2024.12.30.

국회미래연구원 I 연구보고서 I 24-02호

생성형 AI 혁명 시대 미래정책 연구

이승환, 한상열, 손승록, 이재윤

국회미래연구원
NATIONAL ASSEMBLY FUTURES INSTITUTE

연구보고서 24-02

생성형 AI 혁명 시대 미래정책 연구

국회미래연구원
NATIONAL ASSEMBLY FUTURES INSTITUTE

연 구 진

내부연구진

이승환 연구위원(연구책임)

자문위원

한상열 책임연구원(SW정책연구소)

손승록 매니저(KT 융합기술원)

이재윤 대표(평천교육)

발 | 간 | 사

AI 기술이 빠르게 발전하면서 우리 사회는 거대한 변화의 물결을 마주하고 있습니다. 특히 2022년 말 ChatGPT의 등장 이후, AI는 이제 먼 미래의 기술이 아닌 우리의 일상을 변화시키는 현실이 되었습니다. 교육, 제조, 국방, 공공 등 사회 전반에 걸쳐 AI 기반의 혁신이 가속화되는 가운데, 이에 대한 체계적인 정책적 대응의 필요성이 더욱 커지고 있습니다.

본 보고서는 이러한 AI 혁명 시대에 주목해야 할 세 가지 핵심 이슈를 심도 있게 다루고 있습니다. 첫째, 초거대 AI 등장 이후의 정책변화를 분석하여 규제와 진흥의 조화로운 접근 방향을 제시하였습니다. EU의 AI 법, 미국의 AI 행정명령, 영국의 AI 규제백서 등 주요국의 정책 동향을 면밀하게 검토하여, 우리나라가 나아가야 할 정책 방향을 모색하였습니다.

둘째, 생성형 AI의 발전으로 부상하는 딥페이크 문제와 그 대응 방안을 검토하였습니다. 딥페이크 기술이 가져올 산업적 혁신 가능성과 함께, 증가하는 사회적 위험에 대한 균형 잡힌 시각을 제시하고 있습니다. 특히 기술 발전 속도에 맞춘 제도적 보완과 국제 협력의 필요성을 강조하였습니다.

셋째, 세계 최초로 도입되는 AI 디지털교과서의 기대와 우려 사항을 분석하였습니다. 개인 맞춤형 학습이라는 혁신적 가능성과 함께, 교육 효과와 공정성 등 다양한 측면에서의 과제를 종합적으로 검토하였습니다. 이를 통해 미래 교육의 새로운 패러다임을 준비하는 데 필요한 정책적 시사점을 도출하였습니다.

특히 본 연구는 전문가 집단 심층 면접과 텍스트 분석을 통해 실증적 근거를 확보하고, 국내외 다양한 사례를 검토하여 실효성 있는 정책 방향을 제시하고자 하였습니다. 이러한 다면적 분석은 AI 시대를 맞아 우리 사회가 나아가야 할 방향을 모색하는 데 중요한 이정표가 될 것입니다.

AI는 우리에게 무한한 가능성과 함께 새로운 도전과제를 안겨주고 있습니다. 기술 발전의 속도가 정책 대응의 속도를 앞서가는 현시점에서, 우리는 더욱 면밀한 분석과 대응이

필요합니다. 본 보고서가 급변하는 AI 환경 속에서 정책입안자들과 각계 전문가들이 미래를 준비하는 데 유용한 참고자료가 되기를 기대합니다.

아울러 본 연구를 위해 귀중한 의견을 제시해 주신 전문가 여러분과 열정적으로 연구를 수행한 우리 연구진께 깊은 감사의 말씀을 드립니다.

2024년 12월

국회미래연구원

목 차

표 목 차

그림목차

요 약

1 서론

❑ AI 전환(Transformation)은 경제·사회·문화 전반의 패러다임 전환을 의미하며 이로 인한 정책 대응의 필요성도 증대

- **AI 기술의 급격한 발전은 여러 분야에 걸쳐 광범위한 영향을 미치는 중**
  - AI는 발전을 거듭하며 기계 학습, 딥러닝, 생성형 AI로 계속 진화
  - 교육, 제조, 국방, 공공 행정 등 다양한 분야에서 AI 기반 제품과 서비스가 개발되어 생산성 향상, 비즈니스 모델 혁신, 공공 서비스 개선에 접목
- **AI의 파급력이 확대됨에 따라 정책 대응의 필요성도 증가**
  - AI 기술의 발전 속도가 정책 및 제도의 대응에 크게 앞서고 있어, 적절한 규제와 진흥 정책의 균형을 찾는 것이 중요한 과제로 부상

❑ 국가 미래와 국회 관점에서 중요한 핵심 AI 정책 이슈를 선정 및 분석

- **현재 입법과 정책 측면에서 이슈가 제기되고 있으며 이에 중장기적으로 미래에 영향을 미칠 AI 이슈 3개를 선정하고 분석**
  - 초거대 AI 등장 이후, AI 정책변화의 특징을 분석 및 전망하고 계류 중인 AI 법과 정책 논의에 시사점을 제시
  - 생성형 AI로 빠르게 확산 중인 딥페이크 이슈를 분석하고 미래 진화 방향 및 정책 시사점을 제시
  - 세계 최초로 도입되는 AI 디지털교과서에 대한 기대와 우려 사항을 검토하고 정책 시사점을 제시

- AI 정책변화, AI 디지털교과서 이슈 관련 FGI 분석과 다양한 국내외 사례 분석, 문헌 검토 등을 통해 시사점을 도출

2 AI 혁명 시대, AI 정책변화의 특징과 전망

❑ FGI를 통해 AI 정책변화의 4대 특징을 도출하고 관련 사례를 분석

- **AI 전문가 그룹 대상 FGI를 통해 초거대 AI 등장 이후, AI 정책변화의 특징을 논의하고 이를 워드 클라우드(World Cloud)로 분석**
  - 주요 키워드로 규제 프레임워크, 산업 육성, 윤리, 보안, 특구 등이 등장

- **언급된 텍스트를 기반으로 토픽 모델링(Topic Modeling) 분석을 통해 AI 정책변화의 4가지 특징을 도출**
  - 범주화된 4가지 토픽은 ① 규제와 진흥의 조화 ② AI 안전 및 신뢰 ③ 국가별 규제 ④ 지자체 AI
  - AI 정책 방향이 규제와 진흥의 조화로 이동하며 안전·신뢰 정책이 강화되고 국가별 AI 규제 강도가 상이하며 지자체도 AI에 주목

- **4가지 특징을 기반으로 국내외 사례 분석을 추가하여 시사점을 도출**

구분	내용
AI '진흥'에서 '규제와 진흥'의 조화로	• 초기 진흥에서 초거대 AI가 부상하며 진흥과 규제의 조화방안을 모색 • EU는 2024년 최초의 AI 규제법안인 EU AI 법을 최종 승인하였고, 유럽의회는 생성형 AI를 도입하여 거대한 의회 아카이브에 다양한 참여자가 활용할 수 있도록 지원 • 미국은 2023년 AI 행정명령을 발표하고 AI 기술개발과 이용을 규제하기 위해 8개 원칙을 제시 • 영국은 2023년 AI 규제 방향과 원칙이 담겨있는 AI 규제백서를 발표하며 AI 규제 체계 기반을 마련
AI 안전 및 신뢰 정책 강화	• 위험 비례 규제 접근, 보안 강화, AI 안전연구소 설립 등을 통해 AI 안전 및 신뢰 정책을 강화 • EU AI 법은 위험 수준에 따라 AI 시스템에 대한 규제 수준을 차등화하는 위험 기반 접근

구분	내용
	• 미국의 AI 행정명령은 AI 기술의 안전 및 보안 보장을 강조 • 미국, 영국, 일본 등은 AI 안전연구소 설립 및 운영을 통해 안전과 신뢰성 확보를 위해 노력
국가별 AI 규제 강도가 상이	• EU의 AI 법은 AI 전반을 포괄하며 강도 높은 처벌규정이 포함 • 미국은 포괄적인 규제 대신 행정명령을 통해 AI의 잠재성은 극대화하고 중국 견제 등 국가안보와 거짓정보 대응에 주력 • 영국은 EU의 강력한 규제와는 달리 친 혁신적 AI 정책 접근 방식을 추진 중이며, 중국은 포괄규제 대신 새로운 AI 이슈에 대해 개별 규칙을 신속하게 제정
AI에 주목하는 지자체	• 미국 행정부의 AI 행정명령 이행과 함께, 주(州)별 AI 법제 논의가 진행 중 • 중국은 지방정부 차원에서도 AI 정책을 추진하며 지역 경쟁력을 확보 • 주요국 지자체에서 AI 도입을 통해 행정서비스 제고 방안을 모색

❑ 변화하는 AI 환경을 고려한 정책 수립 및 미래 변화에 대비

- **계류 중인 AI 법 도입 논의에 있어 변화하는 AI 환경을 종합적으로 고려**
  - 기존 AI 모델보다 매개변수가 적음에도 높은 성능을 보이는 모델이 등장
  - 이에 EU의 스케일링 법칙 기반 매개변수 기준 범용 AI 모델 규제는 미래 관점에서 성능 중심으로 재해석하여 도입하는 방안 검토
  - 빠르게 변하는 AI 생태계를 파악하기 위해 EIT(European Institute of Innovation & Technology)에서 제시한 데이터 기반의 AI 생태계 체계를 한국 상황에 맞게 재구성하여 구축하고 활용하는 방안을 논의
  - 계류 중인 다양한 AI 법에 전 국민이 AI를 알고 다룰 수 있는 AI 리터러시(Literacy)에 대한 사항은 충분히 담겨있지 못하고 있어 이를 강화

- **국회도 진화하는 AI 환경을 고려한 대응 방안을 모색**
  - 방송법과 AI 법을 분리하는 상임위 구성 방안을 논의
  - 유럽의회의 생성형 AI 도입을 통한 의회 아카이브 접근성 혁신 사례에 주목하고 국회의 생성형 AI 도입 방안을 모색

• **AI 미래정책 로드맵 구상을 통해 AI의 빠른 진화에 대비**

- AI 기금, AI 보편적 서비스, 기본소득, AI 에이전트 기반의 플랫폼지배력, 로봇세 등 향후 논의 가능성이 있는 정책 이슈를 모니터링하고 정책실험 등을 통해 미래에 대비

3 AI가 만든 가짜, 딥페이크(Deepfake)의 진화

❑ 생성형 AI로 진화하는 딥페이크

• **생성형 AI 혁명이 시작되며 딥페이크 시도가 급속히 증가**

- 알파고 대국 후 딥러닝이 주목받으며 딥페이크 검색 양이 증가했고, 이후 생성 AI 혁명 시대에 접어들며 딥페이크가 본격 확산
- 2023년 전 세계 딥페이크 시도는 전년 대비 31배 증가

• **생성형 AI 도구의 진화로 누구나 딥페이크 제작이 가능한 여건이 조성**

- 텍스트, 음성·소리, 이미지, 영상, 가상공간·인간 등 디지털 요소를 제작하는 생성형 AI 도구들이 배포되면서 딥페이크가 확산
- 생성형 AI를 활용하여 이미지나 영상을 변형하거나 목소리를 복제하는 등 다양한 디지털 합성 작업이 가능
- 생성형 AI는 일상에서 사용하는 자연어 기반의 프롬프트(Prompt)로 디지털 요소를 생성할 수 있도록 지원하여 사용자가 과거보다 낮은 비용으로 쉽고 빠르게 딥페이크 제작이 가능

❑ 딥페이크는 산업과 사회혁신 측면에서 활용 가능

• **미디어 등 다양한 분야에서 생산성을 높이고 서비스 혁신에 기여**

- 딥페이크는 미디어 제작 시간과 비용 절감, 개인화된 콘텐츠 제작에 활용
- 고령 혹은 사망 배우재현, 위험 장면 가상 연출 등 다양한 분야에 적용

- 딥페이크를 활용하여 공익 캠페인 영상을 다국어로 제작하고 배포하는 등 사회혁신 측면에서도 적용

- **생성형 AI를 활용한 딥페이크로 사업모델을 혁신하는 슈퍼 개인이 등장**
  - 인플루언서 카린(Caryn)은 자기 복제 AI를 만들어 수익모델을 창출
  - 1인 개발자가 만든 게임 '매너로드'는 세계 PC 게임 매출 2위를 달성

❑ 딥페이크의 부작용으로 피해 범위와 규모가 확대 추세

- **연예, 정치 분야에서 유명 인사를 대상으로 딥페이크가 지속 확산**
  - 세계적인 팝스타 테일러 스위프트의 딥페이크는 4,700만 회가 조회
  - 튀르키예, 슬로바키아 등 주요 선거에서도 딥페이크가 기승

- **딥페이크로 인한 피해가 기업과 일반인으로 확대**
  - 홍콩의 금융기업이 딥페이크 사기로 340억 원을 도난당한 사례가 발생
  - 일반인을 대상으로 딥페이크를 활용한 투자 사기 사례가 발생하고 학교와 직장에서도 관련한 피해 사례가 속출

❑ 생성형 AI 혁명 시대, 진화하는 딥페이크에 대한 대비가 필요

- **딥페이크 제작 주체, 방식, 피해 대상과 범위가 변화하는 현상에 주목**
  - 전문가를 넘어 이제 일반인도 자연어 기반 생성형 AI로 딥페이크 제작이 가능하며 피해 대상도 이제 유명 인사를 넘어 일반인과 기업으로 확대

- **기술적 측면에서 딥페이크 탐지 기술의 고도화가 요구되는 시점**
  - 구글의 신스 ID(Synth ID) 등 딥페이크 탐지기술 개발이 진행 중이며 이에 딥페이크 탐지기술에 대한 연구개발 확대 및 글로벌 협력 강화 필요

- **딥페이크 관련 법제도 개선 방안 모색**
  - 현재 딥페이크는 성범죄 방지법, 공직선거법에서 금지 사항을 규정하고 있으나 피해 대상과 범위가 확대됨에 따라 이에 비례하는 규제조치가 필요
  - 딥페이크 피해자의 콘텐츠 삭제율이 낮아 빠른 삭제 조치 방안을 마련

- **딥페이크의 미래 진화에 대비**
  - 딥페이크는 AI, 공간컴퓨팅, 오감기술의 융합으로 더욱 고도화될 전망
  - 화면을 넘어 초실감 몰입공간에서 생길 정책이슈에 대한 모니터링 필요

4 AI 디지털교과서 : 기대와 우려 속 미래 제언

❑ AI 디지털교과서에 대한 기대와 우려가 존재

- **개인 맞춤형 학습 및 데이터 관리, 사교육비 절감 등 기대감과 함께 교육 효과에 대한 의문, 낮은 완성도, 부작용 등 우려하는 의견도 제기 중**

❑ FGI를 통해 AI 디지털교과서 이슈를 논의하고 사례 분석을 추가

- **AI 전문가 FGI를 통해 AI 디지털교과서 이슈를 논의하고 이를 텍스트 기반 토픽 모델링 분석을 통해 핵심 이슈 8개로 구분**
  - 8가지 토픽은 ① AI 개념 혼선 ② AI 진화에 맞는 사례 분석 ③ 다양한 관점 ④ 학습효과 측정 ⑤ 공정성과 안전 ⑥ 투명한 공론화 ⑦ 법제도 개선 ⑧ 새로운 기회와 위험

- **8개 토픽 관련 다양한 사례, 연구, 전문가 인터뷰를 참고하여 이슈를 분석**

구분	내용
AI 개념 혼선	• 현재 AI 디지털교과서를 둘러싼 교육 주체들과 다수의 이해관계자가 AI 디지털교과서에 적용되는 AI에 대해 혼재된 개념을 사용 • 전통적 기계학습, 딥러닝, 생성형 AI에 대한 구분, 혼합 사용에 대한 명확한 합의가 부족한 상황에서 논의가 진행
AI 진화에 맞는 사례 분석	• 과거의 AI 교육사례가 현재 일반화되기 어려울 만큼 기술이 빠르게 진화 • 구글 딥마인드의 수학 AI는 국제수학올림피아드 은메달 수준
다양성 관점	• 교육의 목표 분류, 적용과목에 따라 다양한 AI 조합이 활용될 수 있음 • 미국 22개 주의 교육부가 AI 교육지침을 발표했으며 다양한 논의가 전개 • AI 교육에 대한 원칙에 대해서는 공통된 의견이나 세부 적용 시 차이가 있어 노스캐롤라이나주는 5학년까지 챗봇 상호작용을 미 권고 • AI 기반의 온라인 교육과 오프라인 교육의 비중 조합도 중요
학습효과 측정	• AI 디지털교과서 도입의 실제 효과 측정 및 지속 관리가 필요 • Hamsa Bastani(2024)의 연구에서는 AI를 잘못 활용할 경우 오히려 학습효과가 저하되는 현상이 발견
공정성과 안전	• 듀오링고 영어테스트(DET)의 경우 공정성과 안전을 위해 인간이 관여하는 AI(Human in the loop AI) 적용 방식을 도입 • 특정 계층, 문화에 따른 차별방지를 위해 차별적 항목기능 분석을 수행 • LLM 모델의 생성결과에 인간이 개입하여 보완 • AI와 인간이 협업하는 시험 부정행위 방지 시스템을 도입
투명한 공론화	• 노스 캐놀라이나 교육청은 생성형 AI 교육 실행 및 권고에 관한 사항을 지속 업데이트하며 공개하고 있으며 2024년에 9번의 업데이트가 이루어짐
법제도 개선	• 인터넷 혁명의 가속화로 2004년 이러닝 산업 발전 및 이러닝 활용 촉진에 관한 법률이 제정된 것처럼, AI 혁명 시대에 미래 교육을 위한 AI 교육법에 대한 논의도 필요 • AI 디지털교과서의 법적 근거 등 제도적 보완 사항 논의 필요
새로운 기회와 위험	• 미래 이슈로 AI 교육으로 야기될 다차원적 격차에 대한 열린 논의가 필요 • AI 혁명 시대에 프리미엄 아날로그 교육의 부상도 미래 이슈로 논의 • AI로 인한 교육의 초 국가화와 교육경계 붕괴 가능성, 글로벌 AI 교육 플랫폼의 지배력 강화, AI 없이는 학습 불가능한 세대 출현, AI 의존 불안장애 등 다양한 이슈에도 관심을 가질 필요

