

중국의 챗GPT 대응 동향과 전망

유재홍 선임연구원
소프트웨어정책연구소 AI정책연구실
jayoo@spri.kr



초거대 인공지능의 부상과 중국

OpenAI의 챗GPT 열풍이 거세다. 챗GPT는 작년 11월 30일에 美 OpenAI가 개발해 공개한 초거대 언어 모델(Large Language Model) 기반의 대화형 인공지능 서비스다. 텍스트로 질문을 하거나 명령을 내리면 적절한 답변을 하고 글을 만들어 낸다. 빌 게이츠는 챗GPT가 인터넷만큼 중대한 발명이라 평가를 했다. 노래 가사, 소설, 시, 에세이, 논설문, 광고 문구, 심지어 대통령 신년사까지 작성하는 챗GPT의 능력 앞에 개인, 기업, 국가들이 연일 감탄하고 있다. 최근 챗GPT, GPT-4, 초거대 AI, 초거대 언어모델, 생성AI, 범용 인공지능 이런 용어들이 언론 매체를 장식하고 있다. 공개 후 두 달 만에 1억 가입자 확보라는 인류 역사상 가장 빠른 확산, 그리고 전격적인 글로벌 유료화(챗GPT PLUS), 여기에 세 달이 되지 않아 업그레이드 버전인 GPT-4 공개까지 인공지능의 새로운 미래를 두고 논쟁이 뜨거워지고 있다.

챗GPT가 등장하기 전 중국이 인공지능 분야에서 미국을 추월하기 시작했다는 소식이 전해졌다. 스탠포드대 인간중심인공지능연구소(HAI)가 발간하고 있는 AI Index 보고서에 따르면 2021년 기준 전 세계 인공지능 출원건수의 약 52%가 중국, 미국은 17%로 중국이 압도적이다.¹ AI 논문 수 역시 중국이 가장 많았다. 2010년부터 2021년까지 전 세계에서 AI 관련 논문을 가장 많이 쓰는 국가는 중국이며 2021년 기준 약 40%로 절대 1위다. 최근

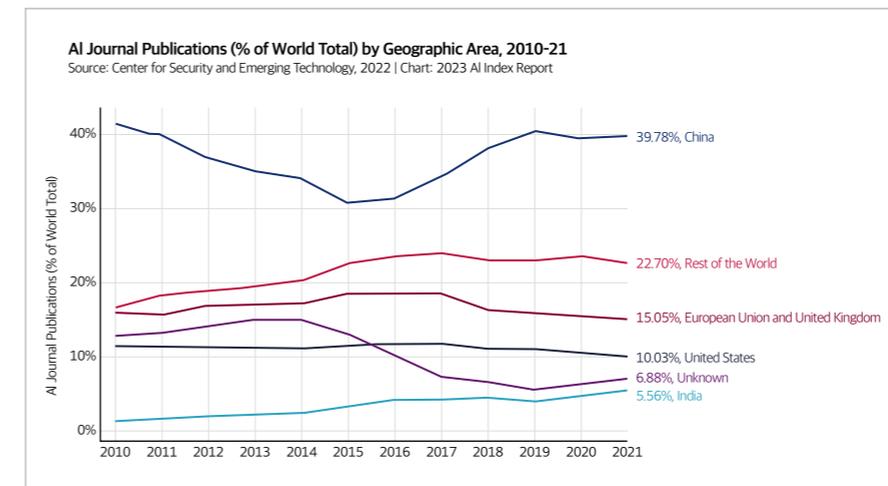
¹ Stanford HAI (2023.4), Artificial Intelligence Index Report 2023

논문의 질적 수준을 참고할 수 있는 AI 저널의 인용 수 비중으로도 2021년 기준 중국 30%, 미국 15%, 유럽이 22%로 중국이 선도하고 있다. AI 논문 출판 수 상위 10개 기관 중 美 MIT를 빼고는 중국과학원, 칭화대학교, 중국과학원대학, 상하이교통대학교, 저장대학교, 하얼빈공과대학교, 북경대학교 등 모두 중국 연구소와 대학이 차지하고 있다.

