Journal of KIIT. Vol. 24, No. 8, pp. 217-230, Aug. 31, 2026. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2026.24.8.217

텍스트 임베딩 및 대규모 언어모델 기반 채용공고-NCS 직무 자동 매핑 방법의 성능 비교 연구

김경하*, 김동호**[1], 유기중**[2]

A Comparative Study of Text Embedding and Large Language Model-based Methods for Automatic Mapping between Job Postings and NCS Job Categories

Kyongha Kim*, Dongho Kim**[1], and Kijung Ryu**[2]

요 약

산업 채용공고와 국가직무능력표준(NCS) 인공지능 세분류 간 의미 기반 자동 매핑 가능성을 검토하고, Ko-SBERT와 GPT-4o-mini를 활용해 임베딩 기반 접근과 LLM 기반 접근의 성능과 오류 특성을 비교하였다. 284건의 채용공고를 대상으로 전문가 검증 기반 기준 라벨을 구축하고, 1:1 및 1:3 매핑 실험을 수행하였다. 실험 결과, 1:1 매핑에서는 LLM 기반 접근이 상대적으로 높은 성능을 보였으며, 1:3 매핑에서는 두 접근 모두 높은 후보 추천 성능을 보였다. 오분류는 주로 의미적으로 인접한 세분류 사이에서 발생했으며, 세분류 간 직무 경계 중첩이 주요 요인으로 작용하였다. 이러한 결과는 채용공고와 NCS 간 의미 정렬 가능성을 보여주며, AI 직무훈련 설계 지원을 위한 기초 근거를 제공한다.

Abstract

This study examines the feasibility of semantic-based automatic mapping between industrial job postings and the AI subcategories of the National Competency Standards (NCS). Using Ko-SBERT and GPT-4o-mini, it compares the performance and error characteristics of an embedding-based approach and an LLM-based approach. A reference label dataset was constructed from 284 job postings through expert validation, and 1:1 and 1:3 mapping experiments were conducted. The results showed that the LLM-based approach achieved relatively higher performance in 1:1 mapping, while both approaches demonstrated high candidate recommendation performance in 1:3 mapping. Misclassifications mainly occurred among semantically adjacent subcategories, and overlapping job boundaries among the subcategories were a major contributing factor. These results demonstrate the feasibility of semantic alignment between job postings and the NCS and provide foundational evidence to support the design of AI job training programs.

Keywords

national competency standards, NCS, Job posting analysis, automatic Job mapping, Ko-SBERT, GPT-4o-mini

* 숭실대학교 IT정책경영학과 박사과정
- ORCID: https://orcid.org/0009-0008-1700-7010
** 숭실대학교 글로벌미디어학부 교수(**[2] 교신저자)
- ORCID[1]: https://orcid.org/0000-0002-8100-9079
- ORCID[2]: https://orcid.org/0009-0007-2862-4092

· Received: May 04, 2026, Revised: May 26, 2026, Accepted: May 29, 2026

· Corresponding Author: Kijung Ryu
Global School of Media, Soongsil University, Seoul, 06978, Korea
Tel.: +82-2-820-0114, Email: kjryu@ssu.ac.kr

Ⅰ. 서 론

최근 생성형 AI를 포함한 인공지능 기술의 확산은 산업 전반의 업무 수행 방식과 인력 수요 구조를 빠르게 변화시키고 있다. 경제협력개발기구(OECD)는 AI 노출도가 높은 직무에서 요구 역량의 중심이 관리, 비즈니스 프로세스, 사회·정서 역량, 디지털 역량으로 이동하고 있음을 보고하였으며, 국제노동기구(ILO) 역시 생성형 AI가 직무 전체를 일괄적으로 대체하기보다 직무를 구성하는 과업과 요구 역량을 재구성하는 방향으로 영향을 미친다고 분석하였다[1][2].

이와 같은 변화 속에서 산업 수요와 국가직무능력표준(NCS, National Competency Standards)을 의미적으로 정렬할 수 있는지를 검토할 필요가 있다. NCS는 산업현장에서 요구하는 지식, 기술, 태도를 국가 차원에서 체계화한 기준으로서, 교육훈련, 자격, 채용, 인사관리 등 다양한 영역에서 활용된다[3][4]. 그러나 인공지능 분야는 기술 변화 속도가 매우 빠르기 때문에, 채용공고에는 최신 기술 스택과 실무 요구가 빠르게 반영되는 반면 NCS는 제도적 안정성과 범용성을 기반으로 구성된다. 이로 인해 두 체계 사이에는 의미적 간극이 발생할 수 있다[1][3]. 특히 이러한 간극의 검토는 NCS 기반 AI 직무훈련 체계와 산업 수요 간 정합성을 확인한다는 점에서 중요하다[4][5].

채용공고는 산업 수요를 직접적으로 보여주는 대표적 자료이다. 세계은행과 유럽직업훈련개발센터(Cedefop)는 온라인 채용공고가 직무 수요와 기술 요구를 시의성 있게 파악할 수 있는 노동시장 정보원으로 활용할 수 있다고 설명한다[6][7]. 국내에서도 관련 연구들은 채용공고를 활용해 산업현장의 필요 역량을 도출하고, 이를 교육과정과 연결하려는 접근을 제시하였다[8][9]. 또한 구인광고와 NCS를 비교하여 정보보호 직무의 필요 지식과 기술을 분석한 선행연구도 있다[10]. 이는 채용공고가 산업 수요와 국가 표준 체계 간 비교를 위한 유의미한 자료가 될 수 있음을 보여준다. 그러나 실제 채용공고는 비정형 자연어 텍스트이며, 하나의 공고에 복수의 직무 역할과 기술 요구가 혼합되어 있는 경우가 많다. 반면 NCS는 구조화된 직무 분류 체계이므로 두 체계를 직접 연결하는 데에는 한계가 있다[7][10][11]. 또한 기존 연구들은 채용공고를 활용한 역량 분석이나 NCS와의 비교 가능성을 주로 제시하였으며, 채용공고를 NCS 인공지능 세분류 직무에 자동 매핑하고 임베딩 기반 접근과 대규모 언어모델(LLM, Large Language Model) 기반 접근의 성능을 직접 비교한 실험 연구는 아직 제한적이다. 따라서 채용공고와 NCS 간 관계를 전문가의 수작업에만 의존하기보다, 자연어 처리 기반의 의미 분석을 통해 자동 매핑하는 접근이 필요하다[11][12].

산업 채용공고와 NCS 인공지능 세분류 간의 의미 기반 자동 매핑 가능성을 검토하기 위해, 한국어 문장 임베딩 기반 의미 유사도 접근인 Ko-SBERT와 대규모 언어모델 기반 문맥 해석 접근인 GPT-4o-mini를 비교하였다. Ko-SBERT는 채용공고와 NCS 세분류 설명 간 의미적 유사도를 정량적으로 산출하는 방식으로 활용하였고, GPT-4o-mini는 NCS 세분류 설명과 채용공고 텍스트를 함께 고려하여 문맥 기반 분류를 수행하는 방식으로 활용하였다. 최고 성능 모델 탐색보다, 임베딩 기반 접근과 LLM 기반 접근이 채용공고 - NCS 매핑 문제에서 보이는 적용 가능성, 성능 차이, 오류 특성을 분석하는 데 초점을 두었다. 그림 1은 실험 절차를 간략히 도식화한 것이다.

