 사용자, 의사소통 원활하신분, 컴퓨터 사용 원활하신분”이라는 문장에 형태소 분석을 적용하면 “회계프로그램/NNG”, “사용자/NNG”, “의사소통/NNG”, “원활하/VV”, “신분/NNG”, “컴퓨터/NNG”, “사용/NNG”으로 분리가 된다. 문장을 형태소로 분리한 후, 조사와 동사 같은 불필요한 품사를 제외한다. 본 연구에서는 한국어 전용 형태소 분석기인 Konlpy의 Kkma 모듈을 사용한다. 형태소로 분리된 단어 집합에 TF-IDF를 적용해 2차원 행렬로 구성된 벡터를 만든다. 이때, 채용공고 데이터와 구직 데이터 각각 TF-IDF를 적용하여 2개의 임베딩 행렬을 생성한다. 구직 추천에서 채용공고 데이터는 2차원 행렬로 구성되고, 구직자 데이터는 벡터 형태로 처리된다. 이후, 구직자 벡터와 각 채용공고 행렬 간의 유사도를 계산한다. 마지막으로 유사도 점수가 가장 높은 상위 k개의 공고문을 추천한다. 반대로 구인 추천의 경우에는 구직자 데이터가 2차원 행렬로 구성되고, 채용공고 데이터가 벡터 형태로 처리된다.

2) SBERT 기반 추천 알고리즘

본 연구에서 사용한 SBERT 모델은 한국어를 지원하는 사전학습 모델 중 RoBERTa 모델을 적용한 jhgan/ko-sroberta-multitask을 적용하였다. SBERT 기반 추천 알고리즘은 Algorithm 2에 나타나 있으며, 자세한 구현 방법은 다음과 같다. 우선, 채용공고 데이터를 SBERT에 입력해 1차원 행렬로 구성된 벡터(1 x 768)를 생성한다. 해당 벡터는 채용공고 정보가 압축되어 숫자로 표현된 형태이다. 각각의 채용공고 벡터를 합치면 (N x 768) 차원의 2차원 벡터가 된다. 다음으로, 해당 벡터에 유사도를 적용하면 TF-IDF와 마찬가지로 채용공고와 구직자 간의 유사도 점수를 가진 2차원 벡터가 생성된다.

3) 코사인 유사도(Cosine Similarity)

두 벡터 간의 유사도 측정 방법은 코사인 유사도를 사용한다. 코사인 유사도는 내적 공간에서 두 벡터 간 각도의 코사인 값을 이용하여 벡터 간의 유사도를 측정하는 방법이다. 두 벡터 간의 각도가 0도

일 때 코사인 값은 1이 되며, 이는 두 벡터의 방향이 같음을 의미한다. 반대로 각도가 90도일 경우 코사인 값은 0이 되어 벡터 간에 유사성이 없음을 나타낸다. 코사인 유사도는 주어진 벡터 A와 B에 대해 식 (4)와 같이 계산된다. 여기서 는 두 벡터의 내적(Inner Product)을 의미하고, 와 는 각각 벡터 A와 B의 크기(Norm)을 나타낸다.

$$\text{Similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} \quad (4)$$

Algorithm 1 TF-IDF Algorithm for Job Recommendation

1. **Input:** Document: D (JD(or CV)), Query: Q (CV(or JD))
 2. **Output:** TF-IDF vector for CV(or JD) Q and JD(or CV) similarity score
 3. **Step 1: Preprocessing**
 4. Remove punctuations and stop words from the input query Q and document D .
 5. Tokenize both the query and the document.
 6. **Step 2: Compute Term Frequency (TF)**
 7. **for each** word w **in** query Q **do**
 8. $TF(w, D) \leftarrow \frac{\text{Number of times word } w \text{ appears in document } D}{\text{Total number of tokens in } D}$
 9. **end for**
 10. **Step 3: Compute Inverse Document Frequency (IDF)**
 11. **for each** word w **in** query Q **do**
 12. $IDF(w, D) \leftarrow \log \frac{N}{|d \in D: w \in d|}$
 13. N : Total number of documents in dataset D
 14. **end for**
 15. **Step 4: Compute TF-IDF for each word**
 16. **for each** word w **in** query Q **do**
 17. $TF-IDF(w, D) \leftarrow TF(w, D) \times IDF(w, D)$
 18. **end for**
 19. **Step 5: Compute Document Similarity**
 20. Calculate the cosine similarity between the TF-IDF vector of query Q and the document D .
 21. **Step 6: Return similarity score**
 22. **return** Cosine similarity score between query Q and document D .
-

