

AI 파운데이션 모델의 발전과 영국의 경쟁 이슈 검토 동향

Developments of AI Foundation Models and Review of Competition Issues in the UK

설성호 (S.H. Seol, ssh1517@etri.re.kr) 산업분석연구실 책임연구원

ABSTRACT

This paper examines the trends of AI Foundation Model development and the competition to lead the related ecosystem, which have been rapidly unfolding since the emergence of ChatGPT, focusing on big tech companies in the United States. Based on this understanding of background knowledge, I analyzed and presented the main contents of the initial report reviewed by the UK competition authority, CMA, on potential competition issues that may arise in the process of innovations resulting from FM development. In addition, the trend and background of the CMA's investigation into the OpenAI-Microsoft partnership, whose importance has recently been highlighted, were also explained. It is expected that a reasonable domestic policy plan will be established by referring to these UK policy trends and monitoring & analyzing domestic industries.

KEYWORDS AI, CMA, foundation model, GPT, LLM

1. 서론

이미 널리 알려진 바와 같이 GPT-3의 등장(20.05) 이후 대형 기술기업 간에 초거대 AI 모델 구축 선도 경쟁이 활발하게 전개되기 시작하였다. 또한, 이보다 능력이 개선된 GPT-3.5 모델을 기반으로 하는 대화형 챗봇 ChatGPT의 출시(22.11)는 지구촌 곳곳에서 폭발적인 소비자 반응을 불러일으켰다.

ChatGPT 열풍이 폭발적으로 나타나자 인공지능

을 적극적으로 활용하여 디지털 시장의 판도 자체를 바꾸려는 마이크로소프트의 공세(GPT-4 모델 공개(23.03) 등)와 디지털 시장 지위를 방어하고 반격을 가하려는 구글의 대응(PaLM 2 모델 공개(23.05) 등)가 한동안 이어졌다. 이어 메타는 LLaMA 1, 2 모델을 오픈소스 기반으로 공개하는 등 독자적인 길을 모색하였고, 국내 기업 네이버는 하이퍼클로바 X 모델을 공개(23.08)하는 등 다양한 AI 파운데이션 모델(Foundation Model)들이 많이 등장하였다.

* DOI: <https://doi.org/10.22648/ETRI.2024.J.390206>



본 저작물은 공공누리 제4유형
출처표시+상업적이용금지+변경금지 조건에 따라 이용할 수 있습니다.

© 2024 한국전자통신연구원

여기서 파운데이션 모델이란 방대한 양의 데이터로 학습되어 다양한 분야의 과업 및 운영에 적용될 수 있는 AI 기술의 한 유형을 말한다[1]. 즉, 파운데이션 모델이란 방대한 데이터로 사전학습된(Pre-trained) 트랜스포머 아키텍처 기반의 대규모 언어모델(LLM) 또는 멀티모달화 된 모델을 의미하며 각 도메인에 특화된 데이터로 추가 학습하는 미세조정(Fine-tuning)을 거치면 다양한 다운스트림 과업에 활용될 수 있다.

파운데이션 모델의 주요 특징은 초거대 규모의 신경망 구조를 가지는 것과 함께 하나의 모델을 통해 다양한 능력과 기능을 발휘할 수 있는 범용성에 있다. 예를 들어, 최대 1,750억 개의 파라미터를 가진 GPT-3는 신경망의 규모를 키워 모델의 성능, 특히 퓨샷러닝을 획기적으로 개선한 연구였다[2]. ChatGPT는 채팅 형식을 통해 사용자 친화적으로 질의응답, 번역, 요약, 오픈 도메인 질문 응답, 감정 분석, 테이블/텍스트 변환, 코딩 등의 작업을 수행할 수 있다[3].

GPT-3 이후 등장한 파운데이션 모델들은 그 규모가 점점 더 커지고 있으며 구현을 위해서는 엄청난 양의 컴퓨팅 자원을 필요로 한다. 예를 들어, ChatGPT는 엔비디아 GPU인 A100 칩이 1만 개 정도 사용된 것으로 알려지고 있으며 개발자들은 마이크로소프트의 클라우드 서비스인 Azure OpenAI service를 통해 파운데이션 모델(GPT-3.5 또는 GPT-4)에 접속할 수 있다.

이와같이 파운데이션 모델이 가진 다양한 능력과 기능들은 기존 제품 및 서비스에 활용되거나 또는 새로운 제품 및 서비스 창출에 기여함으로써 여러 다운스트림 시장(예: 마케팅, 금융, 교육 등)의 경쟁에 큰 영향을 미칠 수 있다. 또한, 파운데이션 모델의 개발 및 구현에는 막대한 컴퓨팅 인프라 자원과 비용이 요구되므로 자칫 소수 기업에 의해 그 시장

(즉, 업스트림 시장)이 독과점화될 가능성이 있고, 이는 여러 다운스트림 시장의 경쟁에도 악영향을 미칠 수 있다.

따라서 이러한 잠재적 가능성을 미리 파악하기 위하여 영국의 CMA(기관 성격이 우리나라 공정거래위원회에 해당)는 AI 파운데이션 모델을 둘러싼 여러 경쟁 이슈들과 소비자 보호 문제들에 대하여 선도적으로 검토하여 초기 보고서를 발간(23.09)하였다. 또한, 경쟁 이슈의 하나로써 그 중요성이 최근에 부각되고 있는 오픈AI-마이크로소프트 간의 전략적 파트너십 관계에 대한 조사에 착수하였다.¹⁾

이와 같은 상황에서 본고는 국내의 관련 산업발전과 경쟁정책 수립에 도움을 주려는 의도로, AI 파운데이션 모델을 둘러싼 경쟁 이슈들에 대한 영국 CMA의 검토 내용과 오픈AI-마이크로소프트 간의 관계에 대한 조사 착수 배경에 대하여 서술한다. 이를 서술하기에 앞서 독자들의 이해와 효과적인 설명을 위하여 먼저 AI 파운데이션 모델들의 최근 발전 동향 및 주요 특징에 대해 여러 문헌 자료에 기반하여 미국 빅테크 중심으로 정리한다(참고로 다음에서 사용되는 FM이란 약어는 파운데이션 모델을 의미한다).²⁾

II. FM들의 발전 동향과 주요 특징

1. 산업 측면

먼저 AI 파운데이션 모델 산업은 유재홍 등(23.11)이 제시한 바와 같이 AI반도체-클라우드-파운데이션 모델-AI 애플리케이션 서비스의 가치사

