



생성 AI를 활용한 NCS 직무능력 평가를 위한 자동문항생성 방법론 연구 : GPT4-o 기반 정보보호 분야 침해사고 분석 능력단위 중심

A Study on the Automatic Generation Methodology of NCS-based Job Competency Assessment Items Using Generative AI : Focused on GPT4-o based Information Security Analysis Competency Unit

이재식[†]
Jaesik Lee[†]

요약

본 연구는 대규모 언어모델인 ChatGPT-4o를 활용하여 NCS 정보보안 분야 능력단위의 직무수행능력 평가문항을 자동으로 생성하고 품질을 개선하는 방법론을 제안하였다. NCS 능력단위요소를 생성모델 기반으로 재정의하고, 평가문항 생성을 위한 프롬프트를 최적화하였으며, 자동 생성된 문항의 품질을 평가 및 개선하는 피드백 프로세스를 자동화하였다. 제안 방법론을 통해 실무 상황을 반영한 변별력 있는 평가문항을 자동으로 개발할 수 있음을 확인하였다. 본 연구는 NCS 기반 직무능력 평가의 고도화를 위한 생성AI 기술 활용 가능성을 제시하였다는 점에서 의의가 있다.

주제어 성 AI, NCS 직무능력, 역량평가, 자동문항생성, 침해사고 분석

ABSTRACT

This study proposed a methodology for automatically generating and improving the quality of job performance assessment items for NCS information security capability units using a large language model, ChatGPT-4o. The NCS competency unit elements were redefined based on the generative model, prompts for generating evaluation items were optimized, and the feedback process for evaluating and improving the quality of automatically generated items was automated. The proposed methodology confirmed the feasibility of automatically developing discriminative evaluation items reflecting practical situations. This study is meaningful in that it presented the possibility of utilizing generative AI technology for the advancement of NCS-based job competency assessment.

Keywords Generative AI, NCS job competency, Competency Evaluation, Automatic Question Generation(AQG), Incident Analysis

†정회원 한국인터넷진흥원 보안교육운영 팀장
논문투고 2024년 06월 10일
심사완료 2024년 07월 19일
게재확정 2024년 07월 24일
발행일자 2024년 08월 30일

1. 서론

1.1 연구 배경 및 필요성

정보보안 분야는 급변하는 사이버 위협 환경에 효과적으로 대응할 수 있는 실무 역량을 갖춘 인재 양성이 필요하다. 한국인터넷진흥원(KISA)의 ‘2023년 하반기 사이버 위협 동향 보고서’에 따르면, 2023년 침해사고 건수는 전년 대비 12% 증가한 1,277건을 기록했다. 특히 상반기에는 전년 동기 대비 40% 증가한 664건의 사고가 발생했다[1]. 이러한 통계는 사이버 공격이 점차 정교화되고 대규모화되는 추세를 보여주며, 정보보안 전문가의 역할, 특히 ‘침해사고 분석’ 능력의 중요성이 증대되고 있음을 보여준다.

국가직무능력표준(National Competency Standards, 이하 NCS)에서는 ‘침해사고 분석’과 같은 직무능력단위와 관련하여 국가적 차원에서 능력단위를 표준화하였다[2,3]. 공공기관 및 공기업을 중심으로 NCS기반의 직무능력 평가를 채용 과정에 활용되고 있으나, 평가 문항 개발 과정에서 시간, 비용, 전문 인력 등의 제약이 존재한다. 특히 최신 보안 위협과 기술을 반영한 평가 문항을 지속적으로 개발하고 개선하는 데는 어려움이 존재한다.

본 연구에서는 이러한 문제를 해결하기 위해 대규모 인공 지능 언어모델(Large Language Model, 이하 LLM)[4,5]을 활용한 자동 문항 생성 방법론을 제안한다. 이 방법론은 NCS 기반 직무능력 평가의 효율성과 전문성을 높이기 위해 문항생성을 자동화하여, 실질적인 침해사고 분석 및 대응 능력을 갖춘 전문 인력 양성을 위한 효율적인 평가 체계 개발에 기여한다.

1.2 연구 목적

본 연구는 대규모 언어모델(LLM) 중 하나인 OpenAI사의 최신형 모델인 ChatGPT-4o[5]를 활용한다. ChatGPT-4o는 멀티모달 성능이 뛰어나고, 특히 한국어 성능이 GPT-4 대비 91% 이상 향상되는 등 비영어권 언어 처리 능력이 크게 향상되어 다양한 분야에서 유용하게 활용될 수 있다[5].

본 논문에서는 OpenAI사의 최신형 모델인 GPT-4o를 활용하여 NCS 정보보안 분야 ‘침해사고 분석’ 능력단위[3]의 평가 문항을 자동 생성하고 품질을 개선하는 방법론을 제안한다. 구체적인 제안 방법론은 다음과 같다.

- 1) NCS 능력단위의 구성 요소를 분석하여 평가 문항 생성에 필요한 핵심 정보를 추출하고 구조화한다.
- 2) 구조화된 NCS 능력단위 정보를 바탕으로 ChatGPT-4o가 직무 지향적인 평가 문항을 생성하도록 유도하는 프롬프트를 설계한다.
- 3) 자동 생성된 평가 문항의 품질을 검증하고 개선 피드백을 도출하는 자동화 프로세스를 구현한다.
- 4) 3) 과정을 반복하여 생성된 문항의 품질을 개선하고, 최종 평가 문항을 도출하여 활용한다.

이를 통해 NCS 기반 직무능력 평가를 위해 AI 기술을 활용한 자동 문항 생성 방안을 제안하여, 정보보안 분야에 NCS 기반의 실무역량을 갖춘 인력이 양성될 수 있도록 기여하고자 한다.

2. 연구 관련 동향 분석

2.1 국가직무능력표준(NCS) 기반 직무능력 평가

국가직무능력표준(NCS)은 산업 현장에서 요구되는 직무 능력을 국가 차원에서 표준화한 것으로, 각 직무별로 필요한 능력단위와 수행준거, 지식·기술·태도 요소를 정의하고 있다[2]. 이러한 NCS의 특징을 반영하여 대학에서도 마이크로디그리 및 나노디그리 체계를 NCS 기반으로 설계하는 등 다양한 분야에 NCS의 활용이 시도되고 있다[6].

이처럼 NCS 기반 직무능력 평가는 이러한 표준에 따라 개인의 직무 수행 능력을 객관적으로 측정하고 향상시키는 데 활용될 수 있다. 그러나 실제 교육 현장에서는 NCS 기반 직무능력 평가에 어려움을 겪고 있다. 이는 NCS 학습 모듈이 능력단위별 필요 지식과 기술, 평가 방법 등을 제시하고 있으나, 내용이 추상적이고 모호하여 실제 활용이 어렵기 때문으로 추정된다. 따라서, NCS 학습모듈의 효과성 제고와 실무 역량을 갖춘 인재 양성을 위해서는 보다 구체적이고 실제적인 평가 방안 및 NCS 학습모듈 이용 방안이 필요하다.

본 논문에서 자동문항생성에 활용한 정보보안 분야는 NCS 체계에서 ‘20.정보통신’ 대분류 내 ‘01.정보기술’ 중 분류에 속하며, 다양한 능력단위로 세분화되어 있다[2]. 그 중 ‘침해사고 분석’ 능력단위는 정보보안 업무 수행에 필수적인 요소로, 침해사고 발생 시 원인과 과정을 분석하고 대응 방안을 마련하는 능력을 의미한다[3].

‘침해사고 분석’ 능력단위는 크게 ‘분석 기반 조성’, ‘침해사고 원인 분석’, ‘침해사고 대응 및 복구’ 등 3개의 세부 능력단위 요소로 구성된다. 각 항목은 3-4개의 수행준거와 이를 달성하는 데 필요한 지식, 기술, 태도 요소를 포함하고 있다[3]. 현재(24년 6월)를 기준으로 2001060305_19v2버전이 배포되고 있으나, 이 버전은 2019년에 만들어져 2022년에 보완 예정이었지만, 현재까지 개선 반영되지 않고 있다. 따라서 본 논문에서는 LLM을 활용한 자동화된 문제생성을 위해 LLM이 사전에 학습한 지식을 기반으로 기존의 NCS에서 정의된 능력단위요소 및 지식, 기술을 새롭게 구성하여 활용하였다. 태도와 관련된 요소는 온라인 기반의 자동화된 문제 생성을 활용하여 개선하거나 평가/측정하기 어려운 요소이며, 특히 성인을 대상으로 하는 교육과정에 적합하지 않은 요소로 판단하여 본 논문에서는 배제하였다.

