



챗GPT의 등장과 인공지능 분야의 과제

정준화

챗GPT는 초거대 인공지능 모델 GPT-3.5를 누구나 쉽게 사용할 수 있도록 만든 챗봇이다. 간단한 질문이나 명령만 하면 챗GPT가 체계적 구성을 가진 문서를 만들어 준다. 이러한 생성형 인공지능의 확산에 대비하여 대량의 데이터를 빠르게 학습하고 처리하는 컴퓨팅파워 확보, AI 반도체 연구개발, 데이터 유통·거래 활성화, 인공지능을 고려한 저작권 정비, 이용자의 활용 역량 강화와 같은 준비가 필요하다.

1 서론

2022년 11월 30일 미국의 인공지능(artificial intelligence: AI) 기업 오픈에이아이(OpenAI)가 '챗GPT(ChatGPT)'를 시장에 내놓았다. 챗GPT는 오픈에이아이의 AI 모델¹⁾인 'GPT²⁾-3.5'를 대화 방식으로 사용할 수 있게 만든 챗봇(chatbot)이다.

챗GPT는 언어와 문자로 표현할 수 있는 것이라면 논문, 보고서, 문학, 코딩 등 분야를 가리지 않고 수준 높은 결과물을 만든다. 그래서 챗GPT 또는 GPT-3.5를 '생성형 AI(generative AI)'라고 한다.

PC나 스마트폰만 있으면 누구나 챗GPT에 접속하여 간단한 채팅만으로 GPT-3.5의 기능을 이용할 수 있다. 월 20달러에 달하는 유료 서비스가 있지만 무료로도 사용할 수 있어서 비용 부담도 낮다. 덕분에 챗GPT는 출시 두 달 만에 전 세계 월간 활

성 사용자 수(MAU) 1억 명을 돌파하여 인스타그램(2년 반 소요), 틱톡(9개월 소요) 등이 가지고 있던 종전 기록을 단숨에 갈아치웠다.³⁾ 이는 챗GPT가 거둔 성과이지만, 한편으로는 생성형 AI의 대중화 가능성을 확인하는 계기로서도 의미가 크다.

챗GPT가 열어준 생성형 AI 시대를 맞이하여 우리는 어떠한 준비를 해야 할까? 이 글을 통해 챗GPT의 특징과 국내외 동향을 살펴보고 인공지능 분야에서의 향후 과제를 모색하고자 한다.

2 챗GPT의 특징

(1) 초거대 AI

오늘날 AI 알고리즘에는 인간의 뇌를 모방한 기계학습 방식인 심층학습(deep-learning)이 적용된다. 학습을 통해서 도출된 값을 파라미터(parameter, 매개변수)라고 하는데, 파라미터가 많을수록 AI 모델의 성능이 좋아진다.

1) 알고리즘(algorithm)은 데이터를 학습하여 그 내용을 추론하는 절차 또는 체계를 말한다. 모델(model)은 학습이 끝난 알고리즘을 말하며, 사용자는 모델에 요구사항을 입력하고 모델이 연산 결과를 제시한다. 양자는 개념상 미세한 차이가 있지만 실제로 혼용되는 경우가 많다.

2) GPT는 '생성형 사전학습 트랜스포머(Generative Pretrained Transformer)'의 약자이다.

3) Krystal Hu, "ChatGPT sets record for fastest-growing user base - analyst note", *Reuters*, 2023.2.3.



2020년 기준으로 마이크로소프트의 AI 모델인 'MS Turing-NLG'에는 170억 개의 파라미터가 사용되었는데, 챗GPT의 두뇌에 해당하는 GPT-3.5에는 이것보다 열 배 이상 많은 1,750억 개의 파라미터가 사용되었다. GPT-3.5와 같이 대량의 파라미터를 적용하는 AI 모델을 '대규모 언어 모델 (large language model: LLM)' 또는 '초거대 AI'라고 한다. GPT-3.5 외에도 구글의 '람다 (LaMDA)', 네이버의 '하이퍼클로바(Hyper CLOVA)'와 같은 생성형 AI는 대부분 초거대 AI를 지향한다. 현재 오픈에이아이 개발중인 'GPT-4'에는 100조 개 이상의 파라미터가 사용될 것으로 예상되어 생성형 AI의 대형화는 당분간 지속될 전망이다.

