

Chat GPT의 원리, 활용, 한계와 업무효율화

오상훈 *



1. 들어가며

OpenAI에서 2022년 11월에 발표된 Chat GPT는 인간처럼 자연어 문장을 생성하는 인공지능 서비스로써 그 파급력은 날로 증대되고 있다.¹⁾ 이전에 개발되어 발표된 인공지능 관련 서비스들은 아주 협소한 영역에서 동작하기에 ANI(Artificial Narrow Intelligence)라고 지칭되었다. ANI 분야의 인공지능 시스템들을 살펴보면 필기체 숫자인식, 음성인식, 영상인식, 불량품 검출 같은 인식능력을 지닌 인식(Classification) 인공지능 시스템; 주식시장 예측, 곡물가격 예측, 홍수 예측, 인구 예측처럼 시간에 따라 변하는 데이터의 예측(Time Series Prediction)을 수행하는 예측 인공지능 시스템;

* 목원대학교 게임소프트웨어공학과 교수

1) <https://openai.com/ChatGPT>, 2024.02.15.

SNS 사용자의 정치적 성향 분류, 전자상거래 이용자의 성향 분류 등 데이터의 비슷한 부류를 찾아내는 군집화(Clustering) 인공지능 시스템; 로봇자율주행, 헬리콥터 비행, 게임실행 등 제어(Control) 인공지능 시스템 등으로써 이들은 모두 특정한 목적을 대상으로 개발된 인공지능 기법이었다.²⁾

이와 달리 Chat GPT는 거의 모든 영역에서 사람이 관심을 가지는 요청에 대한 응답을 보이는 인공지능 서비스이기에 ANI보다는 AGI(Artificial General Intelligence) 성격에 많이 가까워진 서비스이다.³⁾

이러한 특성 때문에 사용자가 급격히 증가하면서 사회적인 파급력도 아주 크게 작용하고 있다. 따라서, 이 논문에서는 이제 일반인들도 흔히 접하고 사용하게 된 Chat GPT를 소개하고자 한다. 우선 자연어 처리의 기본 개념을 기반으로 자연어의 임베딩과 언어모델, 그리고 Chat GPT의 기본 원리를 소개한다. 다음으로, Chat GPT를 활용하는 방법에 대하여 알아보자 한다. 그리고, Chat GPT가 가져온 충격과 악용 사례, 사용상의 한계, 업무 효율화 방안에 대하여 설명한다.

2. 자연어 처리와 Chat GPT

1) 자연어 처리

먼저 사람과 대화를 주고받는 인공지능 시스템을 만들려면 어떤 기술이 필요한지 알아보자. 이는 크게 세 단계로 나누어 볼 수 있는데, 첫 단계로 사람이 말하는 음성을 문장으로 전환하여주는 음성인식(Automatic

- 2) Osmo Kuusi·Sirkka Heinonen, “Scenarios from artificial narrow intelligence to artificial general intelligence—Reviewing the results of the International Work/Technology 2050 study”, *World Future Review*, vol.14, 2022, pp.65-79.
- 3) Mostafa M. Amin·Erik Cambria·Bjorn W. Schuller, “Will affective computing emerge from foundation models and general artificial intelligence? A first evaluation of Chat GPT”, *IEEE Intelligent Systems*, vol. 38, 2023, pp.15-23.

Speech Recognition, ASR)이 필요하다.⁴⁾ 이를 통하여 사람이 말하는 내용을 문장으로 전환시켜주어야만 그 다음 단계에서 여러 가지 형태로 사람이 말한 내용에 대한 답을 할 수 있을 것이다. 이렇게 음성인식이 이루어지면, 그 다음 단계로 인식된 문장에 기반 하여 자연어 처리(Natural Language Processing, NLP)가 이루어진다.⁵⁾ 자연어 처리 단계에서는 사람이 말한 문장을 인식한 내용에 대한 대응 차원에서 각종 일 혹은 작업(Tasks)들이 이루어진다. 세 번째 단계로 필요한 기술은 자연어 처리 결과로 얻은 내용을 다시 음성으로 전환하여 사람에게 들려주는 음성합성(Text to Speech, TTS) 기술이다. 이렇게 크게 세 단계를 거치면서 사람이 말한 내용을 인식하고 그에 대한 대응으로 문장을 만들은 후 이를 음성으로 사람에게 들려주게 된다.

이들 세 가지 단계는 각각 상당히 고난이도의 기술이다. 음성인식은 잡음이 없는 환경에서는 잘 동작하는 수준까지 개발되었지만, 잡음이 있는 경우 인식성능이 많이 저하되므로 잡음을 분리하거나 제거하기 위한 기술이 지속적으로 개발되고 있다.⁶⁾⁷⁾ 음성합성은 사람이 말하듯이 자연스러운 음성을 만들어 내기 위한 기술로써, 상용화 단계에서 이미 많은 서비스들이 이루어지고 있다.⁸⁾ 여기서는 Chat GPT와 관련이 깊은 자연어 처리에 대

-
- 4) M. Benzeghiba et al., "Automatic speech recognition and speech variability: A review", *Speech Communication*, vol. 49, 2007, pp.763-786.
- 5) Prakash M. Nadkarni·Lucila Ohno-Machado·Wendy W. Chapman, "Natural language processing: an introduction", *Journal of the American Medical Informatics Association*, vol. 18, 2011, pp.544-551.
- 6) Hyung Min Park·Sang Hoon Oh·Soo Young Lee, "Adaptive noise cancelling based on independent component analysis", *IEE Electronics Letters*, vol. 38, no. 15, 2002, pp.832-833.
- 7) Hyung Min Park·Sang Hoon Oh·Soo Young Lee, "A bark-scale filter bank approach to independent component analysis for acoustic mixtures", *Neurocomputing*, vol. 73, 2009, pp.304-314.
- 8) <구글 딥마인드>, <https://deepmind.google/discover/blog/wavenet->

하여 좀 더 자세히 살펴보자.

