

혁신정책

생성형 AI 관련 주요 이슈 및 정책적 시사점

KISTEP 과학기술정책센터 고윤미 · 심정민



생성형 AI 관련 주요 이슈 및 정책적 시사점

(‘23.4.13, 과학기술정책센터 고윤미, 심정민)

1 검토배경

- 세계 빅테크 기업에서 초거대 생성형 AI 서비스인 ChatGPT가 출시되면서 최근 전 세계적인 화두로 등장
 - ChatGPT*는 OpenAI에서 개발한 GPT-3.5 아키텍처를 기반으로 한 대화형 인공지능
 - * ‘22년 11월 3.5버전을 발표하였으며, 단 4개월만에 GPT 4.0 공개(‘23.3.)로 전 세계의 이목을 집중
 - ※ GPT(Generative Pre-trained Transformer)-3(‘20.6.)는 대규모 텍스트 데이터 셋을 사용하여 머신러닝을 통해 문맥의 의미를 이해한 사전학습 후, 특정한 태스크에 대해 파인튜닝을 수행하는 자연어 이해 및 생성능력을 갖춘 인공지능 모델로 기존 버전에 비해 큰 주목을 받음
 - ChatGPT로 촉발된 언어모델 기반 서비스의 가능성이 가시화됨에 따라 빅테크 기업들은 대규모 언어모델¹⁾과 생성형 AI 챗봇 서비스 출시 계획을 발표
 - 국내 대기업에서도 한국어 기반의 초거대 언어모델 개발을 적극 추진하고 있으며, 다수의 스타트업들에서도 다양한 AI 서비스를 제공 중

〈표 1〉 국내 기업의 초거대 AI 모델 개발 추진현황

기업명	모델명	주요 특징
네이버	하이퍼클로바 (HyperCLOVA)	<ul style="list-style-type: none"> • 국내 기업 최초 자체 개발 인공지능 모델 • 2,040억개의 매개변수, 5,600억개의 한국어 토큰 • GPT-3보다 한국어 데이터를 6천5백배 이상 학습 • 하이퍼클로바X 7월 공개 예정
카카오	코지피티 (KoGPT)	<ul style="list-style-type: none"> • GPT-3의 한국어 특화 언어 모델 • 60억개의 매개변수, 2천억개의 한국어 토큰 • 390억 개 매개변수를 학습한 코(Ko)GPT-3.5 공개 예정
LG전자	엑사원 (EXAONE)	<ul style="list-style-type: none"> • 국내 최대 매개변수 보유(3천억개) • 언어, 이미지 이해·생성, 데이터 추론
SKT	에이닷(A.)	<ul style="list-style-type: none"> • 한국어 GPT-3 기술을 적용한 AI 비서서비스 제공(‘22.5.) • 장기기억 기술 및 멀티모달 기술 도입 예정

* 출처 : 글로벌 과학기술정책정보 서비스 235호, KIAT 산업기술정책 애자일 2023-제1호 재인용

- 본 고에서는 생성형 AI에 관한 주요 이슈와 과학기술적 과제를 정리

1) 대형 언어모델은 방대한 대규모 텍스트를 기반으로 훈련된 모델로서 대규모(초거대)에 대한 명확한 기준은 없으나, 최근 발표된 GPT-4의 파라미터 수는 정확하게 공개되지는 않았지만 GPT-3의 약 1,750억보다는 훨씬 많을 것으로 추정(KISTEP 사과플러스 제10호)

