

ChatGPT 관련 논문의 연구동향과 성능평가로 본 시사점

(Research Trends in ChatGPT-related Articles and
Implications from Performance Evaluation of ChatGPT)

김 상 준 한국생명공학연구원

Sang-Jun Kim(Korea Research Institute of Bioscience and Biotechnology)

[초 록]

ChatGPT 이용 관점에서 성능 정보를 공유할 목적으로, DB에 수록된 ChatGPT 논문을 검토해 리뷰로 작성하였다. 출판동향 분석결과, ChatGPT 출현 초기에는 연구논문보다 속보성의 논문이 많이 출판되었다. ChatGPT 관련 논문으로 보면 교육계는 미래 교육 방향을 정비하고, 저작권과 특허계는 AI 생성물 소송을 주시하고, 출판계는 AI의 저자 기준을 확립하고, 의료계는 의료인 지원 가능성을 탐구하고 있었다. 이처럼 성능 정보를 기반으로 예상되는 문제를 극복 하면서 ChatGPT를 이용한다면, 더 유용한 생성형 AI 시대가 기대될 것이다.

For the purpose of sharing performance information from the perspective of using ChatGPT, the articles in three DBs were investigated and wrote a review article. According to the analysis, more breaking news articles were published than research articles in the early days. The education community was establishing the future direction of education, the copyright and patent community was monitoring AI product litigation, the publishing community was establishing authorship guidelines for generative AI, and the medical community was exploring the possibility of supporting medical professionals. If ChatGPT's performance information is used to overcome these challenges, it will be expected in a more useful era of generative AI.

키워드 : 인공지능; 서지분석; 생성형 챗봇; 성능; 출판; Artificial intelligence; Bibliometrics; ChatGPT; Performance; Publishing



ChatGPT의 등장과 치열한 생성형 AI 경쟁

2022년 12월부터 본격적으로 시작된 생성형 Artificial Intelligence(AI) 챗봇인 OpenAI(<https://openai.com/>)의 ChatGPT는, Chat와 Generative Pre-trained Transformer(GPT)의 합성어로 2023년 1월에 1억 명 이상의 사용자를 조기에 확보하였다. ChatGPT는 시와 에세이 등의 글쓰기는 물론 시험문제, 상식 정보, 음악 작곡, 예술 창작, 게임, 프로그래밍 등 다양한 분야에서 프롬프트(prompt)에 정교한 답변을 생성한다. 그래서 AlphaGo가 이세돌과의 바둑 대결에서 승리한 후 다시 한번 AI 기술의 발전을 누구나 쉽게 체감하도록 만들었다. 이렇게 많은 사람의 관심을 단기간에 집중적으로 받은 ChatGPT는, Google이 개발한 Transformer를 기반으로 주어진 문장의 앞부분을 활용하여 다음 단어를 확률적으로 예측하는 deep learning된 자연어 처리 모델이다. ChatGPT는 Large Language Model(LLM)인 GPT-3.5를 기반으로 reward model을 포함해 인간 AI 훈련사가 관여된 지도학습과 Reinforcement Learning from Human Feedback(RLHF)을 사용해 파인튜닝 되었다(Wikimedia Foundation 2023). 다른 챗봇과 달리 ChatGPT는 학습한 내용은 물론 대화와 질문내 문맥을 기억하여 인간처럼 상세하고 논리적인 글과 답변을 생성하지만, 거짓된 답변을 회피하도록 설계되었음에도 그럴듯해도 정확하지 않거나 황당한 답변을 하기도 한다. 또한 2021년 9월 이후의 사건과 지식에 대해서는 학습이 제한(GPT-4 Turbo는 2023년 4월까지로 확장)되고 알고리즘 편향에 따른 위험하고 비윤리적인 답변을 생성한다는 학습상의 문제도 있어서, 그 이용에 대한 찬성과 반대가 분분하다.

이러한 ChatGPT의 부분적 한계에도 불구하고 IT 업체 사이에는 생성형 AI 기술의 주도권을 잡기 위한 세계적 경쟁이 치열하다. 2023년 2월 Google이 LLM인 LaMDA를 기반으로 한 Bard를 공개하자, Meta는 LLM인 LLaMA를 2월에 출시한 후 공개된 코드와 모델로 누구나 AI를 만들 수 있는 오픈소스 기반의 LLaMA 2를 7월에 공개했다. 중국의 바이두(Baidu)는 2023년 3월 Ernie Bot을 출시했고, 한국의 네이버도 HyperCLOVA X를 운영하기 시작하였고, 애플(Apple)도 Apple GPT를 개발하고 있다고 2023년 7월에 발표했다. 이들 외에도 Amazon Q, Claude, Jasper AI 등이 개발되어 여러 목적으로 AI가 널리 이용되고 있다. 그 중 Stat, Uptrends, Alpha 등은 통계 데이터나 주식시장 분석 등 다양한 분야에서 데이터를 빠르고 정확하게 분석해 해안을 제공하고 있다. 실제로 삼성 갤럭시 S24 핸드폰에 디바이스(device) 기반으로 통역과 번역 등의 생성물에 AI 기술이 적용되면서 인기를 끌고 있다. 한편 Elicit, Scite, Consensus 등에서는 무료 과학 DB나 출판사와 협력해 유료 논문 정보를 요약하거나 연구를 식별하는 AI 기능이 운영되고 있다. 의료 전문팀이 학습시킨 Google의 Med-PaLM 2는, 방대한 의학 정보로부터 수집한 인사이트를 요약하고 관련 질문에 대한 답변을 제공할 수 있다. GPT-3.5를 사용해 2013년 이후 Scopus 논문의 초록

을 학습한 Scopus AI는(Van Noorden 2023a) 자연어 요약과 함께 참고문헌(5~10개의 초록을 참조해야만 답변을 생성) 제공을 기반으로 시각화 기능과 추가 질문의 예시 제공 기능을 2024년부터 판매하고 있다. 또한 Digital Science는 Dimensions용 LLM 어시스턴트의 비공개 시험판을 발표했고, Clarivate는 Web of Science(WoS)에 LLM을 도입하기 위해 논문 초록을 학습한 생성형 AI 개발 후 테스트 중이라고 한다. 이러한 서지 DB 기반의 생성형 AI는, 연구논문의 배경과 선행연구 부분의 초안 작성에 큰 도움이 될 것으로 보인다. 기존 언어 모델을 통합해 복잡한 실험을 AI가 설계, 계획, 수행하는 GPT-4 기반의 Coscientist가 개발되어 금년부터 무인(無人) 연구소를 열 예정으로, 실제로 AI가 문헌을 검색해 분자를 만들기 위한 실험을 설계하고 시약을 결정해 합성한 후 화합물을 검사하는 논문이 *Nature* 저널에 수록되어서 무인(無人) 연구실이 현실로 다가오고 있다(Boiko *et al* 2023).

