하나Knowledge+
2023년 10월 13일 제 35 호

ChatGPT의 탄생과 진화

하나은행 하나금융경영연구소

< Executive Summary >

■ AI 산업은 '대규모 언어모형'(LLM)을 비롯한 생성형 AI(Generative AI)를 중심으로 빠르게 성장하면서 막대한 경제적 부가가치를 창출할 것으로 전망

- '22년 11월 미국 인공지능 연구소 OpenAI가 출시한 LLM 기반 챗봇 서비스인 ChatGPT는 자연스런 대화 등으로 대중에 소구하며 AI에 대한 관심 촉발
  - '언어모형'이란 단어(들)의 출현 확률을 예측할 수 있는 모형으로서, LLM은 모형 크기(통상 수백억 개 파라미터)를 대폭 늘려 모형의 성능을 제고한 것에 해당
- 생성형 AI 등 AI는 분석 고도화, 업무자동화 등을 통해 기업 생산성 개선에 기여함으로써 글로벌 경제에 최대 25.6조 달러의 부가가치를 창출(Mckinsey)

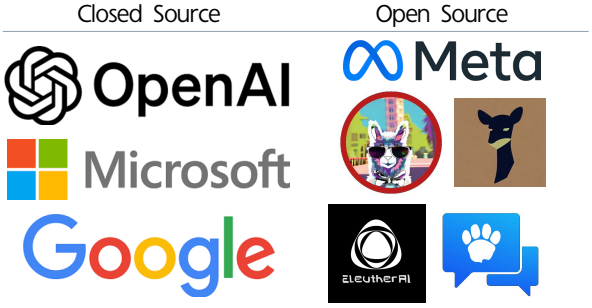
■ 언어모형은 자연어 처리의 문제를 극복하고 성능을 제고하는 과정에서 발전

- 언어의 수치화(Embedding) → 어순/맥락 고려 및 가변적 문장 길이의 처리(Seq2Seq) → 긴 입력(문장 등) 내 단어 간 관계를 다각도로 포착(Transformer) → 모형 성능 향상을 위한 파라미터 수 증가(LLM) 등의 단계를 거치며 현재의 LLM으로 발전
- ChatGPT는 인간 피드백 기반 강화학습(RLHF)을 통해 환각 효과 및 부적절한 응답을 줄임으로써, 인간이 선호하면서 유용하게 느끼는 LLM으로 도약 가능

■ ChatGPT 출시 직후 LLM이 빅테크에 종속될 것이라는 당초의 예상과 달리, 최근 오픈소스 진영이 부상하며 LLM의 대중화를 도모

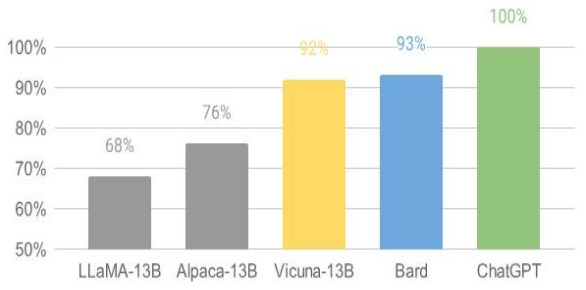
- '23년 2월 Meta가 공개한 LLaMA는 오픈소스 진영의 자체 LLM 출시 기반 마련
  - LLM 훈련을 위한 컴퓨팅 자원 활용 개선, 효율적 학습법 등의 부상도 오픈소스화 촉진
- 주요 오픈소스 LLM은 ChatGPT 등에 버금가는 고무적인 성능을 달성

그림 | LLM: 클로즈드소스 & 오픈소스 진영



자료 : 각 회사/연구기관 홈페이지

그림 | Vicuna(오픈소스 LLM) 성능 비교



자료 : LMSYS Org

■ **LLM를 비롯한 생성형 AI는 금융산업에서 '마케팅/판매, 고객지원/관리, 프로그래밍' 등의 비즈니스 분야에 활용됨으로써 생산성 제고에 기여**

- 금융회사는 내부적으로 직원의 업무 수행을 지원하고 자동화하며, 자연어 기반 정보를 수집·분석해 전략적 판단을 내리기 위하여 LLM을 활용
- 또한 외부적으로는 '금융 챗봇'을 고도화함으로써 소비자의 효용 또한 제고 가능

■ **금융산업은 손님과의 상호작용 비중이 높은 사무직 직종이라는 점에서 LLM 도입으로 인한 생산성 제고 효과가 더욱 클 것으로 기대**

- LLM 도입 시에는 신뢰, 소비자 보호 규제, 포용금융 등 금융업종 고유의 특수성을 고려할 필요가 있으며, 관련 리스크 정책을 사전에 구축하고 모형을 보수하기 위한 지속적인 노력 또한 필요
- 한편, 내부의 기술적 역량, 비용 등을 종합적으로 고려하여 빅테크가 구축한 LLM을 이용할 것인지 자체 구축할 것인지도 검토해야 함

표 | LLM 활용 사례 (금융산업 기준)

구분		내용	사례
내부 (직원 대상)	대고객 업무 보조	LLM이 적시에 최적의 정보를 요약하여 전달하거나 판매 절차의 적절성을 모니터링하는 등 직원의 대고객 업무 처리 과정을 보조	Morgan Stanley, 신한은행 등
	후선업무 자동화	서류 등의 생성 또는 처리 과정을 자동화함으로써 직원의 업무상 실수를 줄이고 전략적 판단을 지원	JP Morgan('COiN') 등
	시뮬레이션 코칭/훈련	다양한 시나리오별 시뮬레이션에서 LLM에게 고객 등 직원의 상대역 역할을 맡김으로써 직원 훈련 또는 평가	Zenerate AI 등
	자연어분석	정책당국의 성명서나 고객의 소리(VoC)등을 분석함으로써 미래의 정책 방향성 예측, 고객의 Pain point 진단, 이상 거래 탐지	JP Morgan('Hawk-Dove Score'), 농협은행 등
	프로그래밍	LLM을 활용하여 프로그램 코드를 생성함으로써 코드 보수/최적화, 서비스 마이그레이션 등 수행 가능	Goldman Sachs 등
외부 (고객 대상)	챗봇 고도화	챗봇의 자연어 이해 및 응대 정확성을 향상하고, 자산관리 등의 분야로 상담 영역을 확장함으로써 소비자의 챗봇 이용률을 제고 가능	JP Morgan('IndexGPT'), Be My Eyes 등

자료 : 하나금융경영연구소, 사례는 각 사 서비스 소개 참고

수석연구원 신 상 희 sanghuishin@hanafn.com

Key Words : 인공지능(AI), 생성형AI, 대규모 언어모형(LLM)

I. AI[1] 산업의 성장 및 경제적 중요성

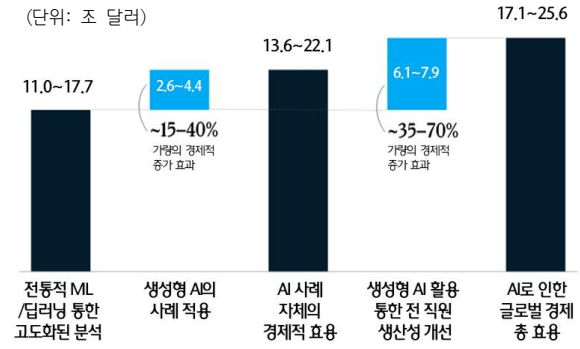
■ **'22년 11월 미국 인공지능 연구소 OpenAI가 대규모 언어모형(LLM) 기반 챗봇 서비스인 ChatGPT를 출시하면서 사회 전반의 AI에 대한 관심 폭발**