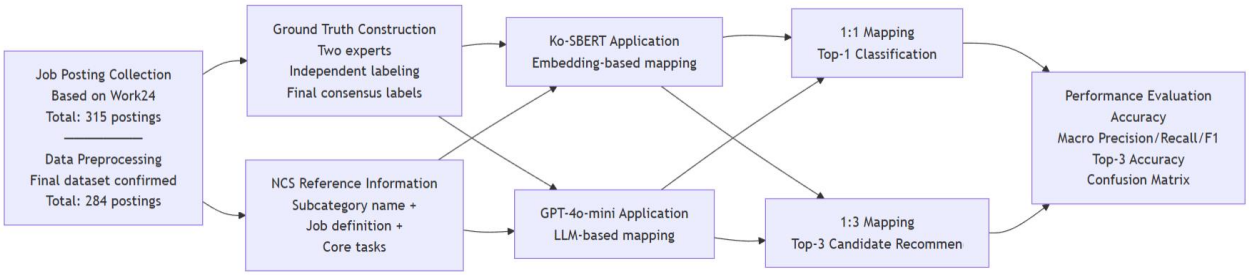

그림 1. 채용공고와 NCS 자동 매핑 연구 절차
Fig. 1. Research workflow for automatic mapping between job postings and NCS

NCS 인공지능 세분류 7개에 해당하는 채용공고 315건을 수집한 후, 전처리를 거쳐 최종 284건의 분석 데이터를 구성하였다. 이후 전문가 검증을 통해 기준 라벨을 구축하고, Ko-SBERT 기반 문장 임베딩 방식과 GPT-4o-mini 기반 대규모 언어모델 방식을 적용하여 1:1 및 1:3 매핑 조건에서 성능을 비교하였다.

연구 질문은 다음과 같다. 첫째, 실제 산업 채용공고 텍스트와 NCS 인공지능 세분류 직무 간의 의미 기반 자동 매핑이 가능한가? 둘째, 임베딩 기반 의미 유사도 접근과 LLM 기반 문맥 해석 접근은 1:1 및 1:3 매핑 조건에서 어떠한 성능 차이를 보이는가? 셋째, 세분류 간 혼동은 어떤 양상으로 나타나며, 주요 오류 원인은 무엇인가?

이를 통해 채용공고와 NCS 간 의미 정렬 가능성을 세분류 수준에서 실험적으로 검토하고, 임베딩 기반 접근과 LLM 기반 접근의 성능 차이를 비교함으로써 자동 직무 매핑 연구의 기초 근거를 제시한다. 또한 세분류 간 혼동 패턴과 오류 원인을 분석하여 산업 수요와 국가 표준 직무 체계 간 대응 관계를 설명하고, 산업 수요 기반 AI 직무훈련 설계 연구에 필요한 시사점을 도출한다.

II. 관련 연구

2.1 국가직무능력표준(NCS) 체계

국가직무능력표준(NCS)은 산업현장에서 직무 수행에 필요한 지식, 기술, 태도를 국가 차원에서 표준화한 체계이다. NCS는 개별 직무를 능력단위 또는 능력단위의 집합 형태로 구조화하여 제시하고, 교육훈련, 자격, 채용, 경력개발 등 다양한 영역의 기준을 제공한다[3].

NCS의 분류 체계는 일반적으로 대분류-중분류-소분류-세분류-능력단위의 위계 구조를 가진다. 이 중에서 대분류 20. 정보통신, 중분류 01. 정보기술, 소분류 07. 인공지능에 속하는 세분류 직무를 분석 대상으로 하였다[3].

채용공고는 구체적인 직무 요구와 기술 조합을 포함하므로, 상위 분류 수준보다 세분류 수준에서 산업 수요와 NCS 간의 정렬 가능성을 검토하는 것이 적절하다. 분석 대상인 NCS 인공지능 세분류 7개의 핵심 직무 내용을 표 1에 요약하였다[3].

2.2 채용공고 기반 직무 분석 연구

온라인 채용공고는 기업이 실제 채용 과정에서 요구하는 수행 업무, 자격 요건, 기술 스택, 우대사항 등을 구체적으로 제시하므로 산업현장의 직무 수요를 반영하는 실증 자료로 적합하다. 세계은행과 Cedefop은 온라인 채용공고가 직무 수요와 기술 요구를 시의성 있게 파악할 수 있는 노동시장 정보원으로 평가하였다[6][7].

표 1. NCS 인공지능 세분류별 직무
Table 1. Job descriptions by NCS AI subcategory

Job label	Summary of job description
AI platform construction	Building and optimizing platforms, infrastructure, and interfaces for AI service implementation
AI service planning	Establishing AI service goals, analyzing requirements, and planning service models and scenarios
AI modeling	Securing and processing training data, extracting features, and developing and optimizing models
AI service operation management	Operating, monitoring, and maintaining the quality of deployed AI services
AI service implementation	Analyzing, designing, developing, testing, and deploying services based on modeling results
AI training data construction	Acquiring, storing, labeling, integrating, transforming, and validating the quality of training data
Generative AI engineering	Selecting and training generative AI models, and developing and validating customized products

Note. The job labels are author-translated from the official Korean NCS subcategory names; the official Korean labels were used as reference labels in the experiment.

국내 연구들도 채용공고를 활용하여 산업현장의 필요 역량을 분석하고, 교육과정과의 연계 가능성을 검토하였다. 예를 들어 SW 분야 채용공고 빅데이터를 통해 직업 및 직무 변화를 분석한 연구와 데이터 과학자 채용정보를 바탕으로 기업의 기술 역량 및 협업 역량 요구를 탐색한 연구는 채용공고가 산업 수요와 직무 역량 구조를 파악하는 실증 자료로 유용함을 보여준다[8][9][13]. 또한 구인광고와 NCS를 비교하여 정보보호 직무의 필요 지식과 기술을 분석한 연구는 채용공고와 국가 표준 직무 체계 간 비교 가능성을 보여주었다[10].

한편 채용 분야에서 NCS 기반 직무 중심 채용이 확산되면서, 관련 연구들은 NCS의 채용 적용 실태와 능력중심 채용모델의 개발·개선 방향을 검토하였다[14]-[16]. NCS 활용 실태를 분석한 연구는 공공기관과 민간기업이 채용 분야에서 NCS를 직무기술서와 직무 단위의 기준으로 사용하는 양상을 분석하였다[14]. 직무능력 채용 표준화 도구 개발·개선 수요조사 연구는 NCS 소분류 기준 직군을 대상으로 신규 및 고도화 직군 수요를 조사하고, 능력중심 채용모델의 정비 방향을 제시하였다[15]. 또한 NCS 직무 중심 채용제도의 활용요인과 고용성과를 분석한 연구는 공공기관 중심의 NCS 기반 능력중심 채용 확산과 고용성과 간 관계를 분석하였다[16].

그러나 기존 접근은 주로 직무기술서 작성, 직무분류, 평가요소 설계 수준에서 채용 직무를 구조화하는 데 강점을 가지지만, 실제 채용공고 본문의 자유서술형 특성, 기업별 표현 차이, 복합 직무 구성, 최신 기술 용어의 빠른 변화까지 충분히 반영하기에는 한계가 있다[11][17]. 특히 기업은 동일하거나 유사한 직무라도 서로 다른 용어와 기술 스택으로 표현하므로, 채용공고를 표준 직무 체계와 직접 연결하려면 기존의 키워드·텍스트마이닝 기반 접근을 넘어 문맥과 의미 유사도를 반영하는 보다 정교한 자동 분석 방법이 필요하다[10][11].

2.3 자연어 처리 기반 직무 분류 및 의미 매핑 연구

채용공고는 대부분 비정형 자연어 텍스트이므로, 최근 직무 분류와 스킬 추출 연구는 자연어 처리 기반 접근을 활발히 활용하고 있다. 특히 직무 및 역량 정보는 동일한 의미를 서로 다른 표현으로 나타내거나, 동일한 용어도 맥락에 따라 다르게 해석되므로 단순 키워드 일치보다 문맥과 의미를 반영한 방법이 필요하다[11][12].

최근 연구들은 채용공고 분류에서 다국어 임베딩과 LLM 기반 접근의 가능성을 보고하였다[11]. 또한 온라인 채용공고를 활용하여 데이터 사이언티스트 직무 역량을 탐색한 연구도 있다[18]. 한국어 문장 유사도 측정을 위한 Sentence-BERT 연구는 문장 임베딩 기반 접근이 한국어 의미 비교 문제에 유효함을 제시하였다[19]. 이러한 연구 흐름은 자연어 처리 기술이 비정형 채용공고 텍스트를 직무 정보로 구조화하고 분류하는 데 효과적임을 보여준다.