ChatGPT는 2023년 2월 유료제 프리미엄 ChatGPT Plus에 이어 2023년 3월 GPT-4의 출시와 함께 API 제공은 물론, 내부와 외부에서 ChatGPT plugins이 추가되고 2023년 5월에 전용 iOS app을 출시하면서 그 이용이 더 확산되었다. ChatGPT의 기반이었던 GPT-3.5의 문자(text) 답변 중심에서 다국어와 이미지 해석도 가능한 멀티모달(multimodal) 기반의 GPT-4를 조기 공개하면서 단시간에 놀라운 기술적 진보를 보여주었다. GPT-4의 성능에 대해서 OpenAI의 자료로 보면 GPT-3.5보다 영어는 정확도가 70.1%에서 85.5% 수준으로 향상되고, 한국어도 77.0% 수준으로 다른 언어와 함께 크게 개선되었으며, 민감정보의 정책적 응답이 29% 증가하면서 불허용 응답도 82%나 줄어서 환각(hallucination) 등이 감소되었다고 한다(OpenAI 2023). 그러나 GPT-4 기반의 ChatGPT에서도 OpenAI(OpenAI 2022)가 이미 밝힌 제한점[그렇듯하게 들리지만 부정확하거나 무의미한 답변을 작성하고, 입력 문구 수정에 민감하거나 동일한 프롬프트를 여러 번 시도하고, 지나치게 장황하거나 특정 문구를 남용하고, 모호한 질문에 대해 사용자가 의도한 바를 추측하고, 때로는 유해한 지침에 응답하거나 편향된 행동을 보임]이 GPT-3.5보다 완화되었을지 몰라도 여전히 숙제이다. 또한 검색 엔진의 종말을 우려하는 Google이 PaLM 2에 이어서 2023년 12월 개발한 생성형 멀티모달 AI인 Gemini는, 대규모 다중작업 언어 이해(MMLU)에서 정답률 90.0%로 GPT-4의 86.4%와 인간 전문가의 89.8%보다 우수해 인간에 버금가는 성능이라고 홍보하고 있다. 그러나 Microsoft 연구진은 Medprompt+를 활용한 GPT-4로 MMLU에서 더 높은 90.1%를 기록해 MMLU뿐만 아니라 GSM8K와 같은 다른 테스트에도 GPT-4가 Gemini보다 성적이 더 좋다는 반론이 있었다. 이러한 생성형 AI의 급성장으로 인해 *Science* 저널은 2023년 10대 과학적 성과로 AI 기상 예측 모델을 제시하였고, *Nature* 저널은 2023년 과학계 10대 인물로 과학계와 사회 전반을 변화시킨 비인간 ChatGPT를 선정했지만, AI 이용으로 인해 많은 연구논문이 철회되었다는 부정적인 소식도 있었다(Van Noorden 2023b).



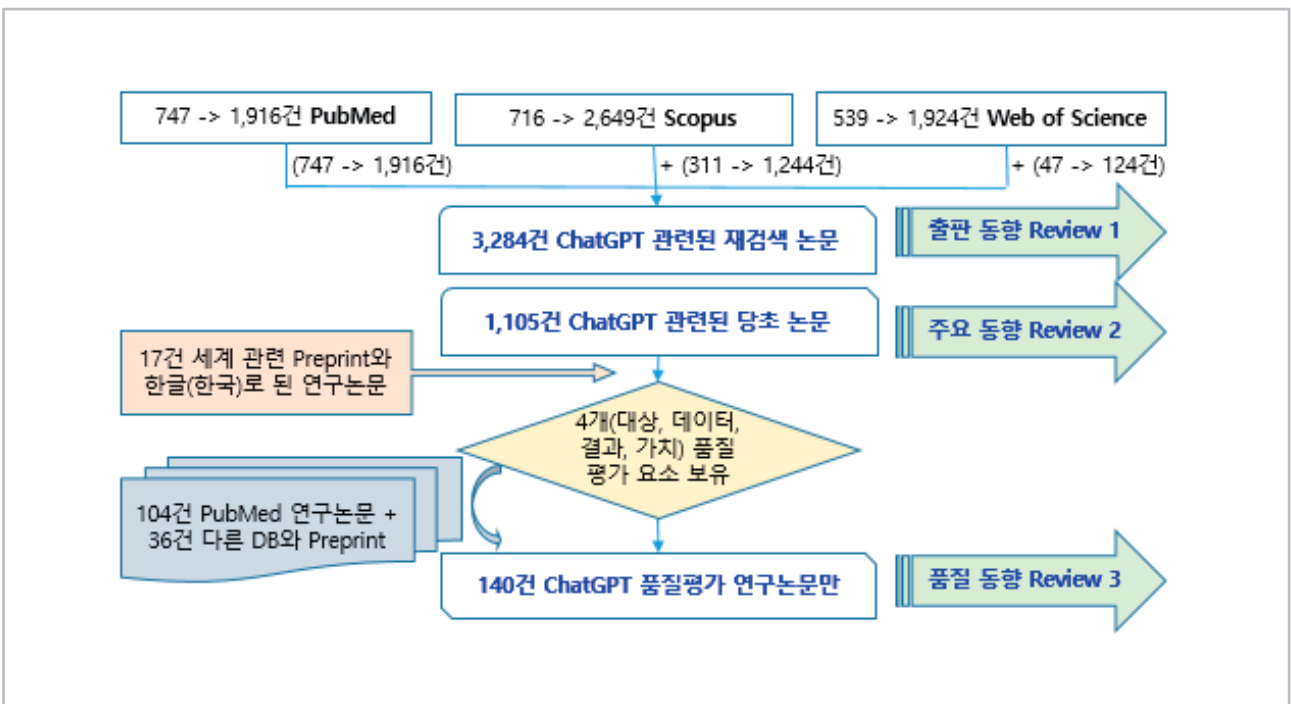
ChatGPT의 학술적 이용 관련 리뷰 방법과 데이터

실제 한국 한의사 시험에 GPT-4가 그냥 응시하면 불합격 수준이지만, 프롬프트 엔지니어링을 활용하면 합격 수준인 것으로 보고되었다(Jang *et al* 2023). 이처럼 GPT-4와 더불어 ChatGPT API와 plugins 등이 발표되면서 ChatGPT의 성능 문제에서 고려 사항이 늘어나고 정확한 품질에 대한 판단도 더 복잡해졌다. 따라서 생성형 AI의 기술적 측면이 아니라 학술적 이용 측면에서 ChatGPT의 사용 여부를 결정하려면, 연구데이터 기반의 연구논문(research article) 위주로 정확한 품질을 검토할 필요가 있다. 이에 따라 ChatGPT 성능에 대해 객관적인 수치로 계량화해서 평가한 연구논문만을 대상으로 그 성능과 품질에 대해 종합적으로 정리하면, 지난 7개월(일부 내용은 12개월) 동안의 폭발적 관심에 따른 성과를 더 확실하게 판단할 수 있다고 보았다. 이를 위해 주요 논문 DB에서 검색된 ChatGPT 관련 논문의 간략한 출판 및 연구 동향과 주요 대응을 먼저 살펴보고, 성능과 품질의 객관적 평가에 부합하는 연구논문만을 대상으로 품질평가 실태를 비교하면 적절할 것이다. 이에 따라 ChatGPT 관련 연구논문의 평가 결과와 가치 등을 4개 항목으로 나누어 비교한 조사 결과를 공유하면, 연구자는 물론 일반인에게도 ChatGPT의 품질 판단에 도움이 될 것으로 판단된다. 따라서 ChatGPT를 언급한 전체 논문의 출판 및 연구 동향과 함께 품질과 성능에 대해 객관적으로 평가한 연구논문에서 시사점을 도출한 연구가 필요하다고 판단되므로, 본 원고는 ChatGPT 사용을 고민하는 사람에게 정확한 판단에 도움이 되도록 품질 판단의 길라잡이가 되기 위한 목적으로 성능을 비교한 결과를 이해관계자와 공유하고자 작성되었다.

연구대상 논문을 수집하기 위해 주요 논문 DB에서 title, abstract, keywords 필드로 제한해 [ChatGPT or “GPT-3.5” or “GPT-4”]를 검색어로 입력하였다. DB 검색에서 기술적 측면을 다루거나 심사 과정이 치밀하지 못한 학술 행사자료(conference proceedings)는 제외하였다. DB에서 index date나 DB 수록일이 2023년 6월 30일까지를 기준으로 하면 PubMed 747건, Scopus 716건, WoS 539건이 7개월간 ChatGPT 관련 논문으로 각각 검색되어 중복을 제외하면 1,105건이 도출되었다. 다시 2023년 11월 30일까지를 기준으로 DB를 검색하면 PubMed 1,916건, Scopus 2,649건, WoS 1,924건이 12개월간 ChatGPT 관련 논문으로 각각 검색되어 3,284건의 비중복 리스트가 만들어졌다. 여기에 영어 외 언어의 ChatGPT 관련 품질평가를 보강하기 위해 한국 논문 등을 추가하려고 Korea Citation Index(<https://www.kci.go.kr/>)와 ScienceON(<https://scienceon.kisti.re.kr/>)을 검색해 보완하였다. 그 후 3개 DB에서 의학용 PubMed 위주로 된 주제 분야의 편중을 완화하기 위해서, Preprint 논문도 인터넷에서 추가로 수집하였다.