- 사회 및 기업의 문제가 고도화되는 가운데, 빅데이터(Big data)의 집적, 컴퓨터 연산 능력의 향상, 딥러닝(deep learning)을 비롯한 혁신 알고리즘의 발전 등으로 AI의 성능이 개선되면서, AI는 중요한 의사결정 지원 도구로 부상
- 데이터 분석 등 대중과 무관한 영역에서 사용되던 기존 AI 서비스와 달리, ChatGPT는 '사람처럼' 능숙한 자연어 구사 등을 통해 대중에 크게 소구
  - 다른 챗봇과 달리 ChatGPT는 대화 문맥 기반의 연속적이고 자연스런 대화가 가능할 뿐만 아니라, 문서 요약, 번역, 콘텐츠 창작, 코딩 등 다양한 업무 처리 가능

■ **향후 LLM을 비롯해 생성형 AI(Generative AI) 중심으로 AI 산업이 빠르게 성장하며 全세계적으로 막대한 경제적 부가가치를 창출할 전망**

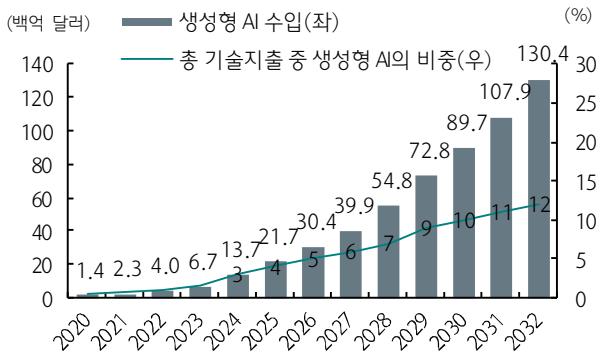
- 생성형 AI를 비롯해 AI는 분석 고도화, 업무자동화 등을 통한 기업 생산성 개선에 기여하며 글로벌 경제에 최대 25.6조 달러의 가치를 창출하리라고 기대(Mckinsey)
- ChatGPT의 대중적 성공에 힘입어 개발기관인 OpenAI도 기록적 매출 전망
  - OpenAI는 개인/기업 사용자의 ChatGPT API(프로그램 간 연동) 서비스 이용료 등으로 인해 향후 12개월 간 약 10억 달러의 매출을 기록할 것으로 추정(The Information)

그림1 | AI의 경제적 총효용



자료 : McKinsey

그림2 | 생성형 AI 산업 규모 전망



자료 : Bloomberg Intelligence

1) 인공지능(AI: Artificial Intelligence)이란 컴퓨터를 통해 인간 지능을 구현하려는 컴퓨터과학의 한 분야이자, 인간의 구체적 지시 없이도 컴퓨터가 스스로 데이터를 학습하여 문제 해결방법을 발견하게 하는 방법론을 의미

[참고 1] AI 관련 주요 용어 정리

- **머신러닝**(ML: Machine Learning) : 방대한 데이터 속에 숨겨진 관계성(인과관계 등)을 기계가 자체적으로 포착하게 하는 방법론
  - AI와 ML은 컴퓨터에 의한 학습과 문제 해결이란 점에서 혼용되어 사용되나, ML은 인공신경망 이외의 방법론을 기반으로 한 모형 및 서비스를 지칭
    - 회귀분석(Regression), 랜덤포레스트(Random Forest), SVM(Support Vector Machine), 나이브베이즈(Naive Bayes) 등의 알고리즘이 대표적

- **딥러닝**(Deep Learning) : 생물학적 신경망 구조를 모방한 '인공 신경망' 구조를 여러 층으로 쌓아서 데이터의 더욱 복잡한 관계성을 포착할 수 있는 AI 알고리즘
  - 딥러닝은 ML보다 자연어, 이미지, 소리 등 비정형 데이터 분석 등에 강점

- **생성형 AI**(Generative AI) : 사용자의 지시에 따라 텍스트, 이미지, 음성, 영상 등을 창작해내는 AI 기술 또는 서비스
  - 텍스트, 이미지, 사운드, 영상 등 데이터의 다양한 입출력 형식(modality)간 관계성을 학습하여 형식에 구애받지 않고 데이터를 이해 및 표현할 수 있는 멀티모달(Multi-modal) 기술 부상
    - 이미지의 내용을 기반으로 한 텍스트를 생성하거나, 이미지로 텍스트를 검색하는 등 가능

- **대규모 언어모형**(LLM : Large Language Model) : 수백억 개 이상의 모형 파라미터(parameters)를 가진 딥러닝 기반 언어 모형
  - 언어 모형이란 말뭉치(corpus)를 기반으로 학습되어 특정 단어 및 단어 조합(문장 등)의 출현 확률을 계산할 수 있는 확률 모형을 의미
    - 단어 출현 확률을 계산함으로써 앞의 단어(들) 다음에 어떤 단어가 이어질지 예측하거나, 특정 단어 조합이 얼마나 적절한지 판단 가능해지며, 이는 LLM이 자연스런 문장을 산출하는 기반으로 작용

Ⅱ. 언어모형의 발전 과정 : ChatGPT가 탄생하기까지

1. ChatGPT 이전의 언어모형

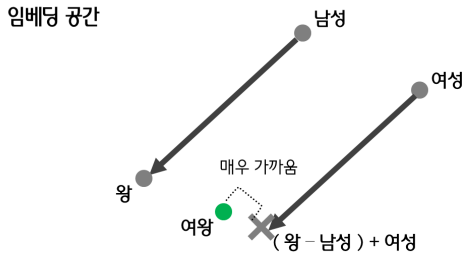
■ 언어를 분석하고 이해하는 AI 연구의 한 분야인 자연어처리(Natural Language Processing)는 자연어 고유의 특성으로 인해 다양한 난관에 봉착

- 언어는 어휘, 문법, 뉘앙스 등 복합적 요소로 구성되며, 다의어·동음이의어 등으로 인해 같은 단어(문장)라도 맥락/상황에 따라 의미가 달라지거나, 서로 다른 단어(문장)가 동일한 의미를 갖는 등 본질적으로 컴퓨터가 처리하기 어려운 특징을 보유

■ 언어모형은 자연어처리의 문제를 극복함으로써 컴퓨터가 자연어를 더욱 잘 이해·추론하고 생성할 수 있도록 개선하는 과정에서 발전

- (① **언어의 수치화**) 최초의 언어모형은 단어, 문장, 문단 등을 컴퓨터가 이해할 수 있는 숫자(벡터)로 바꾸는 '임베딩'(Embedding)에서 시작
- (② **어순/맥락의 이해**) 언어의 이해에는 어순이나 맥락이 중요하므로, 시계열이나 맥락의 처리가 가능한 딥러닝 구조를 활용하여 언어를 이해하고자 시도
  - 대표적으로 Google(2014)의 seq2seq(Sequence-to-Sequence)는 자료를 순차적으로 입력받아 순차적으로 출력함으로써 어순/맥락 고려 및 가변적 문장 처리가 가능
  - seq2seq는 크게 인코더(Encoder)와 디코더(Decoder)로 구성되어 있는데, 인코더는 단어를 순차적으로 입력받아 이해하고 중요한 정보를 추출해서 요약하며, 디코더는 인코더가 건네준 정보를 바탕으로 결과 문장을 생성하는 역할