Algorithm 2 Sentence-BERT Algorithm for Job Recommendation

1. **Input:** Sentence A : JD(or CV), Sentence B : CV(or JD)
 2. **Output:** Cosine similarity score between sentences A and B
 3. **Step 1: Initialize Sentence-BERT model**
 4. Load pre-trained Sentence-BERT model weights.
 5. **Step 2: Encode the sentences**
 6. **function** Encode(sentence)
 7. Tokenize the input sentence.
 8. Pass the tokenized sentence through Sentence-BERT to get token embeddings.
 9. Apply pooling strategy (mean-pooling or [CLS] token) to get a fixed-size sentence embedding.
 10. **return** Sentence embedding.
 11. **end function**
 12. **Step 3: Get embeddings for both sentences**
 13. embedding_A \leftarrow Encode(A)
 14. embedding_B \leftarrow Encode(B)
 15. **Step 4: Compute Cosine Similarity**
 16. cosine_similarity $\leftarrow \frac{\text{embedding}_A \cdot \text{embedding}_B}{\| \text{embedding}_A \| \| \text{embedding}_B \|}$
 17. **Step 5: Return the similarity score**
 18. **return** cosine similarity
-

IV. 연구 결과

1. 추천 성능 평가방법

실험에서 구직 추천과 구인 추천의 평가 기준은 상위 k 개의 추천 리스트 중에서 채용공고의 직종과 구직자의 희망 직종이 일치하는지로 정의한다. 성능 평가 지표로는 Top- k Accuracy와 Mean Average Precision(MAP) 2가지 지표를 사용하였다. Top- k Accuracy는 모델이 예측한 상위 k 개의 추천 결과 중에서 정답이 존재하는 경우 올바른 예측으로 간주하는 방식이다. 가령, Top-3 Accuracy는 모델이 예측한 상위 3개의 추천 결과 중 하나가 정답일 때 정답으로 간주한다. 일반적으로 k 가 커질수록 Top- k 정확도는 높아지게 된다.

$$Top - K Accuracy = \frac{1}{K} \sum_{i=1}^K Hit_i \quad (5)$$

Top-k Accuracy는 추천 리스트에 아이템이 등장하는 순위와 관계없이 사용자가 선호하는 아이템이 추천 결과에 포함되었는지만 평가한다. 그러나 추천된 아이템의 순위(Rank)는 사용자의 만족도와 직접적으로 연관되기 때문에 추천 품질을 평가할 때 중요한 요소이다. MAP는 추천 결과의 순위를 고려하여 사용자가 선호하는 아이템이 상위에 위치할수록 더 높은 점수를 부여한다. MAP는 개별 사용자에게 Average Precision(AP)을 먼저 계산한 후, 모든 사용자의 AP를 평균내어 구한다. Precision@k는 상위 k개의 추천 항목 중에서 사용자가 선호하는 항목의 비율을 의미한다. 예를 들어, 5개의 추천 항목 중 3개가 사용자가 선호하는 항목이라면, Precision@5는 3/5, 즉 0.6이 된다. AP는 개별 사용자가 추천받은 리스트에서 선호하는 항목이 나타나는 순위별로 Precision을 계산하여, 그 평균을 구한 값이다. 관심 있는 항목이 추천 리스트에서 나타날 때마다 그 시점의 Precision을 계산하고, 이를 누적하여 평균을 낸다. 즉 추천된 각 항목이 사용자의 선호 대상인지 여부를 나타내는 가중치로, 순위 추천 항목을 사용자가 선호하면 1, 그렇지 않으면 0을 부여한다. 순위가 높을수록 더 높은 점수를 부여하므로, 추천 리스트 상위에 있을수록 AP 점수가 높아진다. 이것을 모든 사용자 수에 대해 평균한 것이 MAP@k이며, 다음과 같이 계산한다.

$$Precision @ k = \frac{1}{k} \sum_{i=1}^k rel(i) \quad (6)$$

$$AP @ K = \frac{1}{K} \sum_{k=1}^K Precision @ k \cdot rel(i) \quad (7)$$

$$MAP @ K = \frac{1}{|U|} \sum_{i=1}^{|U|} (AP @ K)_i \quad (8)$$

2. 실험 및 분석 결과

본 연구에서는 구인·구직 추천에 적용한 TF-IDF와 SBERT 기반 방법의 성능을 비교 분석하였다. 또한, 추천 결과의 신뢰성과 유효성을 종합적으로 평가하기 위해 정성적 평가를 진행하였다.

첫 번째 실험은 k개의 추천 항목 수(3개와 5개)에 따른 TF-IDF와 SBERT 기반의 구직 추천과 구인 추천 알고리즘의 성능을 각각 비교하였으며, 분석 결과는 <표 4>과 같다. 구직 추천에서 Top-3 Accuracy는 SBERT가 81.5%로 TF-IDF(76.3%)에 비해 6.9% 성능 향상을 보였으며, Top-5 Accuracy에서는 SBERT가 84.2%로 6.6%의 성능 향상을 기록하였다. MAP@3과 MAP@5에서도 SBERT는 각각 73.2%, 71.5%로, TF-IDF 보다 8.44%와 4.82%의 성능이 향상되어, 추천된 직무의 순위 품질 측면에서도 SBERT가 더 나은 성능을 제공하는 것으로 분석된다.