그러나 채용공고와 NCS 인공지능 세분류를 연결하는 과제는 단순한 스킬 추출이나 일반적 직무 분류와 차이가 있다. 산업현장의 최신 수요를 반영한 채용공고를 국가표준직무체계와 연결해야 하므로, 직무 매핑 과정에서 산업 수요와 국가 표준 간 정렬 가능성을 함께 검토해야 한다. 또한 이러한 직무 매핑은 NCS 기반 인공지능 직무훈련 체계와 산업 수요 간 정합성을 검토하는 기초 절차에 해당한다. 채용공고는 실무 용어와 최신 기술 스택을 포함한 자유서술형 텍스트인 반면, NCS는 표준화된 직무 체계를 따른다. 이로 인해 두 체계 간에는 표현과 의미의 간극이 존재한다. 따라서 채용공고와 NCS 간 자동 매핑은 단순 키워드 대응이 아니라 의미 기반 정렬 문제로 접근해야 한다. 이를 위해 Ko-SBERT와 GPT-4o-mini를 활용하여 자동 매핑 가능성과 접근별 성능 특성 및 오류 양상 차이를 비교한다.

2.4 NCS와 채용공고 간 어휘 간극 문제

NCS는 직무를 대분류, 중분류, 소분류, 세분류 및 능력단위 중심으로 체계화한 국가 표준 체계인 반면, 실제 채용공고는 기업의 사업 내용, 프로젝트 맥락, 기술 트렌드, 현장 용어를 반영하는 자유서술형 텍스트이다. 따라서 동일하거나 유사한 직무를

설명하더라도, NCS는 표준화된 직무명과 능력 중심 표현을 사용하고, 채용공고는 기술 스택과 실무 중심 용어를 주로 사용하므로 표현상의 차이가 발생한다[18][20].

예를 들어 채용공고는 특정 도구, 프레임워크, 서비스명, 최신 기술 용어를 직접 제시하는 반면, NCS는 이를 일반화된 직무 개념으로 기술하는 경향이 있다. 이러한 차이는 단순 문자열 일치 방식만으로 두 체계를 정확히 연결하기 어렵게 만든다. 즉, 채용공고와 NCS 간 자동 매핑은 어휘 수준의 대응이 아니라 의미 수준의 매핑 문제로 접근해야 한다[11][12].

종합하면, 기존 연구들은 채용공고를 활용한 직무 및 역량 분석 가능성, 자연어 처리 기반 분류 및 의미 비교의 유효성, NCS와의 비교 필요성을 제시하였다. 그러나 채용공고와 NCS 인공지능 세분류 간 자동 매핑을 수행하고, 임베딩 기반 접근과 LLM 기반 접근의 성능을 직접 비교한 연구는 아직 제한적이다. 따라서 채용공고와 NCS 인공지능 세분류 간 의미 기반 자동 매핑 가능성을 세분류 수준에서 실험적으로 검토할 필요가 있다.

III. 연구 방법

3.1 연구 설계 개요

인공지능 관련 채용공고를 NCS 인공지능 세분류 7개 직무에 자동 매핑할 수 있는지를 검증하기 위해 비교 실험 연구를 설계하였다. 연구 절차는 데이터 수집 및 정제, 기준 라벨 구축, 의미 기반 자동 매핑, 성능 평가의 4단계로 구성하였다. 먼저 고용24를 통해 인공지능 관련 채용공고를 수집하고, 중복 제거와 텍스트 정제를 거쳐 분석용 데이터셋을 구성하였다. 다음으로 NCS 인공지능 세분류 7개 직무를 기준 라벨 체계로 설정하고, 전문가 검증을 통해 기준 라벨을 구축하였다. 이후 채용공고와 NCS 세분류 간 의미 기반 자동 매핑을 수행하기 위해 임베딩 기반 의미 유사도 접근과 LLM 기반 문맥 해석 접근을 비교하였다. Ko-SBERT는 한국어 문장 임베딩을 활용하여 채용공고 텍스트와 NCS 세분류 설명 간 의미적 유사도를 산출하는 방식으로 적용하였다. GPT-4o-mini는 NCS 세분류 설명과 채용공고 텍스트를 함께 고려하여 문맥 기반 분류를 수행하는 방식으로 적용하였다. 두 모델은 특정 모델의 우열을 일반화하기 위한 대상이 아니라, 채용공고 - NCS 매핑 과업에서 서로 다른 두 방법론적 접근의 적용 가능성과 한계를 비교하기 위한 기준으로 활용하였다.

성능 평가는 1:1 및 1:3 자동 매핑 조건으로 구분하였다. 1:1 매핑은 단일 세분류를 예측하는 엄격한 자동 분류 성능을 확인하기 위한 조건이며, 1:3 매핑은 전문가 검토나 교육과정 설계 과정에서 활용 가능한 후보 추천 성능을 확인하기 위한 조건이다. 마지막으로 Accuracy, Macro Precision, Macro Recall, Macro F1-score, Top-3 Accuracy와 confusion matrix를 활용하여 전체 성능, 세분류별 성능, 오분류 패턴을 비교·분석하였다.

이와 같은 설계는 자유서술형 채용공고와 구조화된 국가 표준 직무 체계 간의 의미 정렬 가능성을 실험적으로 검토하기 위한 절차이다. 또한 단순한 모델 성능 비교를 넘어, 산업 채용공고와 NCS 간 자동 매핑 과정에서 임베딩 기반 접근과 LLM 기반 접근이 보이는 성능 차이와 오류 특성을 함께 분석하는 데 초점을 두었다.

3.2 데이터 수집 및 구성

고용노동부 고용24의 채용공고 서비스를 대상으로 데이터를 수집하였다. 고용24는 고용노동부와 한국고용정보원이 운영하는 공공 고용서비스 채널로, 채용정보 상세검색, 직종별·지역별 채용정보, 테마별 채용관 등을 통해 다양한 일자리 정보를 제공한다[21]. 데이터 수집 기간은 2025년 12월부터 2026년 3월까지이며, NCS 인공지능 세분류 7개 직무를 기준으로 직무별 45건씩 총 315건의 채용공고를 수집하였다. 수집된 채용공고는 주요업무, 자격요건, 우대사항을 중심으로 구성하였으며, 중복 게시, 정보 누락, 세분류 판별 가능성 등을 기준으로 정제하였다. 실험 데이터의 구성은 표 2와 같다.

표 2. 실험 데이터 구성
Table 2. Experimental data composition

Item	Description
Unit of analysis	1 job posting
Initial number collected	315
Final sample size	284 (after preprocessing)
Collection period	Dec. 2025–Mar. 2026
Input fields	main duties, qualifications, preferred qualifications
Label system	seven AI subcategories under NCS
Exclusion criteria	duplicate postings, insufficient information, postings not classifiable into a subcategory

각 세분류의 직무명과 관련 기술 용어를 조합하여 채용공고를 검색하였다. 직무별 45건은 초기 검색 단계에서 세분류별 자료를 균형 있게 확보하기 위한 수집 기준이며, 최종 기준 라벨의 분포를 의미하지 않는다. 실제 채용공고는 제목이나 검색 키워드와 본문상의 핵심 직무가 일치하지 않는 경우가 있어, 최종 라벨은 전문가 검증 과정에서 본문 내용을 기준으로 확정하였다. 대표 검색 키워드는 표 3과 같다.

표 3. 데이터 수집 검색 키워드
Table 3. Search keywords for data collection

Job label	Search keywords
AI platform construction	AI infrastructure, model serving, AI systems, AI platforms
AI service planning	AI service planning, AI experience design
AI modeling	Model development, model tuning, AI engineer
AI service operation management	AI service ops, AI platform ops, AI system ops, MLOps
AI service implementation	AI service development, AI applications, model serving
AI training data construction	AI data, dataset development, data preprocessing
Generative AI engineering	LLM development, multimodal AI, LangChain, AI agent

Note. Job labels and search keywords are presented in their original form to ensure reproducibility and preserve meaning.