1) 최근 들어 미국의 반독점기관들(즉, 법무부 및 연방거래위원회)도 오픈AI-마이크로소프트 간의 관계에 대해 조사하려는 유사한 움직임을 보이고 있음

2) FM 개발 경쟁이 상대적으로 활발하지 않은 영국의 입장에서는 미국 빅테크 중심으로 고찰하는 것이 적절하며, CMA 검토 보고서와의 정합성 유지를 위해 본고는 미국 빅테크 중심으로 정리함. 국내 기업의 FM 개발 경쟁은 유재홍 등(23.11)을 참고

슬 속에서 그 위치를 파악할 수 있다(표 1 참고)[4]. 파운데이션 모델 산업의 입장에서는 클라우드가 업스트림에 해당하고 AI 애플리케이션 서비스는 다운스트림에 해당한다(단, 파운데이션 모델 및 AI 애플리케이션 서비스의 경쟁 이슈들에 초점을 맞춘 영국 CMA 보고서는 파운데이션 모델, 클라우드, 데이터 부문을 합쳐 업스트림으로 본다).

AI 파운데이션 모델은 모델이 공개되지 않는 폐쇄소스와 기본 코드, 아키텍처, 학습데이터 등이 공개되는 오픈소스로 구분할 수 있다.³⁾ 폐쇄소스 모델은 다시 오픈AI-마이크로소프트 연합군과 구글이 선도하고 있으며, 오픈소스 모델은 메타와 IBM이 이끌고 있다. 크게 3파전 양상이라는 평가가 나오고 있다.⁴⁾

먼저 오픈AI가 개발한 파운데이션 모델은 GPT-3(20.05), GPT-3.5(22.03), GPT-4(23.03) 순으로 발전해왔다. ChatGPT의 대성공으로 자신감을 갖게 된 마이크로소프트는 New Bing 공개(23.02), ChatGPT Plus 구독서비스 개시(23.02), MS 365 코파일럿 공개(23.03), ChatGPT 플러그인(Plug-in) 출시(23.05)를 거쳐 최근에는 GPT스토어를 개설하였다(표 2 참고). 즉, 기존 디지털 서비스에 AI를 접목한 AI 애플리케이션 서비스들을 새롭게 제공하고 있으며 써드파티들을 자사 생태계에 포함시켜 확장을 시도하고 있다.

이에 맞서 구글이 개발한 파운데이션 모델은 LaMDA(22.01), PaLM(22.04), PaLM 2(23.05), Gemini(23.12) 순으로 공개되어 왔다. 2023년 2월에

3) 일부 오픈소스 모델들은 파라미터(즉, 가중치와 바이어스)까지 공개되어 사전학습 없이도 파운데이션 모델을 그대로 사용하거나 미세조정할 수 있음
 4) 이외에도 폐쇄소스로 접근하고 있는 주요 기업으로는 Claude 모델을 개발한 앤트로픽과 ACT-1을 개발한 Adept 등이 있고, 오픈소스로 접근하고 있는 주요 기업으로는 이미지 생성 모델인 Stable Diffusion을 개발한 Stability AI, Falcon 모델을 개발한 아랍에미리트 TII 연구소 등이 주목할 만하지만 본고에서는 지면상의 한계로 이들을 다루지 않음

표 1 AI 가치사슬

| 구분 | 주요 특징 | 주요 기업 |
|---------------|---|---|
| AI 애플리케이션 서비스 | 파운데이션 모델을 활용해 만든 서비스로 언어, 이미지 생성 등의 기능을 활용해 초기 생태계가 성장 중 | • 해외: Jasper(마케팅 문구), Midjourney(이미지 생성) 등 • 국내: 뽀빠(작문 보조), 잡브레인(자소서 생성), 클라썸(사내업무용 챗봇) 등 |
| 파운데이션 모델 | 대규모 데이터로 학습되어 다양한 과업에 적용될 수 있는 AI 모델. 개발 주체의 정책에 따라 비공개(폐쇄) 또는 공개(오픈) | • 폐쇄: 오픈AI GPT-4, 구글 PALM 2, 네이버 하이버클로바X 등 • 오픈: Stability AI의 Stable Diffusion, 메타 LLaMA 2, 허깅페이스 BLOOM 등 |
| 클라우드 | 인공지능 데이터의 수집, 저장, 학습, 서비스 개발 지원 등의 기능 수행 | • 해외: 아마존 AWS, MS Azure, 구글 GCP 등 • 국내: 네이버 클라우드, KT 클라우드 등 |
| AI 반도체 | AI 모델 연산에 최적화된 반도체로 고성능, 저전력을 지향 | • 해외: 엔비디아, AMD, 인텔, 구글(TPU), 그래픽코어, 세레브라스 등 • 국내: 삼성전자, SK하이닉스, 사피온, 리벨리온 등 |

출처 Reproduced with permission from [4]. 공공누리 4유형.

표 2 오픈AI-마이크로소프트의 AI 생태계 확장

| 활동 | 내용 |
|------------------------|--|
| New Bing 공개(23.02) | 기존 검색엔진 Bing에 GPT-3.5(또는 GPT-4) 기반의 Prometheus 모델을 결합하여 챗봇 형태로 업그레이드 |
| ChatGPT Plus 개시(23.02) | ChatGPT 유료 서비스(월20\$)로 무료 버전에 비해 빠른 응답과 새로운 기능에 대한 우선적인 접속이 가능. 디플트 모드 가입자는 GPT-4 기반 서비스를 이용 |
| MS 365 코파일럿 공개(23.03) | 기존 MS 오피스 365에 GPT-4를 탑재하여 엑셀, 파워포인트 등에서 업무 생산성을 향상. 최근에는 유료 버전의 구독서비스로 기업용 MS 365 Copilot 런칭(23.11) 및 개인용 Copilot Pro 런칭(24.01) |
| ChatGPT 플러그인 출시(23.05) | ChatGPT Plus 구독자가 플러그인(Plug-in) 기능을 통해 써드파티 앱들을 이용할 수 있음 |
| GPT Store 개설(24.01) | 특정 분야에 특화된 맞춤형 AI 챗봇을 사고팔 수 있는 거래장터로서 월20\$ 요금으로 GPT 스토어 이용 가능 |

ChatGPT 대항마로 공개된 챗봇 Bard는 LaMDA 기반으로 운영되었으나, 보다 강력한 모델인 PaLM 2를 거쳐 최근에는 Gemini 기반으로 바뀌었다. 구글은 오픈AI-마이크로소프트 연합에 효과적으로 대

응하기 위해 구글 브레인과 딥마인드를 합쳐 구글 딥마인드로 통합(23.04)하였고, 오픈AI의 라이벌 기업인 앤트로픽에 대한 대규모 투자를 발표(23.10)하였다. 또한, Vertex AI 플랫폼에서 지원 가능한 대형언어모델을 확대하는 등 생태계 확장을 지속하고 있고 써드파티와의 협력을 강화하고 있다.