구인 추천에서는 구직 추천에 비해 TF-IDF와 SBERT의 성능이 전반적으로 낮게 측정되었다. 이는 구인 데이터의 풀이 38개로 매우 제한적인 환경에서 실험이 진행되었기 때문에, 다양한 채용공고에 적합한 인력을 추천하기 어려웠던 것이 주요 원인으로 분석된다. 이러한 제한된 데이터 환경에서도 SBERT는 TF-IDF와 비교하여 훨씬 더 두드러진 성능 차이를 보여주어 유의미한 성능을 입증하였다. Top-3 Accuracy와 Top-5 Accuracy에서 SBERT는 각각 61.1%와 69.5%로, TF-IDF(49.0%와 55.7%)보다 상당한 성능 향상을 기록하였다. MAP@3에 비해 MAP@5에서 TF-IDF와 SBERT의 성능이 더 높게 측정되었는데, 구체적으로, TF-IDF는 MAP@3에서 34.6%였던 성능이 MAP@5에서는 36.3%로 약 1.7% 향상되었고, SBERT는 MAP@3에서 35.7%였던 성능이 MAP@5에서는 39.9%로 약 4.2% 향상되었다.

해당 실험 결과를 종합하면, 모든 지표에서 SBERT가 TF-IDF에 비해 전반적으로 우수한 성능을 보여주며, 복잡한 문맥을 고려해야 하는 추천 시스템에서는 SBERT가 더 효과적일 수 있음을 시사한다. TF-IDF는 단순한 빈도 기반 접근 방식을 사용하므로 문맥 정보를 고려하지 못하지만, SBERT는 문장 수준에서 의미를 파악할 수 있기 때문에 더 정확한 추천 결과를 제공할 수 있다. 특히 구인 추천에서 SBERT가 큰 차이를 보이는 것은, 제한된 구인 데이터에서도 의미를 잘 포착하는 능력이 더 우수하기 때문으로 볼 수 있다.

<표 4> TF-IDF 및 SBERT 추천 성능 비교

Data	Method	Top-3 Acc.	Top-5 Acc.	MAP@3	MAP@5
구직 추천	TF-IDF	0.76315	0.78947	0.67543	0.68278
	SBERT	0.81578	0.84210	0.73245	0.71571
	IMP. (%)	6.90	6.67	8.44	4.82
구인 추천	TF-IDF	0.49047	0.55714	0.34616	0.36336
	SBERT	0.61111	0.69523	0.35740	0.39982
	IMP. (%)	24.60	24.79	3.25	10.03

두 번째 실험에서는 추천 알고리즘이 제공하는 추천 항목이 얼마나 적합한지에 대해 심층적인 분석을 수행하였다. 두 알고리즘은 임베딩 과정에서 동일한 변수를 사용하여 구직자의 장애 유형 및 직무 적합성을 고려하도록 설계되었지만, 실제 추천 리스트에서 차이를 보인다.

〈그림 4〉는 각각 TF-IDF와 SBERT 기반으로 추천된 구직 리스트를 보여준다. TF-IDF 기반 추천 리스트를 살펴보면, 유사도 점수가 0.29에서 0.35 사이로, 비교적 낮은 유사도를 보인다. 추천된 직무들은 대체로 구직자가 제시한 희망 직무와 장애 유형을 반영하고 있으나, 일부 직무는 구직자의 장애 유형에 적합한 직무라고 보기 어려운 경우가 포함되어 있다. 반면, SBERT 기반 추천 결과에서는 유사도 점수가 0.72에서 0.75 사이로 TF-IDF보다 전반적으로 높은 유사도를 보인다. 추천된 직무들은 구체적인 장애 유형과 그에 따른 필요 요건을 보다 잘 반영하고 있으며, 특히 SBERT는 장애 정도와 직무의 상세 요구 사항 사이의 관계를 더 깊이 있게 분석하여 보다 정밀한 추천을 제공하고 있는 것으로 나타난다. 예를 들어, SBERT 기반 추천에서는 구직자가 희망하는 '재택사무'와 '지적장애'나 '뇌병변장애'와 같은 특정 장애에 맞춤형 직무가 추천되고 있다.

〈그림 5〉은 각각 TF-IDF와 SBERT 기반으로 추천된 구인 리스트를 보여준다. TF-IDF 기반 추천 리스트를 살펴보면, 유사도 점수가 0.08에서 0.18 사이로 비교적 낮은 범위를 유지하고 있다. 추천된 인력들은 직무에 적합한 전공이나 자격증을 가지고 있지만, 추천된 순위가 직무의 중요도와 완전히 일치하지 않는 경우가 존재한다. 예를 들어, 2번째로 추천된 인력보다는 3번이나 4번에 추천된 인력이 직무 적합성이 더 높아 보이며, 우선순위가 잘 반영되지 않은 것으로 분석된다. 이는 TF-IDF가 단순 텍스트 유사성에 기반하여 추천하기 때문에 직무와 구직자 간의 세부적인 적합성이나 우선순위를 충분히 반영하지 못한 결과로 볼 수 있다. 반면, SBERT 기반 추천에서는 유사도 점수가 0.65에서 0.73 사이로 TF-IDF보다 전반적으로 높은 유사도를 보인다. SBERT는 전공과 자격증뿐만 아니라 장애 유형과 직무의 상세 요구사항 간의 의미론적 관계를 더 깊이 있게 분석하여 추천을 제공한다. 예를 들어, '사회복지' 전공의 후보자들은 사회복지 관련 직무와 높은 유사도를 보였으며, 우선순위로 추천된 것을 확인할 수 있다. 이는 SBERT가 텍스트의 문맥을 이해하고, 장애 유형과 직무의 세부적 적합성을 반영한 추천이 가능하다는 것을 입증한다. 결과적으로, SBERT는 단순한 키워드 기반의 추천을 넘어, 구직자의 특성과 직무 요구 사항을 심층적으로 분석하여 보다 개인화된 추천을 가능하게 한다. 이와 같은 결과는 추천 시스템의 설계 시 정량적 지표뿐만 아니라 실제 추천의 적합성과 유용성을 평가하는 정성적 접근 또한 고려해야 한다는 시사점을 제공한다.