수집된 315건의 채용공고를 대상으로 중복 게시, 정보 누락, 세분류 판별 가능성을 점검하였다. 동일 기업, 동일 직무, 동일 본문을 가진 공고는 중복으로 간주하여 제거하였고, 주요업무·자격요건·우대사항 중 핵심 직무 정보가 지나치게 부족한 공고는 제외하였다. 그 결과, 최종 분석 데이터는 284건으로 확정하였다. 다만 온라인 채용공고를 산업 수요를 반영하는 분석 자료로 활용하는 과정에서 특정 플랫폼과 공개 채용 중심의 대표성 한계를 전제하였다.

3.3 전문가 검증 기반 기준 라벨 구축

기준 라벨은 수집된 채용공고를 NCS 인공지능 세분류 7개 직무에 전문가가 직접 매핑한 평가 기준 데이터이다. 두 명의 AI 실무 전문가가 284건의 채용공고 전체를 독립적으로 검토한 뒤, 각 공고에 가장 적합한 NCS 인공지능 세분류 1개를 선택하여 라벨을 부여하였다. 판단 기준은 직무명 자체보다 주요업무, 자격요건, 우대사항에 반복적으로 나타나는 핵심 수행 업무와 요구 역량에 두었다. 예를 들어 모델 설계, 학습, 평가가 중심인 경우에는 인공지능모델링으로, 요구사항 분석과 서비스 기획이 중심인 경우에는 인공지능서비스기획으로 우선 매핑하였다. 라벨 체계는 NCS 인공지능 세분류와 관련 직무 정의를 기준으로 설정하였다[5][20]. 복합 직무 공고의 경우에는 텍스트 전체에서 가장 중심적인 업무 비중을 차지하는 세분류를 대표 라벨로 부여하였다.

평가자 간 일치도는 Cohen's kappa 계수로 산출하였다. Cohen's kappa는 범주형 자료에서 두 평가자 간 일치도를 우연 일치를 보정하여 측정하는 대표적 지표이다[22][23]. 초기 독립 라벨링 결과, 전체 284건 중 267건이 일치하여 일치율은 94.01%로 산출되었다. Cohen's kappa 값은 0.929로, Landis와 Koch의 기준에 따르면 거의 완전한 일치 수준에 해당한다[23]. 이후 불일치 항목은 재검토 및 협의를 거쳐 최종 합의 라벨로 확정하고, 이를 최종 기준 라벨로 사용하였다. 기준 라벨 구축 과정에는 SW 개발 및 기획 경력 12년, K-digital AI 분야 강의 경력 6년 이상인 전문가 1인과 SW 개발 및 AI 기반 서비스 개발 경력 20년 이상인 전문가 1인이 참여하였다. 최종 기준 라벨의 세분류별 분포는 표 4와 같다.

표 4. 기준 라벨 데이터셋 구축
Table 4. Reference label dataset composition

Job label	Count	%
AI platform construction	28	9.86
AI service planning	44	15.49
AI modeling	30	10.56
AI service operation management	27	9.51
AI service implementation	55	19.37
AI training data construction	40	14.08
Generative AI engineering	60	21.13
Total	284	100.0

Note. Job labels follow the English labels presented in Table 1.

초기 데이터는 직무별로 45건씩 균등하게 수집하였으나, 전문가 라벨링 결과 실제 직무 내용에 따라 클래스별 표본 수가 달라졌다. 이는 채용공고의 제목이나 검색 키워드만으로 실제 직무 세분류를 정확히 판별하기 어렵고, 본문 기반의 의미 판단이 필요함을 시사한다.

3.4 데이터 전처리

분석 전에 채용공고 텍스트를 전처리하였다. 먼저 특수문자, 중복 공백, 불필요한 줄바꿈을 제거하고, 주요업무·자격요건·우대사항을 하나의 문자열로 통합하였다. 다음으로 'Python', 'PyTorch', 'LLM', 'RAG'와 같은 기술 용어는 의미 손실을 줄이기 위해 원문 표기를 유지하였고, 단순 채용 안내 문구나 복지 정보처럼 직무 매핑과 직접 관련이 낮은 표현은 제외하였다. 한국어와 영어가 혼합된 경우에도 원문을 유지하되, 지나치게 짧은 약어나 의미 해석이 어려운 표기는 최소한으로 정규화하였다. 전처리는 키워드 수준의 단순 일치보다 문장 수준의 의미 보존을 우선하여, 임베딩 기반 비교와 LLM 기반 분류에 적합한 입력 데이터를 구성하였다.

Ko-SBERT는 문장 간 의미 유사도 계산에 강점을 가지며[12], GPT-4o-mini는 긴 입력 문맥을 처리할 수 있는 LLM이다[24]. 이에 따라 과도한 분절을 피하고 문맥 보존 중심의 전처리 전략을 적용하였다. 표 5는 채용공고의 전처리 결과 예시이다. 채용공고 원문에서 불필요한 표현을 제거하고 주요업무, 자격요건, 우대사항을 하나의 입력문으로 통합한 결과를 제시하였다.

표 5. 채용공고 텍스트 전처리 결과 대표 예시
Table 5. Representative examples of preprocessing results for job posting text

No	Job label	Example of preprocessing
1	AI platform construction	[Key Responsibilities] Design a shared GPU platform (GPUaaS) that enables multiple teams to efficiently utilize GPU resources, and develop a reliable API gateway and security filtering system to handle high volumes of AI model requests. Prepare technical presentations and roadmaps based on the proposed architecture for effective communication with clients. [Qualifications] Experience leading infrastructure design or public-sector SI projects as a Project Leader (PL), with an in-depth understanding of Kubernetes (K8s) and containerized environments. Experience building systems in closed-network environments with stringent network security requirements.
2	AI modeling	[Key Responsibilities] Design, train, evaluate, and improve medical imaging AI models, enhance AI model R&D efficiency using ML frameworks, and analyze research trends in medical AI. [Qualifications] Proficiency in developing AI models with PyTorch, experience with development and code management in Git- and cloud-based collaborative environments, and the ability to document work outcomes and ideas and communicate effectively in English. [Preferred Qualifications] Strong interest in the latest medical imaging AI research and development trends, clear communication skills based on logical thinking, and the ability to efficiently develop and validate models using AI tools.

Note. Job labels are shown in English, while examples are retained in Korean as model input texts.

3.5 자동 매핑 모델 설계

채용공고 - NCS 자동 매핑 과업에서 임베딩 기반 의미 유사도 접근과 LLM 기반 문맥 해석 접근의 적용 가능성을 비교하기 위해 Ko-SBERT와 GPT-4o-mini

를 선정하였다. Ko-SBERT는 채용공고 텍스트와 NCS 세분류 설명 간 의미적 유사도를 정량적으로 산출하는 임베딩 기반 접근으로 활용하였고, GPT-4o-mini는 채용공고 텍스트와 NCS 세분류 설명을 함께 고려하여 문맥 기반 분류를 수행하는 LLM 기반 접근으로 활용하였다. 두 모델은 현존 모델 전반의 우열을 일반화하기 위한 대상이 아니라, 동일 과업에서 서로 다른 방법론적 접근의 성능 특성과 오류 양상을 비교하기 위한 기준으로 설정하였다.