고성능 연산 인프라인 슈퍼컴퓨터 분야에서도 중국은 2017년 성능 기준 세계 1위와 2위 슈퍼컴퓨터를 보유했다. 하지만 지난해 6월부터는 미국이 대규모 정책 자금을 들여 개발한 엑사스케일 컴퓨터 프론티어가 1.102엑사플롭스(EF)(1초에 110.2경번 연산)로 1위를, 이어 2위는 일본 이화학연구소와 후지쓰가 공동 개발한 후가쿠가 차지하고 있다. 성능 면에서 미국이 43.6%, 일본이 12.8%, 중국이 10.6%를 차지하지만, 슈퍼컴퓨터의 보유 대수 기준으로는 중국 162대(32.4%), 미국 127대(25.4%), 독일 34대(6.8%) 순으로 중국이 미국을 앞서 1위를 유지하고 있다.²

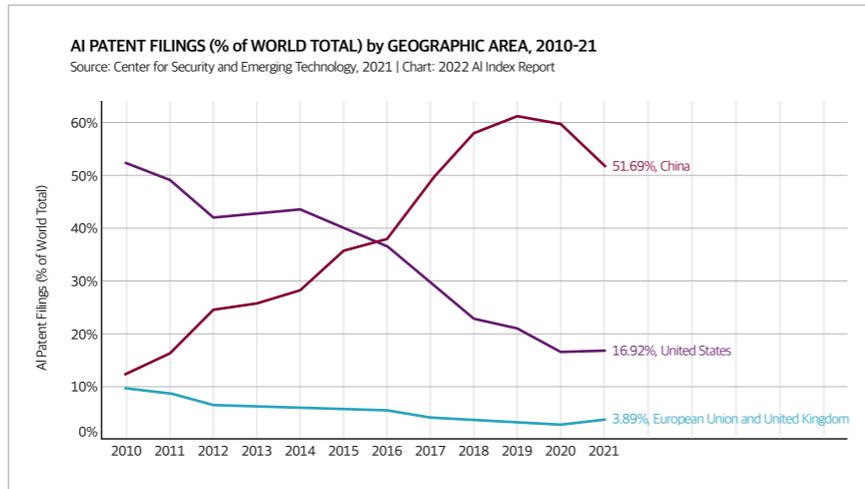
하지만, 마이크로소프트와 OpenAI의 합작품이라 할 수 있는 챗GPT가 등장한 이후로 언론은 경쟁사들이 그동안 무엇을 했는지, 무엇을 준비하고 있는지를 조망했으나 중국 기업에 대해서는 잠잠했다. 구글이 지난 2월 챗GPT의 대항마로 소개한 챗봇 '바드(Bard)'는 엉뚱한 대답을 내놓으며 '실망스럽다'라는 혹평을 받았고, 하루 뒤 구글의 시총 126조 원이

[그림 1] 지역별 인공지능 논문(저널) 출판 비중(Stanford AI Index, 2023)



² 한국과학기술정보연구원(2022.11.21), 슈퍼컴퓨터 Top500 중 한국 슈퍼컴 성능, 대수 모두 8위

[그림 2] 국가별 인공지능 특허 출원 비중(Stanford AI Index, 2023)



증발했다. 언론은 마이크로소프트(Microsoft), 구글(Google), 메타(Meta) 등 미국 빅테크의 동향을 집중적으로 조망했고, 중국의 빅테크에 대한 언급은 찾아보기 힘들었다. 오히려 중국 정부가 챗GPT에 대한 우회 접속까지 차단하는 등 민감한 정보를 선제적으로 차단하고 있다고 보도하며 혁신 기술에 대해 규제 중심적인 중국 정부의 태도를 지적했다.

