

챗GPT, 생성형 AI가 가져올 산업의 변화

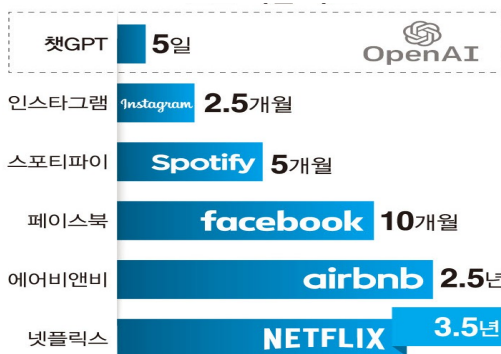
[애자일] KIAT 산업기술정책단 정책기획실('23.02.28)

- ◆ 최근 화제가 된 챗GPT¹⁾로 대표되는 생성형 AI 발전 양상은 초거대 AI의 활용 시대가 임박하였음을 시사
- ◆ 초거대 AI 모델을 활용하여 혁신적인 서비스를 선제적으로 개발하고 선도하는 경쟁의 시기가 도래할 것으로 전망되는 만큼 우리 산업정책에의 시사점 탐색

1. 생성형 AI의 개요

- (생성형 AI, Generative AI) 대형 언어모델(LLM)이나 이미지 생성 모델(IGM)을 활용하여 사용자가 원하는 무언가를 만들거나 생성하는 모든 기술을 의미
 - 텍스트만 입력하면 사용자가 원하는 글과 그림을 만들어주고 영상까지 만들어내는 AI
 - * **대규모 언어 모델(Large Language Model)** : 테라바이트 단위의 대용량의 텍스트 덩어리를 통해 모델을 학습시켜 복잡한 문장을 생성가능하며 대표적인 예시가 GPT-3, 챗GPT
 - * **이미지 생성 모델(Image-Generation Model)** : 텍스트를 입력하면 그에 따라 이미지를 생성하는 AI 기술로 대표적인 예시가 DALL-E2, 미드저니

[서비스별 이용자 100만명 달성 소요기간 비교]



[초거대 AI 보유 기업]

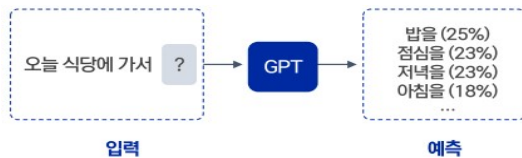
기업명	대표 초거대 AI	학습 매개변수 (파라미터)	주요 특징
오픈AI	GPT-2	15억	*언어생성, 번역, 검색, 기사 작성 등
	GPT-3	1750억	*기존 모든 기능 고도화, 프로그래밍
	GPT-4	100조(추정)	*2023년(예정)
MS	MT-NLG S30B	5300억	*초대형 언어모델
구글	스위치 트랜스포머	1조6000억	*최초 초 단위 파라미터 모델 *연구자에게 논문과 소스코드 형태로 제한적 공개
메타	RoBERTa	3억5500만	*언어 생성, 번역, 검색, 기사 작성 등
네이킹 지위안 인공지능연구원	우디오 2.0	1조7500억	*기존 모든 기능 고도화, 중국어와 이미지 생성
네이버	하이퍼클로바	2040억	*국내 기업 최초 자체 개발 인공지능 모델 *GPT-3보다 6500배 많은 한국어 데이터 학습
카카오	KoGPT	60억	*한국어 특화모델
	MinDALL-E	300억	*이미지 생성 등 멀티 모달
LG	엑사원	3000억	*언어, 이미지 이해·생성, 데이터 추론

출처 : 전자신문('23.2) 右소프트웨어정책연구소, 미래에셋증권취합, 전자신문 재인용('23.1)