검색된 3개 주요 DB의 결과는 DOI, PubMed ID, Article title을 상호 비교하면서 중복을 제외하였다. 그 결과 고유한 논문 1,105건(재검색 후에 3,284건이지만, 7개월간 리스트는 <https://doi.org/10.6087/>

kcse, 321의 연구데이터 참조)의 엑셀 리스트를 만들었다. 이 리스트에는 PubMed 외에 겹치지 않는 Scopus와 WoS의 비의학 분야 논문이 추가되었다. 통합된 12개월 전체 3,284건의 출판동향과 함께, 당초 7개월 동안 1,105건의 논문에 대해 ChatGPT 관련 연구 동향을 간략히 먼저 검토한 후에 성능과 품질 평가를 비교하였다. ChatGPT의 품질 평가 비교에는 수집된 논문의 초록과 본문을 확인해 평가 대상과 목적, 평가에 사용된 데이터, 평가된 주요 결과, 평가 결과의 의미와 가치 등 4개 항목 모두를 요약할 수 있는 140건의 연구논문만(PubMed 기반 104건과 한국 논문 등을 포함해 기타 36건) 대상으로 하였다. 이렇게 조사된 방법과 과정을 그림으로 그려 이해를 쉽게 하면, 다음 <그림 1>과 같다.



<그림 1> 데이터 수집과 분석 과정도



ChatGPT 관련 논문의 출판 동향(3개 DB 12개월 3,284건의 리뷰 1)

DB 수록의 신속성이 우수한 PubMed(김상준 2005)에서 2023년 4월까지의 ChatGPT 관련 논문이 다른 DB보다 압도적으로 많이 검색되었지만, 5월부터는 대표적 논문 서지 DB인 Scopus와 WoS에도 관련 논문이 증가하기 시작했다. ChatGPT가 출현한 지 1년이 지나면서 Scopus의 논문 수가 가장 많아도 생의학 분야에 국한된 PubMed와 모든 주제 분야를 다루는 WoS가 비슷할 만큼 PubMed에 관련 논문이 많이 수록

되었다. ChatGPT는 등장 직후 Editor, Editorial, Letter to editor, Correspondence, Comment, Opinion, News 등의 이름으로 많은 의견 개진이 있었다. ChatGPT 공개 후 지난 1년간 학술계에는 통상적인 연구데이터 기반의 연구논문 보다 초기 도입 단계에서의 찬성과 반대 논의, 사용 후기 성격의 테스트 사례, 그리고 속보성의 짧은 의견·건의·제안을 하는 넓은 의미의 논문이 많이 출판된 것이다. 초기의 이러한 논문에는, OpenAI(OpenAI 2022)가 공식 인정한 제한점을 확인하면서 이용 시 주의 촉구와 함께 다양한 학술적 용도를 엿볼 수 있는 찬성하는 내용도 많았다. PubMed와 달리 DB에 상세 정보가 많은 Scopus와 WoS 데이터를 중심으로, 두 DB에서 수록 정보의 차이도 고려해 확실해 가까운 데이터를 기준으로 ChatGPT 관련 논문의 1년간 출판 동향을 먼저 도출하였다. 1년간 3개 DB의 검색된 전체 논문에서 Scopus와 WoS에서 다르게 분류하거나 불일치하거나 부정확해서 저자가 추가로 확인한 데이터 기반의 간략한 ChatGPT 출판 동향은, 정확성과 완성도가 높지 않아도 대략적인 연구 방향을 엿볼 수 있어서 가치가 있다.

저널이 7개월까지 587종에서 1년이 된 이제 1,491종으로 대폭 늘어나고 다양해져 모두 3,284건의 ChatGPT 관련 논문을 수록하고 있었다. 이 중 다수 논문이 집중적으로 게재된 *Cureus* 저널의 149건 외에 *Ann Biomed Eng*, *Nature*, *Int J Surg*, *JMIR Med Educat*, *Radiology*, *J Med Internet Res*, *Aesthet Surg J*, *Aesthet Plastic Surg*, *IEEE Access*, *Library Hi Tech News* 저널 순으로 논문을 많이 수록하고 있었다. 이렇게 수록된 저널의 357개(출판사 정보가 없는 127건 제외) 출판사 순위는 선행연구(Kim & Park 2023)와 동일한 기준으로 보면, ChatGPT를 특집으로 운영한 *Cureus* 저널을 출판한 Springer(26.0%)의 뒤를 이어서 Elsevier(21.4%), Wiley(6.8%), T&F(6.3%), Kluwer, MDPI, Sage, OUP 순(첫 조사한 7개월 동안에는 Springer, Elsevier, T&F, Wiley, OUP, MDPI, Sage 순)이었다. 1년 동안에 상위 10개 출판사의 논문 비중은 86.4% 정도를 차지해, 선행연구(Kim & Park 2023)보다 대형 출판사에 ChatGPT 논문이 더 집중적으로 출판되었다. 출판 진행 단계가 표시된 2,649건 중 Accepted, Article in press, Advanced print 등 최종 출판되지 않은 것이 493건이고 최종 출판 완료된 것이 2,156건(7개월 동안에는 각각 226건과 490건)으로 구분되어서, 초기에 비중이 높았던 최종 출판 전 논문보다 출판 완료된 논문의 비중이 점차 늘어났다. 논문의 출판된 월을 추정할 수 있는 1,820건 중 2022년 12월 4건에 이어 2023년에는 1월부터 11월까지 각각 27, 41, 107, 141, 134, 201, 258, 229, 291, 255, 105건으로 상반기보다 하반기에 더 많이 출판되어 증가하는 추세였다.

비록 정확한 분류는 아니더라도 논문의 세부 종류가 Scopus와 WoS에서 대략 확인된 2,950건 중 Article 이 51.8%로 가장 많았지만 Letter 18.2%, Editorial 13.6%, Note 8.1%, Review 7.1% 외에도 Comment, News, Opinion 등이 1.2%(7개월 동안에 868건 중 Article 43.9%, Letter 19.5%, Editorial 17.4%, Note

8.6%, Review 5.8%, 기타 2.6%)로 ChatGPT가 시작된 초기에는 속보성 의견 위주의 넓은 범위에서 논문도 많았다. 이처럼 다양한 종류의 논문에 사용된 참고문헌 수는 확인된 1,920건 중 0개가 83건, 1~5개가 350건, 6~10개가 343건, 11~20개가 369건, 21~30개가 232건, 31개 이상이 547건으로 나타났다. 참고문헌이 전혀 없거나 5개 이하로 사용된 경우가 많은 이유는, 연구논문으로 보기 어려운 속보성의 넓은 의미에서의 논문이 많아서인 것으로 추정된다. 수록 저널의 주제 분야는 3,222건 중에서 의학 1,796건, 자연과학 752건, 교육학 219건, 사회과학 380건, 인문학 75건(7개월 전에는 935건 중 의학 668건, 자연과학 156건, 교육학 55건, 사회과학 46건, 인문학 10건)이었다. 7개월 동안 의학 분야가 대세를 이루었지만, 1년이 되면서 의학 논문 비중이 71.4%에서 55.7%로 점차 줄어들었다. 대신 자연과학 논문의 비중이 16.7%에서 23.3%로 자연스럽게 증가하는 등 약간의 변화를 보였다. 특이한 것은 인문학에서 세부 분류한 교육학만의 논문도 많아서, ChatGPT 이용에 대해 의학계는 물론 교육계에서도 관심이 큼을 알 수 있었다. 완전한 Open Access(OA)로 보기 어려운 Bronze OA나 Green OA를 제외하면, 전체 논문 중 Full OA 저널에 수록된 논문 857편과 Hybrid OA 저널에 수록된 논문 272편으로 나타나 OA 논문 비율이 34.4%(7개월 전 263건과 85건으로 31.5%) 정도였다. 이와 같은 결과는 Journal Citation Reports 2021 기준 추정치 38.3%(Kim & Park 2023)보다 낮았다. ChatGPT 관련 논문에 사용된 언어는 2,950건 중 영어가 2,818건이고, 나머지가 모두 132건에 16개 언어여서 대부분 영어로 논문이 작성되었다.