메타(구, 페이스북)는 2023년 2월에 LLaMA 1 모델 시리즈를 발표하였고, 2023년 7월에는 LLaMA 2 모델 시리즈를 각각 발표하였다. LLaMA 1은 온디맨드로 모델 가중치에 대한 정보를 연구자들에게 제공하였고, 이러한 정보 공개는 다른 연구소들도 동참하도록 이끌었다. LLaMA 2는 LLaMA 1과 거의 유사하지만 추가로 미세조정되었으며 대화형 애플리케이션에 최적화되었다. LLaMA 2는 상업적인 이용을 불허했던 LLaMA 1과는 달리 거의 모든 기업 및 이용자가 상업용으로도 이용할 수 있는 권리를 부여했으며, 2023년 9월 기준으로 약 3,200만 건의 다운로드를 기록하였다⁵⁾[5].

IBM은 메타와 함께 2023년 12월에 50여 개의 기업, 정부기관, 대학 등이 참여한 ‘AI 얼라이언스’를 창설하였다. AI 얼라이언스는 규제와 안전 등 6대 분야에 우선 집중해 AI의 안전 및 모델 검증을 위한 도구들을 출시할 예정이다[6]. IBM은 새로운 대형언어모델 Granite를 출시하고 자사의 ‘왓슨x 플랫폼’을 통해 기업용 AI 모델 구축을 위한 모든 구성 요소를 원스톱으로 제공하는 전략으로 나가고 있다[7].

2. 기술 측면

앞서 서론에서 언급한 바와 같이 AI 파운데이션

모델들의 주요 특징의 한 가지는 초거대 규모의 신경망으로 구현된다는 점이다. 표 3은 잘 알려진 몇몇 AI 파운데이션 모델의 학습 데이터, 학습시간, 하드웨어를 요약한 것이다. 예를 들어, 오픈AI가 개발한 GPT-3 175B 모델(즉, 1,750억 개 파라미터)⁶⁾[8]은 3,000억 개의 토큰을 이용하여 34일간의 학습으로 만들어졌으며, 이를 가능하게 하기 위하여 1,024개의 엔비디아 A100 GPU가 사용되었다는 의미이다.

그동안 AI 커뮤니티는 모델이 잘 학습된다면, 파라미터 개수가 모델의 능력에 대한 대리치라는 것을 광범위하게 확인하였다. 그러나, 이와 같은 스케일링 법칙이 향후에도 지속될 것인지는 불확실하다. 마이크로소프트 연구자들은 매우 특수하고 엄선된 데이터로 학습된 소규모 언어모델이 대규모 모델과 잘 경쟁할 수 있다는 것을 보여주었으며 소규모 모델을 효율적으로 만들 수 있는 다양한 방법들도 존재하고 있다[1,5].

한편, 오픈AI는 GPT-3까지는 AI 연구자들에게 유용한 정보를 제공하는 보고서를 발간하였으나,

표 3 FM들의 규모, 학습데이터/시간, 하드웨어

| 모델 | 파라미터 | 학습 데이터 (토큰) | 학습시간 (일) | 하드웨어 (GPUs/TPUs) |
|-----------------------|------|-------------|----------|------------------|
| LLaMA (메타) | 65B | 1400B | 21 | 2048 A100 |
| LaMDA (구글) | 137B | 2801B | 57.5 | 1024 TPU v3 |
| GPT-3 (오픈AI) | 175B | 300B | 34 (추정) | 1024 A100 (추정) |
| MT-NLG (마이크로소프트/엔비디아) | 530B | 270B | 90 | 4480 A100 |

출처 Reproduced from [1]. Open Government Licence.

5) 메타의 약관에 따르면 상업용 애플리케이션이 7억 명 이상의 가입자를 보유하지 않는 한, LLaMA 2 모델을 이용할 수 있음. LLaMA 2는 파라미터 개수에 따라 3가지 모델(7B, 13B, 70B)이 제공되고 있음

6) GPT-3 모델은 총 8개가 있으며 가장 큰 모델은 1,750억 개의 파라미터로 구성됨. 규모가 언급되지 않는 GPT-3는 통상적으로 GPT-3(175B)를 의미함

GPT-4 기술보고서는 형식적인 정보만을 담고 있어 AI 연구를 산업화하겠다는 확실한 의지를 보여주었다. 구글의 PaLM 2 기술보고서도 이와 유사하게 유용한 정보를 포함하지 않았으며 엔트로픽은 Claude 모델에 대한 기술보고서를 공개하지 않았다[5].

비교적 최근까지는 오픈AI의 GPT-4가 현존하는 최고성능 인공지능 모델로 평가되었으나, 구글이 2023년 12월에 새로 공개한 제미니(Gemini)로 인해 최고성능 모델이 무엇인지가 다소 애매해졌다. GPT-4와 제미니는 텍스트 기반의 대규모 언어모델을 넘어서서 멀티모달 모델에 해당한다. 모델의 성능을 평가하는 방법에는 여러 가지가 있지만, 최근에는 다중작업 언어 이해(MMLU)가 가장 많이 주목받고 있다. MMLU는 수학, 역사, 법률 등 57가지의 다중작업 성능을 측정하는 것으로서 GPT-4와 Gemini Ultra⁷⁾의 MMLU 점수는 거의 비슷한 것으로 알려지고 있다.

성능이 매우 뛰어난 파운데이션 모델들의 챗 애플리케이션 서비스들(오픈AI ChatGPT, 메타 LLaMA-2-chat, 구글 Bard 등)은 인간피드백강화학습(RLHF)을 채택하고 있는 것이 특징적이다. ChatGPT의 기반 모델인 GPT-3.5의 경우 약 40명의 전문가가 직접 피드백을 통해 강화학습을 시행한 것으로 알려지고 있다[3]. 인간피드백강화학습은 AI를 안전하고 유용하게 만드는 데 있어 매우 유용하지만, 기술적인 난이도가 높고 비용도 많이 소요된다는 단점이 있다.

