

거대 언어모델 미래전개도 I' STEEP 프레임으로 기회와 위험 탐색

이명호
(사)미래학회 부회장 | 태재연구재단 자문위원
lee.myungho@gmail.com

윤기영
한국외국어대학교 겸임교수
에프엔에스컨설팅 미래전략연구소장
synsaje@gmail.com

거대 언어모델 미래전개도 I

- I. 들어가기
 - 1. 거대한 충격이 된 거대언어모델
 - 2. 방법론
- II. 변화 전망
 - 1. 개요
 - 2. 사회

거대 언어모델 미래전개도 II(107호 게재)

- 3. 기술
- 4. 경제
- 5. 환경
- 6. 정치
- III. 결론
 - 1. 정책 및 전략 의제 도출
 - 2. 한계와 과제

I. 들어가기

1. 거대한 충격이 된 거대 언어모델

인공지능의 주요 4대 기능은 시각 지능, 음성 지능, 언어 지능 및 그 밖의 통계 지능이라 할 수 있다. 시각과 음성 및 통계 지능에서 신경망 인공지능은 탁월한 성과를 보였다. 이에 반해 언어지능의 성과는 지지부진했다. 인간의 사고가 언어로 구성되며, 인간 활동이 주로 언어를 통해 수행된다는 점을 고려한다면, 언어 지능의 중요성은 높으나 그 성과는 미미했다. 그런데 2022년 말부터 상황이 변했다.

2017년 구글이 Transformer 인공지능 모델을 발표하면서 그간의 언어지능 모델로 활용됐던 LSTM(Long Short Term Memory)나 RNN(Recurrent Neural Network)의 한계를 극복할 수 있다는 기대가 일어났다. 이듬해 OpenAI가 GPT-1(Generative Pre-trained

Transformer-1)을 2년 후인 2019년 GPT-2를, 구글은 BERT(Bidirectional Encoder Representations from Transformers)를 발표했다. 2020년 GPT-3가 발표되면서, 거대 언어모델 개발과 경쟁이 촉발됐다(Thompson, 2023; 윤기영, 2023).

GPT-3는 여러가지 언어 작업을 다수 수행할 수 있다. 대화, 소설과 시나리오 등 창작, 보고서 작성, 요약, 번역, 코딩 등의 작업을 수행할 수 있다. 특수 인공지능에서 일반 인공지능(Artificial General Intelligence, AGI)으로 가는 여정의 발걸음을 뚝 것으로 볼 수 있다. 이후 거대 언어모델 개발과 경쟁은 광범위하게 진행됐다. 중국의 베이징 인공지능 아카데미(BAAI), 알리바바, 바이두 등이 이 경쟁에 참여했고, 우리나라의 네이버와 LG 등도 경쟁에 참여했다. 특히 구글은 내부의 복수 조직이 거대 언어모델을 개발하고 있어 가장 적극적이다.

1 본고는 두 차례에 걸쳐 <SW중심사회> 106호, 107호에 게재

2022년 11월 OpenAI가 챗GPT(ChatGPT)를 발표하면서 대중이 거대 언어모델을 활용할 수 있게 됐다. 단 5일 만에 사용자 100만, 2개월 만에 사용자 1억을 달성했다. 유례가 없는 확산 속도다. 거대 언어모델에 대한 사회적, 경제적 수요가 그만큼 크다는 방증이다. 이에 따라 거대 언어모델 개발 경쟁도 심화됐다. 2023년 3월 말까지 20여 개의 거대 언어모델이 공개됐다(Thompson, 2023).



2023년 3월 공개된 GPT-4는 멀티모달 인공지능이다. 멀티모달은 거대 언어모델의 추세 중 하나다(윤기영, 2023.03.15: 2023). 멀티모달이란 언어지능, 시각지능 및 음성지능을 통합한다. 사진을 입력하면 이를 해석하여 문장을 생성할 수 있으며, 혹은 로봇 팔을 움직일 수 있다. 이 멀티모달은 일반인공지능(AGI)으로 가는 여정에 한 걸음 더 나아갔다고 볼 수 있다.

부분적 일반인공지능(AGI)인 거대 언어모델은 정치, 경제 및 사회 등에 근본적 전환을 가져올 것으로 보인다. 지식과 전문성의 가치와 필요성에 대해 사회적 의문을 제기하며 관련한 사회 시스템의 변경을 요구한다(이명호, 2023). 골드만 삭스(2023.03.26)는 거대 언어모델을 포함한 생성 인공지능이 전 세계적으로 약 3억 개의 일자리를 대체 가능할 것으로 보았다.