ChatGPT 관련 논문의 주요 동향(3개 DB 7개월 1,105건의 리뷰 2)

검토된 ChatGPT 관련 논문 리스트는 출판된 참조자료(<https://doi.org/10.7910/DVN/LMPTQH>)와 AccessON 학술지 리포지터리 플랫폼(<https://accesson.kr/kslain/v.1/0/1/34376>)에서 보충자료로 제공되어, 3개 관심 분야의 주요한 참고문헌만 서술하면서 인용하였음을 미리 밝힌다.

교육계 등 비의학 분야의 학술적 이용

ChatGPT 도입 이후 가장 많은 관련 논문을 출판한 의학 분야만큼 교육학 분야에서도 논문이 많이 출판되는 한편, 교육 당국의 대응도 다양하게 보고되었다. 초등·중등·고등학교에서는 에세이나 과제물에 생성형 AI의 대표적인 ChatGPT의 지나친 의존을 우려하면서 대응 전략을 포함해 교육 방향을 고민하고 있었다(Kasneci *et al* 2023; Lo 2023). 대학에서도 수업과 글쓰기와 과제물 작성에서 AI 도구의 활용 지침을 제시하기 시작하면서, 표절 예방을 포함한 윤리적 글쓰기에 노력하고 있었다(Lo 2023; Tlili *et al* 2023).

한편 저작권과 특허권 분야에서는 지적재산권의 권리자격을 AI에게 부여할지가 주요 관심사였다. AI 생성물에 대해 AI가 저작권을 부여받으려면, 저작자인 인간의 사상과 감정이 표현되고 최소한의 창작성이 있어야 한다면, 다수 국가에서는 부정적으로 판단하였다(Lee 2023; Zhuk 2023). 이와 별개로 현재 AI 생성물을 위한 학습과정에 사용된 학습데이터의 저작권 위반 여부가 소송 중이고, 작가 등의 창작자들도 ChatGPT 등장 후 업계의 지나친 AI 의존에 반발해 파업과 함께 소송 중이다. 발명과 예술적 착상을 하는 AI인 DABUS에게 특허권을 보유한 발명자 자격을 부여할 수 있느냐의 다국가 소송도, ChatGPT 등장으로 인해 더욱 큰 관심사가 되었다(Ariyaratne *et al* 2023; Picht & Thouvenin 2023; Zhuk 2023). 비록 ChatGPT는 아니어도 AI가 관여된 발명에 대해 다수 국가는 DABUS에게 자격을 거부했지만, 일부 국가는 형식적인 이유로 특허권을 부여했기 때문이다. 저작권법과 특허법 외에 ChatGPT류 생성형 AI 확산에 따른 사회 문제화에 대비해, AI 관련 윤리 기준 수립은 물론 법적인 규제도 뜨거운 관심사이다. AI 관련 법적 규제 중 가장 앞선 EU의 AI 법안에는, 위협의 정도에 따라 AI 모델을 유형화하면서 고위험 AI 관련 사업자는 영업비밀을 제외한 알고리즘 동작 원리 등을 고지하도록 하는 방안을 추진하고 있었다(Helberger & Diakopoulos 2023; Zhuk 2023).

저널 출판계와 저자 성립 기준

학생들의 학습에 ChatGPT 이용이 화두가 된 교육계와 함께 ChatGPT가 출시된 후 신속하게 반응한 분야는 저널 출판계로, AI의 저자 성립과 사용 허용 문제가 쟁점이 되었다(Stokel-Walker 2023; Thorp 2023). 저널 출판계는 ChatGPT 등장 이전부터 AI에 대한 기준을 두기도 했지만, ChatGPT 등장 후에 이용이 확산되면서 이를 신설하거나 정비할 필요성이 더 커졌다. AI나 ChatGPT를 이용한 논문 작성과 저자 성립에 대해 저널 관련 국제 협의체인 International Committee of Medical Journal Editors(ICMJE), World Association of Medical Editors(WAME), American Medical Association(AMA), Committee on Publication Ethics(COPE), Science/Technology/Medicine(STM) 등의 가이드라인을 요약하면, 다음 처럼 공통적인 기준을 정리할 수 있다(위 단체별 URL은 <https://doi.org/10.6087/kcse.321> 참조).

- AI의 저자 표시는 불허 : AI 도구는 연구 결과물에 대한 책임을 지지 못하고 법인격이 없어서 저자가 될 수 없다.
- 인간 저자에게 전적인 책임 부여 : 논문에 AI 도구가 이용된 경우, 인간 저자는 해당 연구 결과물의 진실성에 대한 모든 책임을 져야 한다.
- 사용한 AI의 적절한 공표 : 논문에 AI 도구가 이용된 경우 도구의 명칭, 버전, 제조사 등의 정보를 Methods 또는 Acknowledgment 등 적절한 위치에 기재해야 한다.

Big 3 출판사인 Elsevier, Springer, Wiley 등은 ChatGPT식 AI가 이용되면 저자의 모든 책임하에 이용

사실을 명시해야 하지만 AI가 직접 저자가 될 수 없다는 입장이었다(Lo 2023). 이처럼 다수 출판사와 관련 단체는 대체로 유사하지만, *Science* 저널은 ChatGPT 등의 AI를 사용할 수 없다면서 이 정책을 위반하는 행위를 과학적 부정행위로 엄격하게 보았다(Stokel-Walker 2023). *Science* 저널은 AI가 논문의 저자가 될 수 없음은 물론 편집자의 명시적인 허가 없이는 AI, 머신 러닝, 이와 유사한 알고리즘 도구로 생성된 텍스트를 논문에 사용할 수 없을 뿐만 아니라 이에 첨부된 그림, 이미지, 그래픽도 AI의 산물일 수 없다는 다른 저널보다 더 완고한 정책을 보유하고 있었다. 그러나 AI가 논문의 저자가 될 수 있다는 반대 의견도 있었다(Valentín-Bravo *et al* 2023).



<그림 2> DB 등에서 ChatGPT가 저자로 조사된 사례

그래서 1년 동안 ChatGPT가 저자로 올라간 논문을 주요 DB에서 검색해 보았다. PubMed는 여전히 0건으로 검색되지만, <그림 2>의 *Oncoscience* 저널처럼 저자로 표시된 논문이 있었다. Scopus는 7개월까지 0

건이었지만, 5개월 후 3건이 추가되었다. WoS에서 일부는 단체(Group) 저자로 일부는 공저자로 모두 8건이 검색되었다. ScienceDirect는 여전히 2건이지만, 추후 공저자에서 제외(correction)한 것이 포함되었다. preprint와 OA 위주의 ScienceOn에는 6건으로 변화가 없지만, 7개월 동안과 5개월 후에는 대상 논문이 서로 달라졌다. 이처럼 1년 동안 ChatGPT가 저자로 검색되는 논문은 preprint를 포함해 16건 중 15건(7개월까지의 10건 그림은 <https://doi.org/10.6087/kcse.321>를 참조)이 고유했다. 결국 ChatGPT식 AI 관련 저자 성립에 대한 정책 기준이 아직 명확하지 않을 뿐만 아니라, 동일 DB에서도 검색 시점별로 대상이 다르거나 저자에서 제외되는 수정(correction)도 있어 변화가 많았다.