이외에도 대규모언어모델의 한계를 극복하기 위한 연구들과 학습데이터 부족 문제를 해소하기 위한 연구들이 꾸준히 진행되고 있다. LLM의 능력을 제한하는 컨텍스트 길이(Context Length)를 늘리기 위

한 혁신적인 방법들이 나타나고 있고, 메모리 요구 급증을 극복하기 위한 경량화 기법들(예를 들면, 양자화, 프루닝 등)과 환각(Hallucination)을 줄이기 위한 기술연구들도 활발하다. 합성데이터(Synthetic Data)는 점점 유용해지고 있지만 아직까지는 이를 이용하는 데 따른 리스크(예를 들면, 모델 붕괴 등)도 상당한 것으로 알려지고 있다.

III. 영국의 FM 경쟁 이슈 검토

영국 CMA는 AI 파운데이션 모델과 관련된 경쟁 이슈 및 소비자 보호 문제들을 검토하기 위하여 다양한 FM 개발자, FM 배포 기업, 소비자 및 산업 조직, 학계를 포함한 총 70개 이상의 이해관계자 참여와 함께 최신 AI 연구 등 공개적으로 이용 가능한 정보를 종합 리뷰하여 초기 보고서를 발간하였다(최종보고서 발간은 2024년 초반 예정). 그 보고서에서는 경쟁 이슈 검토에 있어 업스트림과 다운스트림으로 나누어 분석하였는데, 그 결과를 정리하면 다음과 같다[1].

1. 업스트림 경쟁에 대한 검토

먼저 업스트림의 FM 개발 경쟁에서는 (1) 데이터에 대한 액세스, (2) 컴퓨팅 파워에 대한 요구사항 및 액세스, (3) 대형 기술기업 및 선발사업자가 다른 기업들보다 이점이 있는지 여부, (4) 경쟁력 있는 오픈소스 모델 존재 여부의 4가지가 경쟁적인 시장결과가 나타날지 또는 우려되는 시장결과가 나타날지에 대해 영향을 미친다.

데이터에 대한 액세스는 특히 막대한 데이터가 요구되는 사전학습(Pre-training) 단계에서 문제가 될 수 있다. 현행 추세에 따르면 사전학습을 위해 공개적으로 이용 가능한 데이터가 향후 몇 년 안에 고갈

7) Gemini는 성능과 크기에 따라 Ultra, Pro, Nano의 3가지 모델이 있음

될 수 있다고 한다. 이 경우에 독립적인 FM 개발기업들은 소유자가 있는 데이터(Proprietary Data)에 대한 의존도가 높아져서 개발 비용이 높아질 위험성이 있다(사전학습 과정에서 공개적으로 이용 가능한 데이터 외에도 소유자가 있는 데이터를 추가적으로 이용하는 경향은 이미 나타나고 있다).⁸⁾

만약 독립적인 FM 개발기업들이 FM 개발에 필요한 데이터를 합리적인 조건으로 액세스할 수 있어서 다이내믹한 경쟁이 향후에도 지속된다면 좋은 시장 결과가 나타날 것이고, 다른 디지털 시장에서 우세한 포지션에 있는 기존 기업들이 경쟁 FM 개발기업들의 데이터 액세스를 과도하게 제한한다면 우려되는 시장 결과가 나타날 것이다. 소유자가 있는 데이터에 대한 대안으로 검토되는 합성데이터는 학습비용을 절감시킬 수 있지만, 모델 붕괴와 같은 성능저하 문제를 야기할 위험성이 있다. 이 분야에서는 관련 연구가 지속되고 있지만, 합성데이터가 FM 개발을 위한 데이터 소스로서 어느 정도나 실행 가능한지는 상당히 불확실하다.

컴퓨팅 파워에 대한 요구사항 및 액세스는 스케일링 법칙에 따라 AI 모델이 앞으로도 계속 커질 것인가 여부와 상관이 있다. 만약 FM이 성능 개선을 위해 앞으로도 계속 더 커진다면 이것은 진입장벽을 높이고 FM 개발기업 간의 경쟁을 감소시킬 수 있다. 사전학습에는 엄청난 규모의 컴퓨팅 파워가 필요하다. 데이터센터를 보유하고 있지 않은 대다수 FM 개발기업들은 대형 클라우드 서비스 프로바이더(CSP)와의 계약 또는 파트너십에 의존해야만 한다.

이와 관련된 사항으로 Ofcom이 최근에 수행한

클라우드 서비스 시장 조사에 따르면, 영국에는 2개의 리딩 CSP(아마존 AWS, 마이크로소프트 Azure)가 있고 이들의 합산 점유율은 2022년도 기준으로 70~80% 수준인 것으로 나타났다. CSP 간의 경쟁은 주로 신규 고객을 끌어들이는 데만 집중되고 있고 고객들의 사업자 전환 및 멀티-클라우드화는 어렵기 때문에 경쟁 수준이 제한적인 것으로 Ofcom은 평가하였다[9].

그러나, 특정 애플리케이션에 대하여는 소규모의 저렴한 모델이 최첨단의 대형 모델보다 더 적합할 수가 있고 모델을 더 작고 효율적으로 만들 수 있는 방법에 대한 연구들도 진행되고 있어 진입장벽이 낮아질 가능성이 있다. 정리하자면 만약 모든 FM 개발기업이 과도한 제한 없이 공정한 상업적 약관으로 컴퓨팅 파워에 액세스할 수만 있다면 시장에서는 최신 능력을 갖춘 프론티어 모델, 일상적인 과업을 수행하는 모델을 포함하여 다양한 FM이 지속될 수 있을 것이다.

대형 기술기업 및 선발사업자가 대량의 데이터 및 인프라에 대한 액세스로 인해 소규모의 독립적인 FM 개발기업에 비해 누리는 이점의 정도는 불확실하다. 잠재적인 이점들은 선발사업자로 인한 이점, 펀딩 및 기술 지식상의 이점, 범위의 경제, 그리고 데이터 피드백 루프 효과로 인해 발생한다(표 4 참고). 그러나, 선발사업자 이점은 시행착오(예를 들어, 지식재산권 위반 등) 발생, 오픈소스 FM으로 인한 후발사업자의 빠른 캐치업 등으로 인해 상쇄될 수가 있고 범위의 경제 및 규모의 경제, 데이터 피드백 루프 효과는 그 크기가 상당한지가 확실하지 않다.

