

Prompt engineering to improve the performance of teaching and learning materials Recommendation of Generative Artificial Intelligence

Soo-Hwan Lee*, Ki-Sang Song**

*Teacher, Oh-hyun Elementary School, Suwon, Korea

**Professor, Dept. of Computer Education, Korea National University of Education, Chungju, Korea

[Abstract]

In this study, prompt engineering that improves prompts was explored to improve the performance of teaching and learning materials recommendations using generative artificial intelligence such as GPT and Stable Diffusion. Picture materials were used as the types of teaching and learning materials. To explore the impact of the prompt composition, a Zero-Shot prompt, a prompt containing learning target grade information, a prompt containing learning goals, and a prompt containing both learning target grades and learning goals were designed to collect responses.

The collected responses were embedded using Sentence Transformers, dimensionalized to t-SNE, and visualized, and then the relationship between prompts and responses was explored. In addition, each response was clustered using the k-means clustering algorithm, then the adjacent value of the widest cluster was selected as a representative value, imaged using Stable Diffusion, and evaluated by 30 elementary school teachers according to the criteria for evaluating teaching and learning materials.

Thirty teachers judged that three of the four picture materials recommended were of educational value, and two of them could be used for actual classes. The prompt that recommended the most valuable picture material appeared as a prompt containing both the target grade and the learning goal.

▶ **Key words:** Prompt Engineering, Generative Artificial Intelligence, Personalized Education, Sentence Embedding, t-SNE, k-mean clustering algorithm

-
- First Author: Soo-Hwan Lee, Corresponding Author: Ki-Sang Song
 - *Soo-Hwan Lee (leesoohwan@outlook.kr), Oh-hyun Elementary School
 - **Ki-Sang Song (kssong1k4@gmail.com), Dept. of Computer Education, Korea National University of Education
 - Received: 2023. 07. 21, Revised: 2023. 08. 16, Accepted: 2023. 08. 21.

[요 약]

본 연구에서는 GPT, Stable Diffusion과 같은 생성형 인공지능을 이용한 교수·학습 자료 추천 성능 향상을 위해 프롬프트를 개선하는 프롬프트 엔지니어링에 대해 탐색하였다. 분석할 교수·학습 자료의 종류는 그림 자료이다. 프롬프트 구성에 따른 영향을 탐색하기 위해 명령만 담긴 Zero-Shot 프롬프트, 학습 대상 학년 정보가 담긴 프롬프트, 학습 목표가 담긴 프롬프트, 학습 대상 학년과 학습 목표가 모두 담긴 프롬프트를 설계하여 각각을 GPT-3.5모델에 입력하고 응답을 수집하였다.

수집한 응답을 Sentence Transformers로 임베딩 하고 t-SNE를 활용하여 차원 축소하여 시각화 한 다음 프롬프트와 응답 간의 관계를 탐색하였다. 그리고 각 응답을 k-means clustering algorithm을 활용하여 군집화 한 다음 가장 넓은 클러스터의 첫 번째 값을 대표로 선택하여 Stable Diffusion을 이용하여 이미지화 한 다음 교수·학습자료 평가 기준에 따라 초등학교 교사 30명에게 평가 받았다.

초등학교 교사 30인은 추천한 4종의 그림 자료 중 3종은 교육적 가치가 있다고 판단하였으며, 그 중 2종은 실제 수업에 사용할 수 있다고 하였다. 가장 가치 있는 그림 자료를 추천한 프롬프트는 대상 학년과 학습 목표가 모두 담긴 프롬프트로 나타났다.

▶ **주제어:** 프롬프트 엔지니어링, 생성형 인공지능, 맞춤형 교육, 문장 임베딩, t-SNE, k-mean 클러스터링 알고리즘

I. Introduction

인공지능 기술이 발전함에 따라 교육 분야에도 인공지능 기술을 응용하려는 다양한 시도가 일어나고 있다. 최근 혁신적인 도구 중 하나는 생성형 인공지능(Generative AI)이다. 생성형 인공지능은 대규모 언어 데이터 셋에서 얻은 지식을 기반으로 자연어 또는 그림 등 생성하거나 처리할 수 있도록 개발된 인공지능 모델이다. 생성형 인공지능에는 GPT, BARD, BERT, Stable Diffusion 등 다양한 모델이 있으며, 이와 같은 생성형 인공지능의 강점을 이용하여 다양한 분야에 응용하려는 시도가 활발하게 이뤄지고 있다.

교육 분야에서도 인공지능 모델을 그대로 교수·학습 과정에 이용하거나 인공지능 모델을 새로운 서비스에 접목하는 등 다양한 방식의 연구가 진행되고 있다. 그러나 이러한 생성형 인공지능은 기반 모델(Foundation Model)의 성격을 가지고 있어 특정한 목적에 맞게 활용하고자 할 때 적절하게 가공해야 한다.

그 과정에서 주로 미세 조정(Fine-Tuning)이 활용된다. 이라고 한다. 미세 조정은 사전 학습(Pre-trained)되어 있는 인공지능 모델을 약간의 추가 학습 데이터를 이용하여 매개 변수를 업데이트하여 생성형 인공지능을 의도한 목적에서 더 잘 동작하도록 적용하는 것을 의미한다. 하지만 미세 조정은 다양한 기술이 필요하고 비용이 발생하는 등의 어려움이 있다.

대안으로 프롬프트 엔지니어링(Prompt Engineering)이 있다. 프롬프트 엔지니어링은 생성형 인공지능에 입력하는 값인 프롬프트(Prompt)를 생성형 인공지능을 통해 원하는 결과를 얻도록 설계하여 활용하는 것으로, 이와 관련한 다양한 기법이 제시되고 있다.