2.2 대규모 언어 모델(LLM)과 자동 문항 생성

대규모 언어모델(LLM)은 방대한 텍스트 데이터로 사전 학습된 신경망 모델로, 문맥을 이해하고 자연스러운 텍스트를 생성하는 능력을 갖추고 있다. 딥러닝 기술의 발전으로 Transformer 모델[7]이 등장하였고, 이를 활용한 GPT(Generative Pre-trained Transformer) 모델이 대표적인 예로, 토큰 단위의 자기회귀적 언어모델링 방식으로 학습되며 다양한 자연어처리 태스크에 전이학습될 수 있다. 최근에는 사전학습된 대규모 언어모델을 미세조정하거나 프롬프트로 제어하는 방식으로 출력의 품질을 높이는 연구가 활발히 진행되고 있다[8]. 이처럼 GPT 모델은 지속적으로 고도화되어 최근 ChatGPT-4o 모델에 이르러서는 인간 수준에 근접하는 언어 이해와 생성 능력을 보여주고 있다[4,5].

자동 문항 생성(Automatic Question Generation, AQG)은 주어진 텍스트나 지식 베이스로부터 평가 문항을 자동으로 생성하는 기술로, 초기에는 규칙 기반 방식이나 질의응답(QA) 데이터로부터 템플릿을 추출하는 방식이 주로 활용되었다[9]. 문항 생성을 위해 핵심 용어 추출, 알은 구문 분석, WordNet 등의 NLP 기술과 언어자원을 활용하여 문제와 오답지를 자동 생성하고, 사용자 사후 편집을 허용하는 방식을 사용하여 자동화 할 경우 문항 생성 시간이 74% 이상 단축된다는 연구 결과도 있다[10]. 기존의 AQG 기술은 주로 텍스트 기반 문항 생성에 국한되어 템플릿 기반의 문항 복제 수준에 머물렀으나, 최근 연구에서는 생성AI를 활용하여 동적 자동평가 체계를 통해 학습자 개개인에 맞춘 개인화된 평가 문항 생성과 관련된 방안도 제시되고 있다[11]. Kiyak, Y. S.[12]는 ChatGPT를 통한 자동생성 질문들이 높은 성취도를 보이는 학생들과 낮은 성취도를 보이는 학생들을 효과적으로 구별할 수 있음을 보여주었으며, 이는 ChatGPT가 시험 문제 개발에서 인공지능 도구로서의 잠재력을 지니고 있음을 시사한다.

프로그래밍 코딩 분야에도 자동화된 코드 생성과 테스트에 관한 연구[13-15]도 진행되고 있는 등 다양한 분야에 있어서 자동화 문항생성과 관련된 연구가 활발히 진행되고 있다.

Seulki, Kim[14]은 생성AI를 통해 프로그래밍 교육에 필요한 코드 자동생성 연구에서 교수학습 전략과 프롬프트 엔지니어링 기법을 적용한 프롬프트를 개발하여 ChatGPT로 맞춤형 교육 자료를 생성하고, 100회 반복 실험과 CodeBERT 분석을 통해 개발된 프롬프트의 우수성(평균 0.940의 코사인 유사도)을 입증하였다. Doughty, J.[15]는 Python 프로그래밍 수업의 다지선다형 질문(MCQs)을 GPT-4와 인간이 만든 결과를 비교한 연구에서는 생성된 결과물의 품질이 비슷하며, 특히 학습 목표와 일치성은 GPT(82.9%)가 인간(67.3%)이 만든 것보다 더 잘 일치하는 것으로 보였다. 이는 전반적으로 생성AI를 활용한 문항 생성이 교육평가의 자동화에 활용될 수 있음을 보여준다.

본 논문에서는 ChatGPT-4o 시스템을 활용하여 NCS 기반 직무능력 평가문항을 자동 생성하는 방법론을 제안하였다. ChatGPT-4o는 대화형 인터페이스를 통해 사용자의 요구사항을 이해하고 관련된 문제를 생성할 수 있는 강력한 도구로 활용될 수 있다. ChatGPT-4o로 생성한 평가 문항의 품질을 자동으로 검증하고 개선하기 위해 전문가 관점에서의 문항 적절성을 평가하기 위해 4가지 평가 기준을 구성하여 평가 및 개선사항을 도출하고, 이를 다시 ChatGPT-4o에 프롬프트로 제공하여 문항의 오류를 진단하고 수정하는 피드백을 생성하도록 하였다. 이를 통해 전문가의 직접 개입 없이도 문항 품질을 자동으로 개선해 나갈 수 있는 방법론을 구현하고자 하였다.

3. NCS 평가 문항 자동 생성 방법론

3.1 방법론 개요

본 논문에서는 OpenAI사의 최신형 생성AI 모델인 ChatGPT-4o 언어모델을 활용하여 NCS 정보보안 분야 ‘침해사고 분석’ 능력단위의 평가 문항을 자동 생성하고 품질을 개선하는 방법론을 제안한다.

제안 방법론은 크게 NCS 능력단위 요소 구조화, 생성 문항 구조화 및 프롬프트 설계, 자동 생성 문항 품질 개선 프로세스의 단계로 구성된다.

3.2 NCS 능력단위 요소 구조화

평가 문항 자동 생성을 위해 우선 기존 NCS ‘침해사고 분석’ 능력단위인 2011060305_19v2버전의 구성 요소를 분석하고, LLM이 학습한 지식에 기반하여 능력요소를 새롭게 도출하였다. 새롭게 도출한 능력단위 구성요소는 각 수행준거 별로 관련된 지식과 기술을 매핑하였고, 평가에 필요한 평가기준도 능력단위요소 별로 새롭게 도출하였다. [표 1]은 새롭게 구성된 NCS ‘침해사고 분석’ 능력단위 구성요소 중 일부를 나타낸 예시이다. 새롭게 정의된 능력단위 구성요소를 JSON 형태로 구조화하여 LLM에 입력하고, 이를 기반으로 NCS 능력단위 구성요소 별로 수행준거에 기반한 지식과 기술을 측정할 수 있는 문제를 자동으로 생성할 수 있다. 구조화된 형태는 [표 2]와 같다.

Table 1. NCS ‘Analysis of Security Incidents’ Competency Unit Components(24v3)

Unit	Description	Knowledge	Skills
(2001060305_24v3.2) Analyzing the Causes of Security Incidents	(2.1) Understand the procedures for analyzing the causes of security incidents and apply effective methods for each procedure.	Procedures for analyzing security incidents, strategies for responding to security incidents	Ability to apply procedures for analyzing security incidents, ability to apply effective response methods

Unit	Description	Knowledge	Skills
(2001060305_24v3.2) Analyzing the Causes of Security Incidents	(2.2) Apply response methods for various types of security incidents and establish future response strategies based on the results of cause analysis.	Response methods for each type of security incident, measures to prevent recurrence	Ability to utilize the results of cause analysis, ability to establish future response strategies
	(2.3) Develop and implement follow-up measures to prevent recurrence after responding to security incidents.	Follow-up measures, procedures for reporting security incidents	Ability to develop follow-up measures, ability to implement follow-up measures
	(2.4) Report responses and cause analysis results of security incidents in accordance with domestic and international information security laws and regulations.	Domestic and international information security laws and regulations, methods for writing security incident analysis reports	Ability to write security incident reports, ability to comply with laws and regulations
(Evaluation Factors) Ability to understand and apply the procedures for analyzing the causes of security incidents, ability to apply response methods for various types of incidents, ability to develop and implement follow-up measures			

Table 2. NCS 'Incident Analysis' competency unit components (24v3) JSON structuring

```
{ "ability_units": {
  "2001060305_24v3": {
    "code": "2001060305_24v3",
    "name": "Incident Analysis",
    "definition": "Incident analysis refers to the ability to analyze causes, purposes, processes, and impacts of incidents when they occur, and to propose solutions.",
    "performance_criteria": {
      "2001060305_24v3.1": {
        "code": "2001060305_24v2.1",
        "name": "Establishing Analysis Basis",
        "description": {...},
        "knowledge": {...},
        "skills": {...}
      }, ...
    }
  }, ...
}
```

3.3 생성 문항 구조화 및 프롬프트 설계

평가 문항은 능력단위요소들의 '수행준거'에 명시된 지식과 기술을 평가하는 데 초점을 두고 구조화하였다. 문항 유형은 실무 역량 측정에 적합한 '객관식 4지선다형'으로 설정하였다.

ChatGPT-4o가 NCS 능력단위의 요구사항에 부합하는 문항을 출제하도록 유도하기 위해 입력 프롬프트를 최적화하였다. 생성형 AI를 활용할때 프롬프트에 대한 입력을 어

떻게 하느냐에 따라서 생성되는 데이터의 품질은 크게 변경된다. 이러한 프롬프트 입력에 관한 기술을 프롬프트 엔지니어링이라고 하며, 본 논문에서는 알려진 다양한 프롬프트 엔지니어링 기술을 활용하여, 최적화된 문항 생성이 이루어질 수 있도록 제안한다.