Ko-SBERT 기반 임베딩 접근을 구현하기 위해 Hugging Face에 공개된 한국어 Sentence-BERT 계열 모델인 jhgan/ko-sbert-nli를 사용하였다[25]. GPT-4o-mini 기반 매핑에는 OpenAI API의 gpt-4o-mini 모델을 사용하였다[24]. 모든 채용공고에 동일한 프롬프트 구조를 적용하였으며, 출력 변동성을 최소화하기 위해 temperature를 0.0으로 설정하였다. 프롬프트에는 NCS 인공지능 세분류 7개의 라벨명, 직무 정의, 핵심 업무와 채용공고 입력문을 함께 제시하고, 출력은 사전에 정의한 7개 세분류 라벨 중 하나 또는 상위 3개 후보만 산출하도록 제한하였다. 출력 라벨이 사전에 정의한 라벨명과 부분적으로 불일치하는 경우에는 기준 라벨 체계에 따라 정규화하였다.

Ko-SBERT는 Sentence-BERT 구조를 기반으로 한국어 문장을 밀집 벡터로 변환하고, 문장 간 의미 유사도 계산에 활용하는 문장 임베딩 모델이다[12][19]. GPT-4o-mini는 비용 효율성과 문맥 처리 능력을 고려하여 LLM 기반 비교 모델로 선정하였으며, 복합 업무와 다양한 기술 표현이 혼재된 채용공고를 문맥 수준에서 해석하고 세분류를 판별하는 데 활용하였다[24]. 실험의 공정성을 확보하기 위해 두 모델에 동일한 채용공고 입력 텍스트와 동일한 NCS 참조 정보를 적용하였다. NCS 참조 정보는 세분류명, 직무 정의, 핵심 업무를 결합한 텍스트로 구성하였다. 또한 GPT-4o-mini 입력 과정에서 채용공고 텍스트를 별도로 영어로 번역하지 않고, 한국어 원문과 채용공고에 포함된 영문 기술 용어를 함께 유지하였다.

Ko-SBERT 기반 매핑에서는 채용공고 텍스트와 각 NCS 세분류 참조문을 문장 임베딩으로 변환한 뒤, 코사인 유사도를 계산하였다. 이후 가장 높은 유사도를 가진 세분류를 1:1 매핑의 예측 라벨로 선택하고, 유사도 순위 상위 3개 세분류를 1:3 매핑의 Top-3 후보로 추출하였다[12]. GPT-4o-mini 기반 매핑은 채용공고 텍스트와 NCS 세분류 참조 정보를 프롬프트에 함께 제시하고, 가장 적합한 세분류 1개 또는 상위 3개 세분류를 지정된 형식으로 산출하였다. 프롬프트는 입력 형식을 일관되게 유지하고 자유서술을 최소화하도록 구성하였으며, 복합 업무가 포함된 공고는 전체 문맥에서 중심 직무를 기준으로 분류하였다. 이를 통해 1:1 매핑과 1:3 매핑 결과를 산출하였다. 자동 매핑 모델 설계의 요약은 표 6에 제시하였다.

표 6. 자동 매핑 모델 설계
Table 6. Automatic mapping model design

Model	Ko-SBERT	GPT-4o-mini
Input	Job posting text + NCS reference text	
Output	Top-1 and Top-3 subcategories	
Processing method	Cosine similarity calculation after sentence embedding	Prompt-based classification and candidate inference
Strengths	Consistent similarity calculation, high reproducibility	Contextual reasoning, ability to handle complex expressions

3.6 실험 설계

실험은 1:1 매핑과 1:3 매핑으로 구분하였다. 1:1 매핑은 각 채용공고에 대해 가장 적합한 NCS 세분류 1개를 예측하고, 이를 기준 라벨과 비교하여 일치 여부를 평가하는 방식이다. 이는 자동 분류 시스템의 엄격한 단일 라벨 예측 성능을 확인하기 위한 조건이다. 반면 1:3 매핑은 모델이 제시한 상위 3개 후보 세분류 안에 기준 라벨이 포함되는지를 평가하는 방식이다. 채용공고는 하나의 세분류로 명확히 구분되기보다 복수의 직무 요소를 포함하는 경우가 많고, NCS 인공지능 세분류 간에도 의미적으로 인접한 영역이 존재한다. 따라서 1:3 매핑은 단일 자동 분류 성능이 아니라, 전문가 검토나 교육과정 설계 과정에서 활용 가능한 후보 추천 성능을 확인하기 위한 조건으로 설정하였다. 7개 클래스에서

Top-3는 비교적 완화된 기준이므로, 최종 분류 성능이 아니라 후보군 제시 가능성을 평가하는 보조 지표로 해석하였다. 표 7은 자동 매핑 실험 조건을 정리한 것이다.

실험의 공정성을 확보하기 위해 Ko-SBERT와 GPT-4o-mini에 동일한 전처리 텍스트와 NCS 참조 체계를 적용하였다. 또한 별도의 파인튜닝 없이 사전학습 모델을 그대로 사용하여 추가 학습 데이터의 영향을 배제하고, 동일 조건에서 임베딩 기반 접근과 LLM 기반 접근의 성능을 비교하였다. 성능 비교에서는 전체 성능, 세분류별 성능, confusion matrix, 오류 유형을 함께 분석하였다. 이를 통해 오분류가 집중되는 세분류 쌍과 직무 경계의 모호성이 크게 나타나는 채용공고 유형을 파악하였다.

표 7. 실험 조건
Table 7. Experimental settings

No	Model	Evaluation method	Description
E1	Ko-SBERT	1:1	Comparison between the Top-1 prediction and the reference label
E2	GPT-4o-mini	1:1	
E3	Ko-SBERT	1:3	Evaluation of whether the reference label is included in the Top-3 candidates
E4	GPT-4o-mini	1:3	

3.7 평가 지표

평가 지표는 정확도(Accuracy), 정밀도(Macro Precision), 재현율(Macro Recall), Macro F1-score, Top-3 Accuracy를 사용하였다. Accuracy는 전체 채용공고 중 기준 라벨을 정확히 예측한 비율을 의미하며, 두 접근 방식 간 직관적인 성능 비교에 유용하다. 그러나 다중 범주 분류 과제에서 세분류별 표본 수가 동일하지 않은 경우, Accuracy만으로는 클래스별 성능 차이를 충분히 설명하기 어렵다. 따라서 각 세분류를 동일한 중요도로 반영하기 위해 Macro Precision, Macro Recall, Macro F1-score를 함께 산출하였다. Precision, Recall, F1-score는 정보검색과 분류 문제에서 널리 사용되는 기본 평가 지표이다. Macro 평균은 소수 클래스의 성능을 함께 반영하므로 클래스 분포가 불균형한 다중 범주 분류 문제에 적절하다[26].

1:3 매핑 평가에는 Top-3 Accuracy를 사용하였다. Top-3 Accuracy는 모델이 제시한 상위 3개 후보 세분류 안에 기준 라벨이 포함되는 비율을 의미한다. 이는 최종 단일 분류 성능을 평가하기 위한 지표라기보다, 세분류 간 의미적 인접성을 고려하여 후보 추천 성능을 확인하기 위한 보조 지표이다. 또한 1:1 매핑 결과에 대한 혼동행렬을 분석하여 세분류 간 오분류 양상을 확인하였다. 혼동행렬은 단순 수치 지표만으로 파악하기 어려운 오분류 패턴과 클래스 간 경계 문제를 시각적으로 제시한다. 이를 통해 전체 성능뿐 아니라 세분류별 혼동 양상을 함께 분석하고, 채용공고와 NCS 간 의미 정렬의 가능성과 한계를 논의하였다[26].

IV. 실험 결과 및 분석

4.1 전체 성능 비교

표 8은 임베딩 기반 의미 유사도 접근과 LLM 기반 문맥 해석 접근의 전체 성능 비교 결과이다.