(그림 4) TF-IDF 및 SBERT 기반 구직 추천 결과 예시

Results of Company Recommendation									
CV	분류	희망직종 (전문/자격증/어학)	직역 사항	장래 유형	장래 종류				
TF-IDF	ID	분류	제목	직무	담당업무	우대조건	적합 장애유형	장애인 편의 시설	similarity
	56	경영사무금융보험	[장애인 채용] 사무보조 구인(경영/총무 관련)	사무보조 및 사무관리	원고 작성, 역설, 직접 총 관리, 사무보조 업무		지체장애(경중), 뇌병변장애(경중), 정신장애	장애인 편의 시설	0.3591
	43	경영사무금융보험	[장애인 채용] 일반 사무 업무 (일 8시간 재택 근무)	일반 사무 업무	행정, 배송, 응송 관련 사무 업무		지체장애(경중), 청각장애(경중), 언어장애(경중)	장애인 편의 시설	0.3318
	77	경영사무금융보험	[장애인 채용] 재택 근무 사무보조 사원 채용	사무보조	블로그마케팅 업무 (광고작성)		지체장애(중중), 뇌병변장애(중중), 정신장애	장애인 편의 시설	0.3183
	61	경영사무금융보험	[장애인 채용] 일반 사무 업무 총 (일 8시간 재택 근무)	일반 사무 업무	행정, 배송, 응송 관련 업무		지체장애(경중), 청각장애(경중), 뇌병변장애(경중)	장애인 편의 시설	0.3138
56	경영사무금융보험	[장애인 채용] 오피스 채용 지원 (일 8시간 재택 근무)	행정, 배송, 응송	재택 근무 관련 업무		지체장애(경중), 청각장애(경중), 언어장애(경중)	장애인 편의 시설	0.2958	
SBERT	ID	분류	제목	직무	담당업무	우대조건	적합 장애유형	장애인 편의 시설	similarity
	60	경영사무금융보험	중증장애인 재택근무 행정보조인력 모집	행정보조	주제별 자료조사 및 조사 내용 요약 제출 (원어민 용서 작성능력 필수)	고졸학력, 컴퓨터활용능력 1급 이상	지체장애(경중), 뇌병변장애(경중), 정신장애	기타(재택근무)	0.7598
	88	경영사무금융보험	[중증장애인 채용/재택근무] 장애인 재택 정보입력 및 생산, 접수 업무	정보입력 및 생산, 접수 업무	정보입력 및 생산, 접수 안내 (의사소통 능력 필수), 컴퓨터 활용능력 1급 이상	고졸학력, 컴퓨터활용능력 1급 이상, 인사시스템 활용능력, 컴퓨터 활용능력 1급 이상	지체장애(중중), 뇌병변장애(중중), 정신장애	장애인 편의 시설	0.7531
	77	경영사무금융보험	[장애인 채용] 재택 근무 사무보조 사원 채용	사무보조	블로그마케팅 업무 (광고작성)	중증장애인 필수	지체장애(중중), 뇌병변장애(중중), 정신장애	장애인 편의 시설	0.7438
	42	경영사무금융보험	재택근무 간단한 사무보조 업무 장애인 채용 모집 공고	사무보조	관련 기사스크랩, 용의금 대출, 판매 원고 작성 및 회의 진행	중증 장애인 우대	지체장애(중중), 뇌병변장애(중중)	장애인 편의 시설	0.7314
47	경영사무금융보험	[장애인 채용 / 재택근무] 영입지원 사무보조 1명 모집	사무보조	문서 입력, 자료 정리 등		지체장애(경중), 언어장애(경중), 청각장애(경중)	장애인 편의 시설	0.7275	