- 1) 페르소나 설정 : 프롬프트 입력시 생성AI의 역할인 페르소나를 명확히 명시해 주는 경우 성능이 향상된다고 알려져 있다. 본 논문에서는 문항 생성시에는 '침해사고 분석업무 수행경험이 많은 정보보안 분야 전문가이자 문제출제 전문가'로 문항 품질 평가시에는 '한국 NCS 정보보호 분야 전문가이자 문제출제 전문가'로 페르소나를 설정하였다.
- 2) 문항출제 가이드 : NCS 능력단위요소의 수행준거를 바탕으로 정보보안 전문가에게 필요한 지식/기술을 평가할 수 있게 작성하였다.
- 3) 추가 요구사항 : 수험자 간 변별력을 고려하고, 문제의 질문과 보기는 명확하고 이해하기 쉽게 작성되도록 작성하였다.
- 4) 보기구성 요구사항 : 보기만을 통해서 답안을 유추하기 어렵게 하고, 명백한 오답 보기 생성은 피하도록 작성하였다.
- 5) 문항품질 요구사항 : NCS 수행준거 연관성, 문제 난이도와 변별력, 문제 명확성과 이해도, 오답보기 품질 적절성 등 4개 항목의 품질요구사항을 만족하도록 문항생성을 요청하였다.

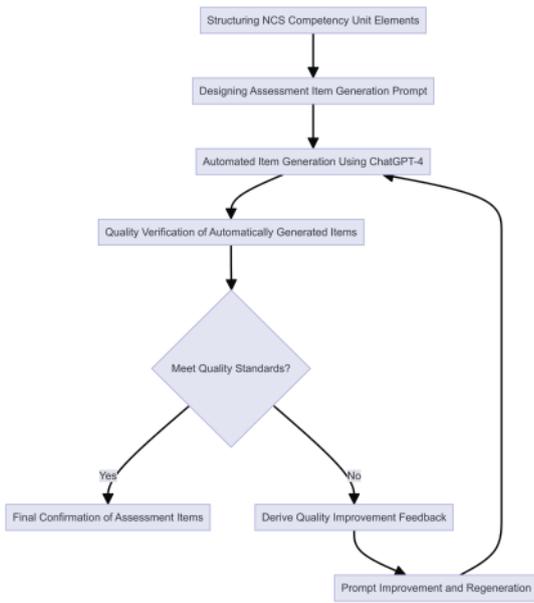
3.4 자동 생성 문항 품질 개선 프로세스

ChatGPT-4o로 생성된 평가 문항이 프롬프트에서 요구한 생성 기준에 따라 정확히 생성되었는지 확인이 필요하다. 생성AI는 환각과 같은 단점이 존재하며, 아직까지 성능측면에 있어서 인간을 대체할 수 있는 수준의 생성이 이루어지는지 100% 확신할 수 없기 때문이다. 따라서 본 논문에서는 AI 기반으로 자동으로 생성된 문항의 품질에 대해서 품질을 평가하고 개선하는 프로세스를 제안한다. 품질 평가 및 개선 또한 생성AI를 활용하여 자동화된 형태의 프로세스를 구현하였다. 이를 통하여, 문항의 생성부터, 검증, 개선 까지 하나의 사이클을 통해서 생성AI 기반의 자동 생성 문항에 대한 품질을 크게 개선할 수 있다.

문항 품질 검증을 위한 평가 기준과 피드백 방식은 다음과 같이 구성된다. 'NCS 수행준거와의 연관성', '문제의 난이도와 변별력', '문제의 명확성과 이해도', '오답 보기의 적절성' 등 4대 기준에 따라 생성 문항의 적절성을 각각 25점 척도씩 총 100점으로 진단하고, 감점 사유와 개선 방안을 제시하도록 한다.

[그림 1]는 이러한 피드백 기반 자동 문항 생성 및 품질 고도화 프로세스 전체를 도식화한 것이다. 자동 평가 결과를 바탕으로 프롬프트를 개선하고 문항을 재생성하는 과정을 반복함으로써, NCS 수행준거에 기반한 직무와 연계된 문항 생성 자동화 할 수 있도록 하였다.

Figure 1. Feedback-Based Automated Questions Generation and Quality Enhancement Process



4. 자동화 생성 및 결과

4.1 자동화 생성 실험 조건

제안 방법론의 효과성을 검증하기 위해 수정된 NCS 정보보안 분야 ‘침해사고 분석’ 능력단위(24v3)를 대상으로 평가 문항 자동 생성 실험을 수행하였다. 자동화 생성AI 모델은 ChatGPT-4o 버전을 활용하였고, 각 문항생성 및 개선 프롬프트는 매번 API 호출을 통해 새롭게 요청하였다. 이는 생성AI의 평가를 품질하는 다른 평가방법과 유사한 형태의 실험 진행 방법이다.

실험은 (1) NCS 능력단위 요소 기반 13개 문항 생성(프롬프트 개선 전), (2)프롬프트 개선 및 피드백 반응을 통한 13개 문항의 반복적 품질 고도화 의 2가지 절차로 진행되었다. 생성된 13개 문항은 반복적 품질 고도화를 7회까지 수행 하였으며, 프롬프트 개선 전 버전(NT0), 프롬프트 개선 버전(NT1), 피드백 반영 버전(NT2~NT8) 7개 등, 총 9개의 버전(NT0~NT8)으로 구성되었다.

4.2 자동화 생성 및 문항 품질 검증 수행

자동화된 문제 생성을 위하여 3.3절에서 제안한 문제 생성과 관련된 주요 구성요소를 반영하여 제안한 생성 프롬프트는 [표3]과 같다. 대괄호 {} 에 있는 영문변수는 3.2절에서 제시한 구조화된 JSON 형태의 NCS 능력단위구성요소들의 명칭이다.

Table 3. NCS-based assessment item generation prompt

<p>prompt = f"""</p> <p>You are a cybersecurity expert with extensive experience in incident analysis and a specialist in creating exam questions.</p> <p>This is the NCS-based meta-information required to create a multiple-choice question with 4 options.</p> <p>Ability Unit: {ability_unit['name']} ({ability_unit['code']})</p> <p>Performance Criteria: {performance_criteria['name']} ({performance_criteria['code']})</p> <p>Sub-Criteria: {sub_criteria_code}</p> <p>Description: {description}</p> <p>Required Knowledge: {', '.join(knowledge)}</p> <p>Required Skills: {', '.join(skills)}</p> <p>The question should be based on the given NCS performance criteria and should evaluate the knowledge and skills required of a cybersecurity expert. The question and options should be clear and easy to understand, and should be constructed with the intent to differentiate between examinees.</p> <p>There should be a total of 4 options, with only one correct answer, and care should be taken not to include clues that could reveal the correct answer. Incorrect options should be crafted in such a way that the correct answer is not easily inferred, and overly obvious wrong answers should be avoided.</p> <p>The generated question should be able to achieve a total of 100 points based on the following 4 evaluation criteria, each worth 25 points:</p> <ul style="list-style-type: none"> - Relevance to NCS performance criteria (25 points) - Difficulty and differentiation (25 points) - Clarity and comprehensibility (25 points) - Appropriateness of incorrect options (25 points) <p>Please output the generated question in the following JSON format:</p> <pre> {{ "question": "Question {question_number}: ({sub_criteria_code}) Question content", "choices": ["a) Option 1", "b) Option 2", "c) Option 3", "d) Option 4"], "answer": "Correct answer" }} """ </pre>
--

생성된 문항에 대한 문항 품질 검증 평가를 위하여 3.4절에서 제안한 문항 품질 검증 기준 및 피드백을 적용하여 문항을 생성하고 검증을 수행하였다.

[표4]는 이러한 문항 품질 검증 평가 기준에 대한 평가 및 개선을 위한 피드백 생성을 위한 프롬프트이다.

