

# 공작기계 매뉴얼 특화 생성형 언어 모델 개발

팀 말랑둥이

( 2023. 09. 29. )

경희대학교 산업경영공학과

지도교수 엄주명, 팀원 김시연, 박주영, 이예원, 조성우

## Abstract

공작기계 산업 현장에서 CNC 컨트롤러를 사용하면서 작업자는 다양한 기계와 소프트웨어에 대한 매뉴얼 북의 내용을 전부 숙지하고 있을 수 없다. 종이로 된 매뉴얼 가이드북을 사용하는 작업자는 다양한 상황에 대처하는 것이 어려울 수 있고 대개 노하우를 기반으로 문제를 해결하게 된다. 노하우 기반의 문제 해결은 휴먼 에러를 유발하고 유희시간 증가와 대기 손실 발생을 일으킨다는 문제점이 있다. 본 연구에서는 이러한 산업 현장에서의 문제를 해결하기 위해 작업자의 질문에 대답함으로써 문제를 즉각적으로 해결할 수 있게 하는 AI 기반 생성형 언어 모델(이하 Manual GPT)을 개발하는 것을 목적으로 한다.

데이터셋은 GPT3.5 Turbo를 사용해 prompt engineering 기반으로 구축하였다. 데이터는 크게 매뉴얼 가이드북의 전반적인 내용에 대한 질문과 답변으로 이루어진 매뉴얼 가이드북 QA 데이터와 공작기계 작동에 있어 발생할 수 있는 알람(alarm)에 대한 질문과 답변으로 이루어진 알람 QA 데이터의 두 가지로 이루어져 있다. Manual GPT의 모델은 수차례의 실험을 통해 Llama2 모델로 선정되었으며 이를 fine-tuning하여 최종 모델을 만들었다. Epoch test를 통해 적절한 하이퍼 파라미터를 결정하였으며 training loss와 답변의 정확도를 척도로 모델을 평가함으로써 최종 모델을 선정하였다. 최종 모델은 모바일 애플리케이션에 탑재하여 작업자가 간편하게 사용할 수 있도록 하였다.

**Key Words:** 공작기계, CNC 컨트롤러, Manual GPT, 생성형 언어 모델, prompt engineering, NLP

## 1. 서 론

### 1.1. 연구의 필요성 및 목적

공작기계 산업의 발전에 따라, 공작기계와 이를 제어하는 CNC (Computer Numerical Control) 컨트롤러의 종류도 다양해지고 있다. 이에 작업자들은 더 많은 공작기계와 CNC의 매뉴얼 가이드북을 숙지해야 한다는 부담을 가지게 된다. 작업자들은 수천 쪽씩 되는 매뉴얼 가이드북을 암기할 수 없으므로 산업 현장에서 문제가 발생할 때마다 두꺼운 매뉴얼 가이드북을 찾아보거나 그동안 쌓은 노하우를 기반으로 이를 해결하곤 한다. 이런 방식의 문제 해결은 또 다른 문제점을 낳게 된다. 책자를 찾으며 문제를 해결하면 그 양이 방대하므로 문제를 해결하는 데 소요되는 시간이 늘어나게 된다. 노하우를 기반으로 문제를 해결할 때는 휴먼 에러가 발생할 수 있고 이에 따른 시간 손실이 발생한다. 문제 해결을 위해 전문 엔지니어를 부르더라도 문제는 곧바로 해결되지 않는다. 관련 논문에 따르면 CNC 기계에 결함이 발생할 때 초급 유지보수 기술자들의 경험 부족과 결함 원인을 신속하게 찾을 수 있는 경험이 풍부한 엔지니어 수의 부족으로 결함 진단에 문제를 겪고 있음을 알 수 있었다[1, 2].

이러한 문제점을 해결하기 위해 작업자가 공작기계와 CNC 컨트롤러를 능숙하게 사용할 수 있도록 도움을 주는 assistant program 및 system이 필요할 것이다. 이에 본 연구에서는 공작기계 산업 현장에서 작업자의 요청에 따라 문제 사항을 파악하고 해결 방법의 즉각적인 제시를 통해 작업의 효율을 극대화하는 AI 기반 생성형 언어 모델, Manual GPT를 개발하고자 한다.

## 1.2. 연구 내용 및 방법

- 1) PDF 텍스트 전처리: 우선, CNC 공작기계의 매뉴얼 가이드북 pdf의 텍스트를 추출하고 전처리한다. 이 단계에서는 텍스트 정제, 토큰화, 및 문장 구조 유지를 위한 작업이 수행된다.
- 2) QA 데이터셋 생성: 전처리한 텍스트를 기반으로 QA(질문-답변) 데이터셋을 생성하기 위해 트랜스포머 모델을 활용한다. 이 데이터셋은 매뉴얼 가이드북의 전반적인 내용과 작업 현장에서 발생할 수 있는 경보를 중심으로 한 QA 데이터로 향후 모델 학습 및 평가에 사용된다.
- 3) 모델 학습 및 평가: 후보 모델(GPT2, Llama2)은 생성된 QA 데이터셋을 기반으로 학습된다. 학습된 모델은 정성적 및 정량적 평가를 거쳐 성능을 평가하고 최종 모델을 선정한다.
- 4) 상용화: 사용자 친화적인 UI를 갖춘 모바일 애플리케이션을 개발하여 최종 모델의 사용성을 향상한다.

## 1.3. 관련 연구

### 1.3.1. 생성형 언어 모델

관련 논문을 통해 우리는 LSTM이나 RNN보다 더 나은 퍼포먼스를 보여주는 트랜스포머 모델에 관한 내용과 이 트랜스포머가 BERT, GPT 등 자연어 처리 모델에 어떻게 적용되고 활용되는지에 대한 내용을 알 수 있다[3, 4]. 또한 현재 시점 NLP의 강자인 OpenAI사에 Microsoft의 지속적인 투자가 이루어지고 있으며 최근 두 회사가 파트너십을 강화했다는 점과 국내 게임사인 NC소프트에서 멀티모달 생성 AI를 개발하였고 이후 한국어/영어 모두 사용 가능한 이중 언어 모델을 공개했다는 점에서 Manual GPT 개발의 필요성을 확보할 수 있었다.