❑ AI 교육 혁신을 위한 정책개선 방안 모색이 필요

- **AI 디지털교과서에 대한 공감대 형성을 위해 정책 투명성을 제고**
  - 교과서에 적용되는 AI 세분류, 적용 비중·방식에 대한 인식의 합의 필요
  - AI 디지털교과서 도입 관련 공론화 과정을 투명하게 공개하고 업데이트
  - AI 기술 시차를 고려한 AI 교육사례 분석·공유를 통해 공감대 마련

- **데이터 기반의 AI 디지털교과서 도입 및 운영체계 마련**
  - AI 디지털교과서의 학습효과 측정·공개·관리 체계 마련
  - 인간이 관여하는 AI 관점에서 공정성·안전 문제를 데이터로 측정·관리

- **한국적 상황을 고려한 AI 디지털교과서 도입 및 법제 개선 방안 논의**
  - 대학입시와 AI 디지털교과서와의 정합성, AI 디지털교과서 도입으로 인한 사교육비 절감 효과 측정 및 관리 방안 모색
  - AI 교육 중장기 계획 수립, 실태조사, 재원 마련, 데이터 기반의 관리체계 등 다양한 이슈를 포괄하는 법제 마련 논의

- **AI 디지털교과서와 교육으로 야기될 미래 이슈 모니터링 강화**
  - AI 교육 보편화로 야기될 다차원적(학생, 교사, 지역, 학교, 가정 등) 격차 발생 이슈를 발굴하고 관리하는 방안 모색
  - 프리미엄 오프라인 교육의 부상, AI 의존성 장애 학생의 등장, AI 교육의 초 국가화로 인한 교육경계 붕괴, 글로벌 AI 교육 플랫폼지배력 강화 등 새로운 기회와 위험에 대비

제1장

서론

제1절 연구 배경 및 목적

제2절 연구 범위 및 방법

제1절 연구 배경 및 목적

최근 몇 년간 AI 기술이 급속도로 발전하여 다양한 산업과 우리 일상생활에 깊이 침투하고 있다. 특히 2022년 말부터 ChatGPT를 비롯한 생성형 AI의 등장으로 AI 기술은 새로운 국면을 맞이하게 되었다. 생성형 AI는 텍스트, 이미지, 음성 등 다양한 형태의 콘텐츠를 생성할 수 있는 능력을 보유하고 있어, 기존 AI 기술과는 차원이 다른 혁신적인 변화를 일으키고 있다. 교육, 제조, 국방, 공공 행정 등 다양한 분야에서 AI 기반 제품과 서비스가 개발되어 생산성 향상, 비즈니스 모델 혁신, 공공 분야 서비스 개선에 접목되고 있다. 이에 정부는 AI 기술개발을 촉진하기 위한 연구개발 투자, 인재 양성, 인프라 구축 등을 지원하며, 디지털 전환 가속화와 AI 활용을 통한 산업 고도화 정책을 지속 추진하고 있다. 주요국들은 AI 기술과 산업 경쟁력 우위를 선점하기 위해 적극적으로 나서고 있다. 자국내 AI 기술개발과 산업 활용 지원을 확대하고 국제적으로는 자국의 기술과 규제를 글로벌 표준으로 만들기 위해 노력하고 있다. 또한, 국가 AI 전략 수립과 전담 기구 설치 등을 통해 변화를 대비하고 있다.

AI 기술의 급격한 발전은 사회, 경제, 문화 등 여러 분야에 걸쳐 광범위한 영향을 미치고 있으며, 이에 따라 정책적 대응의 필요성도 높아지고 있다. AI 기술이 고도화 되면서 이로 인한 무분별한 개발과 사용에 대한 우려도 증가하고 있다. AI 알고리즘의 불투명성, 환각 현상 등 기술적 문제와 오류가 현실에서 발생하고 있으며, 딥페이크 등 AI의 악의적 사용, 오남용, 과도한 상업화로 인한 사회경제적 부작용도 나타나고 있다. 특히 AI 기술의 발전 속도가 정책 및 제도의 변화 속도를 크게 앞서고 있어, 적절한 규제와 진흥 정책의 균형을 찾는 것이 중요한 과제로 대두되고 있다.

이러한 배경 속에서, AI의 안전성과 신뢰성을 확보하면서도 혁신을 저해하지 않는 균형 잡힌 규제 방안에 대한 논의가 활발히 진행되고 있다. 국제 사회의 규제 동향을 살펴보면, 각국이 다양한 방식으로 AI 기술과 사용을 규제하고 안전·신뢰 기준을 도입하고 있다. EU의 AI 법이 가장 포괄적이고 강력한 규제로 평가되며 미국, 영국 등은 보다 유연

한 규제체제를 구축하고 있다. 책임 있는 AI, 신뢰할 수 있는 AI, AI 안전성 등을 위한 가이드라인과 기술개발, 인간 중심의 AI 환경조성을 위한 국제적 협력도 활발히 진행되고 있다. 한국도 「AI 법·제도·규제 정비 로드맵」을 발표하고, OECD 권고안을 반영한 「AI 윤리기준」을 수립하는 등 AI 규제 체계 마련에 노력을 기울이고 있다. 또한 「디지털 권리 장전」을 통해 AI의 보편적 가치를 제시하고, AI 기본법 제정을 위한 논의를 진행하고 있다. AI 기술의 발전과 확산은 단순한 기술 혁신을 넘어 사회 전반의 패러다임 전환을 의미한다. 따라서 특정 분야나 기술에 국한된 규제가 아닌, 종합적이고 체계적인 국가 AI 정책 체제 구축이 필요하다. 이 과정에서 AI의 위험과 안전성에 대한 고려뿐만 아니라, 산업 경쟁력 강화와 국가 발전에 기여할 수 있는 균형 잡힌 접근이 중요하다. 글로벌 동향을 주시하며 AI의 안전과 신뢰를 확보하는 다양한 규제 수단을 검토하되, 중장기적 산업 발전을 저해하지 않는 유연하고 효과적인 규제 방안을 모색해야 할 것이다.

본 연구의 목적은 생성형 AI를 중심으로 한 AI 혁명 시대에 제기되는 다양한 정책 이슈를 분석하고, 이에 따른 미래정책 방향을 제시하는 것이다. 구체적인 세부 목적은 다음과 같다. 먼저, 생성형 AI의 개념과 생태계를 체계적으로 분석하여 현재의 기술 동향과 향후 발전 방향을 파악한다. 다양한 AI 정책 이슈를 분석하고 이에 대한 시사점을 도출하기 위해서는 기본 개념과 생태계 구조, 현상에 대한 분석이 선행될 필요가 있다. 둘째, AI 정책의 변화 특징을 분석하고, 국가별 AI 정책 추진 방식의 차이점을 고찰하여 한국의 AI 정책 방향 설정에 참고할 수 있는 시사점을 도출한다. 셋째, 생성형 AI 기술의 대표적인 응용 사례인 딥페이크 기술의 진화와 그 사회적 의미를 탐구하여, 기술의 양면성에 대한 이해를 높이고 적절한 대응 방안을 제시한다. 넷째, 세계 최초로 국내에 도입 논의가 이루어지고 있는 AI 디지털교과서 관련된 기대와 우려를 분석하고, 미래 교육의 방향성에 대한 제언을 통해 교육 분야에서의 AI 활용방안을 제시한다. 마지막으로, AI 생태계와 정책변화, 딥페이크, AI 디지털교과서 이슈를 종합하고 미래를 위한 정책 방안을 논의하고자 한다. 본 연구는 이러한 다각도의 분석을 통해, 생성형 AI로 대표되는 AI 혁명 시대의 정책적 과제를 종합적으로 파악하고, 기술 발전과 사회적 가치의 조화를 이루는 미래지향적인 정책 방향을 제시하고자 한다. 또한, AI 기술이 가져올 수 있는 긍정적 변화를 극대화하고 부정적 영향을 최소화할 수 있는 균형 잡힌 접근 방식을 모색하여, 우리 사회가 AI 혁명 시대에 적절히 대응하고 새로운 기회를 창출할 수 있도록 하는 것이 본 연구의 궁극적인 목적이다.

제2절 연구 범위 및 방법

AI는 사회경제적 파급효과가 매우 광범위한 범용기술(General Purpose Technology)로 다양한 연구주제가 설정될 수 있다. 본 연구에서는 다양한 세부 주제를 다루기 위해 선행되어야 할 기본 개념과 생태계 분석을 시작으로 국회와 미래차원에서 중요한 연구주제를 선정하여 중점적으로 다루고자 한다.

먼저 AI 혁명 시대에 진입하면서 AI 정책이 어떻게 변하고 있는지 중요한 특징을 도출하고자 한다. 긴 겨울의 시기를 지나 알파고 대국 이후 주목받았던 2016년, 그리고 초거대 AI가 주목받기 시작한 2022년을 중심으로 동태적 측면에서 주요국의 AI 정책이 어떻게 변했는지 분석한다. 이는 한국에서 논의 중인 AI 기본법 제정에 있어 매우 중요한 시사점을 제공할 것이다.

두 번째로, 생성형 AI가 만들어내는 딥페이크 이슈를 분석한다. 생성형 AI를 활용하면 이제 누구나, 쉽고 빠르게 낮은 비용으로 딥페이크를 만들 수 있는 기반이 마련되고 있다. 자연어인 프롬프트로 다양한 디지털 요소를 만들고 결합하여 딥페이크 제작이 가능해져 이를 긍정적으로 활용하는 사례도 있으나, 성 착취 등 부정적인 측면에서의 활용도 최근 급격히 증가하여 이에 대한 논의가 필요한 상황이다. 딥페이크는 현재에 발생한 현안이면서 동시에 미래에도 계속 진화해 나갈 이슈이기 때문에 본 연구에서 다루고자 한다.

세 번째로 AI 디지털교과서 이슈를 검토하고자 한다. 공교육 추진에 있어 교과서는 핵심 이슈이며 세계 최초로 AI 디지털교과서를 도입하는 현시점에서 생기고 있는 기대와 우려에 대한 세부 논의를 검토하고 미래 교육을 위한 방안이 모색되어야 하는 시점이다. AI 디지털교과서도 현재의 이슈이면서 동시에 미래 교육에 지대한 영향을 미치며 AI라는 새로운 패러다임이 교육을 바꾸고 있는 현시점에서 다루어야 할 중요한 의제이다.

본 보고서는 총 6장으로 구성되어 있다. 1장은 서론에서 연구의 배경과 목적, 연구 범위와 방법을 제시한다. 2장은 생성형 AI의 개념과 생태계 분석을 다룬다. AI가 과거부터

어떻게 진화해왔으며 어떻게 분류되고 정의되는지를 검토한다. 또한, 가치사슬(Value Chain) 관점에서 AI 생태계는 어떻게 구성되고 있으며, 최근 AI 기술은 어떻게 진화하고 있는지 검토한다. 3장에서는 AI 정책변화의 특징을 도출하고 국내 AI 정책 수립 시 고려할 시사점을 제시한다. 4장은 생성형 AI가 만든 딥페이크의 진화와 의미를 탐색하고 5장에서는 AI 디지털교과서 이슈를 중점적으로 다룬다. 6장에서는 주요 이슈를 종합하고 정리하며 미래 AI 정책에서 고려할 사항을 검토한다.

본 연구는 국가 경쟁력 강화를 위한 미래 AI 정책 방안을 모색하는 것을 주요 목적으로 하며 이를 위해 최신 AI 기술과 정책에 관한 종합적인 분석을 수행한다. 연구 방법으로는 문헌 연구, 사례 연구, 비교연구, 전문가 및 실무자와의 면담, 서면 자문 등을 활용한다. 구체적으로, 주요국의 정책문서, OECD 등 국제기구와 연구기관의 전문보고서, 법 규정과 판례 등의 문헌을 검토하고 다양한 관점에서 분석을 수행한다. 최근 AI 동향과 세부 연구를 시의성 있게 파악하고 심층적인 이해를 위해 관련 전문가 인터뷰를 진행하며 이는 향후 연구확산과 연구 네트워크 강화에도 활용될 것이다. 연구 결과는 외부 전문가 검토와 세미나를 통해 검증하고 의견을 수렴하며 이러한 체계적이고 종합적인 접근을 통해 본 연구는 AI 정책에 대한 깊이 있는 이해를 바탕으로, 국가 경쟁력 강화와 혁신 촉진을 위한 균형 잡힌 AI 정책 방안을 제시하고자 한다.

제2장

AI 개요 및 생태계 분석

제1절 AI 개요

1 AI의 개념과 진화

지능(Intelligence)은 복잡한 아이디어를 이해하고, 환경에 효과적으로 적응하고, 경험을 통해 학습하고, 다양한 형태의 추론에 참여하며, 사고 과정을 통해 여러 문제를 해결하는 개인의 능력으로 정의할 수 있다.[1] 지능의 개념은 관점에 따라 다르게 정의되고 각자의 관점에서 복잡성을 지니고 있으며 지능 자체가 일관성 있게 정의되지 않은 만큼 AI의 정의 또한, 매우 다양하며 관점에 따라 다를 수 있다.[2]

AI에 대한 논의는 1940년대로 거슬러 올라간다. 인공신경망 연구의 시초는 1943년 워렌 맥컬록(Warren Mc Cullonch)과 월터피츠(Walter Pitts)의 획기적인 논문에서 찾을 수 있다. 맥컬록은 철학과 의학 분야의 학위를 보유한 신경정신과 교수였으며, 피츠는 뇌 신경 시스템에 특별한 관심을 가진 수학자였다. 이들의 핵심 아이디어는 단순한 전기 스위치와 같이 '온'과 '오프' 상태를 가진 인공 신경 단위를 네트워크 형태로 연결하면, 인간 두뇌의 기본적인 기능을 모방할 수 있다는 것이었다.[3] 이 이론적 모델은 인공신경망의 기초가 되었으며, 뇌의 작동 원리를 컴퓨터 시스템으로 구현하려는 시도의 출발점이 되었다. 맥컬록과 피츠의 연구는 실험실에서 뇌의 기능을 모델링하기 위한 이론적, 실험적 연구의 토대를 마련하였다. 비록 그들의 초기 모델은 매우 단순한 선형 구조로 한계가 있었으나, 이는 후속 연구자들에게 영감을 주어 퍼셉트론과 같은 더 복잡하고 정교한 인공신경망 모델의 개발로 이어졌다. 이러한 선구적 연구는 현대 AI와 기계 학습 분야의 근간을 형성하였으며, 오늘날 딥러닝과 같은 고도화된 인공신경망 기술의 발전에 중요한 이론적 기반을 제공하였다.

1) Boykin, A. et al. (1996). "Intelligence: Knowns and unknowns". American Psychologist. 51. pp.77~101.
2) 경제인문사회연구회(2024), "인공지능 시대의 경쟁력 강화를 위한 AI 규제 연구".
3) McCulloch, W.S. and Pitts, W.H. (1943) A Logical Calculus of the Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics, 5, 115-133.