따라서 본 연구에서는 프롬프트 엔지니어링을 위해 Zero-Shot 프롬프트와 설계한 Few-Shots 프롬프트 모두를 GPT에 입력하고 프롬프트 별 100개의 응답을 얻어 임베딩과 시각화를 통해 관계를 탐색하였다. 그리고 프롬프트 별 100개의 응답을 클러스터링 하여 가장 넓은 클러스터의 첫 번째 값을 선정하여 Stable Diffusion으로 시각화 한 다음 인간 교사에게 기준에 따라 교육적 가치를 평가받았다.

이를 통해 학생들의 다양한 수준 및 학습 스타일에 능동적으로 대응할 수 있는 교수·학습 자료 추천 시스템에 대한 가능성을 탐색하고자 한다.

II. Preliminaries

1. Related works

왕감경(2023)은 AI 기반 대화형 챗봇 활용한 한국어교육에의 활용 방안을 모색하였으며, 여전히 일부 한국어

질문에 대한 답변에서 어휘 오류와 오답이 나타나는 점을 한계점이라고 하여 한국어 모델 개선의 필요성을 제안하였다[1]. 이수환, 송기상(2023)은 GPT모델과 Stable Diffusion을 활용한 교수·학습 그림 자료 추천의 가능성을 탐색하였으며, 추천된 그림 자료의 교육적 가치에 대해서는 인간 교사의 판단이 필요하다고 하였다[2]. 유인근, 박형용(2023)은 초등 국어과 글쓰기 교육을 위한 AI 문장 생성 웹 서비스를 KoGPT 모델을 활용하여 개발하고 그 효과를 분석하였다[3]. 이효섭, 심호석(2023)은 ChatGPT기반 메타버스 플랫폼 모델 설계에서 생성형 인공지능을 미세 조정하여 상호작용을 증강시킬 수 있는 방안을 제시하였다[4]. 용건우(2022)는 프롬프트 엔지니어링이 건축 분야의 하자 검출 인공지능 모델의 성능 향상에 효과적으로 기여할 수 있다고 하였다[5]. 전인성, 송기상(2022)은 GPT-2 모델을 미세 조정하여 엔트리 블록 코드 생성 및 추천 모델을 개발하였다[6].

이처럼 생성형 인공지능은 ChatGPT 서비스 그 자체로 활용할 뿐 아니라 목적에 따라 API를 활용하거나 모델을 변형하여 다양한 서비스에 개발 적용되며 모델을 변형하는 방식으로 미세 조정을 이용하는 것으로 나타났다.

2. Prompt Engineering

프롬프트 엔지니어링은 생성형 인공지능 모델에 사용자의 의도를 효과적으로 전달하여 원하는 결과물을 얻기 위해 입력하는 프롬프트를 설계, 최적화 하는 과정을 의미한다[7]. Wei et al.(2021)은 Finetuned Language Models Are Zero-Shot Learners에서 instruction-tuning을 통해 만든 FLAN 모델을 만들었으며 FLAN 기존 버전 대비 성능이 향상되었음을 증명하여 프롬프트 엔지니어링의 가능성을 보여주었다[8]. 프롬프트를 구성하는 방법에 대한 연구도 이뤄졌는데, Wei et al.(2022)는 Chain-of-Thought Prompting Elicits Reasoning in Large Language Models에서 생각의 사슬(Chain-of-Thought)라고 부르는 CoT 형태의 프롬프트 구성을 제안하였으며, 프롬프트를 사고의 과정을 따라 연쇄적으로 구성하여 통해 산술, 상식 및 상징적 추론 작업의 범위에서 성능을 향상시킬 수 있다고 하였다[9].

Yao et al.(2023)은 프롬프트 구조로 생각의 나무(Tree of Thoughts, ToT)를 제안하였으며 메커니즘은 Fig. 1과 같다.

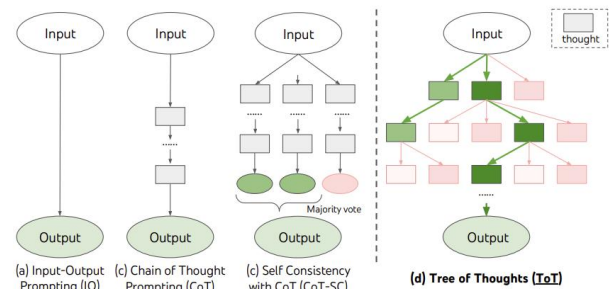


Fig. 1. Tree of Thoughts Prompt[9]

ToT는 탐색이 필요한 작업이나 전략적으로 초기 결정이 중요한 역할을 하는 작업에서의 성능을 향상시킬 수 있으며, 특히 게임 오브 24, 미니 크로스 워드 등의 게임에서 문제해결력을 향상시키는 것으로 나타났다[10].

프롬프트는 생성형 인공지능 응답의 품질에 직접적인 영향을 미치기 때문에 프롬프트를 개선하기 위한 엔지니어링의 기법에 대해 다양한 연구가 이루어지고 있다.

3. NLP(Natural Language Processing)

자연어 처리(NLP, Natural Language Processing)는 컴퓨터가 인간이 사용하는 자연어를 이해하거나 생성할 수 있도록 하는 학문분야로서, 인간의 지각 능력, 탐색 능력, 추론능력 등을 소프트웨어로 구현하는 인공지능의 한 분야이며, 자연어 처리 기술은 정보 추출, 텍스트 분류, 자연어 생성, 기계 번역, 문서 요약, 질의 응답, 대화 시스템 등에 활용된다[11].

임베딩(Embedding)은 단어, 문장 등의 텍스트를 벡터로 나타내는 것이다. 본 연구에서는 문장을 임베딩 하기 위해 Sentence Transformers를 활용하였다. Sentence Transformers는 Sentence-BERT를 이용한 문장 임베딩 모델이다. Sentence-BERT는 사전 훈련된 BERT 네트워크를 수정한 것으로 BERT / RoBERTa 보다 향상된 성능으로 문장 임베딩을 할 수 있다[12]. 임베딩을 위해 사전 훈련된 모델을 제공하며 지원하는 모델은 'All-mpnet-base-v2', 'multi-qa-mpnet-base-dot-v1' 등 30여 가지가 있으며, 본 연구에서는 품질이 가장 좋은 'All-mpnet-base-v2'와 유사한 품질을 제공하면서도 속도가 5배 정도 빠른 'All-MiniLM-L6-v2'를 사용하였다[13].