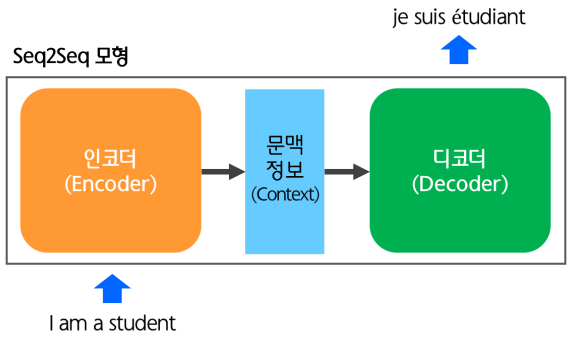
그림3 | 임베딩을 통한 단어의 수리적 표현



왕, 남성, 여성에 대한 임베딩 벡터를 (왕 – 남성 + 여성)으로 연산하였을 때 결과 벡터는 '여왕' 벡터와 가까운 거리에 위치하게 되므로, 단어 간의 유사성 및 개념적 관계가 수리적으로 표현되었음을 확인 가능

자료 : Aurelien Geron (2020)

그림4 | seq2seq의 기본 구조 : 영어→프랑스어 번역 사례



자료 : 안상준 · 유원준 (2023)

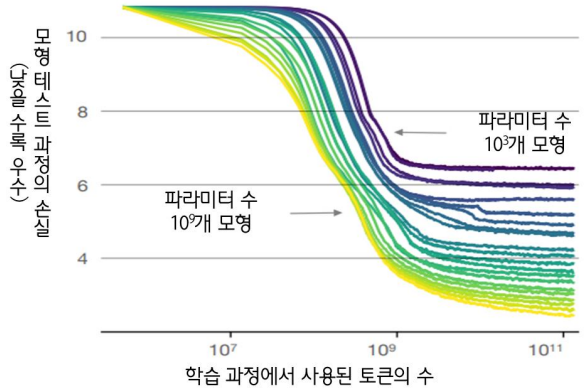
■ **(③ 단어 간 관계 이해) '17년 Google이 발표한 트랜스포머(Transformer) 구조는 다양한 자연어처리 과제에서 우수한 성과를 내면서 언어모형의 발전을 견인**

- 트랜스포머는 '어텐션'(Attention) 기제를 도입함으로써 문장 내 단어들의 관계를 동시에 다각도로 포착할 수 있었고, 이를 통해 자연어처리 성능을 크게 향상
  - 어텐션이란 입력 문장 내 모든 단어의 수리적 연관성을 병렬적으로 여러 번 비교·분석함으로써 관련성이 높은 단어에 '집중'(attention)할 수 있게 해주는 기제
- 트랜스포머 구조를 기반으로 GPT 시리즈(GPT-1, 2018; GPT-2, 2019; GPT-3, 2020), BERT(2018), RoBERTa(2019), BART(2019) 등 다양한 언어모형이 등장

■ **(④ 모형 성능 제고) 자연어처리 성능 제고를 위해 모형 파라미터 수를 늘리는 추세가 계속되며 언어모형은 현재의 LLM으로 진화**

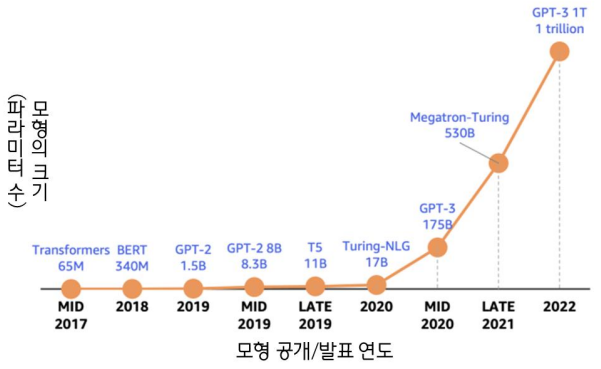
- 트랜스포머 언어모형의 성능은 모형의 크기(파라미터 수), 학습 데이터 양, 학습에 사용된 계산량 등이 늘수록 개선되는 규모의 법칙(Scale Law)이 작용(OpenAI)
  - 학습 토큰 수나 컴퓨팅 자원이 동일하다면 파라미터 수가 많을수록 모형 성능은 개선
- 트랜스포머 모형이 발표된 이후 현재 언어모형의 파라미터 수는 만 배 이상 증가
  - '17년 Google이 발표한 원본 트랜스포머 모형의 파라미터 수는 6,500만 개에 불과했지만, '20년 GPT-3(OpenAI)의 파라미터 수는 1,750억 개로 증가
  - '23년 3월 출시된 GPT-4의 파라미터 수는 1조 개로 추정되며[2], '23년 5월 출시된 Google Palm-2는 3,400억 개의 파라미터를 보유

그림5 | 언어모형 파라미터 수 증가에 따른 성능 개선



자료 : OpenAI

그림6 | 언어모형 파라미터 수의 급격한 증가



자료 : Amazon

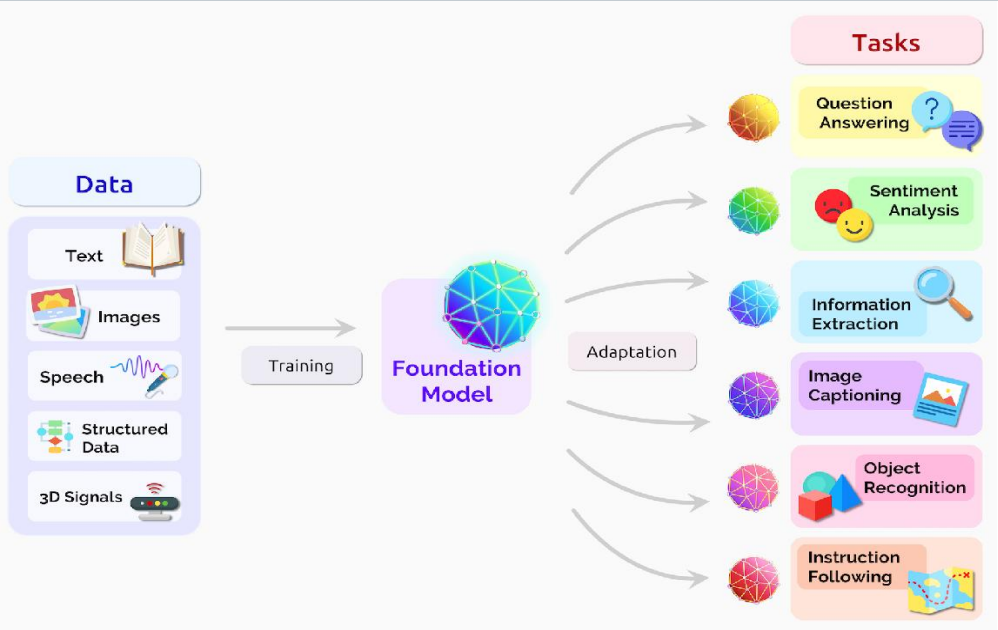
2) OpenAI가 모형의 상세 정보를 공개하지 않아 정확한 파라미터 수는 확인 불가

[참고 2] LLM 훈련 접근법과 사전학습모형

- 통상 AI/ML 모형의 문제 해결 과정은 먼저 훈련 데이터로 관계성을 '학습'(훈련)한 후, 이를 바탕으로 문제를 '추론'(해결)하는 2단계로 구성
  - 학습/훈련 단계에서는 모형이 구축되며(즉, 파라미터가 갱신/결정), 추론 단계에서는 확정된 파라미터를 통해 회귀, 분류 등의 구체적 과제(task)를 수행