### 1.3.2. GPT-2

본 연구에서 다루게 될 언어 모델 중 하나인 GPT-2 모델의 구조와 작동 원리를 분석하였다. GPT-2는 OpenAI에서 개발한 자연어 처리를 위한 딥러닝 언어 모델 중 하나로, 2019년에 공개되었으며 자연어 처리 분야에서 혁신적인 성과를 내는 인공 지능 언어 모델이다. 이 모델의 구조를 이해함으로써 우리는 자연어 생성 및 이해 과제를 보다 효과적으로 해결할 수 있을 것으로 기대한다. GPT-2는 자연어 처리 작업에서 뛰어난 성능을 보이며, 어텐션 메커니즘을 활용하여 문장 내의 단어 간 상호작용을 모델링하는 transformer 아키텍처를 기반으로 한다. GPT-2는 WebText라 불리는 40GB 크기의 거대한 코퍼스와 인터넷에서 수집한 데이터를 결합하여 훈련하였다. 약 117M 개의 가장 작은 GPT-2 모델 파라미터들을 저장하는 데 필요한 저장 공간만 약 500MB이고, 가장 큰 GPT-2 모델 파라미터들을 저장하는 데에는 약 6.5GB의 저장 공간이 필요하다. 다양한 크기와 버전으로 제공되며 작은 모델부터 큰 모델까지 다양한 용도와 요구사항에 맞게 선택할 수 있다. GPT-2와 같은 pre-trained 모델은 특정 작업에 맞게 추가적인 학습을 통해 성능을 향상할 수 있으며, 이로써 새로운 작업에 많은 양의 데이터가 필요하지 않으면서도 효과적인 모델을 생성할 수 있다.

### 1.3.3. Llama2

본 연구에서 다루게 될 언어 모델 중 다른 하나인 Llama2 모델의 구조와 작동 원리 또한 분석하였다. Llama2는 Meta에서 2023년 출시한 오픈소스 언어 모델이다. 공개된 버전은 7B, 13B, 34B, 70B의 4가지로, GPT-2 모델과 비교하여 훨씬 많은 크기의 파라미터를 가지고 있다. Context의 길이는 Llama1 모델보다 2배 정도 늘어난 약 4K 개, 학습된 토큰의 수는 Llama1 모델보다 약 1.5~2배 정도 늘어난 2T 개이다. Llama2 모델은 파라미터의 크기가 큰 모델일수록 train loss가 토큰 처리에 따라 더 급격히 줄어든다. Llama2 모델 또한 GPT-2와 같은 pre-trained 모델로, 특정 작업에 맞게 추가적인 학습을 통해 성능을 향상할 수 있으며, 이로써 새로운 작업에 많은 양의 데이터가 필요하지 않으면서도 효과적인 모델을 생성할 수 있다.

## 2. 본 론

Manual GPT 모델의 모델 아키텍처는 Fig. 1.과 같다.

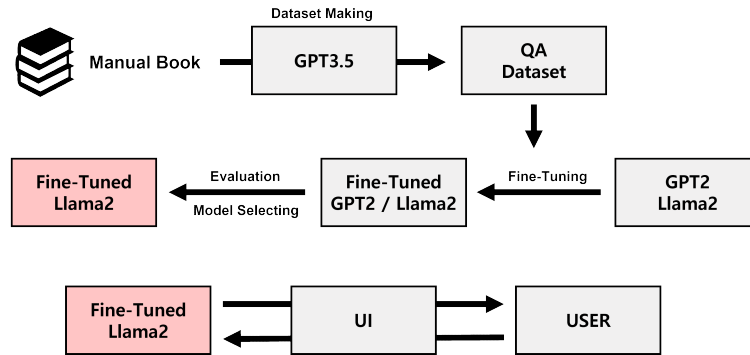


Fig. 1. 모델 아키텍처

## 2.1. 데이터셋

### 2.1.1. 데이터 수집과 전처리

본 연구에서는 공작기계 사용자를 지원하기 위해 pdf 형식의 매뉴얼 가이드북으로부터 텍스트를 추출하여 사용하였다. 연구에 사용된 매뉴얼 가이드북은 'FANUC MANUAL GUIDE i OPERATOR'S MANUAL'이다. FANUC MANUAL GUIDE i는 CNC 기계 도구를 조작하는 데 사용되는 소프트웨어이고, operator's manual은 이 소프트웨어의 사용자 매뉴얼로 CNC 작업자나 프로그래머가 FANUC MANUAL GUIDE i를 사용하여 CNC 기계 작업을 수행할 때 필요한 지침과 도움말을 제공한다. 이 pdf에서 텍스트를 추출하는 것은 자동화된 방식으로 이루어졌으며, 추출된 텍스트 데이터를 이후 활용하기 위해 전처리를 진행하였다.

추출된 텍스트 데이터로부터 패턴을 발견하고, 이를 기반으로 문서의 구조를 이해하였다. 텍스트 데이터는 title, heading, content로 나누어 저장하였다. 이때 content는 이후 GPT-3.5 Turbo 모델에 넣을 것이므로 content의 토큰 수는 모델이 처리할 수 있는 최대 토큰 수인 약 4천 개에 맞추어 정제하였다.

### 2.1.2. 매뉴얼 가이드북 QA 데이터 생성

전처리가 끝난 데이터의 title, heading, content를 합쳐 context를 생성하였다. 이후 OpenAI API를 호출하여 GPT-3.5 Turbo 모델을 사용하여 질문을 생성하였다. 질문 생성에 사용한 프롬프트는 다음과 같다.

You are going to make some questions that may arise using a machine equipped with the FANUC MANUAL GUIDE i. A part of the manual book will be provided. Please write questions based on the context below.

Context:{context}

Questions:

1.

질문 데이터를 생성한 뒤 같은 모델에 또 다른 프롬프트를 사용하여 답변을 생성하였다. 답변 생성에 사용한 프롬프트는 다음과 같다.

Here are some questions that may arise using a machine equipped with FANUC MANUAL GUIDE i. A part of the manual book will be provided. Please answer to the questions based on the context below.

Context:{context}

Questions:

{questions}

Answers:

1.