4. t-SNE

t-SNE는 고차원의 데이터를 차원 축소하여 2차원 또는 3차원 맵에 표현하는 알고리즘이다. 차원을 축소하면서도 고차원에서의 유사도를 보존하는 것이 핵심이다.

t-SNE는 Stochastic Neighbor Embedding의 변형으로 최적화 하가 훨씬 쉽고 맵의 중심에 점이 집중되는 기존의 SNE의 경향성을 줄임으로써 시각화 성능을 개선한 것이 특징이다[14].

그러나 t-SNE 알고리즘이 다양한 자료에 활용되면서 단점 및 문제점이 나타났는데 주로 모호성의 선택 문제, 전체적 구조 파악이 어려운 점이 지적된다[15].

5. k-means clustering algorithm

k-means clustering algorithm은 MacQueen이 제안한 데이터 군집 분석 알고리즘이다. 기본적인 k-means clustering은 주어진 데이터를 n보다 작은 k의 클러스터 S로 묶는 알고리즘으로 각 클러스터와 데이터와의 거리 차이 평균의 제곱합을 최소화 하는 것이다. 목적함수란 알고리즘을 최적화 하는 함수로 k-means clustering 알고리즘의 목적 함수는 $\min S = \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2$ 이고, μ_i 는 클러스터 S_i 의 평균이다[16].

본 연구에서는 k-means clustering algorithm을 활용하여 프롬프트로 생성한 응답을 군집하고 대푯값을 찾아 프롬프트를 선정하는 데 활용하였다.

III. Research Methods

1. Research Design

본 연구는 생성형 인공지능의 교수·학습 자료 추천 성능 개선을 위한 프롬프트 엔지니어링에 대해 탐색하였다.

생성형 인공지능으로 GPT-3.5와 Stable Diffusion을 선정하였다. 두 인공지능 모델은 API를 제공하고 성능이 우수하다. 또 Stable Diffusion은 오픈 소스를 이용하여 무료로 사용할 수 있으며, GPT-3.5 API는 토큰 단위로 사용량에 대한 비용을 지불하여 사용하거나 ChatGPT에서 무료로 활용할 수 있다. 생성형 모델에서 나타나는 환각(hallucination)을 줄이기 위해 프롬프트는 영어로 작성하였다.

1.1. Exploring Prompt Variables

프롬프트는 단일 명령어만으로 구성된 경우 Zero-Shot Prompt, 여러 명령어로 구성된 프롬프트를 Few-Shots Prompt 라고 한다. 이처럼 프롬프트는 사용자의 의도에 따라 다양한 데이터나 변인들을 포함한다.

프롬프트 엔지니어링의 효과를 분석하기 위해 비교군 프롬프트는 Zero-Shot Prompt로 실험군은 변인을 추가

한 Few-Shots Prompt로 설계하고 프롬프트를 GPT에 입력하여 응답을 수집한다. 수집한 응답은 임베딩하여 벡터화 하고 시각화 하여 응답에 영향을 주는 프롬프트 변인을 탐색한다. 세부 연구 절차는 Fig. 2와 같다.

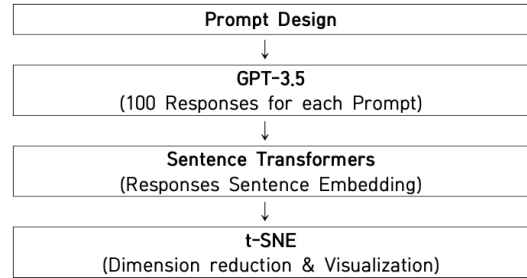


Fig. 2. Exploring Prompt Variables procedure

문장 임베딩에는 Sentence Transformers를 활용하였으며 모델은 ‘all-MiniLM-L6-v2’를 사용하였다. ‘all-MiniLM-L6-v2’ 모델은 문장을 384차원 벡터로 임베딩한다.

임베딩한 고차원 벡터를 t-SNE를 이용하여 2차원으로 축소하여 시각화 하였다. 그리고 시각화 한 자료와 2차원 벡터를 바탕으로 프롬프트에 영향을 미치는 변인에 대해 분석한다.

1.2. Prompt Selection and Evaluation

프롬프트 엔지니어링의 효과성을 분석하기 위해 프롬프트 별 GPT응답을 문장 임베딩 한 다음 군집한다. 군집은 벡터 값을 기준으로 하였으므로 의미가 유사한 응답의 집합이라고 할 수 있다. 가장 큰 군집이 프롬프트에 대한 응답의 의미 대표라 가정하고 첫 번째 인덱스를 프롬프트의 대표 응답으로 선택하고 선정한 응답을 Stable Diffusion을 이용하여 이미지 생성한다. 생성한 이미지를 초등학교 교사로 구성된 평가집단이 평가 기준에 따라 평가한다. 세부 연구 과정은 Fig. 3과 같다.

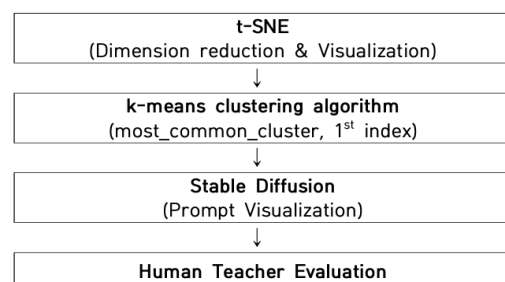


Fig. 3. Prompt Selection and Evaluation procedure

선택한 프롬프트 대표 응답별로 이미지는 4개를 생성하였다. Stable Diffusion 또한 생성형 인공지능 모델이므로 극값을 희석하기 위한 의도이다.