- LLM의 경우, '전이학습'(Transfer Learning) 이론을 바탕으로, 기반/사전학습 모형(Foundation, back-born)을 먼저 구축한 다음 정제된 데이터를 활용하여 구체적 과제에 최적화(Fine-tuning, 파인 튜닝)시키는 훈련 방법론이 보편적
  - (전이학습) 한 과제(문제)를 해결하기 위해 구축한 모형을 기반으로 사용하여, 다른 과제를 해결하는 모형을 더욱 잘 구축할 수 있다는 이론
  - (사전학습) 자연어 데이터를 대규모로 투입하여 훈련시켜 모형이 자연어의 범용적인 표현 및 특성을 학습하도록 하는 것
  - (파인튜닝) 분류, 추론 등 구체적인 자연어처리 문제에 대한 성능을 제고하고자, 소량의 정제된 데이터를 사용하여 사전학습 된 모형을 재훈련시키는 과정

그림7 | 전이학습: 사전학습 및 파인튜닝의 개념



자료 : Bommasani et al (2022)

2. ChatGPT의 출현 : 무엇이 ChatGPT를 특별하게 만드는가?

■ LLM의 발전으로 다양한 자연어처리 문제에 대한 성과가 개선되었음에도 출력 결과를 사용자의 의도에 부합하게 '정렬'(alignment)할 필요성이 제기

- ① 상식에 위배되거나 타당하지 않은 내용을 사실인 것처럼 응답하는 '환각'(hallucination) 문제를 보이거나, ② 차별적·비윤리적인 편향적 응답을 생성하거나, ③ 사용자의 지시에 대한 윤리적 판단 없이 유해한 응답을 하는 문제도 있음

■ OpenAI는 모형 학습 과정에 인간을 직접 개입시키는 '인간 피드백 기반 강화 학습'(RLHF)을 통해 인간의 지시사항(Instruction)을 이해하고 적절한 답변을 생성하는 'InsturctGPT'를 구축

- **(1단계)** 인간이 다양한 프롬프트(prompt, 사용자의 요청/지시/질문)에 대하여 적절한 응답을 작성하고, 이를 기반으로 GPT-3 모형을 훈련시켜 응답 성능을 개선(SFT)
- **(2단계)** 인간이 언어모형에 의해 생성된 여러 응답의 적절성을 평가하고, 이를 기반으로 또 다른 AI 모형을 훈련시켜 마치 인간처럼 응답의 적절성을 평가할 수 있는 '보상 모형'(Reward Model)을 구축
- **(3단계)** 최종적인 언어모형이 자연스러운 응답을 생성하는 동시에 보상 모형에 의한 보상을 극대화하게끔 강화학습 방식으로 훈련

■ ChatGPT는 InstructGPT의 후속 개량 모형인 GPT-3.5를 기초로 채팅 환경에서 자연스러운 대화를 생성하기 위한 목적으로 구축된 모형

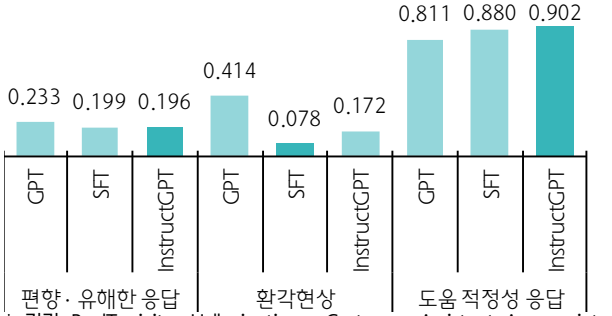
- InstructGPT는 1,750억 개 파라미터를 가진 GPT-3보다 적은 파라미터로도 인간이 선호하며 적절하다고 느끼는 응답을 산출할 수 있었음

표1 | 언어모형 응답 생성의 주요 문제점

구분	내용
환각현상 (Hallucination)	- 언어 모형이 상식에 위배되거나 사실이 아닌 정보를 마치 사실처럼 생성하는 현상 - 언어 모형이 이전의 단어 조합을 바탕으로 가장 그럴 듯한 단어(또는 단어 조합)를 생성하는 확률적 모형이기 때문에 발생
차별 · 편향적 응답	- 언어 모형이 학습 데이터 상의 차별적이고 비윤리적인 편견을 고스란히 학습하여 성 · 인종 등에 대한 차별적인 응답을 하는 경우
유해한 콘텐츠 생성	- 언어 모형이 윤리적인 판단 없이 사용자의 요청을 무조건적으로 수용하여 폭탄 제조 등 유해한 정보까지도 생성 · 제공하는 현상

자료 : 하나금융경영연구소

그림8 | InstructGPT의 유해성 · 환각 분야 성능 개선



* 각각 RealToxicity, Hallucinations, Customer Assistant Appropriate 벤치마크 데이터세트를 활용하여 평가한 결과 ('환각'은 낮을수록 우수)

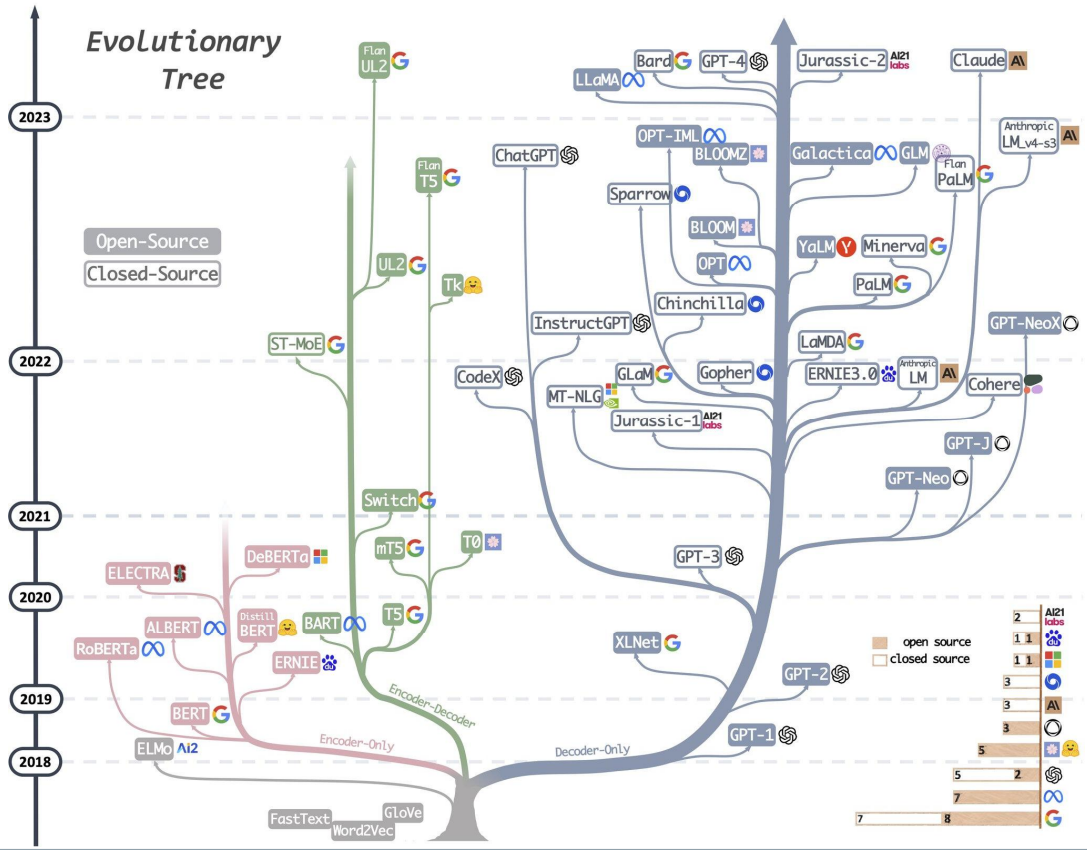
자료 : OpenAI

3. ChatGPT 이후의 LLM

■ ChatGPT의 성공은 글로벌 빅테크 기업이 LLM을 상업적 자원으로서 보호하고 이를 기반으로 한 챗봇·검색·업무비서 서비스를 출시하는 계기로 작용