정리하자면 비록 선발사업자 이점, 범위의 경제 및 규모의 경제, 데이터 피드백 루프 효과가 존재하더라도 만약 신규 진입 기업이 성공적인 FM 개발기업에게 충분한 경쟁제약(Competitive Constraints)을 줄 수 있다면 장기적으로는 좋은 시장 결과가 나타날

8) 메타의 LLaMA 2나 Stability AI의 Stable Diffusion 같은 파운데이션 모델들은 웹 및 기타 공개적으로 이용 가능한 데이터에서 스크랩한 데이터만을 사용해서 사전학습하였음. 그러나, 앞으로는 공개적으로 이용 가능한 데이터만을 이용하여 파운데이션 모델을 개발하는 것이 어려워질 수 있음. 이미 일부 기업들은 모델 훈련을 위해 학술지, 이미지 저장소, 콘텐츠 웹사이트 등의 소스를 사용하기 시작하였음

표 4 대형 기술기업 및 선발사업자의 이점들

| 종류 | 내용 |
|-----------------|--|
| 선발사업자 이점 | 선발사업자 이점은 낮은 입력 투입 비용, 우수한 브랜드 인지 및 고객 로열티, 모델에 대한 실험 및 업데이트에 더 많은 시간을 확보, 견고한 생태계 구축 기회 등이 있음 |
| 편당 및 기술 지식상의 이점 | LLM의 사전학습 및 미세조정에는 많은 비용이 소요되고 특수한 전문지식이 요구됨(소규모 FM 개발기업은 편당 및 기술 지식 확보 측면에서 불리) |
| 범위의 경제 및 규모의 경제 | 높은 FM 개발 비용을 다양한 제품 또는 서비스를 이용하는 많은 고객으로 분산시킬 수 있음 |
| 데이터 피드백 루프 효과 | 다른 디지털 시장에서 획득한 정보 또는 폭넓은 고객 기반으로부터 획득한 정보를 FM 개선에 이용할 수 있음 |

출처 Reproduced from [1]. Open Government Licence.

수 있다. 만약 다른 디지털 시장에서 우세한 위치에 있는 기업이 그 위치를 경쟁 기업을 봉쇄(Foreclose)하는 데 사용하지 않는다면 시장은 긍정적으로 발전하게 될 것이다.

마지막으로 오픈소스 FM 개발 경쟁이 현재까지는 활발하게 지속되고 있으나, 이와 같은 흐름이 향후에도 지속 가능한지는 불확실하다. 오픈소스 모델의 미세조정에서의 발전은 해당 기능을 크게 향상시켰다. 오픈소스로 사전학습된 모델들은 진입 장벽을 낮추고 기업들이 혁신적인 아이디어로 신규 모델을 발전시킬 수 있도록 보장한다. 그러나, 제약이 없는 오픈소스 모델 이용은 안전 우려를 야기할 수 있고, 미세조정된 오픈소스 모델들이 지속적으로 경쟁력을 유지할 수 있을지도 확실하지 않다. 만약 경쟁력 있는, 고품질 오픈소스 모델들이 앞으로도 지속적으로 출시된다면 이는 소비자의 선택을 넓혀 경쟁을 강화시킬 수 있다.

2. 다운스트림 경쟁에 대한 검토

개발된 FM은 (1) 기존 제품 및 서비스의 개선 또는 (2) 신규 제품 및 서비스의 생성을 통해서 다운스

트림의 여러 산업(헬스케어, 금융, 교육, 법률 등)에서 활용될 수 있다.

FM을 기존 프로세스, 제품 및 서비스에 통합시키는 기업들은 생산성과 효율성을 높일 수 있다. 마케팅을 포함한 창조적인 산업들에서 이러한 변화가 일어나기 쉽다. FM은 온라인 검색, 생산성 소프트웨어(Productivity Software)와 같은 분야에서 새로운 기능을 제공함으로써 기존 서비스를 개선하는 데 사용될 수 있다. FM을 이용한 기존 제품 및 서비스의 개선은 소비자에게 좋은 편익을 주지만, 어떤 경우에는 기존 시장 포지션을 강화하거나 또는 약화시킴으로써 경쟁력에 영향을 미칠 수 있다.⁹⁾

FM을 이용하여 생성된 신규 제품 및 서비스로 인해 기존 기업의 인접시장 포지션이 강화될 수도 있지만 반대로 다른 기업들에 의해 기존 기업의 시장 포지션이 약화될 수도 있다. 신규 제품 및 서비스가 효과적으로 경쟁할 수 있는 여러 기업들에 의해 개발된다면 기존 기업의 시장지배력 강화로 이어지는 않을 것이다. 반면, 소유자가 있는 데이터에 대한 특권적인 액세스, 확고한 유통 체계 수립, 통합 생태계의 고객 Lock-in 등 이점 및 시장지배력을 이용할 수 있는 소수의 기업만이 그 신규 제품 및 서비스를 개발하게 된다면 이로 인해 시장이 집중될 가능성이 높아진다.

현재 많은 기업들은 FM을 어떻게 활용할지 검토 중이며 FM이 다운스트림에서 얼마나 중요한 입력이 될 수 있는지를 알기에는 너무나 이른 시점이다. 그러나, FM 활용방식이 무엇이든지에 상관없이 긍정적인 시장 결과로 이어질지 또는 우려스러운 시

9) 구글, 마이크로소프트, 어도비, 슬랙은 생성 AI 기능을 자사의 생산성 소프트웨어 서비스에 통합시킬 계획을 발표하였음. FM이 생산성 소프트웨어 경쟁에 미치는 잠재적인 영향은 불확실하며 여러 요인에 달려있음. 향후 경쟁 다이내믹스가 어떻게 진화하고 현행 시장 포지션이 얼마나 파괴될지는 불분명함

장 결과로 이어질지는 다음과 같은 여러 요인에 달려 있다: (1) 다운스트림 기업들이 다양한 FM에 액세스 가능하고 쉽게 전환할 수 있는가, (2) 이용자가 자신의 니즈에 부합한 최적 제품 및 서비스에 대하여 올바른 정보에 입각한 선택을 할 수 있는가, (3) 수직적 통합이나 전략적 파트너십이 업스트림 및 다운스트림 경쟁자를 봉쇄할 유인이 있는가, (4) 데이터 피드백 효과가 얼마나 중요한가.