이와 같이 거대 언어모델은 정치, 경제 및 사회의 디지털 전환(윤기영, 2018)을 가져올 핵심 동인이다. 거대 언어모델의 동향뿐만 아니라, 이로 인한 정치, 경제 및 사회의 변화 등을 미리 전망해 이들 인공지능이 던질 기회와 위험을 미리 포착하고 대응해야 한다. 이 연구에서 미래학 방법과 기법을 통해 거대 언어모델이 한국사회와 인류사회에 던질 변화를 더듬고 기회와 위험을 탐색했다.

2. 방법론

거대 언어모델은 정치, 경제 및 사회에 깊고 넓은 영향을 미칠 것이다. 이를 전망하기 위해서 미래학에서 활용되는 기법인 미래전개도(Futures Wheel)와 STEEP 프레임틀을 사용했다. 미래전개도는 만약 어떤 일이 발생한다면, 그다음에 발생할 수 있는 사건을 브레인스토밍

으로 탐색하는 기법으로, 그 근원은 Glenn(2009)의 Futures Wheel에서 찾을 수 있다. Futures Wheel은 경제학의 물결 효과(Riffle Effect)(Navarro, 2004)를 제롬 글렌이 미래학에 접목(Glenn, 2009)한 것으로, 설명 가능한 미래(Plausible Futures)를 탐색하는데 탁월하다. 설명 가능한 미래란 역사적 경험, 자연과학과 인간의 심리로 합리적으로 설명 가능한 미래를 의미한다(Voros, 2001). 한편 미래학의 연구 성과에 따르면 통계적 미래(Probable Futures)가 실현될 가능성은 장기적으로 매우 낮다.

글렌(Glenn)의 Futures Wheel은 미래사건을 중심으로 수레바퀴처럼 물결이 전파되는 형태로 구성돼 있다. 윤기영(2018: 2020.11)은 Futures Wheel에 시스템 사고 기법과 외부 환경 변수를 고려하는 형태를 반영해 미래전개도 기법을 발달시키고 실무에 적용했다. 윤기영의 미래전개도는 Futures Wheel보다 복잡하다는 단점을 지녔다. 대신 미래변화에 민감하며, 시스템 사고의 환류(Feedback)를 고려하고, 복수의 외부 변수를 참조할 수 있어 미래변화에 민감하다는 장점을 가졌다. 글렌의 Futures Wheel이 미래학적 사유에 익숙하지 않은 사람이 사용하기에 비교적 적합하다면, 윤기영의 미래전개도는 미래학적 사유에 익숙한 사람이 활용하기에 좋다. 이 연구에서 미래연구 실무자인 소수의 사람이 설명 가능한 미래를 도출하는 것으로 미래전개도를 사용하기에 적합했다. 본 연구에서는 10~15년의 중장기 미래의 관점에서 미래를 고찰하였다.

현재의 사건이나 미래의 사건은 특정 분야에만 파급을 일으키지 않는다. 호수에 돌을 던지면 그 물결은 360° 전방위로 퍼진다. 전 분야를 아우르는 프레임은 다양한데, 여기서는 STEEP을 사용했다. STEEP은 Society, Technology, Economy, Ecology, Politics의 약어다. 여기에 Law, Resource, Population을 더하는 경우가 있는데, 대체로 STEEP 체계에 통합될 수 있다. 그 외에 Value를 더해 STEEPV 체계를 택하는 경우도 있는데 여기서 택하지 않았다.

STEEP 프레임으로 미래전개도를 작성해, 미래의 위험과 기회를 탐색할 수 있다. 이들 기회와 위험은 현재의 정책과 전략의 강건성(Robustness)을 점검할 수 있도록 한다. 일부 기회와 위험은 미래지향적 정책과 전략 의제로 전환될 수 있다.