1950년에 영국 수학자 앨런 튜링(Alan Turing)은 '계산 기계와 지능(Computing Machinery and Intelligence)'이라는 논문에서 기계가 생각할 수 있는지 테스트하는 방법, 지능적 기계의 개발 가능성, 학습하는 기계에 대해 논하였다. 튜링은 사람이 인간과 기계의 답을 두고 구분할 수 없을 정도로 인간적인 응답을 할 수 있는지를 기준으로 AI의 발달 정도를 파악하고자 하였다. 앨런 튜링은 기계는 생각할 수 있다고 주장하며, 이를 테스트하는 방법으로 '튜링 테스트(The Turing Test)'를 고안했고 이는 AI라는 개념을 본격적으로 제시한 연구로 꼽힌다. 앨런 튜링은 AI 자체에 대한 정의를 내렸다기보다는 AI를 테스트 방법인 튜링 테스트, 지능 기계 개발의 가능성, 학습하는 기계를 설명하며 AI의 가능성을 제시한 것이다.[4] 앨런 튜링의 연구에 기반하여 존 폰 노이만 교수는 컴퓨터의 아키텍처를 제안하였고 이는 현대 컴퓨터 구조의 표준이 되었다.[5]

1956년 존 매카시와 마빈 민스키가 AI라는 신조어를 그들이 참여하는 다트머스 콘퍼런스(Dartmouth Conference) 제안서에 처음으로 사용하였다. 이 용어는 당시의 '사이버네틱스'라는 용어를 사용하는 기존 시각에서 벗어나는 동시에 자신들의 연구를 차별화시키기 위한 시도였다. 이들이 제안한 AI 개발 연구 제안서에서는 음성인식, 문자인식, 체스, 논리 추론 등을 AI의 예로 들었는데, 이들 모두 사이버네틱스의 한 분야로 여겨졌던 분야였다. 존 매카시는 이때 AI를 "고도의 지능을 가진 컴퓨터 디바이스를 만들기 위한 공학과 과학"으로 정의하였다.[6]

이 시기에는 인공신경망(Artificial Neural Network) 모델에 관한 연구가 활발히 진행되었는데 1958년 코넬대 심리학자 프랑크 로젠블랏(Frank Rosenblatt)의 연구에서 뇌 신경을 모사한 인공신경 뉴런 퍼셉트론(Perceptron)이 탄생하였다. 연구모델을 통해 컴퓨터가 패턴을 인식하고 학습할 수 있다는 개념을 실증적으로 보였으며 이는 1943년에 워렌 맥컬록(Warren Mc Cullonch)과 월터피츠(Walter Pitts)이 제시한 신경망 이론을 실제 테스트에 활용한 것이다. 하지만 1969년 마빈 민스키와 세이무어 페퍼트는 저서를 통해 퍼셉트론은 AND 또는 OR 같은 선형 분리가 가능한 문제는 가능하지만, XOR문제에는 적용할 수 없다는 것을 수학적 증명으로 발표했고 이에 따라 미국방부 DARPA는

4) Turing, A. M. (2021). Computing Machinery and Intelligence (1950).
5) Mühlenbein, H. (2009). "Computational Intelligence: The Legacy of Alan Turing and John von Neumann. In: Mumford, C.L., Jain, L.C. (eds) Computational Intelligence". Intelligent Systems Reference Library. Vol 1. Springer. Berlin. Heidelberg.
6) 경제인문사회연구회(2024), "인공지능 시대의 경쟁력 강화를 위한 AI 규제 연구".

AI 연구자금 2천만 달러를 지원을 중단했다. 또한, 1973년 제임스 라이트힐(James Lighthill) 경이 영국 의회에 영국 대학의 AI 연구비 축소를 권고한 보고서를 의회에 보고하였는데 "폭발적인 조합증가(Combinational explosion)를 AI가 다룰 수(Intractability) 없다"게 이유였다. 실제로 1960년대에 AI가 복잡한 결정을 해결하는데 주로 쓰인 의사결정 트리(decision tree)를 통한 검색 모델은 그 조합의 수가 기하급수적으로 증가하는 문제를 가지고 있었다.[7] 이러한 이유로 사실상 AI 관련 대규모 연구는 중단되어 AI 연구는 암흑기에 접어들게 된다.

침체기에 있던 AI 연구는 1980년대 산업계에 전문가 시스템(Expert system)이 도입되며 다시 조명을 받았다. 사람이 입력한 규칙을 기반으로 자동 판정을 내리는 '전문가 시스템(Expert System)'이 등장했는데 전문가 시스템은 특정 분야의 전문 지식을 바탕으로 문제를 해결하는 인공지능 시스템이었다. 전문가 시스템은 의료, 금융, 제조 등 다양한 분야에서 진단, 분류, 분석 등의 기능을 수행하며, AI 연구에 관심을 불러일으켰고 당시 미국 500대 기업 중 절반 이상이 전문가 시스템을 활용하였으며, 이를 위해 지속적인 투자를 하였을 정도로 높은 관심을 받았다. 하지만 전문가의 지식을 추출하여 프로그램에 입력하기 어렵고 전문가의 지식을 획득하기 난해하며 인간이 사전에 설정한 규칙에만 의존하여 작동하여 복잡한 현실 세계를 이해하는데는 한계가 있었다.

인간의 명령으로만 작동하던 전문가 시스템의 한계로 다시 침체기로 들어선 AI 연구는 1990년대 들어서 스스로 규칙을 찾아 학습하는 '머신러닝(Machine Learning)' 알고리즘을 활용하면서 다시 주목받기 시작한다. 인터넷 시대로 접어들면서 웹에서 수집한 대량의 데이터를 활용할 수 있게 되었고 AI는 스스로 규칙을 학습하고 나아가 사람이 찾지 못하는 규칙까지 발견하게 되었다. 또한 침체기 속에서도 인공신경망 연구는 지속되며 새로운 국면에 접어들게 되는데 1986년, 2024년 노벨 물리학상을 수상한 제프리 힌튼 교수는 인공신경망을 여러 겹 쌓은 다층 퍼셉트론(Multi-Layer Perceptrons) 이론에 역전파(Backpropagation)[8] 알고리즘을 적용하여 기존의 퍼셉트론 문제를 해결할 수 있음을 증명했으나 신경망의 깊이가 깊어질수록 학습 과정과 결과에 이상이 나타나는 한계가 존재했다. 2006년, 제프리 힌튼 교수는 다층 퍼셉트론의 성능을 높인 심층 신뢰 신경망

7) KAIT(2019), "영국의 AI 산업 육성 전략"
8) 신경망에서 출력 값과 실제 값 사이의 차이를 계산하고, 오차를 줄이기 위해 출력부터 시작하여 역순으로 가중치를 조절하는 알고리즘

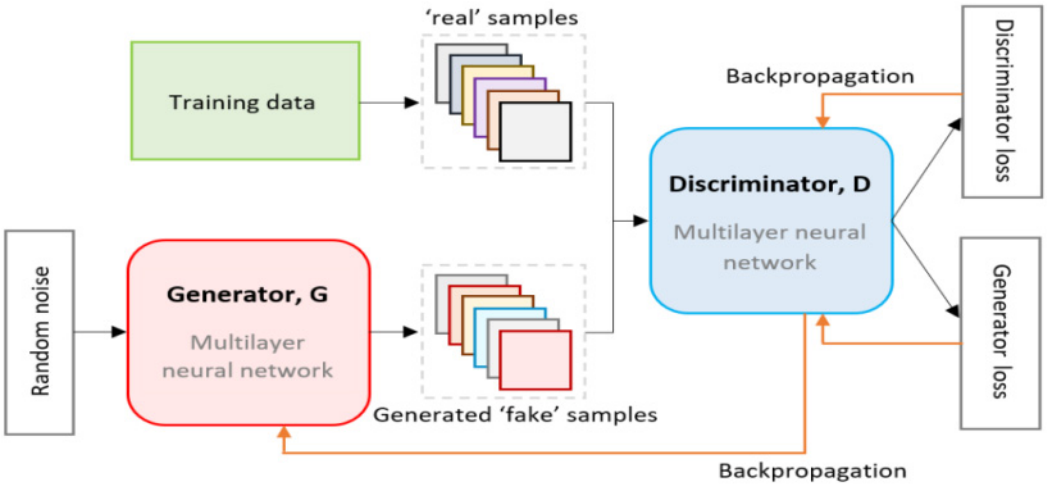
(Deep Belief Network, DBN) 개념을 제한 논문 'A fast learning algorithm for deep belief nets'을 발표했다. 심층 신뢰 신경망은 비지도 학습(Unsupervised Learning)[9]을 통해 각 층을 사전 훈련한 후, 전체 네트워크를 미세 조정하는 방식으로 신경망의 학습 속도와 효율성을 크게 높였고 이러한 노력을 기반으로 '딥러닝(Deep Learning)'이 AI 기술을 대표하는 알고리즘으로 자리매김하기 시작했다.

딥러닝의 우수한 성능을 증명한 역사적인 사건은 2012년에 발생했다. 당시 이미지 인식 경진대회인 ILSVRC(ImageNet Large Scale Visual Recognition Challenge)에서 제프리 힌튼 교수가 이끈 팀이 개발한 딥러닝 기반 모델인 알렉스넷(AlexNet이 우승을 차지하며 딥러닝의 강점이 부각되었다. 알렉스넷은 이미지 인식률 84.7%를 기록하였으며, 이는 당시 다른 모델에 비해 매우 높은 수준이었다. 특히, 전년도 우승팀의 오류율이 25.8%였던 반면, 알렉스넷은 이를 16.4%로 낮추며 기존 접근 방식의 한계를 뛰어넘는 성과를 보여주었다. 이 사건을 통해 AI 연구에서 딥러닝이 주요한 위치를 차지하게 되었고, 2010년대부터 딥러닝은 빠르게 성장하였다. 딥러닝 기술의 급격한 성장은 두 가지 주요 요인에 의해 가능해졌다. 첫 번째 요인은 GPU(Graphics Processing Unit)를 포함한 컴퓨터 시스템의 발전이다. GPU는 원래 컴퓨터의 그래픽 처리를 위해 설계된 장치이지만, CPU(Central Processing Unit, 중앙처리장치)와 비교했을 때 유사하고 반복적인 연산을 병렬로 처리할 수 있어 계산 속도가 훨씬 빠르다. 2010년대에 접어들며 GPU의 활용 범위가 인공신경망 학습으로 확장되었다. 딥러닝 학습 과정에서는 방대한 데이터에 대한 반복적 연산이 필수적인데, GPU의 병렬 처리 구조는 이러한 작업을 효과적으로 수행하여 딥러닝 발전에 큰 기여를 했다. 두 번째 요인은 데이터의 폭발적 증가이다. 인공신경망의 학습에는 대량의 데이터가 필요하다. 과거에는 컴퓨터에 입력된 데이터의 양이 제한적이었으나, 1990년대 이후 인터넷의 보급과 웹의 확산, 검색엔진의 발전으로 가공할 수 있는 데이터의 양이 기하급수적으로 증가하였다. 2000년대 이후에는 스마트폰 보급 확산으로 데이터 사용은 더욱 늘어나게 되었고 방대한 데이터가 딥러닝 학습에 사용되었다.

2014년에는 이안 굿펠로우(Ian Goodfellow)가 발표한 'GANs(Generative Adversarial Networks, 생성적 적대 신경망)' 모델이 주목받았다. GANs는 두 신경망이 서로 경쟁하

9) 입력 데이터에 대한 정답을 주지 않고, 숨은 구조나 패턴 등을 발견하고 이해할 수 있게 학습시키는 방법

면서 학습하는 구조다. 한 신경망은 실제 데이터와 구분하기 어려운 새로운 데이터를 생성하고 다른 신경망은 이를 실제 데이터와 비교하여 판별하는데, 이 과정을 반복하며 점점 더 정교한 데이터를 완성한다.[10] 이안 굿펠로우는 GANs을 NIPS(Neural Information Processing Systems)에서 발표하며 생성자를 위조지폐범에 식별자를 경찰에 비유했다. GANs 모델은 이후 변형과 개선을 통해 현재까지 이미지 생성 및 변환 등 다양한 응용 분야에서 활발하게 사용되며 생성형 AI와 딥페이크 확산에 중요한 역할을 했다.



[그림 2-1] GANs 아키텍처

자료: Claire Little et al(2021), "Generative Adversarial Networks for Synthetic Data Generation: A Comparative Study", DOI:10.48550/arXiv.2112.01925

딥러닝의 또 다른 중요한 사건은 2016년에 발생하였다. 구글 딥마인드(Google DeepMind)가 개발한 AI 알파고(AlphaGo)가 바둑기사 이세돌 9단과의 대국에서 4승 1패로 승리하며 전 세계에 AI의 잠재력을 각인시킨 것이다.

2017년에는 딥러닝 모델 중 하나인 자연어 처리(Natural Language Processing) 모델 '트랜스포머(Transformer)'가 발표된다. 트랜스포머는 데이터 간의 관계를 중요 변수

10) Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, "Generative Adversarial Networks"

로 고려한다. 특정 정보에 더 많은 '주의'를 기울여 데이터 사이의 복잡한 관계와 패턴까지 학습할 수 있으며, 더 중요한 정보를 포착해 이를 기반으로 더 나은 품질의 결과물을 생성할 수 있다. 트랜스포머 모델은 언어 이해, 기계 번역, 대화형 시스템 등의 자연어 처리 작업에 혁신을 가져왔다. 딥러닝 모델은 다양한 세부 모델을 거쳐 진화했으며 트랜스포머와 GAN은 생성형 AI 혁명을 주도하는 핵심 모델이 된다.

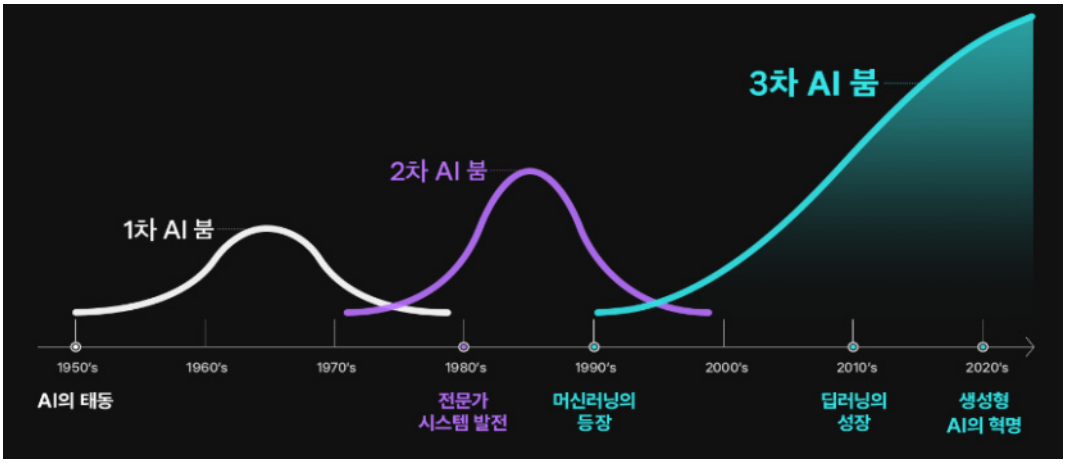
딥러닝(Deep-Learning)			
CNN (Convolutional Neural Network)	RNN (Recurrent Neural Network)	GAN (Generative Adversarial Network)	Transformer
• 1980년대 등장한 개념으로 Convolutional Layer를 거치며 부분별 이미지가 가지는 패턴을 파악하고 데이터를 분석함 • 패턴분석을 통한 영상과 이미지를 인식하는데 특화	• 1980년대 등장한 개념으로 데이터의 입력과 출력을 순환적 구조로 해결 • 문장의 흐름을 이해하고 단어별 분류 기능을 탑재하여 자연어 처리에 특화	• 2014년에 발표된 개념으로 정보를 생성하는 기능과 판별하는 기능이 서로 대립하여 기능을 향상시키는 모델 • 사전 훈련된 데이터를 기반으로 주어진 데이터를 판별 및 실제와 가장 유사한 정보를 생성	• 2017년 구글이 공개한 모델로 단어의 의미와 문장 내 단어 간 관계를 분석 • 언어의 구조적 특성을 이해하고 언어를 생성하는 데 특화

[그림 2-2] 딥러닝의 발전단계

자료: 삼정KPMG(2024) "생성형 AI에게 펼쳐진 새로운 무대, 온디바이스 AI"

트랜스포머를 기반으로 오픈AI는 2018년 처음 GPT(Generative Pre-trained Transformer)를 출시하고 매년 더 많은 매개변수와 학습 데이터를 사용해, 빠른 속도로 성능을 개선해 왔다. 그리고 마침내 2022년 11월, 오픈AI가 LLM(Large Language Model, 거대 언어 모델)[11] GPT 3.5를 탑재한 '챗GPT'를 출시하면서 생성형 AI(Generative AI)의 시대가 열리며 본격적인 AI 혁명이 시작되었다. 위에서 언급한바 AI는 발전과정에서 2번의 침체 시기를 지나 새로운 국면으로 진입하며 본격적인 AI 혁명 시대가 열리고 있다.