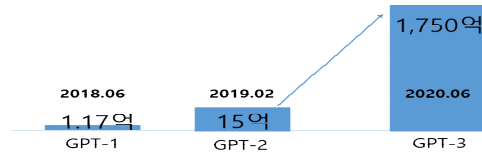
1) OpenAI(비영리기관)는 2018년 GPT-1 버전을 선보인 후, GPT-1보다 1,500배 많은 매개변수 (1,750억개)를 활용한 GPT-3버전을 2022년 발표하며 크게 주목. GPT-3은 대용량 데이터를 학습한 초거대 AI의 원형이라 할 수 있음

- 챗GPT는 OpenAI에서 개발한 GPT-3.5를 기반으로 만들어진 대화형 AI로 사람의 피드백을 통한 강화학습을 적용하여 사람이 쓴 것처럼 의미있는 텍스트를 생성 및 상호소통 가능
- GPT는 주어진 텍스트의 다음 단어를 예측하는 태스크를 학습한 딥러닝 방식의 AI 모델
- * 'Generative Pre-trained Transformer'의 약자며, Transformer'는 문장 속의 단어와 같은 순차적인 데이터 내의 관계를 추적해 맥락과 의미를 학습하는 신경망 모델(Google, 2017)

[GPT의 개념의 예시]



[GPT 시리즈의 파라미터 개수]



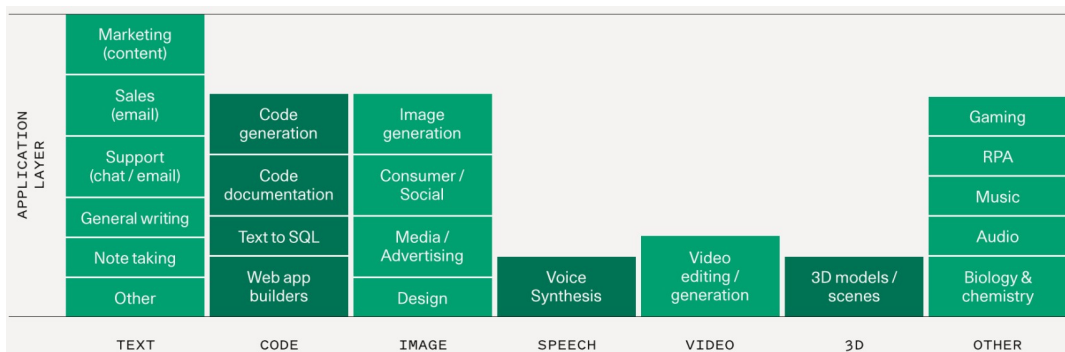
출처 : NIA IT&Future Research 11호('22)

- AI 모델의 대형화 및 성능이 향상에 주목하는 시기는 지나고 있으며, 챗GPT와 같은 초거대 AI를 활용한 혁신적인 서비스 개발 등 응용 분야 성과 창출이 활발해질 전망

2. 생성형 AI의 산업적 응용

- (생성형 AI의 산업적 응용) AI 모델이 하나의 과업을 수행하던 방식을 벗어나 하나의 모델이 여러 과업에 활용할 수 있는 전산업의 기반으로 진화하여 AI R&D 시대에서 응용의 시대로 전환
- AI의 자동화 운영, 전망 및 분석, 텍스트나 이미지 생성, 음성 합성, 3D 모델링 등의 다양한 기능은 제품 및 서비스 개발·판매 등에 다양하게 응용

[생성형 AI의 산업 활용 사례]



출처 : Sequoia Capital, Generative AI ('22)

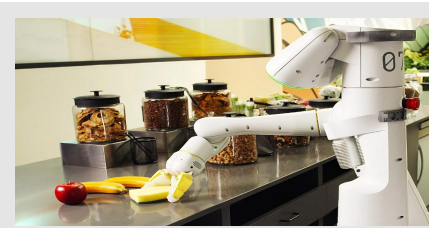
- **(작업 효율화)** 기업 내 필수업무인 판매, 보안, 마케팅 등에서 AI 적용이 가능해짐에 따라 해당 기능에 대한 효율성 제고
- 챗GPT와 같은 챗봇은 일상적인 업무를 자동화하거나 이메일을 활용한 세일즈, 코드를 수정하거나 고객지원을 개선하는 등 고객과의 커뮤니케이션 제고