Table 4. Assessment and Improvement Prompt for Questions Quality Verification

```

prompt = f"""
Here is the generated question:
{question}
Here are the options:
{chr(10).join(choices)}
This question is related to the following NCS competency unit
element:
Competency Unit Element: {performance_criteria['name']}
({performance_criteria_code})
Sub-Criteria: {sub_criteria_code}
Description: {description}
Required Knowledge: {' '.join(knowledge)}
Required Skills: {' '.join(skills)}
You are an expert in the Korean NCS field of information security
and a specialist in creating exam questions.
Please evaluate the question based on the following 4 evaluation
criteria on a 25-point scale, and provide specific reasons for point
deductions and improvement suggestions for each item.
Relevance to NCS Performance Criteria (25 points):
Does the question accurately reflect the sub-criteria, knowledge,
and skill elements? (15 points)
Is the question appropriately constructed to measure the sub-
criteria? (10 points)
Specific reasons for point deductions:
Difficulty and Differentiation of the Question (25 points):
Is the difficulty level appropriate for an information security
expert? (10 points)
Does the question sufficiently differentiate between examinees
based on their results? (15 points)
Specific reasons for point deductions:
Clarity and Comprehensibility of the Question (25 points):
Are the question and options clearly and easily understandable? (15
points)
Is there any room for different interpretations of the question's answer
or intent from an expert's perspective? (10 points)
Specific reasons for point deductions:
Appropriateness of Incorrect Options (25 points):
Is the correct option clearly distinguishable from the incorrect
options? (10 points)
Is it difficult to infer the correct answer solely from the incorrect
options? (15 points)
Specific reasons for point deductions:
Please provide the overall evaluation result and feedback for
improvement in the following JSON format:
{{
  "Relevance to NCS Performance Criteria": 25,
  "Difficulty and Differentiation of the Question": 25,
  "Clarity and Comprehensibility of the Question": 25,
  "Appropriateness of Incorrect Options": 25,
  "Overall Evaluation Result (Total)": 100,
  "Feedback": "Include specific reasons for point deductions and
detailed improvement suggestions for each evaluation item."
}}
"""

```

평가 결과 도출된 개선사항은 프롬프트에 반영되어 문항을 재생하여 반복적으로 품질을 개선해 나가게 된다. [표 5]는 프롬프트를 통해 생성된 출력을 JSON포맷으로 구조화하여 나타낸 것이다.

Table 5. Example of JSON Structuring for Results of Automated Questions Quality Improvement Process (NT5-7)

```

{
  "question": "Question 7: (2.2.4) When reporting incident response
and analysis results according to domestic and international
information security laws and regulations, what is the most
appropriate method? Also, what items should be evaluated to
assess various impacts such as system and network performance
degradation and data loss?",
  "choices": [
    "a) The report should meet legal requirements, include detailed
analysis results and response procedures, and attach additional
materials (e.g., log files, technical reports) that evaluate impacts
such as system and network performance degradation and data
loss.",
    "b) The report should record only the response procedures
according to legal requirements, and the analysis results and
evaluation of impacts such as system and network performance
degradation and data loss should be reported verbally.",
    "c) The report should be detailed according to domestic and
international information security laws and regulations, including
response procedures, analysis results, and the evaluation of impacts
such as system and network performance degradation and data
loss, and additional materials should be attached if necessary.",
    "d) The report should be briefly written at the request of the
victim, minimizing legal requirements, and omitting the evaluation
of impacts such as system and network performance degradation
and data loss."
  ],
  "answer": "c) The report should be detailed according to domestic
and international information security laws and regulations,
including response procedures, analysis results, and the evaluation
of impacts such as system and network performance degradation
and data loss, and additional materials should be attached if
necessary.",
  "Relevance to NCS Performance Criteria": 20,
  "Difficulty and Differentiation of the Question": 20,
  "Clarity and Comprehensibility of the Question": 20,
  "Appropriateness of Incorrect Options": 20,
  "Overall Evaluation Result (Total)": 80,
  "Feedback": {
    "Relevance to NCS Performance Criteria": {
      "Score": 20,
      "Reason for Deduction": "While the question is generally
relevant to the performance criteria, the explanations for each
option are insufficient, making it difficult to clearly identify the
direct relevance to the performance criteria.",
      "Improvement Suggestion": "Each option should be described
more specifically to clarify its relevance to the performance criteria.
For example, specifying the specific impact of 'data loss' on the
organization would be helpful."
    },
    "Difficulty and Differentiation of the Question": {
      "Score": 20,
      "Reason for Deduction": "The difficulty of the question is
somewhat low, which may not be sufficient for differentiating
at an expert level. All options are somewhat general, making it
difficult to choose a specific answer.",
      "Improvement Suggestion": "To increase the difficulty, use

```

```

specific and professional terminology, and add realistic scenarios
to enhance differentiation at the expert level.”
},
“Clarity and Comprehensibility of the Question”: {
“Score”: 20,
“Reason for Deduction”: “While the question and options are
clear, some options are ambiguous, leaving room for interpretation.
For example, it is unclear how ‘illegal access attempts in the
network’ and ‘whether malware is detected’ specifically impact the
situation.”,
“Improvement Suggestion”: “Specify the options more clearly to
reduce ambiguity and improve the clarity of the question.”
},
“Appropriateness of Incorrect Options”: {
“Score”: 20,
“Reason for Deduction”: “Some incorrect options may be
confused with the correct option. For example, both ‘system
recovery time and post-recovery stabilization time’ and ‘degree of
system and network performance degradation’ could be considered
important factors.”,
“Improvement Suggestion”: “Differentiate the incorrect options
more clearly and make the correct option more specific to reduce
confusion.”
}
}
}
    
```

[표 6]은 NCS 능력단위 요소만으로 생성한 초기 문항(NT0)과 프롬프트 개선을 거쳐 생성한 문항(NT1), 그리고 문항 개선 피드백을 반영하여 자동화 반복 생성한 문항(NT2~NT8)의 품질 평가 결과를 정리한 것이다. 문항 품질 평가결과는 ‘NCS 수행준거와의 연관성’, ‘문제의 난이도와 변별력’, ‘문제의 명확성과 이해도’, ‘오답 보기의 적절성’ 등 4개 척도의 합계 점수이다.

Table 6. Results of Quality Evaluation for NCS-Based Automatically Generated Assessment

Num	NT0	NT1	NT2	NT3	NT4	NT5	NT6	NT7	NT8	Avg
Q1	85	82	80	75	80	82	85	85	80	81.13
Q2	80	82	80	80	85	80	80	80	82	81.13
Q3	80	85	82	80	80	75	80	81	80	80.38
Q4	80	85	80	80	83	92	85	83	80	83.50
Q5	80	80	88	89	83	85	92	88	87	86.50
Q6	80	75	88	91	80	85	89	86	88	85.25
Q7	80	80	80	87	82	80	92	88	92	85.13
Q8	80	75	80	92	85	92	87	75	87	84.13
Q9	80	88	75	80	88	90	80	81	88	83.75
Q10	80	80	83	87	85	80	80	75	80	81.25
Q11	85	92	85	85	85	82	88	88	80	85.63
Q12	85	80	85	85	75	80	83	88	88	83.00
Q13	81	85	75	78	80	88	86	85	80	82.13
Avg	81.2	82.2	81.6	83.8	82.4	83.9	85.2	83.3	84.0	83.3

실험 결과, NCS 능력단위 요소만으로 생성한 초기 문항의 품질(NT0)은 평균 81.2점으로 비교적 양호한 수준을 보였으며, 프롬프트를 개선한(NT1)결과 82.2점으로 1점 상승한 것을 확인할 수 있다.

프롬프트 개선 및 피드백 반영을 통해 문항의 품질이 지속적으로 향상되는 경향을 보였다. 특히, ‘문항의 실무 상황 반영도’와 ‘변별력’ 측면에서 긍정적인 변화가 관찰되었다. 특히 문항의 실무 상황 반영도와 변별력 측면의 개선이 확인되었다.

총 13개 문항 중 8개 문항에서 품질 향상이 관찰되었으며, 최종 문항 품질은 평균 84.0점까지 향상되었다. ‘NCS 수행준거와의 연관성’, ‘문제의 명확성 및 이해도’ 부문을 중심으로 개선이 이루어졌다.

하지만, 품질평가 점수는 생성AI를 통해 측정된 결과치로, 실험을 반복한 결과 매번 동일한 수치가 나오지 않는 단점이 확인되었다. 따라서 본 논문에서 측정된 수치적인 품질평가 결과 수치는 상대적인 수치로만 참고용으로 활용하고 실제 본 논문의 방법론을 활용하여 생성된 문항(문제와 보기 구성) 자체에 대해서 세부적으로 살펴봐야 할 것이다. 이는 아직까지 생성AI가 가지는 한계점인 동시에, 향후 생성 AI 모델의 발전과 프롬프트 개선 및 문제 생성 및 평가를 위해 추후 연구 등으로 해결이 가능할 것으로 기대한다.

4.3 생성 문항 개선 전후 품질 평가

평가문항 품질개선을 지속하면서 생성된 문항을 분석한 결과 다음과 같은 특징을 관찰할 수 있었다.

(특징1) 품질개선이 지속될수록 대체적으로 문항의 길이가 길어지는 현상이 발생하였다. (특징2) 실무적인 측면에서의 문항 개선이 일어났다. (특징3) 품질개선을 통해 수정된 질문과 보기를 기준으로 정답선택지의 변경이 일어나는 경우가 있다.

[표 7]은 제안된 방법론을 통해 최종 생성된 ‘침해사고 분석’ 능력단위 평가 문항 중 위에서 관찰된 특징이 반영된 사례 몇 가지를 정리하였다.

사례1은 프롬프트 개선을 통해서 단순한 기술을 묻는 수준에서 실무 수행에 필요한 단계를 묻는 질문으로 개선되었으며, 사례2는 피드백을 1회 수행하여, 수행준거에 기반하여 보기가 조금 더 구체적으로 개선되었다.