자연어 처리는 다시 세부적으로 자연어 이해(Natural Language Understanding, NLU), 자연어 처리(Natural Language Processing, NLP), 자연어 생성(Natural Language Generation, NLG)으로 나눌 수 있다. 이중에 가장 핵심은 자연어 처리로써, 컴퓨터가 인간의 언어를 이해하고 모사하도록 하는 학문분야이다. 이를 바탕으로 각종 정보를 처리하는 데, 언어학, 컴퓨터공학, 인공지능 분야가 복합적으로 융합하여 발전되어 왔다. 특히, 결정된 구조에 맞추어진 정형 데이터의 영역을 벗어나서 구조화 되지 않은 비정형 텍스트 데이터를 처리하기 위한 기술이 필수적이다. 자연어 이해는 자연어 형태의 문장을 이해하는 기술로써 형태소 분류, 개체명 인식, 문서분류 등의 기능에 해당한다. 자연어 생성은 사람이 말하는 듯이 자연어 문장을 생성하여주며 번역, 문서요약, 챗봇 등의 기능에 해당한다.

자연어 처리의 작업(Tasks) 영역을 더 자세히 알아보자. 먼저, 기계번역(Machine Translation)은 원천언어(Source)를 목표언어(Target)로 자동번역 하는 기술이다. 한국어를 일본어나 영어로 자동 번역하거나 다른 외국어를 한국어로 번역하는 기술에 해당한다. 감정분석(Sentiment Analysis)은 주어진 문장의 표현이 긍정/부정 뿐 만 아니라 즐거움, 슬픔, 분노, 놀람, 두려움 등의 감정 상태에 해당하는지를 분석한다. 이는 리뷰, 설문응답, 온라인이나 소셜 미디어를 통한 고객의견 자료 분석 등에 널리 활용된다. 질의자동응답(Question Answering)은 인간이 자연어로 묻는 질문에 자동으로 답을 하는 것으로써, 다양한 유형의 질문에 응답을 할 수 있는 것이 중요하다. 자동 문서요약(Automatic Text Summarization)은 긴 문서를 단락 혹은 문장 수준으로 요약을 해주는데 핵심 내용을 축약하는 기술

agenerative-model-for-raw -audio/, 2024.

이 중요하다. 개체명 인식(Named Entity Recognition)은 문장에서 사람, 이름, 동물, 장소, 지명, 시간, 단위 등 개체명을 인식하는 작업이다. 텍스트 분류(Text Classification)는 텍스트로 주어진 데이터에서 스팸 메일을 분류하거나, 뉴스 기사를 정치, 경제, 사회, 문화 등의 기사로 분류하는 것, 혹은 텍스트의 감정을 분석하는 작업에 해당한다. 스펠 체크(Spell Checking)는 텍스트 내에서 오타를 교정하거나 먹는 사과(Apple)과 회사 'Apple'을 구분해주는 것처럼 문맥을 고려한 부정확한 단어 사용의 수정 등을 자동으로 해준다.

2) 자연어의 임베딩(Embedding)과 언어모델(Language Model)

앞에서 설명한 바와 같이 사람이 말한 자연어를 처리하기 위해서는 자연어를 컴퓨터가 이해하도록 표현해주어야 한다. 컴퓨터는 인간이 사용하는 언어 그 자체를 이해하는 게 아니어서, 자연어를 숫자로 표현하여 컴퓨터가 처리할 수 있게 표현을 바꾸어주어야 한다. 즉, 표현력이 무한한 인간의 언어를 숫자가 나열된 벡터 형태로 표현하여야 하는데 이를 임베딩(Embedding)이라고 한다.⁹⁾ 임베딩을 잘하여야만 성능이 좋은 자연어 처리를 할 수 있다. 또한, 자연어 처리 모델의 학습 속도도 빨라진다.

임베딩은 단어 수준 임베딩과 문장 수준 임베딩으로 나누어진다. 단어 수준 임베딩이란 각 단어를 하나의 벡터로 표현하는 것이다. 사람의 몸무게나 키 아니면 자동차의 속도 같은 하나의 수치로 주어진 데이터를 스칼라(Scalar)라고 하며, 어느 학생의 국어, 영어, 수학 점수를 모은 것처럼 숫자 여러 개를 모아서 표현한 것이 벡터이다. 이때 나열된 숫자의 개수를 벡터의 차원이라고 한다. 예를 들어, 단어를 3차원 벡터로 표현한다면 각 단어가 3차원 공간상의 하나의 점으로 나타내어지는 것이다. 이렇게 되면

9) Jihyeon Roh et al., "Unsupervised multi-sense language models for natural language processing tasks," *Neural Networks*, vol. 142, 2021, pp.397-409.

‘hi’와 ‘hello’는 그 의미가 비슷하므로 3차원 벡터 공간 상에서도 가까운 위치에 존재하게 된다. ‘dog’과 ‘cat’도 마찬가지이다. 이렇게 각 단어를 하나의 벡터로 표현하는 단어 수준 임베딩은 임베딩을 학습하는 과정에서 학습대상으로 보여 지지 않은 단어를 표현하지 못한다. 또한, 문장의 문맥 (Context)이 단어 수준 임베딩에는 사용되지 않았다. 즉, 먹는 사과를 뜻하는 ‘Apple’과 회사명 ‘Apple’이 구분되지 않는다.

문장 수준 임베딩은 문맥을 활용하여 임베딩을 한다. 이렇게 하면 단어가 가진 의미를 잘 파악할 뿐 아니라 단어가 어떤 문맥에서 나왔는지도 반영하여 임베딩이 만들어진다. 그렇지만 문장이 길어지면 가까이 있는 단어의 의미만을 잘 반영하게 된다. 이 단점을 해결하고자 트랜스포머 (Transformers)가 제안되었다.¹⁰⁾ 트랜스포머는 긴 문장에서도 각 단어의 연관성을 학습한다. GPT(Generative Pre-Trained Transformer)는 아주 많은 문장들로 미리 학습을 한 트랜스포머로써, 문장에서 이전의 단어 열이 입력되면 그 다음에 올 수 있는 단어를 문맥 정보를 활용하여 확률적으로 예측하고 생성해주는 언어모델이다. 이렇게 대량의 문서를 학습한 언어모델을 대규모 언어모델(Large Language Model, LLM)이라고 한다.

3) GPT(Generative Pre-Trained Transformer)와 Chat GPT

OpenAI에서 2018년 6월에 GPT-1을 1.17억 개의 파라미터를 지닌 모델로 발표하였다. 여기서 트랜스포머가 지닌 파라미터 수가 인공지능 모델의 성능에 직결된다고 볼 수 있다. 2019년 2월에는 파라미터 수가 12.8 배 증가하여 15억 개의 파라미터를 지닌 GPT-2가 발표되었다. 2020년 6 월에 116.7배의 파라미터 증가를 이뤄내어 1750억 개의 파라미터를 지닌 GPT-3가 570GB의 텍스트를 대상으로 학습되어 발표되었다.

