

Reddit 소셜미디어를 활용한 ChatGPT에 대한 사용자의 감정 및 요구 분석[☆]

Analysis of Users' Sentiments and Needs for ChatGPT through Social Media on Reddit

나 혜 인^{1,2} 이 병 회^{1,2*}
Hye-In Na Byeong-Hee Lee

요 약

ChatGPT는 생성형 인공지능(Generative AI) 기술을 활용한 대표적인 챗봇으로서 과학기술 영역뿐만 아니라 사회, 경제, 산업, 문화 등 다양한 분야에서 유용하게 활용되고 있다. 본 연구는 글로벌 소셜미디어 레드딴(Reddit)을 활용해 ChatGPT에 대한 사용자의 감정과 요구에 대한 탐색적인 분석을 수행한다. 이를 위해, 2022년 12월부터 2023년 8월까지의 댓글 10,796건을 수집하여 키워드 분석, 감정 분석, 니즈마이닝(Needmining) 기반 토픽모델링을 실시하였다. 분석 결과, ChatGPT에 대한 댓글에서 출현 빈도가 가장 높은 단어는 "time"으로 답변의 신속성, 시간 효율성, 생산성 향상을 강조한 것으로 나타났다. 사용자들은 ChatGPT에 대해 신뢰와 기대의 감정과 동시에 사회적 영향에 대한 두려움과 분노의 감정을 표현하였다. 또한, 토픽모델링 분석을 통해 잠재적 니즈(Needs)를 포함한 14개의 주제를 도출하였고, 사용자들이 특히 ChatGPT에 대한 교육적 활용과 사회적 영향에 많은 관심을 보였다. 또한, ChatGPT와 관련된 언어모델, 직업, 정보, 의료, 서비스, 게임, 규제, 에너지, 윤리적 문제 등 다양한 주제들이 논의된 것을 알 수 있었다. 분석 결과를 바탕으로 사용자들의 요구를 반영하여 향후 실행계획의 방향을 제시하였다. 본 연구는 향후 ChatGPT를 이용하여 제품과 서비스를 개선하고, 새로운 서비스 플랫폼 기획 단계에서 유용한 정보를 제공할 것으로 기대된다.

☞ 주제어 : ChatGPT, 레드딴, 니즈마이닝, 감정분석, 토픽모델링

ABSTRACT

ChatGPT, as a representative chatbot leveraging generative artificial intelligence technology, is used valuable not only in scientific and technological domains but also across diverse sectors such as society, economy, industry, and culture. This study conducts an explorative analysis of user sentiments and needs for ChatGPT by examining global social media discourse on Reddit. We collected 10,796 comments on Reddit from December 2022 to August 2023 and then employed keyword analysis, sentiment analysis, and need-mining-based topic modeling to derive insights. The analysis reveals several key findings. The most frequently mentioned term in ChatGPT-related comments is "time," indicative of users' emphasis on prompt responses, time efficiency, and enhanced productivity. Users express sentiments of trust and anticipation in ChatGPT, yet simultaneously articulate concerns and frustrations regarding its societal impact, including fears and anger. In addition, the topic modeling analysis identifies 14 topics, shedding light on potential user needs. Notably, users exhibit a keen interest in the educational applications of ChatGPT and its societal implications. Moreover, our investigation uncovers various user-driven topics related to ChatGPT, encompassing language models, jobs, information retrieval, healthcare applications, services, gaming, regulations, energy, and ethical concerns. In conclusion, this analysis provides insights into user perspectives, emphasizing the significance of understanding and addressing user needs. The identified application directions offer valuable guidance for enhancing existing products and services or planning the development of new service platforms.

☞ keyword : ChatGPT, Reddit, Needmining, Sentiment Analysis, Topic Modeling

1. 서 론

최근 생성형 AI(Generative AI)는 사회·경제적 구조의 전반에 영향을 미칠 '게임 체인저(Game changer)' 기술로

1 NTIS center, Korea Institute of Science and Technology Information, Daejeon, 34141, Korea.

2 Science and Technology Management Policy, University of Science and Technology, Daejeon, 34113, Korea.

* Corresponding author (bhlee@kisti.re.kr)

[Received 8 November 2023, Reviewed 23 November 2023(R2 6 February 2024), Accepted 15 February 2024]

☆ 이 논문은 2024년도 한국과학기술정보연구원(KISTI)의 기본 사업으로 수행된 연구입니다(과제번호: K24L1M3CS).

부상하여 산업의 새로운 핵심동력이 되고 있다. 생성형 AI는 기존 텍스트, 이미지, 동영상 등 데이터에서 학습한 패턴과 구조를 기반으로 새로운 데이터를 생성하는 인공지능 모델이다[1]. 딥러닝 기법을 사용한 인공신경망 중 방대한 양의 매개변수(Parameter)를 보유한 대규모 언어 모델(Large language model; LLM) 및 이미지 생성모델(Image generative model; IGM) 등을 이용하여 사용자의 특정한 요구를 충족시키는 창작물을 만드는 기술을 포함한다[2]. 이러한 생성형 AI 기술이 구현된 대표 서비스는 OpenAI의 ChatGPT와 DALL-E2, Microsoft의 Bing Chat, Google의 Bard, IMAGEN, Stable Diffusion, Midjourney 등이 있다.

생성형 AI의 세계 시장 규모는 2023년 130억 달러에서 연평균 35.6% 성장하여 2030년에 1,093억 7천만 달러에 이를 것으로 예상된다[3]. 생성형 AI 개발은 초기 단계에 있으나 산업 전(全) 분야로 확산되고 있다. 즉, 마케팅 및 영업, 제품·서비스 기획, 서비스 운영계획 등에 관한 문서 작성, 고객 및 동향 분석, 제품 디자인 설계 등 여러 분야에서 이미 활용되고 있다. 글로벌 기업들은 생성형 AI를 생산 원가절감과 새로운 비즈니스 발굴 및 수익 창출을 이끌 중요 기술로 평가하며 투자를 확대하고 있다. 향후 여러 도메인에서 제품 개발 주기를 최적화하고, 새로운 제품 기능을 도입하는 과정에서 생성형 AI 기술 수요가 지속될 것이다[4].

생성형 AI 기술은 2022년 11월말 OpenAI에서 출시한 초거대 AI 대화형 챗봇인 ChatGPT의 대중화와 함께 더욱 주목을 받기 시작하였다. ChatGPT는 GPT-3.5(Generative pre-trained transformer 3-5) 아키텍처를 기반으로 한 대형 언어모델을 적용하여 사용자 질의에 자연스러운 답변을 제공한다. 그리고 책, 기사, 블로그, 온라인 대화 및 리뷰 등 약 1,500억 개 이상 콘텐츠에서 학습된 딥러닝과 강화 학습 알고리즘으로 구성되어 있다[5].

가장 최근에 추가된 GPT-4는 2023년 3월 14일에 공개되었고, 뛰어난 성능과 멀티모달(Multi Modal) 기능으로 주목을 받고 있다. 수조 개의 매개변수가 있을 것으로 추정되며, 텍스트와 시각적인 입력 모두 효율적으로 처리할 수 있어 AI 응용분야 및 생성 콘텐츠를 확장할 수 있다[6]. ChatGPT는 자연어의 패턴과 맥락을 인간 수준으로 이해하고, 동일한 질문을 반복하더라도 사전에 학습된 데이터 셋을 응용하여 새로운 답변을 스스로 만들 수 있다. 이러한 ChatGPT는 인간과 기술이 상호작용 하는 방식에 혁신을 가져왔고, 과학기술 분야를 넘어 교육, 의료, 법률, 금융, 소매 등 다양한 분야에서 유용하게 활용되고 있다. 특히, 정보 검색, 문서 작성 및 요약, 번역, 프로그래밍 등

창의성이 요구되는 콘텐츠 제작 관련 업무와 직종에서 획기적인 변화를 가져올 것으로 전망된다.

반면에 ChatGPT 사용을 둘러싸고 유용성과 편의성에 대한 비판적인 의견도 있다. 즉, 학습 데이터 부족, 설명력 부족, 오작동 가능성, 인권 침해나 차별을 조장할 수 있는 정보 생성의 위험성 등을 우려하며, 생성형 AI의 한계점을 해결하고 안정성을 확보하는 것이 요구된다[7-8].

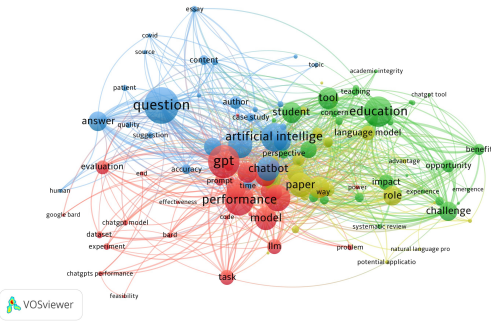
니드마이닝(Needmining)은 제품, 서비스, 기업, 개인, 사건, 애플리케이션을 대상으로 개인의 관점, 태도, 감정 등 관련 속성들을 평가하는 과정이다. 고객의 잠재된 니즈(Needs)를 파악하는 것은 새로운 제품과 서비스 혁신 프로세스에 영향을 주며, 의사결정자의 우선순위 결정에 도움을 줄 수 있다. 따라서 화두로 떠오른 ChatGPT 기술 및 서비스에 대한 답론을 이해하고 지속적으로 대응하기 위해 사용자들의 관심과 이슈를 신속하게 파악하여 신규 플랫폼 기획 및 실행계획 방향성을 모색할 필요가 있다.