(1)과 관련하여 만약 다운스트림 기업들이 선택 가능한 FM 옵션들이 있고 FM 옵션 간에 상호운용할 수 있다면 FM 전환 및 멀티호밍(Multi-homing)이 가능하게 된다. 즉, 전환 및 멀티호밍이 용이하다면 다운스트림 기업들이 하나의 생태계에 묶이지 않을 것이며 긍정적인 시장 결과로 이어지게 될 것이다. 현재까지는 다운스트림 기업들의 전환 및 멀티호밍이 쉬운 편이라는 견해가 제시되었지만, 미래에도 계속 그럴지 여부는 불분명하다.

(2)를 살펴봄에 있어 우선 소비자의 FM 서비스 선택은 서비스가 어떻게 제시되는가(즉, Choice Architecture)와 여러 FM 서비스의 통합 제공 여부 및 고객화된 서비스 제공이 주는 편익에 따라 영향을 받는다는 것을 감안해야 한다. 서비스 선택 아키텍처와 서비스의 통합 제공 여부가 소비자에게 어떤 영향을 미칠지는 불분명하지만, 고객화된 서비스 제공이 주는 편익은 데이터 이동성을 통해 유지될 수 있고, 이로 인해 시장에서 고객들은 특정 생태계에 Lock-in되는 리스크를 피할 수 있다.

(3)과 관련하여 다운스트림 시장을 독점하려는 시도를 통해 얻는 이익이 FM 라이선싱을 통해 얻을 수 있는 이익보다 크다면 업스트림 기업이 다운스트림 시장을 봉쇄할 유인을 갖게 된다. 이미 가치사슬 상에서 여러 기업들이 다양한 형태의 수직적 통합 현상을 보이고 있으며 전략적 파트너십들도 나타나고 있다. 이러한 수직적 통합의 유형과 사례를

표 5 AI 가치사슬에서 수직적 통합 유형과 사례

| 수직적 통합 유형 | 내용 및 사례 |
|------------------------|---|
| FM 및 FM 개발자 도구 공급 | 구글과 오픈AI는 자체 FM을 개발하고 그것을 자사 서비스(Bard, ChatGPT 등)에 이용 중임. 또한, 써드파티에게는 자사 FM(PaLM 2, GPT-4 등)을 제공하며 개발자에게는 FM을 관리하고 배포하기 위한 다양한 도구를 제공함 |
| 플러그인 | 플러그인 프로바이더는 자사 서비스에 이용자 친화적으로 액세스할 수 있게 만들어주는 자연어 인터페이스와 같은 FM 서비스를 추가(Add-on)하기 위해 플러그인 호스트에 의존하게 될 수 있음. 다운스트림에서 플러그인 프로바이더가 플러그인 호스트와 경쟁할 수 있음 |
| FM 서비스 추론을 위한 클라우드 컴퓨팅 | 아마존, 마이크로소프트, 구글 등은 다운스트림 기업들에게 추론을 위한 FM 서비스의 실행을 위하여 클라우드 컴퓨팅을 제공하면서 동시에 FM 서비스 시장에서 다운스트림 기업들과 경쟁함 ¹⁰⁾ |
| 전략적 파트너십 | 마이크로소프트와 오픈AI 간의 파트너십, 구글과 엔트로픽 간의 파트너십 등 |

출처 Reproduced from [1]. Open Government Licence.

정리하면 표 5와 같다.

봉쇄 유인은 부분적으로는 FM 서비스의 수익화 방법에 달렸는데, 아직은 기업들의 수익화 전략이 진화 중이므로 다양한 FM 서비스에 걸쳐 이를 일반화하기가 어렵다. 만약 간접네트워크 효과가 실현된다면 특히 플러그인 호스트가 플러그인 프로바이더를 봉쇄할 수 있는, 시장지배력을 가진 앱스토어 스타일의 플랫폼이 될 가능성이 있다.¹¹⁾ 자기 우대(Self-preferencing), 끼워팔기(Tying)와 같은 반경쟁 행위가 발생하지 못하게 하고 수직적인 통합이나 전략적 파트너십으로 인한 편익만이 구현될 수 있게

10) 예를 들어 아마존은 기업용 AI 플랫폼인 Bedrock을 통해 AWS에 저장된 클라이언트 데이터를 연결하여 사전 훈련된 아마존 FM을 물론, 엔트로픽, Stability AI와 같은 다른 기업들의 FM들을 미세조정함

11) 특정 플러그인 호스트(예를 들어, ChatGPT)가 제공할 수 있는 플러그인이 많을수록 더 많은 고객이 해당 호스트로 유인될 수 있고 그것은 더 많은 플러그인 프로바이더가 그 플러그인 호스트에 대한 플러그인을 더 많이 생성하도록 유도하는 양면 네트워크 효과가 나타날 수 있음. 이로 인해 시장이 특정 플러그인 호스트쪽으로 쏠리는 현상이 발생할 수 있음

하는 것이 바람직한 시장 결과를 가져올 수 있다.

수직적으로 통합된 경우가 그렇지 않은 경우에 비해 더 강한 데이터 피드백으로 인한 혜택을 누릴 것으로 생각되나 이러한 효과의 크기는 FM이 사용되는 특정 상황에 달렸으며 다운스트림 데이터의 실시간 자동 피드백 여부를 단정할 수 없기 때문에 데이터 피드백 효과가 큰지 여부에 대한 결론 도출은 아직은 시기 상조이다. 만약 시장들이 개방되고 경쟁적이어서 FM 개발자들이 데이터 피드백 효과나 선발사업자 이점을 약화시키는 경쟁 제약에 직면하게 된다면, 시장은 보다 긍정적인 방향으로 발전할 것이다.

IV. 오픈AI-MS 파트너십 조사

그동안 마이크로소프트는 장기적인 파트너십을 맺어 오픈AI에 다년간 수십억 달러를 투자(총 130억 \$로 추정됨)하고 AI 슈퍼컴퓨팅과 연구 전반에 걸쳐 협력하기로 합의한 것으로 알려지고 있다(표 6 참고) [10-13]. 그러나, 양사의 파트너십 약관이 공개되지 않아 영국 경쟁 당국 CMA는 이들의 관계가 실질적으로 인수합병에 해당하는지 여부를 조사하기 시작하였다.