2 생성형 AI 관련 주요 이슈

□ (산업적 영향) 업무 효율화를 극대화하고 신규 창작물을 빠른 속도로 생성하면서 산업생태계 전반에 걸쳐 영향을 미칠 것으로 예상

○ 콘텐츠 산업, IT 산업, 제조업을 중심으로 다양한 측면에서 응용이 가능

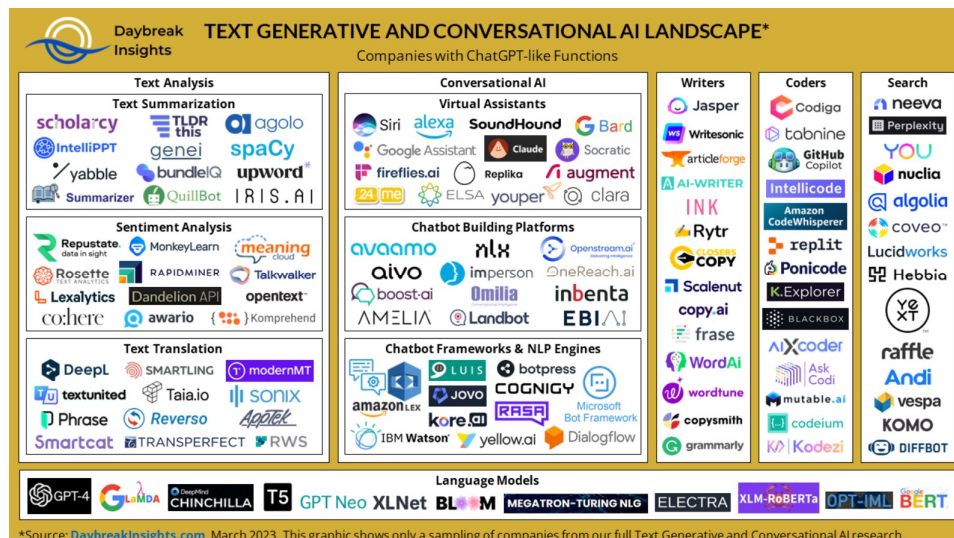
〈표 2〉 ChatGPT의 산업적 활용(예시)

구분	주요 내용
콘텐츠 산업	광고·미디어: 사용자가 원하는 방식으로 “빨리”, “맞춤형”의 다양한 콘텐츠 생성 가능
	가상세계: 메타버스 사용자 서비스 향상, 가상공간 개발 등 지원 가능
IT 산업	반도체: AI 및 데이터센터용 반도체에 대한 수요 증가 예상
	사이버 보안: 테스트, 공격, 사고에 대한 분석자료 작성, 분석한 데이터를 기반으로 고위험군 이벤트 선별, 신·변종 위협 탐지 등 가능
제조업	신약 개발: 연구진이 원하는 모양, 크기, 기능의 제약을 설정하고 단백질 생성 가능
	재료 설계: 데이터 기반 재료(materials informatics)로 특정 속성을 대상으로 하는 새로운 재료 구성 가능
	합성데이터: 컴퓨터 시뮬레이션이나 알고리즘에서 생성된 데이터로 실데이터를 낮은 비용으로 대체 가능

* 출처 : 삼일Pwc경영연구원(2023), KIAT 산업기술정책 애자일 2023-제1호 재인용

○ 기술연구전문 데이브레이크인사이드가 생성형 AI 업계를 조사한 결과, 최소 700개 이상의 기업이 경쟁 중²⁾으로 시장선점을 위한 기업경쟁 심화

※ (1)텍스트 요약 (2)감정 분석 (3)텍스트 번역(이상 텍스트 분석) (4)가상 비서 (5)챗봇 구축 플랫폼 (6)챗봇 프레임 워크 및 NLP 엔진(이상 대화형 AI) (7)글쓰기 도구 (8)코드 생성 (9)검색 (10)언어 모델 등 10개 카테고리로 분류



[그림 1] 텍스트 생성 및 대화형 AI 지형도

* 출처 : 데이브레이크인사이드, 시타임스('23.3.20.)

2) 시타임스('23.3.20.)

- 글로벌시장조사기관인 CB insights가 발표('23.2.)한 생성형 AI 관련 글로벌 스타트업 250개 중 한국기업은 3개*에 불과, 미국기업이 다수 차지³⁾

※ 미국 126개, 인도/영국 14개, 이스라엘 12개, 캐나다 10개, 프랑스 6개, 중국/호주/일본 5개, 네덜란드/스페인 4개, 한국 3개로 13위를 차지

* 시각미디어 분야의 딥브레인AI, 디오비스튜디오, 클레온 등

- (기술적 역량) 우리나라의 생성형 AI 관련 기술적 역량은 압도적 선두국가인 미국, 중국과의 격차가 크게 벌어진 상황

▶ 분석대상 : 최근 5년간(2018~2022년) 생성형 AI분야의 Web of Science Core Collection 중 book을 제외한 논문 54,899건

▶ 분석기관 : 클래리베이트('23.3.)