ChatGPT 출시 이후, 사용자 수가 급격히 증가하면서 소셜미디어 커뮤니티를 중심으로 서비스에 대한 의견이 공개적으로 표출되고 있다. 이에 본 연구에서는 ChatGPT에 대한 사용자의 경험과 인식을 조사하기 위해 글로벌 소셜미디어 플랫폼인 레딧(Reddit)을 활용하여 사용자의 감정과 요구사항을 분석한다. 레딧 댓글의 텍스트 분석을 통해 긍정 및 부정의 감정 변화를 확인하고, 사용자들이 무엇을 기대하고 우려하는지 파악한다. 또한, 니즈마이닝 기법을 적용한 토픽모델링을 통해 ChatGPT에 대해 논의되는 잠재적 니즈(Needs)를 포함한 토픽들을 도출하고, 주요 이슈를 제시한다. 나아가 사용자들이 무엇을 필요로 하는지 또는 요구하는지에 대한 시사점을 도출하고, 향후 실행계획의 방향성을 제안한다.

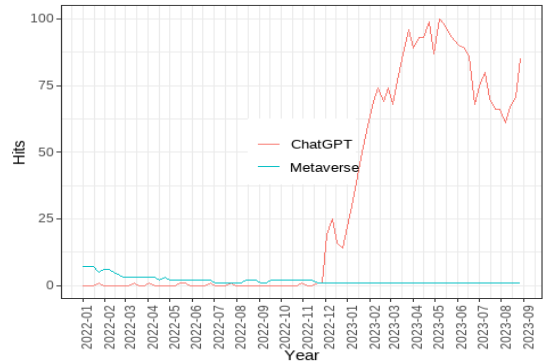
2. 관련 연구

2.1 ChatGPT 연구 및 검색 동향

ChatGPT는 자연어 처리(NLP) 작업, 비지도 사전학습, 미세조정을 통해 사람과 유사한 수준으로 텍스트를 이해하고 생성할 수 있도록 설계된 트랜스포머(Transformer) 기반 신경망 아키텍처를 변형한 대규모 언어모델이다[9]. 특히, 번역, 질의응답, 에세이 작성, 프로그래밍 코딩 등 광범위한 자연어 이해 및 처리 작업에서 우수성이 입증되었다[10]. 적은 양의 특정 데이터 셋을 대상으로 모델을 미세조정 하거나 전이학습을 적용한 연구를 통해 성능을 지속적으로 향상시키고 있다. 이러한 발전은 개인 맞춤형



(그림 1) ChatGPT 연구키워드 네트워크
(Figure 1) ChatGPT research keyword network



(그림 2) ChatGPT와 Metaverse 구글 트렌드
(Figure 2) Google trends of ChatGPT and Metaverse

답변을 생성하여 높은 만족도의 사용자 경험을 제공할 수 있다. ChatGPT는 언어의 이해, 다중모드 처리, 분야별 미세조정으로 이어져 연구개발과 산업영역의 혁신적인 변화를 주도할 것으로 예상된다.

세계적인 학술 논문 및 연구 자료를 검색하는데 사용되는 Google Scholar를 통해 ‘ChatGPT’를 키워드로 포함한 996편의 최근 논문을 수집해 제목과 초록에서 추출한 단어들의 동시출현관계를 시각화 프로그램 VOSviewer로 그림 1과 같이 나타냈다. 이를 통해 해외 ChatGPT에 관한 최신 학술연구 동향을 파악할 수 있다. 키워드 네트워크에서 연결 관계가 높은 단어들은 question, performance, GPT, artificial intelligence, chatbot, tool, paper, education, challenge, impact 등으로 확인된다. 키워드 군집 구성은 AI 챗봇의 특징, ChatGPT 언어모델의 성능 평가, 교육적 응용과 과제, 산업 및 비즈니스 활용에 관한 연구에 집중된 것으로 파악된다.

선행연구를 살펴보면, Ray[1]은 290편의 ChatGPT 관련 논문을 통해 개발 배경과 응용분야를 탐색하고, 데이터 편향성, 윤리 문제, 기술적 한계를 분석해 사회와 산업적 영향을 연구하였다. Guo et al.[11]은 ChatGPT가 생성한 자료를 탐지하기 위해 금융, 의료, 법률, 심리학 데이터셋을 구축해 전문가들의 텍스트와 비교하는 연구를 수행하였다. Tlili et al.[12]는 교육 현장에서 ChatGPT 사용에 대한 트윗 소셜 네트워크 분석, 인터뷰 내용 분석, 사용자 경험을 조사하여 새로운 교육 철학 및 역량강화 방안을 제시하였다. 또한, Haque et al.[13]은 ChatGPT를 파괴적 기술로 정하고 초기 사용자들이 3일 동안 게시한 트윗의 주요 주제와 감성을 분석하여 사회적·윤리적인 영향을 파악하였다. Feng et al.[14]은 레딧과 트위터 소셜미디어 정보를 활용하여 ChatGPT가 스트리밍 미디어에 미치는

영향을 파악해 클라우드소싱 데이터 기반 프레임워크를 제안하였다. 이외에 ChatGPT를 적용한 전자상거래, 금융, 서비스, 마케팅 등 비즈니스 사례 분석을 기반으로 응용 프로그램 탐색[15], 환자교육, 임상연구, 원격의료, 건강 정보 제공 등 의료서비스 영역에 ChatGPT를 적용한 연구 [16], 비즈니스 프로세스를 혁신하기 위한 ChatGPT 미래 로드맵 및 전망을 제시한 연구도 있다[17].

그림 2는 구글 검색에서 ‘ChatGPT’와 ‘Metaverse’ 키워드의 검색량 지표를 0-100 사이 값으로 정규화 하여 시각화한 구글 트렌드(Google trends) 그래프이다. 최근 두 키워드에 대한 검색량을 비교하여 대중의 ChatGPT에 대한 관심도가 높다는 것을 확인할 수 있다. 메타버스는 2020년에 코로나19 팬데믹 시기에 비대면 방식이 선호되면서 주목받았으나 점차 그 인기가 시들해졌다. 이는 메타버스 콘텐츠가 교육, 게임, 패션 등 적용분야가 제한적이며 모든 연령층의 이용자들이 체감할 수 있는 킬러 콘텐츠(Killer content)가 부족하기 때문으로 파악된다. 비용 측면에서 메타버스 콘텐츠를 구현할 기기의 대중화가 아직 어렵다는 점도 요인으로 여겨진다[18].

ChatGPT는 2022년 11월말 출시된 지 5일 만에 100만 명의 사용자를 달성할 정도로 전 세계적으로 폭발적인 인기를 끌며 검색량이 급격히 증가하여 2023년 5월에는 정점에 이르렀다. ChatGPT는 대규모의 사전 훈련된 언어 모델로 인간과 유사한 텍스트 생성과 이전 대화를 기억하여 복잡한 질문에 대한 답변이 가능하여 누구나 쉽게 서비스를 이용해 다양한 영역에 활용 가능하므로 높은 인기를 얻고 있다. 또한, ChatGPT는 텍스트뿐만 아니라 음성, 이미지, 영상, 음악 등 멀티모달 영역까지 확대되고 있다. 향후 ChatGPT와 같은 초거대 및 생성형 AI는 메타

버스 가상공간의 일방적·단방향 정보전달에서 사용자와 상호작용 하는 방법을 개선할 새로운 기술로 주목받고 있다. 나아가 메타버스 플랫폼, 위성통신, 지상 유·무선 통신, 초고속 인터넷망 등을 활용한 공간 서비스의 융합 및 응용연구가 발전할 것으로 예상된다.

2.2 니드마이닝(Needmining)

기업의 제품과 서비스 혁신은 고객의 요구사항을 조사 하는 것에서부터 시작된다. 고객의 니즈(Needs)는 특정 제품이나 서비스에서 보완이 필요한 요구사항을 잠재적으로 표현한다. 니드마이닝(Needmining)은 정보나 데이터를 수집 및 분석하여 고객의 특정 요구사항을 파악하는 과정이며[19], 사용자가 생성한 콘텐츠에서 니즈를 추출하여 내용을 분석하기 시작한 이후 발전하였다.

니드마이닝은 주로 경영학 분야에서 시장 조사, 제품·서비스 아이디어 도출, 마케팅 전략 수립 등의 업무에서 응용된다. 전통적으로 고객의 니즈를 예측하기 위해 설문 조사나 인터뷰 방법을 사용했으나 디지털 기술이 발전과 함께 온라인 텍스트 데이터를 대상으로 고객의 니즈를 파악하기 위한 다양한 방법이 연구되고 있다.

사용자가 생성한 텍스트 데이터에서 특정 제품, 기술, 서비스에 대한 수용 태도, 감정, 의견 등을 분석한 니드마이닝 관련 연구는 다음과 같다. 온라인 리뷰에서 제품의 특징을 요약하고 고객 감성을 분류하는 방법으로 데이터 마이닝과 자연어 처리를 혼합한 방법론 연구[20], 트위터(Twitter) 데이터를 수집해 항공 산업 관련 트윗(Tweet)의 감성분석을 통해서 고객 경험에 영향을 주는 서비스 연구[21], 사물인터넷(IoT)에 대한 트윗의 토픽모델링과 감성을 분류한 연구[22] 등이 있다.