(그림 5) TF-IDF 및 SBERT 기반 구인 추천 결과 예시

Results of Job Seeker Recommendation									
JD	분류	제목	직무	담당업무	자격	우대조건	적합 장애유형	장애인 편의시설	
	교육·법률·사회복지·경찰·소방 및 안전	2024 상용 장애인재용장성지원센터 홍보 담당가 모집	홍보 담당	홍보상임을 통한 사색관리 및 내방자 유무조사 및 대상자 발굴	사회복지사 2급, 홍보상임가 양성 과정 이수자(필수)		[지체장애(경중), 지체장애(중중), 뇌병변장애(경중), 뇌병변장애(중중)]	장애인을 주차장, 건물 내부 경사로 장애인용 화장실	
TF-IDF	분류	희망직종	직공	자격증/어학	장래 유형	장래 종류	적합 장애유형	장애인 편의 시설	similarity
	10	교육·법률·사회복지·경찰·소방 및 안전	재택사무직	사회복지학	-사회복지사 1급wn-정소년지도사 2급	뇌병변장애(편마비)	중중	0.201561	
	31	경영사무금융보험	고객상담	전기컴퓨터공학	-컴퓨터활용능력 1급(23.06)wn-정보처리기사(22.07)	지체장애	중중	0.659295	
	25	교육·법률·사회복지·경찰·소방 및 안전	사무보조	사회복지학	-컴퓨터활용능력 2급wn-사회복지사 2급	뇌병변장애	중중	0.626152	
	24	경영사무금융보험	상담직	관광중국어과	-사회복지사 2급wn-노인심리상담사wn-ITQ 한글 PPTwn-워드프로세서 2급	지체장애(천주)	중중	0.627649	
1	경영사무금융보험	사무지원		-ITQ파워포인트(22.04)wn-ITQAI레완글(21.07)wn-간호조무사(21.03)	지적장애	중중	0.619538		
SBERT	분류	희망직종	직공	자격증/어학	장래 유형	장래 종류	적합 장애유형	장애인 편의 시설	similarity
	24	경영사무금융보험	상담직	관광중국어과	-사회복지사 2급wn-노인심리상담사wn-ITQ 한글 PPTwn-워드프로세서 2급	지체장애(천주)	중중	0.193179	
	23	경영사무금융보험	고객상담직	미디어공학		청각장애	중중	0.127647	
	25	교육·법률·사회복지·경찰·소방 및 안전	사무보조	사회복지학	-컴퓨터활용능력 2급wn-사회복지사 2급	뇌병변장애	중중	0.118006	
	31	경영사무금융보험	고객상담	전기컴퓨터공학	-컴퓨터활용능력 1급(23.06)wn-정보처리기사(22.07)	지체장애	중중	0.083771	
10	교육·법률·사회복지·경찰·소방 및 안전	재택사무직	사회복지학	-사회복지사 1급wn-정소년지도사 2급	뇌병변장애(편마비)	중중	0.082068		

V. 결론 및 논의

본 연구의 목적은 장애인을 대상으로 하는 맞춤형 구인·구직 추천 알고리즘을 구현하여 장애인 취업 시장에 존재하는 정보격차를 해소하고, 장애인 인력 고용 활성화를 도모하는 것이다. 본 연구에서는 보다 개인화된 추천을 위해서, LLM의 프롬프트 엔지니어링을 통해 데이터의 구조화를 수행하고, 채용공고에 직접적으로 명시되지 않은 적합 장애 유형 등의 은닉 정보를 추출하여 활용하였다. 또한, 개인별 직무수행 능력과 업무별 요구되는 신체 능력을 고려하기 위해서, 개인의 희망 직무, 장애 유형 및 정도와 기업이 제시하는 직무와 업무 환경 등을 추천 시스템의 요소로 활용하여 추천을 진행하였다.

본 연구에서는 수행한 실험의 평가를 위해, 상위 k개의 추천 리스트 중에서 채용공고의 직종과 구직자의 희망 직종 일치 여부를 비교하여 Top-k Accuracy와 MAP를 산출하였다. SBERT를 기반으로 하였을 때, 구직 추천의 경우 0.842의 Top-5 Accuracy를, 0.716의 MAP@5를 달성하였으며, 구인 추천의 경우 0.695의 Top-5 Accuracy를, 0.4의 MAP@5를 달성하였다. 전반적으로 SBERT의 결과가 TF-IDF의 결과보다 우수함을 확인할 수 있었는데, 이는 SBERT가 복잡한 문맥을 고려해야 하는 추천 시스템에서 효과적일 수 있음을 시사한다. 또한, 추천 결과를 직접 비교 및 분석하여 정성 평가를 수행하였는데, 결과적으로 수요자의 특성에 맞는 적절한 구인·구직 추천이 이루어진 것을 확인할 수 있었다. 본 연구를 통해, 장애인 노동시장의 수요자 맞춤형 구인·구직 추천 알고리즘에 LLM의 적용 가능성을 확인하였고, 수요자의 특성과 요구사항이 반영되어 보다 개인화된 추천이 가능함을 알 수 있었다.