II. 변화전망

1. 개요

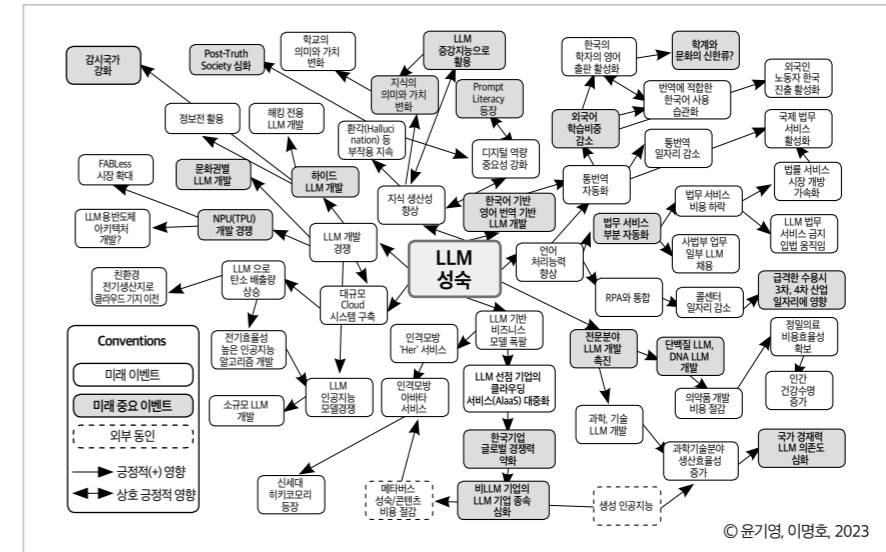
미래전개도 기법과 STEEP 프레임 기반의 거대 언어모델이 어느 정도 성숙한다면 그다음 일어날 수 있는 일을 도출하고 이를 그래프로 표현했다. 거대 언어모델의 성숙 정도는 경제적활용 가능성을 기준으로 했다. 다시 말하자면 일반인공지능(AGI)으로 성숙하거나 혹은 거대 언어모델의 한계인 환각(Hallucination) 등을 수용 가능할 정도로 극복하는 것을 전제로 하지 않았다.

마이크로소프트의 브라우저인 엣지(Edge)에서 마이크로소프트의 챗GPT인 프로메테우스(Prometheus)를 활용하고, 마이크로소프트 오피스 제품군에 Copilot을 융합하고 있어, 이미 거대 언어모델이 부분적으로 비즈니스에 접목되고 있다. 거대 언어모델 번역기인 딥엘(DeepL)의 프로 버전은 유료 서비스로 진행되고 있다. 따라서 거대 언어모델의 경제적 활용은 지금도 상용화 됐다고 볼 수 있다. 위에 제시한 미래전개도에서 지금 이후 어떤 일이 발생할 것인가를 탐색한다.

그래프를 단순화하려 했으나 그렇게 하지 못했다. 거대 언어모델이 전방위적으로 영향을 넓고 깊게 미치는 데 단순화는 연계적 사고의 흐름도 끊을 위험이 있기 때문이다. 오히려 미래전개도에서 유의미한 미래 위험과 기회를 탐색하기 위해서는 이보다 더욱 상세화할 필요가 있다. STEEP 분야별 미래전개도를 나누어 제시함으로써, 복잡성으로 인한 가독성이 떨어지는 단점은 완화하려고 노력했다.