(2) 퓨샷러닝과 챗봇 방식

초거대 AI의 장점은 소량(few)의 학습데이터만 있어도 원하는 결과를 얻는 '퓨샷러닝(few-shot learning)'이 가능하다는 점이다.⁴⁾ 기존에는 개별 목적에 따라 각각 데이터셋을 학습시켜 전용 AI 모델을 만드는 미세조정(fine-tuning) 방식이 일반적이었지만, 초거대 AI는 대규모 데이터⁵⁾를 사전학습(pre-training)시켜 공통적인 기본 AI 모델을 만든 다음 그 위에 소량의 데이터만 추가 학습시켜도 정확도가 높은 결과를 제시할 수 있다.

챗GPT는 소량의 학습 데이터를 확보하기 위해 챗봇 방식을 적용한다. 이용자가 대화창에 간략한 지시어(prompt)를 입력하면 챗GPT는 정리된 결과물(텍스트)을 생성한다. 지시어가 효과적일 경우에는 추가적 설명과 학습 없이(zero-shot) 한 번의 채팅만으로 원하는 결과를 얻을 수도 있다.

4) Brown, Tom, et al., "Language models are few-shot learners", *Advances in neural information processing systems*, 33, 2020, pp. 1877-1901.

5) GPT-3.5에는 인터넷에서 수집(크롤링)한 데이터 4,100억 개, 논문·보고서와 같은 웹텍스트 190억 개, 인터넷 기반의 말뭉치 670억 개, 인터넷 사전(위키피디아) 정보 30억 개의 학습데이터가 사용되었다.

(3) 생성형 트랜스포머 모델

기존 AI 모델은 여러 장의 사진 중에서 고양이를 찾는 것과 같은 식별(recognition) 기능에 초점을 두었다. 이와 달리 챗GPT는 글, 문장, 단어, 부호와 같은 언어로 결과물을 생성하는 AI 모델이다. 예를 들어 챗GPT에 '고양이는 어떤 모양이냐'고 물어보면 '고양이는 네 발로 서 있는 동물로, 작은 머리와 둥글고 작은 귀, 눈과 코, 그리고 길고 부드러운 꼬리를 가지고 있습니다'라는 문장을 만들어 낸다.

챗GPT가 문장을 생성하는 원리는 특정 단어 다음에 올 가장 적합한 단어를 예측하고 이것을 연결하여 문장을 만드는 것이다. 다음에 올 단어 예측에는 '트랜스포머(transformer)'라는 AI 모델이 적용된다.⁶⁾ 트랜스포머 모델은 데이터의 입력 정보와 출력 정보를 동시에 처리하는 방식으로, 입·출력 정보를 순차적으로 처리하던 기존의 순환신경망(recurrent neural network: RNN) 방식에 비해 계산 효율성을 크게 높일 수 있다.

3 챗GPT 등장 이후 국내외 동향

(1) AI 공급 분야 : 검색 서비스의 변화 예상

챗GPT 등장 이후 가장 적극적으로 대응하는 기업은 미국의 마이크로소프트다. 자사의 검색엔진 '빙(bing)'과 챗GPT를 결합하여 새로운 검색 경험을 제공하고, 클라우드 환경에서 AI를 사용할 수 있도록 클라우드서비스 '애저(Azure)'에 GPT-3.5를 탑재하는 방안을 모색하고 있다. 이를 통해 궁극적으로 구글이 독점하는 검색 시장과 아마존클라우드가 우위를 점하는 클라우드 시장에서 마이크로소프트의 경쟁력을 높이고자 한다.

구글은 지금까지 적극적으로 AI 투자를 해 왔고, 초거대 AI 모델인 람다도 보유하고 있지만 상용 AI

6) 이상근, 『초대규모 AI 모델(GPT-3)의 부상과 대응 방안』(IT & Future Strategy 제2호), 한국지능정보사회진흥원, 2021.

서비스인 ‘바드(Bard)’ 발표는 오픈에이아이보다 늦었다. 대중적인 AI 서비스에 대해서는 신중하게 접근하자는 내부 판단에 따른 것이었다. 그러나 오픈에이아가 챗GPT를 발표한 이후 내부적으로 위기 상황임을 뜻하는 ‘코드 레드(Code Red)’를 선포하고 AI 대응 전략을 준비한 것으로 보아 구글도 챗GPT가 검색 시장에 미칠 영향을 상당히 위협적으로 느끼는 것으로 볼 수 있다.