GPT-3는 본질적으로 입력된 단어 열을 기반으로 다음에 오게 될 단

10) Ashish Vaswani et al., “Attention is all you need”, 2017.

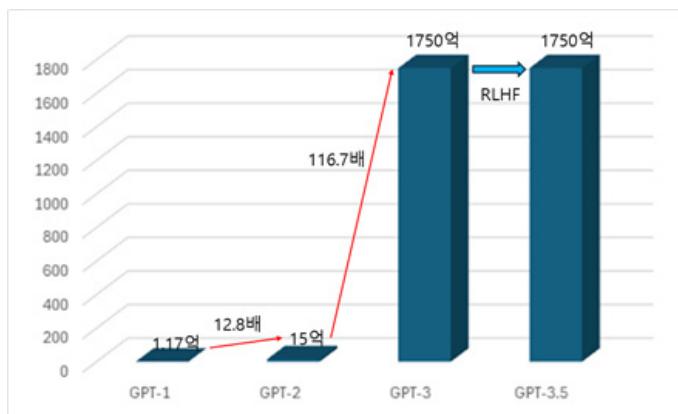
어를 예측하여 생성해주는 생성모델이다. 따라서, 채팅 시 사용자의 명령(Instruction)을 이해하고 그에 적합한 답변을 생성하는 데는 적합하지 않다. 특히, 채팅 시에는 입력된 단어들을 기반으로 다음 단어를 예측하는 게 중요한 것이 아니라 입력된 명령을 이해하고 이에 맞는 응답을 생성하여야 한다. 이를 위하여 개발된 것이 InstructGPT이다. AI가 채팅을 잘 하도록 하려면, 채팅 창에 입력된 명령(Instruction)을 학습한 AI가 생성한 글이 입력된 명령에 적합한지를 점수를 부여하고 그 점수를 기반으로 학습을 하여야 한다. 그렇지만, 생성된 답변에 손실함수(Loss Function)을 통하여 점수를 매기는 것은 정의되기 어려우므로, 사람이 직접 피드백(Feedback)을 주도록 하고, 이를 손실함수로 활용하여 강화학습을 수행도록 하였다. 이를 인간 피드백을 통한 강화학습(Reinforcement Learning from Human Feedback, RLHF)이라고 하며, 40명의 레이블러가 투입되어 학습된 것이 InstructGPT인데, 이는 GPT-3보다 모델은 작지만 더 좋은 성능을 보였다. 마찬가지로 GPT-3에 RLHF를 적용하여 GPT-3.5가 개발되었다. Chat GPT 가 GPT-3.5를 활용하여 2022년 11월30일에 발표되어, 사용자로 부터 입력 받은 문장을 이해하고 사람과 대화하는 것처럼 일상적인 언어를 사용한 의사소통을 할 수 있게 되었다. [그림 1]은 GPT-1에서 GPT-3.5로 발전된 관계를 도식적으로 보여준다.¹¹⁾

OpenAI에서 발표한 Chat GPT는 다양한 활용 가능성을 보여주어 100만명의 사용자를 확보하는 데 단 5일이 소요되었다. 이는 3.5년이 걸린 Netflix, 2.5년이 소요된 Airbnb, 2년의 Twitter, 10개월의 Facebook, 7개

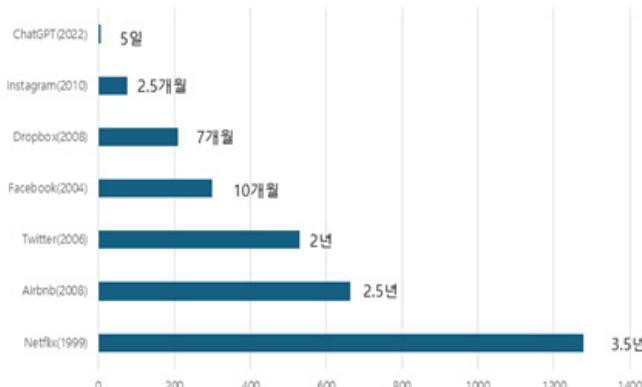
11) 김태원, 「Chat GPT는 혁신의 도구가 될 수 있을까?: Chat GPT 활용 사례 및 전망」, 『THE AI REPORT』, 한국지능정보사회진흥원, 2023.

12) Katharina Buchholz, "Threads Shoots Past One Million User Mark at Lighting Speed", <Statista: AI World Wide Statistics & Facts>, 2023.07.07.