표 8. 전체 성능 비교 결과
Table 8. Overall performance comparison by mapping condition

Model	Evaluation method	A	MP	MR	MF	T3A
Ko-SBERT	1:1	0.412	0.545	0.408	0.391	-
GPT-4o-mini	1:1	0.567	0.640	0.572	0.557	-
Ko-SBERT	1:3	-	-	-	-	0.842
GPT-4o-mini	1:3	-	-	-	-	0.838

A: Accuracy, MP: Macro Precision, MR: Macro Recall, MF: Macro F1-score, T3A: Top-3 Accuracy

1:1 매핑(Top-1)에서는 GPT-4o-mini 기반 LLM 접근의 성능이 Accuracy 0.567, Macro F1-score 0.557로 나타나, Ko-SBERT 기반 임베딩 접근의 Accuracy 0.412, Macro F1-score 0.391보다 상대적으로 우수하였다. 다만 1:1 매핑의 절대 성능은 완전 자동 분류보다는 후보 추천 및 전문가 검토를 결합하는 방식

이 더 적절함을 시사한다. 이는 단일 세분류를 선택하는 엄격한 분류 조건에서 LLM 기반 문맥 해석 접근이 채용공고의 복합적 직무 맥락을 해석하는 데 상대적으로 유리하게 작용했을 가능성을 보여준다. 반면 1:3 매핑에서는 Top-3 Accuracy가 Ko-SBERT 기반 임베딩 접근이 0.842, GPT-4o-mini 기반 LLM 접근이 0.838을 기록하였다. 두 접근 방식 모두 비교적 높은 후보 추천 성능을 보였으며, 두 접근 방식 간 차이는 크지 않았다. 이는 단일 라벨 자동 분류에서는 LLM 기반 접근이 상대적으로 높은 성능을 보였지만, 전문가 검토에 활용 가능한 후보군 제공 관점에서는 임베딩 기반 접근도 유사한 수준의 활용 가능성을 보여준다. 따라서 1:1 매핑 결과는 자동 분류 성능의 차이를, 1:3 매핑 결과는 후보 추천 방식의 실무적 활용 가능성을 보여주는 결과로 해석할 수 있다.

4.2 세분류별 성능 분석

표 9는 세분류별, 모델별 성능 비교 결과를 나타낸다. Ko-SBERT 기반 임베딩 접근은 인공지능학습데이터구축(F1=0.699)에서 가장 높은 성능을 보였고, 인공지능플랫폼구축(F1=0.466)과 인공지능모델링(F1=0.473)에서도 비교적 안정적인 성능을 나타냈다. 반면 GPT-4o-mini 기반 LLM 접근은 인공지능서비스기획(F1=0.876), 인공지능모델링(F1=0.658), 인공지능서비스운영관리(F1=0.565), 인공지능서비스구현(F1=0.533)에서 Ko-SBERT 기반 임베딩 접근보다 높은 성능을 보였다. 이는 세분류별 직무 특성과 채용공고의 표현 구조에 따라 두 접근 방식의 성능 차이가 다르게 나타남을 보여준다.

세분류별 오차 양상은 Precision과 Recall의 차이에서도 확인된다. Ko-SBERT는 인공지능서비스기획에서 Precision 0.800, Recall 0.182로 나타나, 해당 세분류로 예측한 경우의 정확성은 높았으나 실제 인공지능서비스기획 공고를 충분히 포착하지 못하였다. GPT-4o-mini는 인공지능학습데이터구축에서 Precision 1.000, Recall 0.250으로 나타나, 해당 클래스로 예측한 사례는 정확했지만 실제 학습데이터 구축 공고 중 상당수를 다른 세분류로 오분류하였다. 이러한 결과는 세분류별 성능 차이가 표본 수뿐 아니라 채용공고의 표현 방식, 직무 내용의 명확성, 인접 세분류와의 의미적 중첩에 영향을 받았음을 시사한다.

표 9. 세분류별, 모델별 성능 비교
Table 9. Performance comparison by job label and model

Job label	Model	P	R	F1	N
AI platform construction	Ko-SBERT	0.378	0.607	0.466	28
	GPT-4o-mini	0.304	0.500	0.378	
AI service planning	Ko-SBERT	0.800	0.182	0.300	44
	GPT-4o-mini	0.867	0.886	0.876	
AI modeling	Ko-SBERT	0.520	0.433	0.473	30
	GPT-4o-mini	0.543	0.833	0.658	
AI service operation management	Ko-SBERT	0.600	0.111	0.188	27
	GPT-4o-mini	0.684	0.481	0.565	
AI service implementation	Ko-SBERT	0.571	0.145	0.232	55
	GPT-4o-mini	0.450	0.655	0.533	
AI training data construction	Ko-SBERT	0.674	0.725	0.699	40
	GPT-4o-mini	1.000	0.250	0.400	
Generative AI engineering	Ko-SBERT	0.275	0.650	0.386	60
	GPT-4o-mini	0.632	0.400	0.490	

P: Precision, R: Recall, F1: F1-score, N: sample size

4.3 1:1 매핑 결과의 혼동행렬 및 오류 분석

두 접근 방식의 오분류는 무작위로 분산되기보다 의미적으로 인접한 세분류 사이에 집중되는 경향을 보였다. 오분류 사례를 해석하기 위해 오류 유형을 유사 직무 혼동과 기타로 구분하였다. 유사 직무 혼동은 실제 라벨과 예측 라벨이 NCS 직무 정의 또는 수행 업무 측면에서 인접한 경우를 의미하며, 기타는 유사 직무 혼동으로 보기 어려운 잔여 사례를 의미한다. 그림 2와 그림 3은 1:1 매핑(Top-1) 결과에 대한 혼동행렬을 나타낸다.

혼동행렬을 기준으로 세분류별 오차 분포를 살펴보면, 주요 오류는 일부 인접 세분류 쌍에서 반복적으로 발생하였다.

Ko-SBERT 기반 임베딩 접근은 인공지능서비스구현과 인공지능서비스기획을 생성형AI엔지니어링으로 분류하는 경향이 두드러졌다. 이는 서비스 개발, 모델 적용, API 연동, LLM 활용 등 채용공고에 포함된 기술 표현이 생성형 AI 관련 직무 표현과 의미적으로 가깝게 나타났기 때문으로 해석된다.

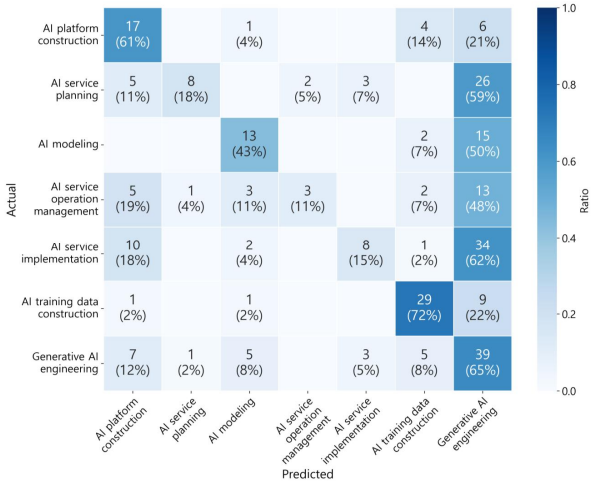

그림 2. Ko-SBERT의 혼동행렬
Fig. 2. Confusion matrix of Ko-SBERT

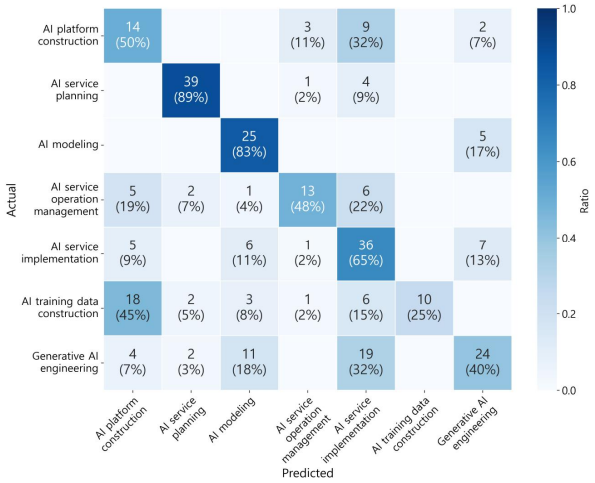

그림 3. GPT-4o-mini의 혼동행렬
Fig. 3. Confusion matrix of GPT-4o-mini

반면 GPT-4o-mini 기반 LLM 접근은 생성형AI엔지니어링을 인공지능서비스구현으로, 인공지능학습데이터구축을 인공지능플랫폼구축으로 분류하는 사례가 상대적으로 많았다. 이는 LLM 기반 접근이 문맥상 중심 업무를 해석하는 데 강점을 가지지만, 데이터 구축, 플랫폼 구축, 서비스 구현, 생성형 AI 개발 요소가 함께 포함된 복합 공고에서는 인접 세분류 간 경계를 일관되게 구분하는 데 한계가 있음을 보여준다.