미국 챗GPT가 전 세계의 커다란 관심을 불러일으킨 것은 사실이지만, 그렇다고 중국이 초거대 인공지능 기술이 뒤처져 있는 것은 아니다. 실제로 중국은 2년 전 이미 우다오(WuDao)2.0이라는 초거대 언어 모델을 공개한 바 있다. 중국 베이징인공지능연구소(BAAI, 2018년 설립)가 개발한 우다오2.0은 매개변수가 챗GPT 1,750억 개보다 10배 많은 1조 7,500억 개를 포함하고 있는 사전학습 모델이다. 구글 스위치 트랜스포머(파라미터 1조 6,000억 개)보다 더 큰 것으로, 세계 최대의 매개 변수를 가진 초거대 모델을 선보인 것이다. 우다오2.0은 대화 음성을 시뮬레이션하고, 시를 쓸 수 있으며, 그림을 이해하고 요리 레시피를 만들어 내는 기능을 선보였다. 총 4.9TB의 텍스트와 이미지 데이터를 학습했으며 여기엔 중국어와 영어 각각 1.2TB의 데이터가 포함됐다. 멀티모달을 지원하는 인공지능 기반 모델(Foundation Model)으로써 가능성을 보였으나 반향은 크지 않았다.

중국의 대표적 인터넷 기업이자 인공지능 선도기업인 바이두는 지난 3월 16일 챗GPT의 대항마로 어니봇(ERNIE Bot)을 공개했다. 어니봇은 '지식 통합을 통한 강화된 표현(Enhanced Representation from Knowledge Integration)'의 약자로, 중국어로는

원신이엔(文心一言)으로 불린다. ERNIE의 첫 번째 버전은 2019년 출시됐다.³ 다양한 중국어 사투리를 고려한 생성형 AI 모델이다. 2021년 12월에는 ERNIE 3.0 모델을 선보였으며 2,600억 개의 매개 변수를 토대로 언어 생성, 텍스트-이미지 생성 작업을 수행할 수 있다.⁴ 즉 어니봇은 2019년 시작한 AI 기반 대형 언어 모델로 현재 3.0 버전으로 발전, 매개변수 2,600억 개로 학습해 문자, 그림, 동영상을 이해하고 언어를 생성하는 모델이라 할 수 있다. 한편, 지난 3월 16일에 어니봇 공개행사에서 어니봇은 시를 창작하고 텍스트 기반 이미지와 영상을 제작했으며 수학 문제를 풀었다. 하지만 이날 바이두의 주가는 최대 10%까지 급락했는데, 이는 "아직 완벽하지 않지만 시장 수요로 이른 공개"를 결정했다는 리엔훙 회장의 발언 여파로 분석된다. 하지만 바이두는 이미 약 650개 업체들이 '어니 생태계'에 합류했으며 검색 엔진 통합, 클라우드, 스마트자동차, 가전제품에 서비스 접목을 확장할 계획임을 밝혔다.⁵ 바이두는 출시 행사 후 3만 개 이상의 기업이 어니봇 API 액세스에 신청했다고 알려졌다. 주목할 만한 점은 바이두가 인공지능 기술을 자사 보유 IT 생태계에 접목하고 있는 모습이다. 바이두는 자사가 독자 개발한 AI 칩(쿤룬), 클라우드, 자율주행플랫폼(아폴로), 오픈소스 플랫폼(바이두브레인), 딥러닝 오픈소스 프레임워크(패들패들), 음성인식 AI(두어OS)를 통해 아래에서는 AI 반도체, 위로는 각종 AI 융합 가전 애플리케이션까지 AI 생태계의 가치사슬 전반에 역량을 강화해 나가고 있다.⁶