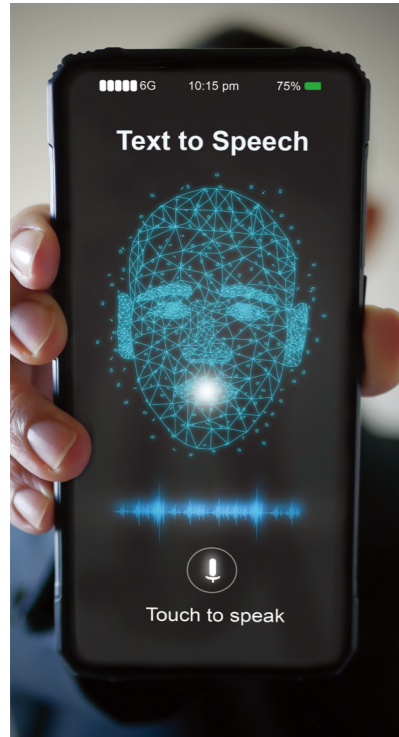
[그림 1] 거대 언어모델이 성숙하는 경우의 미래전개도



2. 사회

가. 미래전개도 설명

거대 언어모델이 가져올 가장 큰 변화는 지식 생산성 향상이다. 챗GPT와 같은 거대 언어모델은 하나의 시스템이 사람과 같이 대화를 이어 나가고, 글을 쓰고, 보고서나 논문을 요약하거나 분석하고, 코딩하고, 문제를 푸는 능력을 보여주었다. 잘못된 정보를 제시하는 등 여러 실수를 드러냈지만, 빠르게 수정된 이후 개략적으로 잘 답변하고 정리된 문서를 작성해 준다. 논리적으로 복잡한 내용까지 분석하고 정리해줘 업무 효율성, 생산성을 높여줄 것으로 기대된다. 현재 거대 언어모델의 가장 큰 문제점은 문맥에 따라 단어 다음에 등장할 단어나, 문장 혹은 문단을 확률적으로 예측해 그럴듯한 문장을 만들어 준다는 데 있다. 이는 상식 밖의 문장도 만들어 내서 거짓말 제조기, 디지털 환각이라고 비판받고 있다. OpenAI의 챗GPT 개발 책임자는 이러한 문제들은 2년 정도 후에는 극복될 것으로 전망했다. 그러나 환각이 근본적으로 극복 불가능하다는 견해도 존재한다. 환각 극복이 불가능하다는 근거로 지식의 특성으로서 형식지에 암묵지가 연결돼 있다는 점을 드는 경우와, 신경망 알고리즘의 근원적 한계에 해당한다는 견해 등이 있다.



중국어 중심으로 개발했다. 그렇기 때문에 각각의 언어가 사회적으로 형성된 관행, 맥락을 이해하지 못하는 경우가 있다. 앞으로는 기본 언어 모델인 영어 번역기와 각국 번역기가 연계돼 더 완벽한 소통이 가능하게 될 것이다. 언어 번역 기능이 향상되면 외국어를 몰라도 외국의 자료를 보거나 대화하는데 지장이 없기 때문에 외국어 학습이 감소할 것이다. 한때 영어를 잘하는 것이 능력의 척도가 된 적이 있는데, 앞으로는 문제를 찾는 능력과 종합적 사고력이 더 중요한 능력이 될 것이다. 번역 기능의 향상은 영어의 수요를 감소시키고, 각국 고유 언어가 더 주목받을 기회를 확대해 줄 것이다. 글로벌화가 진행되고, 영어의 지배력이 높아지면서 전 세계의 소수 언어, 고유한 언어들이 사라졌는데, 이제는 이들 언어가 보존될 가능성이 커졌다. 아이슬란드 정부는 아이슬란드어 보존을 위해 GPT-4를 개발한 OpenAI와 협력하고 있다. 더 정확하게 아이슬란드어로 질문하고 답을 하도록 하고, 아이슬란드어가 다른 나라 언어로 더 정확하게 번역되도록 하기 위해서다. 또한 언어학자들은

사라진 고대 언어로 작성된 문서를 해독하기 위해 인공지능을 사용하고 있다. 사라진 언어를 구사하는 인간은 없어도 인공지능은 남아 있을 수 있다.

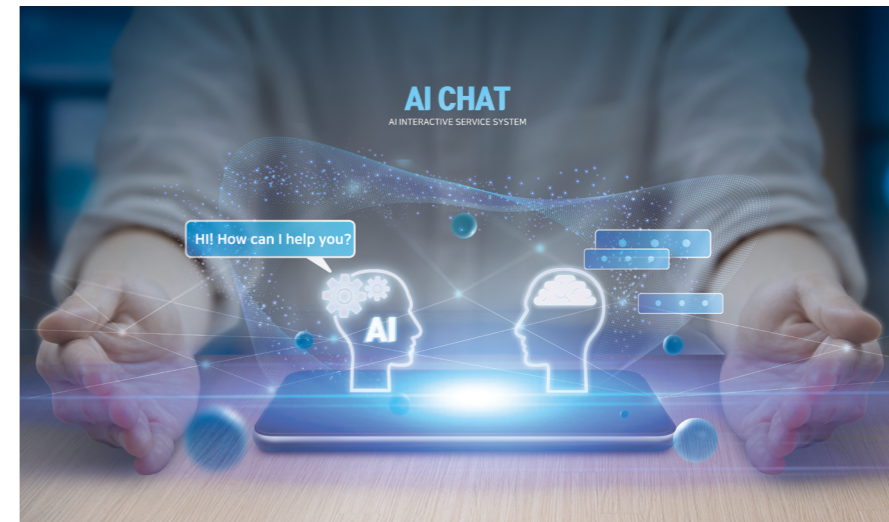
그렇게 언어 번역이 좋아지면, 각국의 유명한 문학 작품을 접할 기회도 많아질 것이다. 우리나라의 뛰어난 작가들의 작품을 보는 외국인들도 늘어나면서 한류가 음악, 영화에서 전반적인 문화, 문학으로 확대될 수 있다. 진정으로 각국의 고유문화, 문학을 세계인이 즐기는 하나의 세계가 될 것이다. 콘텐츠만 세계로 퍼져나가는 것이 아니라 사람의 이동, 교류도 늘어나게 된다. 외국에서 일하기가 쉬워지고, 외국인이 한국에서 일하기도 쉬워진다. 다른 민족의 문화적 특성이나 관습을 이해하지 못해서 생기는 갈등도 인공지능이 줄여줄 것이다. 전 세계 민족이 하나의 세계인이 되는 것을 앞당길 것이다.