[주요 생성형 AI 서비스]

- **(비휴먼, bHuman)** 이미 만들어진 동영상에 고객 이름 같은 개별적인 부분을 시가 채워줄 수 있어 개인화된 대량 이메일을 발송가능
- **(달리, DALL-E2)** 사용자의 설명을 듣고 지시에 따라 새로운 그림을 그려주며, 사용자가 이미지를 업로드하여 달리에게 수정요청 가능

- **(인간-로봇 인터페이스)** 증강현실, 동작 컴퓨팅, 로봇 등의 분야에서 언어와 시각(이미지 및 비디오)가 결합하는 단계를 넘어 오디오, 3D, 로보틱스와 같은 개별 분야까지 파급력 확장
- 현재 로봇 프로그래밍의 과제는 엔지니어가 먼저 로봇이 수행해야 할 작업의 요구사항을 로봇 시스템을 위한 적절한 코드로 변환 후 코드를 적용한 후 로봇의 동작을 관찰하고 잘못된 동작일 경우 코드를 다시 수정해야 로봇의 행동을 교정할 수 있다는 점
 - * 시간이 오래 걸리고 비용이 많이 소요
- 챗GPT를 활용하면 미래에는 기술지식이 없는 사용자라도 로봇의 동작을 관찰하고 자연어로 챗GPT를 통해 로봇의 잘못된 동작을 바로잡을 수 있음

[인간-로봇 인터페이스 사례 : PaLM-SayCan]



- **PaLM-SayCan(Google, 22.08)** 로봇과 구글의 대규모 언어 모델을 결합하여 자연어 지시어를 해석하고 행동기능한 로봇
 - 주방에서 로봇에 연결된 노트북에 '나 배고파'라고 입력하면 로봇이 근처의 조리대를 확인하고 플라스틱 집게로 과자봉지를 집어 화자에게 간식 제공

- **(콘텐츠 제작)** 영화 시나리오, 소설, 노래 가사, 광고카피 등의 다양한 콘텐츠를 창의적 형태로 사용자가 원하는 방향으로 제작 가능
- 기존의 미디어 산업에서는 기자들이 정보를 수집하고 작성했지만 생성형 AI를 이용하면 대량의 데이터를 기반으로 자동으로 기사 작성
 - * 번역가, 음성 배우, 일러스트레이터 등의 업무 수행의 다수가 AI 기술에 의해 대체될 가능성이 높아짐
- **(프로그래밍)** 간단한 프로그램 코드짜기, 주식달기, 코드 상 오류 찾기, 에러코드 수정 등 다양한 프로그래밍 작업 수행 가능,
 - * 소프트웨어 개발의 전체 비용을 낮추는 효과가 있으며 Github사의 Copilot 프로그램은 개발자 코드의 최대 40%가 자동 완성되도록 지원