도서관 분야의 이용과 사서

출판 동향 분석에서 크게 드러나지 않았지만, *Int J Info Manag* 저널에 43개(Dwivedi et al 2023)의 논문과 *IEEE Tran Comput Soc Syst* 저널에도 32개(Wang et al 2023)의 논문이 ChatGPT 관련해 집중적으로 수록되어서 현황 파악에 유익하였다. 도서관과 직접 관련된 분야의 저널인 *Library Hi Tech News* 특집호에도 8개 이상의 논문이 집중 수록되었다. 도서관에서 ChatGPT 활용 예로 검색 및 발견, 참고 및 정보 서비스, 메타데이터 생성, 콘텐츠(보고서, 요약, 서평 등) 생성, 문서 작성의 자동화(Kesselman 2023), 사용자의 작문 과제 지원, 디지털 리터러시 교육(Oyelude 2023) 등이 언급되고 있으므로, ChatGPT 등 AI가 생산한 데이터의 신뢰성이나 품질을 확인하면서 도서관 업무에 이를 활용해야 한다는 의견이었다(Zhang 2023). AI 기술의 핵심에는 데이터, 정보원, 정보 탐색이 있어서 정보 제공자인 도서관은 정보를 저장, 검색, 사용하는 방식을 최신 상태로 유지해야 하므로, 사서도 ChatGPT 기술의 장점과 단점을 모두 이해하고 숙지해 활용하는 것이 필수라는 의견이었다(Frederick 2023). 특히 ChatGPT는 사용자와 소통하고, 개인화된 지원을 제공하고, 관련 정보를 검색하는 기능(Panda & Kaur 2023)이 더욱 정교해질 것이므로 부정확한 질의응답, 오용, 제한된 이해력, 입력 제한, 기술 의존성 등의 위험 때문에 인간 사서의 대체보다는 보완적으로 활용해 업무량을 줄이고 복잡한 업무에 시간을 더 할애해 심층 서비스(Adetayo 2023)를 제공하는 데 도움을 줄 수 있는 잠재력을 가진 대안으로 보았다. 또한 ChatGPT는 학습된 양과 출처 및 언어에 따라 성능에 차이가 다소 있어서, 언어의 편중된 학습과 이용에 따라 자연적으로 발생하는 언어나 지역적 편향 등의 문제가 ChatGPT의 리터러시 문제와 함께 자주 제기되었다.



ChatGPT 관련 연구논문의 품질 동향(3개 DB 등 7개월 140건의 리뷰 3)

ChatGPT의 등장으로 인해 신체 기반의 직업군보다 경력이 적은 프로그래머를 포함해 전문가를 보조하는 지식형 직업군에서 일자리 감소의 위험성을 더 우려스럽게 보는 연구논문이 많았다. 그런데 학술 과정에서 사용될 ChatGPT의 품질과 성능에 대해 조사된 연구논문을 보면, 의료 직군에서 가장 큰 관심을 보여 절반 이상을 차지하였다. 따라서 ChatGPT의 품질 평가 연구논문은 각종 시험의 통과 여부(전문자격과 대학 수학), 의료 업무에서의 지원 가능성(의학과정 시험, 진료와 의사, 환자와 간호), 논문과 글쓰기 지원 등으로 크게 구분할 수 있었다. 아래에는 저자가 분류한 위 3개 분야별 ChatGPT 관련 성능과 품질 평가 연구논문 140건(아래의 참고문헌 번호 21부터 160번까지는 <https://doi.org/10.7910/DVN/LMPTQH>의 참고문헌 번호로 대체함)의 주요 내용을, 결과 위주로 간략히 정리해 비교하면서 살펴보았다.

전문자격과 대학수학 시험

전문자격 시험 : 의사, 레지던트, 변호사, 로스쿨 입학 등의 전문자격 관련 시험 중 United States Medical Licensing Examination(USMLE)에서 GPT-3.5[21-22, 55]는 통과 수준이었으나 응시자 평균에는 미치지 못하였다. 하지만, GPT-4[53]에서 합격선과 응시자 평균을 상회하는 우수함을 보였고, 이러한 사실은 일본의 의사 자격시험[45, 54]에서도 확인되었다. 의사 중 경험과 경력이 더 필요한 전문의 자격을 위한 각종 시험에서 합격 여부가 분명하지 않게 기술된 연구논문[25, 28, 38, 47]도 있지만, 불합격 수준[23-24, 26-27, 30-31, 33, 37, 40, 44, 46], 합격에 근접[29, 32, 36, 42], GPT-3.5에서 합격[35]이거나 GPT-4에서 합격[39, 43] 수준 등으로 다양하게 평가되었다. 합격 여부가 의문스러운 독일어 전문의 자격 시험[41]도 있지만, 대만의 약사 시험[34]에는 합격에 근접한 수준이었다. 또한 변호사 자격 부여를 위한 Bar 시험[49, 51]과 Certified Public Accountant(CPA) 시험[52]에서는 인간 수준의 성능을 보였다. 미국의 Law School Admission Test(LSAT)에서는 간신히 통과하는 수준[48]이었지만, 이 시험에 해당하는 한국의 Legal Education Eligibility Test(LEET)[50] 시험에서는 문제의 유형에 따라 큰 격차를 보였다. 조사된 전문자격 시험 분야의 성능 평가 결과, 일부 분야에서 응시자의 평균에는 미치지 못하더라도 대체로 합격할 수준이었으며, GPT-4에서는 인간의 평균을 뛰어넘는 경우도 간혹 보였다[49].

대학수학 시험 : 대학교 입학과 수업과정 및 졸업 관련 시험에서 풀이 과정이 복잡한 수학과 화학 시험에서는 다소 아쉬운 성적[56, 59]이었다. 다른 대학과정 시험에는 간신히 통과할 성과[63], 분야별로 다양한 성과[57, 61, 64], 거뜬히 통과하거나 놀라운 성과[58, 60, 62, 65] 등 평가자는 물론 대상 및 평가 방법별

로 다양한 결과를 보이면서 인간 대비 충분한 경쟁력을 보여주었다. 이처럼 대학수학 과정의 각종 시험에서 ChatGPT는 과정별로 다양하게 평가되었지만, 일부를 제외하고 다수 분야에서 응시자와 유사한 수준과 능력을 보유한 것으로 평가되었다. 다만, GPT-4로 평가한다면, 실망스러웠던 일부 분야도 저조한 성적을 극복할 가능성이 기대되었다.

의료 업무에서의 지원 가능성

의학과정 시험 : 대학수학 과정 중의 시험에서 너무 많아서 별도로 분류된 다양한 의학과정 시험에서 보면 GPT-3.5는 합격하지 못 할 수준[67], 부분적인 통과[68-69, 73, 75], 완전히 통과[70-71, 78-79], 통과와 함께 인간 이상[72], 전문의 시험 대비용으로도 충분[76]하다고 나타나 평가 사례별로 다양한 수준과 결과를 보였다. 의학 분야의 전문 지식 생성 능력에서도 ChatGPT는 상당한 수준을 보여, 이를 근거로 하면 의학 교육의 보조 도구로 사용될 가능성을 보유한 것으로 일반적으로 평가되었다. 특히 GPT-3.5에서는 한국 의대생의 평균에 비해 크게 낮은 기생충학 성적[66]을 보였지만, GPT-4에서도 불합격[74]과 전문의 시험 대비용으로도 충분[77]하다고 평가 결과가 상반되어서 이 문제가 얼마나 극복되었는지 기대가 크다.