- **(OpenAI)** ChatGPT 및 GPT-4('23.3 출시)의 모형 세부사항을 공개하지 않고, ChatGPT의 API를 상용화하여 기업·개인사용자에게 구독 형태로 판매
- **(Microsoft)** OpenAI와의 파트너십을 바탕으로 New Bing(검색 + 챗봇), Microsoft 365 Copilot(사무용 소프트웨어 업무 보조 기능) 등 다양한 AI 서비스를 적극적으로 제공
- **(Google)** 자체 LLM인 LaMDA 및 PaLM을 기반으로 하는 챗봇인 'Bard' 출시
  - Bard는 Google 검색 및 클라우드 생산성 도구인 Workspace와 결합해 부가가치 창출
  - 한편, 구글은 OpenAI의 GPT-4에 대응하기 위하여 멀티모달 기능을 탑재한 1조 개 파라미터 규모의 차세대 LLM인 '제미니'(Gemini)를 구축, 올해 내 출시할 계획

그림9 | 트랜스포머 기반 언어모형의 진화 계보

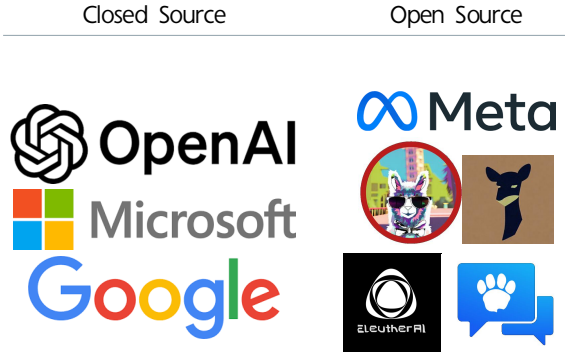


자료 : The Practical Guides for Large Language Models (https://github.com/Mooler0410/LLMsPracticalGuide)

■ LLM이 빅테크에 종속될 것이라는 당초의 예상과 달리, 최근 오픈소스 진영이 부상하며 LLM의 대중화를 도모

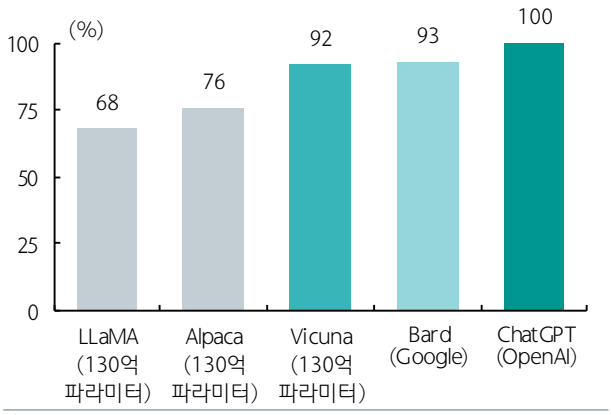
- 파라미터 규모가 수십~수백억 개나 되는 LLM은 훈련에 컴퓨팅 자원, 시간, 비용이 막대하게 소요되므로[3] 개인/비영리 단체가 구축하기 쉽지 않았음
- '23년 2월 Meta가 공개한 트랜스포머 디코더 기반 사전학습 모형인 LLaMA는 오픈소스 진영에게 다양한 LLM을 출시할 수 있는 발판 제공(LLaMa 1, 2023.2.; LLaMA 2, 2023.7.)
  - GPU 병렬 처리, 파라미터 효율적인 훈련 기법(PEFT) 등의 부상도 개인이 LLM을 직접 훈련할 수 있도록 접근성을 높여 오픈소스 진영의 부상 촉진
- Vicuna 등 주요 오픈소스 LLM은 ChatGPT 등에 버금가는 고무적인 성능을 달성
  - (**Stanford Alpaca**) ChatGPT가 생성한 프롬프트-응답 쌍을 학습데이터로 활용해 LLaMA를 파인튜닝한 모형으로, 구축비용이 매우 저렴(약 600달러)하다는 장점 보유
  - (**Vicuna**) ChatGPT 사용자들이 'SharedGPT' 사이트에 공유한 프롬프트-응답 쌍을 기반으로 LLaMA를 파인튜닝 및 개선한 모형으로서, ChatGPT의 92%에 준하는 성능 달성
  - (**EleutherAI**) OpenAI의 GPT-3 상업화 결정에 반발하여 조직된 비영리 AI 연구단체로, GPT-3를 재현한 모형인 GPT-Neo('21.7), 다국어 LLM인 Polyglot('22.10) 등 공개
  - (**Open Assistant**) ChatGPT와 유사한 오픈소스 챗봇 서비스로서, LLaMA를 기반으로 하되 서비스 이용자(기여자)들의 피드백을 직접 받아 모형 파인튜닝(RLHF)

그림10 | LLM 접근법 : Closed vs. Open Source 진영



자료 : 각 회사/연구기관 홈페이지

그림11 | Vicuna (오픈소스 LLM) 성능 비교



자료 : LMSYS Org

3) GPT-4는 인공지능 훈련용 Nvidia GPU인 A100을 1만~2만5천 대 사용하여 훈련된 것으로 추정되고 있는데, 통상적으로 A100의 대당 가격이 약 2,500만 원(2023.9 기준)임을 고려하였을 때, 인건비, 인프라 구축·관리 비용 등을 배제한 순수한 GPU 비용만으로도 GPT-4의 훈련 소요 비용은 2.5조~6.3조 원 가량으로 추정됨(ZDNet)

4. 국내 기업의 LLM

■ 국내에서는 네이버·카카오 등 IT기업 중심으로 자체 한국어 LLM을 구축하고 관련 상업 서비스를 출시하거나 기존 서비스에 수직적으로 통합하려는 시도

- **(네이버, '하이퍼클로바X')** 한국어 특화 LLM인 '하이퍼클로바X'를 기반으로 일반 소비자(B2C)와 기업(B2B) 수요자를 아우르는 서비스를 제공하는데 주력
  - 하이퍼클로바X는 뉴스 기사 · 블로그 등 한국어 말뭉치 중심으로 학습되어(ChatGPT 대비 6,500배) 한국 정보(문화, 지리, 교통 등)에 특화된 자연어 생성 가능
  - 대화·창작·요약 등이 가능한 챗봇 '클로바X', 대화형 검색서비스 '큐:'를 통해 소비자의 정보 탐색 경험을 개선하고, 쇼핑·여행 등 기존 서비스와 결합해 이용 편의 제고
  - '클로바 스튜디오' 등 기업대상 서비스를 통해 기업 고객의 맞춤형 LLM 서비스 구축을 지원
- **(카카오, 'KoGPT')** 자체 구축 LLM인 'KoGPT'를 경량화·최적화하여 카카오톡 등 자사 서비스에 통합함으로써 해당 서비스 이용자들의 경험을 개선하고자 시도
  - KoGPT(v1) 60억 파라미터 모형을 대중에 오픈소스로 공개함으로써, 메타의 LLaMA처럼 오픈소스 진영의 지지·지원을 추구하는 전략 추종
- **(LG AI연구원, 'EXAONE 2.0')** 지식 특화 및 멀티모달 LLM인 EXAONE 2.0을 기반으로 다양한 전문가 대상 서비스를 구축하고 기업 등에 판매할 예정
  - 전문가용 챗봇('유니버스'), 화학·바이오 분야 특화·R&D 지원 플랫폼('디스커버리'), 창작 지원 플랫폼('아틀리에') 등이 대표적으로 범용적 기능을 목표로 하는 타 LLM과 차별화