다만, 이러한 결과에도 본 연구의 한계점은 존재한다. 우선, 장애인 관련 분야에서의 제한된 데이터 문제이다. 기본적으로 특수한 분야이다 보니, 비교적 데이터의 확보가 쉽지 않으며, 확보된 데이터마저도 좋은 품질을 기대하기 힘든 실정이다. 다수의 결측치가 존재하며, 구조화되어 있지 않아 활용도가 떨어진다. 일반적으로 수기로 작성하여 형식이 정형화되어 있지 않은 경우가 많은데, 이러한 데이터를 인공지능에 적용하기 위해서는 더미변수로의 전환 및 범주형 태깅 등 정형화된 형태로 변환해야 한다(김필호·정규형, 2020). 또한, 이미지로 작성된 일부 데이터에서 내용 추출 및 처리에 한계가 있을 만큼 큰 이미지에 과도하게 많은 정보를 담은 경우가 다수 발견되었는데, 장애인 대상의 채용공고인 만큼 디지털 정보의 접근성을 더욱 고려하여 개선되어야 할 것이다. 추후에 다양한 양질의 데이터를 충분히 확보할 수 있다면, 시스템의 성능 개선에 큰 도움이 될 것으로 보인다.

다음으로, 데이터의 구조화 및 은닉 정보 추출을 위한 전처리 과정에서 LLM을 활용하였는데, 생성형 AI의 특성상 같은 입력에도 매번 다른 결과를 출력할 수 있다는 점이다(Radford, A. et al., 2019). 이것은 LLM 자체의 한계로도 볼 수 있는데, 이를 보완하기 위해 도출한 결과의 일관성 확보를 위한 검증 과정이 필요할 것으로 보인다. 뿐만 아니라, LLM의 환각(Hallucination) 현상에 대한 문제점도 고려되어야 한다. 연구에서 사용한 데이터는 한정적이기 때문에, 본 연구에서는 LLM이 생성한 결과에 대한 데이터 검증 작업을 직접 수행하였으나, 향후 연구에서 데이터를 추가적으로 수집하여 활용할 경우에는 별도의 모델 신뢰도 검증 방안을 검토할 필요가 있을 것이다. 그럼에도 본 연구에서 LLM을 활용한 은닉 정보의 추출과 그 활용에 대한 가능성을 확인하였고, 일관성 및 신뢰도 보장에 대한 한계를 보완한다면 향후 연구에서 LLM의 활용성을 확장시킬 수 있을 것이다.

또한, LLM을 활용한 데이터 가공 등 직접 데이터를 LLM에 입력해야 하는 작업 시에는 개인정보 유출 등 보안과 관련하여 주의를 기울여야 한다(Bommasani, R. et al., 2021). 본 연구에서는 GPT-4o-mini API(Application Programming Interface)의 유료 버전 활용으로 보안 문제를 해결했지만,

개인정보유출을 대비해 자체 LLM을 구축하여 활용하는 것이 바람직해 보인다.

한국장애인고용공단을 비롯한 정책기관에서 장애인들의 채용 및 사회 진출을 꾸준히 지원하고 있으나, 대내외적 이유로 여전한 편견과 비장애인에 비해 작은 규모의 채용시장을 겪고 있다(김필호·정규형, 2020). 계속해서 빠르게 발전하는 디지털 시대에서 장애인들이 느끼는 정보격차는 커질 수밖에 없고(김영호 외, 2019), 더불어 구인·구직에 있어 수급 불균형이 심화되고 있다(전희주 외, 2021). 본 연구의 경우에도, 경기복지재단을 비롯한 경기도 산하기관 및 각종 장애인 복지 관련 기관과 협업하여 데이터를 수집하고 연구를 수행하였으나, 데이터의 확보가 쉽지 않고 관련 정보가 여전히 부족한 상황이다. 이러한 문제들을 해소하기 위해 본 연구에서는 기존의 일자리 추천 시스템을 넘어, 개인의 특성과 희망 직무, 장애 유형 및 정도를 고려하여 맞춤형 구인·구직 정보를 추천하는 시스템을 제안하였다. 고질적인 사회 문제 해결을 위해 최신 생성형 인공지능 모델을 기반으로 명시되지 않은 은닉 정보를 도출하여 추천에 활용한다는 점에서 다른 연구와의 차별점을 확인할 수 있다. 또한, 개인뿐만 아니라 장애인 인력 고용을 희망하는 기업에게도 적절한 인재 탐색에 도움을 줄 수 있을 것으로 보여, 장애인 채용 시장의 심각한 수급 불균형 해결에 기여할 것으로 예상된다.

향후 추천 시스템의 실사용을 위한 플랫폼 구축 시, 장애인 구직 희망자 데이터 풀의 양과 질 향상이 필요하고, 다양한 데이터 확보를 통해 기업의 채용 패턴 파악이 필요할 것이다(김필호·정규형, 2020). 또한, 장애인을 대상으로 하는 만큼 사용자의 접근성이 용이한 인터페이스의 구축이 필요할 것으로 예상된다. 본 연구에서는 장애 유형 및 정도, 희망 직무 등의 항목을 추천에 활용하였는데, 추후 개인의 경력 정보를 획득하여 추천 시스템에 활용한다면, 비장애인들에게는 상당히 보편화되어 있는 이직 시장에서도 장애인의 능동적인 구직 활동을 기대해 볼 수 있을 것이다.