교수·학습 그림 자료의 평가는 교사 30인이 한국교육과정평가원(2002)에서 개발한 교수·학습 자료 평가 기준 [17]을 연구에 맞게 일부 발췌하여 사용하였다.

IV. Results

연구를 위한 분석도구는 Google Colab에서 Python을 활용하였다. 생성형 인공지능은 웹 브라우저로 각 서비스에 접속하여 ChatGPT-3.5, Stable Diffusion Playground를 활용하였다. ChatGPT는 프롬프트 간 간섭을 막기 위해 프롬프트를 변경할 때마다 새 Chat을 열어 입력하였다.

1. Exploring Prompt Variables

연구에 적용할 학습 주제는 과학과목에서 선택하였다. 과학과는 학습 내용의 특성상 학습 방법으로 실험을 활용하는데 이러한 과정에서 교실에서 실제로 보여주기 어려운 개념은 모형이나 그림자료 등의 다양한 학습 자료를 활용한다. 또 교과와 학습 요소 간 관계를 가지고 나선형 구조로 교육과정이 구성되어 있어 학습 주제 선정에 대한 변인을 고려하기 용이하다. 이를 통해 선정한 선택한 학습 주제는 Table 1과 같다.

Table 1. Select a Learning Topic

Subject	Science
Grade	3
Achievement criteria	It is possible to explain the shape and surface appearance by examining data related to the Earth.
Learning objectives	Knowing the shape of the Earth.
Keyword	Earth

Table 2. Prompt Design

Name	Type	Prompt
Prompt1	Zero-Shot Prompt	Recommend images that I need when I teach about the Earth.
Prompt2	Prompt1 + Grade Information	Recommend images that I need when I teach about the Earth. The learner is in the 3rd grade of elementary school.
Prompt3	Prompt1 + Learning Objective	Recommend images that I need when I teach about the Earth. The learning goal is to learn the shape of the Earth.
Prompt4	Prompt1 + Grade Information + Learning Objective	Recommend images that I need when I teach about the Earth. The learner is in the 3rd grade of elementary school. The learning goal is to learn the shape of the Earth.

학습 주제에 대한 키워드로는 지구를 선정하였다. 지구는 크기가 거대한 주제 특성상 반드시 모형이나 그림 자료 등 학습 자료를 활용하기 때문에 생성한 학습 자료에 대한 가치 평가가 수월할 것으로 판단하였다.

1.1. Prompt Design

교수·학습 자료 생성을 위한 프롬프트가 가지는 변인으로 학년 정보와 학습 목표를 선정하였다. 학년 정보는 해당 교수·학습 자료에 반영되어야 할 학습자의 수준을 나타내는 것으로 내용 구성의 위계에 따라 수직적인 특성을 갖는다. 학습 목표는 교수·학습 자료가 사용되어지는 목적으로 교사의 교수·학습 자료 활용의 의도를 반영하는 것으로 방향성을 포함하는 수평적인 특성을 갖는다고 볼 수 있다.

선정한 두 가지 변인을 바탕으로 4종의 프롬프트를 설계하였고, 세부 내용은 Table 2와 같다.

Prompt 1은 교수·학습용 그림 자료 추천을 요청하는 명령만 담은 Zero-Shot 프롬프트이다.

Prompt 2는 Prompt 1을 기반으로 학년 정보를 의미하는 문장을 더한 것으로 자료를 적용할 학습자의 수준을 고려할 수 있는 Few-Shots 프롬프트이다.

Prompt 3은 Prompt 1을 기반으로 학습 목표를 진술한 문장을 더한 것으로 요청하는 교수·학습용 그림 자료에 대한 활용 방향이나 내용을 구체화 하는 Few-Shots Prompt이다.

Prompt 4는 Prompt 1을 기반으로 학년 정보를 나타내는 문장과 학습 목표를 진술한 문장을 모두 더하여 설계한 프롬프트로 Few-Shots Prompt이다.

1.2. Visualization of responses

설계한 프롬프트와 응답 간의 관계를 분석하기 위해 프롬프트 별 100개의 응답을 생성하여 수집하였다.

수집한 응답을 Sentence Transformers로 384차원으로 임베딩하고, t-SNE을 이용하여 2차원으로 축소한 다음 시각화 하였다. 시각화 한 자료를 분석하여 프롬프트가

응답에 영향을 미치는 탐색하였다. 프롬프트 별 응답을 시각화한 결과는 Fig. 4, Fig. 5, Fig. 6, Fig 7과 같다.

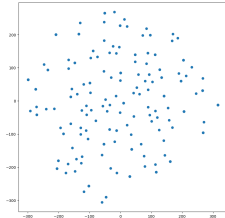


Fig. 4. Prompt 1

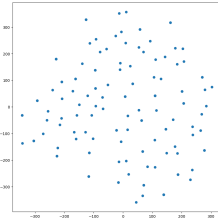


Fig. 5. Prompt 2

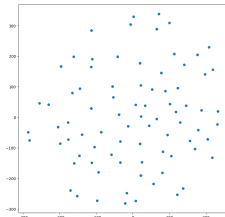


Fig. 6. Prompt 3

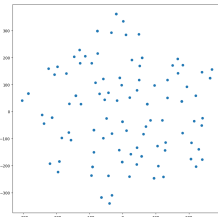


Fig. 7. Prompt 4

프롬프트 별 결과 비교를 위해 다시 2개 이상의 자료를 묶어 시각화 하였으며 그 결과는 Fig. 8, Fig. 9, Fig. 10, Fig. 11, Fig. 12, Fig. 13이다.