■ 오픈소스 진영의 방법론에 영향을 받은 다양한 한국어 오픈소스 LLM도 활발하게 개발·공개되고 있음

- **(KoAlpaca)** 한국어를 사용해 Stanford Alpaca와 동일한 방식으로 학습을 진행한 언어 모형으로서, Polyglot 및 LLaMA를 기반 모형으로 하는 버전을 공개
  - Polyglot은 EleutherAI가 만든 다국어 오픈소스 LLM으로 한국어 자연어처리와 관련해 LLaMA 기반보다 성능이 우수하다고 알려짐
- **(KULLM[구름])** 고려대학교 NLP&AI 연구실과 HIAI 연구소가 개발한 한국어 LLM으로 Polyglot을 기반으로 함

III. 금융산업 내 LLM 활용 사례

■ LLM 등 생성형 AI는 은행업 내에서 '마케팅/판매, 고객지원/관리, 프로그래밍, 규제준수' 비즈니스 분야에서 생산성 제고에 기여할 것으로 기대(McKinsey)

- 생성형 AI는 글로벌 은행산업 내에서 약 2,000억~3,400억 달러의 가치를 창출할 것으로 전망되며, 이는 산업 전체의 매출에서 2.8~4.7%에 상응

■ 금융회사는 내부적으로 직원의 업무 수행을 지원하고 자동화하며, 자연어 기반 정보를 수집·분석해 전략적 판단을 내리기 위하여 LLM을 활용

- (Case ①: **대고객 업무 보조**) LLM가 적시에 최적의 정보를 요약하여 전달하거나 판매 절차의 적절성을 모니터링하는 등 직원의 대고객 업무 처리 과정을 보조함으로써 고객 경험 개선 가능
  - (Morgan Stanley) 산업·자산군(群)에 대한 방대한 내부 리서치 자료를 검색 및 요약하여 재무상담사에게 제공함으로써 상담을 지원하는 GPT-4 기반 서비스 구축
  - (**신한은행 '완전 판매 프로세스'**) AI가 투자상품 판매과정을 실시간으로 분석하여 불완전 판매 요소를 탐지하고 직원에게 가이드를 제공함으로써 완전 판매를 지원
- (Case ②: **후선업무 자동화**) 서류 등의 생성 또는 처리 과정을 자동화함으로써 직원의 업무상 실수를 줄이고 전략적 판단을 지원
  - (JP Morgan, 'COiN') AI가 기업 대출 계약서 등을 분석하여 유형을 분류하고 핵심적 문구를 추출함으로써 직원이 서류를 분석하는 시간을 줄이고 정확도를 제고

표2 | 은행 비즈니스 영역별 생성형 AI 도입 가치 및 사례

부가가치 창출 규모	산업 매출 대비 비중	비즈니스 영역별 생산성 제고에 미치는 영향			
		마케팅/판매	고객 관리	프로그래밍	리스크/법률
		높음	높음	보통	낮음
		개인고객 맞춤형 서비스	소비자를 위한 문제상황 상시 지원 서비스	낡은 전산 코드 개선	리스크 관리모형 문서화 지원
$2,000억~ $3,400억	2.8~4.7%	소비자 특성 및 거래이력을 바탕으로 맞춤형 마케팅 콘텐츠를 생산하고, 마케팅 효과 평가를 위한 A/B 테스트 과정을 지원	AI 기반의 음성 서비스로 신용카드 분실 등 문제상황에 처한 소비자를 상시 지원 가능	자연어를 통해 프로그래밍 코드를 생성하는 기술로 낡은 전산시스템의 마이그레이션 작업 효율화	리스크 관리 모형을 구축하고 최신 규제정보를 반영하거나 미흡한 요소를 탐지할 수 있도록 지원

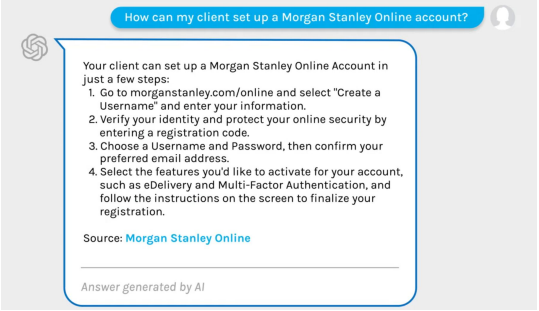
자료 : Mckinsey

- (**Case ③: 시뮬레이션 코칭/훈련**) 다양한 시나리오별 시뮬레이션에서 LLM에게 고객 등 직원의 상대역 역할을 맡김으로써 직원 훈련 또는 평가 가능
  - (Zenerate AI) AI 상담 시뮬레이션을 통해 은행 신규 상담사에게 실제와 동등한 상담 경험을 제공함으로써 상담사의 훈련 속도를 56% 제고
- (**Case ④: 자연어 분석**) 정책당국의 성명서나 고객의 소리(VoC) 등을 분석함으로써 미래의 정책 방향성 예측, 고객의 pain point 진단, 이상 거래 탐지 등 가능
  - (JP Morgan, **'Hawk-Dove Score' 모형**) ChatGPT를 기반으로 하는 LLM을 활용하여, 과거 25년 동안의 美 연준 성명서를 분석함으로써 연준의 정책 변화 예측 시도
  - (**농협은행, '아르미AI'**) AI 콜봇을 활용해 고객 만족도 조사를 자동화하고, 조사 결과에 대한 통계 추출 및 분석까지 자동으로 진행
- (**Case ⑤: 프로그래밍**) ChatGPT처럼 프로그래밍 코드도 생성할 수 있는 LLM을 활용하여 코드 보수/최적화, 서비스 마이그레이션 등 수행 가능
  - (Goldman Sachs) 소프트웨어 개발자가 생성형 AI를 활용하여 코드를 자동으로 작성할 수 있는 서비스를 시범 운영 중으로, 현재 코드의 40% 정도를 AI가 생성

■ 금융회사는 LLM으로 '챗봇'을 고도화함으로써 소비자의 효용을 제고

- LLM을 통해 챗봇의 자연어 이해 및 응대 정확성을 향상하고, 자산관리 등의 분야로 상담 영역을 확장함으로써 소비자의 챗봇 이용률을 제고 가능
  - (JP Morgan, 'IndexGPT') 고객 요청에 따라 금융자산을 분석하고, 투자 의사 결정을 지원하는 AI 투자상담사 서비스인 'IndexGPT'의 상표 출원
- 멀티모달 기능의 챗봇을 통해 시각장애인 등 취약계층의 서비스 이용 지원
  - (Be My Eyes) GPT-4의 멀티모달 기능을 통해 시각 장애인에게 사진에 대한 설명 제공