이외에도 시장 구조 분석 마케팅 조사에 고객 리뷰를 바탕으로 제품 속성과 특징 추출 및 분석 방법 연구[23], 머신러닝을 적용하여 트윗에 포함된 고객 니즈 정보를 자동으로 예측한 연구[24]가 있다. 이러한 니드마이닝은 텍스트 정보를 바탕으로 더 많은 고객들의 의견을 분석하여 혁신 아이디어를 찾거나 시장에서 성공 가능성을 예측하는데 효과적이다.

2.3 감성분석과 토픽모델링

텍스트 정보를 분석하는 방법으로 감성분석(Sentiment analysis)은 텍스트에 내재한 특정 주제에 대한 주관적인 의견, 평가, 감정, 느낌 등을 추출하여 긍정과 부정으로

감정을 판별하는 기법이다[25]. 감성분석은 분석 범위에 따라 문서, 문장, 속성 단위로 나뉘며, 분석 방법은 크게 기계학습 기법(Machine learning approach), 어휘기반 기법(Lexicon-based approach), 통합 기법(Hybrid approach) 기법으로 구분된다. 기계학습 기법은 감성극성이 지정된 훈련 자료를 학습하여 감성분류를 실시하고, 어휘기반 기법은 미리 구축된 감성사전(Sentiment lexicon)과 매칭하여 감성 분류와 문서 내용의 통계적 분석에 따른 말뭉치(Corpus) 기반 유형이 있다[26]. 감성분석은 소셜미디어 데이터를 대상으로 특정 제품, 시장, 산업에 관한 인식을 분석하여 이슈와 관심사를 파악한 연구가 활발하다[27-28].

텍스트를 분석하는 다른 기법으로 토픽모델링(Topic modeling)은 비구조화 된 텍스트 정보에 숨겨진 주제를 비지도 학습을 통해 자동 추출하는 자연어 처리 및 분석 방법이다[29]. 대표적인 잠재디리클레할당(Latent Dirichlet Allocation; LDA) 기법은 가장 많이 사용되며, 문서에 포함된 단어의 확률 분포에 따라 문서의 잠재 토픽을 추론한다[30]. 그러나 LDA는 문서의 메타데이터를 활용하지 않고, 문서에 포함된 단어 빈도 기반으로 토픽을 추출해 결과가 정확하지 않을 수 있는 한계가 존재한다. 이에 Roberts[31]은 문서의 단어 빈도와 메타데이터를 활용해 토픽별 단어 분포를 결정하는 LDA 확장 모형인 구조적 토픽모델링(Structural Topic Modeling; STM)을 제안하였다. STM은 토픽 간 상관관계를 추정해 토픽의 관계를 해석할 수 있다[34-35]. STM은 사회과학을 중심으로 교육학, 경영학, 공학, 보건·의료 등 다양한 분야에서 잠재적 주제와 패턴을 찾는 데 활용되고 있다[36-38].

위와 같이 ChatGPT는 다양한 영역에 적용되어 사회 및 산업을 재편하고 있다. 그러나 국내에서는 ChatGPT 기술 도입과 활용에 대한 사용자 또는 대중의 의견을 분석한 연구는 부족한 상황이다. 이에 본 연구에서는 레드 커뮤니티 내 ChatGPT 사용자들의 인식과 주제별 논점 및 니즈를 탐색하고자 한다.

3. 데이터 수집 및 연구방법

3.1 데이터 수집

본 연구에서는 글로벌 온라인 커뮤니티인 레드(Reddit)에서 주제별 서브레딧(Subreddit)의 포스트(Posts)와 댓글(Comments)을 분석대상으로 설정한다. 레딧은 회원들이 관심 있는 특정 주제에 대해 글자 수 제한이 없어 자유롭게 의견을 표현하고, 토론할 수 있으므로 대중들의 인식

을 파악하기에 적합한 데이터로 간주된다. 또한, 레딧은 소비자들이 실제 경험을 바탕으로 작성한 후기를 실시간 수집해 사용 가능하며, 높은 외부 타당성 및 정당성, 이용자 만족 및 관련 속성 등에 대한 다차원 정보를 제공한다 [39]. 이러한 데이터 수집을 위해 R 프로그램의 Reddit API(Application Programming Interface)를 활용하였다.

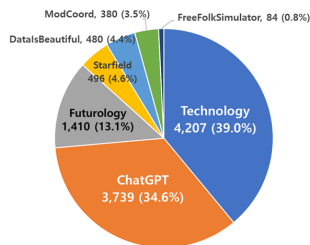
ChatGPT 발표 이후인 2022년 12월 1일부터 자료 수집 시점인 2023년 8월 31까지 ‘ChatGPT’ 단어를 포함한 총 11,194건의 데이터를 수집해 삭제된 댓글을 제외한 총 10,796건을 추출하였다. 분석에 사용한 항목들은 이용자 ID, 게시물의 날짜, 제목, 내용, 댓글의 날짜, 내용, 점수, 서브레딧, 추천, 찬반 여부 등이 있다.

표 1은 ChatGPT에 대한 월별 댓글과 이용자의 현황을 보여준다. 2023년 1월과 5월에 가장 많은 댓글이 작성되었고, 7월에는 댓글 수가 현저하게 줄었다가 최근 다시 증가하는 추세를 보인다. 이는 앞선 구글 트렌드의 검색 관심도 변화와 유사한 것을 확인할 수 있다.

(표 1) 월별 레딧의 댓글 현황
(Table 1) Comments by Month on Reddit

Month	Comment	User	Month	Comment	User
2022-12	559	106	2023-05	2,846	2,316
2023-01	2,350	1,739	2023-06	378	332
2023-02	938	757	2023-07	109	77
2023-03	945	782	2023-08	1,263	1,055
2023-04	1,408	1,095	Total	10,796	8,559

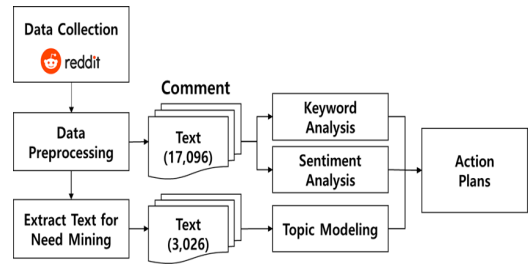
ChatGPT 관련 댓글들을 서브레딧 기준으로 분류하면, 그림 3과 같이 과학기술, ChatGPT 및 AI, 미래 기술 발전 및 예측, 게임, 데이터 시각화 등의 주제들을 포함한다. 커뮤니티 활동이 활발한 분야는 Technology, ChatGPT, Futurology 순으로 댓글의 약 87%가 집중되어 있다.



(그림 3) 서브레딧의 커뮤니티별 댓글 현황
(Figure 3) Comments by Subreddit Communities

3.2 연구방법

본 연구는 레딧에 작성된 댓글을 분석하여 ChatGPT에 대한 대중들의 의견과 니즈를 탐색하여 향후 실행계획의 방향을 제안한다. 본 연구의 전체 연구 절차는 그림 4와 같다. 댓글 텍스트 분석은 데이터 수집, 데이터 전처리, 키워드 분석, 감정분석, 토픽모델링의 5단계로 진행된다.



(그림 4) 연구의 절차
(Figure 4) Research procedures

수집된 댓글을 대상으로 토큰화, 불용어 제거, 어간 및 표제어 추출 등 전처리(Preprocessing) 과정을 수행하였다. 특히, 소셜미디어 텍스트의 특성상 비속어, 줄임말, 이모티콘 등이 다수 포함되어 있어 보다 세심한 댓글의 정제 과정을 거쳐야 한다. 정제된 단어들은 빈도수를 기준으로 키워드 분석, 각 단어의 감성점수(Sentiment Score)를 산출하여 긍정(+)과 부정(-)으로 감성극성을 분류하였다. 즉, 시간 흐름에 따른 대중들의 감성 변화를 확인하기 위해 댓글의 단어별 감성점수를 합산하여 댓글의 총 단어의 수로 나눈 값을 월별로 묶어 산술평균 하여 감성점수를 시계열 그래프로 나타냈다. 또한, 감성어휘사전(Sentiment lexicon dictionary)을 바탕으로 세부 감정들을 기쁨(joy), 슬픔(sadness), 놀람(surprise), 신뢰(trust), 기대(anticipation), 분노(anger), 혐오(disgust), 두려움(fear)으로 분류하였다.

레딧 댓글에 나타난 ChatGPT 사용자들의 관심 주제와 잠재적 니즈 파악을 위해 니즈마이닝 기법을 적용하였다. 즉, 기대, 필요, 요구 등을 의미하는 ‘need’, ‘want’, ‘wish’, ‘expect’, ‘anticipate’, ‘require’, ‘demand’, ‘desire’ 등 니즈와 관련된 어간(word stem)을 포함한 3,026개 댓글을 추출하였다. 이후 추출된 댓글들은 구조적 토픽모델링을 통해 니즈 토픽을 도출하고, ChatGPT에 대한 기대, 논란, 요구 등의 특징을 파악하여 시사점을 제시하였다.