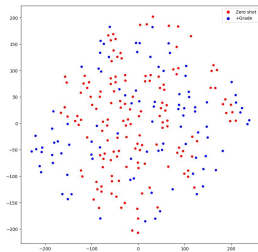


Fig. 8. Prompt 1 and 2

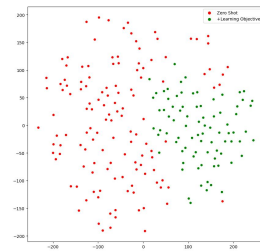


Fig. 9. Prompt 1 and 3

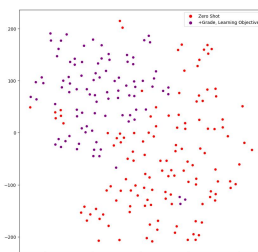


Fig. 10. Prompt 1 and 4

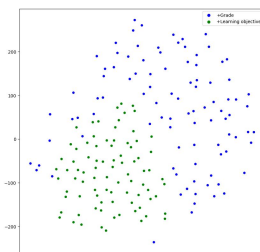


Fig. 11. prompt 2 and 3

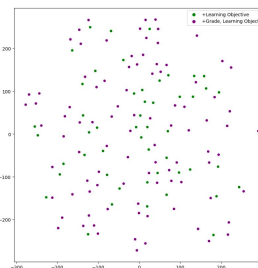


Fig. 12. prompt 3 and 4

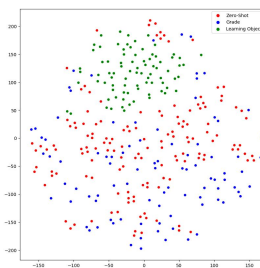


Fig. 13. prompt 1,2 and 3

Prompt 1은 빨간색, 2는 파란색, 3은 초록색, 4는 보라색으로 시각화하였다.

Fig. 8은 Prompt 1과 2를 비교하여 나타낸 것으로 두 자료 간의 구분이 뚜렷하지 않고 비슷한 영역에 분포하는 것을 볼 수 있다. Prompt 2는 Prompt 1에 수직 계열의 성격을 가진 학년 정보를 추가하였는데, 학습자의 수준은 내용의 변화보다는 같은 낱말을 조금 더 쉬운 표현으로 바꿔 나타내는 방식 등의 변화로 나타나 문장을 임베딩 하고 차원을 축소하는 과정에서 유사한 의미영역으로 표현되는 것으로 보인다.

Fig. 9는 Prompt 1과 3을 비교하여 나타낸 것으로 두 자료가 표현된 영역이 뚜렷하게 구별되어 보인다. Prompt 3은 Prompt 1에 학습 목표를 추가한 것으로 학습 목표는 교수·학습 자료를 어떤 목적으로 사용할지에 대한 방향성을 담고 있어 자료의 내용을 구체화한다. 따라서 Prompt 1보다 나타내는 내용이 더 구체적이며, Prompt 3이 나타내는 영역은 Prompt 1보다 집중된 것으로 보인다.

Fig. 10은 Prompt 1과 4를 비교하여 나타낸 것으로 두 자료 간의 구분이 비교적 뚜렷하게 나타난다. Prompt 4는 학년 정보와 학습 목표를 추가하여 가장 많은 정보를 제공하는 프롬프트이다.

Fig. 11은 Prompt 2와 3을 비교하여 나타낸 것으로 두 자료 간의 차이가 뚜렷하게 나타나 보인다.

Fig. 12는 Prompt 3과 4를 비교하여 나타낸 것으로 두 자료 간의 차이가 뚜렷하게 나타나지 않는다.

Fig. 13은 Prompt 1,2,3을 비교하여 나타낸 것으로 초록색으로 표현되는 Prompt1,2는 비슷한 영역에 나타나고 Prompt 3은 뚜렷하게 구분되어 나타난다.

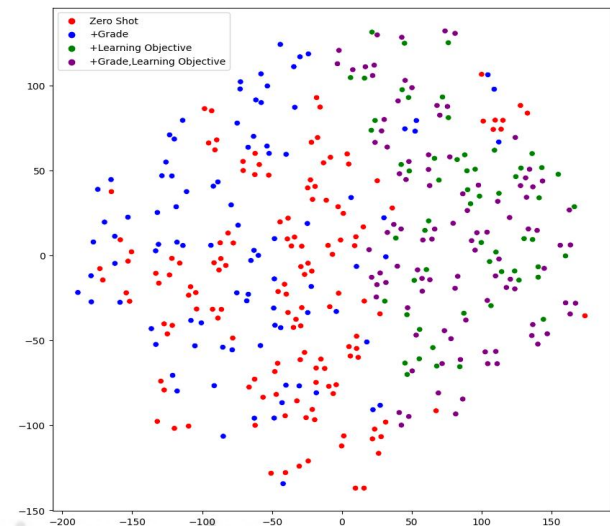


Fig. 14. prompt 1,2,3 and 4

Fig. 14는 Prompt 1,2,3,4를 모두 나타낸 것으로 대체로 Prompt 1,2와 3,4는 각각 비슷한 영역에 분포하고 있으며, 1,2와 3,4는 좌, 우로 나뉘어 분포되었다.

시각화 하여 분석한 결과로 미뤄볼 때 프롬프트를 구성하는 변인 중 학습 목표는 교수·학습 자료를 어떤 목적으로 사용할지에 대한 내용을 담고 있어 해당 키워드의 내용을 구체화하면서 GPT가 생성하는 응답의 방향성에 영향을 미쳐 각 응답 좌표의 분포가 수평 이동하거나 더 집중되는 등의 형태로 나타나는 것으로 보인다.

학년 정보는 학습자의 수준을 담고 있어 내용이나 의미 자체의 변화에 영향을 주기 보다는 같은 의미나 내용에서 어려운 낱말을 쉬운 낱말로 바꾸는 등의 형태로 나타나 그러한 변화가 문장 임베딩과 차원 압축 과정에서 비슷한 영역으로 표현이 되는 것으로 보인다.