QA 데이터 생성 과정에서는 GPT-3.5 Turbo 모델에 사용한 프롬프트에 FANUC MANUAL GUIDE i를 언급함으로써 현장의 작업자에게 유용한 질문과 답변을 자동으로 생성하였다. Table 1은 자동으로 생성된 QA 데이터 중 하나를 가져온 것이다.

Table 1. 매뉴얼 가이드북 QA 데이터 예시

Q	I would like to ask a question about MANUAL GUIDE i of FANUC. How is the bottom size/angle determined for turning when [WIDTH] is specified?
A	The bottom size/angle is determined for turning when [WIDTH] is specified by the difference between the groove entry of the side face on the reference position side and the width of the groove bottom. The inclination angle of the side face on the reference position side is determined when [ANGLE] is specified. If no value is entered, it is considered as 0 for both.

### 2.1.3. 알람 QA 데이터 생성

알람에 대한 부분은 작업자가 가장 많은 오류를 만날 수 있는 곳 중 하나이므로, 더 많은 데이터를 생성하였다. 매뉴얼 가이드북의 'Appendix'의 'Alarms' 섹션에서 알람 번호, 발생원인, 해결 방법을 추출하여 데이터프레임으로 만들었다. 이를 기반으로 두 가지 유형의 QA를 생성하였다.

첫 번째는 알람 발생원인이 주어졌을 때 해결 방법을 묻는 질문에 대해 해결 방법을 답변으로 제공한다. Table 2는 해당 QA 데이터에 대한 예시이다.

두 번째는 알람 번호가 주어졌을 때 해결 방법을 묻는 질문에 대해 발생원인과 해결 방법을 답변으로 제공한다. 이 때는 첫 번째 경우보다 다양한 질문을 생성하고자 GPT-3.5 Turbo 모델에 prompt one-shot 방식을 사용하였다. 사용한 프롬프트는 다음과 같다.

You will be provided with an example that contains a context, a question, and an answer. Please make some questions and answers based on this example.

The number of questions and answers cannot be less than the number of warning numbers.  
You will first be provided with an example. Consider this example:

Context:

Appendix

Alarms

B. ALARMS APPENDIX Alarm 16i 30i Description Cause The first feed override is invalid. 3062 3562  
Action A value not specifiable as turning or other first feed overrides is entered. Modify the machining program to specify an appropriate value.

Questions:

- 1.What should I do when I receive warning number 3062?
- 2.What should I do when I receive warning number 3562?

Answers:

1. Cause: The first feed override is invalid.  
Action: A value not specifiable as turning or other first feed overrides is entered. Modify the machining program to specify an appropriate value.
2. Cause: The first feed override is invalid.  
Action: A value not specifiable as turning or other first feed overrides is entered. Modify the machining program to specify an appropriate value.

Now, please write questions and answers based on the context below.

Context:{context}

Questions:

- 1.

위 프롬프트를 이용하여 53가지의 질문 형식을 만들었고 해당 형식을 모든 알람 번호에 적용하여 두 번째 알람 QA 데이터를 만들었다. Table 3은 해당 QA 데이터에 대한 예시이다.

Table 2. Cause - Action 대응 QA 데이터 예시

Q	I received an alarm says, "An arbitrary figure is not closed." What should I check when this warning occurs?
A	An arbitrary figure entered for facing, pocketing, or turning is not closed. Modify the machining program to specify a closed figure that uses the same point both as the start and end points.

Table 3. Alarm Number - Cause &amp; Action 대응 QA 데이터 예시

Q	I would like to ask a question about MANUAL GUIDE i of FANUC. I received a warning number 3503. What should I do about this?
A	There is no area that can be machined. The tool is too large for the specified machining area. Modify the machining program to use a smaller tool.

#### 2.1.4. 최종 데이터셋

2.1.2.와 2.1.3.에서 만든 QA 데이터를 통합하여 최종 학습 데이터와 검증 데이터를 만들었다. 이때 검증 데이터는 기존 QA 데이터를 특정 비율로 단순히 분리할 경우, 매뉴얼 가이드북의 챕터별 데이터 수나 알람 번호별 데이터 수가 균등하지 않아질 수 있으므로 이 점에 유의하여 랜덤하게 만들어 주었다. 최종 데이터셋은 약 20,000개의 학습 데이터와 약 7,500개의 검증 데이터로 이루어져 있다. 학습 데이터의 매뉴얼 가이드북 QA 데이터와 알람 QA 데이터는 각각 약 10,000개이고 검증 데이터의 매뉴얼 가이드북 QA 데이터는 약 4,500개, 알람 QA 데이터는 약 3,000개이다.

## 2.2. 모델 학습

Fine-tuning은 이미 학습된 모델을 특정 작업이나 특정 분야에 더 알맞게 추가적인 학습을 통해 성능을 최적화하는 과정이다. 우리는 GPT-2와 Llama2의 2가지 모델을 fine-tuning할 것이고, 해당 2가지 모델은 서로 다른 구조와 학습 방식을 가지고 있으므로 기존에 데이터를 전처리한 것에 추가적인 전처리 과정이 필요하다. 이에 따라 각 모델의 특성과 요구사항에 맞게 데이터를 조정하고 가공하였다. 일반적으로 전처리한 데이터는 question과 answer 열로 구성되어 있으며, 앞서 데이터셋 부분에서 설명한 Table 1, 2, 3이 그 예시이다.

### 2.2.1. GPT-2 Fine-Tuning

초기 단계에서는 원본 데이터를 텍스트 파일 형태로 전환해야 한다. 또한 학습 데이터의 구조나 형식을 조정해야 한다. 이를 위해 우리는 질문 데이터와 답변 데이터에 특별한 토큰을 추가했다. 'question' 열에 있는 모든 질문 앞에는 '[Q]'라는 표시를 붙여주었다. 이 표시는 해당 텍스트가 질문임을 모델에 알려주는 역할을 한다. 마찬가지로 'answer' 열에 있는 답변의 시작 부분에는 '[A]'라는 표시를 붙여주었다. 이 표시는 해당 텍스트가 답변임을 모델에 알려주는 역할을 한다. 이러한 방식으로 모델이 질문과 답변을 명확히 구분하도록 하였다. Table 4는 새로 전처리된 GPT-2의 학습 데이터 예시이다.