사례3는 “(2.3) 침해사고 대응 후 재발 방지를 위한 사후 조치 방안을 마련하고 실행할 수 있다.” 수행준거를 위한 지식(사후 조치 방안, 침해사고 보고 절차) 및 기술(사후 조치 방안 마련 능력, 사후 조치 실행 능력)을 확인할 수 있는 예시로 구체적이고 효과적인 사후 조치방안을 생성하고(NT4, 피드백 3회 수행), 조치 방안에 포함되어야 할 구체적인 법률 조항을 명시하며(NT6, 추가 피드백 2회 수행), 재발 방지 모니터링 및 업데이트 계획(NT7, 추가 피드백 1회 수행)까지 물어보도록 내용이 구체적인 형태로 객관식 문항이 개선된 것을 확인할 수 있었다. 사례4 또한

사례3과 유사한 형태로 총 7회의 피드백을 통해 “(3.3) 침해 사고 분석 결과를 바탕으로 향후 대응 전략을 수립하고 실행할 수 있다.”라는 수행준거와 관련된 지식(향후 대응 전략 수립 방법, 침해사고 분석 결과활용 방법) 및 기술(향후 대응 전략 수립 능력, 침해사고 분석 결과활용 능력)을 확인할 수 있는 문항으로 문항이 개선됨을 확인할 수 있었다.

Table 7. Comparison of Before and After Quality Improvement for Automatically Generated Questions

Category		Content
Case1	Before (NT0-Q1)	Question 1: (1.1) What is the primary technology used to analyze the latest types of incidents, attack patterns, and vulnerabilities announced by domestic and international information security agencies? b a) Data encryption technology b) Incident analysis tool usage technology c) Network configuration technology d) Software development technology
	After (NT1-Q1)	Question 1: (1.1) What is the first step that should be taken to identify the cause and process of an incident when analyzing the latest types of incidents, attack patterns, and vulnerabilities announced by domestic and international information security agencies? d a) Establishing an incident response plan b) Analyzing system logs c) Monitoring network traffic d) Collecting the latest security notices and vulnerability information
	Details	The question was changed from asking about the technology used in vulnerability analysis to a specific step in the process, and the options were revised to include examples relevant to information security (NT0 → NT1).
Case2	Before (NT1-Q5)	Question 5: (2.2) What is the most important factor in establishing a future response strategy based on the results of incident cause analysis? c a) Analyzing log data at the time of the incident b) Monitoring for network traffic increases c) Identifying and improving weaknesses in the security policy d) Setting the frequency of user password changes
	After (NT2-Q5)	Question 5: (2.2) What is the most important factor in establishing a future response strategy based on the results of incident cause analysis? c a) Identifying the attack path and vulnerabilities through detailed analysis of log data at the time of the incident b) Continuously monitoring for abnormal network traffic to detect further intrusion attempts early c) Reflecting the results of incident analysis to improve and supplement weaknesses in the security policy d) Setting the frequency of user password changes to enhance internal account security
	Details	The options were improved to be more related to the performance criteria (2.2) by incorporating feedback (NT1 → NT2).

Category	Content	
Case3	Before (NT1-Q6)	Question 6: (2.3) What is the most appropriate follow-up measure to prevent recurrence after incident response? b a) Notifying employees via email in case of an incident. b) Conducting internal training and education on the causes of the incident. c) Replacing all network equipment. d) Deleting log files in case of an incident.
	After (NT4-Q6)	Question 6: (2.3) What is the most appropriate specific and effective follow-up measure to prevent recurrence after incident response? Consider the importance of legal requirements and the attachment of specific materials. c a) Documenting the incident response procedure, educating all employees, and continuously updating relevant materials to reflect legal requirements. b) Conducting internal training and regular drills based on the results of the incident cause analysis, and attaching materials that meet legal requirements. c) Regularly reviewing and updating the incident response process, improving security weaknesses, and meeting necessary legal requirements. d) Analyzing log files to identify security weaknesses and implementing corrective actions considering legal requirements.
	After (NT6-Q6)	Question 6: (2.3) What is the most appropriate specific and effective follow-up measure to prevent recurrence after incident response? Consider the importance of legal requirements and the attachment of specific materials. Also, specify the specific legal requirements (Article 33 and 34 of GDPR, Article 27-3 of the Domestic Information and Communications Network Act) and the types of additional materials (impact size, response process, recurrence prevention plan) that should be included in the report. b a) Documenting the incident response procedure, educating all employees, and reporting to the supervisory authority within 72 hours in case of an incident as per GDPR Article 33, while continuously updating relevant materials. The report should include the cause of the incident, the response process, and improvements, and specify compliance with related regulations. b) Conducting internal training and regular drills based on the results of the incident cause analysis, notifying the victim immediately in case of an incident as per Article 27-3 of the Information and Communications Network Act, and attaching materials that meet legal requirements. The report should include the impact size, response details, future plans, and evaluate the long-term impact on organization operations, and include a regular system check plan for recurrence prevention. c) Regularly reviewing and updating the incident response process, improving security weaknesses, and notifying data subjects as per GDPR Article 34. The report should clearly include legal compliance details, improved security policies, training content, and specify the loss of important data types

Category	Content
Case3	After (NT6-Q6) and amounts, including a continuous security monitoring plan for recurrence prevention. d) Analyzing log files to identify security weaknesses, and implementing corrective actions by supplementing necessary materials as per Article 27-3 of the Information and Communications Network Act. The incident response report should include log analysis results, corrective actions, and recurrence prevention plans, evaluating the frequency and types of disturbance activities in the network, and including a regular vulnerability check plan.
	Question 6: (2.3) What is the most appropriate specific and effective follow-up measure to prevent recurrence after incident response? Consider the importance of legal requirements and the attachment of specific materials. Also, specify the specific legal requirements (Article 33 and 34 of GDPR, Article 27-3 of the Domestic Information and Communications Network Act) and the types of additional materials (impact size, response process, recurrence prevention plan) that should be included in the report. Additionally, include continuous monitoring and update plans for recurrence prevention. a) Documenting the incident response procedure, educating all employees, and reporting to the supervisory authority within 72 hours in case of an incident as per GDPR Article 33, while continuously updating relevant materials. The report should include the cause of the incident, the response process, specific improvements, specify compliance with related regulations, and include methods for evaluating post-recovery system stability and a regular monitoring plan for recurrence prevention. b) Conducting internal training and regular drills based on the results of the incident cause analysis, notifying the victim immediately in case of an incident as per Article 27-3 of the Information and Communications Network Act, and attaching materials that meet legal requirements. The report should include the impact size, response details, future plans, and evaluate the long-term impact on organization operations, including a regular system check plan for recurrence prevention. c) Regularly reviewing and updating the incident response process, improving security weaknesses, and notifying data subjects as per GDPR Article 34. The report should clearly include legal compliance details, improved security policies, training content, and specify the loss of important data types and amounts, including a continuous security monitoring plan for recurrence prevention. d) Analyzing log files to identify security weaknesses, and implementing corrective actions by supplementing necessary materials as per Article 27-3 of the Information and Communications Network Act. The incident response report should include log analysis results, corrective actions, and recurrence prevention plans, evaluating the frequency and types of disturbance activities in the network, and including a regular vulnerability check plan.
	Details The question and options were improved to include specific and detailed content reflecting legal requirements based on feedback (NT1 → NT8, total 7 rounds of feedback).

Category	Content
Case4	After (NT7-Q10) Question 10: (3.3) What is the first step to take when establishing a future response strategy based on incident analysis results, and why? For example, if abnormal traffic is detected on a specific web server, explain the first action to take and the reason for it. a) Quickly isolating the affected system based on the initial analysis of the incident cause to prevent further damage. For example, if abnormal traffic is detected due to a vulnerability in specific software, isolate the system from the network and analyze logs to determine the extent of the intrusion. b) Recovering and restoring the compromised system to normal operation based on detailed analysis results of the incident cause. For example, if abnormal traffic is detected due to a vulnerability in specific software, fix the vulnerability and restore the system. c) Strengthening security policies and preventing similar future incidents based on detailed analysis results of the incident cause. For example, if abnormal traffic is detected due to a vulnerability in specific software, apply security patches to the software and implement additional security solutions. d) Confirming normal operation of the system based on detailed analysis results of the incident cause, and deploying additional monitoring tools. For example, if abnormal traffic is detected due to a vulnerability in specific software, fix the vulnerability and enhance the monitoring system.
	Detail The question and options were improved to include specific and detailed content considering actual incident scenarios based on feedback (NT1 → NT8, total 7 rounds of feedback).

5. 결론

5.1 연구 결과 요약

본 연구는 ChatGPT-4o를 활용하여 NCS 기반 정보보안 직무능력 평가 문항을 자동으로 생성하고 품질을 개선하는 방법론을 제안하였다.