- 최근 5년간 생성형 AI 분야에서 발표된 논문을 분석한 결과, 우리나라의 전체 논문 수는 총 2,682건으로 전체 5위를 차지

※ 중국 19,318건(1위), 미국 11,624건(2위), 인도 4,058건(3위), 영국 3,484건(4위)

- 피인용 상위 1% 논문은 총 70건으로 조사대상국 중 7위 수준⁴⁾으로 1위(미국), 2위(중국)와 격차가 큰 편

※ 미국 691건(1위), 중국 565건(2위), 영국 144건(3위), 독일 107건(4위), 호주 93건(5위)

- 세부 기술분류별*로 살펴보면, 우리나라는 이미지 생성형 AI 분야의 논문이 가장 많이 게재되었고 다음으로 자연어처리(NLP), 비디오의 순

* 자연어처리(NLP), 음성, 이미지, 비디오, 신약개발 등 5개 세부 기술로 분류

- 인용수 대비 상위 1% 논문은 이미지 생성형 AI 분야에 집중, 자연어 처리(NLP) 분야의 경우 전체 논문수에 비해 인용수 대비 상위 1% 논문 비중이 저조한 편

- 미국은 자연어처리(NLP) 생성형 AI분야에 강점을 보유하고 있으며, 중국은 이미지 생성형 AI 분야의 논문 수가 가장 높게 나타남

3) 머니투데이(2023.2.22.)

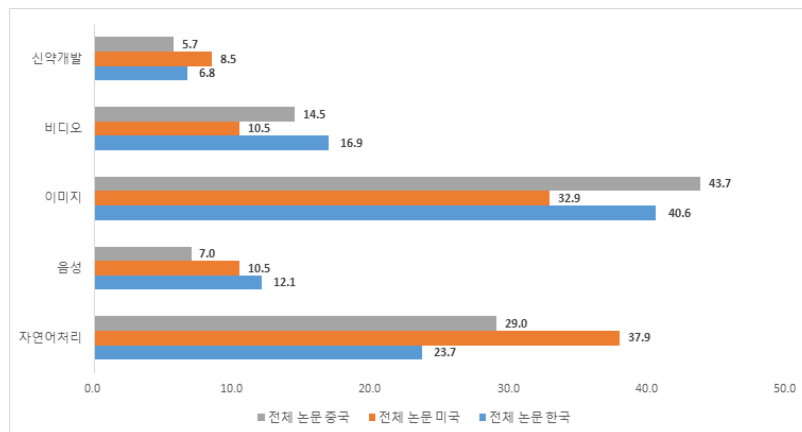
4) 클래리베이트 분석결과('23.3.)

〈표 3〉 생성형 AI 분야의 국가별 세부 기술분류별 논문 수 현황

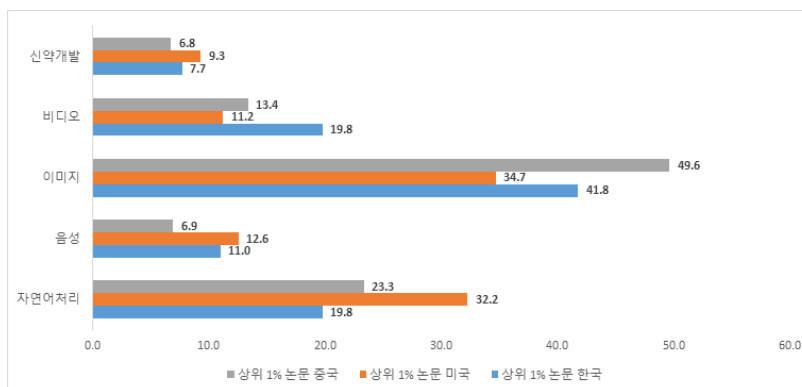
구분 (단위:건,%)	한국		미국		중국	
	전체 논문 수	인용수 대비 상위 1% 논문 수	전체 논문 수	인용수 대비 상위 1% 논문 수	전체 논문 수	인용수 대비 상위 1% 논문 수
자연어처리 (NLP)	712 (23.7)	18 (19.8)	4,865 (37.9)	267 (32.2)	5,961 (29.0)	152 (23.3)
음성	363 (12.1)	10 (11.0)	1,347 (10.5)	104 (12.6)	1,439 (7.0)	45 (6.9)
이미지	1,219 (40.6)	38 (41.8)	4,223 (32.9)	287 (34.7)	8,986 (43.7)	323 (49.6)
비디오	508 (16.9)	18 (19.8)	1,347 (10.5)	93 (11.2)	2,976 (14.5)	87 (13.4)
신약개발	203 (6.8)	7 (7.7)	1,091 (8.5)	77 (9.3)	1,181 (5.7)	44 (6.8)
총계 ¹⁾	3,005 (100.0)	91 (100.0)	12,873 (100.0)	828 (100.0)	20,543 (100.0)	651 (100.0)

* 주¹⁾ : 여러 기술 분야에 속하는 논문들이 다수 존재하므로 국가별 총계와 다름

* 출처 : 클래리베이트 분석결과('23.3.)