Table 4. GPT-2 학습 데이터 예시

[Q]	I would like to ask a question about MANUAL GUIDE i of FANUC. What should be referred to for functions other than those described in this manual?
[A]	The operator's manual for the specific CNC model should be referred to for functions other than those described in this manual.

Hugging Face의 transformers 라이브러리를 사용하여 GPT-2 모델을 불러온 후 학습을 진행했다. 모델 학습은 Google Colab 환경에서 진행했다. GPU는 Colab의 T4를 사용했다. GPT-2 모델 학습에 사용된 주요 파라미터로는 epoch와 batch size가 있다. Epoch는 모델 학습의 기본 단위다. 이는 전체 학습 데이터셋을 한 번 훑고 학습하는 과정을 의미하는데, 예를 들어서 만약 10,000개의 샘플이 담긴 데이터셋을 가지고 있다면, 이 10,000개의 샘플 전부를 사용하여 학습을 완료하면 이것을 1 epoch라고 한다. 모델을 학습시킬 때, 여러 번의 epoch를 거치며 학습을 반복하게 되면, 모델의 성능은 일반적으로 점차 개선된다. 하지만 epoch 수를 지나치게 늘리면, 과적합(overfitting)이라는 문제에 직면할 가능성이 있다. 이는 모델이 학습 데이터에 너무 특화되어 실제 환경에서의 성능이 떨어지는 현상을 말한다. GPT-2 모델 fine-tuning 과정에서는 epoch를 4로 설정하였다. 그 이유는 검증 데이터셋을 통해 학습 과정을 모니터링 해보았을 때, 5번째 epoch부터는 과적합 현상이 나타났기 때문이다. 그리고 batch size는 한 번의 파라미터 업데이트를 할 때 사용하는 샘플의 수를 말한다. batch size는 8로 설정하였다. Fig. 2는 GPT-2모델 학습 데이터와 검증 데이터의 epoch당 손실을 나타낸 그래프이다.

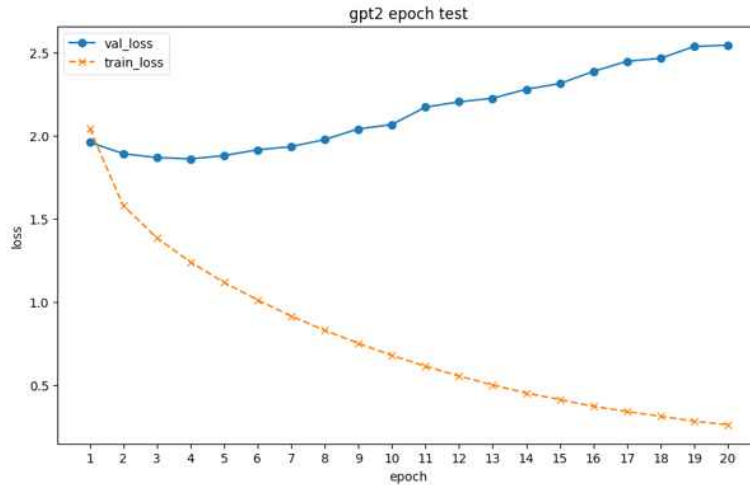


Fig. 2. GPT-2 학습 데이터 및 검증 데이터 손실 그래프

### 2.2.2. Llama2 Fine-Tuning

우선, 질문 텍스트를 처리할 때, 질문의 시작 부분에는 [INST]라는 태그를 붙였고, 끝부분에는 [/INST] 태그를 붙였다. 이렇게 해서 질문의 시작과 끝을 명확하게 구분할 수 있게 했다. 그 후에, 해당 질문에 대한 답변 텍스트를 바로 연결해 하나의 통합된 열을 생성했다. 이렇게 만들어진 통합 열의 시작 부분에는 <s> 태그를, 끝부분에는 </s> 태그를 추가했다. 여기서 <s> 태그는 문장이 시작된다는 것을 의미하며, [INST]는 사용자의 입력 부분을 알려주는 태그다. 예를 들어, Table 5에서는 <s>와 </s> 태그 사이에 전체 질문과 답변 내용이 포함되어 있고, [INST]와 [/INST] 사이에는 질문 부분만 들어가 있다. 마지막으로, 이렇게 전처리된 데이터는 원래 csv 파일 형식으로 저장되어 있었지만, Hugging Face 라이브러리와 호환되도록 하기 위해서는 이를 HuggingFace의 'datasets' 형태로 변환해야 한다. 요약하면, Llama2 모델 학습을 위한 데이터 전처리는 기존 데이터에 특정 태그를 추가하여 문장의 시작, 끝 및 사용자 입력 부분을 구분하고, 이를 HuggingFace 라이브러리와 호환되는 형태로 변환하는 작업을 포함하고 있다.

Table 5. Llama2 학습 데이터 예시

<s>[INST] I would like to ask a question about MANUAL GUIDE i of FANUC. What should I review if I receive warning number 3594? [/INST] The end face specification is invalid. A value not specifiable as the end face of the cycle machining is entered. Modify the machining program to specify an appropriate value. </s>
---

먼저, 'datasets' 라는 라이브러리는 Hugging Face에서 제공하는 것으로, 주로 학습 데이터를 처리하고 변환하는데 사용된다. 학습 데이터를 효과적으로 다루기 위한 다양한 도구와 기능들이 포함되어 있어, 데이터의 형식 변환이나 적재 과정에서 유용하다. 다음으로, 'peft'는 모델 학습에 있어서 효율적인 방법을 제공하는 라이브러리이다. 전체 모델의 모든 매개변수를 미세 조정하는 대신, 일부 매개변수만을 조절하게 되면서 계산량이나 저장 공간 등의 자원을 크게 절약할 수 있게 된다. 이로 인해, Fine-tuning 과정이 더 빠르고 경제적으로 진행될 수 있다. 'Bitsandbytes'는 모델의 크기 최적화를 위한 도구로써, 4bit 양자화 기술을 활용해 모델의 크기를 줄이는 데 효과적이다. 이렇게 함으로써 메모리 제약이 있는 환경에서 매우 유용하다. 마지막으로, 'trl' 라이브러리는 특별한 학습 방법론을 제시한다. 'Transform', 'retrieve', 'label'의 세 가지 과정을 거쳐 학습을 진행하는데, 이는 원본 텍스트 데이터를 변환하고, 그 중에서도 가장 중요하고 관련 높은 정보를 추출하여, 그 정보에 적절한 라벨을 부여하는 과정을 통해 진행된다.