현재까지 장애인의 자립 생활을 위한 지원은 주로 지원금 지급 등 단기적인 문제 해결에 집중되어 왔다. 그러나 이러한 접근 방식은 장애인 관련 사회 문제의 근본적인 해결책이 되기 어렵다. 본 연구는 이러한 문제를 개선하기 위한 출발점으로, 경기도 내에서 장애인들이 제약 없이 구인·구직 활동에 적극 참여할 수 있는 환경을 조성하는 것을 목표로 한다. 이를 위해, 경기도 내 유관 기관들과 지속적으로 협력하고, 사용자 피드백을 시스템에 반영하는 프로세스를 구축할 예정이다. 특히, 경기도는 전국에서 가장 인구가 많은 광역자치단체로서, 본 연구에서 개발된 장애인 구인·구직 추천 시스템의 전국적 확산 가능성에 중요한 기반이 될 수 있다. 이러한 시스템의 확산을 통해 장애인 인력의 고용기회가 증대되고, 체계적인 장애인 일자리 제공의 기틀이 마련되어, 장애인들이 장애에 구애받지 않고 개인의 역량을 발휘할 권리가 보장되기를 기대한다.

■ 참고문헌 ■

- 이수련·양성원(2023). “장애인 근로자의 노동시장 이행행태 분석”, 『노동정책연구』, 23(2): 1–22.
- 이윤지 외(2023). 『2023 하반기 장애인경제활동실태조사』, 한국장애인고용공단.
- Qin, Chuan, et al.(2023). “A comprehensive survey of artificial intelligence techniques for talent analytics.”, *arXiv preprint*, arXiv:2307.03195.
- Wu, Likang, et al.(2024). “A survey on large language models for recommendation.”, *World Wide Web*, 27,5: 60.
- Fang, C., et al.(2023). “Recruitpro: A pretrained language model with skill-aware prompt learning for intelligent recruitment.”, In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, p. 3991–4002.
- Guan, Zhihao, et al.(2024). “JobFormer: Skill-Aware Job Recommendation with Semantic-Enhanced Transformer.”, *arXiv preprint*, arXiv:2404.04313.
- 조신영 외(2023). 『2023 기업체장애인고용실태조사』, 한국장애인고용공단.
- 전희주 외(2021). 『장애인 구인·구직 미스매치 해소를 위한 디지털 기반 강화방안』, 한국장애인고용공단.
- Young, T., et al.(2018). “Recent trends in deep learning based natural language processing”, *IEEE Computational Intelligence Magazine*, 13(3): 55–75.
- Almeida, F., & Xexeo, G.(2019). “Word embeddings: A survey.”, *arXiv preprint*, arXiv:1901.09069.
- Wang, S., et al.(2020). “A survey of word embeddings based on deep learning.”, *Computing*, 102(3): 717–740.
- Tata, S., & Patel, J. M.(2007). “Estimating the selectivity of tf-idf based cosine similarity predicates.”, *ACM Sigmod Record*, 36(2): 7–12.
- Devlin, J.(2018). “Bert: Pre-training of deep bidirectional transformers for language understanding.”, *arXiv preprint*, arXiv:1810.04805.
- Liu, Y.(2019). “Roberta: A robustly optimized bert pretraining approach.”, *arXiv preprint*, arXiv:1907.11692.
- Sun, K., et al.(2022). “A survey of pretrained language models.”, In *International Conference on Knowledge Science, Engineering and Management*, Cham: Springer International

- Publishing, p. 442–456.
- Reimers, N.(2019). “Sentence–BERT: Sentence Embeddings using Siamese BERT–Networks.”, *arXiv preprint*, arXiv:1908.10084.
- Han, Yuexing.(2024). “Automatic Pipeline for Information of Curve Graphs in Papers Based on Deep Learning.”, *Research Square*, <https://doi.org/10.21203/rs.3.rs-3825977/v1>.
- De Ruijt, Corne, & Sandjai Bhulai.(2021). “Job recommender systems: A review.”, *arXiv preprint*, arXiv:2111.13576.
- A, Patil et al.(2023). “A Survey on Artificial Intelligence (AI) based Job Recommendation Systems.”, *2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India*, p. 730–737.
- Mashayekhi, Yoosof, et al.(2024). “A challenge–based survey of e–recruitment recommendation systems.”, *ACM Computing Surveys*, 56.10: 1–33.
- 손지은 외(2015). “추천 시스템 기법 연구동향 분석”, 『대한산업공학회지』, 41(2): 185–208.
- Guo, X. et al.(2014). “An analysis framework for content–based job recommendation.”, In *Proceedings of the 22nd International Conference on Case–Based Reasoning (ICCBR), Cork, Ireland, 29 September – 01 October 2014*.
- Alsaif, S.A. et al.(2022). “Learning–Based Matched Representation System for Job Recommendation.”, *Computers*, 11(11): 161.
- Mishra, R. & Rathi, S.(2020). “Efficient and Scalable Job Recommender System Using Collaborative Filtering.”, In *ICDSMLA 2019: Proceedings of the 1st International Conference on Data Science, Machine Learning and Applications*, Springer Singapore, p. 842–856.
- Patel, R. & Vishwakarma, S.K.(2020). “An Efficient Approach for Job Recommendation System Based on Collaborative Filtering.”, In *ICT Systems and Sustainability: Proceedings of ICT4SD 2019, Volume 1*, Springer Singapore, p. 169–176.
- Saha, D. et al.(2023). “Job Recommendation a Hybrid Approach Using Text Processing.”, In *Doctoral Symposium on Human Centered Computing, Singapore: Springer Nature Singapore*, p. 74–85.
- Gupta, A. & Garg, D.(2014). “Applying data mining techniques in job recommender system for considering candidate job preferences.”, In *2014 international conference on advances*