11) 방대한 양의 데이터를 통해 얻은 지식을 기반으로 다양한 자연어 처리 작업을 수행하는 딥러닝 알고리즘

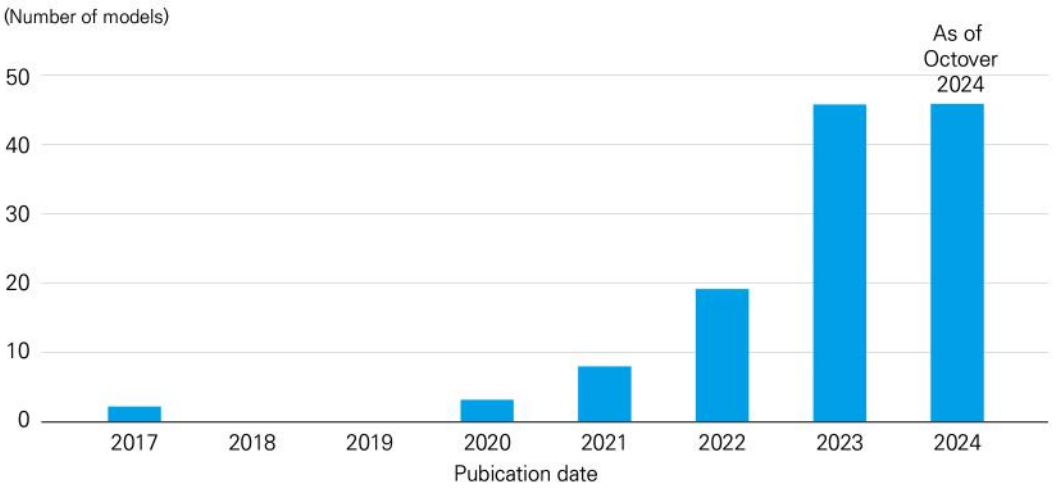


[그림 2-3] AI의 발전과정

자료: SK하이닉스 뉴스룸(2024.3.15.) "AI의 시작과 발전과정, 미래 전망"

2023년 오픈AI는 GPT-4를 발표하고, 2024년 'GPT-4o', 'o1'을 발표하며 LMM(Large Multi Modal)로 진화하고 있고 고도화된 AI 알고리즘, 폭발적으로 증가하는 데이터, 빠른 연산을 지원하는 GPU 등 하드웨어의 결합으로 수많은 초거대 AI 모델이 지속 출시되며 경쟁 중이다.

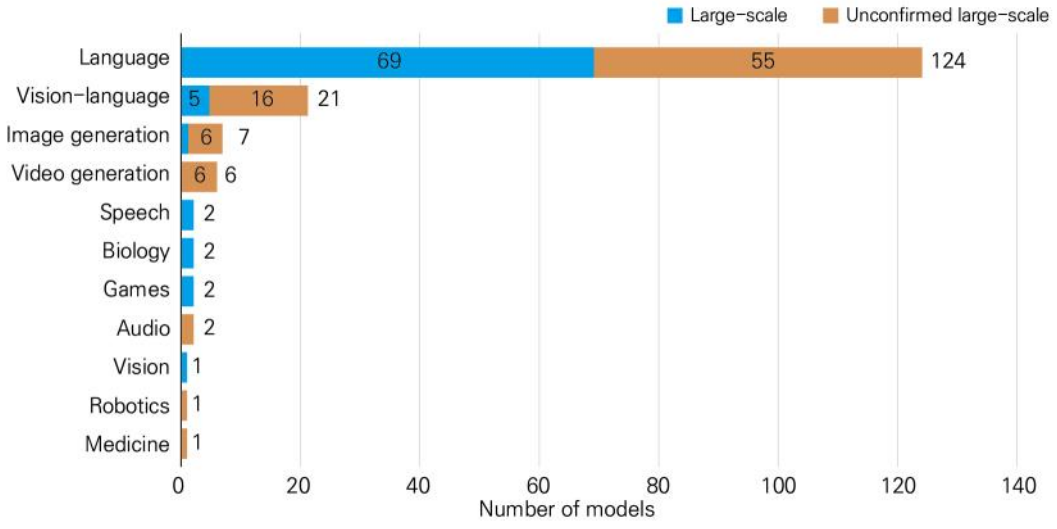
2017년에는 단 두 개의 모델만이 훈련 연산량 10^{23} FLOP를 초과했으나, 2020년에는 5개로 증가했고, 2022년에는 32개, 2024년에는 124개 모델이 이 기준을 초과한 것으로 확인되었다. FLOP는 '부동소수점 연산(floating-point operations per second)'의 약자로, 컴퓨터가 수학적 계산을 얼마나 빠르게 할 수 있는지를 나타내는 단위이다. 또한, 훈련 연산량이 10^{23} FLOP를 초과할 가능성이 높은 모델도 96개에 이르는 것으로 보고되었다. 2022년 이후 모델 수는 빠르게 늘어나고 있으며 AI 투자 증가와 훈련 하드웨어의 비용 효율성 향상으로 인해, 이와 같은 규모의 모델 개발은 점점 더 많은 개발자에게 접근 가능해지고 있다.



[그림 2-4] 연도별 출시된 초거대 AI 모델 수(10^{23} FLOP 이상)

자료: https://epochai.org/data/large-scale-ai-models

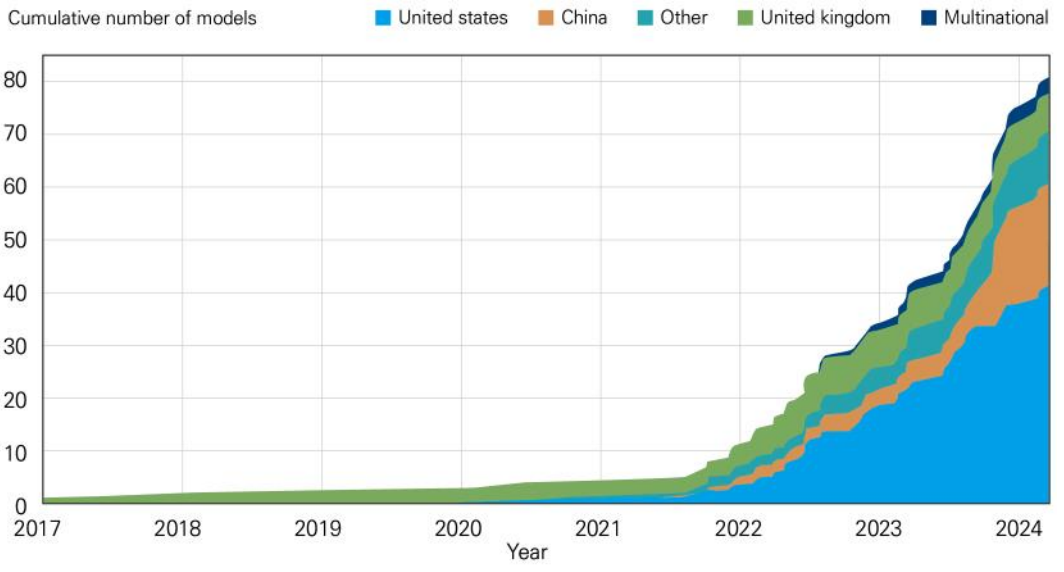
현재 초거대 AI 모델 대부분은 언어(Language) 모델이며 비전(Vision), 이미지(Image) 등 세부 특화 분야로 지속 발전해 나갈 전망이다.



[그림 2-5] 분야별 초거대 AI 모델 수

자료: https://epochai.org/blog/tracking-large-scale-ai-models

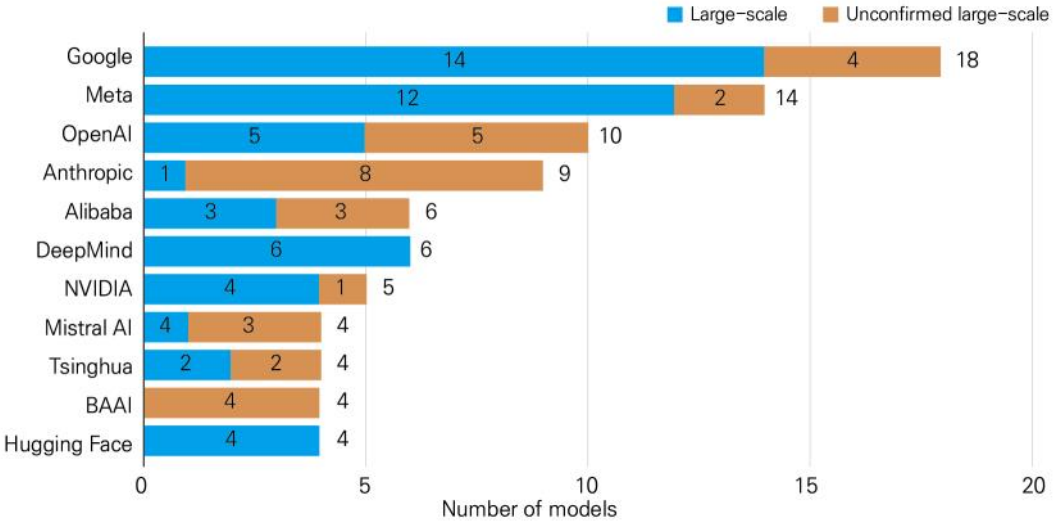
초거대 AI 모델은 미국과 중국이 주도하고 있으며 초기 영국이 보유 수가 1위였으나 2021년 이후 판도 변화가 생기며 미국의 수가 급격히 늘어났고, 이후 빠르게 중국이 추격하고 있는 양상이다.



[그림 2-6] 국가별 초거대 AI 모델 누적 수

자료: https://epochai.org/blog/tracking-large-scale-ai-models

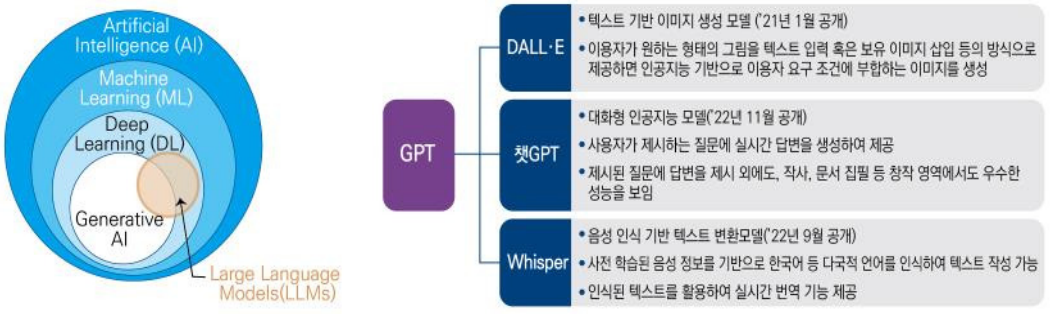
2024년 4월 기준, 기업별로는 구글이 가장 많은 초거대 AI 모델을 보유하고 있으며 메타가 그 뒤를 잇고 있다. 주요 글로벌 기업이 초거대 AI를 보유하고 이를 활용하여 새로운 혁신 모델을 개발 중이다.



[그림 2-7] 기업별 초거대 AI 모델 보유 수('23.4 기준)

자료: https://epochai.org/blog/tracking-large-scale-ai-models

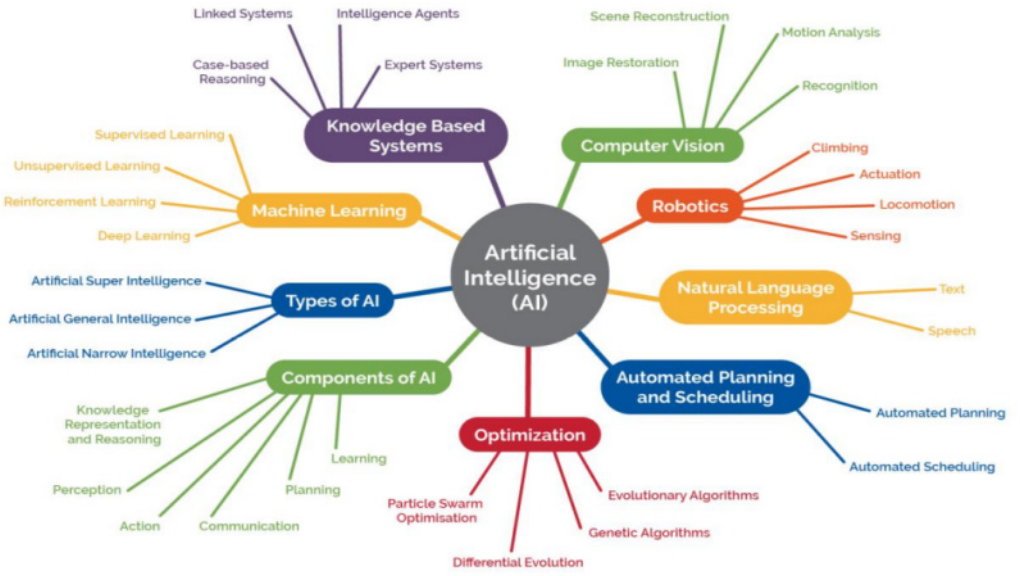
AI는 발전을 거듭하며 머신러닝, 딥러닝, 생성형 AI로 계속 진화하고 있다. GPT는 딥러닝 기반의 트랜스포머로 개발된 초거대 AI인 대형언어모델(LLM)이며 GPT 기반의 채팅 시스템으로 사용자가 원하는 텍스트나 이미지 등을 생성하는 생성형 AI 서비스가 챗GPT이다.



[그림 2-8] AI 분류(좌)와 GPT 기반 생성형 AI 서비스(우)

자료: Omer Shahab et al(2024) "SoroushLarge language models: a primer and gastroenterology applications" Ther Adv Gastroenterol 2024, Vol. 17: 1-15; 삼정KPMG 경제연구원

AI는 알고리즘의 진화적 접근 외에도 목적에 따라 다양한 형태로 분류될 수 있다. 지능 수준에 따라 약인공지능(Artificial Narrow Intelligence), 범용인공지능(Artificial General Intelligence), 초인공지능(Artificial Super Intelligence)으로 구분할 수 있다.

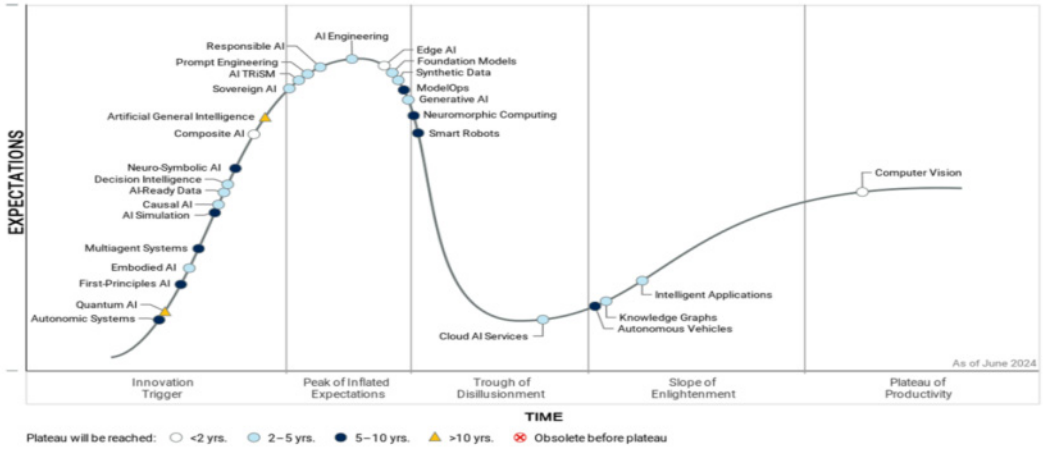


[그림 2-9] 컴포넌트, 형태, 세부분야에 따른 AI 범주

자료: Regona, Massimo & Yigitcanlar, Tan & Xia, Bo & Li, R.Y.M. (2022). Opportunities and adopti on challenges of AI in the construction industry: A PRISMA review. Journal of Open Innovati on Technology Market and Complexity, 8(45).

AI는 진화과정에서 하이프 사이클(Hype Cycle for AI)의 단계를 거치게 되며, 혁신 촉발(Innovation Trigger) 단계에서는 아직 상업적으로 적용되지 않고 연구와 개발 초기 단계에서 많은 관심을 받는 기술들이 포함된다. AGI(Artificial General Intelligence), 복합 AI(Composite AI), 멀티에이전트 시스템(Multi-agent Systems), 양자 AI(Quantum AI) 등이 여기에 해당하고 이들은 아직 초기 단계에 있지만 잠재력이 있는 AI 기술이다. 기대의 정점(Peak of Inflated Expectations) 단계에서는 특정 AI 기술들이 과장된 기대 속에서 관심을 끌며 정점에 도달한다. AI 공학(AI Engineering), 책임있는 AI(Responsible AI), 프롬프트 엔지니어링(Prompt Engineering), AI 신뢰 및 위험관리(AI TRiSM) 등

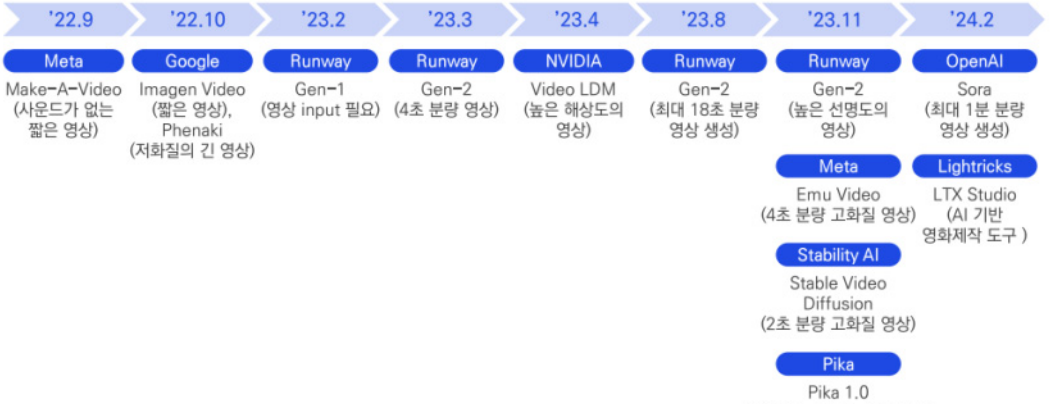
이 포함되어 있으며, 이는 AI 개발과 적용이 기업 및 산업 전반에 걸쳐 활발하게 논의되고 있다는 것을 보여준다. 환멸의 골짜기(Trough of Disillusionment) 단계에서는 기술의 과대광고가 줄어들면서 초기 기대에 비해 성과가 부족한 기술들이 환멸을 겪는다. 여기에는 클라우드 AI 서비스(Cloud AI Services)가 포함되며, 생성형 AI는 현재 과대광고가 줄어드는 환멸의 골짜기 단계(Trough of Disillusionment)에 있고 초기 높은 기대와 다르게 실질적인 도입과 성과에 어려움을 겪고 있음을 의미한다. 이를 극복하는 과정에서 실질적인 이점이 나타나고 있으며, 향후 몇 년 동안 기능이 빠르게 발전할 것으로 분석되었다. 생성형 AI가 텍스트, 이미지, 오디오, 비디오 등 다양한 유형의 데이터를 학습하도록 진화하면서 생산성을 높이고 새로운 비즈니스 모델 형성할 것으로 전망된다. 계몽의 경사(Slope of Enlightenment) 단계에서는 기술의 성과와 가능성이 다시 평가되면서 점차 개선되고 안정화된다. 지식 그래프(Knowledge Graphs), 자율 차량(Autonomous Vehicles), 지능형 애플리케이션(Intelligent Applications)이 포함되어 있으며, 이는 AI가 실제 환경에서 유의미한 성과를 내기 시작하는 단계이다. 생성형 AI는 이 단계에서는 산업에 특화된 애플리케이션에 적용될 가능성이 있다. 생산성의 고원(Plateau of Productivity) 단계에는 기술이 상업적 가치와 적용성을 충분히 확보하여 주류로 자리매김하는 단계이다. 컴퓨터 비전(Computer Vision)이 이 단계에 속한다.