Epoch는 5로 설정하고 Batch size는 4로 설정했다. 학습률(learning\_rate)은  $2e-4$ 로, 가중치 감소율(weight decay)은 0.001로 설정하였다. 이 두 파라미터는 모델 학습의 방향성과 속도를 조절하는 데 영향을 미친다. 최적화 알고리즘으로는 AdamW를 사용하였다. AdamW는 Adam 알고리즘의 변형으로, 가중치 감소가 적용된 알고리즘이다. 이는 모델의 일반화를 도와 과적합을 방지한다. 학습률의 변화를 조절하는 방식으로는 cosine 기반의 스케줄링을 선택하였다. Cosine 기반 스케줄링은 학습률이 훈련 과정에서 cosine 함수를 따라 감소하게 하여, 초기에는 전역 최적점을 찾기 위해 빠르게 탐색하고, 나중에는 지역 최적점에 더 세밀하게 접근하는 방식이다. 이 방법은 모델이 좋은 성능에 도달하면 학습률을 줄여, 성능의 향상을 안정화시킬 수 있다. Llama2는 GPT-2와 마찬가지로 Google Colab 환경에서 진행하였고 GPU는 T4를 사용했다. Fig. 3은 Llama2 모델에 대한 학습 데이터와 검증 데이터의 epoch당 손실을 나타낸 그래프이다.

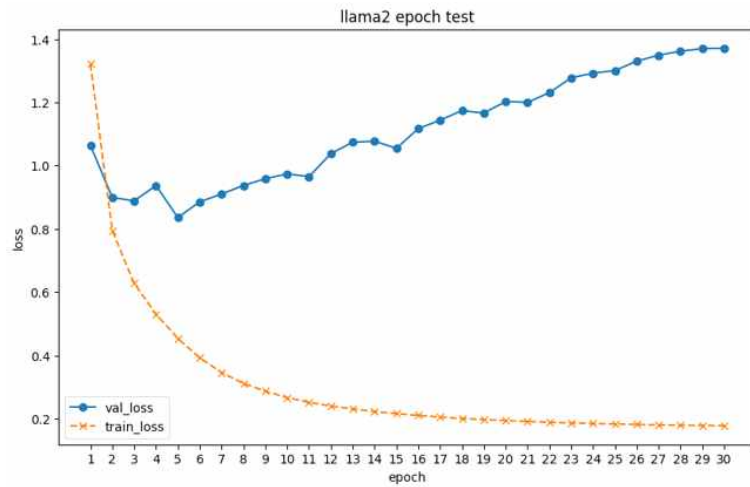


Fig. 3. Llama2 학습 데이터 및 검증 데이터 손실 그래프

### 2.2.3. 모델 학습 결과

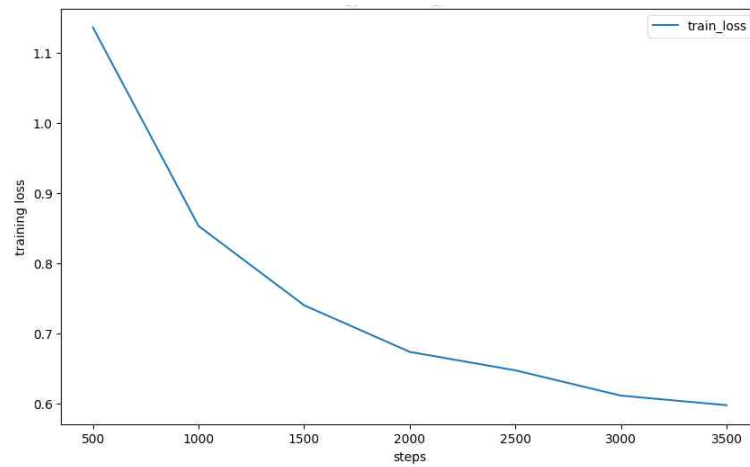


Fig. 4. GPT-2 training loss 시각화

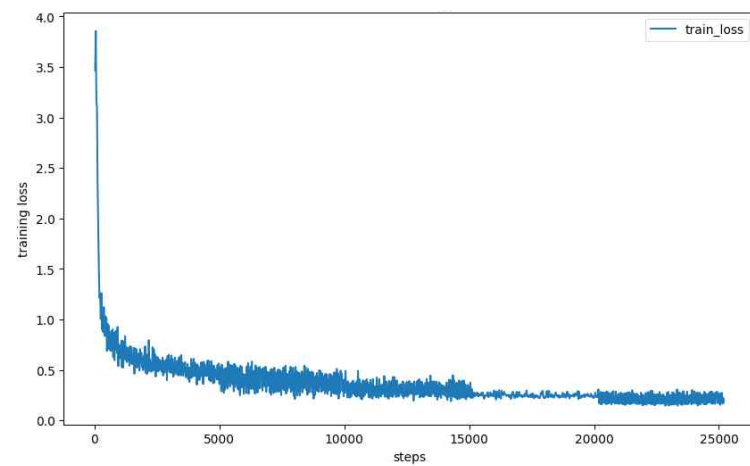


Fig. 5. Llama2 training loss 시각화

GPT-2 모델의 학습 결과를 확인한 결과, 학습 단계(step)가 증가함에 따라 학습 손실(training loss)이 감소하는 것이 관찰되었다. 또한, 이전에 GPT-2 모델에 대하여 적정 epoch을 4로 결정한 이후에는, 모델 학습 과정에서 검증 데이터셋을 별도로 사용하지 않았다. Fig. 4는 GPT-2 모델 학습 결과이다.

Llama2 모델 또한 학습 단계가 증가함에 따라 training loss가 감소한다. 마찬가지로 적정 epoch를 5로 결정한 이후에는 모델 학습 과정에서 검증 데이터셋을 별도로 사용하지 않았다. Fig. 5는 Llama2 모델 학습 손실 결과이다.