3. 생성형 AI와 제조업

- 제조업에서 생성형 AI를 활용 시 제품설계, 제조 프로세스, 품질관리, 공급망 관리, 로봇공학 및 자동화 개선 가능
 - **신약 개발** : 기존 단백질과 유사점이 전혀 없고 자연에 없는 새로운 단백질 설계용 프로그램을 통해 연구진이 원하는 모양, 크기, 기능의 제약을 설정하고 단백질을 생성하여 효과적인 신약 개발에 활용
 - * 바이오테크 스타트업 제너레이트 바이오 메디슨의 크로마와 워싱턴 대학의 '로제타폴드 디퓨전'
 - **재료 설계**(Materials Informatics, MI) : 데이터 기반 재료 설계로 특정 속성을 대상으로 하는 완전히 새로운 재료를 구성할 수 있음
 - AI는 소재 물성 DB에서 소재(조성 및 구조)정보와 물성 간의 상관관계를 찾고 이를 기반으로 새로운 소재의 물성을 예측하는 방식으로 기존에 비해 월등히 빠른 예측 속도 제공
 - **반도체 설계** : 수백억이 투입되는 고성능 반도체 칩 설계에서 강화학습을 통해 생산성을 극대화할 전망으로 반복 작업을 줄여 원가를 낮추고 설계 인력 간 협업 효율성 강화
 - AI가 설계 도면을 만들어내면 엔지니어가 상벌을 주는 방식으로 혼잡도, 밀도, 소비전력, 면적을 줄여 최적화에 수렴하면 상을 주고 성공하지 못하면 벌을 주는 방식으로 학습
 - * Google은 TPU 반도체 칩 설계 시 AI를 적용하여, 수개월 소요되던 배치 작업을 6시간 이내 수행(21)
 - **합성데이터** : 문제 상황 및 활용목적에 따라 데이터를 직접 획득하는 대신 컴퓨터 시뮬레이션이나 알고리즘에서 생성된 데이터로 실데이터를 낮은 비용으로 대체해줌
 - 수학적으로나 통계적으로 실제 데이터의 분포를 반영하여 AI 모델을 훈련하는데 적합
 - 실제 데이터는 데이터를 수집한 이후 전처리 과정을 거쳐서 개인정보 제거, 오류 제거 및 서로 다른 데이터 형식들도 통일해야 하므로 비용 증가
 - * 가트너(2021.6)에 따르면 2030년에 이르면 AI에 사용되는 데이터 대부분이 규칙, 통계모델, 시뮬레이션 등을 통해 인위적으로 생성될 것으로 전망

[주요 합성데이터 활용 사례]

- 제조분야에선 상대적으로 발생 빈도가 낮은 불량 케이스에 대한 합성 데이터 생성을 통한 효과적인 불량 분석 및 탐지 모델링 수행
- 금융/보험 분야 : 고객 데이터의 외부 반출 및 활용 시 개인식별이 불가능하도록 민감한 개인정보를 비식별화한 합성데이터 활용
- 3D 시뮬레이션을 사용해 제품 이미지 5개만으로도 1,000개 이미지의 합성 데이터 세트 생성 가능

4. 한국형 챗GPT 현황과 정책과제

- 오픈 AI의 GPT-3 발표('20.6) 이후 초거대 AI의 기록적인 성능 향상을 경험한 빅테크 기업들은 경쟁적으로 대규모 자본을 투입하여 초거대 AI 모델 개발 중
 - 미국, 한국, 중국 등 각국에서 모국어 버전의 초거대 모델을 구축 중이며 우리나라도 언어 AI 주권 확보 차원에서 네이버, 카카오, LG전자 등에서 적극적으로 초거대 AI 모델 개발
 - * 미국 : GPT-3(오픈AI), LaMDA2, PaLM(Google), OPT-175B(메타)
 - * 중국 : Wudao2.0(BAAI), Pangu Alpha(화웨이)

[한국어 버전 초거대 AI 모델 개발 현황]

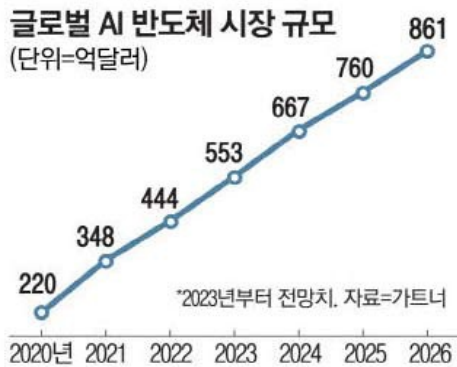
기업	모델명	특 징
네이버	하이퍼클로버	<ul style="list-style-type: none"> • 2,040억개의 매개변수, 5,600억개의 한국어 토큰 • GPT-3보다 한국어 데이터를 6천5백배 이상 학습
카카오브레인	KoGPT	<ul style="list-style-type: none"> • 60억개의 매개변수, 2천억개의 한국어 토큰
LG전자	엑사원	<ul style="list-style-type: none"> • 3천억개의 매개변수 • 6천억개의 말뭉치, 2억5천만개의 이미지를 학습한 멀티모달 모델
SKT	에이닷	<ul style="list-style-type: none"> • AI 비서 서비스로 한국어 GPT-3 기술 적용해서 B2C 서비스 상용화('22..5)