월의 Dropbox, 5개월의 Spotify, 2.5개월의 Instagram과 비교하면 과학 혁명적인 변화라 볼 수 있다.¹²⁾



[그림 1] GPT의 매개변수 비교



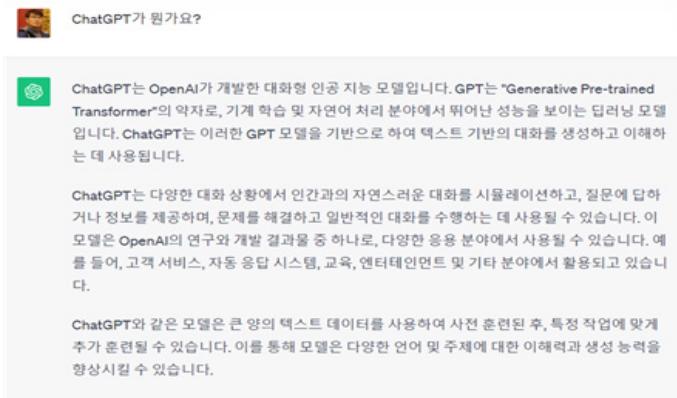
[그림 2] 100만 명 사용자 도달 시간 비교

3. Chat GPT의 활용과 한계

1) Chat GPT의 활용법

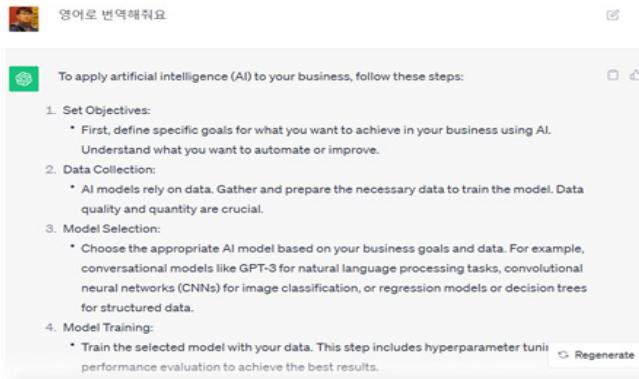
Chat GPT의 사용은 질의 프롬프트에 원하는 사항을 입력하여 응답을

받는 방식으로 하면 된다. 활용의 방법으로 우선 질의응답을 알아보자. 질의응답은 알고자 하는 사항을 프롬프트로 입력하면 된다. [그림 3]은 질의응답의 한 예시로써, Chat GPT를 설명해달라고 요청하면 설명을 해주는 것을 보여준다. 다소 어려운 주제도 답을 해주지만 막대한 양의 텍스트를 학습한 내용을 기반으로 답을 생성하기에 텍스트로 제시된 내용의 중요도를 기반으로 답을 하는 것이 아니어서 아주 전문적인 수준의 답을 요청하기에는 무리가 있다.



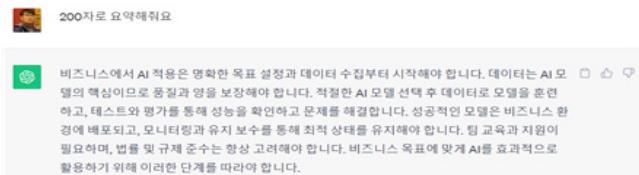
[그림 3] Chat GPT의 활용법: 질의응답 예시

그 다음으로 한글로 작성한 문장의 번역을 요청할 수 있다. [그림 4]는 Chat GPT에 “AI를 업무에 적용하려면 어떻게 해야 하나요?”라고 질의한 다음에 그 응답을 영어로 번역하기를 요청한 내용으로 얻은 답변의 일부분이다. 이는 영어 문서의 작성에 아주 유용한 활용법이다. 작성한 영어 문서의 교정을 요청하여도 무방하다.



[그림 4] Chat GPT의 활용법: 번역

또한, 답변을 얻은 내용을 원하는 양으로 요약을 요청하여도 [그림 5]와 같이 요약문을 얻을 수 있다. 문서의 요약문을 작성할 때, 요약문의 양이 제한되어 있는 경우, 일단 내용을 작성한 후 원하는 분량으로 요약을 요청하면 아주 쉽게 원하는 양에 맞는 요약문을 얻을 수 있다. 이외에도 문장을 제시하고 다시 재작성을 요청하거나 입력된 문서의 감정분석을 요청하여도 된다. 표로 제시된 문서를 문장으로 변경하도록 요청하여도 되고, 이의 역과정을 요청하여도 응답을 얻을 수 있다. 문장을 제시하고 개체명을 인식하는 요청도 수행한다.



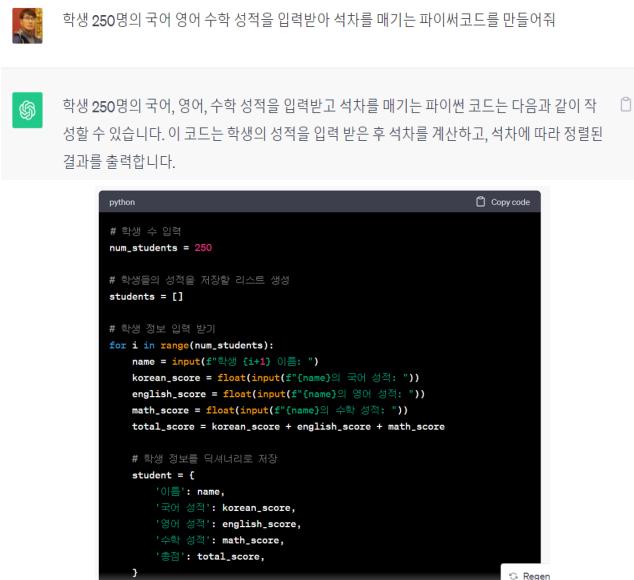
[그림 5] Chat GPT의 활용법: 요약

Chat GPT가 코딩을 하는 능력은 프로그래머의 영역까지도 AI가 수행할 수 있음을 보여주어 충격적이다. 이를 자세히 알아보기 위하여 프롬프트에 “Chat GPT 관련된 네이버 뉴스를 크롤링하는 파이썬 코드를 짜줘”라

고 입력하였다. Chat GPT 발표 초기에는 이런 요청에 코드를 제공하여 주었으나, 그 이후 지적 재산권 관련 논란이 일어났으며 이제는 코드를 제공하지 않는다는 답변을 제시한다. [그림 6]은 이를 보여준다. 지적 재산권과 상관없는 코드를 요청한 경우에는 답변으로 코드를 제공함을 [그림 7]이 보여준다.



[그림 6] Chat GPT의 활용법: 코딩1



```

# 출점을 기준으로 학생 정렬
students.sort(key=lambda x: x['총점'], reverse=True)

# 석차 매기기
for i, student in enumerate(students):
    student['석차'] = i + 1

# 결과 출력
print("\n학생 석차표:")
for student in students:
    print(f"이름: {student['이름']}, 석차: {student['석차']}, 국어: {student['국어']}")

```

이 코드를 실행하면 학생들의 성적을 입력하고, 종점에 따라 석차를 매긴 결과를 출력할 수 있습니다. 입력할 학생 수 및 성적은 필요에 따라 조절할 수 있습니다.

[그림 7] Chat GPT의 활용법: 코딩2

이상과 같이 Chat GPT를 활용함에 있어, 역할을 설정하여 답변을 요청할 수 있다. 특정 분야 전문가의 입장에서 답변을 요청할 수도 있고, 또는 아주 쉬운 설명을 얻기 위하여 유치원생에게 설명하기 위한 답변을 요청할 수도 있다. 또, 원하는 답변이 나오지 않을 경우 자료내용을 일부 전달하여 주어진 자료를 활용하여 답변을 하도록 요구할 수 있다. 질문 역시 어떤 결과물을 원하는지를 명확히 제시하는 닫힌 질문 형태로 제시하는 것이 원하는 답변을 얻기에 더 좋은 방법이다. “~입니다.”, “~임” 등과 같이 답변의 형식도 지정할 수 있다.