CMA는 2023년 12월 8일에 이해관계자들에게 코멘트로의 초대(ITC)를 요청하였다. 이 코멘트로의 초대는 최근 발전상황을 포함하여 오픈AI-마이크로소프트 간의 파트너십이 2002년 기업법의 합병 조항에 따라 심사대상 합병(Relevant Merger Situation)¹²⁾이 발생했는지 여부와 만약 심사대상 합병이 발생했다면 그 합병 효과가 영국의 경쟁에 미치는 영향에 관하여 이해관계자가 자신의 의견과 관련 증거를 제시할 수 있는 기회를 얻게 되는 것들의

표 6 오픈AI-MS 간의 파트너십 확대 동향

| 연도 | 내용 |
|--------------|--|
| 2019 | 비영리 기업이었던 오픈AI는 제한된 영리법인(Capped Profit Company)인 '오픈AI 글로벌'을 자회사로 보유한 모기업으로 지배구조가 변경('19.03)되었고 마이크로소프트는 오픈AI(정확히는 오픈AI 글로벌)에게 10억 \$를 투자하기로 하고 다년간의 전략적 파트너십을 체결함('19.07). 전략적 파트너십에는 다음 사항들이 포함됨: (i) 마이크로소프트와 오픈AI는 새로운 애저 AI 슈퍼컴퓨팅 기술을 공동으로 구축하고, (ii) 오픈AI는 마이크로소프트 애저에서 실행되도록 서비스를 포팅할 예정이고 이를 통해 새로운 AI 기술을 개발하고 일반인공지능(AGI) 약속을 이행하며, (iii) 마이크로소프트는 새로운 AI 기술 상용화에 있어 오픈AI의 우선적인 파트너가 됨 |
| 2019~2023 초반 | 20억\$를 추가로 투자 |
| 2023 | 이번 합의('23.01)를 통해 AI 슈퍼컴퓨팅과 연구 전반에 걸쳐 진행 중이던 협력이 더욱 확대되고 그 결과로 생성된 고급 AI 기술을 각자가 독립적으로 상용화할 수 있게 됨: (i) 마이크로소프트는 오픈AI의 획기적인 독립적인 연구를 가속화하기 위해 특수 슈퍼컴퓨팅의 개발 및 배포에 대한 투자를 늘리고, (ii) 마이크로소프트는 소비자 및 기업용 제품 전반에 오픈AI의 모델들을 배포하고 새로운 범주의 디지털 경험을 도입하며, (iii) 오픈AI의 독점적인 클라우드 프로바이더로서 애저는 연구, 상품 및 API 서비스에 걸쳐 오픈AI의 모든 워크로드를 지원함. 마이크로소프트는 오픈AI에 대한 10억\$ 투자 계획을 확인(구체적인 액수는 밝히지 않음) |

출처 Reproduced from [10-13].

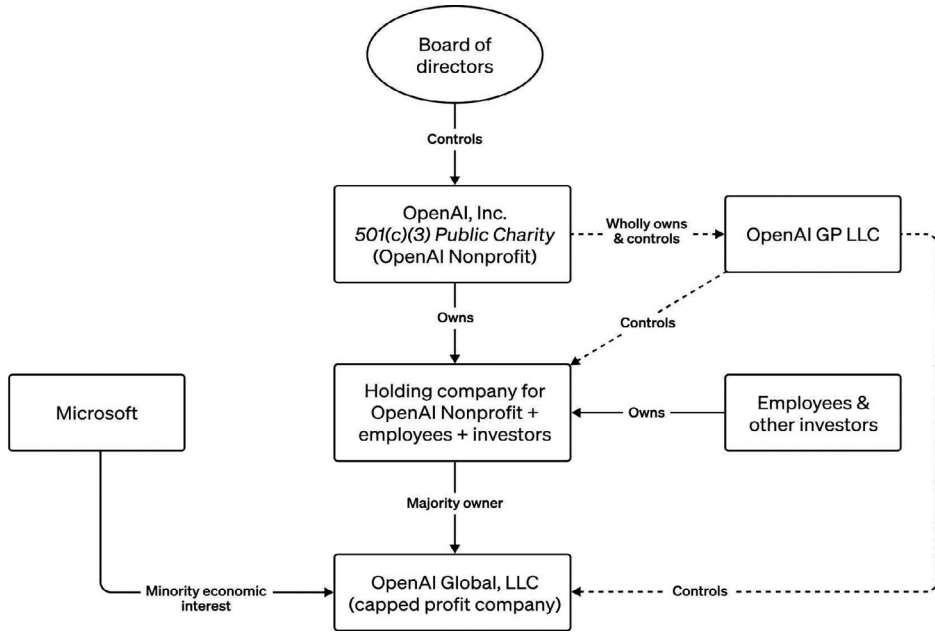
미한다. 이것은 공식적인 1단계 조사(Formal Phase 1 Investigation)에 앞서 영국의 경쟁 당국이 관련 정보를 수집하는 첫 단계에 해당한다[14].

CMA의 ITC가 발표된 지 하루 만에 마이크로소프트와 오픈AI의 관계는 '소수 주주(Minority Owner)'에서 '소수의 경제적 이익(Minority Economic Interest)'을 보유하고 있는 자로 문구가 정정되었다¹³⁾(그림 1 참고). 영국의 경쟁 당국이 양사 간의 잘 이해되지 않는, 불분명한 파트너십을 조명하려고 시도함에 따라 이러한 표현 차이는 매우 중요할 수도 있다[15].

다가오고 있는 조사에 대한 발표는 오픈AI의 최

12) 심사대상 합병은 2002년 기업법 제23조에서 규정

13) 마이크로소프트는 오픈AI 글로벌의 49% 지분을 가지고 있는 것으로 알려지고 있음



출처 Reprinted from [10], © Copyright Open AI.

그림 1 오픈AI의 지배구조

고경영자인 샘 알트만이 이사회에 의해 해고되었다가 며칠 뒤 오픈AI 회사직원들과 투자자들(여기에는 마이크로소프트 최고경영자인 사티아 나델라도 포함됨)의 압력으로 복직된 직후에 나왔다. 그러나, 마이크로소프트나 다른 후원 캐피털들은 비영리법인인 오픈AI의 지분을 소유하고 있지는 않으며 대신에 오픈AI 자회사(오픈AI 글로벌)로부터 특정 한도까지 이익 일부를 받을 자격이 있을 뿐이다.