4. 데이터 분석 결과

4.1 키워드 빈도 분석

ChatGPT에 대한 키워드 분석을 위해 R의 형태소 분석 패키지를 통해 표제어가 추출된 단어는 총 9,558개이다. 표 2는 빈도에 따른 상위 30개 키워드를 나타낸다.

출현 빈도가 가장 높은 단어는 ‘time’이며, ‘real’, ‘job’, ‘human’, ‘write’ 순으로 확인된다. 이는 ChatGPT에 대해 즉각적인 답변, 작업 시간 단축, 직업의 변화, 구직 활동, 글쓰기 등 실제 활용 사례와 사회적 영향에 관한 주제를 중심으로 질문, 대답, 토론이 활발히 진행된 것으로 파악된다. ChatGPT 언어모델의 일반적인 특징은 나타내는 ‘learn’, ‘answer’, ‘create’, ‘understand’, ‘generate’, ‘question’, ‘search’, ‘model’, ‘prompt’ 단어가 상위에 포함된다. 이는 대화형 질의응답, 자연어 학습 및 이해, 언어 생성 모델의 특징, 프롬프트 유형, 창작물 제작 및 저작권 문제 등 신규 출시된 응용프로그램에 대한 작동원리와 사용법을 공유하는 것으로 이해할 수 있다. 또한, ‘paper’, ‘text’, ‘essay’, ‘game’, ‘code’ 등 자료 형태에 관한 단어들을 고려했을 때, ChatGPT가 창의적 글쓰기, 게임 콘텐츠 제작, 프로그래밍 코드 생성에 유용한 도구로서 커뮤니티 내 높은 관심을 받으며, 이슈가 된 것으로 볼 수 있다.

댓글에서 자주 사용된 ‘student’, ‘teach’, ‘tool’, ‘test’, ‘class’, ‘school’, ‘information’ 등을 통해 ChatGPT가 교육 분야에서 다양한 논의가 진행되는 것을 알 수 있다. 학교

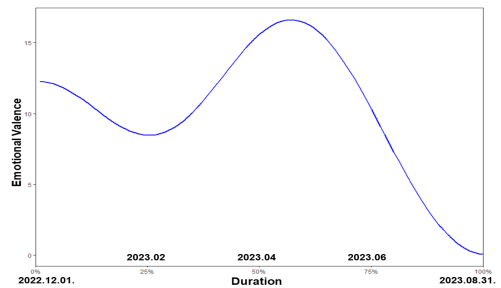
(표 2) 댓글의 단어 빈도 (상위 30)
(Table 2) Frequency of Words in Comments (Top 30)

No.	Word	Freq.	No.	Word	Freq.
1	time	1,147	16	question	509
2	real	1,068	17	tool	506
3	job	834	18	paper	478
4	human	810	19	text	476
5	write	792	20	view	476
6	learn	786	21	search	475
7	student	769	22	test	472
8	actual	731	23	essay	468
9	teach	648	24	game	464
10	IoT	582	25	read	463
11	answer	564	26	school	451
12	create	546	27	model	446
13	understand	540	28	information	404
14	generate	516	29	code	404
15	source	513	30	prompt	400

현장에서 교수학습 활동을 보조하며 풍부한 학습 경험을 제공하는 측면과 가짜 정보, 부정행위, 표절 등의 문제점을 토론의 주제로 다루고 있다. 또한, ChatGPT의 사용 지침 마련을 위한 의견 제안과 표절 탐지 및 평가 방법에 대한 대안 등과 밀접한 연관이 있는 것으로 파악할 수 있다.

4.2 감성분석

감성어휘사전을 기반으로 레딧 커뮤니티의 ChatGPT 관련 게시글에 대한 2022.12.01~2023.8.31. 기간 동안의 시계열 감성분석 결과는 그림 5와 같다.



(그림 5) 댓글의 감성 변화 추이
(Figure 5) Sentiment Change Trends of Comments

2022년 11월 말 출시된 ChatGPT에 대한 레딧 사용자들의 감성점수는 2022년 12월에 높은 긍정지수로 시작했으나 2023년 2월 초까지 다소 하락하였다. 전 세계 ChatGPT 사용자 수가 약 18억 명으로 정점에 이르렀던 2023년 5월에는 긍정지수가 가장 높았고, 최근에는 다시 하락세를 보인다. 이는 ChatGPT를 활용하는 것이 업무의 효율성뿐만 아니라 정보의 신뢰성, 데이터 보안, 윤리적 문제 등의 우려들이 제기되면서 긍정 및 부정적 주제가 함께 논의된 것으로 이해할 수 있다.

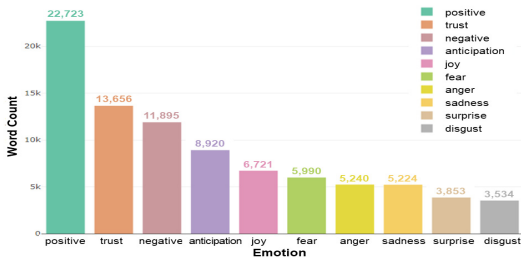
긍정적 댓글에 자주 등장한 단어들은 ‘understanding’, ‘excellent’, ‘productivity’, ‘fast’, ‘intelligence’, ‘quality’, ‘wonder’ 등이며, ChatGPT 대규모언어모델의 기능적인 특성과 우수한 생산성을 대표한 것으로 파악된다. 반면, 부정적 댓글에 등장한 단어들은 ‘cheating’, ‘plagiarism’, ‘error’, ‘biases’, ‘racist’, ‘conservative’, ‘lies’, ‘stupid’으로 확인된다. ChatGPT가 생산한 정보로 인해 발생하는 여러 문제들을 함축하고 있으며, 다수 논의된 것으로 보인다.

다음 그림 6은 ChatGPT 댓글에서 추출된 단어와 NRC 감성어휘사전의 단어들을 비교하여 세부 감정을 10가지 범주의 단어로 분류한 결과이다. 긍정과 부정의 양극의

감정을 제외하면, 새로운 서비스에 대한 관심을 촉발하고 수용하는 과정에서 경험하는 두드러진 감정은 신뢰(trust), 기대(anticipation), 기쁨(joy)임을 알 수 있다.

ChatGPT에 대한 의견은 기능과 사용 방법에 초점을 맞춰 긍정적인 감정을 중심으로 전개되는 경우가 많다. 이러한 의견은 시험 문제 풀이, 질병 진단, 창의적 활동과 같은 작업에서 속도와 정확성을 자주 강조하고 있다. 사용자가 복잡한 명령어 없이 간단한 질의를 통해 필요한 답변을 얻을 수 있다는 장점이 있기 때문으로 볼 수 있다.

ChatGPT 사용이 증가함에 따라 동시에 두려움(fear)과 분노(anger)의 감정이 공존한다. 데이터 침해, 잠재적인 실직, 학문적 부정행위, 표절, 윤리적 문제 등이 포함되며, 부정적 감정이 복잡하게 혼재하여 나타난다. 이러한 개인 정보 및 저작권 보호와 관련된 문제, 특정 직업군의 대체 가능성, 정치적 관점에 미치는 영향, 교육에서의 활용 및 영향 등에 대한 내용을 다루고 있다.



(그림 6) 댓글 단어의 감정 분포

(Figure 6) Sentiment Distribution of Words in Comments

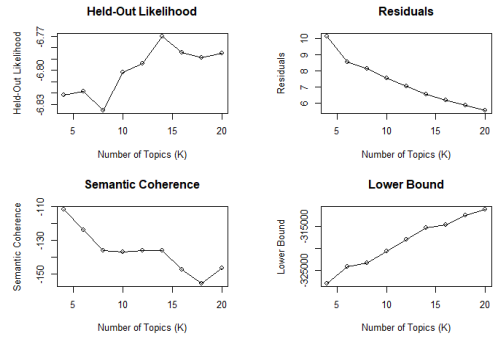
4.3 토픽모델링 분석

4.3.1 최적의 토픽 수(K) 도출

니드마이닝을 통해 추출한 댓글 데이터 3,026건의 주제를 파악하기 위해 STM 분석을 실시하였다. 먼저, 토픽모델링은 비지도 학습을 통해 문서의 잠재 토픽을 추출하여 군집을 정하기 때문에 훈련 및 데이터 셋이 존재하지 않아서 모형의 성능 평가가 어렵다. 따라서 적절한 토픽 수(K)를 정해하는 것이 중요하다[40].

본 연구에서는 최적의 토픽 수(K)를 찾기 위해 Roberts et al.[41]의 STM 패키지에서 제공한 최대우도 추정치(Hold-out likelihood), 의미적 일관성(Semantic Coherence), 하한(Lower Bound), 잔차(Residuals) 등 4개 지표를 검토하였다. 그림 7은 토픽 수를 2부터 20까지

변화시켰을 때 산출되는 4개의 지표 값의 결과를 보여 준다. 이에 따라 의미적 일관성이 최소화 되지 않으며, 최대우도 추정치와 잔차가 극소에 가까운 지점을 기준으로 최종적으로 토픽 수를 14개로 정하였다.