국내 기업들은 한글 서비스에 관해서는 우리 기업이 비교우위가 있다고 보고 초거대 AI 대응 전략을 마련하고 있다. 네이버는 2021년 자체적으로 초거대 AI 모델인 하이퍼클로바를 개발하였고, 2023년 상반기에는 한국어에 특화된 생성형 AI 챗봇인 ‘서치GPT’를 선보일 예정이다. 카카오는 초거대 AI 모델인 ‘KoGPT’를 챗봇 조르디, 카카오톡 채널 등과 결합하여 상용 AI 서비스를 제공할 예정이다.

챗GPT는 이용자의 요구 또는 질문에 대한 답을 제시하는 것이어서 기존의 검색 서비스와 지향하는 바가 같다. 차이가 있다면 검색 서비스는 정보의 출처를 제시하여 이용자가 각각의 정보를 찾아 비교·평가하도록 하는데, 챗GPT는 AI가 이 과정을 대신해 준다. 그 결과 챗GPT는 검색 서비스 방식과 주요 사업자의 변화를 초래할 것으로 예상된다.

(2) AI 이용 분야 : 인간의 보완 수단으로 기대

챗GPT 이용자들은 대부분 그 능력에 대해서 긍정적으로 평가하며 다양하게 활용하고 있다. 예를 들어, 미국에서는 챗GPT가 의사·변호사·MBA 시험을 통과하는 수준이라는 평가가 있었고,⁷⁾ 국내에서는 챗GPT에게 2023년도 수능 문제를 풀게 해 보니 외국어는 수능 2등급 수준이었다는 분석도 있었다.⁸⁾ 챗GPT를 이용하여 보고서 초안을 만들거나 챗GPT만 이용해서 책을 출판하는 등 실생활에 직

7) 김성민, 「美 MBA·로스쿨·의사면허 ‘3관왕’ 가능…이 천재의 정체는」, 『조선일보』, 2023.2.2.

8) 허진, 「챗GPT는 수포자?…수능 수학 낙제점」, 『서울경제』, 2023.2.6.

접 활용하는 사례들도 나타나고 있다.⁹⁾

그러나 모든 분야에서 챗GPT가 환영을 받는 것은 아니다. 특히 대학과 같은 교육기관에서 챗GPT 사용은 논쟁적이다. 일부 대학은 모든 수업·과제·평가에 챗GPT 사용을 금지하고, 허가를 받지 않은 챗GPT 이용은 표절로 간주한다고 밝혔다.¹⁰⁾ 이와 달리 어떤 교수는 AI 시대를 살아가는 인간의 역량 함양을 위해 챗GPT 사용을 권장하기도 한다.¹¹⁾

챗GPT 이용에 따른 긍정적 효과와 부정적 효과가 병존하지만, 전반적으로 본다면 지금까지의 디지털 혁신과 마찬가지로 인간을 보완하는 역할이 더 클 것이다. 다만, 이용자가 챗GPT를 오·남용할 경우 사회적 가치와 충돌하는 문제를 초래하고, 경우에 따라서는 직업 기회 감소와 같은 실질적 위협에 직면할 수도 있다. 따라서 AI 이용 확대에 따른 사회 질서의 정립과 자원의 재분배 원칙 수립 등이 병행되어야 한다.

4 AI 정책의 과제

(1) 컴퓨팅파워 구축과 AI 반도체 연구개발 강화

초거대 AI 시대의 기업경쟁력, 국가경쟁력을 강화하기 위해서는 대규모 학습데이터를 빠르게 처리하고, 수많은 이용자 질문에 실시간으로 대답할 수 있는 컴퓨팅파워(computing power)를 확보하는 것이 중요하다.¹²⁾ 이를 위해 정부는 국가 슈퍼컴퓨터 자산을 확충하고 민간이 이를 초거대 AI 개발에 활용할 수 있도록 「국가초고성능컴퓨팅 혁신전략」에 반영해야 한다. 또한 민간이 컴퓨팅파워 구축에

9) 한순천, 「"30시간 만에 책 집필" 챗GPT, 출판계 판도 흔들까」, 『서울경제』, 2023.2.18.

10) 신정은, 「홍콩대, 챗GPT 사용 금지령…"허가 없으면 표절"」, 『한국경제신문』, 2023.2.19.

11) 박동휘, 「"챗GPT 안쓰면 감점"…AI 전면허용한 대학 교수는」, 『서울경제』, 2023.2.26.

12) 컴퓨팅파워 확보를 위해 챗GPT와 GPT-3.5는 마이크로소프트의 애저 AI 슈퍼컴퓨터에서 학습을 했고, 네이버는 하이퍼클로바를 위해 2020년에 700PF(페타플롭스) 성능의 슈퍼컴퓨터를 구축했다.