2) Chat GPT가 가져온 충격과 악용사례

Chat GPT가 엄청나게 빠른 속도로 사용자를 확보하여 나간 만큼 사회에 미친 충격 역시 엄청났다. 제일 먼저 불거진 이슈는 아무래도 산업계의 민감한 데이터가 OpenAI로 넘어간다는 보안이슈였다.¹³⁾ 심지어 삼성전자는 데이터 유출을 방지하고자 자체적인 LLM 개발에 나섰다.¹⁴⁾ 이탈리아는 국가적 차원에서 국민의 데이터가 학습을 위하여 무단으로 수집되고 있다면서 Chat GPT 접속을 차단하고 나섰다.¹⁵⁾ 이는 정치적 이유로 차단조치를

13) 정두용, “[단독] 우려가 현실로...삼성전자, 챗 GPT 빗장 풀자마자 ‘오남용’ 속출”, <이코노미스트>, 2023.03.30.

14) 임경업, “[단독] 삼성, 챗 GPT 대항마 개발 나섰다.”, <조선일보>, 2023.06.09.

15) 김성민, “이탈리아는 챗 GPT 금지...‘데이터 무단 수집 안돼’”, <조선일보>,

한 북한, 중국, 이란, 러시아, 아프리카 일부국가를 제외하면 서방 국가 중 첫 조치이었다.

과학계는 Chat GPT를 활용하여 논문을 집필한 경우 저자에 포함하여 야 하느냐는 이슈가 불거졌다. Chat GPT가 교신저자로 참여한 논문이 발표되기도 하였으며,¹⁶⁾ 처음에는 Chat GPT를 공동저자로 포함시켰다가 나중에 저자명에서 제외시킨 경우도 발생하였다.¹⁷⁾ 이러한 논란은 Nature 와 Science에서 논문에 대한 책임을 거론하면서 LLM(Large Language Models)을 연구논문의 저자로 인정하지 않겠다고 발표하고 나서야 정리되었다.¹⁸⁾

Chat GPT가 예술계에 미친 영향은 일자리와 관련된 형태로 아주 심각하게 일어났다. 작곡가는 종말이 온 것 같다고 탄식을 하였고,¹⁹⁾ 영화계는 Chat GPT가 감독으로 데뷔하고 각본과 연출 작업까지 수행하고 심지어는 영화촬영 목록, 카메라 위치, 배우 위치, 조명, 등장인물의 의상, 표정, 작은 소품까지도 생성해줌으로 일자리 위협을 크게 느끼게 되었다. 이러한 여파로 할리우드에서는 영화업계 종사자들의 대대적인 파업도 일어났었다.²⁰⁾

무엇보다도 일반인들에게 가장 크게 다가온 Chat GPT의 충격은 사람이 작성한 문서인지 컴퓨터가 Chat GPT로 생성해낸 문서인지 구분할 수

2023.04.03.

- 16) Siobhan O'connor·Chat GPT, "Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse?", *Nurse Education in Practice*, vol. 66, 2023, pp.103-573.
- 17) Tiffany H Kung et. al, "Performance of Chat GPT on USMLE: Potential for AI-assisted medical education using large language models", vol. 2, no. 2.
- 18) 김동원, "네이처·사이언스 'AI 논문 저자로 인정 안해'", <THE AI>, 2023.1.29.
- 19) James Vincent, "Nick Cave says imitation Chat GPT song is 'a grotesque mockery of what it is to be human'", <The Verge>, 2023.01.17.
- 20) Dani Anguiano·Lois Beckett, "How Hollywood writers triumphed over AI-and why it matters", <The Gurdian>, 2023.10.01.

없는 단계에 이르렀다는 것이다. OpenAI에서 AI가 쓴 글인지 사람이 쓴 글인지 구분하는 AI 텍스트 분류기를 발표하였는데 AI가 쓴 글의 26% 만을 찾아내었고 사람이 쓴 글의 9%는 AI가 쓴 글로 잘못 분류하는 수준으로, 문서 작성자가 사람인지 AI인지 제대로 구분을 할 수 없다.²¹⁾ 이러한 여파로 2023년에 미국 출판사가 뽑은 올해의 단어는 “진짜(Authentic)”이다.²²⁾ 이 제까지 검색 시장을 이끌어 왔던 Google 역시 Chat GPT의 큰 도전을 받게 되었고, 마침내 BARD를 발표하여 생성형 AI 시장에 대응하고 나섰다.²³⁾ 물론 우리나라 기업들도 미국 빅테크의 식민지를 우려하여 초거대 AI 개발에 나서고 있다.²⁴⁾

Chat GPT가 사용자의 요청에 따라 다양한 작업을 해줌으로 나타나는 충격의 또 다른 양상은 사이버 범죄에 악용하는 경우이다. 기술적인 전문 지식이나 코딩 능력이 없더라도 피싱메일, 악성코드, 랜섬웨어 등을 만들 수 있게 되었다.²⁵⁾ 이는 누구든지 마음만 먹으면 전문적인 지식이 없더라도 사이버 범죄를 일으킬 수 있는 능력을 지니게 되었음을 보여준다.

3) Chat GPT의 한계

Chat GPT는 이미 앞에서 설명한 바와 같이 확률적으로 문장을 생성한다. 이러한 단점 때문에 질의 내용에 일부러 함정을 파두고 질의를 하면 질의가 사실인 것처럼 착각하여 응답을 생성하기 때문에 사실이 아닌 내용도 사실인 양 생성하여 답변으로 제시한다. Chat GPT 발표 초기에 이러한 약

21) <오픈 AI>, <https://openai.com/blog/new-ai-classifier-for-indicatingai-written-text>, 2024.02.15.

22) 최진석, “美 출판사가 뽑은 올해의 단어 ‘진짜’”, <한국경제>, 2023.11.28.

23) 김태종, “구글 바드 ‘제미나이’로 명칭 변경...최상위 모델 올트라 출시”, <연합뉴스>, 2024.02.08.

24) 김아람, “‘초거대 AI’ 국내기업 경쟁 격화…LG·네이버 등 신모델 공개”, <연합뉴스>, 2023.07.19.

25) Michael Hill, “‘Chat GPT, 곧 사이비 공격 대량생산에 쓰일 것’ 영국 IT 의사결정자 경고”, <IT World>, 2023.02.06.