(그림 7) 최적의 토픽 수(K)

(Figure 7) Optimal Number of Topics K

4.3.2 댓글의 니즈 토픽 분석

ChatGPT에 대한 사용자 경험과 소셜미디어에 공개된 대중의 관심과 요구사항을 조사하기 위해 댓글의 토픽을 분석하였다. 표 3은 니드마이닝 접근 방법을 STM 분석에 적용하여 추출한 총 14개 토픽의 결과를 보여준다.

토픽들은 레드 사용자가 ChatGPT에 관한 요구사항을 표현한 댓글 3,026건을 분석대상으로 도출하였다. 토픽에 부여된 명칭은 각 토픽에서 기여도가 높은 단어(Highest Prob)에 중점을 두고 연구자가 공동으로 논의하여 정하였다.

전체 댓글 수가 가장 많은 게시물은 ChatGPT를 통해 생성된 법률 문서에 조각된 판례가 포함된 사례, 생성형 AI가 직업 환경과 노동 시장에 미치는 영향, 생성형 AI 표절 탐지 기술 개발 계획에 관한 내용으로 확인되었다. 일반 대중의 관심과 참여도가 가장 높았던 니즈 토픽은 글쓰기 도구(T1)이며, 전체 댓글의 14%를 차지하고 있다. 다음으로 직무 능력 개발(T2), 프로그래밍 학습(T3), 정보 편향(T4) 순으로 댓글 비율이 높게 나타난다.

ChatGPT에 대한 레드 댓글에 나타난 니즈 토픽들의 세부 내용을 살펴보면 다음과 같다. 토픽 1은 학생들의 다양한 글쓰기 과정에서 ChatGPT가 작문 도구로써 유용하게 활용되고 있는 내용을 강조한다. 즉, 논문과 에세이 작성 과정에서 새로운 아이디어를 탐색하고, 특정 주제에 대한 정보 요약 및 생성, 글쓰기 문법 및 맞춤법의 오류 검사, 번역, 편집 등 ChatGPT의 도움을 받았던 경험들을

(표 3) STM 기반 토픽 추출 및 키워드
(Table 3) STM based Topic Extraction and Keywords

Topic Name	Major Keyword	Prop.
T1 Writing Tool	write, student, know, ask, question, teacher, tool, essay, solve, answer	14%
T2 Job Skills Development	people, job, require, skill, make, level, come, content, keep, critical	10%
T3 Programming Learning	learn, make, give, way, good, work, code, generate, prompt, wrong	9%
T4 Information Bias	read, information, people, potential, source, bad, good, provide, bias, search	8%
T5 Language Model	model, great, take, datum, world, case, technology, fact, language, new	7%
T6 Digital Healthcare	time, pay, live, doctor, better, improvement, huge, sound, app, health	7%
T7 Customer Service	human, work, point, world, client, response, machine, lack, feel, talk	7%
T8 Information Productivity	find, tell, mean, look, specific, comment, time, grade, claim	6%
T9 Educational Technology	use, school, create, network, software, papers, day, internet, access, training	6%
T10 Game Entertainment	bot, game, review, know, story, feel, character, space, video, play	6%
T11 Legal Regulations	make, require, social, government, law, exist, money, medical, right, lawyer	6%
T12 Education System	class, system, education, result, test, pass, university, cheat, create, expert	5%
T13 Energy Industry	energy, industry, change, power, work, content, drive, nuclear, produce, demand	5%
T14 Ethical Concerns	people, know, save, take, claim, information, ethic, prompt, real, work	4%

공유하고 있다. 반면, 학생들이 ChatGPT에 지나치게 의존하는 경우에 자기주도적인 창의적 글쓰기 능력 향상을 저해할 수 있음을 지적하고, ChatGPT와 교육과정을 연계하는 효과적인 교수학습법의 필요성이 요구된다.

토픽 2는 직무능력개발 과정에서의 ChatGPT 역할과 직업적 환경 변화에 미치는 영향이다. ChatGPT는 반복적이고 규칙적인 콘텐츠 생성을 자동화하고, 업무의 효율성 향상에 유용하게 활용된다. ChatGPT 사용자들은 시·소설·시나리오 작업, 문서 작성 및 편집, 데이터 분석 결과 보고서 작성, 고객 지원 업무, 소프트웨어 개발 등의 실제 업무 활용 사례와 방법들을 공유하고 있다. 그러나 일부 대중들은 작가, 기자, 사무원, 회계사, 고객 상담사, 프로그래머 등의 직업의 역할을 대체 및 일자리 감소에 대한 사회적 변화에 대한 대응방안 마련을 요구한다.

토픽 3에서는 ChatGPT가 프로그래밍 학습을 보조하는 혁신 도구로서 중요성이 강조된다. ChatGPT는 학습자와 대화 형식의 상호작용을 통해 프로그래밍 지식과 기술을 효과적으로 제공하고, 코딩 능력 향상을 지원하는 강력한 도구로 활용된다. 프로그래밍 언어, 알고리즘, 코딩 패턴 등을 이해하도록 개념과 원리를 설명하고, 코딩 오류를 찾아 해결방법의 피드백과 코드 예제를 생성하여 제공해 준다. ChatGPT 사용자들은 코딩 시간을 단축하고, 잘못된 코딩에 대한 정확한 피드백을 받고자 도움을 요청하는 최적의 코딩 프롬프트 작성법을 공유하고자 한다.

토픽 4는 ChatGPT가 제공하는 정보의 편향성에 대한

문제 인식과 개선방안이다. ChatGPT는 사용자의 질문 수준을 파악하여 맞춤형 정보를 제공하지만 정보의 출처 및 편향성, 학습 데이터의 품질 및 신뢰성 등에 대해 직접 판단하기 어렵다는 한계를 지적한다. 이에 사용자들은 ChatGPT가 생산한 정보의 신뢰성을 어떻게 평가할 것인지에 대한 기준을 포함한 가이드라인 마련의 필요성을 제시한다.

토픽 5는 ChatGPT의 자연어 처리 기능과 언어모델의 특징이다. ChatGPT는 대규모 데이터 셋의 학습을 통해 자연어 이해 능력이 향상되어 인간과 유사한 수준으로 자연스러운 대화가 가능해져 기존 챗봇과 차별화된 기술 혁신의 측면에서 높은 관심을 보인다. 대중들은 학습한 데이터의 규모가 커질수록 성능을 향상시키고, 특정 도메인의 전문지식을 학습하여 새로운 기능 추가 또는 새로운 서비스 플랫폼 개발에 응용되기를 기대한다.

토픽 6은 디지털 헬스케어(Digital Healthcare) 분야에서 ChatGPT를 어떻게 활용할 것인지를 다룬다. ChatGPT는 대량의 의학적 지식과 임상 보고서를 기반으로 미국 의사 면허시험(USMLE)의 통과 기준치인 60점을 넘을 정도로 높은 성능을 보였다[42]. 대중들은 환자들의 증상에 대한 실시간 의료상담과 빠른 진단, 신약 및 치료법 연구 지원, 모바일 애플리케이션을 이용한 건강관리 등 의료서비스 영역에 대한 접근성을 높이고, 시간과 비용절감에 도움이 될 것으로 기대한다. 그러나 ChatGPT의 의료정보에 대한 정확성과 신뢰성 검증, 개인의료정보 유출, 의료 행위에

대한 책임 소재 불분명 등이 우려되므로 실제 의료 과정에서는 의사의 진단과 처방이 필요하다는 의견이다.

토픽 7은 인간과 기계의 상호작용을 통한 대화형 AI 모델인 ChatGPT가 고객센터 지원 업무에서 응용되는 내용을 포함한다. ChatGPT는 복잡한 용어가 아닌 일상 언어를 사용하여 고객과 소통 가능하며, 고객의 질문에 신속한 응답 및 실시간 문제해결 지원, 고객별로 맞춤형 마케팅 콘텐츠 제작, 고객 의견을 분석하여 제품·서비스 개선 반영 등에 활용되고 있다. 그러나 고객의 감정 이해 및 인식의 한계, 고객별 요구에 대한 완벽한 대응 어려움, 고객 정보에 관한 데이터 관리 문제, 응답의 일관성 유지, 기존 서비스와의 통합 등 지속적인 개선이 필요하다.

토픽 8은 방대한 양의 다양한 주제와 정보를 빠르게 검색하고 요약하여 제공하는 ChatGPT의 정보 생산성에 관한 주제를 다룬다. 사용자가 특정 정보나 문서에 대한 질문에 대해 논문, 뉴스, 웹 페이지 등 여러 유형의 자료에서 구체적인 정보를 탐색하여 맞춤형 정보를 제공하는 부분에 집중한다. ChatGPT는 검색 속도와 정확성면에서 높은 성능을 보이는 사용자 상호작용 모델을 통해 기존 검색 엔진과 지식 공유 플랫폼의 대체품으로 간주된다.

토픽 9는 ChatGPT 기술을 학교 교육시스템과 연계 및 활용하는 ‘에듀테크(Educational Technology)’와 관련이 있다. 이는 교육 현장에 ChatGPT를 효과적으로 적용하기 위해 교육자료 및 교수학습법 개발, 교육 소프트웨어 및 애플리케이션 개발, 전문 학술자료에 대한 접근성 향상, 디지털 리터러시 교육적 활용 등 대중의 기대와 수요를 포함한다.