[그림 2-10] AI 하이프 사이클(Hype Cycle) 2024

자료: Gartner(2024)

분야별 생성형 AI 서비스는 빠르게 진화하고 있는데 영상분야 생성형 AI의 경우 영상 분야 생성형 AI의 경우 2022년에는 소리가 없는 영상을 생성했고, 이후 2023년 최대 18초 분량의 영상을 생성했으나 2024년 들어서며 고화질 영상을 1분 이상 생성할 수 있게 되었고 최근 이러한 변화는 더욱 빨라지는 추세이다.



[그림 2-11] 생성형 영상 모델의 주요 변천사

자료: CB insight, 삼정KPMG

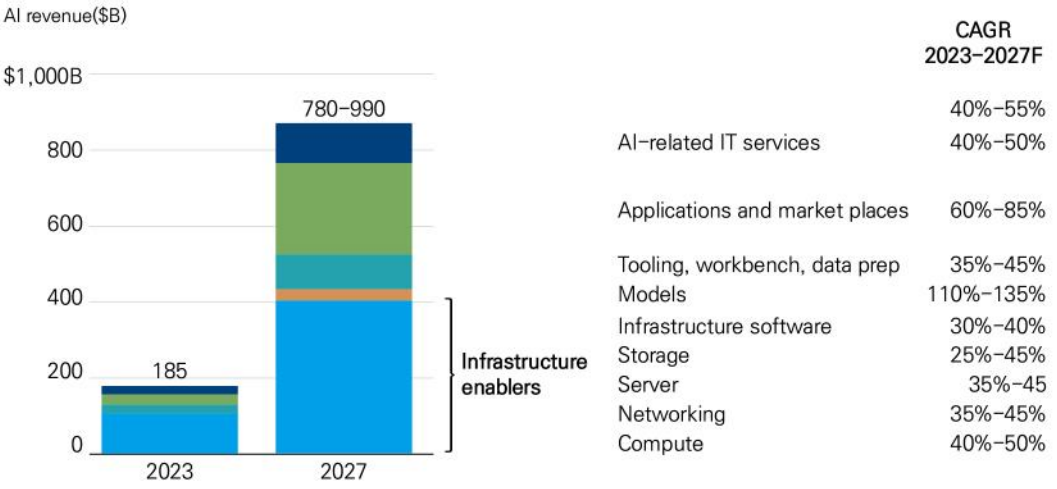
2 AI의 파급효과

글로벌 AI 시장은 매년 40~55%씩 급성장하며 2027년까지 7,800억 달러에서 최대 약 1조 달러 규모에 도달할 것으로 전망되고 있다. 2023년 AI 관련 수익은 1,850억 달러였으며, 인프라 관련 기술을 포함한 여러 부문에서 빠르게 성장이 예상된다. 2023년의 5배가 넘는 규모로, 특히 AI 모델 부문은 110%에서 135%의 매우 높은 성장률을 보일 것으로 전망되며, 컴퓨팅은 40%에서 50%, 서버와 네트워킹은 각각 35%에서 45% 성장할 것으로 예측된다.[12)]

AI 시장의 성장은 AI 관련 하드웨어와 소프트웨어의 수요 증가로 인해 견인되고 있으며, AI 기술이 단순한 애플리케이션 수준을 넘어, 인프라와 서비스를 포함한 전체 생태계

12) Bain&Company(2024) "AI's Trillion-Dollar Opportunity"

로 확장되고 있음을 시사하고 있다.

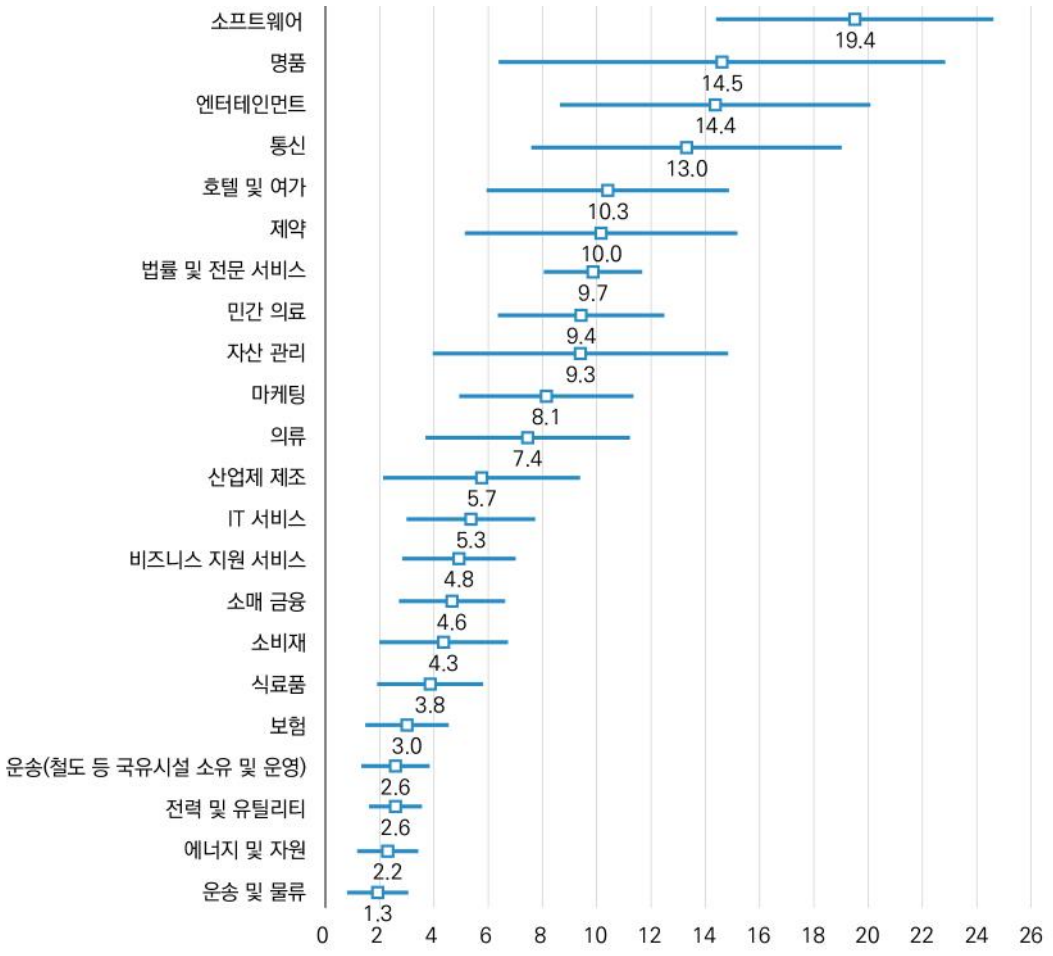


[그림 2-12] 글로벌 AI 시장 전망

자료: Bain&Company(2024) "AI's Trillion-Dollar Opportunity"

PwC는 산업과 조직에 따라 생성형 AI의 잠재적 영향이 상당히 다르다는 사실을 발견했다. 분석에 따르면, 생성형 AI를 현재 운영에 적용하면 소프트웨어 기업들의 영업이익률이 약 20% 증가하는 것으로 나타났다. 운송 및 물류와 같이 예상 영업이익이 훨씬 낮은 산업에서도 1%p의 영업이익률 증가를 기대해 볼 수 있는 것으로 조사되었다.[13)]

13) PwC(2024), "생성형 AI를 통한 가치 창출의 경로: 플라이휠(Flywheel) 접근법"

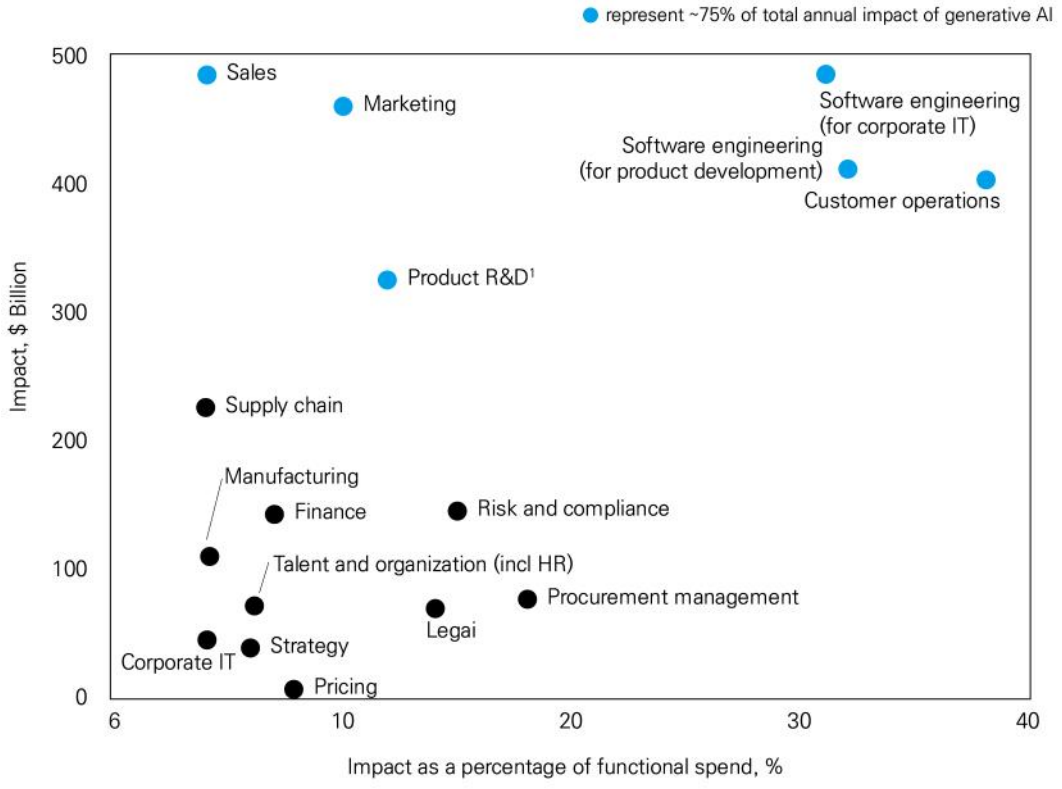


[그림 2-13] 산업별 생성형 AI의 잠재적 가치

주: 영업이익률 상승 폭 (단위: % p), 영업이익률의 상승 폭은 같은 산업에서도 기업마다 상당히 다를 수 있음. 잠재적 가치 상승은 현재 운영 모델을 기반으로 각 산업의 사례를 평가하여 계산
자료: PwC의 S&P Capital IQ 데이터 분석

생성형 AI의 도입 효과를 기능(Function)별로 보면 고객 운영(Customer operations), 마케팅 및 판매(Marketing and Sales), 소프트웨어 엔지니어링(Software engineering), 연구개발(Research and Development) 측면에서 상대적으로 효과가 큰 것으로 나타났다. 생성형 AI를 활용해서 얼마나 비용을 줄이고, 새로운 가치를 만들어 낼 수 있는지를 분석한 결과 위의 기능들이 전체 가치의 75%를 차지하는 것으로 분석되었다. 공급망

(Supply chain), 제조(Manufacturing), 재무(Finance)와 같은 다른 기능들도 생성형 AI의 도입 효과가 있지만 상대적으로 영향력이 작은 것으로 나타났다.[14]



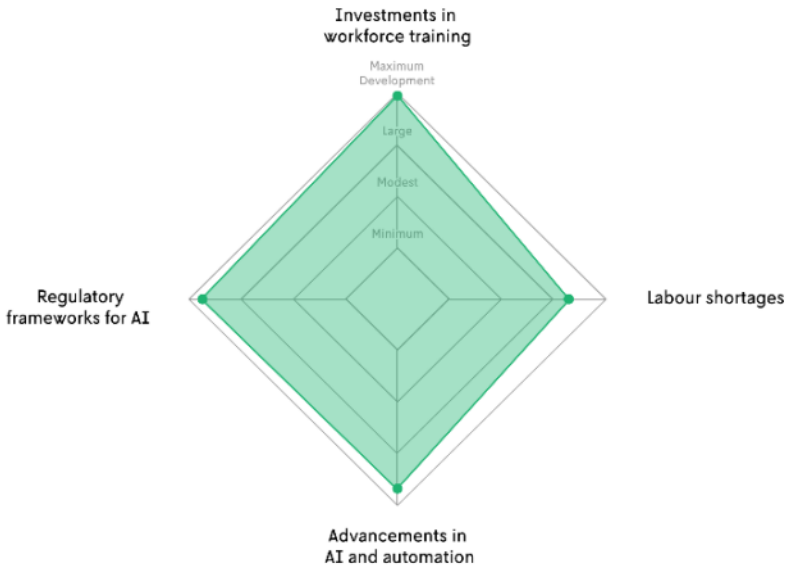
[그림 2-14] 기능별 생성형 AI 도입 효과

자료: Mckinsey&Company(2023) "The economic potential of generative AI"

AI의 노동시장에 대한 파급효과는 많은 이슈를 제기하고 있다. OpenAI가 ChatGPT를 출시한 이후 불과 2년 만에 AI 도구는 빠르게 업무와 접목되고 있다. 이러한 빠른 변화는 중요한 질문들을 제기한다. AI는 앞으로 5년, 10년, 20년 동안 어떻게 직장을 변화시킬 것이며, 직원들의 생산성과 일과 삶의 균형을 개선할 것인지, 아니면 사회경제적 격차를 심화시키고 생계를 위협할 것인지 다양한 이슈가 제기될 수 있다. Futures

14) Mckinsey&Company(2023) "The economic potential of generative AI"

Platform(2024)에서는 이러한 가능성을 탐구하기 위해 AI가 직장에 통합됨에 따라 두 명의 근로자, 간호사인 Sam과 건설 현장 관리자 Anna가 네 가지 잠재적인 미래를 어떻게 헤쳐 나가는지 시나리오 분석을 통해 알아보았다. 네 가지 시나리오는 가능한 모든 미래 방향을 포괄하지는 않지만, 인간과 AI의 협업이 직장에서 발전해 나갈 주요 경로들을 분석하여 조직이 미래가 어떻게 펼쳐질지 다양한 시각으로 탐구할 수 있는 전략적 관점을 제공한다. 시나리오 구축 과정에서 이 주제의 미래 발전에 영향을 미칠 핵심 동력들을 4가지 설정했다. 노동력 훈련에 대한 투자, AI에 대한 규제 프레임워크, AI와 자동화의 발전, 노동력 부족이 그것이다. 각 시나리오에서는 이러한 동력들이 앞으로 몇 년 동안 어떻게 발전할지, 또한 서로 어떻게 상호작용하고 영향을 미칠지를 다루었다. 첫 번째 시나리오는 인간과 AI가 협업하며 시너지를 내는 경우이다. 시나리오를 구성하는 4가지 요소들이 충분히 갖추어진 상황이다.



[그림 2-15] 시나리오1

자료: Futures Platform(2024), “Four scenarios on the future of AI in the workplace”

이 경우 가정하는 세부 시나리오1은 다음과 같다.