진료와 의사 : 학술적 이용 측면에서 가장 치열하게 논의 중인 의료분야의 ChatGPT 성능 평가 결과, 만족하는 경우도 있지만 대체적으로 인간 의사보다는 아직 역량이 부족하다는 결과였다. 하지만 역량이 부족하다는 평가[90, 92, 94, 96, 106, 108], 주제별로 다양한 수준[81], 부분적으로 의료 현장에 수용이 가능한 수준[80, 82-83, 85, 89, 97, 99, 100-105], 아쉬움도 있지만 충분히 이용할만한 성과[84, 86, 88, 91, 93, 95, 98, 106-107, 109], 인간 의사보다 더 나은 수준[87] 등 다양한 결과로 나타났다. 의료 정보에서 진단의 정확성과 신뢰성(진단 근거)은 매우 중요한 기본적이고 필수적인 문제이므로, ChatGPT를 사용할 때 어느 정도 유용성과 이점은 있어도 발생할 수 있는 잠재적 위험을 신중하게 고려하면서 우려를 해결해야 할 것으로 평가되었다. 특히, 인체 해부학이나 약학과 같은 정교하고 복잡한 지식과 진단에는 여전히 ChatGPT가 한계를 드러내는 것으로 보였다. 결국 ChatGPT는 인간 의료 전문가를 대체하는 용도보다는, 의료 전문가의 업무를 보강하고 지원하여 환자에게 더 나은 치료와 그 기회를 제공할 수 있도록 의료인을 지원하고 보조하는 도구로 자리매김해야 한다는 연구논문이 많았다. 직업 특성상 지식정보의 확보 수준이 매우 높고 인명과 관련되어 엄격한 의료분야에서도, 이처럼 ChatGPT를 이용한 의료진의 업무 지원 가능성을 조심스럽게 그러나 가장 적극적인 태도와 관심으로 이용을 모색하고 있었다.

환자와 간호 : 의료분야 중 의사보다는 전문성과 필요한 지식수준이 낮을 것으로 보이는 환자에 대한 정보 지원과 간호 지식에서 ChatGPT는 인간 의료진보다는 역부족이어도 도움이 되는 수준[116, 122-124,

127], 만족할만한 수준[110-113, 115, 117-121, 125-126] 등으로 조사되었다. 환자와 관련된 간호와 환자 질문에 대한 답변 정보 생성에 대해서는, 대부분 만족할만한 수준으로 평가되었다. 하지만, 간호 업무 현장에서 ChatGPT의 즉각 이용에는 의사처럼 조심스러운 접근이 필요하다는 의견이 대세를 이루었다. 따라서 병원 방문 전후의 환자에게 제공될 의학 정보 작성과 서비스 등에는, ChatGPT가 도움이 될 여지가 많다고 일반적으로 평가되었다.

논문과 글쓰기 등

논문 작성이나 글쓰기와 관련된 다양한 평가 중 긍정적[136] 평가가 있는 한편 문제가 있어서 확인이 필요한 경우[137, 140-141]도 있었다. 글쓰기에서 연구주제 선정과 생성[132, 147], 보고서 작성 능력[131, 133, 142-143, 150-151], 내용 요약 능력[129, 138, 144, 147, 154], 적절한 번역 능력[128], 개선된 문법적 글쓰기[130, 155], 심사자가 AI 지원 생성임을 알아채기 어려운 글쓰기[135], 글쓰기 채점의 부분적인 자동화 가능성[148] 등 다양한 용도의 활용성을 보인 것으로 평가되었다. 그러나 생성된 답변의 근거에서 일부 내용이 허위로 가공되거나 형편없는 참고문헌 제시 능력[134, 139, 145] 문제도 확인되었다. 그럼에도 논문 작성과 글쓰기에서 ChatGPT는 용도에 따라 다양하게 도움을 받을 수 있다고 다수가 평가하므로, 그 이용이 더 확산될 것으로 보였다.

그러나 글쓰기에서 같은 의미의 다른 표현으로 고쳐쓰기를 해 표절 탐지에서 벗어나는 행위는, 현재 기술로 완벽하게 대응하기가 쉽지 않은 문제이다. ChatGPT 류의 기계적 AI 도구가 손쉽게 지원되어 글쓰기 결과물로 제공되면, 일부 약탈적 저널에서 부실하고 비윤리적인[77] 생성물의 출판 가능성이 더 높아지므로, 세계적 대응과 사회적 관심이 필요하다. AI 챗봇을 통해 인간의 참여와 개입이 낮은 저품질 원고를 빠르게 생성할 수 있다는 점을 악용한다면, 부실한 논문 출판이 확산되면서 내용의 진위 판별에 소요되는 사회적 에너지 낭비가 크기 때문이다. MDPI식(Kim & Park 2023)의 지나친 신속 출판과 더불어 ChatGPT류가 생성한 결과를 이용한 논문 투고가 더 늘어나면서 저널, 편집자, 독자 등 출판계의 이해관계자 모두에게 작성자와 내용의 진위 여부 확인에 시간 낭비가 우려된다. 따라서 글쓰기 평가자나 논문 심사자들은 이러한 AI 생성물 탐지 도구의 기술적 지원과 강력한 성능을 기대하고 있었다. 이러한 지원에 제공될만한 탐지 도구가 유효하다는 평가[146, 149, 152]도 있지만, 더 깊이 들여다보면 아직 기술적 한계[153]도 보여 신중한 판단이 필요하다. 한편 5건의 기타 분야로 분류할만한 내용은, ChatGPT 자체의 성능 평가나 인간과 비교한 경우였다. 이러한 범주의 성능 평가는 ChatGPT의 이용적 측면보다는 다분히 기술적 측면의 평가 결과여서, 본 연구의 범위를 벗어나므로 상세한 기술은 생략한다.



ChatGPT 관련 논문의 학술적 이용 리뷰에서 본 시사점

위에서 검토한 ChatGPT 관련 논문의 학술적 이용 측면에서의 출판 동향을 보면, 시간이 지나면서 점차 preprint 수록이 감소되고 초기의 PubMed 위주에서 점차 Scopus와 WoS 수록 건수가 늘어나 비의학 분야의 논문도 증가하면서 초기의 의견성이 강한 넓은 의미의 논문 위주에서 실질적 연구 데이터에 근거한 연구 논문이 증가하였다. 분야별로 보면 이론과 실무가 중요한 의학계에서 가장 많은 관심과 적극성을 보였고, 저널 출판계도 저자 자격 부여 문제에 대해 매우 신속한 입장을 드러냈다. 반면에 교육계는 의학계 다음으로 논문이 많이 출판되었지만 논문 속의 대응은 다양하였고, 저작권과 특허권 분야에서는 AI의 지나친 활용과 지적재산권 권리자 자격에 대한 소송이 있었다. ChatGPT의 성능을 평가한 연구논문의 다수는 학술 과정에서 이용을 찬성하는 것이 대부분을 차지하는 가운데, 문제점도 다양하게 지적되고 있었다. 다수의 연구논문에서 적극적이지만 한편으로는 조심스럽게 문제점을 극복하면서 ChatGPT의 이용을 기대하는 경우가 많아 이용을 고려하는 사람은 이를 참고할 필요가 있다.

각종 시험에서 ChatGPT의 성능을 평가한 결과, 일부(Suchman *et al* 2023)를 제외하고 응시자 수준을 넘어서면서 GPT-4에서는(Cai *et al* 2023) 더 큰 개선이 기대되고 있다. 연구논문마다 다양한 방법과 기준으로 성능과 품질 평가가 이루어져 이용 여부의 판단이 쉽지는 않지만, 평가 결과 일부 문제가 있어도 인간의 지적 활동을 지원하고 보조함에 있어서 ChatGPT의 이용을 긍정적으로 보는 의견이 더 많았다. 물론 진단 등 진료 영역에서 의료진의 완전 대체보다는 보조나 지원 역할이 적절해 진료 현장에 당장 투입하기엔 아직 어려움이 많지만, 그래도 부정적 의견보다는 긍정적 분위기가 강하였다. 특히 환자용 정보 제공과 정보비서 역할은 큰 문제가 없는 것으로 평가되어서, 이 분야에서 가장 적극적인 도입이 예상된다. 문제는 ChatGPT가 인간의 지적 업무에 도움은 되지만, 정확성을 요하는 경우에는 생성된 오류나 문제를 찾아 수정하는 새로운 과업이 필요하다는 점이다. 지금까지 인간만이 생산하던 지식정보를 이제는 기계류 AI가 쉽게 생성하게 되었다. 따라서 만들어진 지식정보 중에서 인간 영역과 기계 영역을 구분하는 것이 중요한데, 이를 확실하게 분류하려면 많은 시간과 돈이 소요되어야 한다는 새로운 도전이 우리 앞에 생겼다. ChatGPT를 이용한 글이 쉽게 생성될 경우 파일 버전, 작성자, 작성된 내용의 진위 여부를 가릴 수 있는 기술적 조치가 조기에 확보되지 않으면, 논문 철회(Van Noorden 2023b)처럼 많은 사회적 비용과 혼란이 발생할 것으로 예상되기 때문이다.