### 2.3. 성능 평가 결과

두 가지 모델을 각각 fine-tuned 버전과 baseline 버전으로 분리하여 총 네 가지의 모델에 대한 성능 평가를 진행하였다. 기존에는 baseline GPT-2, baseline Llama2, fine-tuned GPT-2, fine-tuned Llama2의 네 가지 버전을 모두 같은 기준으로 평가하려고 하였다. 그러나 평가를 위해 뽑은 30개의 데이터를 모델에 입력하여 답변을 얻은 결과 baseline의 두 가지 모델은 학습된 모델들에 비해 한눈에 봐도 알 수 있을 정도로 답변의 질이 떨어지고, 모델의 성능이 부족함을 알 수 있었다. 따라서 우리는 fine-tuned GPT-2와 fine-tuned Llama2의 두 가지 버전을 평가하였다. Table 6은 한 가지 예시 질문에 대한 네 가지 모델의 답변 예시이다. 해당 예시를 통해 학습된 모델에 비해 baseline 모델의 성능이 얼마나 부족한지를 알 수 있다.

Table 6. 모델별 질문에 대한 답변 예시

Question	I would like to ask a question about MANUAL GUIDE i of FANUC. How is the "U* ESCAPE AMOUNT" measured for inner surface roughing in turning?	
Answer (Fine-tuned GPT-2)	The "U* ESCAPE AMOUNT" for inner surface roughing in turning is measured by specifying the distance between the reference position of two adjacent grooves, specified as a positive radius value (radius value).	
Answer (Fine-tuned Llama2)	The "U* ESCAPE AMOUNT" for inner surface roughing in turning is measured as the distance between the current position and the position of the "cut end point" along the tool axis. It represents the distance by which the tool retracts from the cutting surface after completing the roughing operation. The "U* ESCAPE AMOUNT" is used to determine the clearance distance between the cutting surface and the position of the "cut end point" to prevent interference or collisions during the retract operation. It is a positive value that represents the distance by which the tool moves away from the cutting surface.	
Answer (Baseline GPT-2)	I had a "U* ESCAPE AMOUNT" of 100x500i (5mm) and 10ml. Then i used a 5mm Follower to turn it down into a smooth groove which is 4mm square in the first half of the turn. I then used 1 1/8" U* Sleeve (not 6.5mm) and a small tool (3.5mm diameter) to turn it down into a smooth grooved curve in the second half. I then used a small piece of 3mm U* Sleeve with an Follower. I started to find ways that Follower is not the cause, it is just the fact that on my current "turn off" car seat you would have to drive a little bit over 15km.	
Answer (Baseline Llama2)	The "U* ESCAPE AMOUNT" in FANUC's MANUAL GUIDE i for inner surface roughing in turning is measured by calculating the ratio of the tool's cutting edge radius to the workpiece's surface roughness. This ratio is then used to determine the appropriate value for the "U* ESCAPE AMOUNT" parameter in the machining process. The exact method for measuring this value may vary depending on the specific application and workpiece material.	

Fine-tuned GPT2 모델과 Fine-tuned Llama2 모델을 비교 평가하는 방법으로는 예시 질문에 대한 답변이 모범 답변과 비교하여 얼마나 정확한지에 대한 정성적 평가 기준을 부여하여 팀 구성원 4명이 각각 평가하고, 해당 평가에 대한 과반의 의견을 점수로 환산하였다. 평가 기준은 다음과 같다. 1) 질문에 대하여 정확히 맞는 답변 또는 해결 방안을 제시하는 경우: O, 1점 부여. 2) 질문에 대하여 정확히 맞는 답변을 포함하지만 관련 없는 내용을 함께 포함한 경우, 질문에 대한 답변이 틀리지는 않았으나 표현이나 용어 선택이 애매하여 작업자의 판단에 의존해야 하는 경우: △, 0.5점 부여. 3) 질문에 대한 답변이 잘못된 경우: X, 0점 부여.

Table 7은 질문에 대한 fine-tuned GPT-2, Llama2 모델의 답변 예시 및 평가 결과이다. Table 8은 평가 결과를 종합한 값이다.

Table 7. GPT-2, Llama2 모델 답변 예시 및 평가 결과

Question #1	I would like to ask a question about MANUAL GUIDE i of FANUC. I received an alarm says, "G codes that cannot be combined together. A figure block not available to the machining type block is specified." What should I check when this warning occurs?	
Answer about Q #1 (Fine-tuned GPT-2)	G codes that cannot be combined together. A figure block not available to the machining type block is specified. Modify the machining program to specify a correct combination of machining type block and figure block.	O
Answer about Q #1	To resolve this warning, check the machining program to ensure that the	O



(Fine-tuned Llama2)	necessary machining type block is available. Modify the program by replacing the machining type block with an appropriate one.	
Question #2	I would like to ask a question about MANUAL GUIDE i of FANUC. What should I do if alarm number 3535 is raised?	
Answer about Q #2 (Fine-tuned GPT-2)	The groove depth specification is invalid. A value not specifiable as turning or other cut directions is entered. Modify the machining program to specify an appropriate cut direction.	X
Answer about Q #2 (Fine-tuned Llama2)	The thread lead specification is invalid. A value not specifiable as the thread lead of a screw thread is entered. Modify the machining program to specify an appropriate value. (Note) The thread lead is the angle under which a thread is made. It is a positive value.	O

Table 8. 모델별 질문에 대한 답변 예시

	GPT-2	Llama2		GPT-2	Llama2		GPT-2	Llama2
1	O	O	11	X	O	21	X	O
2	X	△	12	X	O	22	O	O
3	X	X	13	X	O	23	△	O
4	X	X	14	X	O	24	O	△
5	X	X	15	X	X	25	O	O
6	X	O	16	X	O	26	O	△
7	X	X	17	X	△	27	O	O
8	△	O	18	X	O	28	O	O
9	X	X	19	X	O	29	X	△
10	△	△	20	X	O	30	O	O
GPT-2				Llama2				
O		8		O		18		
△		3		△		6		
X		19		X		6		
점수		9.5		점수		21		

평가 결과, GPT-2 모델은 9.5점, Llama2 모델은 21점으로 Llama2 모델이 GPT-2 모델에 비하여 월등히 우수한 성능을 보이므로 최종 모델로 Llama2 모델을 선정하게 되었다.