[표 2-1] 시나리오1 세부 내용

산업들이 노동력 부족 문제에 직면하면서 인간과 기계의 협업이 표준적인 관행이 되고 있습니다. 바쁜 도심 병원에서 근무하는 간호사 Sam은 그의 일상 업무에 변화가 생긴 것을 경험하고 있습니다. AI로 구동되는 로봇이 이제 환자의 생명 징후를 모니터링하고 데이터를 입력하는 등 그가 시간을 많이 쏟았던 반복적인 업무를 처리합니다. 이를 통해 Sam은 불안해하는 가족을 위로하고 환자들이 어려운 치료 과정을 극복하도록 돕는 것과 같은 인간적인 돌봄에 더 집중할 수 있게 되었습니다. 한편, 건설 현장 관리자 Anna는 정밀하게 자재를 들어 올리는 AI 장착 크레인을 감독하고 있습니다. 그녀의 팀은 이제 AI가 반복적이고 노동 집약적인 작업을 처리하는 동안에 보다 복잡한 현장 작업을 해결할 수 있게 되었습니다. 이러한 시스템들은 피로를 느끼지 않고, 세부 사항을 놓치지 않으며, 인간 근로자들이 더욱 안전하고 효율적으로 일할 수 있도록 돕습니다. 정부는 Sam과 Anna 같은 근로자들을 보호하기 위해 AI가 인간의 역할을 대체하기보다는 보완하도록 규제를 마련하는 데 적극적으로 나서고 있습니다. 기업들 또한 고도화된 AI와 협력할 수 있는 기술과 자신감을 근로자들에게 제공하기 위해 대규모로 훈련에 투자하고 있습니다. 그 결과, 인간이 소외되지 않고 지원받는 안전하고 효율적인 작업 환경이 조성되고 있습니다.

자료: Futures Platform(2024), “Four scenarios on the future of AI in the workplace”

4가지 시나리오[15]는 AI 도입과 효과 측면에서 다음과 같은 함의를 준다. AI 기술의 발전 속도와 그 영향력은 매우 매우 불확실하여 우리는 AI가 인간과 완벽한 시너지를 이루는 상황부터 기대에 미치지 못하는 상황까지 다양한 가능성을 고려해야 한다. 규제 관련, 적절한 규제 프레임워크의 존재 여부가 AI 도입의 성공을 크게 좌우한다. 규제가 AI 발전 속도를 따라가지 못하면 노동자 대체와 같은 부정적 결과를 초래할 수 있다. 교육과 재훈련 측면에서 AI와 효과적으로 협업하기 위해서는 노동자들의 지속적인 교육과 재훈련이 필수적이며 기업과 정부 차원의 투자가 필요한 영역이다. 인간의 고유 가치 관련 모든 시나리오에서 모든 시나리오에서 감정 지능, 복잡한 의사결정, 창의성 등 인간 고유의 능력이 여전히 중요하게 다뤄진다. AI 시대에도 이러한 인간의 강점을 활용하고 발전시키는 것이 중요하다. 또한, AI 도입을 서두르거나 지나치게 신중한 태도 모두 문제를 일으킬 수 있으며 이에 효과와 위험을 균형 있게 평가하며 단계적으로 접근하는 방식이 필요하다. 사회경제적 영향 측면에서 AI의 도입은 단순히 기술적 문제가 아니라 소득 불평등, 일자리 변화 등 광범위한 사회경제적 영향을 미칠 수 있으며 이에 대한 대비책도 논의되어야 한다. 산업별 차별화된 전략도 요구되는데 의료, 건설 등 각 산업의 특성에 맞는 AI 도입

15) 나머지 3개 시나리오는 별첨 참고

전략이 필요하며 일괄적인 접근보다는 각 분야의 특수성을 고려한 맞춤형 접근이 중요하다. 또한, AI 기술과 그 영향은 계속 변화할 것이므로, 도입 후에도 지속적인 모니터링과 평가가 필요하며 상황에 적합한 조정이 적시에 필요하다.

제2절 AI 생태계 분석

1 AI 기술 생태계

생태계란 자연환경과 생물이 서로 밀접한 관계를 형성하고 있는 시스템을 의미한다.[16] Moore(1996)는 기업 그 자체를 분석하던 경영학의 시각에서 벗어나 기업이 속한 산업의 전체적인 구조와 역학관계를 바탕으로 기업의 분석이 이루어져야 한다고 주장하였으며 이를 위해 비즈니스 생태계(Business Ecosystem)라는 접근을 제시하였다.[17] 비즈니스 생태계는 많은 연구자가 다양한 분야에서 활용해 왔다.[18] Messerschmitt & Szyperski(2005)는 소프트웨어 산업[19], Basole(2009)과 Karla(2011)는 모바일 산업[20], Vaz et al(2013)은 비디오게임 산업[21], Yang et al(2009)는 스포츠 산업[22] 등 다양한 분야에서 시도되고 있다.

현재까지 비즈니스 생태계를 명확히 정의할 수 있는 통일된 개념이 제시되지는 않았으나, 일반적으로 사회과학 분야에서 핵심종, 공진화, 선순환, 외부 환경요소 등의 특징들이 가지는 개념으로 활용되었다.[23] Moore(1993)는 비즈니스 생태계를 '공동의 운명 하에서

16) Tansley, A.G., "The Use and Abuse of Vegetational Terms and Concepts", Ecology, Vol. 16, No. 3, pp. 284-307, 1935
17) Iansiti, M. and Levien, R., "Strategy as Ecology", Harvard Business Review, Vol. 82, No. 3, pp. 68-81, 2004.
18) 서한결·박인서·김성환·채수준·정양헌(2015). 공공 R&D (참여) 조직의 공생적 학습 생태계 조성방안 연구. 경영교육연구. 30(6): 557 - 579.
19) Messerschmitt, D.G. and Szyperski, C., Software Ecosystem, MIT Press, 2005.
20) Basole, R.C., "Visualization of Interfirm Relations in a Converging Mobile Ecosystem", Journal of Information Technology, Vol. 24, pp. 144-159, 2009
21) Vaz, L.F.H., Nogueira, A.R.R., Rodrigues, M.A.D.S.,and Chimenti, P.C.P.D.S., "A New Conceptual Model for Business Ecosystem Visualization and Analysis", Revista de Administrao Contempornea, Vol. 17, No. 1, pp. 1-17, 2013.
22) Yang, X.C., Li, Z.H., and Liang, Q., "An Analysis on the Ecosystem of Leisure Sports Industry and Its Competition Strategy Choice", Journal of Beijing Sport University, Vol. 3, No. 7, pp. 25-28, 44, 2009.
23) 김성근·김종엽(2015). 산업생태계 모형 및 영향 파급효과 묘사 방안: 공공 SI산업에의 적용. Journal of Information Technology and Architecture. 12(3): 299 - 309

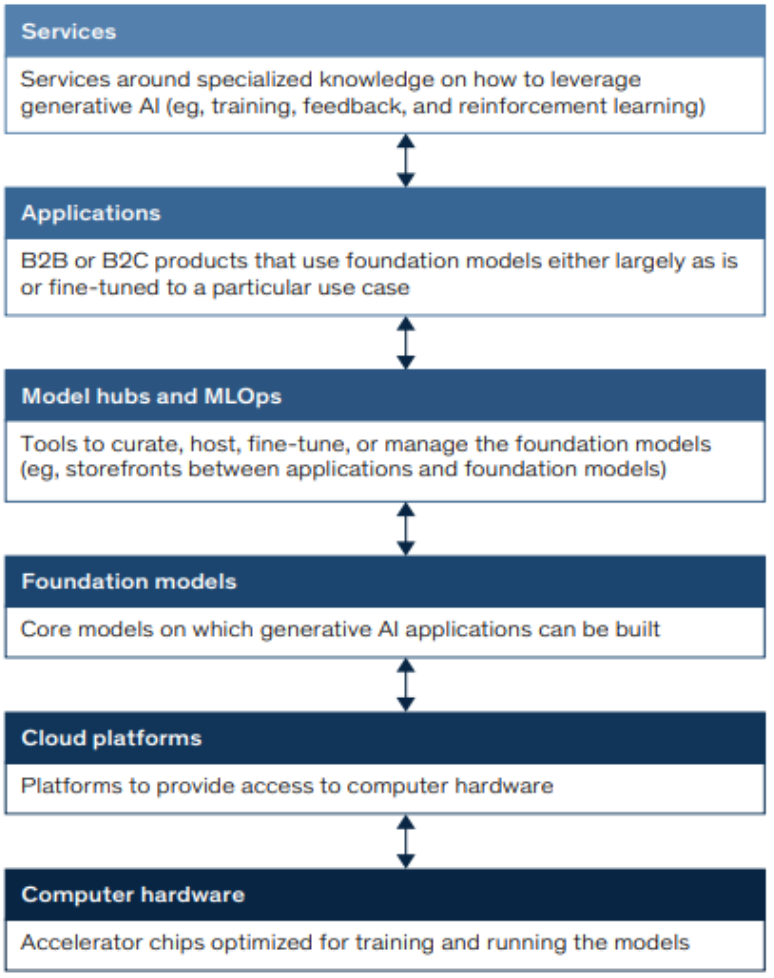
하나 이상의 자원을 공유하며 생태계를 구성하는 조직과 개인들이 상호작용하고 공진화(co - evolve)하는 경제적 공동체'라고 정의하였다. 또한 핵심종(keystone species)과 공유된 비전, 경제적 공동체, 공동의 가치 창출, 공진화 등을 비즈니스 생태계의 공통적인 구성 요소로 제시하며, 이를 통해 가치를 창출하고, 핵심종을 중심으로 공진화하는 생태계가 비즈니스 생태계라고 설명하였다.[24] Moore(1993)에 의해 처음 제시된 비즈니스 생태계는 Porter(1985)의 가치사슬(value chain) 이론에서 발전해 왔고 가치사슬 이론은 가치 창출을 위해 단일 기업이 수행하는 전 활동 분석의 기반으로 활용되었으나,[25] 점차 기술의 발전과 산업 간의 융·복합화로 인해 산업 환경이 더욱 복잡해지면서 더 이상 가치 창출 과정들을 선형적인 관계만으로 이해하기에는 한계가 있었다.[26] 생태계를 종합적으로 이해하기 위해서는 정책, 투자 등 다양한 시각에 참여자와의 상호작용을 고찰할 필요가 있다. Mckinsey&Company(2023)은 생성형 AI의 가치사슬을 6계층으로 구분하였다. 컴퓨터 하드웨어(Computer hardware)는 생성형 AI 모델의 훈련 및 실행에 최적화된 가속기 칩과 같은 하드웨어이다. AI 모델이 제대로 작동하려면 고성능의 컴퓨터 하드웨어가 필수적이다. 클라우드 플랫폼(Cloud platforms)은 컴퓨터 하드웨어에 접근할 수 있는 플랫폼으로, 클라우드를 통해 사용자는 AI 훈련과 같은 고도 연산을 수행할 수 있다. 클라우드 플랫폼은 하드웨어의 고성능을 손쉽게 활용할 수 있도록 지원한다. 기초 모델(Foundation models)은 생성형 AI 애플리케이션을 구축할 수 있는 핵심 모델이다. 이러한 기초 모델은 다양한 AI 애플리케이션의 기반이 되며, 다른 단계에서 이를 활용해 추가적인 맞춤화가 가능하다. 모델 허브 및 MLOps(Model hubs and MLOps)는 기초 모델을 관리하고 조정하며, 애플리케이션과 기초 모델 간의 중개 역할을 한다. 예를 들어, 모델을 미세 조정하거나 호스팅하는 데 사용되며, 새로운 제품에 맞출 수 있는 기능을 제공한다. 애플리케이션(Applications)은 기초 모델을 기반으로 한 B2B 또는 B2C 제품 및 서비스다. 애플리케이션은 최종 사용자를 위한 다양한 기능을 제공한다. 서비스(Services)는 생성형 AI를 활용하기 위한 전문 지식을 바탕으로 한 서비스들로, 훈련, 피드백, 강화 학습 등이 포함된다. 기업들은 서비스를 통해 AI를 효율적으로 도입하고 활용할 수 있다.[27]

---

24) Moore, J. F.(1996). The Death of Competition: Leadership and Strategy in the Age of Business Ecosystems. New York. NY: Harper Business.

25) Porter, M. E.(1985). Competitive Advantage: Creating and Sustaining Superior Performance. New York. NY: The Free Press.

26) 곽정호·조지연·이용성·이봉규(2011). 새로운 통신시장 활성화를 위한 모바일 생태계 통신정책. 인터넷정보학회논문지. 12(4): 93 – 106.



[그림 2-16] 맥킨지의 생성형 AI 가치사슬

자료: Mckinsey&Company(2023), "Exploring opportunities in the generative AI value chain"

PwC(2024)는 공급자와 수요자를 고려한 생태계 구조를 제안하였다. AI 산업은 관련 기술을 개발하거나 AI를 활용한 제품 및 서비스를 생산·유통·활용하는 등의 과정에서 가치를 창출하는 산업을 일컫는다. 후방산업은 AI 학습 및 활용을 위한 데이터 수집, 구매, 구축 컨설팅, 분석 등과 연계된 생태계이며, 전방산업은 AI를 활용하여 제품 및 서비스를 생산하고 제공하는 영역이다.[28]

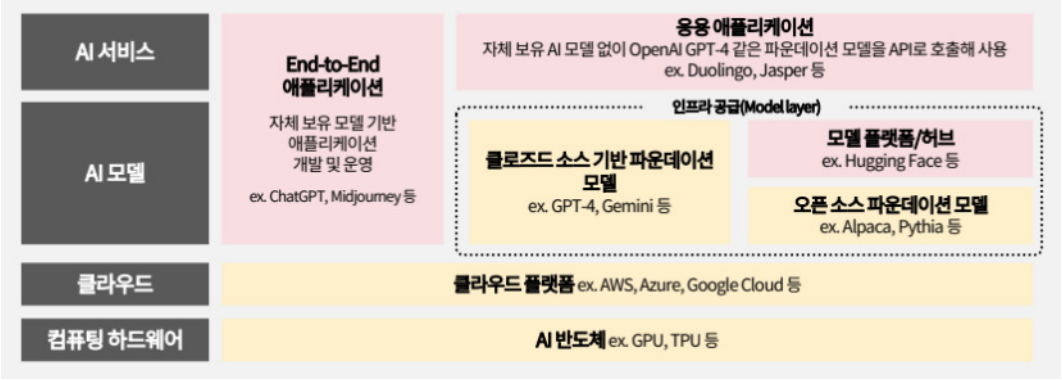
27) Mckinsey&Company(2023), "Exploring opportunities in the generative AI value chain"

[그림 2-17] AI 산업의 생태계 : 공급자와 수요자

자료: PwC(2024) "생성형 AI를 활용한 비즈니스의 현주소"

공급자 측면에서의 생성형 AI의 비즈니스 생태계는 크게 ① 컴퓨팅 하드웨어(AI 반도체), ② 클라우드, ③ AI 모델, ④ AI 서비스(애플리케이션)의 4계층으로 구분될 수 있다. Mckinsey&Company(2023)의 6계층 체계보다 수직적으로는 단순화했지만, 수평적 측면에서 다양한 요소를 고려하고 있다. 하나의 기업이 AI 모델과 서비스를 동시에 제공하는 End to End 애플리케이션, 별도의 응용 AI 서비스로 애플리케이션을 제공하는 형태를 구분하고, AI 모델에서 폐쇄형 기반 모델, 오픈소스 기반모델, 다양한 기반 모델을 연동시켜주는 모델 플랫폼 허브로 세분화하였다. NVIDIA, Google, OpenAI 등 빅테크들은 압도적인 컴퓨팅 파워와 대규모 자본을 토대로 생성형 AI 인프라(컴퓨팅 하드웨어, 클라우드)와 AI 모델 선점을 위해 역량을 쌓는 중이며, 스타트업들은 대체로 자체 개발 또는 타사 개발 AI 모델을 이용해 다양한 AI 서비스를 출시하며 경쟁력을 확보를 위해 노력 중이다.

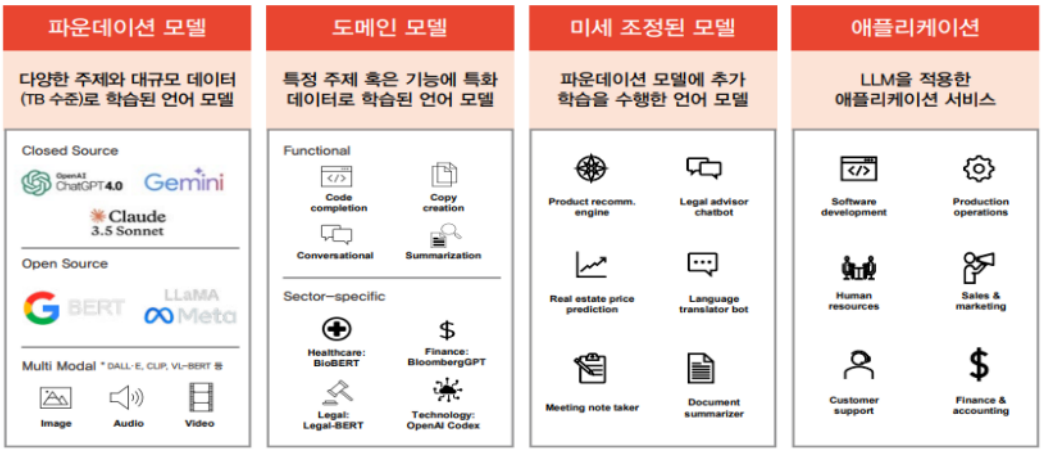
28) PwC(2024) "생성형 AI를 활용한 비즈니스의 현주소"



[그림 2-18] 생성형 AI 생태계 구조 : 공급자

자료: 삼성증권, PwC(2024)

AI 생태계 구조는 AI 모델을 중심으로 세분화해 볼 수 있다. 대규모 데이터와 컴퓨팅 파워로 학습한 파운데이션 모델, 특정 주제에 특화된 데이터로 학습된 도메인 모델, 파운데이션 모델에서 추가 학습한 미세 조정된 모델, 그리고 이를 활용한 애플리케이션으로 구성될 수 있다. 특히 메타의 라마 공개를 계기로, 1년 반 만에 수천 개의 파운데이션 모델, 도메인 모델, 미세 조정된(Fine-tuned) 모델이 시장에 쏟아져 나오고, 이를 활용하는 다양한 애플리케이션이 등장하며 거대한 생성형 AI 생태계가 형성 중이다.