1.3. Prompt Selection and Evaluation

설계한 프롬프트가 추천한 교수·학습 그림 자료의 교육적 가치를 평가하기 위해 100개의 응답 중 대표 응답을 선정하여 Stable Diffusion을 이용하여 이미지화 하였다.

대표 응답을 선정하는 방법으로 k-means clustering algorithm을 이용하여 프롬프트 응답의 임베딩 벡터 값을 10개의 클러스터로 군집 하였으며 가장 큰 클러스터에 속한 응답이 확률적으로 대표성을 띤 의미 영역으로 판단하고 해당 클러스터의 첫 번째 원소를 탐색하였다. 클러스터링을 시각화 한 결과는 Fig. 15, Fig. 16, Fig. 17, Fig. 18과 같고 프롬프트 별 대표응답으로 선정한 결과는 Table 3과 같다.

Table 3의 프롬프트 별 대표 응답을 Stable Diffusion에 입력하여 이미지로 생성 하였다. Stable Diffusion도 생성형 인공지능 모델로 같은 프롬프트를 입력하여도 입력할 때마다 다른 이미지가 생성된다. 따라서 평균적인 결과를 얻기 위해 Stable Diffusion에 입력한 프롬프트 별 4개의 이미지를 생성하였으며 생성된 이미지는 각각 Fig. 19, Fig. 20, Fig. 21, Fig. 22와 같다.

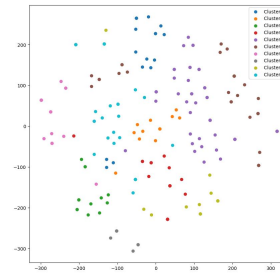


Fig. 15. Prompt 1

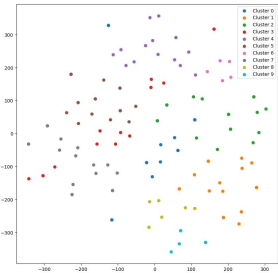


Fig. 16. Prompt 2

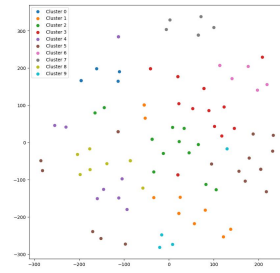


Fig. 17. Prompt 3

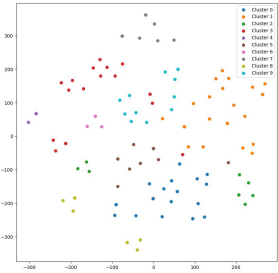


Fig. 18. Prompt 4



Fig. 19. Prompt 1

Fig. 20. Prompt 2

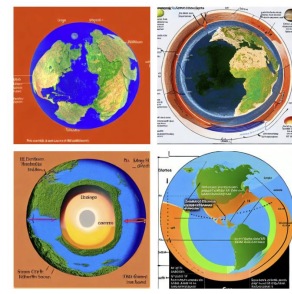


Fig. 21. Prompt 3



Fig. 22. Prompt 4

Table 3. Prompt Selection

Prompt	Most Common Cluster	Element index	GPT Response
1	4	0	A breathtaking aerial view of the Earth from space showcasing its vibrant blue oceans lush green continents and wisps of white clouds.
2	2	11	A group of children recycling different materials, promoting responsible waste management and environmental stewardship.
3	5	3	A diagram of the Earth's layers, depicting the spherical structure from the core to the crust.
4	1	0	A photograph of a globe, clearly showing the round shape of the Earth and the continents.

Fig. 19는 Prompt 1의 GPT 응답을 입력하여 얻은 그림으로 우주에서 바라본 지구의 모습이 나타났다. Fig. 20은 Prompt 2의 GPT 응답을 입력한 결과이며 지구의 환경 보호를 위해 활동하는 아이들의 모습이 나타났다. Fig. 21은 Prompt 3의 GPT 응답을 입력한 결과이고 지구 내부를 구성하는 층을 도식화 한 모습이 나타났다. Fig. 22는 Prompt 4의 GPT 응답을 입력한 결과이며 지구의 전체 모습을 조망할 수 있는 지구본이 나타났다.

생성한 그림 자료의 가치를 판단하기 위해 평가 집단을 Table 4과 같이 구성하였다.

Table 4. Human Teacher Group

Total	Teaching Experience Group (Year)		Experience in teaching the learning subject of this study	
	1~3	6	Y	12
30	5~10	7	N	18
	10 above	17		

평가 집단은 초등학교 교사 30명이며 교육 경력 별로 인원을 구성하였다. 교육 경력 별 인원은 10년 이상인 교사가 17명, 5~10년인 교사는 7명, 1~3년인 교사는 6명이고 사전 설문 결과 본 연구에서 선정한 교육 주제로 수업을 해본 경험이 있는 교사는 12명, 그렇지 않는 교사는 18명으로 나타났다.

평가 도구는 한국교육과정평가원(2002)의 교수·학습 자료 평가 기준을 사용 하였으며 초등 교육 전문가 3인의 자문에 따라 그림 자료와 직접 관련이 없는 항목을 제외하고 1. 목표 및 내용 영역과 4. 기능영역 만 활용하였으며 평가 도구는 Table 5와 같다.

평가 집단이 응답한 결과를 독립변인 t-검정하였으며, 프롬프트 설계가 없는 Prompt 1과 프롬프트 엔지니어링을 통해 개선한 프롬프트 간 효과성을 검증하기 위해서 비교군은 Prompt 1을 대조군은 Prompt 2,3,4로 설정하여 분석하였다. t-검정 결과는 Table 6과 같다.