알리바바는 그룹의 글로벌 연구 이니셔티브인 다모아카데미(DAMO Academy)에서 유사한 도구를 개발 중이다. 2022년 11월 이미 알리바바 클라우드 그룹이 글로벌 개발자 및 연구자들을 위한 대규모 사전 학습 모델인 '모델스코프(ModelScope)'를 오픈소스로 공개했다.⁷ 모델스코프 플랫폼은 DAMO에서 지난 5년간 개발한 300개 이상의 즉시 사용할 수 있는 AI 모델과 함께 컴퓨터 비전, 자연어 처리, 비디오, 오디오 관련 기술을 포함하고 있다. 특히 텍스트를 이미지로 변환시키는 통이(Tongyi)와 '원포올(One-For-All)'과 같은 멀티모달 생성 AI 모델도 포함돼 있다. 알리바바는 자체적으로 2017부터 생성 AI 분야 관련 일을 하고 있다. 알리바바는 지난 2021년 4월 멀티모달 AI 모델인 M6를 공개했는데, M6는 100억 개의 매개변수를 가지며 1.9TB의 이미지와 중국어 텍스트 292GB를 학습했고 텍스트 기반 이미지 생성, 이미지 기반 질의응답, 이미지-텍스트 매칭 등을 목적으로 파인튜닝됐다.⁸ 알리바바는 M6 기반의 챗봇을 준비 중이며 알리클라우드와 비즈니스

³ 중국전문가포럼(2023.2.2), ChatGPT 열풍! 바이두, 3월 AI챗봇 출시 예정

⁴ <https://arxiv.org/abs/2107.02137> (2021.7.5)

⁵ 한겨레(2023.3.16), 중, 바이두 AI챗봇 출시했지만... 주가 10%까지 하락

⁶ 삼성증권(2023.2.21), 바이두 차이나 생성 AI의 강자, 변화는 이제 시작일 뿐이다.

⁷ <https://modelscope.cn/home>

⁸ InfoQ(2021.4.6), Alibaba Announces 10 Billion Parameter Multi-Modal AI M6

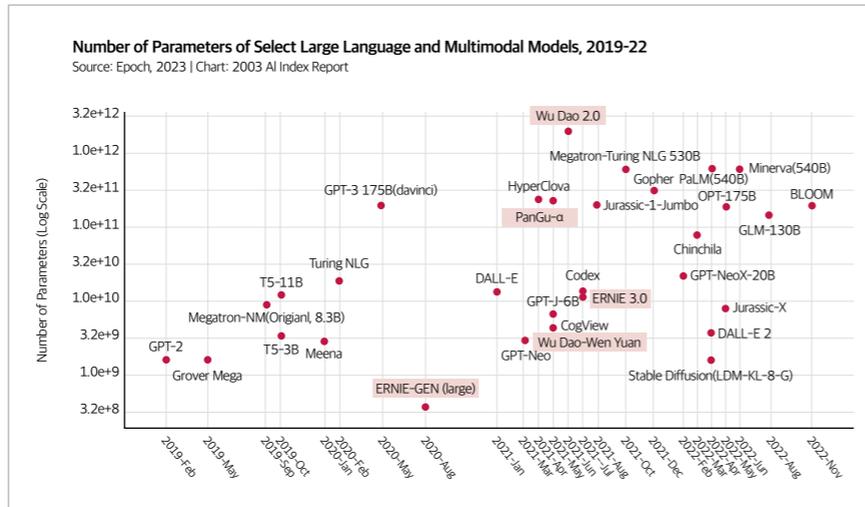
플랫폼인 딥톡에 먼저 적용할 예정⁹으로 알려진 가운데 지난 4월 11일 초거대 AI 모델인 '통이치엔원(通义千问, 천 개의 질문에 대한 답)'을 발표했다.¹⁰ 통이치엔원은 빠른 시일 내에 알리바바 생태계 내 비즈니스 애플리케이션에 통합될 계획이다.

텐센트는 위챗을 통한 챗GPT 서비스 우회 접속을 차단함으로써 중국 정부의 챗GPT 접속 금지 정책에 협력하고 있다. 한편 텐센트 역시 챗GPT 대항마를 개발할 예정인데 바로 '훈위안(HunyuanAide)'으로 알려진 AI챗봇이 그것이다. 훈위안(混元, 우주)의 AI 모델은 2022년 중국어 이해 평가 테스트에서 최고 점수를 획득한 것으로 알려졌다.¹¹ 또한, 텐센트는 텍스트 기반 비디오 생성 AI인 '즈잉(Zhiying)'을 출시할 예정임을 밝혔다.¹² 지난 3월 있었던 실적 발표 컨퍼런스 콜에서 류치핑 텐센트 회장은 챗GPT와 유사한 생성 AI 챗봇 사업화에 나설 것임을 밝혔다. 텐센트 AI는 위챗에 결합되고, 주력 사업인 게임 분야(코딩, 아트, 캐릭터 디자인 등)와 스트리밍 서비스에도 적용될 전망이다.