토픽 10은 게임 엔터테인먼트에서 ChatGPT의 잠재적 활용 가능성에 주목한다. ChatGPT는 게임 유저들의 선호 스타일에 맞는 게임 추천, 다양한 게임 리뷰, 플레이 방법 및 캐릭터의 특징 등의 정보를 제공하고 있다. 게임 기획 및 개발자들은 게임 캐릭터 및 세계관 구축, 스토리보드 작성, 게임 레벨 설계, 스페이스 디자인, 게임 개발방법 및 코드 작성 등의 아이디어부터 프로그래밍 테스트에 이르는 모든 개발 과정에서 ChatGPT를 활용하고 있다. 게임 유저들은 게임 환경에서 캐릭터와 서로 대화하거나 언어가 다른 유저들과 자유롭게 의사소통하기를 원한다.

토픽 11은 ChatGPT 기술과 정보 활용에 따른 사회적 영향을 고려하여 법·제도적 규제의 필요성을 포함한다. 특히, 레드 커뮤니티에서 이슈가 된 ChatGPT가 알려준 거짓 판례를 법정에 인용한 변호사 사례는 ChatGPT 활용의 잠재적 위험성을 보여준다. 일반 대중은 다양한 법률 정보에 대한 접근성이 높아졌지만, 동시에 정보의 진위

여부를 확인해야 하는 책임감을 강조한다. 이는 ChatGPT 정보를 의료, 법률, 비즈니스 등 여러 분야에서 활용하기 위한 규제 방안 모색의 중요성을 시사한다.

토픽 12는 ChatGPT 기술이 교육 및 평가방식에 주는 영향을 논의하고, 학교 교육 시스템과의 연계하는 방안 중점을 둔다. ChatGPT는 학습의 보조도구인 동시에 대학에서 논문 표절, 과제물 작성 대체, 시험 부정행위 등을 우려되는 문제를 지적하고, 해결방안을 논의한다. 최근 대학에서의 ChatGPT 사용금지 및 사이트 차단과 관련된 사례를 대상으로 찬성과 반대로 나뉘어 토론하고, 교사와 학생들을 대상으로 ChatGPT 활용 교육과정, 교수학습법, 평가기준, 표절 검증 시스템 등의 필요성을 요구한다.

토픽 13은 에너지 산업에서 ChatGPT의 역할과 영향에 대한 주제를 다루고 있다. ChatGPT는 에너지 산업 정보와 지식 제공, 산업 환경 변화 분석 및 예측, 에너지 수요 및 공급 분석, 연구 및 기술개발 지원, 시스템 자동화 및 관리 등 분야에 활용 가능하다. 그러나 ChatGPT와 같은 대규모언어모델은 데이터 학습 및 운영 과정에서 많은 전략 공급이 필요하기 때문에 대중들은 에너지 소비 및 온실가스 배출량이 증가되는 상황을 우려하고 있다.

토픽 14는 ChatGPT가 잘못된 지식을 학습하거나 거짓 정보의 생성이 초래하는 사회적 부작용에 대한 이슈를 중점적으로 다룬다. 인종차별, 성차별, 혐오 등의 내용과 정치적 성향이 편향된 답변을 차단하고, 윤리적 중립을 지키기 위한 노력의 중요성을 언급한다. ChatGPT 활용이 일상화됨에 따라 학습 데이터 품질 관리, 사용자 재교육, 윤리적 지침 개발 등의 필요성을 시사한다.

전체 토픽들의 관계성을 파악하기 위해서 토픽 간의 상관 계수를 계산하고, 네트워크 그래프를 통해 연결성을 확인한 결과, 글쓰기 도구(T1), 프로그래밍 학습(T3), 정보 생산성(T8), 에듀테크(T9), 교육 시스템(T12)의 니즈 토픽들이 서로 연결되었다. 이를 통해 레드 댓글에서 ‘교육적 활용 및 영향’에 관련된 유사한 주제가 함께 논의된 것을 알 수 있었다. 서로 연결되지 않은 7개의 토픽들은 ChatGPT와 관련된 언어모델, 직업, 정보, 의료, 서비스, 게임, 규제, 에너지, 윤리 등 중복되지 않는 독립된 주제들을 논의한 것으로 이해할 수 있다.

5. 결과 종합 및 시사점

ChatGPT의 인기가 높아짐에 따라 사용자들 사이에서 ChatGPT 기술을 어떻게 인식되고 있는지 감정과 니즈를

파악하여 반영하여 향후 서비스 기획 및 개발 관점에서 시사점을 제시한다. 표 4는 니즈 토픽과 활용분야를 고려하여 제시한 실행계획에 대한 내용을 보여준다.

(표 4) 니즈 토픽에 따른 ChatGPT 실행계획
(Table 4) ChatGPT Action Plans based on Needs Topics

Topic	Field	Action Plan
T1, T3, T12, T8, T9	교육	· 학년별 교육과정 및 학습관리시스템 연계 · 학생별 학습 로드맵 설계 및 정보 제공 · 메타버스 융합형 교육 플랫폼 개발 · 교육용 데이터 검증 도구 및 활용 지침 개발
T2	직업	· 직무 능력 평가 및 피드백 기능 추가 · 인간 중심의 작업 환경 조성
T4	정보	· 편향된 정보 검출 알고리즘 도입 · 사용자의 정보 필터링 기능 추가
T5	모델	· 프롬프트 작동 원리 및 내부 메커니즘 공유 · 사용자 맞춤형 모델 조정 가이드라인 제공
T6	의료	· 개인 건강정보 기반 대화형 서비스 개발 · 대화형 원격의료 시스템 개발
T7	서비스	· 텍스트 감정 인식 알고리즘 도입 · 개인 프로필 기반 서비스 제공
T10	게임	· 게임 스토리 및 캐릭터 아이디어 찾기 · 메타버스 융합 개인 맞춤형 게임 생성
T11	규제	· 특정 전문분야 법률 전문가 네트워크 · 실시간 법률 서비스 플랫폼 구축 및 테스트
T13	에너지	· 에너지 산업 정책 정보 플랫폼 · 효율적인 연산 알고리즘 개발
T14	윤리	· 윤리위원회 설립 및 가이드라인 수립 · 사용자 피드백 수집 및 모니터링

T1, T3, T12, T8, T9와 같이 ChatGPT는 글쓰기, 코딩, 정보 생산, 에듀테크, 교육시스템을 고려하여 학교 현장에서 학년별 교과교육과정과 학습관리시스템(LMS)을 연계한 학교용 ChatGPT 챗봇 서비스 및 튜토리얼을 개발한다. 또한, ChatGPT를 통해 학생의 학습목표, 학습수준, 관심영역을 반영한 학습 로드맵을 설계하고, 프로젝트 기반 학습 모듈 개발, 학습 정보 추천, 과목별 퀴즈 등 다양한 콘텐츠를 추가한다. ChatGPT와 메타버스를 융합해 수업 시간 이외에 교사와 학생들이 참여하는 플랫폼을 구축하여 온라인 그룹 학습 활동, 학습 정보 공유 및 피드백을 진행하도록 지원한다. 그리고 ChatGPT에서 제공한 학습 정보의 사실 여부를 판단할 수 있는 교육용 검증 도구를 개발하고, ChatGPT 정보 활용 지침, 학생 개인정보 보안 정책 및 기술도 도입한다.

T2에서와 같이 ChatGPT에 질의응답을 통해 사용자의 직무나 관심 분야에 따라 실무에 적용할 수 있는 맞춤형

교육 과정과 콘텐츠를 추천하는 직업 상담과 직무 능력 수준을 평가하여 피드백 제공 기능을 도입한다. 그리고 ChatGPT가 업무 자동화를 통해 직업의 일부를 대체할 수 있으나 인간 중심의 작업 환경을 유지하며 보조도구로 활용하는 가이드라인을 개발하여 적용한다.

T4에서와 같이 ChatGPT에 편향된 정보를 검출 가능한 알고리즘을 적용해 응답 전에 편향된 내용을 감지할 수 있도록 한다. 개인 맞춤형으로 정보 출처, 자료 유형 등을 선택하고, 필요와 선호에 따라 정보를 필터링할 수 있는 기능을 추가해 불필요한 정보를 제외해 사용자가 입력한 주제에 맞는 정보 및 관점을 탐색하여 제공한다.

T5에서와 같이 다양한 산업분야에 적용할 수 있도록 언어모델 사용 사례 라이브러리를 구축해 공유하거나 ChatGPT와 유사한 언어모델을 활용하여 사용자가 특정 분야에 맞게 조정할 수 있는 방법 및 가이드라인을 제공한다. 또한, 언어모델의 프롬프트 작동 원리 및 내부 메커니즘 관련 정보를 제공하여 기술 활용을 확대한다.

T6에서와 같이 ChatGPT는 병원 및 약국과 연계하여 온라인 진료 예약, 실시간 의료 상담, 약물 정보 조회 등 통합 건강정보를 제공한다. 개인의 건강 상태를 모니터링 하는 웨어러블 디바이스의 정보를 ChatGPT를 통해 건강 관리 서비스 플랫폼 개발도 요구된다. 또한, 개인의 건강 정보 보호를 위한 보안 시스템과 규정을 마련하고, 헬스케어 및 디지털 진료 정보를 확인하기 위해 의사를 통해 건강상태에 대한 진단을 받도록 시스템을 개발해야 한다.