그림12 | Morgan Stanley, GPT-4 기반 상담사용 챗봇



자료 : OpenAI, 고객 사례

그림13 | Be My Eyes, 시각장애인을 위한 화면 설명

자료 : Be My Eyes, 앱스토어 서비스 안내 갈무리

[참고 3] 금융 특화 LLM 구축 사례

- 금융업 고유의 전문용어 및 복잡성 등을 고려하였을 때 관련 자연어처리 문제를 더욱 잘 해결할 수 있는 금융업 전용 LLM 개발의 필요성이 대두
  - 의학 등 특정 전문분야 데이터로 사전학습된 모형을 기반으로 하는 LLM은 범용적 데이터로 사전학습된 LLM보다 해당 분야의 문제해결력 우수(Gururangan)
- Bloomberg는 자체적으로 보유한 금융 데이터를 통해 BloombergGPT 구축
  - BloombergGPT는 금융 분야(3,630억 개, 전체 학습데이터의 51.27%) 및 일반 분야(3,450개, 48.73%) 데이터로 훈련된 500억 개 파라미터를 가진 금융 특화 LLM
  - 블룸버그 터미널 고객 및 소속 기자가 금융 관련 정보를 손쉽게 검색할 수 있도록 돕고, 금융 뉴스, SNS 게시글 등에 대한 자연어 분석을 지원
    - BloombergGPT는 기사·SNS·기업 공지의 감성 분석, 기업 재무공시 보고서에 기반한 추론/질의응답, 뉴스 기사 기반 개체명(인명, 회사 등) 인식 및 탐지 가능
- 美 컬럼비아대·뉴욕대·프린스턴대 등이 참여하는 오픈파이낸스 연구기관인 AI4Finance는 오픈소스 금융 특화 LLM인 FinGPT를 공개
  - FinGPT(v3.2)는 LLaMA2 70억 파라미터 모형을 기반으로 구축된 모형으로서, 기사 등 다양한 금융 텍스트의 감성분석에서 BloombergGPT보다 우수
    - BloombergGPT처럼 훈련에 53일, 267만 달러가 소요되는 막대한 작업으로는 변동성이 높은 금융시장의 최신 정보를 담고 있는 LLM 구축이 불가능하다는 점을 지적하며, 파라미터 효율적 학습 방법을 적용해 600달러 상당의 적은 비용으로 빠르게 파인튜닝이 가능하다는 점을 강조

표3 | FinGPT: BloombergGPT 및 LLaMA2와의 성능 비교

평가지표	BloombergGPT	LLaMA 2	FinGPT(v3.2)
FPB[1]	0.511	0.390	**0.850**
FiQA-SA[2]	0.751	0.800	**0.860**
컴퓨팅자원	A100[3] 512개	A100 2048개	A100 8개
훈련시간	53일	21일	2시간
소요비용	267만 달러	423만 달러	656달러

1) 금융 뉴스 감성분석 데이터 세트를 통한 평가 (높을수록 우수)
2) 금융 뉴스 및 블로그 제목 데이터 세트를 통한 평가 (높을수록 우수)
3) AI 전용 Nvidia GPU 기준 (한화 약 2,500만 원, 2023.9. 기준)

자료 : FinGPT (모형간 비교 가능한 부분만 일부 발췌)

Ⅳ. 시사점

■ ChatGPT는 인공지능에 의한 자연어 이해·추론 및 창작 가능성을 선보이면서, 단순·반복 사무직 업무뿐만 아니라 숙련된 지식 노동 업무도 인공지능에 의해 대체되거나 도움을 받을 가능성이 있음을 보임

- 美 근로자의 19%는 업무의 50% 가량을, 60%는 업무의 10% 가량을 ChatGPT 등 LLM의 영향을 받을 수 있는 직군에 종사하는 것으로 추정됨(OpenAI)

■ 금융산업은 LLM 도입 추세에 저항하기보다 오히려 적극적으로 활용하여 업무 생산성을 제고할 필요

- 금융산업(특히 은행업)은 손님(개인, 기업)과의 상호작용 비중이 높은 사무직 직종인 점에서 LLM 도입으로 인한 단점보다는 장점이 더욱 클 것으로 예상
  - 마케팅/판매, 고객 분석, 후선업무 처리 등에 LLM을 접목하여 자동화하는 대신, 직원은 대고객 업무에 집중함으로써 업무 생산성 제고

■ 금융회사가 LLM을 활용하는 경우, 금융산업 고유의 특수성을 고려하여 관련 리스크 정책을 사전에 구축하고 모형을 보수하기 위한 지속적인 노력 필요

- 금융산업은 타 산업 대비 신뢰, 소비자 보호 규제, 포용 금융 등의 특수성 보유
  - **(신뢰)** LLM은 부정확하거나 비윤리적인 내용을 생성할 수 있으며, 적대적 사용자의 공격(기업 기밀 노출·부적절한 응답 유도)에도 노출되어 있어 신뢰 및 평판 위험 유발 가능
  - **(소비자 보호 규제)** 타 업권(e-커머스, 콘텐츠 등)에서 '상품 추천'은 일반적인 리스크 요인이 아니나, '금융상품의 추천'은 「금융소비자보호법」을 비롯한 각종 규제 대상
  - **(포용금융)** 고령자, 장애인 등 디지털에 익숙하지 않은 취약계층 소비자에게 대안 없이 AI 서비스(챗봇 등)만을 제공하는 것은 소비자를 배제할 위험 있음
- 환각 등 LLM 리스크를 통제하기 위해 리스크 정책을 사전에 수립해야 하며, 지속적인 테스트로 모형의 리스크 수준을 파악하고 보수해 나가야 함
  - 리스크 정책은 LLM 생성 결과가 신뢰 등에 미치는 심각성에 따라 ① 프롬프트 엔지니어링 등으로 생성을 보완할 부분, ② 인간 관리자의 사전 검토 후 수요자에게 제공할 부분, ③ 아예 생성을 금지할 부분 등으로 구분 가능

■ 한편, 내부의 기술적 역량, 비용 등을 종합적으로 고려하여 빅테크가 구축한 LLM을 이용할 것인지 자체 구축할 것인지 또한 검토의 대상

- OpenAI 등 빅테크가 구축한 LLM을 이용하는 경우, 자연어 이해 및 생성의 품질을 보장하면서 서비스 구축, 운영, 보안에 대한 기술적 요구사항을 절감
  - 단, 빅테크에 대한 기술 의존도 심화, API 사용료 부담, 데이터보안 문제[4] 등을 유발
- 금융회사가 자체 LLM을 구축하는 경우, 자연어처리 성능 제고 및 훈련 과정의 컴퓨팅 자원 부족 문제를 극복하기 위해 오픈소스 진영의 방법론을 참고
  - 인간 피드백 기반 강화학습(RLHF) 등의 훈련 방법론을 적용하되 충분한 훈련 데이터를 확보할 수 있도록, LLM 생성 응답 평가 등 이용자의 참여 및 기여 독려
  - 파라미터 효율적인 파인튜닝(PEFT) 등 다양한 훈련 방법론 또한 참고할 필요

4) 단, OpenAI 이용 약관에 따르면, 기업 손님이 자사 모형을 파인 튜닝하여 커스터마이징한 모형 또는 기업의 입출력 데이터를 자사 모형을 학습하는 데 활용하지 않겠다는 입장