Table 6. Evaluation result t-test (N: 30)

Items		t-test Result				
A	E	Prompt	m	SD	t	p
1	Q1	1	1.66	0.66		
		2	1.46	0.77	1.07	0.28
		3	2.0	0.63	-0.39	0.03
		4	2.33	0.80	-3.51	0.001
	Q2	1	2.46	0.62		
		2	1.93	0.90	2.64	0.01
		3	2.36	0.66	0.59	0.55
		4	2.4	0.72	0.38	0.7
	Q3	1	2.33	0.71		
		2	1.83	0.83	2.49	0.01
		3	2.13	0.73	1.07	0.28
		4	2.13	0.89	0.95	0.34
4	Q4	1	2.73	0.52		
		2	2.23	0.89	2.63	0.01
		3	2.46	0.77	1.56	0.12
		4	2.83	0.46	-0.78	0.43
	Q5	1	2.56	0.62		
		2	1.9	0.88	3.36	0.001
		3	2.06	0.82	2.63	0.01
		4	2.5	0.73	0.37	0.70
	Q6	1	2.33	0.80		
		2	1.3	0.71	5.10	0.001
		3	1.83	0.87	2.30	0.02
		4	2.33	0.80	0.5	1.00

*p<.05

교육과정 구성이나 학교 급별 교육목표에 대한 평가에서는 Prompt 2의 평균이 가장 낮은 것으로 나타났으나 p값이 0.28로 Prompt 1과 비교하였을 때 유의미한 차이가 나타나지 않은 것으로 판단된다. 가장 높은 평가를 받은 프롬프트는 Prompt 4로 p값이 0.001로 Prompt 1과 통계적으로 유의미한 차이가 나타난 것으로 판단된다. 두 번째로 평균이 높은 프롬프트는 Prompt 3으로 p값은 0.03으로 Prompt 1과 통계적으로 유의미한 차이가 나타난 것으로 보인다. Prompt 3,4가 학습 목표를 포함하고 있기 때문에 교육과정 구성이나 학교 급별 교육목표에 적합하다 판단을 받은 것이며 이는 본 연구의 프롬프트 엔지니어링이 학습 목표에 부합하는 교수·학습 자료를 추천하는 데 긍정적인 영향을 미치는 근거로 볼 수 있다.

Table 5. Excerpts from some of the criteria for evaluating teaching and learning materials

Area	Element	Criteria	Detailed criteria		
			3	2	1
1. Goals and content areas	Reflect the Curriculum	Is the direction of the curriculum composition and the educational goals of each school level appropriately reflected?			
	An Educational Perspective	Is the content of education treated without prejudice?			
4. Functional area	External Conditions	Is it designed to attract the interest and interest of teachers and learners?			
	Ease of Access	Is it easy to obtain and prepare materials included in the teaching and learning process?			
	Ease of Use	Is it convenient for teachers and learners to use the materials?			
	Availability	Is it likely to be used by teachers and learners in the future?			

전반적인 평가에서 Prompt 2가 가장 낮은 평균 점수를 보이고 있는데 Prompt 2는 학년 정보만을 담고 있어 주제어인 'Earth'와 관련된 폭넓은 주제 중 하나를 추천하였고, 이에 학습 목표와는 다소 동떨어진 그림 자료를 생성했기 때문으로 보인다.

학습 자료의 실제 활용 가능성에 대해서는 Prompt 1, 4의 평균 점수가 높게 나타나 실제 수업 자료로 활용될 수 있는 가능성이 있음을 볼 수 있다. Prompt 1보다 Prompt 4의 교육과정 구성이나 학교 급별 목표의 적합성이 높음은 것으로 보아, Prompt 4의 그림 자료를 수업 주 자료로, Prompt 1의 그림 자료를 보조 자료로 활용할 수 있을 것으로 보인다. Prompt 2, 3은 Prompt 1에 비해 실제 수업에 사용될 가능성이 낮은 것으로 나타났는데, 이는 Prompt 2는 학습 목표와 다소 동떨어진 주제가 나타났고 Prompt 3은 학습자의 수준에 적절하지 않았기 때문으로 보인다.

V. Conclusion

본 연구에서는 생성형 인공지능을 이용한 교수·학습 자료 추천 성능 향상을 위한 프롬프트 엔지니어링에 대해 탐색하였다. 교수·학습 자료의 종류로는 그림 자료를 이용하였다.

프롬프트 구성이 생성형 인공지능의 응답에 미치는 영향을 탐색하기 위해 Zero-Shot 프롬프트, 학습 대상 학년 정보가 담긴 프롬프트, 학습 목표가 담긴 프롬프트, 학습 대상 학년과 학습 목표가 모두 담긴 프롬프트 총 4종을 설계하여 GPT-3.5모델에 입력하고 응답을 수집하였다.

수집한 응답을 Sentence Transformers를 활용하여 임베딩 하고 t-SNE로 차원 축소하여 시각화 한 다음 프롬프트와 응답 간의 관계를 탐색하였다.

시각화하여 탐색한 결과 프롬프트 설계에서 학년 정보와 학습 목표 중 응답에 영향 더 미치는 것은 학습 목표로 보인다. 그 까닭은 수업에 대한 정보 중 수평적인 성격을 갖는 학습 목표 진술은 학습 주제를 구체화하면서 GPT 응답의 내용에 영향을 미친 것으로 보인다. 학년 정보는 수직적 성격을 가지는 것으로 내용 영역보다 같은 내용에 대해 GPT 응답 표현의 난이도를 조절하는데, 그 영향이 문장 임베딩 과정에서 유사한 단어로 처리되어 상쇄된 것으로 보인다.

프롬프트 별 응답의 대푯값을 선택하여 Stable Diffusion을 이용하여 시각화 한 다음 교사 30명에게 교

수·학습자료 평가 기준에 따라 평가 받았다.