그럼에도 불구하고 ChatGPT의 긍정적이고 유용한 성능과 혜택만큼 그늘진 단점과 한계(환각, 학습 정보원의 저작권 침해 우려, 생성된 답변의 표절과 저작권 위반 우려, 전면 학습 후 새로운 정보의 추가 학습에 대한 시간적 지체, 개인정보 보호와 의료 관련 법적 문제 등)에도 관심을 더 기울여, 생성형 AI 이용에 따

른 윤리 확립과 법적 제도화도 같이 노력해야 한다. 이처럼 사회적 논란이 심한 ChatGPT의 성능 평가에는 다른 관점과 이해관계와 결과의 다양성도 존재하므로, 세계적으로 동일한 눈높이와 시각으로 AI의 성능과 품질을 정확하게 판단할 수 있도록 평가 방법도 더 정교해지고 표준화될 필요가 있다. ChatGPT의 성능 평가에서 프롬프트를 사용한 횟수는 어떻게 결정하나?, 복수의 프롬프트를 사용할 경우 어떤 프롬프트를 가지고 측정하나?, 계속 기술이 진보되는 과정에서 어느 시점의 어느 버전을 가지고 평가하나?, OpenAI나 외부 협력사의 API 혹은 plugins 환경 중 어디에서 평가하나?, GPT-3.5에서 이미지가 중요한 문제를 포함한다면 어떻게 text에 가깝게 전환하여 평가하나? 등에 따라서 평가 결과가 확연히 달라질 것이므로 평가 방법과 기준의 표준화가 매우 중요하기 때문이다. 그것도 단일 학문 분야의 문제가 아니라 이 리뷰의 연구 데이터에서 보듯이, ChatGPT는 여러 학문 분야에서 관심을 갖는 문제여서 많은 이해관계자의 갈등 속에서 학계의 다수가 동의하는 평가 방안을 확립해야 하므로 더 어려운 문제이다.

많은 성능 평가 연구논문 중 GPT-4는 거의 완벽한 성과와 연구 아이디어 생성에서 GPT-3.5보다 개선되지 않은 성과(Gupta *et al* 2023)로 평가가 상반되지만, GPT-4로 된 ChatGPT의 성능은 대체로 더 강력해지고 문제점은 더 해소되고 있었다. OpenAI의 발표(OpenAI 2023)처럼 종전 GPT-3.5보다 GPT-4에서 여러모로 개선되었음이 확인되지만, GPT-4 기반의 성능과 품질 평가가 지금보다 더 늘어나면 GPT-4에 대한 더 정확한 판단과 실태 확인이 가능할 것이다. 연구된 본 리뷰의 ChatGPT 성능 평가 연구논문 140건 중 7개월 동안 29건이 GPT-4를 기반으로 평가된 결과였다. 먼저 검색한 7개월이면 GPT-4가 공개된 지 3개월 지난 시점이어서 적지는 않지만, 앞으로 더 많은 평가가 최신 기술이 반영된 GPT-4를 위주로 평가되어야 현재 모습에 더 가깝게 될 것이다. 특히, GPT-3.5에서는 도표와 그림 정보 등에 대한 평가가 제한되어 text 위주로 평가되었지만, 멀티모달 기능이 지원되는 GPT-4에서는 이제 품질 평가를 하기가 더 쉬워지면서도 한편으로 더 복잡하게 되었다. 당분간 GPT-3.5와 GPT-4가 공존하는 상황이어서, AI의 성능 평가 연구논문에 사용된 버전과 평가 환경의 구분도 중요한 문제로 부각되고 있다.

요약과 결론

ChatGPT 관련 논문이 나날이 증가하는 지금 그 성능과 품질을 더 정확히 판단하기 위한 목적으로, 본 원고는 preprint를 포함해 주요 DB에서 출판된 논문을 1,105건의 엑셀 리스트로 만들어 연구 동향(출판 동향은 3,284건을 재조사)을 살펴보고, 그 중 객관적이고 계량적 방법으로 ChatGPT의 성능과 품질 평가를 다룬 140건의 연구논문을 비교 분석하였다. 출시된 지 7개월이라는 ChatGPT 시작 초기에 주요 DB 검색 결

과를 반영한 ChatGPT 관련 연구 동향과 품질을 비교 검토하는 리뷰를 작성하면서, 연구데이터를 근거로 한 연구논문보다는 각종 속보성의 넓은 범위의 논문이 더 많이 출판된 것으로 분석되었다. 분야별로 보면 교육계는 문제에 대응해 교육 방향을 정비하려는 모습을 보였고, 저작권과 특허계에서는 AI 생성 창작물의 지적재산권 보유 여부에 관한 소송을 주시하면서, 사회적 문제화에 대비해 정부의 법률적 제도화도 추진되고 있었다. 그리고 저널 출판계에는 AI 이용에 따른 조치와 저자 인정 기준이 대체로 확립되었고, 의료계에서는 의료인 지원 가능성을 탐구하는 많은 논문이 적극적으로 출판되고 있었다. 연구 결과, ChatGPT는 인간의 지적 활동과 의료 등의 실무 및 학술 과정에서 유용한 지원자이자 조력자의 역할이 충분함을 다수 논문에서 인정하고 있었다. 따라서 지금은 ChatGPT 도입에 대해 반대하거나 머뭇거릴 것이 아니라, 문제점과 한계를 제도와 기술로 극복해 인간의 지적 활동을 더 원활하게 도와주는 보조도구로 적극 활용하자는 의견이 많았다. 그러나 학습된 정보원을 공개하지 않는 상태에서 OpenAI도 알고 있는 생성형 챗봇에 태생적으로 발생할 수 있는 환각(hallucination)과 저작권 위반 가능성은 물론 근거 기반의 참고문헌 제시 능력에도 개선이 필요하다는 논문이 많았다.

ChatGPT의 등장으로 인해 Google식 웹 페이지 검색 시대의 input keywords 대신 prompt sentences를, displaying web pages 대신 writing answers를, only text search 대신 multi-modal query로 변모되면서, AI가 생성한 결과에 기존 검색을 보완하는 시대로 변화되고 있다. 여기서 중점적으로 다른 학술적 이용이 아니라 갤럭시 S24처럼 실생활과 업무에 ChatGPT가 적용된다면, 더 많은 변화가 일어날 것이다. GPT-3.5나 GPT-4와 같은 파운데이션 모델은 주로 인터넷에 공개된 데이터를 특정 시점까지만 학습하므로 조직이나 저작권이 보호된 전문 데이터를 학습하지 못해서, 생성하는 정보의 전문성과 최신성에 약점을 갖기 때문에 선별된 데이터로 더 학습시키는 미세조정과 분야별 전문 AI도 필요하다. 따라서 기술적 측면이 아니라 이용자 측면에서 ChatGPT의 유용성과 그 수준을 가늠해야 하는 일반인은, 본 리뷰(review)와 같이 ChatGPT 이용 시 장단점과 성능 및 품질에 대해 정확한 정보를 가지고 판단하고 대응할 필요가 있다. 다만 이 리뷰는 주요 DB에서 수록된 정보의 불일치를 완전히 확인하지 못했고, 선행연구의 다수자 의견 위주로 정리한 연구라는 제한점이 있다. 그럼에도 학계와 일반인의 ChatGPT 채택 여부의 판단에 본 연구가 조금이라도 도움이 되길 희망하면서, 정보 리터러시 문제를 포함해 파악된 ChatGPT의 많은 문제점이 앞으로 더 개선되어서, 인류에게 파멸이 아니라 더 발전하고 편리한 생성형 AI 시대로 진화하는 세상이 도래하길 기대해 본다.