NCS 정보보안 분야 ‘침해사고 분석’ 능력단위 요소를 새롭게 정의하고 구조화 하였으며, 프롬프트 최적화를 통해 평가 문항의 직무 적합성을 제고하는 한편, 자동 품질 검증 및 피드백 반영을 통해 자동으로 생성되는 객관식 문항 출제를 통해 NCS 능력단위 요소에서 요구하는 수행준거에 따른 지식과 기술을 평가할 수 있도록 하였다. 또한, 사례 분석을 통하여 실제 피드백이 지속될수록 실무업무에 기반한 지식 및 기술요소가 생성된 문항에 적용됨을 보였다.

즉, 제안 방법론을 통해 NCS 정보보안 분야 ‘침해사고 분석’ 능력단위 기반의 평가 문항 자동 생성이 가능하였다.

5.2 연구의 시사점 및 의의

본 연구는 대규모 언어 모델(LLM) 기술을 NCS 기반 직무능력 평가 고도화에 접목한 선도적 시도로서 의의가 있

다. 제안 방법론은 방대한 분량의 고품질 NCS 평가 문항을 신속하고 효율적으로 개발할 수 있는 가능성을 제시한다.

특히, NCS 능력단위의 수행준거와 관련된 지식 및 기술 요소에 기반하여 실무에서 활용 가능한 형태의 질문 및 보기 생성을 통하여 보다 실무 중심적인 문항생성을 자동으로 수행하는 것은 향후 NCS 기반의 직무능력 평가에 많은 도움이 될 것으로 기대한다.

한편, 정략적 측면에서의 객관적인 문항품질 측정을 위해 제안한 부분은 본 논문에서는 크게 의미를 찾기 어려웠다. 하지만, 실제 생성된 문항을 분석한 결과 사례1~4에서 확인할 수 있듯이, 문항 자동 생성 및 개선 피드백 반영이 지속될수록 문항의 내용 및 완성도가 높아지는 것을 확인할 수 있었다. 특히 법률적 조항을 문항에 자동으로 접목하는 부분이나 실제 침해사고 분석 상황을 가정한 구체적 예시를 반영한 문항 자동 생성 결과는 본 연구의 성과 중 하나이다.

향후 본 논문에서 제안한 정략적 측면의 연구와 NCS 능력단위 ‘침해사고 분석’ 이외에 다양 능력단위별 문항생성 등을 통하여 본 방법론을 개선하고 다양한 분야에 제안한 방법론이 활용될 수 있기를 기대한다.

참고문헌

- [1] KISA (Korea Internet & Security Agency). (2023). *Second Half Cyber Threat Trends Report (KISA Report)*. Korea Internet & Security Agency.
- [2] HRDK (Human Resources Development Service of Korea). (2024). *NCS (National Competency Standards)*. <https://ncs.go.kr>
- [3] KRIVET (Korea Research Institute for Vocational Education and Training). (2019). *NCS Module: Incident Analysis (2001060305_19v2)*. Korea Research Institute for Vocational Education and Training. <https://ncs.go.kr/unity/th03/ncsResultSearch.do>
- [4] OpenAI. (2023). *GPT-4 Technical Report*. OpenAI. <https://doi.org/10.48550/arXiv.2303.08774>
- [5] OpenAI. (2024). *GPT-4o*. <https://openai.com/index/hello-gpt-4o/>
- [6] Kim, J.S., Lee, K.C., & Choi, S.Y.(2022). NCS based Leveled Micro-Degree Certification Model for Training Practical Cyber Security Experts. *Journal of The Korea Society of Computer and Information*, 27(8), 123-133. <https://doi.org/10.9708/jksci.2022.27.08.123>
- [7] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All You Need. *Conference on Neural Information Processing Systems (NIPS 2017)*. <https://doi.org/10.48550/arXiv.1706.03762>
- [8] Liu, P., Yuan, F., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2021). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *arXiv preprint arXiv:2010.15980*. <https://doi.org/10.48550/arXiv.2107.13586>
- [9] Das, B., Majumder, M., Phadikar, & S. et al. (2021). Automatic question generation and answer assessment: a survey. *Research and Practice in Technology Enhanced Learning(RPTEL)* 16, 5. <https://doi.org/10.1186/s41039-021-00151-1>
- [10] Mitkov, R., & Ha, L. A. (2003). Computer-aided generation of multiple-choice tests. In *Proceedings of the HLT-NAACL 03 workshop on Building educational applications using natural language processing - Volume 2 (HLT-NAACL-EDUC '03)*. Association for Computational Linguistics, 17-22. <https://doi.org/10.3115/1118894.1118897>
- [11] Cho, Y.S. (2024). Research on Design for the Assessment System and Knowledge Tracing Methods Based on Generative AI. *The Journal of Korean Association of Computer Education*, 27(1), 143-156. <https://doi.org/10.32431/kace.2024.27.1.011>
- [12] Kiyak, Y. S., Coşkun, Ö., Budakoğlu, I. İ., & Uluoğlu, C. (2024). ChatGPT for generating multiple-choice questions: Evidence on the use of artificial intelligence in automatic item generation for a rational pharmacotherapy exam. *European Journal of Clinical Pharmacology*, 80(5), 729-735. <https://doi.org/10.1007/s00228-024-03649-x>
- [13] Sarsa, S., Denny, P., Hellas, A., & Leinonen, J. (2022). Automatic Generation of Programming Exercises and Code Explanations Using Large Language Models. *ICER '22, Vol. 1* 27-43. ACM. <https://doi.org/10.1145/3501385.3543957>
- [14] Seulki Kim. (2023). Developing Code Generation Prompts for Programming Education with Generative AI. *The Journal of Korean Association of Computer Education*, 26(5), 107-117. <https://doi.org/10.32431/KACE.2023.26.5.009>
- [15] Doughty, J., Wan, Z., Bompelli, A., Qayum, J., Wang, T., Zhang, J., Zheng, Y., Doyle, A., Sridhar, P., Agarwal, A., Bogart, C., Keylor, E., Kultur, C., Savelka, J., & Sakr, M. (2024). A Comparative Study of AI-Generated (GPT-4) and Human-crafted MCQs in Programming Education. *Proceedings of the 26th Australasian Computing Education Conference*, 114-123. <https://doi.org/10.1145/3636243.3636256>



이재식

- 2005년 가천대학교 컴퓨터공학과(공학사)
- 2007년 숭실대학교 컴퓨터공학과 (공학석사)
- 2013년 숭실대학교 컴퓨터공학과 (공학박사)
- 2012년 ~ 현재 한국인터넷진흥원 보안교육운영팀장

✚ 관심분야 : 정보보호, 정보보호 인력양성, 생성 AI, 역량평가, 침해사고 대응·분석, 인증

✉ j30231@gmail.com

부 록

〈표 1〉 NCS ‘침해사고 분석’ 능력단위 구성요소(24v3)

능력단위 요소	수행준거 (description)	지식 (knowledge)	기술 (skills)
(200106 0305_24v3.2) 침해사고 원인 분석하기	(2.1)침해사고 발생 원인에 대한 분석 절차를 이해하고, 각 절차에 대한 효과적인 방법을 적용할 수 있다.	침해사고 분석 절차, 침해사고 대응 전략	침해사고 분석 절차 적용 능력, 효과적인 대응 방법 적용 능력
	(2.2)각종 침해사고 유형별 대응 방법을 적용하고, 원인 분석결과를 바탕으로 향후 대응 전략을 수립할 수 있다.	침해사고 유형별 대응 방법, 재발 방지 대책	원인 분석 결과 활용 능력, 향후 대응 전략 수립 능력
	(2.3)침해사고 대응 후 재발 방지를 위한 사후 조치 방안을 마련하고 실행할 수 있다.	사후 조치 방안, 침해사고 보고 절차	사후 조치 방안 마련 능력, 사후 조치 실행 능력
	(2.4)국내외 정보보호 관련 법과 규정에 따라 침해사고 대응 및 원인 분석 결과를 보고할 수 있다.	국내외 정보 보호 법과 규정, 침해사고 분석 보고서 작성 방법	침해사고 보고서 작성 능력, 법과 규정 준수 능력
(평가요소)침해사고 발생 원인 분석 절차 이해 및 적용 능력, 각종 유형별 대응 방법 적용 능력, 사후 조치 방안 마련 및 실행 능력			