점을 파고들면서 오답을 요구하여 Chat GPT의 능력을 비판한 사례들이 많이 보도 되었었다.²⁶⁾ 이를 Chat GPT의 환각(Hallucination) 현상이라고 한다. 확률적으로 답변을 생성하기에 같은 질문에 대한 답변도 달리 하는 경우도 [그림 8]과 같이 나타난다. 따라서, 사용자들은 Chat GPT가 거짓 정보를 제공할 수도 있기 때문에 답변을 완전히 신뢰할 수 없다는 전제하에 반드시 답변을 확인해보아야 한다. 2023년 3월 14일에 발표된 GPT-4를 선택한 Chat GPT가 환각 현상을 줄였다고 볼 수도 있지만 완전히 환각현상이 제거된 것은 아니어서 사용자의 환각현상에 대한 주의는 여전히 필요하다.²⁷⁾

Chat GPT는 질의에 대한 답변을 잘 생성하기 위하여 RLHF을 통한 강화 학습 시 진실에 대한 소스가 없으며 아무 대답이나 하지 않도록 안전하게 학습을 시킨 것이 오히려 질문에 대한 답을 회피하는 방식으로 작동하기도 한다. 또한, RLHF를 통한 사람의 안내에 따르다 보니 GPT 모델 자체의 추론에 제한이 가해지기도 한다.

기본적으로 질의에 대한 응답을 확률적으로 생성해주는 Chat GPT의 특성 때문에 원하는 답을 얻기 위한 프롬프트가 상당히 까다로우며, 심지어 동일한 의미지만 단어의 구성에 따라 원하는 답이 생성되기도 하고 안 되기도 한다. RLHF과정에 투입된 레이블러가 긴 대답을 선호한 특성으로 때론 Chat GPT가 장황한 답변을 늘어놓기도 한다. OpenAI는 정책적으로 폭력성, 성차별, 인종차별, 혐오 답변을 못하도록 하였지만 이러한 기능이 제대로 동작하지 않는 경우도 있음을 사용에 주의하여야 한다.

26) 김도형, “‘세종대왕 맥북 던짐사건 알려줘’ 물었더니...Chat GPT의 엉뚱 답변 ‘밈’으로 유행 중”, <한국일보>, 2023.02.23.

27) Antonio Di Nicola, “GPT-4 vs GPT-3.5: 깊이있는 비교”, <Kanries>, 2023.07.31.



세종과학예술영재학교는 어디 있나요?



세종과학예술영재학교는 대한민국 경기도 수원시 팔달구 장다리로 231(매탄동)에 위치한 고등학교입니다. 이 학교는 경기도 수원시 팔달구에 위치하고 있습니다.



세종과학예술영재학교는 어디 있나요?



제가 가지고 있는 정보에 따르면, 세종과학예술영재학교는 대한민국 경기도 성남시 분당구 구미로 237에 위치한 학교입니다. 위치 정보는 변경될 수 있으므로 정확한 주소 및 위치를 확인하려면 학교 웹사이트나 관련된 공식 소식을 참고하는 것이 좋습니다.

[그림 8] Chat GPT의 한계: 잘못된 정보 제공

근본적으로 Chat GPT의 언어모델로 채택된 GPT는 인간 언어를 확률적으로 흉내를 낸 것임을 유의하여야 한다. 사람처럼 단어의 심오한 의미 혹은 뉘앙스를 알지 못하고 확률적으로 단어를 생성하기 때문에 사람에게 답변을 얻는 것과 같은 신뢰도를 부여하면 안 된다. GPT에 학습시킨 텍스트가 어느 방향으로 편향되어 있는지에 따라 답변도 편향성을 지님에 유의하여야 한다. 물론, 영어로 학습이 많이 되었고 한국어로 학습시킨 문서의 비중은 작기 때문에 질의도 한국어로 하는 것 보다 영어로 하는 것이 질적으로 우수한 답변을 얻을 수 있다. GPT-3.5는 2021년까지의 데이터를 기반으로 학습하였기 때문에 최신 정보를 활용한 답변을 요구하면 잘못된 답을 얻게 된다. 한마디로 정보의 신뢰성에 대하여 주의하여 사용하여야 한다. 물론 GPT-4를 채택한 유료 Chat GPT는 더 최신의 데이터를 학습하였지만 즉각적인 최신 데이터의 학습이 이루어지는 것은 아님을 주의하여야 한다.

4) Chat GPT 사용 업무 효율화 방안

질의에 대한 응답을 자동으로 생성해주는 Chat GPT는 단순히 사람의 호기심을 충족시키는 단계를 넘어 업무의 영역까지도 도움을 받을 수 있기 때문에 단기간에 사용자들이 더 많이 늘어났다고 볼 수 있다. 여기서는 Chat GPT를 실제 업무에 활용하는 방안에 대하여 설명하고자 한다. 먼저,

각종 보고서를 작성하는 경우 문서 작성은 요청하면 내용을 작성하여 준다. 그런데, 이미 Chat GPT의 환각현상이 있음을 언급한 바와 같이, Chat GPT가 작성한 보고서는 허위사실이 포함되어 있을 수 있으므로 반드시 내용을 확인하고 다듬는 작업이 필요하다. 따라서, 비전문가가 어떤 전문적인 내용의 보고서 작성은 Chat GPT를 통하여 출력한 경우, 내용의 진위 혹은 미흡한 부분을 확인하지 않고 타인에게 제출하면 허위 보고서 작성으로 곤란을 겪을 수 있다. 전문가들은 Chat GPT로 작성된 문서는 초안이라고 생각하고 진위여부 확인 및 미흡한 내용의 보충을 거쳐 사용하여야 한다.

Chat GPT는 생성형 AI이지 검색기가 아니다. 그래서 법령이나 규정 등의 내용을 요청하면 실제 법령이나 규정을 알려주지 않고 마치 해당 내용인 것처럼 문서를 생성하여 준다. 실제 미국에서 변호사들이 Chat GPT에게 판례를 요청하여 생성된 내용을 재판부에 제출하였다가 가짜 판례 제출로 벌금을 부과 받은 사례도 있다.²⁸⁾ Chat GPT는 많은 문서들을 학습한 내용을 토대로 답변을 확률적으로 생성하기 때문에, 사실관계와 무관하게 많은 아이디어를 도출하여 보는 것은 유용하다. 본인이 생각하지 못하는 부분도 생성하여 주기 때문에 큰 도움을 받을 수 있다. 물론, 도출된 아이디어에 대하여 실제 적용 가능한지는 반드시 확인하여야 한다. 문서의 작성 내용을 토대로 적당량으로 요약을 하는 기능도 논문이나 보고서의 요약 작성에 큰 도움을 줄 수 있다. 특히, 이제까지 영어문서의 교정을 비싼 비용을 지불하여 한 경우는 Chat GPT를 통하여 아주 편리하면서도 전문적인 영어 문서 교정을 할 수 있다.