- in computing, communications and informatics (ICACCI)*, IEEE, p. 1458–1465.
- Martinez–Gil et al.(2018). “Recommendation of job offers using random forests and support vector machines.”, In *Proceedings of the of the EDBT/ICDT joint conference*.
- Chen, W. et al.(2018). “Tree–based contextual learning for online job or candidate recommendation with big data support in professional social networks.”, *IEEE Access*, 6: 77725–77739.
- Wang, P. et al.(2016). “The analysis and design of the job recommendation model based on GBRT and time factors.”, In *2016 IEEE International Conference on Knowledge Engineering and Applications (ICKEA)*, IEEE, p. 29–35.
- Gugnani, A. & Misra, H.(2020). “Implicit skills extraction using document embedding and its use in job recommendation.”, In *Proceedings of the AAAI conference on artificial intelligence*, p. 13286–13293.
- Khaouja, I. et al.(2021). “A survey on skill identification from online job ads.”, *IEEE Access*, 9: 118134–118153.
- 개인정보보호법, 법률 제 19234호.
<https://www.selenium.dev/> Selenium,
<https://openai.com/chatgpt/> OpenAI,
 장애인복지법 시행규칙, 보건복지부령 제1050호.
<https://huggingface.co/jhgan/ko-sroberta-multitask> Hugging Face.
- Mikolov, T. et al.(2013). “Efficient estimation of word representations in vector space.”, *arXiv preprint*, arXiv:1301.3781.
- Mitra, B. et al.(2017). “Neural models for information retrieval.”, *arXiv preprint*, arXiv:1705.01509.
- 김필호·정규형(2020). “장애인의 직업능력 유형화 및 취업 간의 관계.”, 『장애와 고용』, 30(3): 109–130.
- Radford, A. et al.(2019). “Language Models are Unsupervised Multitask Learners.”, *OpenAI blog*, 1(8): 9.
- Bommasani, R. et al.(2021). “On the Opportunities and Risks of Foundation Models.”, *arXiv preprint*, arXiv:2108.07258.
- 김영호 외(2019). “디지털 시대의 장애인 정보격차 해소 방안 연구”, 『정보사회연구』, 25(3): 55–78.

원 고 접 수 일 | 2024년 9월 30일

심 사 완 료 일 | 2024년 11월 8일

최종원고채택일 | 2024년 11월 11일

임윤정 yunjeong.im@snu.ac.kr

2021년 경기대학교 경영정보학 학사학위를 받았다. 현재 차세대융합기술연구원 재난안전융합연구센터 데이터과학연구부문 연구원으로 재직 중이며, 성균관대학교 인공지능융합학 석사과정에 재학 중이다. 관심분야는 데이터 과학, 기계학습, 딥러닝, 멀티모달 등이며, 저술은 “Perceiving Conflict of Interest Experts Recommendation SystemBased on a Machine Learning Approach”(2023), “Deep Learning Methods for Predicting Tap-Water Quality Time Series in South Korea”(2022) 등 다수 발표하였다.

권지우 jzyuu98@snu.ac.kr

2024년 국립부경대학교 인공지능융합학 석사학위를 받았다. 현재 차세대융합기술연구원 재난안전융합연구센터 데이터과학연구부문 연구원으로 재직 중이다. “Significant Wave Height Prediction from X-Band Marine Radar Images using Deep Learning with 3D CNN”(2023), “딥러닝 신경망을 사용하는 벤더-계슈탈트 검사의 자동채점: 도형의 배열순서”(2023) 등의 논문을 발표하였다. 주요 관심분야는 데이터 과학, 딥러닝, 패턴인식, 컴퓨터비전, 멀티모달 등이다.

송규원 gyuwon.song@snu.ac.kr

2016년 과학기술연합대학원대학교(KIST) HCI 박사학위를 받았다. KIST 영상미디어연구센터 연구원을 역임했으며, 현재 차세대융합기술연구원 재난안전융합연구센터 데이터과학연구부문 책임연구원으로 재직 중이다. “Perceiving Conflict of Interest Experts Recommendation SystemBased on a Machine Learning Approach”(2023) 등 다수 논문을 발표하였고, 관심분야는 인공지능 및 데이터 과학, 클라우드-엣지컴퓨팅, 블록체인 등이다.