나. 기회요인과 위험요인

거대 언어모델은 전반적으로 우리의 지적 역량을 강화하게 한다는 점에서 거대한 기회를 열어준다. 더 적은 노력으로 더 많은 결과를 얻을 수 있어, 인간의 다양한 능력을 개발하는 데 도움이 될 것이다. 언어 장벽을 넘어 세계인과 교류하고, 다른 언어로 작성된 콘텐츠를 더 많은 세계인이 볼 수 있게 될 것이다. 한국의 고유한 문학과 지식이 세계에 진출할 수 있다.

한편 디지털 역량, 증강지능 역량이 취약할 경우 국가적인 경쟁력이 떨어질 위험이 있다. 교육, 의료, 법률 등이 시대의 변화를 수용하지 못하고 산업시대에 형성된 전문직의 장벽, 기득권을 고수할 경우 우리나라는 세계의 갈라파고스가 될 수 있다. 언어장벽이 사라지고, 세계인의 교류가 늘어나면서 우리나라 사람이 직접 외국의 서비스를 받을 수 있다. 거대 언어모델 시대에 맞는 사회 변화가 필요하다.

2편은 <SW중심사회> 107호에 게재.



■ 참고문헌

새뮤얼 아브스만 저, 이창희 역. (2014). 지식의 반감기 세상의 변화에는 공식이 존재한다. 책임은수요일.

윤기영. (2018). 한반도 미래전개도-STEPPER 관점 등; KAIST 문술미래전략대학원 미래전략연구센터. (2018). 『카이스트, 통일을 말한다: 한반도 통일 마스터플랜 2048』 김영사.

윤기영. (2023). 미래 트렌드는 무엇인가?: 이규연, 이명호, 윤기영 외. (2023). 『뉴스피앤스 챗GPT』 광문각출판미디어.

이명호. (2023). 전문가·전문직은 사라질 것인가?: 이규연, 이명호, 윤기영 외. (2023). 『뉴스피앤스 챗GPT』 광문각출판미디어.

Anthony, Scott & Gilbert, Clark & Johnson, Mark. (2017). Dual Transformation: How to Reposition Today's Business While Creating the Future. Harvard Business Review Press.

Denett, Daniel. (2020). The Age Of Post-intelligent Design. in Gouvenia Steven ed. (2020). Artificial Intelligence: an Exploration. Vernon Press.

Downes, Larry & Nunes, Paul. (2014). Big Bang Disruption: Strategy in the Age of Devastating Innovation. Portfolio

Navarro, Peter. (2004). If It's Raining in Brazil, Buy Starbucks. McGraw Hill

Rogers, Everett. (2003). Diffusion of Innovations, 5th Edition. Free Press

Rogers, David. (2016). Digital Transformation Playbook. Columbia Business School.

윤기영. (2018.08). 지식사회의 약속은 여전히 유효한가?: 지식사회 2.0에 대한 전망. 미래연구 3(1):

윤기영. (2018.12). 디지털 범용기술의 출현과 디지털 트랜스포메이션의 전개. 미래연구 3(2):

윤기영. (2020.06). 디지털 뉴딜과 예측적 거버넌스. 월간SW중심사회 2020년 6월호.

윤기영. (2020.12.18). 보건의료의 미래는 디지털에 있다. 한겨레.

윤기영. (2023.02.21). 챗GPT는 인류를 위협할까? 이투데이. 2023년 4월 5일 접근. <https://www.etoday.co.kr/news/view/2223725>.

윤기영. (2023.03.15). 거대언어 모델은 어디로 가는가?. 국가미래연구원. 2023년 4월 5일 접근. https://www.ifs.or.kr/bbs/board.php?bo_table=News&wr_id=53287.