## 2.4. 상용화 (UI)

### 2.4.1. 프로토타입 제작

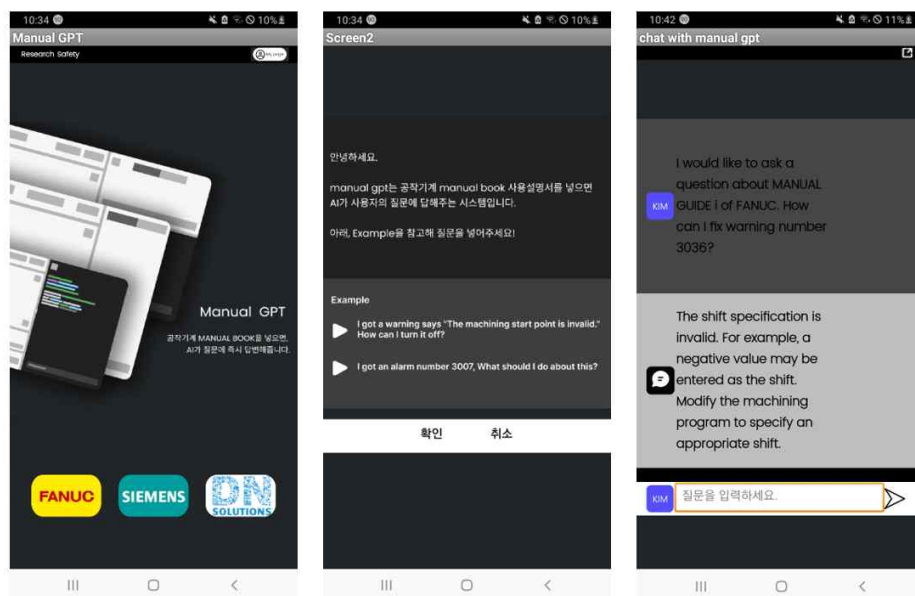


Fig. 6. Manual GPT Prototype

사용자는 모바일 애플리케이션(application, 이하 앱)을 통해 공작기계에 관한 질문을 하고, 그에 대한 답변을 받을 수 있다. 앱 디자인은 구글과 MIT가 만든 안드로이드 앱 프로토타입 툴인 'App Inventor'를 이용해서 진행했다. 앱에 접속하면, 공작기계 3사의 로고가 표시된다. 사용자가 원하는 공작기계에 해당하는 회사의 로고를 클릭하면, 안내 페이지로 넘어간다. 안내 페이지에서는 인사말과 Manual GPT가 어떤 앱인지에 대한 간단한 설명, example 부분의 예시를 참고해서 Manual GPT에 질문을 넣으면 된다는 설명이 제공된다. 이후 '확인' 버튼을 누르면 채팅창이 뜨고, 모델에 질문을 넣었을 때 답변을 받는 채팅이 진행된다. Fig. 6은 Manual GPT 프로토타입 이미지이다.

#### 2.4.2. 서버 컴퓨터 연동

사용자가 앱 입력창에 질문을 입력하면 해당 값을 서버 컴퓨터로 전송해서 fine-tuning한 Llama2 모델을 통해 답변을 생성한다. Llama2 모델이 답변을 생성하면 앱에 전송해 사용자의 화면에 보여준다. 이를 진행하기 위해 Fast API를 사용하여 서버 컴퓨터에 웹 애플리케이션을 만들었고, App Inventor에서 구현하였다.

Fast API는 Python 기반의 웹 프레임워크로, Fast API를 사용한 서버 컴퓨터와 앱 간의 연동은 실제 산업에서 다수의 이점을 제공한다. 먼저, 앱 내에서 모델을 탑재하는 경우 데이터가 변경될 때마다 fine-tuning을 해야 하는 데 시간이 많이 소모된다는 한계점을 극복할 수 있다. 서버 컴퓨터에 fine-tuning된 모델을 탑재하면 작업이 용이해지고 시간을 절약할 수 있게 된다. 또한 복잡한 모델의 경우, 사용자 기기에서 실행하기 어렵기 때문에 서버 컴퓨터에서 실행함으로써 최소한의 리소스 사용이 가능하다. 이렇게 중앙 집중적으로 모델을 관리한다면 데이터를 보호할 수 있고, 운영 및 유지보수가 더 효율적으로 이루어질 수 있다.

이러한 빠른 응답, 데이터 보안, 리소스 최적화, 효율적인 운영 및 유지보수 등의 장점을 가진 Fast API를 사용하여 서버 컴퓨터에 웹 애플리케이션을 구축하였다. Fast API를 이용하여 서버를 설정하고 App Inventor에 입력된 데이터를 받아온다. 받아온 데이터는 Llama2 모델에 입력되어서 출력하도록 설계하였다. 사용자의 질문을 입력으로 받아, 이를 json 형식으로 변환한 후 구축한 서버 컴퓨터 웹페이지로 전송한다. 이후, 서버에서는 Llama2 모델을 활용하여 입력 값을 처리하고, 모델의 출력 값을 가져와 앱 화면에 보여주도록 구현하였다.

### 3. 결 론

#### 3.1. 기대 효과

Manual GPT가 상용화 이후 보급되면 공작기계를 비롯한 산업 현장에서 발생하는 문제점에 대한 빠른 지원 및 문제 해결로 생산성을 향상할 수 있을 것이다. 또한 전문 지식의 공유 및 보존의 장점이 있으며, 특히 새로운 작업자에 대한 교육 및 훈련시간을 절감할 수 있다는 기대효과가 있다. 앞서 서론에서 언급한 다양한 상황대처의 어려움, 노하우 기반의 문제 해결, 휴먼 에러 발생, 유희시간과 대기 손실의 발생 및 증가에 대한 문제를 해결할 수 있을 것이다.

#### 3.2. 한계점 및 향후 연구 방향

본 연구에서 사용한 매뉴얼 가이드북은 약 1,000쪽의 pdf로 FANUC의 CNC 기계에 한정되어 있다. 더 다양한 pdf를 수집하여 모델을 학습한다면 모델의 범용성을 높일 수 있을 것이다. 또한 다양한 pdf를 사용하는 만큼 pdf로부터 텍스트를 추출했을 때 그 패턴을 자동으로 파악하여 QA 데이터를 만들 수 있도록 하는 것도 좋을 것이다.