적극적으로 투자할 수 있도록 현재의 AI 사업 규제를 점검하고 과감하게 개선해야 한다.

중장기적 관점에서 AI 반도체 역량 확보도 관건이다. 현재 사용하는 GPU는 가격이 비싸고 전력소모가 많은 고비용 방식이기 때문에 컴퓨팅파워 확대에 어려움이 크다. 향후 예상되는 컴퓨팅파워 수요 증가에 효율적으로 대응하기 위하여 AI 모델에 최적화된 AI 반도체 연구개발 지원이 필요하다.

(2) 데이터 구축 확대와 유통·거래 활성화

GPT-3.5와 같은 범용성을 갖춘 초거대 AI 모델을 만들기 위해서는 다양한 분야의 학습데이터를 확보해야 한다. 정부는 지금까지 ‘데이터 댐’ 사업을 통해 AI 허브¹³⁾에 한국어 데이터 93종, 영상이미지 78종, 헬스케어 67종, 재난안전환경 59종, 농축수산물 41종, 교통물류 46종의 AI 학습데이터를 구축했는데 이는 초거대 AI 학습용으로는 여전히 부족하다. 따라서 민간에서 직접 공급하기 어려운 데이터 수요를 파악하여 정부와 공공기관이 보다 적극적으로 학습데이터를 구축할 필요가 있다.

이와 함께 민·관의 다양한 분야에 축적되어 있는 방대한 데이터가 초거대 AI 모델 학습에 활용될 수 있도록 데이터 유통·거래를 활성화시켜야 한다. 「데이터 산업진흥 및 이용촉진에 관한 기본법」상 민간 데이터 거래소의 역량을 강화하고, 웹사이트에 공개된 데이터에 대한 학습용 수집(크롤링) 허용 조건도 명확하게 정립해야 한다.¹⁴⁾

(3) 안심하고 AI를 이용할 수 있는 환경 조성

생성형 AI와 관련된 저작권 규정을 합리적으로 정비할 필요가 있다. 특히 AI 학습에 사용된 데이터의 저작권 허용 범위, 생성형 AI가 만든 결과물의 저작권 보호 범위가 중요 논의 대상이다.

13) <https://aihub.or.kr>

14) 정준화·박소영, 『데이터 거래 활성화를 위한거래소·거래사·크롤링의 현황과 개선과제』(NARS 현안분석 제254호), 국회입법조사처, 2022.

이용자의 정보보호 인식과 실천도 중요하다. 생성형 AI를 이용하는 과정에서 무분별하게 본인 또는 주변인의 개인정보, 기업·기관의 비밀 등을 입력하지 않아야 한다. 입력한 정보는 어떠한 형태로든 기록에 남고 재생산될 수 있기 때문이다.

이용자의 비판적 활용 능력을 높여야 한다. 생성형 AI는 확률적으로 ‘적절’한 표현을 생성하는 것이지 ‘정확’한 내용을 찾아주는 것은 아니다. 전혀 존재하지 않는 내용을 그럴듯하게 표현하는 환각(hallucination)이 종종 발생하기도 한다. 따라서 이용자는 생성형 AI의 결과물이 사실과 다를 수 있다는 점, 그래서 책임 있는 결정에 직접 활용하기 어렵다는 점을 명확하게 인지하고 있어야 한다.

마지막으로, 학습한 데이터의 편향성·불충분성 등으로 인해 인간에게 유해하거나 윤리적이지 못한 내용이 여과 없이 표출되는 문제를 경계해야 한다. 따라서 AI 사업자는 AI 윤리규범을 생성형 AI의 개발 및 사업화에 적극적으로 반영할 필요가 있다.

5 결론

챗GPT가 생성형 AI 시장을 선점하여 이용자의 초기 피드백을 독점하고 있지만, 아직 시장은 초기 단계이기 때문에 우리 기업에도 충분한 기회가 있다. 한글에서의 비교우위를 발판으로 국내 기업들이 생성형 AI를 준비 중이지만, 장기적으로 우리나라를 넘어 비영어권 시장으로의 확산도 노려 볼 만하다. 기업의 꾸준한 노력, 정부의 안정적인 환경 조성, 우수한 인재의 육성과 유치, 이용자의 활발한 사용이 모두 필요한 때이다.

『이슈와 논점』은 국회의원의 입법활동을 지원하기 위해 최신 국내외 동향 및 현안에 대해 수시로 발간하는 정보 소식지입니다. 이 보고서의 내용은 국회의 공식 입장이 아니라 국회입법조사처의 조사분석 결과입니다.