환각현상을 방지하기 위하여 검색 증강 생성(Retrieval Augmented Generation, RAG)이 있는 생성형 AI를 활용하는 것도 업무에 큰 도움이 된

28) 장재은, “미 변호사들, 재판서 챗 GPT가 쓴 가짜 판례 제시했다 벌금”, <연합뉴스>, 2023.06.23.

다. 검색 증강 생성이란 사전에 학습된 대규모 언어 모델(Large Language Model, LLM)의 기능을 외부 검색 메카니즘과 결합하는 방법으로써, 외부의 방대한 문서에서 정보를 가져와 생성 AI의 성능을 향상 시켜준다.²⁹⁾ 물론 최신 정보도 검색을 통하여 제공해준다.

LLM이 방대한 양의 문서를 학습에 사용하였다고 하지만 아주 심도 깊은 전문적인 문서들을 모두 학습에 사용한 것은 아니기에 일반적인 답변을 제공하기에는 괜찮지만 특정 영역의 전문적인 답변을 기대하기는 힘들다. 이런 연유로 파라미터 수를 줄여 비용을 절감하면서 특정 분야 전문적인 문서들의 미세조정 학습을 통해 정확도를 향상시킨 sLLM(Small Large Language Model)이 개발되어 제공되고 있다.³⁰⁾ 따라서, 전문적인 분야에 대한 생성형 AI의 도움이 필요하면, 기업체 내부에서 sLLM을 별도로 구축하거나 혹은 sLLM 회사의 서비스를 받아보는 것도 유용한 방법이다.

OpenAI는 Chat GPT의 기능을 지속적으로 향상시키고 있다. 특히, 사용자 개개인의 과거 질문을 기억해 답변의 정확도를 높이면서 연산량을 줄이도록 하겠다고 발표하였다.³¹⁾ 이 기능에 따르면 과거 대화를 기억해 반복적인 답변 없이 더 정확하고 새로운 정보를 얻을 수 있고, 사용자가 선호하는 문체나 문서 양식도 기억하였다가 제공받을 수 있다.

4. 맷음말

사람들은 오래 전부터 사람과 자연스럽게 대화하는 인공지능 시스템을 만들어보고자 하였다. OpenAI가 개발하여 공개한 Chat GPT는 이러한 사

29) Patrik Lewis et al., “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”, *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp.9459-9474.

30) 임대준, “슈퍼컴 필요없는 소형 언어모델 'sLLM' 급부상”, <AI 타임즈>, 2023.04.07.

31) 윤민혁, “챗 GPT에 ‘기억력’ 생겼다”, <서울경제>, 2024.02.14.

람들의 희망에 아주 근접한 서비스이어서 폭발적인 사용자 증가와 함께 사회적인 파급력도 엄청났다. 이 논문에서는 Chat GPT에 대한 원리와 활용법에 대하여 전반적으로 살펴보았다.

인공지능을 ANI와 AGI로 구분하여, Chat GPT 가 사람들에게 큰 호응을 받게 된 원인을 살펴보았으며, 자연어 처리를 임베딩과 언어모델로 소개하고, 언어모델로 제안된 GPT와 대화형 언어모델인 Instruct GPT의 차이를 설명하였다. 그리고, GPT-3에 Instruct GPT가 채택한 RLHF를 적용하여 성능이 개선된 GPT-3.5를 기반으로 Chat GPT가 개발되었음을 설명하였다.

Chat GPT의 활용법과 충격 및 악용사례도 살펴보았으며, 확률론적으로 생성된 언어모델인 Chat GPT의 한계를 설명하였다. 마지막으로, Chat GPT를 업무에 활용하기 위한 방안도 제시하였다.

■ 참고문헌

- 김도형, “‘세종대왕 맥북 던짐사건 알려줘’ 물었더니...Chat GPT의 엉뚱 답변 ‘밈’으로 유행 중”, <한국일보>, 2023.02.23.
- 김동원, “네이처·사이언스 ‘AI 논문 저자로 인정 안해’”, <THE AI>, 2023.01.29.
- 김성민, “이탈리아는 챗 GPT 금지...‘데이터 무단 수집 안돼’”, <조선일보>, 2023.04.03.
- 김아람, “‘초거대 AI’ 국내기업 경쟁 격화…LG·네이버 등 신모델 공개”, <연합뉴스>, 2023.07.19.
- 김태원, 「Chat GPT는 혁신의 도구가 될 수 있을까?: Chat GPT 활용 사례 및 전망」, 『THE AI REPORT』, 한국지능정보사회진흥원, 2023.
- 김태종, “구글 바드 ‘제미나이’로 명칭 변경...최상위 모델 울트라 출시”, <연합뉴스>, 2024.02.08.
- 윤민혁, “챗 GPT에 ‘기억력’ 생겼다”, <서울경제>, 2024.02.14.
- 임경업, “[단독] 삼성, 챗 GPT 대항마 개발 나섰다.”, <조선일보>, 2023.06.09.
- 임대준, “슈퍼컴 필요없는 소형 언어모델 'sLLM' 급부상”, <AI 타임즈>, 2023.04.07.
- 장재은, “미 변호사들, 재판서 챗 GPT가 쓴 가짜 판례 제시했다 벌금”, <연합뉴스>, 2023.06.23.
- 정두용, “[단독] 우려가 현실로...삼성전자, 챗 GPT 빗장 풀자마자 ‘오남용’ 속출”, <이코노미스트>, 2023.03.30.
- 최진석, “美 출판사가 뽑은 올해의 단어 ‘진짜’”, <한국경제>, 2023.11.28.
- Aaron van den Oord et al., “WaveNet: A generative model for raw audio”, 2016.

Antonio Di Nicola, “GPT-4 vs GPT-3.5: 깊이있는 비교”, <Kanries>, 2023.07.31.