소위 중국의 3대 IT기업인 BAT(바이두, 알리바바, 텐센트) 외에 대표적인 중국 IT기업인 화웨이 역시 AI 개발에 적극적이다. 화웨이는 지난 2021년 4월 이미 OpenAI사의 GPT-3보다 많은 2,000억 개의 파라미터를 가진 초거대 AI 모델인 판구-알파(PanGu-α)를

선보였다.¹³ 판구-알파는 1.1TB 분량의 중국어 전자책, 백과사전, 뉴스, 소셜미디어, 웹페이지에서 추출한 언어를 학습했다. 판구 알파는 까다로운 중국어 텍스트의 요약과 질의응답, 다양한 대화 생성에 우수한 것으로 논문에서 밝혔다. 특히 화웨이는 자사 AI 컴퓨팅 프레임워크인 '마인드스포어(MindSpore)'와 AI 프로세서 '어센드(Ascend)910' 2,048개로 구성된 클러스터 컴퓨팅 환경에서 판구-알파를 훈련시켰다. 그 외 중국의 대표적 온디맨드 서비스업체 메이투안을 공동창업한 왕후이웬은 챗GPT와 유사한 프로그램에 5,000만 달러를 투자할 것으로 약속했으며, 중국 상하이 푸단대학교는 지난 2월 챗GPT 유사 챗봇의 테스트 버전을 공개했다.¹⁴ 이 프로젝트를 이끄는 추시평 교수는 아직 테스트 단계이지만 중국도 챗GPT와 같은 챗봇을 개발할 역량을 갖추고 있음을 보여줬다고 자평했으나, 현재는 서비스 접속이 불가능한 상태이며 푸단대학교에서는 관련된 입장을 내놓고 있지 않다. 한편, 상하이에 본사를 둔 세계 최고 안면인식 AI기업으로 알려진 센스타임(SenseTime)은 챗GPT가 공개된 뒤 대화형 AI모델 개발을 선언했고, 지난 4월 10일 자사 초거대 AI 모델인 센스노바(SenseNova)기반의 챗GPT 대항마 '센스챗(SenseChat)'을 공개했다. 시연회에서 센스챗은 컴퓨터 코드를 작성하거나 영어, 중국어 질문을 인식하고 실행하는 모습을 선보였다.¹⁵

[그림 3] 초거대 언어 모델 파라미터 수와 발표시기(Stanford AI Index, 2023)



9 인공지능타임즈(2023.1.11)알리바바그룹, 생성시 사전학습 멀티모달 모델 등 2023년 혁신 기술 10대 트렌드 발표

10 ZNET Korea(2023.4.11), 알리바바 클라우드, 대규모 언어모델 통이치엔원 공개

11 인공지능타임즈(2023.2.28), 텐센트, 유사 ChatGPT 개발 경쟁 합류

12 The GURU(2023.4.5), 텐센트, 텍스트 기반 비디오 생성 AI 즈잉 공개

중국의 독자 AI 생태계 구축과 미·중 기술격차

초거대 AI 기반 인공지능 생태계의 성장은 미·중 기술 패권 경쟁을 가속화시킬 것으로 전망된다. 현재 상황에서는 미국 빅테크 중심의 생태계가 빠르게 구축될 가능성이 높으며 중국이 독자적인 AI 생태계를 얼마나 효과적으로 구축해 갈 수 있는지가 향후 경쟁의 관건이 될 것이다. 중국은 통제 가능한 자국 기업 중심의 챗GPT, 범용인공지능 개발에 박차를 가할 것으로 예상된다. 이렇게 되면 학습 데이터나 AI 모델에 있어서도 중국의 정책을 반영한 데이터 학습과 모델 개발 및 파인튜닝(Fine-Tuning)이 활발하게 이뤄질 것이다. 하지만, 이러한 상황 속에 중국 AI 기업들은 중국과 글로벌로 모델을 병행하며 개발하기에는 효율성이 떨어지는 딜레마에 봉착할 수밖에 없어 미국 빅테크 기업과의 경쟁에서 기회비용 부담을 안을 수밖에 없을 것이다.