- AI 초거대화에 대한 연구 및 생성형AI가 가져올 산업 변화에 대한 대응 필요
 - AI 주권 차원에서 한국어 기반의 초거대 AI 모델 경쟁력을 확보하고, 대규모 데이터를 빠르게 분석하기 위한 AI 반도체 개발, 클라우드 등 컴퓨팅을 담당하는 기반 기술 지원
 - **(초거대 AI 모델 경쟁력 확보) 대학·스타트업 등 컴퓨팅 자원 지원** : AI 신기술 등장 주기가 짧아지고 모델이 커지면서 컴퓨팅 성능과 자원 등의 한계를 극복하기 위한 정책 필요
 - 대규모 연산을 위한 AI 서버를 자체적으로 구축이 어려운 스타트업, 중소기업, 학교 연구소 등이 연구할 수 있도록 API(응용프로그램 인터페이스) 비용 지원 검토
 - 정부 GPU 인프라의 유휴 자원 분석하여 민간에게 개방 검토
 - **(초거대 AI 모델 경쟁력 확보) 데이터 확보 지원** : 대량의 한국어 언어모델 학습데이터 및 AI 응용 프로젝트에 활용되는 고품질 데이터 확보, 합성데이터 상용화를 위한 기업·기술개발 지원
 - 현재 정부 주도의 학습데이터 구축사업*이 진행 중이나 보편성이 높은 초거대 AI모델을 위한 셋이 아닌 도메인 특화 데이터 셋 위주

* NIA의 인공지능 학습용 데이터 구축사업 진행 중('17년~'25년, 약 2조 6천억원 규모, 381종의데이터 구축 완료)

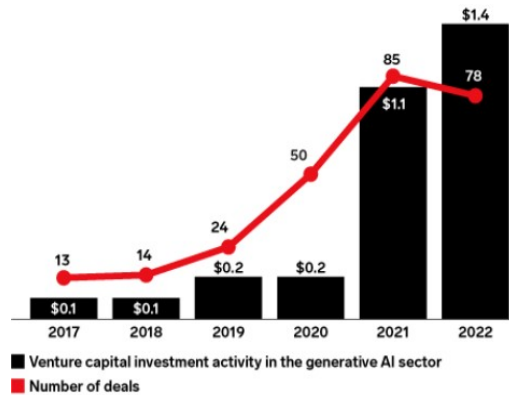
- 한국어 언어모델 개발을 위한 대규모 데이터셋은 글로벌 초거대 AI 모델의 영어 데이터셋과 비교시 상대적으로 부족
- AI 응용 프로젝트 개발에 필요한 시간의 약 80%는 데이터 정리, 표준화에 소요되며, 머신러닝 구현 시 정확성을 높이기 위해서는 고품질 데이터 중요(한국지능화정보사회진흥원, 2021)
 - * AI 데이터 확보 및 품질에 대해서 AI 관련 기업부설연구소의 44.6%, 대학부설연구소의 50%가 문제가 있다고 응답('21년, ETRI)
- 2030년에 이르면 AI에서 사용되는 데이터의 대부분이 합성데이터로 전망되는 만큼 합성데이터 전문 스타트업 지원 및 기술개발 지원, 데이터 구축사업에서 합성데이터 활용 적극 검토

[글로벌 AI 반도체 시장 규모]