즉, 오픈AI의 주요 후원기업인 마이크로소프트는 투표권이 없는 관찰자 지위(Non-voting Observer Position)만을 가지고 있다. 그러므로 마이크로소프트 대리인이 오픈AI 이사회에 참석하고 기밀정보를 액세스할 수는 있지만 이사 선출을 포함한 문제에 대해서는 투표할 수가 없다[16]. 더구나 이 지위도 샘 알트만의 복귀 이후인 최근 시점(23.11)에 획득한 것이다.

CMA의 조사 착수 배경에 대하여 런던에 있는 로펌기업 Fladgate의 파트너인 알렉스 하프너는 “여기서 흥미로운 점은 CMA가 ‘최근 발전상황을 포함하여(Including Recent Developments) 오픈AI-마이크로소프트 간의 파트너십’이라는 표현을 사용하고 있다.”라는 것이며, 이는 “본질적으로 CMA는 파트너십에 어떤 변화가 있었는지를 알아보고 더 넓은 맥락에서 AI 부문에 대한 규제방안을 검토 중인 CMA와 영국 정부가 이에 대해 더 자세히 조사할 가치가 있는지 여부를 알아보고 싶어 한다.”라고 말했다[17].

V. 결론

오픈AI에 의해 공개된 ChatGPT에 대한 폭발적인 소비자 반응은 혁신 경쟁에 대한 촉매제가 되어 여러 기업은 더 우수한 파운데이션 모델(FM)과 관련

서비스의 개발에 나서고 있다. 미국의 대형 기술 기업들 간에는 FM 개발 및 관련 생태계 확장을 위한 선도 경쟁이 치열하게 전개되고 있으며, 여타 국가의 많은 기술기업들도 이를 맹렬히 추격 중이다. 본고는 여러 문헌자료에 기초하여 이러한 혁신 경쟁의 전개 양상과 기술 발전 추세를 보다 단순화하여 미국 빅테크 중심으로 정리하였다.

한편, 영국의 경쟁 당국인 CMA는 FM 발전에 따른 기업 간의 혁신 경쟁이 진행되는 과정에서 잠재적으로 발생할 수 있는 경쟁 이슈 및 소비자 보호 문제를 검토하여 초기 보고서를 최근에 발간하였다. 이 보고서는 다이내믹한 경쟁이 바람직한 시장결과를 도출한다는 경제학적 관점에서 업스트림과 다운스트림의 경쟁에 영향을 미치는 요소들과 각각의 요소들에 담긴 불확실성에 대하여 검토한 바를 정리하였다. 그리고 이 검토 결과를 바탕으로 FM 개발 및 전개에 있어 경쟁 촉진을 위한 중요한 원칙들(예를 들어, 데이터/컴퓨팅 자원 등에 대한 과도한 제한이 없는 액세스 보장, FM 선택에 대한 다양한 형태의 옵션 보장, 수직적 통합에서 발생할 수 있는 반경쟁행위의 방지 등)을 제시하였다.

본고는 또한, CMA 보고서가 검토한 여러 경쟁 이슈 중의 하나로서 특히 최근에 중요하게 부각되고 있는 전략적 파트너십에 대한 대표사례에 해당하는 오픈AI-마이크로소프트 파트너십에 대한 CMA의 조사 착수 움직임과 그 배경에 대하여도 설명하였다.

FM 발전에 따라 발생 가능한 다양한 경쟁 이슈에 대한 영국 CMA의 선제적 검토와 오픈AI-마이크로소프트 파트너십에 대한 조사 착수는 이와 관련된 국내 정책 마련에 있어 좋은 자료로서 참고할 가치가 있다고 생각된다. 이러한 정책 동향을 참고하고 FM 관련 국내 산업에 대한 모니터링 및 분석을 통해 합리적인 국내 정책방안을 마련할 수 있을 것으

로 기대된다.

약어 정리

| | |
|-------|--|
| AI | Artificial Intelligence |
| CMA | Competition & Market Authority |
| CSP | Cloud Service Provider |
| FM | Foundation Model |
| GPT | Generative Pre-trained Transformer |
| GPU | Graphic Processing Unit |
| ITC | Invitation To Comment |
| LLaMA | Large Language Model Meta AI |
| LLM | Large Language Model |
| MMLU | Massive Multitask Language Understanding |
| PaLM | Pathways Language Model |
| RLHF | Reinforcement Learning from Human Feedback |
| TPU | Tensor Processing Unit |

참고문헌

- [1] CMA, AI Foundation Models: Initial Report, Sept. 2023.
- [2] 임준호, "GPT-3 소개(부제: 이게 된다고?)," AI 프렌즈 세미나 #51, 2020. 8. 12.
- [3] 고우영, "ChatGPT 톨아보기," AI 프렌즈 세미나 #132, 2023. 2. 15.
- [4] 유재홍 외, "생성 AI 산업 생태계 현황과 과제," SPRI 이슈보고서, 2023. 11.
- [5] N. Benaich and Air Street Capital Team, "State of AI 2023," Oct. 2023.
- [6] 이코리아, "'AI동맹'이 났다, 생성 AI '개방형 vs. 폐쇄형' 경쟁 치열," 2023. 12. 6.
- [7] AI타임즈, "IBM, '경량 LLM' 포함 기업용 '토털 AI 서비스' 라인업 공개," 2023. 9. 8.
- [8] <https://lambdalabs.com/blog/demystifying-gpt-3>
- [9] Ofcom, "Cloud Services Market Study," Oct. 2023.
- [10] <https://openai.com/our-structure>
- [11] <https://news.microsoft.com/>
- [12] Forbes, "Microsoft Confirms Its \$10 Billion Investment Into ChatGPT, Changing How Microsoft Competes With Google, Apple and Other Tech Giants," 2023. 1. 27.
- [13] TechCrunch, "Microsoft Invests Billions More Dollars in OpenAI, Extends Partnership," 2023. 1. 23.

- [14] CMA, "CMA Seeks Views on Microsoft's Partnership with OpenAI," Press Release, 2023. 12. 8.
- [15] Financial Times, "How Microsoft's Multibillion-dollar Alliance with OpenAI Really Works," 2023. 12. 16.
- [16] Reuters, "Microsoft, OpenAI Tie-up Comes under Antitrust Scrutiny," 2023. 12. 9.
- [17] Guardian, "UK Watchdog to Examine Microsoft's Partnership with OpenAI," 2023. 12. 8.