Ko-SBERT 기반 임베딩 접근은 인공지능서비스구현을 생성형AI엔지니어링으로 분류한 사례가 가장 많았고(34건), GPT-4o-mini 기반 LLM 접근은 생성형AI엔지니어링을 인공지능서비스구현으로 분류한 사례가 가장 많았다(19건). 또한 인공지능학습데이터구축과 인공지능플랫폼구축 간 혼동도 반복적으로 나타났다. 표 10은 대표적인 오분류 사례를 정리한 것이다.

표 10. 주요 오분류 패턴(1:1 매핑 기준)
Table 10. Major misclassification patterns in 1:1 mapping

Model	Actual	Predicted	Freq.
Ko-SBERT	AI service implementation	Generative AI engineering	34
Ko-SBERT	AI service planning	Generative AI engineering	26
GPT-4o-mini	Generative AI engineering	AI service implementation	19
GPT-4o-mini	AI training data construction	AI platform construction	18

표 11의 오류 유형 분석 결과, Ko-SBERT 오분류의 67.1%, GPT-4o-mini 오분류의 66.7%가 유사 직무 혼동에 해당하였다.

표 11. 오류 유형별 건수와 비율(1:1 매핑 기준)
Table 11. Counts and proportions by error type in 1:1 mapping

Model	Similar job confusion	Other	Total
Ko-SBERT	112 cases (67.1%)	55 cases (32.9%)	167
GPT-4o-mini	82 cases (66.7%)	41 cases (33.3%)	123

이는 자동 매핑의 주요 어려움이 단순한 모델 성능 부족보다 채용공고의 복합 직무성, NCS 세분류 간 의미적 중첩, 기술 중심 표현과 직무 중심 표현 간 불일치에 있음을 보여준다. 본 실험에서는 채용공고를 별도로 영문 번역하지 않고 한국어 원문과 영문 기술 용어를 유지하였으므로, GPT-4o-mini의 오분류를 번역 과정의 의미 손실로 보기는 어렵다. 다만 데이터 규모, 전처리 수준, 한영 혼합 표현은 성능에 영향을 줄 수 있는 요인으로 남는다.

4.4 연구 결과 논의

이상의 결과는 채용공고 기반 NCS 인공지능 세분류 자동 매핑이 일정 수준 가능함을 보여준다.

1:1 매핑에서는 GPT-4o-mini 기반 LLM 접근이 Ko-SBERT 기반 임베딩 접근보다 상대적으로 우수한 성능을 보였다. 이는 단일 세분류를 선택하는 엄격한 분류 조건에서 LLM 기반 문맥 해석 접근이 채용공고의 복합적 직무 맥락을 해석하는 데 더 유리하게 작용했을 가능성을 시사한다. 반면 1:3 매핑에서는 두 접근 방식 모두 0.830 이상의 Top-3 Accuracy를 나타내어, 완전 자동 분류보다 전문가 검토와 결합한 후보 추천 방식에서 활용 가능성이 있음을 보여준다.

오분류 분석 결과, 상당수 사례가 유사 직무 혼동에 해당하였다. 이는 자동 매핑의 한계가 단순한 모델 성능 부족만이 아니라, 채용공고의 복합 직무성, NCS 세분류 간 의미적 중첩, 기술 중심 표현과 직무 중심 표현 간 불일치와 관련됨을 의미한다. 특히 생성형AI엔지니어링, 인공지능서비스구현, 인공지능플랫폼구축과 같이 기술 키워드와 수행 업무가 중첩되는 세분류에서는 단일 라벨 분류의 한계가 나타날 수 있다.

향후 NCS의 다른 세분류로 매핑 대상을 확대할 경우, 클래스 수 증가에 따라 인접 직무 간 혼동이 커질 가능성이 있다. 이를 완화하기 위해서는 NCS의 대분류 - 중분류 - 소분류 - 세분류 구조를 반영한 계층적 매핑 방식, 클래스 수와 활용 목적에 따른 Top-k 후보 추천 방식, 주요업무·자격요건·우대사항에 서로 다른 가중치를 부여하는 필드 가중 매핑 방식, 다중 라벨 기반 평가 체계를 함께 검토할 필요가 있다.

Ⅴ. 결론 및 향후 과제

채용공고 텍스트를 국가직무능력표준(NCS) 인공지능 세분류에 의미 기반으로 자동 매핑할 수 있는지를 검토하고, 임베딩 기반 의미 유사도 접근과 LLM 기반 문맥 해석 접근의 성능 차이를 비교하였다. 이를 위해 고용24에서 수집한 인공지능 관련 채용공고 315건 중 전처리를 거쳐 최종 284건을 분석 대상으로 구성하고, 전문가 합의 기반 기준 라벨을 구축하여 1:1 및 1:3 매핑 실험을 수행하였다. 실험 결과, 1:1 매핑에서는 GPT-4o-mini 기반 LLM 접근이 Accuracy 0.567, Macro F1-score 0.557로 Ko-SBERT 기반 임베딩 접근의 Accuracy 0.412, Macro F1-score 0.391보다 상대적으로 우수한 성능을 나타냈다. 반면 1:3 매핑에서는 Ko-SBERT와 GPT-4o-mini가 각각 0.842, 0.838의 Top-3 Accuracy를 기록하여 두 접근 방식 모두 비교적 높은 후보 추천 성능을 보였다. 세분류별 성능과 혼동 패턴을 분석한 결과, 생성형AI엔지니어링, 인공지능서비스구현, 인공지능플랫폼구축, 인공지능학습데이터구축 등 의미적으로 인접한 세분류 사이에서 혼동 사례가 반복적으로 발생하였다. 이는 세분류 간 직무 경계의 중첩이 자동 매핑의 주요 어려움으로 작용하며, 채용공고와 NCS 간 연결 문제가 단순 키워드 일치가 아니라 의미 기반 정렬의 문제에 해당함을 보여준다.

채용공고와 NCS 간 의미 정렬 가능성을 세분류 수준에서 실험적으로 검토하고, 임베딩 기반 접근과 LLM 기반 접근의 성능 차이 및 오류 특성을 비교하였다는 점에서 의의가 있다. 또한 자동 매핑 결과와 오분류 패턴을 분석하여 산업 수요와 국가 표준 직무 체계 간 대응 관계를 이해하기 위한 기초 자료를 제시하였다. 다만 분석 대상이 284건으로 제한되어 있고, 복합 직무 공고를 단일 세분류 라벨로 정리하였으며, 비교 대상 모델이 Ko-SBERT와 GPT-4o-mini에 한정되었다는 점은 한계로 남는다. 향후 연구에서는 데이터 규모를 확대하고, NCS 계층 구조를 반영한 단계적 매핑, Top-k 후보 추천, 다중 라벨 기반 매핑 체계, NCS 능력단위 및 직무기술서 정보를 반영한 정교한 매핑 방식을 검토할 필요가 있다. 이러한 후속 연구는 산업 수요 기반 AI 직무훈련 커리큘럼 설계 지원 체계를 고도화하는 데 활용될 수 있다.

References

[1] A. Green, "Artificial intelligence and the changing demand for skills in the labour market", OECD Artificial Intelligence Papers, No. 14, OECD Publishing, Paris, pp. 1-55, Apr. 2024. https://doi.org/10.1787/88684e36-en.