T7에서와 같이 ChatGPT는 텍스트 기반 감정 인식 알고리즘을 도입해 사용자의 텍스트에서 표현되는 감정을 분석하고, 적절한 반응과 정확한 응답으로 대화의 품질을 향상시켜야 한다. 고객 프로필과 과거 대화 이력을 관리하는 데이터베이스를 연동하여 개인화된 서비스를 제공하며, 이는 개인정보 관리 데이터 보안 정책이 함께 추진되어야 한다. 그리고 ChatGPT를 통해 해결하기 어려운 문제나 복잡한 상황은 자동으로 상담원과 연결되는 전환 시스템을 구축해야 한다.

T10에서와 같이 게임에서는 ChatGPT를 이용해 가상 NPC(Non-Player Character)를 생성하고, 메타버스 내에서 게임 유저들이 가상 캐릭터가 상호작용할 수 있는 대화 시스템을 도입한다. 그리고 ChatGPT로부터 게임 스토리, 가상공간, 캐릭터 등에 관한 아이디어를 얻어 콘텐츠를 생성한다. 유저가 원하는 게임 플레이 도움말, 리뷰 요약, 스토리와 캐릭터, 맞춤형 게임 및 동영상 추천 등 정보를 제공할 수 있다. 유저는 입력한 내용에 맞춰 기존 게임의 환경에서 새로운 스토리나 이벤트를 추가 생성하여 보다

개인화된 경험을 느낄 수 있을 것이다.

T11에서와 같이 정부 규제, 법률, 금융, 경제, 의료 등 특정 전문분야의 최신 관련법과 제도 데이터를 수집 및 연동하여 제공하는 서비스 플랫폼을 구축한다. 각 분야의 전문가 그룹과 협력하여 오류가 없는 최신 정보를 제공할 수 있게 법률적 검토를 거치고, 지속적으로 테스트를 진행하여 잘못된 정보를 수정하도록 한다.

T13에서와 같이 ChatGPT는 에너지 관련 시장 및 동향 변화를 시각적으로 제공할 수 있고, 기업과 기관이 보유한 에너지 정보를 연계해 에너지 소비 예측, 공급 최적화, 유지보수 계획, 시장분석 등 에너지 산업에 다양한 방식으로 영향을 준다. 이러한 과정에서 ChatGPT는 대규모 연산을 수행하므로 많은 전력이 소비되므로 전력 효율이 높은 서버와 프로세스의 하드웨어를 사용하고, 연산 요구 사항이 적은 알고리즘의 개발이 요구된다. 또한, 재생에너지 사용을 확대해 온실가스 배출 감소와 같은 환경에 미치는 영향을 최소화할 수 있는 방안이 필요하다.

T14에서와 같이 ChatGPT의 윤리적 사용과 관련된 가이드라인 및 정책을 수립하는 위원회를 설립하여 윤리적 질문에 어떻게 응답하는지에 대한 가이드라인을 수립하여 배포하고, 사용자 교육 프로그램을 실시한다. 윤리적 문제에 대한 사용자 의견과 피드백을 지속적으로 모니터링하고 대응할 수 있는 체계도 필요하다.

6. 결론 및 향후 연구과제

본 연구는 ChatGPT에 대한 대중의 인식과 사용자의 경험 및 의견에 담긴 니즈를 파악하여 시사점을 도출하기 위해 수행되었다. 이를 위해 2022년 12월 1일부터 2023년 8월 31일까지의 글로벌 온라인커뮤니티 레딧의 ChatGPT 관련 댓글을 수집하여 주요 키워드 및 감정 분석, 니드마이닝 기반 토픽모델링을 수행하였다. 앞선 연구의 결과를 요약하면 다음과 같다.

첫째, ChatGPT 관련 전체 댓글에서 가장 빈번하게 등장한 키워드는 'time'으로 확인되며, 답변의 신속성, 시간 효율성, 생산성 향상에 관한 댓글이 많은 비중을 차지하였다. 또한, 댓글의 주요 키워드 분석을 통해 ChatGPT 사용 방법과 활용 사례, 직업 및 사회적 영향, 기술의 작동 원리, 자료 유형, 교육적 활용 및 해결과제 등의 대중의 관심사가 반영된 것을 알 수 있었다.

둘째, 대중의 ChatGPT에 대한 감정 변화를 분석한 결과, 출시 초기에는 높은 긍정지수로 시작하여 다소 하락했다가 사용자 수가 약 18억 명으로 가장 많았던

2023년 5월에 정점을 찍고, 이후부터 하락세를 보였다. 감정어휘사전 기반 세부 감정 분석을 통해 ChatGPT의 새로운 기능과 서비스에 대한 신뢰, 기대, 기쁨을 표현하는 긍정감정 단어와 함께 ChatGPT의 사회적 영향에 관한 이슈를 다루면서 두려움과 분노가 섞인 부정적인 감정을 표현한 단어들도 나타났다.

마지막으로 니드마이닝으로 추출한 댓글 3,026건에 대한 토픽모델링 분석 결과, 잠재적 니즈를 포함한 총 14개 토픽이 도출되었다. T1(글쓰기 도구)이 전체 토픽에서 가장 많은 비중을 차지하였고, 프로그래밍, 정보 생산, 에듀테크, 교육시스템 등을 포함한 ChatGPT의 '교육적 활용 및 영향'에 관한 주제들이 논의된 것으로 분석되었다. 이외에도 ChatGPT에 관한 언어모델, 직업, 정보, 의료, 게임, 규제, 에너지, 윤리적 문제 등 다양한 사회적 이슈에 대한 사용자 의견과 요구사항을 포함한 니즈 토픽들이 도출되었다. 이러한 토픽들은 앞으로도 대중들의 관심 주제로 계속 부각될 것으로 예상된다. 따라서 ChatGPT가 혁신적인 도구로서 제대로 작동하기 위해서는 활용 과정에서 발생할 수 있는 부작용을 최소화하려는 지속적인 노력이 필요하다.

본 연구에서는 글로벌 소셜미디어 데이터를 근거로 전 세계 대중 및 사용자들의 ChatGPT에 대한 인식을 종합적으로 분석했으므로 향후에는 국내 사용자들의 의견을 분석 및 비교하는 연구가 수행될 필요가 있다. 특정 목적에 맞게 생성형 AI와 ChatGPT 관련 기술을 조정 또는 접목하기 위해 고객 및 시장 세분화를 통해 사용자의 요구사항을 심층적으로 이해하는 것도 필요하다. 또한, 본 연구는 감정어휘사전 기반 감정 분석과 토픽모델링 분석에서 연구자의 해석과 주관이 개입될 가능성이 있다는 점이 한계 요인이 될 수 있다. 따라서 텍스트의 문맥과 뉘앙스를 파악하는데 향상된 성능을 보이는 BERT 등 최신 언어모델을 이용하여 분석하고, 그 결과를 비교하는 후속 연구도 필요하다.

참고문헌(References)

- [1] Ray, P. P., "ChatGPT: A comprehensive review on background, applications, key ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope," Internet of Things and Cyber-Physical Systems, Vol.3, pp.121-154, 2023.