[그림 2] 생성형 AI 분야의 국가별 세부 기술분류별 전체 논문의 비중 현황



[그림 3] 생성형 AI 분야의 국가별 세부 기술분류별 인용수 대비 상위 1% 논문의 비중 현황

* 출처 : 클래리베이트 분석결과('23.3.)

□ (해결과제) 생성형 AI 활용과정에서 법적·윤리적 측면의 해결해야 할 과제가 존재

- 허위정보, 저작권, 논문 표절, 개인정보보호 및 보안, 답변의 윤리 이슈, 부정행위, 학습능력 및 창의성 저하 등 예기치 않은 위험 이슈가 존재
- 일부 개발자 및 기업 영향력이 집중되면서 인간의 편견이 시스템에 반영되거나 이를 활용하는 방식에 따라 국가안보 문제가 발생할 가능성 존재
- 주요국에서는 적극적으로 규제조치를 시행하였고, 국제학술지 네이처, 사이언스에서도 ChatGPT를 논문 저자로 인정하지 않는 가이드라인을 제시('23.1.)

※ (미국)연방거래위원회는 A테크 기업 조사 착수, 국가통신정보국(NTIA)은 AI 모델 출시 전 위험성 확인인증에 대한 규제사전검토 시작(4월), (이탈리아)데이터 보호청은 국가 차원에서 자국 내 접속을 차단(3월), (영국)AI 산업 및 규제 가이드라인이 포함된 AI 백서 발간(2월)

3 시사점

□ 초거대 AI는 산업의 지형을 완전히 바꿀 파괴적 혁신으로 전망되므로 우리나라 초거대 AI산업이 글로벌 시장을 주도할 수 있도록 정부의 정책적 지원 강화 필요

- 정부에서는 인공지능 경쟁력 강화를 위한 정책 추진* 및 생성형 AI에 관한 R&D사업을 추진 중

* 인공지능 일상화 및 산업 고도화 계획(과기정통부, '23.1.)을 통해 인공지능 10대 핵심프로젝트 추진('23년 약 7,129억원 투입) 중

※ '한국어 대형 언어모델 개발사업'을 통해 10억개 수준의 매개변수를 가진 트랜스포머를 개발하고 있으며, '23년도에 3억개 수준의 매개변수를 가진 한국어 언어모델 공개 예정

〈표 4〉 생성형 AI 관련 주요 정부 R&D 사업 현황

내역사업명	주요 내용	'23년도 예산(백만원)
사람중심인공지능 핵심원천기술개발	현 AI 기술의 한계를 극복하여 인간이 효과적으로 활용가능한 인공지능 구현으로 차세대 인공지능 핵심원천기술 확보	49,850
한국어 대형 언어모델 기술개발	대용량 한국어 데이터로부터 범용적 의미표현을 사전학습하고 이를 응용 분야에 적용·활용하기 위한 한국어 대형 언어모델 기술개발	3,000
차세대인공지능 핵심원천기술개발	차세대 인공지능 핵심원천기술 확보를 통한 인공지능 글로벌 기술 경쟁력 제고	5,000

* 출처 : 정보통신기획평가원(IITP) 홈페이지

- 민간과 협업하여 혁신적 기술의 발전 속도에 발맞추어 이머징 기술을 선제적으로 발굴할 수 있는 체계 및 기초·원천연구 강화
- AI 분야의 R&D 투자 현황 및 부처별 전략·계획에 대한 분석을 바탕으로 기술공백 영역이 발생하지 않도록 지원 및 투자필요영역 발굴