장비나 기계를 다룰 때는 가장 작은 오류나 부정확한 정보조차도 큰 문제와 손실을 가져올 수 있다. 따라서 정보 제공 시 그 내용의 정확성과 신뢰성은 매우 중요한 요소로 작용한다. 질문과 답변 데이터를 생성할 때 사용한 모델은 해당 매뉴얼 가이드북을 학습한 모델이 아니다. 그래서 기본적인 정보나 지식을 가지고는 있지만, 특정 매뉴얼에 대한 구체적인 세부적인 지식은 부족한 상황이다. 이로 인해 모델 답변에 대한 신뢰성 문제가 생길 가능성이 있다.

생성형 언어 모델은 자연스러운 문장을 만들어내는 데에는 능숙하지만, 너무 긴 문장을 출력하는 경향이 있다. 이런 특성은 일반적인 상황에서는 유용할 수 있으나, 실무 환경에서는 정보의 정확성과 함께 간결함도 중요하게 여겨진다. 긴 설명보다는 짧지만 분명하고 정확한 답변이 필요한 경우가 많기 때문이다. 이를 고려하여, 현재 모델의 한계를 극복하고, 답변의 길이뿐만 아니라 내용의 신뢰성도 높여보고 시도할 것이다.

앱 내에 피드백을 제출할 수 있는 기능을 추가할 예정이다. 이를 통해 앱을 사용하며 어떤 경험을 하고 있는지, 어떤 점이 편리하게 느껴지는지를 알 수 있게 된다. 사용자들은 앱 사용 중에 발생한 경험을 손쉽게 공유할 수 있으며, 우리는 이러한 피드백을 받아들여 앱을 지속해서 개선해 나갈 것이다.

## 후 기

김시연: 공작기계 매뉴얼 가이드북 파일을 읽고 전처리하는 과정에서 매뉴얼 가이드북이 상당히 정형화되어있음을 알 수 있었다. 또한 학습 모델 선정 과정에서 어텐션 메커니즘과 트랜스포머 기반의 다양한 NLP 모델을 접하게 되었고 자연어 처리 분야에 큰 흥미를 느끼게 되었다. 앞으로의 진로 분야를 선택하는 데 큰 도움이 되었던 프로젝트를 하게 되어 뜻깊은 시간이었다.

박주영: 다양한 대규모 언어 모델을 탐색하며, 각 모델의 특성과 장단점을 비교하였다. 그 과정에서, 단순히 성능이 우수한 모델을 선택하는 것이 아니라, 해당 비즈니스의 요구사항과 상황에 가장 적합한 모델을 선정하는 것이 중요하다는 인식을 얻었다. 초기 모델 학습 시에는 기대한 결과를 얻지 못했지만, 모델의 매개변수를 조절하고, 데이터를 세심하게 전처리함으로써, 답변의 질이 향상되는 것을 확인하며 뿌듯함을 느꼈다.

이에원: 프로젝트를 진행하면서 힘들었지만, 즐거움을 더 크게 느낄 수 있었다. 모든 과정이 의미 있었지만, 특히 앱을 제작했던 과정이 가장 기억에 남는 것 같다. 앱을 제작하면서 떠올렸던 아이디어가 실제로 구현되어 가는 것들을 실시간으로 확인하면서 큰 즐거움을 느꼈다. 조금 더 시간이 주어졌다면, 앱 내에 피드백을 제출할 수 있는 기능을 추가하고 싶었지만 하지 못한 것이 가장 아쉬움이 남는다. 공모전이 끝난 이후라도 이 기능을 개발해보고 싶다.

조성우: Manual GPT 프로젝트를 진행하면서 인공 지능뿐만 아니라, 공작기계 전반에 대한 이해와 관심이 늘었다. 공장방문을 통해 산업 현장이 어떤 프로세스로 구성되어 있고 진행되는지 명확히 알 수 있었다. 공작기계 전반에 관한 관심 중 특히, HMI 중 디지털 트윈이 흥미로웠다. Manual GPT 프로젝트가 끝나면 디지털 트윈에 대한 공부와 연구를 해보고 싶다.

## 참고문헌

- [1] Anzari, F., Sugiarto, D., & Luis, M. (2023). A fault diagnosis system for CNC hydraulic machines: a conceptual framework. *SINERGI*, 27(1), 65-72.
- [2] Michael, G., Peng, J., Marinov, M. B., & Ovtcharova, J. (2021, September). Research on fault diagnosis expert system of cnc machine tool based on expert knowledge. In *2021 XXX International Scientific Conference Electronics (ET)* (pp. 1-4). IEEE.
- [3] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- [4] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2020, October). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations* (pp. 38-45).
- [5] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- [6] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.
- [7] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.
- [8] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.
- [9] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... & Lample, G. (2023). Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.
- [10] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.
- [11] Liu, X., Zhang, F., Hou, Z., Mian, L., Wang, Z., Zhang, J., & Tang, J. (2021). Self-supervised learning: Generative or contrastive. *IEEE transactions on knowledge and data engineering*, 35(1), 857-876.
- [12] Khan, J. Y., & Uddin, G. (2022, October). Automatic code documentation generation using gpt-3. In *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering* (pp. 1-6).
- [13] Anagnoste, S., Biclesanu, I., D'Ascenzo, F., & Savastano, M. (2021). The role of chatbots in end-to-end intelligent automation and future employment dynamics. In *Business Revolution in a Digital Era: 14th International Conference on Business Excellence, ICBE 2020, Bucharest, Romania* (pp. 287-302). Springer International Publishing.
- [14] Li, C., Park, J., Kim, H., & Chrysostomou, D. (2021, March). How can i help you? an intelligent virtual assistant for industrial robots. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 220-224).
- [15] Liu, R., & Shah, N. B. (2023). ReviewerGPT? An Exploratory Study on Using Large Language Models for Paper Reviewing. *arXiv preprint arXiv:2306.00622*.
- [16] Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., ... & Yih, W. T. (2020). Dense passage retrieval for open-domain question answering. *arXiv preprint arXiv:2004.04906*.