[2] P. Gmyrek, J. Berg, K. Kamiński, F. Konopczyński, A. Ładna, B. Nafradi, K.

Rosłaniec, and M. Troszyński, "Generative AI and jobs: A refined global index of occupational exposure", ILO Working Paper, No. 140, ILO, Geneva, pp. 1-72, May 2025. https://doi.org/10.54394/HETP0387.

[3] National Competency Standards (NCS), "What is NCS (National Competency Standards)?", Human Resources Development Service of Korea, https://www.ncs.go.kr/th01/TH-102-001-01.scdo. [accessed: Mar. 15, 2026]

[4] Korea Software Industry Association, "[2025] SQF-based Training Course Design Guide for National Key and Strategic Industry Occupation Training", https://www.sw.or.kr/common/board/Download.do?bcIdx=64865&cbIdx=308&streFileNm=000bf289-cb9e-4494-ad47-6cee56c3e976.pdf. [accessed: Mar. 15, 2026]

[5] National Competency Standards (NCS), "Sectoral Qualifications Framework (SQF)-based Training Course Design Guide for Artificial Intelligence SW Development", https://ncs.go.kr/web/sqf/img/guide01/02/인공지능SW개발.pdf. [accessed: Mar. 15, 2026]

[6] World Bank Group, "Real-time labor market data - The new frontier?", Knowledge4Jobs Newsletter: https://thedocs.worldbank.org/en/doc/d377e231649363d9fd9f07b6044d3131-0350072023/original/K4J-April-2023.pdf. [accessed: Mar. 15, 2026]

[7] Cedefop, "Online job vacancies and skills analysis: a Cedefop pan-European approach", Publications Office of the European Union, Luxembourg, pp. 1-24, Apr. 2019. https://doi.org/10.2801/097022.

[8] H. J. Kim and H. W. Kim, "A Curriculum Study to Strengthen AI and Data Science Job Competency", Informatization Policy, Vol. 28, No. 2, pp. 34-56, Jun. 2021. https://doi.org/10.22693/NIAIP.2021.28.2.034.

[9] S. J. Lee, J. Y. Kim, J. S. Kang, and J. H. Lee, "Research and practice-required competencies of People Analytics through text-mining", Journal of Human Resource Management Research, Vol. 31, No. 3, pp. 33-54, Sep. 2024. https://doi.org/10.14396/jhrmr.2024.31.3.33.

[10] H. J. Jun, T. S. Kim, and B. J. Park, "Information Security Job Skills Requirements: Text-mining to Compare Job Posting and NCS", Information Systems Review, Vol. 25, No. 3, pp. 179-197, Aug. 2023. https://doi.org/10.14329/isr.2023.25.3.179.

[11] H. Kavas, M. Serra-Vidal, and L. Wanner, "Enhancing Job Posting Classification with Multilingual Embeddings and Large Language Models", Proc. Tenth Italian Conf. on Computational Linguistics (CLiC-it 2024), Pisa, Italy, CEUR Workshop Proceedings, pp. 440-450, Dec. 2024. https://aclanthology.org/2024.clicit-1.53/.

[12] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks", Proc. 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int. Joint Conf. on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, pp. 3982-3992, Nov. 2019. https://doi.org/10.18653/v1/D19-1410.

[13] J. M. Kim and C. Hur, "Analysis of Occupational and Job Changes in the SW Field Based on Job Posting Big Data", Communications of the Korean Institute of Information Scientists and Engineers, Vol. 36, No. 9, pp. 15-21, Sep. 2018.

[14] J. S. Kim, "Analysis of NCS application for the Recruitment fields in Korea", Journal of Employment and Career, Vol. 11, No. 3, pp. 1-20, Sep. 2021. https://doi.org/10.35273/JEC.2021.11.3.001.

[15] J. S. Kim, "Demand survey for development and improvement of job competency recruitment standardization tools in Korea - Focusing on new and advanced job groups and needs of the '23 competency-based recruitment model -", Journal of Employment and Career, Vol. 13, No. 4, pp. 69-88, Dec. 2023. https://doi.org/10.35273/JEC.2023.13.4.004.

[16] J. Hong and J. Hur, "The Implementation of National Competency Standard(NCS) Recruitment

Practice: Antecedents and the Impact on its Employment Performance", Quarterly Journal of Labor Policy, Vol. 20, No. 2, pp. 73-103, Jun. 2020. https://doi.org/10.22914/JLP.2020.20.2.003.

[17] J. Napierala, V. Kvetan, and J. Branka, "Assessing the representativeness of online job advertisements", Cedefop Working Paper, No. 17, Publications Office of the European Union, Luxembourg, pp. 1-26, Dec. 2022. https://doi.org/10.2801/807500.

[18] X. Jin and S. I. Baek, "Exploring the Job Competencies of Data Scientists Using Online Job Posting", The Journal of Society for e-Business Studies, Vol. 27, No. 2, pp. 1-20, May 2022. https://doi.org/10.7838/jsebs.2022.27.2.001.

[19] B. M. Kim and S. B. Park, "Sentence BERT for Measuring Korean Sentence Similarity", Proc. Korea Software Congress 2020, pp. 1376-1378, Dec. 2020.

[20] National Competency Standards (NCS), NCS and Learning Module Search, https://www.ncs.go.kr/unity/th03/ncsSearchMain.do?tabNo=1. [accessed: Mar. 15, 2026]

[21] Ministry of Employment and Labor and Korea Employment Information Service, Detailed Job Search, https://www.work24.go.kr/wk/a/b/1200/retriveDtlEmpSrchList.do. [accessed: Mar. 15, 2026]

[22] J. Cohen, "A Coefficient of Agreement for Nominal Scales", Educational and Psychological Measurement, Vol. 20, No. 1, pp. 37-46, Apr. 1960. https://doi.org/10.1177/001316446002000104.

[23] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data", Biometrics, Vol. 33, No. 1, pp. 159-174, Mar. 1977. https://doi.org/10.2307/2529310.

[24] OpenAI, GPT-4o mini: advancing cost-efficient intelligence, Jul. 2024. https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/. [accessed: Mar. 15, 2026]

[25] jhgan, "jhgan/ko-sbert-nli", Hugging Face, https://huggingface.co/jhgan/ko-sbert-nli. [accessed: Mar. 15, 2026]

[26] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks", Information Processing & Management, Vol. 45, No. 4, pp. 427-437, Jul. 2009. https://doi.org/10.1016/j.ipm.2009.03.002.

저자소개

김 경 하 (Kyongha Kim)

1998년 2월 : 국립금오공과대학교 컴퓨터공학과(공학사)
2000년 2월 : 국립금오공과대학교 컴퓨터공학과(공학석사)
2025년 3월 ~ 현재 : 숭실대학교 IT정책경영학과 박사과정
2021년 8월 ~ 현재 : 고용노동부 KDT 전문강사
관심분야 : 인공지능, 자연어처리, 데이터마이닝, AI 에이전트, 온톨로지, SW공학

김 동 호 (Dongho Kim)

1990년 2월 : 서울대학교 전자공학과(공학사)
1992년 2월 : KAIST 전기 및 전자학과(공학석사)
2003년 1월 : George Washington University 전산학과(공학박사)
2003년 9월 ~ 현재 : 숭실대학교 글로벌미디어학부 교수
관심분야 : 가상현실, 메타버스, 스포츠 융합, 디지털콘텐츠

유 기 중 (Kijung Ryu)

2020년 2월 : 숭실대 정보과학대학원 소프트웨어공학(공학석사)
2023년 2월 : 숭실대학교 IT정책경영학과(공학박사)
2023년 11월 ~ 현재 : ㈜소와소 기업부설연구소 연구소장
2026년 3월 ~ 현재 : 숭실대학교 글로벌미디어학부 교수
관심분야 : 메타버스, 디지털트윈, SW공학, 인공지능, 빅데이터