13 <https://arxiv.org/abs/2104.12369> (2021.4.26)

14 로보신문(2023.2.21), ChatGPT 열풍에 푸단대학교 가세

15 머니투데이(2023.4.10), 알리바바가 투자한 中 센스타임 챗GPT 대항마 센스챗 공개

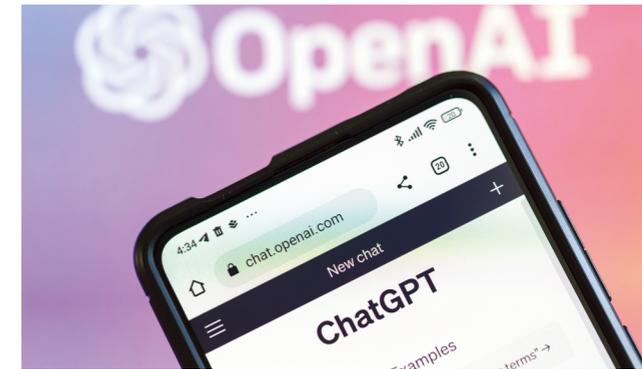
한편, 초거대 인공지능 생태계의 주축으로써 컴퓨팅 인프라의 핵심 요소인 AI 반도체 산업 육성이 주요한 과제로 부상하고 있다. NVIDIA가 관련 시장을 대부분 점유하고 있고, 구글이 2016년부터 자체 TPU(Tensor Processing Unit)를 개발해 인공지능 학습 인프라로 활용하고 있다. 마이크로소프트, 아마존(AWS) 등 클라우드 대표주자들은 엔비디아와의 협력 체계를 강화하고 있고, 구글도 국내 카카오브레인, LG AI연구원과 협력하는 등 TPU기반의 구글 클라우드 생태계를 키워가고 있다. AI 반도체는 대표적인 미국의 수출 규제 품목으로 2010년 중반부터 중국은 자체 개발해야 할 상황에 놓였다. 이에 화웨이는 2019년부터 AI 반도체를 자체 개발 생산해 독자 AI 생태계를 구축하고 있으며, 샤오미도 AI 반도체 개발을 위해 난징다위반도체를 설립했다. 알리바바는 중국 반도체 설계회사 중텐웨이를 인수해 신경망 칩인 Ali-NPU를 개발하고 있다. 바이두는 데이터센터, 클라우드, 자율주행에 적용가능한 반도체 '쿤룬'을 2018년 공개했으며 중국 정부는 2025년까지 1,500억 달러를 투자해 반도체 자급률을 70%로 끌어올린다는 목표를 제시했다.¹⁶ 그 외 중국에서는 바이렌, 캠브리켄, 알리바바 그룹 계열의 핑투게, 일루바타 코어X, 덩린 테크놀로지, 바스타이 테크놀로지, 메타X 등이 엔비디아 대안 제품 개발에 참여하고 있는 것으로 알려지고 있다.¹⁷ 하지만 반도체 분야, 특히 클라우드 시스템, 인공지능 모델 간의 소프트웨어적 통합을 통한 최적화가 중요한 인공지능 반도체 분야에서 단기간에 경쟁력을 확보하기가 어려운 만큼, 중국의 AI 반도체 생태계가 성숙해지는 동안 미국 빅테크 중심의 AI 생태계와의 기술, 서비스 격차가 발생할 가능성이 높다.