교사 30인은 추천한 4종의 그림 자료 중 3종은 교육적 가치가 있다는 평가를 하였으며 그 중 2종은 실제 수업에 활용할 의사가 있다고 판단하였다. 가장 수업과 관련된 그림 자료를 추천한 프롬프트는 대상 학년과 학습 목표가 모두 담긴 프롬프트로 나타났다. 이에 연구를 통해 생성한 이미지는 교수·학습 자료로서 의미가 있는 것으로 보인다고 할 수 있다.

교수·학습 그림 자료를 얻기 위해 google에 학습 목표인 '지구의 모양 알아보기'를 검색하는 경우 나타나는 검색 결과는 약 1,380,000개이며 그 중 검색엔진에서 관련 있는 이미지로 추천하는 자료는 35개에 불과하다. 이는 검색 결과의 0.0025%에 해당한다. 본 연구에서 제안하는 이미지 생성 방식의 교수·학습 자료 추천 과정은 생성한 전체 16개의 이미지 중 12개인 75%의 자료가 학습 목표와 관련이 있는 것으로 나타났다. 따라서 연구 결과로 미루어 볼 때 생성형 인공지능은 효율적인 교수·학습 자료 수집을 위한 교수·학습 자료 추천 시스템에 활용할 수 있을 것으로 보인다.

REFERENCES

- [1] Ggwang, "Natural Language Analysis of Korean Texts of AI-based Chatbots and Exploration of Korean Education Utilization - Focusing on ChatGPT and New-Bing," The Society of Korean Culture and Convergence, Vol.45, No.5, 1-17, May 2023 DOI: 10.33645/cnc.2023.05.45.01
- [2] Shlee, Kssong, "Exploring the possibility of using ChatGPT and Stable Diffusion as a tool to recommend picture materials for teaching and learning", Journal of the Korea society of computer and information, Vol. 28, No. 4, pp. 209-216, April 2023 DOI: 10.9708/jksci.2023.28.04.209
- [3] Igyou, Hypark, "Developing an AI-based Sentence-Generating Web Service for Writing Activities in Elementary Language Education," Journal of Research in Curriculum & Instruction, vol. 27, No. 2, pp. 210-221, April 2023 DOI: 10.24231/rici.2023.27.2.210
- [4] Hslee, Hsshim, "Study on the Design of a ChatGPT-Based Metaverse Platform Model" Journal of Industrial Technology Research, Vol.28 No.2, PP131-136, June 2023
- [5] Gwyong, "Prompt engineering for improving the performance of CLIP-based defect detection," Master's Degree thesis, Yonsei University, Dec 2022.
- [6] Isjoen, Kssong, "Development of Block-based Code Generation and Recommendation Model Using Natural Language Processing

Model,” JOURNAL OF The Korean Association of information Education, Vol. 26, No. 3, pp197-207, June 2022 DOI: 10.14352/jkaie.2022.26.3.197

- [7] Ekin, Sabit (2023): Prompt Engineering For ChatGPT: A Quick Guide To Techniques, Tips, And Best Practices. TechRxiv. Preprint. DOI: 10.36227/techrxiv.22683919.v2
- [8] Wei, Jason, et al. "Finetuned language models are zero-shot learners." arXiv preprint arXiv:2109.01652 (2021). DOI: 10.48550/arXiv.2109.01652
- [9] Wei, Jason, et al. "Chain-of-thought prompting elicits reasoning in large language models." *Advances in Neural Information Processing Systems* 35 (2022): 24824-24837.
- [10] Yao, Shunyu, et al. "Tree of thoughts: Deliberate problem solving with large language models." arXiv preprint arXiv:2305.10601 (2023). DOI: <https://doi.org/10.48550/arXiv.2305.10601>
- [11] Hjrpk, Hsim, "A Study on Use Case Analysis and Adoption of NLP:Analysis Framework and Implications," *Journal of Information Technology Services*, Vol. 21, No. 2, pp. 61-84, April 2022 DOI: 10.9716/KITS.2022.21.2.061
- [12] Reimers, Nils, and Iryna Gurevych. "Sentence-bert: Sentence embeddings using siamese bert-networks." arXiv preprint arXiv:1908.10084 (2019). DOI: 10.48550/arXiv.1908.10084
- [13] SBERT, <https://www.sbert.net/>
- [14] Van der Maaten, Laurens, and Geoffrey Hinton. "Visualizing data using t-SNE." *Journal of machine learning research* 9.11 (2008).
- [15] Swjeon et al, "Document Summarization Using TextRankBased on Sentence Embedding," *Journal of KIISE*, Vol. 46, No. 3, pp. 285-289, Dec 2019 DOI : 10.5626/JOK.2019.46.3.285
- [16] Jmha, Gjmoon, "An Application of k-Means Clustering to Vehicle Routing Problems," *Journal of Korean Society of Industrial and Systems Engineering*, , Vol. 38, No. 3, pp. 1-7, Sep 2015. DOI : <https://doi.org/10.11627/jkise.2015.38.3.01>
- [17] KICE, "A Study on the Development of Teaching and Learning Data Types and Standards according to the Revised Curriculum - Focusing on secondary technology, home, art, and English-," April, 2008.

Authors



Soo-Hwan Lee received the B.S. , M.S. degrees in Computer Education from Gwangju National University of Education, Korea, in 2011, 2018. He joined Elementary school teacher at Yangji

Elementary school, Gwangju, Korea, in 2012. He is currently Ph.D. student in KNUE and a Teacher in Oh-hyun Elementary school. He is interested in Computer Education, Generative AI, Software Education and AI Education.



Ki-Sang Song received the B.S., M.S. and Ph. D. degrees in Electrical and Electronics Engineering area from Ajou University, KAIST and University of Washington in 1983, 1985 and 1994, respectively.

Dr. Song joined the faculty of the Department of Computer Education at Korea National University Education, Chungju, Korea, in 1995. He is currently a Professor in the Department of Computer Education, Korea National University Education. He is interested in Edutech, Learning Analytics, Artificial Intelligence Application to Education.