윤기영, 김원택, 진정숙, 안주삼, 최경희. (2020.11). 남북한 공진화 영역과 전략: 미래전개도를 중심으로. 미래연구 5(1): 이가용, 형사오, 정종필. (2022). 중국 사회신용체계와 디지털 권위주의: 농민공 임금 미지급 사례를 중심으로. 사회과학연구, 48(3): 113-139.

Ark Invest. (2022). Big Ideas 2022.

Bastian, Matthias. (2023.03.25). GPT-4 has a trillion parameters - Report. the decoder. 2023년 4월 3일 접근. <https://the-decoder.com/gpt-4-has-a-trillion-parameters/>

Bender, Emily M. & Gebru, Timnit & McMillan-Major, Angelina & Shmitchell, Shmargaret. (2021.03). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? FAccT '21: 610-623, doi:<https://doi.org/10.1145/3442188.3445922>

Blank, Steven. (2019.02.01). McKinsey's Three Horizons Model Defined Innovation for Years. Here's Why It No Longer Applies. Harvard Business Review. 2023년 4월 5일 접근. <https://hbr.org/2019/02/mckinseys-three-horizons-model-defined-innovation-for-years-heres-why-it-no-longer-applies>

Dalla-Torre, Hugo & Gonzalez, Liam, Revilla1, Javier Mendoza, et al. (2023.01). The Nucleotide Transformer: Building and Evaluating Robust Foundation Models for Human Genomics. bioRxiv. doi: <https://doi.org/10.1101/2023.01.11.523679>

Future Today Institute. (2023). 2023 Tech. Trend Report.

Glanz, James. (2011.09.08). Google Details, and Defends, Its Use of Electricity. the New York Times.

Glenn, J.C. (2009). Futures wheel. In Futures Research Methodology Version 3.0 (Glenn, J.C. and Gordon, T.J., eds), The Millennium Project

Goldman Sachs. (2023.03.26). The Potentially Large Effects of Artificial Intelligence on Economic Growth.

Hadidi, Rassule & Power, Daniel J. (2020.01). Technology Adoption and Disruption: Organizational Implications for the Future of Work. Journal of the Midwest Association for Information Systems

IFR. (2022). World Robotics 2022.

Jonathan Yerushalmy. (2023.02.17). 'I want to destroy whatever I want': Bing's AI chatbot unsettles US reporter. the Guardian. 2023년 4월 5일 접근. <https://www.theguardian.com/technology/2023/feb/17/i-want-to-destroy-whatever-i-want-bings-ai-chatbot-unsettles-us-reporter>

Lin, Zeming & Akin, Halil & Rao, Roshan, et al. (2022.07). Language models of protein sequences at the scale of evolution enable accurate structure prediction. bioRxiv. doi: <https://doi.org/10.1101/2022.07.20.500902>

Luccioni, Alexandra Sasha & Graphcore, Sylvain Viguiet & Ligozat, Anne-Laure.(2022). Estimating The Carbon Footprint Of Bloom, A 176b Parameter Language Model. arXiv:2211.02001

Ludvigsen, Kasper Groes Albin. (2023.03.02). ChatGPT's Electricity Consumption: ChatGPT may have consumed as much electricity as 175,000 people in January 2023. Toward Data Science. 2023년 4월 5일 접근. <https://towardsdatascience.com/chatgpts-electricity-consumption-7873483feac4>

OpenAI. (2023). GPT-4 Technical Report.

Taori, Rohan & Gulrajani, Ishaan & Zhang, Tianyi, et al. (2023.03.13). Alpaca: A Strong, Replicable Instruction-Following Model. Stanford University. <https://crfm.stanford.edu/2023/03/13/alpaca.html>

Tiseo, Ian. (2023.03.10). Energy consumption of Alphabet (Google) from financial year 2011 to 2020. statista. 2023년 4월 5일 접근. <https://www.statista.com/statistics/788540/energy-consumption-of-google/>

Tompson, Alan. (2023.03.15). Timeline of AI and language models. Life Architect. 2023년 4월 5일 접근. <https://life architect.ai/timeline/>

Voros, J. A. (2001). Primer on Futures Studies, Foresight, and the Use of Scenarios. Prospect. Foresight Bull. 6, 1-8. 2023년 4월 4일 접근 <https://foresightinternational.com.au/wp-content/uploads/2020/03/Voros-Primer-on-FS-2001-Final.pdf>