※ 2023년도 KSLA 하계 학술세미나 발표 후 본 원고의 상당 부분은 박계숙 前 KSLA 회장의 조언을 거쳐 *Science Editing* 저널의 리뷰 논문(Trends in research on ChatGPT and adoption-related issues discussed in articles: a narrative review)으로 투고해 2024년에 출판(<https://doi.org/10.6087/kcse.321>)되었지만, 회원을 위해 일부를 보완해 KSLA의 승인에 따라 수정 게재한다.

참고문헌

- 김상준. “학술 웹 데이터베이스의 품질 비교 평가 : NDSL, PubMed, Scopus와 Web of Science를 중심으로.” 『정보관리연구』 2005;36(3):127-165. <https://doi.org/10.1633/JIM.2005.36.3.127>
- Adetayo AJ. “Artificial intelligence chatbots in academic libraries: the rise of ChatGPT.” *Library Hi Tech News*. 2023;40:18-21. <http://dx.doi.org/10.1108/LHTN-01-2023-0007>
- Ariyaratne S, Botchu R, Iyengar KP. “ChatGPT in academic publishing: an ally or an adversary?.” *Scott Med J*. 2023;68:129-130. <https://doi.org/10.1177/00369330231174231>
- Boiko DA, MacKnight R, Kline B, Gomes G. “Autonomous chemical research with large language models.” *Nature*. 2023;624:570-578. <https://doi.org/10.1038/s41586-023-06792-0>
- Cai LZ, Shaheen A, Jin A, et al. “Performance of generative large language models on ophthalmology board-style questions.” *Am J Ophthalmol*. 2023;254:141-149. <https://doi.org/10.1016/j.ajo.2023.05.024>
- Dwivedi YK, Kshetri N, Hughes L, et al. ““So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy.” *Int J Info Manag*. 2023;71:102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Frederick DE. “ChatGPT: a viral data-driven disruption in the information environment.” *Library Hi Tech News*. 2023;40:4-10. <http://dx.doi.org/10.1108/LHTN-04-2023-0063>
- Gupta R, Herzog I, Najafali D, et al. “Application of GPT-4 in cosmetic plastic surgery: does updated mean better?.” *Aesthet Surg J*. 2023;43(8):NP666-669. <https://doi.org/10.1093/asj/sjad132>
- Helberger N, Diakopoulos N. “ChatGPT and the AI Act.” *Internet Policy Rev*. 2023;12:28. <https://doi.org/10.14763/2023.1.1682>
- Jang D, Yun TR, Lee CY, et al. GPT-4 can pass the Korean National Licensing Examination for Korean medicine doctors. *PLoS Digit Health*. 2023;2:e0000416. <https://doi.org/10.1371/journal.pdig.0000416>
- Kasneci E, Sessler K, Küchemann S, et al. “ChatGPT for good? on opportunities and challenges of large language models for education.” *Learn Individ Differ*. 2023;103:102274. <http://dx.doi.org/10.1016/j.lindif.2023.102274>
- Kesselman MA. “LHTN special issue on ChatGPT editor’s introduction.” *Library Hi Tech News*. 2023;40:1-3. <http://dx.doi.org/10.1108/LHTN-04-2023-139>
- Kim SJ, Park KS. “Publishing trends of journals and articles in Journal Citation Reports during the COVID-19 pandemic: a descriptive study.” *Sci Ed*. 2023;10:78-86. <https://doi.org/10.6087/kcse.300>
- Lee JY. “Can an artificial intelligence chatbot be the author of a scholarly article?.” *Sci Ed*. 2023;10:7-12. <https://doi.org/10.6087/kcse.292>
- Lo CK. “What is the impact of ChatGPT on education? a rapid review of the literature.” *Educ Sci*. 2023;13:410. <http://dx.doi.org/10.3390/educsci13040410>
- OpenAI. “GPT-4 [Internet].” OpenAI; 2023 [cited 2023 Jul 8]. Available from: <https://openai.com/research/gpt-4>

- OpenAI, "Introducing ChatGPT [Internet]." OpenAI; 2022 [cited 2023 Jul 8]. Available from: <https://openai.com/blog/chatgpt>
- Oyelude AA. "Much ado about ChatGPT: libraries and librarians perspectives." *Library Hi Tech News*. 2023;40:15–17. <http://dx.doi.org/10.1108/LHTN-03-2023-0050>
- Panda S, Kaur N. "Exploring the viability of ChatGPT as an alternative to traditional chatbot systems in library and information centers." *Library Hi Tech News*. 2023;40:22–25. <http://dx.doi.org/10.1108/LHTN-02-2023-0032>
- Picht PG, Thouvenin F. "AI and IP: theory to policy and back again – policy and research recommendations at the intersection of artificial intelligence and intellectual property." *IIC - Int Rev Intell Prop Compet Law*. 2023;54:916–940. <https://doi.org/10.1007/s40319-023-01344-5>
- Stokel-Walker C, Van Noorden R. "What ChatGPT and generative AI mean for science." *Nature*. 2023;614:214–216. <http://dx.doi.org/10.1038/d41586-023-00340-6>
- Suchman K, Garg S, Trindade AJ. "Chat Generative Pretrained Transformer fails the multiple-choice American college of gastroenterology self-assessment test." *Am J Gastroenterol*. 2023;118:2280–2. <https://doi.org/10.14309/ajg.0000000000002320>
- Thorp HH. "ChatGPT is fun, but not an author." *Science*. 2023;379:313. <http://dx.doi.org/10.1126/science.adg7879>
- Tlili A, Shehata B, Adarkwah MA, et al. "What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education." *Smart Learn Env*. 2023;10:15. <http://dx.doi.org/10.1186/s40561-023-00237-x>
- Valentín-Bravo FJ, Mateos-Álvarez E, Usategui-Martín R, et al. "Artificial Intelligence and new language models in Ophthalmology: complications of the use of silicone oil in vitreoretinal surgery." *Arch Soc Espan Oftalmol*. 2023;98:298–303. <https://doi.org/10.1016/j.oftale.2023.04.011>
- Van Noorden R. "ChatGPT-like AIs are coming to major science search engines." *Nature*. 2023;620:258. <https://doi.org/10.1038/d41586-023-02470-3>
- Van Noorden R. "More than 10,000 research papers were retracted in 2023 – a new record." *Nature*. 2023;624:479–481. <http://dx.doi.org/10.1038/d41586-023-03974-8>
- Wang FY, Li J, Qin R, et al. "ChatGPT for computational social systems: from conversational applications to human-oriented operating systems." *IEEE Tran Comput Soc Syst*. 2023;10:414–425. <https://doi.org/10.1109/TCSS.2023.3252679>
- Wikimedia Foundation, Inc. "ChatGPT [Internet]." Wikimedia Foundation; 2023 [cited 2023 Jul 8]. Available from: <https://en.wikipedia.org/wiki/ChatGPT>
- Zhang B. "ChatGPT, an opportunity to understand more about language models." *Med Ref Serv Q*. 2023;42:194–201. <http://dx.doi.org/10.1080/02763869.2023.2194149>
- Zhuk A. "Navigating the legal landscape of AI copyright: a comparative analysis of EU, US, and Chinese approaches." *AI Ethics*. Online published: 30 May 2023. <https://link.springer.com/article/10.1007/s43681-023-00299-0>

Korea Special Library Association

KSLA Insight

발행일 2024년 2월 21일

발행인 이갑훈

편집위원 KSLA 편찬위원회
심유정(위원장), 이경미(부위원장), 박희숙, 송유영, 신정은, 이진솔, 황혜경

발행처 한국전문도서관협회

I S S N 3022-6198