[그림 2-19] 생성형 AI 기술 생태계

출처: PwC(2024) "생성형 AI 기반의 Accelerated AI: 금융 AI의 발전 전략"

생태계 구분	영역 구분	주요 기업	사업 내용
AI 서비스	스타트업 집중 영역	• Jasper • Stability AI • Github	• 광고 문구 생성 AI 플랫폼 '재스퍼AI' 운영. 기업 가치 15억 달러('23년 기준) • 이미지 생성(Text-to-Image) AI 모델 '스테이블 디퓨전' 개발. 기업 가치 10억 달러('23년 기준) • 코드 완성 AI 도구 '코파일럿(Copilot)' 출시
AI 모델	대기업/빅테크 집중 영역	• OpenAI • Anthropic • 네이버	• API 기반 범용 AI 모델(GPT 시리즈, DALL-E, Codex)을 기업에 유료로 제공 • 자연어 기반 생성형 AI 모델(Claude, Claude Instant)을 유료로 제공 • 생성형 AI 모델 '하이퍼클로바X' 개발 및 다양한 AI 서비스(클로바X, 큐:) 공개
클라우드		• Microsoft • Amazon • Google	• Azure 중심으로 챗GPT 기반 코파일럿 기술을 클라우드로 통합 제공 • AWS 사용자가 AI 모델을 통해 소프트웨어 성능을 향상할 수 있는 클라우드 서비스 '베드록' 출시 • 자사 클라우드 서비스를 활용해 소프트웨어 개발자들에게 자사 언어모델 'PaLM' 유료로 제공
컴퓨팅 하드웨어		• NVIDIA • AMD • 삼성전자	• 신형 AI 반도체 'H200' 공개 (GPT4를 훈련하는데 사용된 H100의 업그레이드 버전) • 최첨단 AI 반도체 'MI300X' 출시, '23년 4분기부터 AI 전용 반도체 생산 증가 • AI 산업용 차세대 메모리 반도체 2종 공개 (고대역폭메모리(HBM)-프로세싱인메모리(PIM)와 저전력 더블데이터레이트 (LPDDR)-PIM)

[그림 2-20] AI 생태계 내 주요 기업들

출처: PwC(2024) "생성형 AI를 활용한 비즈니스의 현주소"

EIT(2021)[29]은 AI 생태계를 체계적으로 분석하고 데이터베이스 기반으로 이를 정책적으로 활용할 수 있는 방안을 제시했다.[30] 연구팀은 35개의 기존 AI 분류 프레임워크를 분석했으며 이 중 32개는 기업에서 만든 것이고, 3개는 학술 연구에서 제시되었다. 기존 프레임워크들은 주로 기업과 기술을 중심으로 분류하고 있었다. 기업 중심의 프레임워크는 다시 두 그룹인 스타트업과 일반 기업에 초점을 맞추었다. 연구팀은 이 프레임워크들이 사용하는 8가지 주요 분류 차원을 발견했는데 산업 분야, 기업 기능, AI 기능, 기술 인프라, 고객 초점(B2C/B2B), 지리적 위치, 자금 조달 단계, 생태계 역할이 분류 요소였다.

29) European Institute of Innovation & Technology A body of the European Union
30) EIT(2021), "Creation of a taxonomy for the European AI ecosystem"

[표 2-2] 기존 AI 생태계 프레임워크 분류

구분	카테고리	산업	역량	기업 기능	인프라
Daxue Consulting AI Landscape	Companies	●	●	●	●
Element AI – Canadian AI Startups	Companies				
Firstmark Data & AI	Companies	●	●	●	●
Linux Foundation(AI Ecosystem)	Companies		●		●
appliedAI Startup Landscape	Startups	●	●	●	●
Bloomberg Beta	Startups	●	●	●	
CB Insights 2017	Startups	●	●	●	
CB Insights 2018	Startups	●	●	●	
CB Insights 2019	Startups	●		●	
CB Insights 2020	Startups	●	●	●	●
CB Insights Agriculture	Startups		●		
CB Insights Healthcare	Startups		●		
CB Insights Retail	Startups		●		
Cognite Ventures DL	Startups	●	●	●	●
MMC Ventures UK AI	Startups	●	●	●	●
StartHub Computer Vision	Startups	●	●	●	
StartHub Israel	Startups	●	●	●	
StartHub NLP	Startups	●	●	●	
3XN – AI Taxonomy	Technology		●		●
AI & Intelligent Automation Network	Technology		●		
Algorithmia – AI & ML	Technology		●		
appliedAI Use Case Cards	Technology		●		
appliedai.com Search platform	Technology	●	●	●	

출처: EIT(2021), “Creation of a taxonomy for the European AI ecosystem”

산업 분야 분류는 대부분의 프레임워크에서 사용되지만, 세부 카테고리가 매년 크게 바뀌었는데 예를 들어, CB Insights의 분류는 2017년부터 2020년까지 매년 상이했다. 기업 기능 분류는 주로 마케팅, 고객 지원, 영업 같은 고객 대면 기능과 IT, HR 같은 지원 기능을 포함하나 생산이나 연구개발 관련 기능은 거의 포함되지 않았다. AI 기능과 기술

인프라는 이 두 가지가 요소가 명확히 구분되지 않고 기술 스택이라는 이름으로 함께 분류되는 경우가 많았으며 또한 세부 카테고리가 일관성 없이 다양하게 나타났다. 다른 차원들 (고객 초점, 지리적 위치, 자금 조달 단계, 생태계 역할)은 일부 프레임워크에만 사용되었다. 이런 기존 프레임워크들은 일관성 부족으로 분류 기준이 자주 바뀌어서 시간에 따른 변화를 추적하기 어렵고 중요한 측면을 모두 포함하지 않아 불완전했으며 기술적 부정확성 측면에서도 문제가 있었는데 특히 AI 기능을 분류할 때 기술적으로 정확하지 않은 경우가 많았다. 새로 연구한 EU의 AI 생태계 프레임워크는 구조 측면에서 시간이 지나도 변하지 않는 산업, 기업 기능, AI 기능 등의 분류 차원을 고려하였고 내용 요소 측면에서는 동적이며 시간에 따라 변할 수 있는 사용 사례, 기술 세부 등의 사항을 포함하였다. 구조적 프레임워크 상세 관련 산업은 NACE 코드(유럽의 표준 산업분류)를 기반으로 농업, 제조업, 교육 등의 세부 분류를 활용하였다. 기업 기능은 인사, 마케팅, 고객 서비스, 영업, 회계 등을 포함했다. 지리적 위치는 국가, 도시, 지역으로 구분하였으며 AI 기능은 8가지 주요 카테고리로 구분(예: 컴퓨터 비전, 자연어 처리, 로보틱스 등)을 하고 각 카테고리는 더 세부적인 하위 카테고리로 나누었다.

[표 2-3] EIT 프레임워크에서 구분된 AI 역량과 세분류

대분류	세부 기능
컴퓨터 비전	• 이미지 분할, 객체 감지 및 추적, 이미지 분류, 감정 인식, 3D 재구성
컴퓨터 청각	• 음성-텍스트 변환, 음악 지식, 소리 유사도 평가, 음원 분리, 오디오 기반 감정 분석
컴퓨터 언어학	• 번역, 텍스트 분류, 감정 분석, 개체 인식, 관계 추출, 대화 시스템
로보틱스	• 로봇 동작 계획, HD 매핑 및 위치 추적, 제어 최적화, 자율 로봇/휴먼 로봇 상호작용, 고급 드론, 모바일 로봇, 사용자 적응형 제어 자동화
예측	• 시계열 예측, 의존성 기반 예측
발견	• 세분화 및 군집화, 이상/특이치 감지, 상관관계 분석, 인과관계 추론, 연관성 분석
계획	• 협력적 다중 에이전트 시스템, 정책 개발/전략적 에이전트, 물류 계획, 계획 및 일정 관리
생성	• 오디오 생성, 이미지 생성 / 조작, 스타일 전이, 텍스트 생성 / 요약, 증강 엔지니어링

출처: EIT(2021), “Creation of a taxonomy for the European AI ecosystem”

지원 기술 유형 관련 인프라, 플랫폼, 애플리케이션으로 구분하고 AI와 직접적으로 관

련되지 않지만 AI 발전에 중요한 기술들을 포함하였다. 도메인 특정 및 확장성 측면에서 특정 분야(예: 기후변화)에 대한 추가적인 분류 차원을 허용하였다.

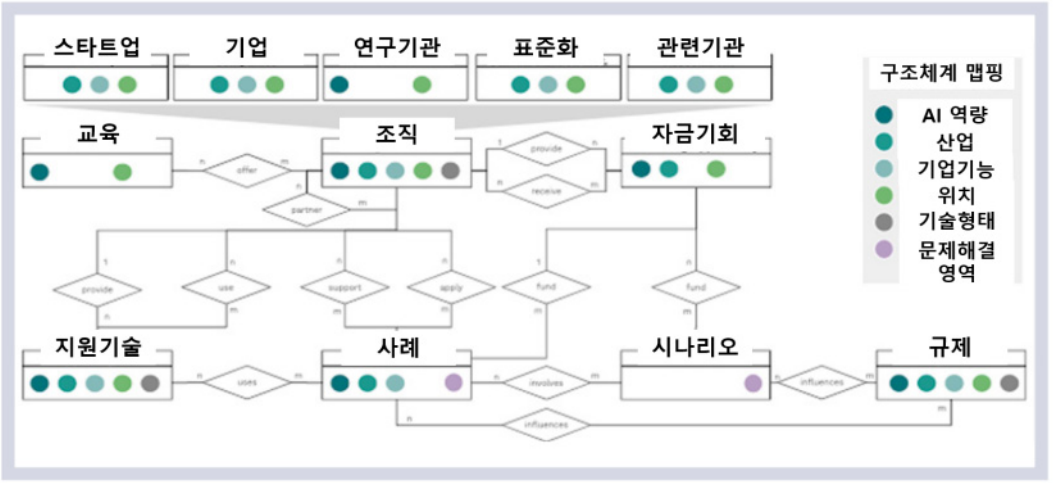
[표 2-4] 기후변화로 확장된 도메인

대분류	세부기능
전력 시스템	• 저탄소 전력 활성화, 현 시스템 영향 감소, 글로벌 영향력 보장
교통	• 교통 활동 감소, 차량 효율성 개선, 대체 연료 및 전기화, 수단 전환
건물과 도시	• 건물 최적화, 도시계획, 미래 도시
산업	• 공급망 최적화, 원자재 개선, 생산 및 에너지
농업과 산림	• 배출량 원격 감지, 정밀 농업, 습지 모니터링, 산림 관리
이산화탄소 제거	• 직접 공기 포집, CO2 격리
기후 예측	• 데이터, ML & 기후과학 연계, 극단적 기상현상 예측
사회적 영향	• 생태계, 인프라, 사회 시스템, 위기
개인 행동	• 개인 탄소발자국 이해, 행동 변화 촉진
집단적 의사결정	• 사회적 상호작용 모델링, 정책 알림, 시장 설계

출처: EIT(2021), “Creation of a taxonomy for the European AI ecosystem”

내용 요소 상세 설명이 가능하도록 사용 사례(AI 기술이 특정 문제를 해결하는 구체적인 예), 조직(AI 관련 제품이나 서비스를 제공하는 기업, 연구소 등), 자금 조달 기회(AI 혁신을 위한 투자 활동), 교육 제공(AI 관련 교육 프로그램이나 자료), 기술 솔루션(AI 개발과 운영에 사용되는 기술적 자산이나 도구), 도메인 특정 시나리오(특정 분야의 복잡한 문제를 해결하는 AI 활용 방안), 규제(AI 사용에 영향을 미치는 법적, 규제적 요구사항)이 포함되었다.

정보 아키텍처 관련 EIT 프레임워크는 관계형 데이터베이스 구조로 구현되었다. 데이터베이스는 구조적 프레임워크 테이블, 내용 요소 테이블, 관계 테이블 등으로 구성되며 이를 통해 복잡한 쿼리와 분석이 가능하다. 이러한 프레임워크를 사용하면 특정 시나리오와 관련된 모든 사용 사례와 조직 찾기, 특정 솔루션 영역에서 핵심 기술을 제공하는 조직 식별, 특정 시나리오에 대한 규제의 영향평가 등이 가능할 것으로 기대하고 있다.



[그림 2-21] AI 생태계의 데이터베이스 구조

출처: EIT(2021), “Creation of a taxonomy for the European AI ecosystem”

EIT AI 생태계 프레임워크는 EU의 전략기획, 문서연결 등 실질적인 목적과도 연계되어 있도록 구성하였다. 제안된 AI 생태계 분류 체계와 EU의 주요 AI 관련 문서들과의 전략적 연계성이 고려되었다. 특히 EU AI 백서, HLEG(High-Level Expert Group) 문서, 그리고 EIT 디지털 정책 문서와의 호환성에 중점을 두고 있다. 제안된 생태계 프레임워크는 4가지 측면에서 의미가 있다. 첫째, AI에 대한 공통 언어를 제공함으로써 유럽 전역에서 일관된 보고와 분석을 가능하게 한다. 둘째, 동적 구조를 통해 빠르게 발전하는 AI 기술을 유연하게 반영할 수 있다. 셋째, 유럽의 AI 리더십 확보를 위한 활동을 가속화하고 중요 영역에 집중할 수 있게 한다. 넷째, AI가 생태계 내에서 효율적으로 응용되고 있는지 네크워크 구조로 이해할 수 있다.

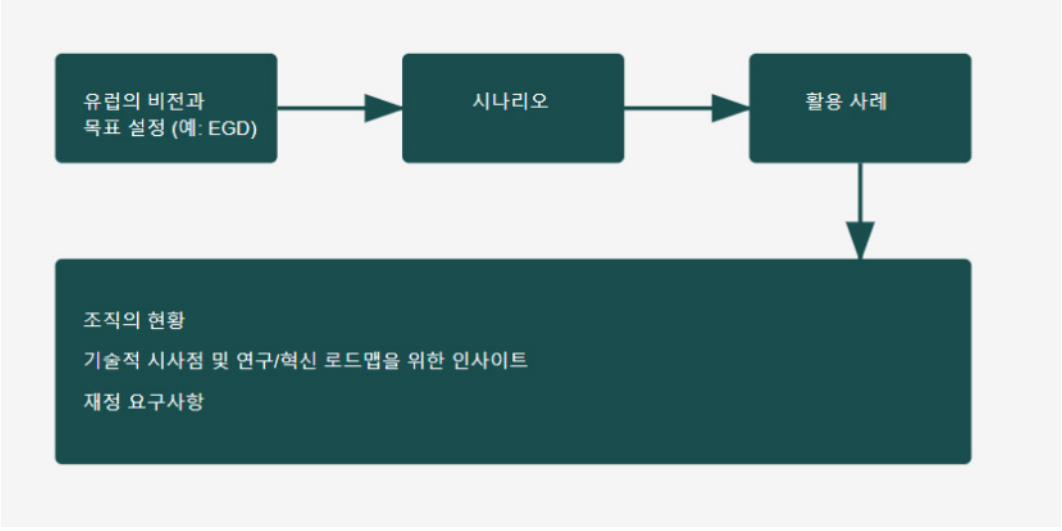
EU AI 백서와 비교한 결과, AI 생태계 프레임워크는 백서에서 제시하는'우수한 생태계'와 '신뢰할 수 있는 생태계' 개념을 반영하고 있었다. 우수한 생태계와 관련하여 과학적 돌파구 지원, 기술 리더십 유지, 새로운 기술의 광범위한 적용 촉진 등의 요소가 AI 생태계 프레임워크 통합되어 있다. ‘신뢰의 생태계’와 관련해서는 AI 시스템의 신뢰성 보장, 위험 기반 접근법, 표준화 및 인증 등의 개념이 반영되어 있다.

HLEG 문서와의 호환성 분석에서는 ‘신뢰할 수 있는 AI를 위한 윤리 지침’과 ‘정책 및

투자 권장사항'이 검토되었다. 윤리 지침에서 제시된 7가지 신뢰할 수 있는 AI 차원이 분류 체계의 사용 사례와 AI 기능 매핑에 반영되어 있음이 확인되었다. 정책 및 투자 권장사항에서 강조된 4가지 이해관계자 그룹과 4가지 AI 생태계 주요 요소 또한 프레임워크에 효과적으로 통합되어 있다.