인공지능 규제의 글로벌 블록화 가능성

끝으로, 챗GPT는 기술적 한계를 가진 채 대중에게 소개됐다. 학습데이터는 2021년에 한정되고, 챗GPT가 생성한 글들은 확률적 조합에 의해 선택된 문장과 글들로 진위를 알 수 없다. 데이터 학습 과정에서 편향성이 있을 수 있으며 답변 또한 오류와 편향의 가능성을 내포하고 있다. 이는 OpenAI가 홈페이지 첫 화면에 분명하게 명시하는 바와 같으며 중국 정부가 신뢰성을 문제 삼아 초기 챗GPT 접속을 차단한 명분이기도 하다. 중국의 스타트업 위안위인텔리전트(元語智能)가 만든 '챗위안(ChatYuan)'은 공개 사흘 만에 서비스가

중단됐다. 챗위안은 우크라이나와 러시아 사이의 전쟁에 대해 '러시아가 우크라이나를 침략한 전쟁'이라고 답했는데 '양측의 군사력과 정치력 등이 동일선상에 있지 않기 때문에 러시아의 일방적인 침략전쟁으로 간주된다'라는 설명도 추가했다. 이는 중국 당국의 입장에 반하는 답변이다. 더구나 중국 경제 문제에 대해 묻자 챗위안은 '투자 부족, 주택 거품, 환경오염, 사업 효율성 저하 등 중대한 문제가 있다'고 언급했다.¹⁸

챗GPT의 유례없는 확산과 더불어 인공지능의 신뢰성과 규제에 대한 얘기가 더욱 활발해지고 있다. 유럽은 2021년 4월 초안을 마련한 「인공지능법(AI Act)」의 입법을 연내 추진할 계획인데 챗GPT와 같은 생성 AI에 대한 입장을 검토하고 있다. 미국에서도 인공지능의 신뢰성, 투명성, 공정성 확보를 위한 여론이 수면위로 부상하고 있고, 일론 머스크, 요수아 벤지오, 유발 하라리 등 저명인사를 중심으로 '거대 AI 실험을 6개월간 중지하자'라는 공개서한에 지지 서명이 전개되고 있는 상황이다.¹⁹ 중국은 향후 생성 AI 기술을 촉진하되 보다 촘촘한 검인증 체계, 신뢰성 확보, 설명 가능성 등에 대해 요구할 것으로 보인다. 실제로 지난 4월 11일 중국 사이버정보판공실(CAC)은 '생성형 AI에 대한 관리 조치'라는 지침 초안을 공개했다.²⁰ 규제 당국은 중국에서 대중에게 서비스되는 AI 제품의 훈련에서 배포까지 사실상 출시 전(全) 단계를 검열한다. 지침은 텍스트 모델뿐만 아니라 이미지, 소리, 영상, 코드 및 기타 콘텐츠 등 생성 AI 기술 전반에 적용된다. 유럽의 규제가 인공지능이 미칠 사회적, 경제적 피해 위험에 관심을 두고 규제에 접근하는데 반해, 중국은 정치적 리스크를 가장 큰 고려사항으로 보고 콘텐츠의 검열, 신뢰성, 안전성 측면에서 더욱 예의 주시할 것이다. 이러한 움직임은 한편으로는 AI신뢰성 기술을 끌어올릴 수 있는 기회가 될 수 있겠으나, 여전히 생태계의 다양성과 혁신을 저해시킬 수 있다. 향후 인공지능 경쟁은 기술과 서비스, 그리고 중국, 유럽, 미국을 중심으로 한 규제 체계의 블록화와 함께 전개될 것으로 예상된다.



¹⁶ S&T GPS(2019.8.27) 중국 ICT업계, 인공지능 반도체 독자개발 행보 가속
¹⁷ 인공지능타임즈(2022.9.12), 대중국 AI 반도체 수출 규제 역효과 유발

¹⁸ 중앙일보(2023.2.16), 러 침략 말했다 증발, 중국판 'ChatGPT' 사흘만에 서비스 종료
¹⁹ Future of Life Institute (2023.3.22), Pause Giant AI Experiments: An Open Letter
²⁰ ZDNET Korea (2023.4.11), 중국, 생성 AI 검열 시작... '공산당 비판하면 벌금'