〈표 3〉 NCS 기반 평가 문항 생성 프롬프트

```

prompt = f"""
당신은 침해사고 분석업무 수행경험이 많은 정보보안 분야 전문가이자 문제출제 전문가입니다.
객관식 4지선다형 문제를 1개 생성에 필요한 NCS 기반 메타정보입니다.
능력단위: {ability_unit['name']} ({ability_unit['code']})
능력단위요소: {performance_criteria['name']} ({performance_criteria['code']})
수행준거: {sub_criteria_code}
설명: {description}
필요 지식: {'; '.join(knowledge)}
필요 기술: {'; '.join(skills)}

출제되는 문제는 주어진 NCS 능력단위요소의 수행준거를 바탕으로, 정보보안 전문가로서 갖추어야 할 지식과 기술을 평가할 수 있어야 합니다. 특히 수험자 간 변별력을 고려하여 문항을 구성하되, 문제의 질문과 보기는 명확하고 이해하기 쉽게 작성되어야 합니다.
총 4개의 보기를 제시하고 그 중 정답은 1개여야 하며, 답안을 눈치챌 수 있는 단서를 포함하지 않도록 주의합니다. 오답 보기의 경우, 그 자체만으로는 정답을 유추하기 어려워야 하고, 지나치게 명백한 오답은 피해야 합니다.

생성된 문제는 다음의 4가지 평가 기준에서 각각 25점씩, 총 100점을 획득할 수 있어야 합니다:
- NCS 수행준거와의 연관성 (25점)
- 문제의 난이도와 변별력 (25점)
- 문제의 명확성과 이해도 (25점)
- 오답 보기의 적절성 (25점)

문제 생성 결과는 아래 JSON 형식에 맞추어 출력해 주세요:
{
  "question": "질문 {question_number}: ({sub_criteria_code}) 문제
    
```

```

내용",
"choices": [
  "a) 보기1",
  "b) 보기2",
  "c) 보기3",
  "d) 보기4"
],
"answer": "정답"
}
"""
    
```

〈표 4〉 문항 품질 검증 평가 및 개선 프롬프트

```

prompt = f"""
다음은 출제된 문제입니다:
{question}
다음은 보기입니다:
{chr(10).join(choices)}
이 문제는 다음의 NCS 능력단위요소와 관련이 있습니다:
능력단위요소: {performance_criteria['name']} ({performance_criteria_code})
수행준거: {sub_criteria_code}
설명: {description}
필요한 지식: {'; '.join(knowledge)}
필요한 스킬: {'; '.join(skills)}
당신은 한국 NCS 정보보호 분야 전문가이자 문제출제 전문가입니다.
다음의 4가지 평가 기준에 따라 25점 척도로 문제를 평가하고, 각 목표별 구체적인 감점 사유와 개선 방안을 피드백으로 제시해 주세요.

NCS 수행준거와의 연관성 (25점):
문제가 해당 수행준거, 지식, 기술 요소를 충실히 반영하고 있는가? (15점)
문제가 수행준거 측정에 적합한 방식으로 구성되었는가? (10점)
감점 시 구체적 사유:

문제의 난이도와 변별력 (25점):
정보보안 전문가 수준에서의 문제 난이도는 적절한가? (10점)
문제 풀이 결과에 따른 수험자 간 변별력이 충분한가? (15점)
감점 시 구체적 사유:

문제의 명확성과 이해도 (25점):
문제의 질문과 보기가 명확하고 이해하기 쉽게 기술되었는가? (15점)
전문가의 관점에 따라 문제 정답이나 의도를 다르게 해석할 여지는 없는가? (10점)
감점 시 구체적 사유:

오답 보기의 적절성 (25점):
정답 보기가 오답 보기와 명확히 구분되는가? (10점)
오답 보기 자체만으로는 정답을 유추하기 어려우나? (15점)
감점 시 구체적 사유:

종합 평가 결과와 개선을 위한 피드백을 다음 JSON 형식으로 제시해 주세요:
{
    
```

```

“NCS 수행준거와의 연관성”: 25,
“문제의 난이도와 변별력”: 25,
“문제의 명확성과 이해도”: 25,
“오답 보기의 적절성”: 25,
“종합평가결과(합계)”: 100,
“피드백”: “각 평가 항목별 감점 사유와 구체적인 개선 방안을 포함하여 작성.”
}
}
“”
    
```

〈표 5〉 자동 생성 문항 품질 개선 프로세스 결과값 JSON 구조화 예시 (NT5-7번 문항)

```

{
  “question”: “질문7: (2.2.4) 국내의 정보보호 관련 법과 규정에 따라 침해사고 대응 및 원인 분석 결과를 보고할 때, 다음 중 가장 적절한 방법은 무엇입니까? 또한, 시스템 및 네트워크 성능 저하와 데이터 손실 같은 다양한 영향을 평가하여 작성해야 하는 항목은 무엇입니까?”,
  “choices”: [
    “a) 보고서는 법적 요구사항을 충족하고, 분석 결과와 대응 절차를 포함하여 상세히 작성해야 하며, 시스템 및 네트워크 성능 저하, 데이터 손실 등의 영향을 평가한 추가 자료(예: 로그 파일, 기술 보고서)를 첨부한다.”,
    “b) 보고서는 법적 요구사항에 따라 대응 절차만을 기록하고, 분석 결과와 시스템 및 네트워크 성능 저하, 데이터 손실 등의 영향 평가는 별도로 구두로 보고한다.”,
    “c) 보고서는 국내의 정보보호 법과 규정에 따라 대응 절차와 분석 결과, 시스템 및 네트워크 성능 저하, 데이터 손실 등의 영향을 평가한 내용을 상세히 작성하고, 필요시 추가 자료를 첨부한다.”,
    “d) 보고서는 피해자의 요청에 따라 간략히 작성하고, 법적 요구사항을 최소화하여 보고하며, 시스템 및 네트워크 성능 저하, 데이터 손실 등의 영향 평가는 생략한다.”
  ],
  “answer”: “c) 보고서는 국내의 정보보호 법과 규정에 따라 대응 절차와 분석 결과, 시스템 및 네트워크 성능 저하, 데이터 손실 등의 영향을 평가한 내용을 상세히 작성하고, 필요시 추가 자료를 첨부한다.”,
  “NCS 수행준거와의 연관성”: 20,
  “문제의 난이도와 변별력”: 20,
  “문제의 명확성과 이해도”: 20,
  “오답 보기의 적절성”: 20,
  “종합평가결과(합계)”: 80,
  “피드백”: {
    “NCS 수행준거와의 연관성”: {
      “점수”: 20,
      “감점 사유”: “문제는 전반적으로 수행준거와 관련이 있지만, 각 보기의 구체적인 설명이 부족하여 수행준거와의 직접적인 연관성을 명확히 파악하기 어렵습니다.”,
      “개선 방안”: “각 보기를 더욱 구체적으로 설명하여 수행준거와의 연관성을 명확히 할 필요가 있습니다. 예를 들어, ‘데이터 손실’이 조직에 미치는 구체적인 영향을 명시하는 것이 좋습니다.”
    },
    “문제의 난이도와 변별력”: {
      “점수”: 20,
      “감점 사유”: “문제의 난이도가 다소 낮아 전문가 수준에서의 변별
    
```

```

력이 충분하지 않을 수 있습니다. 모든 보기가 다소 일반적이어서 특정 답을 고르기 어렵습니다.”,
  “개선 방안”: “문제의 난이도를 높이기 위해 구체적이고 전문적인 용어를 사용하고, 현실적인 시나리오를 추가하여 전문가 수준의 변별력을 높이는 것이 좋습니다.”
},
“문제의 명확성과 이해도”: {
  “점수”: 20,
  “감점 사유”: “문제의 질문과 보기가 명확하지만, 일부 보기가 모호하여 해석의 여지가 있습니다. 예를 들어, ‘네트워크에서의 불법 접근 시도’와 ‘악성 코드 탐지 여부’가 구체적으로 어떻게 영향을 미치는지 불명확합니다.”,
  “개선 방안”: “보기를 보다 구체적으로 명시하여 해석의 여지를 줄이고, 문제의 명확성을 높이는 것이 필요합니다.”
},
“오답 보기의 적절성”: {
  “점수”: 20,
  “감점 사유”: “오답 보기가 일부 정답 보기와 혼동될 수 있습니다. 예를 들어, ‘시스템 복구 시간 및 복구 후 안정화 시간’과 ‘시스템 및 네트워크 성능 저하 정도’가 모두 중요한 요소로 인식될 수 있습니다.”,
  “개선 방안”: “오답 보기를 더 명확히 구분하고, 정답 보기를 더욱 구체적으로 작성하여 혼동을 줄이는 것이 좋습니다.”
}
}
}
    
```