- [2] J. H. Yang and S. H. Yoon, "Beyond ChatGPT and into the era of generative AI: Case studies of media and content generative AI services and ways to secure competitiveness," *Media Issues & Trends*, Vol.55, No.3, pp.62-70, 2023.
- [3] Grand View Research, "Generative AI Market Size, Share & Trends Analysis Report by Component (Software and Services), by Technology (Generative Adversarial Networks (GANs), Transformers), by End-use, by Region, and Segment Forecasts, 2022-2030," Grand View Research, 2021.
- [4] Michael Chui, "The state of AI in 2023: Generative AI's breakout year," McKinsey & Company, 2023.
- [5] Dowling, M., and Lucey, B., "ChatGPT for (finance) research: The Bananarama conjecture," *Finance Research Letters*, Vol.53, 103662, 2023.
<https://doi.org/10.1016/j.frl.2023.103662>
- [6] Mogavi, R. H., Deng, C., Kim, J. J., Zhou, P., Kwon, Y. D., Metwally, A. H. S., ... & Hui, P., "Exploring user perspectives on chatgpt: Applications, perceptions, and implications for ai-integrated education," *arXiv preprint arXiv:2305.13114*, 2023.
<https://doi.org/10.48550/arXiv.2305.13114>
- [7] Legal News, "Trends in the Enactment of the 'Act on Fostering the Artificial Intelligence (AI) Industry and Creating a Trust Foundation' and What Businesses Should Pay Attention to," 2023.04.24.
<https://www.lawtimes.co.kr/news/187090>.
- [8] Stokel-Walker, C., and Van Noorden, R., "What ChatGPT and generative AI mean for science," *Nature*, Vol.614, pp.214-216, 2023.
<https://doi.org/10.1038/d41586-023-00340-6>
- [9] Shahriar, S., and Hayawi, K., "Let's have a chat! A Conversation with ChatGPT: Technology, Applications, and Limitations," *arXiv preprint arXiv:2302.13817*, 2023.
<https://doi.org/10.47852/bonviewAIA3202939>
- [10] Dahmen, J., Kayaalp, M. E., Ollivier, M., Pareek, A., Hirschmann, M. T., Karlsson, J., and Winkler, P. W., "Artificial intelligence bot ChatGPT in medical research: the potential game changer as a double-edged sword. *Knee Surgery*," *Sports Traumatology, Arthroscopy*, Vol.31, No.4, pp.1187-1189, 2023.
<https://doi.org/10.1007/s00167-023-07355-6>
- [11] Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J., and Wu, Y., "How close is chatgpt to human experts? comparison corpus, evaluation, and detection," *arXiv preprint arXiv:2301.07597*, 2023.
<https://doi.org/10.48550/arXiv.2301.07597>
- [12] Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., and Agyemang, B., "What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education," *Smart Learning Environments*, Vol.10, No.1, 15.
<https://doi.org/10.1186/s40561-023-00237-x>
- [13] Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., and Ahmad, H., "I think this is the most disruptive technology: Exploring Sentiments of ChatGPT Early Adopters using Twitter Data," *arXiv preprint arXiv:2212.05856*, 2022.
<https://doi.org/10.48550/arXiv.2212.05856>
- [14] Feng Y, Poralla P, Dash S, Li K, Desai V, and Qiu M., "The impact of chatgpt on streaming media: a crowdsourced and data-driven analysis using twitter and reddit," In 2023 IEEE 9th International conference on Big Data security on cloud (BigDataSecurity), IEEE international conference on high performance and smart computing, (HPSC) and IEEE international conference on intelligent data and security (IDS). IEEE, pp.222-227, 2023.
- [15] George, A. S., and George, A. H., "A review of ChatGPT AI's impact on several business sectors," *Partners Universal International Innovation Journal*, Vol.1, No.1, pp.9-23, 2023.
<https://doi.org/10.5281/zenodo.7644359>
- [16] Javaid, M., Haleem, A., and Singh, R. P., "ChatGPT for healthcare services: An emerging stage for an innovative perspective," *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, Vol.3, No.1, 100105, 2023.
<https://doi.org/10.1016/j.tbench.2023.100105>
- [17] Gill, S. S., and Kaur, R., "ChatGPT: Vision and challenges," *Internet of Things and Cyber-Physical Systems*, Vol.3, pp.262-271, 2023.

- <https://doi.org/10.1016/j.iotcps.2023.05.004>
- [18] A. H. Lee, "Metaverse Concept and Key Technology Developments for Popularization," Information-Communication Planning and Evaluation Institute, Weekly Tech Trends, No.2044, pp.2-14, 2022.
- [19] Kühn, N., and Satzger, G., "Needmining: Designing digital support to elicit needs from social media," arXiv preprint arXiv:2101.06146, 2021.
<https://doi.org/10.48550/arXiv.2101.06146>
- [20] Hu, M. and Liu, B., "Mining and Summarizing Customer Reviews," Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.168-177, 2004.
<https://doi.org/10.1145/1014052.1014073>
- [21] Misopoulos, F., Mitic, M., Kapoulas, A., and Karapiperis, C., "Uncovering customer service experiences with Twitter: the case of airline industry," Management Decision, Vol.52, No.4, pp.705-723, 2014.
<https://doi.org/10.1108/MD-03-2012-0235>
- [22] Bian, J., Yoshigoe, K., Hicks, A., Yuan, J., He, Z., Xie, M., Guo, Y., Prospero, M., Salloum, R. and Modave, F., "Mining Twitter to assess the public perception of the Internet of Things," PloS one, Vol.11, No.7, e0158450, 2016.
<https://doi.org/10.1371/journal.pone.0158450>
- [23] Lee, T. Y. and Bradlow, E. T., "Automated marketing research using online customer reviews," Journal of Marketing Research, Vol.48, No.5, pp.881-894, 2011.
<https://doi.org/10.1509/jmkr.48.5.881>
- [24] Kuehl, N., Scheurenbrand, J., and Satzger, G., "Needmining: Identifying micro blog data containing customer needs," arXiv preprint arXiv:2003.05917, 2020.
<https://doi.org/10.48550/arXiv.2003.05917>
- [25] Fang, X., and Zhan, J., "Sentiment analysis using product review data," Journal of Big Data, Vol.2, No.5, 2015.
<https://doi.org/10.1186/s40537-015-0015-2>
- [26] Xu, Q. A., Chang, V., and Jayne, C., "A systematic review of social media-based sentiment analysis: Emerging trends and challenges," Decision Analytics Journal, Vo.3, 2022.
<https://doi.org/10.1016/j.dajour.2022.100073>
- [27] W. H. Kim and E. H. Park, "Exploring Game Consumers' Perception in Online Game Community using Text Mining Approach: The Case of Reddit," Journal of Consumption Culture, Vol.24, No.1, pp.73-83, 2021.
<http://dx.doi.org/10.17053/jcc.2021.24.1.004>
- [28] Jeong, B., Yoon, J., and Lee, J. M., "Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis," International Journal of Information Management, Vol.48, pp.280-290, 2019.
<https://doi.org/10.1016/j.ijinfomgt.2017.09.009>
- [29] Hu, N., Zhang, T., Gao, B., and Bose, I., "What do hotel customers complain about? Text analysis using structural topic model," Tourism Management, Vol.72, pp.417-426, 2019.
<https://doi.org/10.1016/j.tourman.2019.01.002>
- [30] Blei, D. M., Ng, A. Y., and Jordan, M. I., "Latent dirichlet allocation," Journal of Machine Learning Research, Vol.3, pp.993-1022, 2003.
- [31] Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder Luis, J., Gadarian, S. K., Albertson, B., and Rand, D. G., "Structural topic models for open ended survey responses," American journal of political science, Vol.58, No.4, pp.1064-1082, 2014.
<https://doi.org/10.1111/ajps.12103>
- [32] H. S. Park, D. H. Kim and S. J. Chang, "Research trend analysis on smart city based on structural topic modeling," Journal of Digital Contents Society, Vol.20, No.9, pp.1839-1846, 2020.
<http://dx.doi.org/10.9728/dcs.2019.20.9.1839>
- [33] Bagozzi, B. E., "The politics of scrutiny in human right monitoring: Evidence from structural topic models of US State Department human rights report," Political Science Research and Methods, Vol.6, No.4, pp.661-677, 2018.
<https://doi.org/10.1017/psrm.2016.44>
- [34] S. B. Jo, and S. H. Ha, "Analysis of Open Government Data Demand Using Structural Topic Modeling," Journal of Information Technology and Architecture, Vol.17, No.2, 2020, pp.103-118, 2020.

- <https://doi.org/10.22865/jita.2020.17.2.103>
- [35] Chen, X., Zou, D., Cheng, G., and Xie, H., “Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of Computers & Education,” *Computers & Education*, Vol.151, 103855, 2020.
<https://doi.org/10.1016/j.compedu.2020.103855>
- [36] Jo, W., Lee, J., Park, J., and Kim, Y., “Online information exchange and anxiety spread in the early stage of the novel coronavirus (COVID-19) outbreak in South Korea: structural topic model and network analysis,” *Journal of medical Internet research*, Vol.22, No.6, e19455, 2020.
<https://doi.org/10.2196/19455>
- [37] Curiskis, S. A., Drake, B., Osborn, T. R., and Kennedy, P. J., “An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit,” *Information Processing & Management*, Vol.57, No.2, 2020.
<https://doi.org/10.1016/j.ipm.2019.04.002>
- [38] Moro, S., “Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent dirichlet allocation,” *Expert Systems with Application*, Vol.42, No.3, pp.1314-1324, 2015.
<https://doi.org/10.1016/j.eswa.2014.09.024>
- [39] Roberts, M. E., Stewart, B. M., and Tingley, D., “stm: An R package for structural topicmodels,” *Journal of Statistical Software*, Vol.91, No.2, pp.1-40, 2019.
<https://doi.org/10.18637/jss.v091.i02>
- [40] Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., ... and Tseng, V., “Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models,” *PLoS digital health*, Vol.2, No.2, e0000198, 2023.
<https://doi.org/10.1371/journal.pdig.0000198>

● 저 자 소 개 ●



나 혜 인(Hye-In Na)

2011년 전남대학교 경제학과(경제학사)
 2013년 전남대학교 기술경영전문대학원 기술경영학과(경영학석사)
 2019년~현재 과학기술연합대학원대학교 과학기술경영정책전공(박사과정)
 관심분야 : 과학기술정책, 인공지능, 빅데이터, 텍스트마이닝, 네트워크 분석, etc.
 E-mail : nanahi@kisti.re.kr



이 병 희(Byeong-Hee Lee)

1992년 충남대학교 컴퓨터공학과(공학사)
 1994년 충남대학교 대학원 컴퓨터공학과(공학석사)
 2002년 충남대학교 대학원 컴퓨터공학과(공학박사)
 2002년~현재 한국과학기술정보연구원 NTIS센터 책임연구원
 2012년~현재 과학기술연합대학원대학교 과학기술경영정책전공 교수
 관심분야 : 과학기술경영정책, 인공지능, 빅데이터, 텍스트마이닝, 위성통신, etc.
 E-mail : bhlee@kisti.re.kr