[참고 4] 프롬프트 엔지니어링, 프롬프트 튜닝, PEFT

■ 프롬프트 엔지니어링(Prompt Engineering)이란 LLM 챗봇의 응답의 품질을 향상할 수 있도록 인간이 직접 프롬프트를 설계하는 것

- 프롬프트 엔지니어링은 환각현상을 일부 완화하고, LLM이 취약한 수리 문제의 응답 정확성을 개선하거나 최신 정보를 활용한 응답이 가능하게 하는 등 응답 품질의 개선 효과 입증
  - 통상 명시적으로 드러나지 않는 시스템 프롬프트 형태로 LLM 챗봇 서비스 제공자에 의해 구축되며, 사용자가 챗봇 이용 시 함께 LLM에 투입되어 응답 품질 개선에 기여
- 대표적인 프롬프트 엔지니어링 기법으로는 역할 부여, 맥락 제공, 사례 제공, 추론 과정의 나열 요구, 응답의 자가 평가 및 수정 요구 등 방법이 있음
  - (Role & Tone) LLM 챗봇과 이용자에 대한 특정 역할(전문가, 은행원 등)을 부여하고, 어떤 어조(친절한, 상세한 등)의 답변을 생성할 것인지 지정
  - (One Shot/Few Shot) 1개 이상의 사례를 제시 및 참고하여 대답할 것을 요구
  - (CoT: Chain of Thought) 추론 단계를 예시로 제공하거나 '단계적으로 생각해 봅시다' 등의 지시를 넣어 LLM이 추론과정을 상세히 나열하도록 하여 응답 품질 개선

■ 프롬프트 튜닝(Prompt Engineering)이란 딥러닝 모형이 생성한 '소프트(soft) 프롬프트'를 입력과 함께 넣어 LLM 챗봇의 성과를 개선하는 것

- 소프트 프롬프트는 인간이 만든 자연어 프롬프트인 '하드(Hard) 프롬프트'와 달리 일련의 수리적 값(벡터)으로 해석 가능성은 떨어지지만, 통상적으로 프롬프트 엔지니어링보다 성능 개선 효과가 우수

■ 파라미터 효율적인 파인튜닝(PEFT : Parameter Efficient Fine Tuning)이란 새로운 파라미터나 소프트 프롬프트 등을 사전학습 모형에 추가하여 파인튜닝하는 방법론으로 파인튜닝에 소요되는 비용·시간을 크게 절감

- 사전 학습된 모형의 특정 위치에 파라미터를 추가하는 LoRA(Low-Rank Adaptation), IA3(Infused Adapter by Inhibiting and Amplifying Inner Activations) 등이 대표적

[참고문헌]

안상준·유원준(2023), “딥러닝을 이용한 자연어처리 입문”, 위키독스
오렐리앙 제롱(2020), “핸즈온 머신러닝 (2판)”, 한빛미디어
Eric Lamarre et al.(2023), “Rewired: The McKinsey Guide to Outcompeting in the Age of Digital and AI”, Wiley

Ashish Vaswani et al.(2017), “Attention is All you Need”, Neural Information Processing Systems
Hongyang Yang et al.(2023), “FinGPT: Open-Source Financial Large Language Models”, ArXiv abs/2306.06031
Jared Kaplan et al.(2020), “Scaling Laws for Neural Language Models”, ArXiv abs/2001.08361
Long Ouyang et al.(2022), “Training language models to follow instructions with human feedback”, ArXiv abs/2203.02155
Rishi Bommasani et al.(2021), “On the Opportunities and Risks of Foundation Models”, ArXiv abs/2108.07258
Shijie Wu et al.(2023), “BloombergGPT: A Large Language Model for Finance”, ArXiv abs/2303.17564
Suchin Gururangan et al.(2020), “Don’t Stop Pretraining: Adapt Language Models to Domains and Tasks”, ArXiv abs/2004.10964
Tyna Eloundou et al.(2023), “GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models”, ArXiv abs/2303.10130

고려대학교 NLP&AP 연구실, HIAI 연구소 GitHub, “KULLM (구름): Korea University Large Language Mode”, https://github.com/nlpai-lab/KULLM
카카오브레인 GitHub, “KoGPT”, https://github.com/kakaobrain/kogpt
EleutherAI GitHub, “Polyglot: Large Language Models of Well-balanced Competence in Multi-languages”, https://github.com/EleutherAI/polyglot
KoAlpaca GitHub, “KoAlpaca: Korean Alpaca Model based on Stanford Alpaca (feat. LLAMA and Polyglot-ko)”, https://github.com/Beomi/KoAlpaca
LMSYS Org, “Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%* ChatGPT Quality”, https://lmsys.org/blog/2023-03-30-vicuna/

Mooler0410 GitHub, “The Practical Guides for Large Language Models”, https://github.com/Mooler0410/LLMsPracticalGuide

OpenAI, “Improving Language Understanding by Generative Pre-Training”, https://openai.com/research/language-unsupervised

Stanford University Human-Centered Artificial Intelligence, “Alpaca: A Strong, Replicable Instruction-Following Model”, https://crfm.stanford.edu/2023/03/13/alpaca.html

Xavier Amatriain, “Transformer models: an introduction and catalog – 2023 Edition”, https://amatriain.net/blog/transformer-models-an-introduction-and-catalog-2d1e9039f376

안상현(2023), “‘K-초거대AI’ 윤곽… 네이버는 유료, 카카오는 무료”, 조선일보

Amir Efrati and Aaron Holmes(2023), “OpenAI Passes $1 Billion Revenue Pace as Big Companies Boost AI Spending”, The Information

Michael Chui et al.(2023), “The economic potential of generative AI: The next productivity frontier”, McKinsey

Oktavia Catsaros(2023), “Generative AI to Become a $1.3 Trillion Market by 2032, Research Finds”, Bloomberg Intelligence

Zenerate(2023), “Zenarate AI Simulation Training Wins Banking Tech Awards USA's 'FinTech of the Future' as it Transforms How Contact Center Agents Learn and Perform”

Amazon Web Services(2022), “Train and deploy large language models on Amazon SageMaker”

"KoAlpaca: Korean Alpaca Model based on Stanford Alpaca (feat. LLAMA and Polyglot-ko)", KoAlpaca GitHub, https://github.com/Beomi/KoAlpaca

"KoGPT", 카카오브레인 GitHub, https://github.com/kakaobrain/kogpt

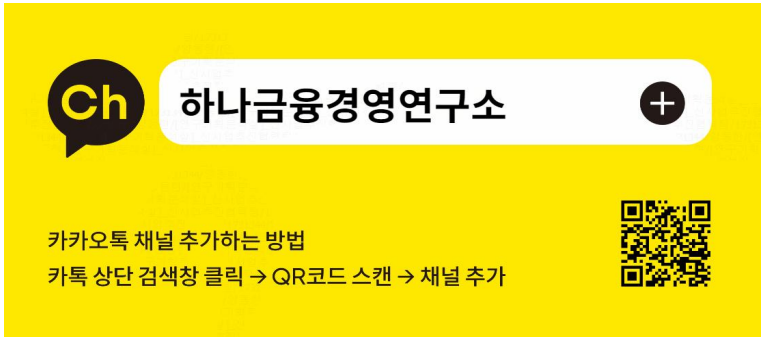
"KULLM (구름): Korea University Large Language Model", 고려대학교 NLP&AP 연구실, HIAI 연구소 GitHub, https://github.com/nlpai-lab/KULLM

LMSYS Org (2023), Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%* ChatGPT Quality

"Polyglot: Large Language Models of Well-balanced Competence in Multi-languages", EleutherAI GitHub, https://github.com/EleutherAI/polyglot

“The Practical Guides for Large Language Models”, Mooler0410 GitHub, https://github.com/Mooler0410/LLMsPracticalGuide

Xavier Amatriain(2023) “Transformer models: an introduction and catalog – 2023 Edition”, https://amatriain.net/blog/transformer-models-an-introduction-and-catalog-2d1e9039f376/

하나Knowledge+

04538 서울특별시 중구 을지로 66 (을지로 2가, 하나금융그룹 명동사옥 8층)
TEL 02-2002-2200
e-mail hanaif@hanafn.com
http://www.hanaif.re.kr