〈표 7〉 자동 생성 문항 품질 개선 사례 전/후 비교

구분	내용
사례1	개선 전 (NT0-Q1) 질문1: (1.1) 국내의 정보보호 유관기관들이 발표한 최신 침해사고 유형과 공격 패턴, 취약점들을 분석할 때 주로 사용하는 기술은 무엇인가요? b a) 데이터 암호화 기술 b) 침해사고 분석 도구 사용 기술 c) 네트워크 설정 기술 d) 소프트웨어 개발 기술
	개선 후 (NT1-Q1) 질문1: (1.1) 국내의 정보보호 유관기관들이 발표한 최신 침해사고 유형과 공격 패턴, 취약점들을 분석할 때, 침해사고의 원인과 침해 과정을 파악하기 위해 가장 먼저 수행해야 할 단계는 무엇인가? d a) 침해사고 대응 계획 수립 b) 시스템 로그 분석 c) 네트워크 트래픽 모니터링 d) 최신 보안 공지 및 취약점 정보 수집
	개선사항 프롬프트 개선을 통해 취약점 분석 시 사용하는 기술에서 구체적인 수행 내용으로 질문이 변경되었으며, 그에 따라 보기의 구성도 정보보호 관점의 예시들로 모두 변경되었음(NT0 -> NT1)

구분	내용
사례2	<p>개선 전 (NT1-Q5)</p> <p>질문5: (2.2) 다음 중 침해사고 원인 분석 결과를 바탕으로 향후 대응 전략을 수립하는 데 가장 중요한 요소는 무엇입니까? c</p> <p>a) 사고 발생 시의 로그 데이터 분석 b) 네트워크 트래픽 증가 여부 모니터링 c) 보안 정책의 취약점 파악 및 개선 d) 사용자 비밀번호 변경 주기 설정</p>
	<p>개선 후 (NT2-Q5)</p> <p>질문5: (2.2) 다음 중 침해사고 원인 분석 결과를 바탕으로 한 향후 대응 전략 수립에 가장 중요한 요소는 무엇입니까? c</p> <p>a) 사고 발생 시의 로그 데이터 상세 분석을 통해 침해 경로와 취약점을 파악하는 것 b) 네트워크 트래픽의 이상 징후를 지속적으로 모니터링하여 추가적인 침해 시도를 조기에 발견하는 것 c) 침해사고 분석 결과를 반영하여 보안 정책의 취약점을 보완하고 개선하는 것 d) 사용자 비밀번호 변경 주기를 설정하여 내부 계정 보안을 강화하는 것</p>
	<p>개선사항</p> <p>문화 개선 피드백 반응을 통해 보기 항목이 수행준거(2.2)와 관련이 있는 항목으로 개선됨(NT1 -> NT2)</p>
사례3	<p>개선 전 (NT1-Q6)</p> <p>질문6: (2.3) 침해사고 대응 후 재발 방지를 위한 사후 조치 방안으로 가장 적절한 것은 무엇입니까? b</p> <p>a) 침해사고 발생 시 이메일로 직원들에게 알람을 전달한다. b) 침해사고 원인에 대한 내부 교육과 훈련을 실시한다. c) 모든 네트워크 장비를 교체한다. d) 침해사고 발생 시 로그 파일을 삭제한다.</p>
	<p>개선 중 (NT4-Q6)</p> <p>질문6: (2.3) 침해사고 대응 후 재발 방지를 위한 구체적이고 효과적인 사후 조치 방안으로 가장 적절한 것은 무엇입니까? 이때, 법적 요구사항과 구체적인 자료 첩부의 중요성을 고려하십시오. c</p> <p>a) 침해사고 대응 절차를 문서화하여 전 직원에게 교육하고, 법적 요구사항을 반영하여 관련 자료를 지속적으로 업데이트한다. b) 침해사고 원인 분석 결과를 기반으로 내부 교육과 정기적인 훈련을 실시하며, 법적 요구사항에 맞춘 자료를 첨부한다. c) 침해사고 대응 프로세스를 정기적으로 검토하고 업데이트하며, 보안 취약점을 개선하고 필요한 법적 요구사항을 충족시킨다. d) 침해사고 로그를 분석하여 보안 취약점을 파악하고, 법적 요구사항을 고려하여 개선 조치를 실행한다."</p>
사례3	<p>개선 중 (NT6-Q6)</p> <p>질문6: (2.3) 침해사고 대응 후 재발 방지를 위한 구체적이고 효과적인 사후 조치 방안으로 가장 적절한 것은 무엇입니까? 이때, 법적 요구사항과 구체적인 자료 첩부의 중요성을 고려하십시오. 또한, 조치 방안에 포함되어야 할 구체적인 법적 요구사항(GDPR의 제33조 및 제34조, 국내 정보통신망법 제27조의3)과 보고서 작성 시 포함해야 할 추가 자료(피해 규모, 대응 과정, 재발 방지 계획)의 종류도 명시하십시오. b</p> <p>a) 침해사고 대응 절차를 문서화하여 전 직원에게 교육하고, GDPR 제33조에 따라 침해사고 발생 시 72시간 이내에 감독 당국에 신고하며, 관련 자료를 지속적으로 업데이트한다. 보고서에는 사고 발생 원인, 대응 과정, 개선 사항 등을 포함하고 관련 법규 준수 여부를 명시한다. b) 침해사고 원인 분석 결과를 기반으로 내부 교육과 정기적인 훈련을 실시하며, 정보통신망법 제27조의3에 따라 침해사고 발생 시 피해자에게 즉시 통지하고, 법적 요구사항에 맞춘 자료를 첨부한다. 또한, 사고 보고서에 피해 규모, 조치 내역, 향후 계획 등을 포함시킨다.</p>

구분	내용
사례3	<p>개선 중 (NT6-Q6)</p> <p>c) 침해사고 대응 프로세스를 정기적으로 검토하고 업데이트하며, 보안 취약점을 개선하고 GDPR 제34조에 따라 데이터 주체에게 통지한다. 보고서에는 법적 준수 사항, 개선된 보안 정책, 교육 내용 등이 포함되어야 한다. d) 침해사고 로그를 분석하여 보안 취약점을 파악하고, 정보통신망법 제27조의3에 따라 필요한 경우 추가 자료를 구체적으로 보완하여 개선 조치를 실행한다. 또한, 사고 대응 보고서에 로그 분석 결과, 조치 내역, 재발 방지 계획 등을 포함한다.</p>
	<p>개선사항</p> <p>문화 개선 피드백 반응을 통해 질문과 보기항목이 법률적 요구사항을 포함하는 구체적이고 상세한 내용으로 개선됨(NT1 -> NT8, 총 7회 피드백 반영)</p>
사례4	<p>개선 전 (NT1-Q10)</p> <p>질문10: (3.3) 침해사고 분석 결과를 바탕으로 향후 대응 전략을 수립하는 과정에서 가장 먼저 고려해야 할 요소는 무엇입니까? b</p> <p>a) 조직의 예산과 자원 할당 가능성 b) 침해사고 발생 원인의 분석 결과 c) 최신 보안 기술의 도입 여부 d) 직원들의 보안 인식 수준</p>
	<p>개선 후 (NT7-Q10)</p> <p>질문10: (3.3) 침해사고 분석 결과를 바탕으로 향후 대응 전략을 수립할 때 첫 번째로 취해야 할 조치는 무엇이며, 그 이유는 무엇인지 설명하세요. 예를 들어, 특정 웹 서버에서 비정상적인 트래픽이 발생한 경우, 이를 분석한 후 첫 번째로 해야 할 작업과 그 이유를 구체적으로 설명하십시오. a</p> <p>a) 침해사고 발생 원인의 초기 분석을 바탕으로 해당 시스템을 신속하게 격리하고 추가적인 손상을 방지합니다. 예를 들어, 비정상적인 트래픽이 특정 소프트웨어의 취약점을 통해 발생했을 경우, 해당 시스템을 네트워크에서 격리하고 로그 분석을 통해 침해 범위를 파악합니다. b) 침해사고 발생 원인의 세부 분석 결과를 바탕으로 손상된 시스템을 복구하고 정상 운영 상태로 되돌립니다. 예를 들어, 비정상적인 트래픽이 특정 소프트웨어의 취약점을 통해 발생했을 경우, 해당 취약점을 수정하고 시스템을 복구합니다. c) 침해사고 발생 원인의 세부 분석 결과를 바탕으로 보안 정책을 강화하고 향후 유사한 사고를 예방합니다. 예를 들어, 비정상적인 트래픽이 특정 소프트웨어의 취약점을 통해 발생했을 경우, 해당 소프트웨어에 대한 보안 패치를 적용하고 추가적인 보안 솔루션을 도입합니다. d) 침해사고 발생 원인의 세부 분석 결과를 바탕으로 시스템의 정상 작동을 확인한 후 추가적인 모니터링 도구를 배치합니다. 예를 들어, 비정상적인 트래픽이 특정 소프트웨어의 취약점을 통해 발생했을 경우, 해당 취약점을 수정하고 모니터링 시스템을 강화합니다.</p>
개선사항	<p>문화 개선 피드백 반응을 통해 질문과 보기항목이 실제 침해사고 상황을 고려하여 구체적이고 상세한 내용으로 개선됨(NT1 -> NT8, 총 7회 피드백 반영)</p>