Ashish Vaswani et al., “Attention is all you need”, 2017.

Dani Anguiano·Lois Beckett, “How Hollywood writers triumphed over AI-and why it matters”, <The Gurdian>, 2023.10.01.

<https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text>, 2024.02.15.

<https://openai.com/Chat GPT>, 2024.02.15.

Hyung Min Park·Sang Hoon Oh·Soo Young Lee, “A bark-scale filter bank approach to independent component analysis for acoustic mixtures”, *Neurocomputing*, vol. 73, 2009, pp.304-314.

Hyung Min Park·Sang Hoon Oh·Soo Young Lee, “Adaptive noise cancelling based on independent component analysis”, *IEE Electronics Letters*, vol. 38, no. 15, 2002, pp.832-833.

James Vincent, “Nick Cave says imitation Chat GPT song is ‘a grotesque mockery of what it is to be human’”, <The Verge>, 2023.01.17.

Jihyeon Roh et al., “Unsupervised multi-sense language models for natural language processing tasks”, *Neural Networks*, vol. 142, 2021, p.397-409.

Katharina Buchh0lz, “Threads Shoots Past One Million User Mark at Lighting Speed”, <Statista: AI World Wide – Statistics & Facts>, July 7, 2023.

M. Benzeghiba et al., “Automatic speech recognition and speech variability: A review”, *Speech Communication*, vol. 49, 2007, pp.763-786.

Michael Hill, “‘Chat GPT, 곧 사어비 공격 대량생산에 쓰일 것’ 영국 IT 의사결정자 경고”, <IT World>, 2023.02.06.

Mostafa M. Amin·Erik Cambria·Bjorn W. Schuller, “Will affective computing emerge from foundation models and general artificial intelligence? A first evaluation of Chat GPT”, *IEEE Intelligent Systems*, vol. 38, 2023, pp.15-23.

Osmo Kuusi·Sirkka Heinonen, “Scenarios from artificial narrow intelligence to artificial general intelligence—Reviewing the results of the International Work/Technology 2050 study”, *World Future Review*, vol. 14, 2022, pp.65-79.

Patrik Lewis et al., “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”, *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp.9459-9474.

Prakash M. Nadkarni·Lucila Ohno-Machado·Wendy W. Chapman, “Natural language processing: an introduction”, *Journal of the American Medical Informatics Association*, vol. 18, 2011, pp.544-551.

Siobhan O’connor·Chat GPT, “Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse?”, *Nurse Education in Practice*, vol. 66, 2023, pp.103-573.

Tiffany H Kung et. al, “Performance of Chat GPT on USMLE: Potential for AI-assisted medical education using large language models”, vol. 2, no. 2, 2023, PLOS Digital Health, <https://doi.org/10.1371/journal.pdig.0000198>.

■ 국문초록

본 논문에서는 대규모 언어모델인 GPT를 활용한 Chat GPT의 원리와 활용방법, 그리고 사용상의 한계 및 업무 효율화 방안을 소개하였다.

이를 위하여 자연어 처리의 작업 영역을 자세히 알아보고, 컴퓨터에서 자연어 처리의 가장 첫 단계로 임베딩이 이루어짐을 설명하였다. 문장 수준의 임베딩을 위하여 제시된 트랜스포머와 GPT의 관계도 설명하였다. 마침내, RLHF를 통하여 GPT-3의 성능을 개선한 GPT-3.5를 기반으로 Chat GPT가 개발되어, OpenAI에 의해 발표되었음도 알아보았다.

Chat GPT는 질의응답, 번역, 교정, 요약, 표 데이터 처리, 코딩 등의 영역에 활용할 수 있으며, 역할설정 및 자료 전달을 통하여 원하는 답변을 얻기 위한 유도도 할 수 있다. 한편으로, Chat GPT의 활용 방안이 광대한 만큼 사회에 가져온 충격도 막대함을 알아보았다. 기업체 혹은 국가 차원의 보안이슈가 제기되어 사용금지 조치가 이루어졌으며, 과학계의 저자 인정 여부와 문화예술계는 예술적 가치에 대한 논쟁과 더불어 일자리 이슈가 제기되면서 파업도 일어났다. 생성 AI에 의한 가짜 문제가 제기되었고, 사이버 범죄의 수단으로 악용됨도 알아보았다. Chat GPT가 지닌 환각현상 등의 한계도 알아보았다. 마지막으로 Chat GPT를 업무에 활용하기 위한 방안으로 환각현상을 염두에 둔 내용확인 필요성, 검색기가 아님을 주의하여 RAG 기능이 있는 생성형 AI의 활용하는 것과 전문적인 영역을 위하여 개발된 sLLM 사용 등을 제시하였다.

주제어 ● 트랜스포머, 대규모언어모델, 챗 GPT, 활용방안, 한계, 업무 효율화

■ Abstract

The Principles, Applications, Limitations, and Business Efficiency of Chat GPT

Oh, Sang Hoon
/ Mokwon University

This paper introduces the principles and applications of Chat GPT, which utilizes the large-scale language model GPT. It first discusses the detailed workings of natural language processing, where embedding serves as the initial step in computer-based language processing. The relationship between transformers and GPT is explained for sentence-level embedding. Then this paper explores the development of Chat GPT based on GPT-3.5, which has been improved via reinforcement learning from human feedback or RLHF, and released by OpenAI.

Chat GPT can be utilized in various domains such as question answering, translation, correction, summarization, tabular data processing, and coding. It can guide users to obtain desired responses through role-setting and data delivery. However, its broad application is having significant societal impacts. Issues related to corporate or national security have led to usage restrictions, while debates over authorship recognition in the scientific community and discussions about artistic value in the cultural and arts sector have sparked strikes due to job concerns. The emergence of fake problems produced by generative AI, along with its misuse for cybercrime, is also highlighted. Such limitations of Chat GPT, including hallucination phenomena, are discussed in this work.

Lastly, suggestions are made for utilizing Chat GPT in business that include emphasizing the need for content verification with hallucination in mind, cautioning against relying solely on search engines, utilizing generative AI with retrieval-augmented generation(RAG) functionality, and employing

small large language models (sLLMs) developed for professional domains.

Keyword • Transformer, LLM, Chat GPT, Applications, Limitations, Business Efficiency

■ 논문투고일:2024.02.15. ■ 심사완료일:2024.03.20. ■ 게재확정일:2024.03.20. ■