- 우리의 AI 기술적 역량은 선도국과 격차가 큰 상황으로 AI 경쟁력 고도화를 위해 요구되는 산업 생태계 조성을 위한 기반 마련 필요
 - AI 기술 초격차를 위한 생성형 AI 기술개발 우수인력 확보 및 양성 강화
 - 생성형 AI 관련 데이터 구축·개방, 컴퓨팅 파워 및 자원 제공, 윤리 및 신뢰성 확보 등 산업생태계 기반조성
 - AI 관련 원천기술을 고도화하기 위한 산학연 협력 및 딥테크 창업기업 지원을 강화할 필요
- 생성형 AI에 대한 진흥을 강화하는 한편, 장단기적으로 미래 위험 요인에 대응하기 위한 정책적 조치를 검토할 필요
 - AI 튜터 도입, 딥페이크 등 잘못된 정보에 따른 미디어보호 등을 통해 AI의 전반적 활용을 강화
 - 새로운 기능 도입에 따른 출시 전 테스트 강화, 국가 안보 위험성 사전 검토 등을 통해 생성형 AI의 안전성을 국가 차원에서 보장

참고문헌

- 글로벌 과학기술정책정보 서비스(S&T GPS) 235호, 이미지 인식 등 한층 고도화된 GPT-4... 업계 도입 경쟁 활발, 2023.3.14.
- KISTEP 사과플러스 제10호, 생성형 AI로 인한 사회변화와 대응방향, 2023.3.31.
- KIAT 산업기술정책 애자일 2023-제1호, 챗GPT, 생성형 AI가 가져올 산업의 변화, 2023.2.28.
- 삼일Pwc경영연구원, ChatGPT, 기회인가 위협인가-ChatGPT 이해와 영향 분석, 2023.3.
- 매일경제, '한국형 챗GPT' 호들갑 떨지만... 기초실력은 '부실', 2023.3.26.
- 머니투데이, 글로벌 250대 생성AI 스타트업 절반은 美.... 한국, 中·日보다 적어
- AI타임스, 텍스트 생성 및 챗봇 AI 지형도... 10개 카테고리서 700여 기업 경쟁, 2023.3.20.
- 클래리베이트 분석결과, 2023.3.

[KISTEP 브리프 발간 현황]

발간호 (발행일)	제목	저자 및 소속	비고
57 (23.01.06.)	MZ세대를 위한 미래 기술	지수영·안지현 (KISTEP)	미래예측
- (23.01.20.)	KISTEP Think 2023, 10대 과학기술혁신정책 아젠다	강현규·최대승 (KISTEP)	이슈페이퍼 (제341호)
58 (23.02.02.)	세계경제포럼(WEF) Global Risks 2023 주요내용 및 시사점	김다은·김유신 (KISTEP)	혁신정책
59 (23.02.07.)	미국의 「오픈사이언스의 해」 선포와 정책적 시사점	이민정 (KISTEP)	혁신정책
- (23.02.21.)	‘데이터 보안’ 시대의 10대 미래유망기술	박창현·임현 (KISTEP)	이슈페이퍼 (제342호)
60 (23.03.06.)	연구자산 보호 관련 주요국 정책 동향 및 시사점	유지은·김보경 (KISTEP)	혁신정책
61 (23.03.20.)	美 「과학적 진실성 정책 및 실행을 위한 프레임워크」의 주요 내용 및 시사점	정동덕 (KISTEP)	혁신정책
- (23.03.29.)	우리나라 바이오헬스 산업의 주력산업화를 위한 정부 역할 및 지원방안	홍미영·김주원 안지현·김종란 (KISTEP)	이슈페이퍼 (제343호)
62 (23.03.30.)	2021년 한국의 과학기술논문 발표 및 피인용 현황	한혁 (KISTEP)	통계분석
63 (23.03.30.)	2021년 신약개발 정부 R&D 투자 포트폴리오 분석	강유진·김종란 (KISTEP)	통계분석
- (23.04.03.)	국방연구개발 예산 체계 진단과 제언	임승혁·안광수 (KISTEP)	이슈페이퍼 (제344호)
64 (23.04.06.)	2023년 중국 양화의 주요 내용 및 과학기술외교 시사점	강진원·장지원 (KISTEP)	혁신정책
65 (23.04.10.)	2023 인공지능 반도체	채명식·이호운 (KISTEP)	기술동향
66 (23.04.13.)	생성형 AI 관련 주요 이슈 및 정책적 시사점	고윤미·심정민 (KISTEP)	혁신정책