출처 : 가트너('23)

[글로벌 VC의 생성형 AI 투자 현황]



출처 : PitchBook('23)

- (AI 반도체, 클라우드 등 기반 기술 확보) 대규모 데이터를 빠르게 분석하기 위해 컴퓨팅을 담당하는 반도체 등 하드웨어의 중요성 또한 증대
- (AI 반도체) △AI 서비스 고도화 △비용 효율화 △전력 소모 절감에서 높은 성능을 구현하는 AI 반도체 개발을 위한 AI반도체 스타트업 지원 및 생태계 구축 필요
 - * 현재 전세계적으로 GPU가 AI 반도체 산업을 주도하고 있지만 GPU는 그래픽 처리를 위한 목적으로 만들어져 텍스트·음성 데이터 등을 다룰 때 시간·비용·전력 효율성이 떨어짐
- 각 산업군에서 경쟁력있는 AI 반도체 칩 개발이 가능하도록 AI 반도체 전문 팹리스들과 클라우드 서비스 운영 기업과의 공동 개발 필요
 - * 아마존 AWS는 자체 개발한 AI 반도체를 자사 데이터 센터에 적용해 컴퓨팅 서비스 비용을 GPU 사용 시 대비 최대 70% 절감, 영상인식 서비스 속도는 8배 향상
 - * Google은 자사의 서비스에 최적화한 AI반도체인 TPUv4를 개발하여 Google colab 등을 통하여 실제 서비스 중
- (AI와 연계된 기술과의 적극 연계 정책) AI기술 하나만을 보는 것이 아닌 양자컴퓨팅, 반도체, 로봇, 제조 등 연계된 미래 기술과의 적극 연계 필요

- AI 서비스의 성능은 획기적으로 향상된 한편, 정보의 신뢰성 판별, 저작권 문제, 제도 정비, 윤리적 이슈, 일부 직무의 AI 대체 가능성 등 총체적인 연구 필요
- 챗GPT의 등장에서 볼 수 있듯이 이제는 쉽게 생성형 AI를 구축할 수 있는 환경이 마련된 만큼 파급력을 갖춘 한국형 초거대 AI 응용서비스 개발 방안 필요

〈참고문헌〉

- 과학기술정보통신부, 2021 인공지능산업 실태조사 2022.04
- 딜로이트 인사이트, 차세대 반도체, 인류에 새로운 세상 열어준다 2023.01
- 전자신문, 스페셜리포트 챗GPT열풍, AI시장을 뒤흔든다 23.1.30
- 정보통신기획평가원, 주간기술동향 초거대 AI 모델 연구 동향 2023.2.8.
- 하이투자증권, 챗GPT, AI 시대의 게임체인저 2023.01.25
- 한국경제, 구글·애플·아마존 AI 반도체 직접만든다 23.2.12
- 한국산업기술진흥원 이슈페이퍼, AI 반도체 기술발전 전망 및 정책적 시사점 2022.09
- 한국전자통신연구원 기술정책이슈, 국내AI 연구기관의 R&D 관련 특성과 시사점 2022
- 한국지능정보화진흥원 IT& Future Research, 데이터 분석 기반 우리나라 AI이슈와 과제, 2021
- 한국지능정보화진흥원 IT& Future Research, 현대 인공지능의 역사적 사건 및 산업·사회 변화 분석, 2022
- IT World Korea, “생성형 AI의 현주소” 주요 생성형 AI 서비스 둘러보기 2023.02.22
- MIT Technology Review, 신약 개발을 돕는 단백질 구조 설계용 AI 2022.12
- Sequoia Capital, Generative AI, 2022.09

※ 한국산업기술진흥원 산업기술정책단 정책기획실 김미현 선임연구원 / mihyun.kim@kiat.or.kr
 ※ 본 자료에 수록된 내용은 작성자의 개인 의견으로 기관의 공식 견해가 아님을 밝힙니다.