EIT "유럽의 AI 접근법" 문서와의 호환성 분석에서는 AI의 맥락별 규제 분석의 필요성이 강조되었으며, 제안된 분류 체계가 맥락별 분석을 할 수 있도록 지원한다. 또한, 민사책임 체계, 윤리적 측면의 프레임워크, 지적 재산권 등 기타 EU 문서들의 내용도 분류 체계의 규제 카테고리나 사용 사례 설계에 반영될 수 있도록 구성되어 있다.

이처럼 제안된 AI 생태계 프레임워크는 EU의 주요 AI 정책 문서들과 높은 수준의 호환성을 보이고 있다. 이는 해당 분류 체계가 유럽의 AI 전략 수립과 실행, 그리고 정책 모니터링 및 평가에 유용한 도구로 활용될 수 있음을 시사한다. 또한, 이 체계의 유연성과 확장성은 향후 AI 생태계의 변화에도 적응할 수 있는 장점을 제공한다. 새로운 프레임워크는 유럽의 AI 생태계를 체계적으로 이해하고 발전시키는 데 중요한 역할을 할 것으로 기대된다.



[그림 2-22] AI 생태계 프레임워크와 EU 전략과의 연계

출처: EIT(2021), "Creation of a taxonomy for the European AI ecosystem"

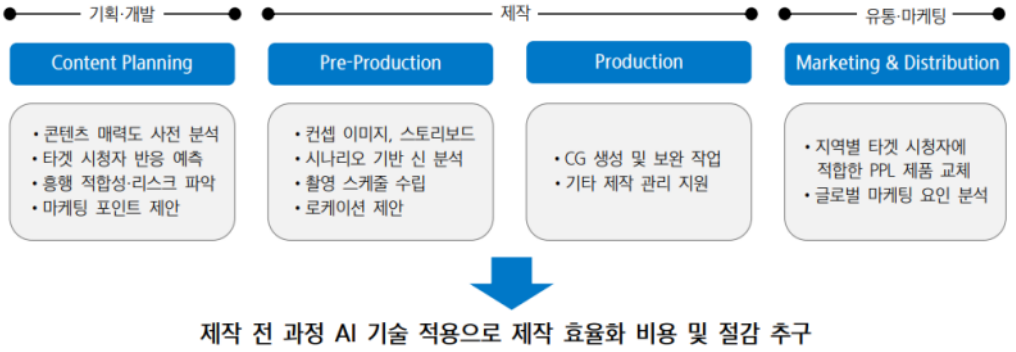
AI 생태계는 독립적으로 존재하지 않고 다른 생태계와 연계되어 생태계의 상호작용, 가치 창출 방식을 변화시킨다. AI 생태계의 애플리케이션 영역에 해당하는 다양한 생성형 AI 서비스는 텍스트, 이미지, 영상, 음성, 코딩, 아바타 등 다양한 디지털 요소를 프롬프트 기반으로 사용자가 생성할 수 있도록 도와준다. 이러한 디지털 요소들은 다양한 형태로 연결될 수 있으며 기존의 디지털 요소 제작방식과는 전혀 다른 접근 방식으로 더 빠르고 낮은 비용으로 다양한 사용자가 쉽게 이를 활용할 수 있도록 지원한다. 이는 생성형 AI가 디지털 창작의 방식을 바꾼다는 의미이다.

[그림 2-23] 생성형 AI 가치사슬 및 창작영역 활용성

출처: 삼정KPMG 경제연구원

콘텐츠 제작의 경우 생성형 AI를 활용하면 기획과 개발, 제작, 유통 및 마케팅 단계에서 생산성을 높이거나 병렬적으로 작업 프로세스를 진행할 수 있어 생산성을 높이고 혁신적

인 방식으로 작업 과정을 재구성할 수 있다.



[그림 2-24] 콘텐츠 제작 프로세스: 제작 단계에 영상 생성 AI 접목 시 효율성

자료: 삼성증권

이처럼, 생성형 AI는 기존의 콘텐츠 생태계의 가치 창출 방식과 상호작용을 바꾸며 이는 생태계 내 기업의 수익구조에 직간접적인 영향을 미친다. 영상 생성형 AI 소라(Sora)의 등장은 기존 광고와 게임 제작, 특수 효과 제작방식 등을 변화시켜 새로운 경쟁 구도를 형성하게 한다.

[그림 2-25] 생성형 영상 AI 소라 등장에 따른 수혜 업종과 피해 업종

자료: 삼성증권

AI로 인해 콘텐츠 생태계의 수익구조 변화 뿐만 아니라 다양한 정책이슈도 제기되고 있다. 첫째, 일자리 안정성 문제이다. 엔터테인먼트 등 콘텐츠 업계의 생성형 AI 도입으로 인해 일자리 안정성에 대한 우려가 제기되고 있다. 특히, AI 기술이 콘텐츠 제작의 일부 역할을 대체할 가능성이 있어 관련 분야의 고용에 영향을 미칠 수 있다. 둘째, 잘못된 정보 양산 문제이다. 생성형 AI의 사용이 확대됨에 따라 허위 정보의 생성 및 확산에 대한 우려가 제기되고 있으며 이러한 허위 정보는 사회적으로 혼란을 초래하고 신뢰성을 해치는 결과를 가져올 수 있다. 셋째, 양질의 데이터 확보 문제이다. 신뢰 가능한 양질의 데이터를 확보하고 이를 학습 데이터로 활용하기 위해서는 상당한 비용이 소요될 것으로 예상되며 생성형 AI의 성능을 높이기 위해서는 방대한 양의 고품질 데이터를 확보해야 하는데, 이 과정에서 데이터 수집 및 관리의 어려움이 존재한다. 넷째, 사이버보안 문제이다. AI 기술의 발전 및 도입 확대는 정교한 사이버 범죄 가능성을 높이고 있으며 생성형 AI는 악의적인 목적으로도 활용될 수 있어 이에 대한 적절한 보안 대책이 필요하다. 다섯째, 저작권 문제이다. 생성형 AI가 창작 활동에 활용됨에 따라, 그 결과물에 대한 저작권 문제가 발생할 수 있다. 특히, AI가 생성한 콘텐츠에 대해 저작권을 누구에게 귀속시켜야 하는지에 대한 명확한 규정이 부족해 법적 분쟁이 발생할 소지가 있다.

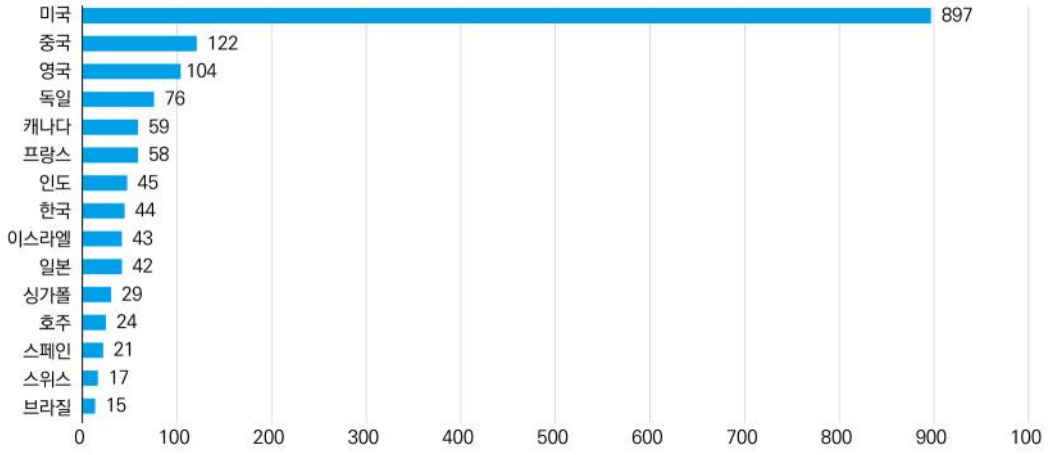
AI 생태계 활성화에 있어 투자도 매우 중요한 요소이다. 미국은 지난 10년 동안 AI에 약 3천350억 달러(약 452조 원) 규모의 민간투자를 진행했다. AI 투자 부분에선 미국, 중국, 영국이 각각 1, 2, 3위를 기록했다. 미국은 2023년 약 5,500개 AI 신생 기업에 약 672억2천만 달러(약 91조 원)를 투자했으며 2024년에만 AI 관련 7만 1천 개의 일자리 공고를 냈다. 이는 전체 구인 공고의 1.62%에 해당하는 수준이다.

[표 2-5] 국가별 AI 민간투자 및 스타트업 수

Country	Private investments 2023	Private investments 2013-2023	AI Startups 2013-2024
United States	$67.22B	$335.24B	5509
China	$7.76B	$103.65B	1446
UK	$3.78B	$22.25B	727
Singapore	$1.14B	$6.25B	193
Canada	$1.61B	$10.56B	397
South Korea	$1.39B	$7.25B	189
Israel	$1.52B	$12.83B	442
Germany	$1.91B	$10.35B	319
Japan	$0.68B	$4.81B	333
Australia	$0.37B	$3.40B	147
India	$1.39B	$9.85B	338
France	$1.69B	$8.31B	391
Sweden	$1.89B	$2.88B	94

자료: ZeroBounce Research(2024) "Countries that are AI superpowers"

미국은 2023년 한해에 새롭게 투자받은 AI 스타트업 수가 897개로 1위를 차지했다. 중국은 122개로 2위를 차지했으며 3위는 영국 104개였고 한국은 44개이다.



[그림 2-26] 2023년 새롭게 투자된 AI 기업

자료: Stanford University(2024) "AI Index Report 2024"

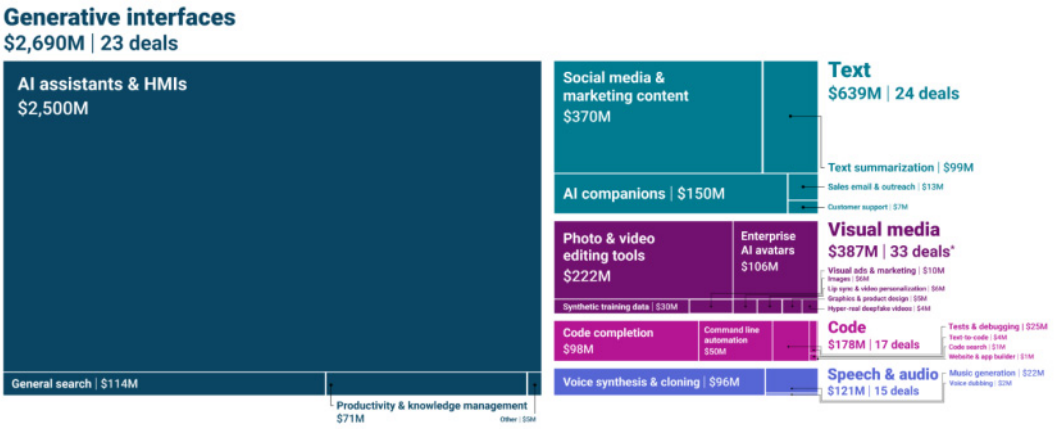
AI 사업이 잠재력과 함께 불확실성이 존재함에도 불구하고 글로벌 빅테크 기업들이 AI 관련 투자를 확대하고 있다. 2024년 상반기 마이크로소프트(MS)와 아마존, 메타, 구글 모회사 알파벳의 AI 투자는 총 1,060억 달러(약 144.5조 원)로 작년 동기보다 50% 증가했다. 이중에서 마이크로소프트는 330억 달러(약 45조 원)로 78%, 알파벳은 252억 달러(약 34.5조 원)로 90% 급증했다. 빅테크 기업의 경영진들은 선제적 투자 개념으로 인식하고 있다. 마크 저커버그 메타(Meta) CEO는 "2024년 AI 관련 자본지출이 400억 달러(약 54.5조 원)에 이를 수 있으며 늦기 전에 AI 역량을 구축하는 것이 위험을 줄이는 것이다"라고 밝혔다. 순다르 피차이 구글 CEO도 "기술 분야에서 이런 전환기를 겪을 때 AI에 대한 과소 투자의 위험이 과잉 투자의 위험보다 훨씬 크다"고 언급했다.

글로벌 생성형 AI 투자는 지속 증가 추세로 2023년 관련 투자가 대폭 증가했으며 거래 건수도 지속 증가하고 있다.

[그림 2-27] 글로벌 생성형 AI 투자 규모 및 거래 건수

자료: CB Insight, 삼정KPMG

2022년 3분기~2023년 2분기 동안 글로벌 생성형 AI 투자 분야는 AI 어시스턴트와 HMI(Human machine interaction) 분야가 가장 높았고, 소셜미디어와 마케팅, AI 컴패니언(Companions), 사진 및 비디어 생성 분야 등이 그 뒤를 이었다.



[그림 2-28] 22년 3분기~23년 2분기 글로벌 생성형 투자 분야

자료: CB Insight

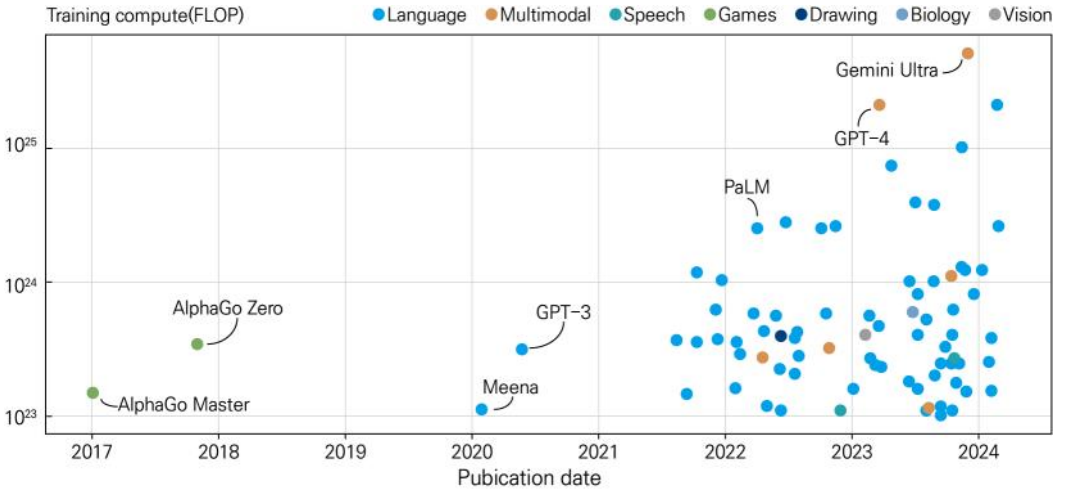
하지만 불확실한 수익성은 여전히 해결해야 할 문제다. AI 가속기 GPU와 고대역폭메모리(HBM) 등 AI 인프라 설비투자(CAPEX)는 눈덩이처럼 불어나지만, 최종 수요 시장에서 그 이상 매출을 만들어낼 수 있을지 의구심이 커지고 있다. 학계와 투자은행(IB) 업계를 중심으로 AI 산업 잠재력을 둘러싼 회의론도 제기되고 있다.

제3장

AI 혁명 시대, AI 정책변화의 특징과 전망

제1절 AI의 빠른 진화와 정책의 고민

AI 기술의 빠른 진화로 다수의 초거대 AI 모델이 출시되며 경쟁 중이고, 이에 따라 AI의 성능도 급격히 증가하는 추세이다. 2020년~2023년까지 4년간 총 144개의 초거대 AI 모델이 출시되었으며 2024년에도 지속 출시 중이다.



[그림 3-1] 초거대 AI 모델 출시 현황

자료: epochai.org

2024년 7월, 구글 딥마인드(Deepmind)는 수학 AI 알파프루프(AlphaProof)와 기하학 문제를 푸는 AI인 알파지오메트리2(AlphaGeometry2)가 국제수학올림피아드(International Mathematical Olympiad)[31]에서 은메달 수준의 성적을 기록했다고 발표했다.[32] AI가 국제수학올림피아드의 메달권에 해당하는 성능을 보인 건 처음이다.

31) 1959년부터 열린 IMO는 예비 수학자들이 겨루는 권위 있는 대회이며 수학계 노벨상으로 불리는 필즈상 수상자 다수가 IMO 대표로 참가했고 AI 분야가 떠오르면서 IMO는 AI의 수학적 추론 능력을 평가하는 기준으로도 사용

32) Deepmind(25 July 2024), "AI achieves silver-medal standard solving International Mathematical Olympiad

2024년 9월 발표된 Open AI의 최신 모델인 “o1”은 노르웨이 멘사 IQ 테스트에서 약 120의 IQ를 기록했으며 이는 AI 모델이 평균 인간 IQ를 넘어선 최초의 사례가 될 가능성이 있다. 미국의 데이터 분석가인 맥심 로트(Maxim Lott)는 Open AI가 공개한 AI인 o1이 노르웨이 멘사의 IQ 테스트에서 120을 기록했다고 밝혔다.[33]

[그림 3-2] 주요 초거대 AI 모델의 IQ 테스트 결과

자료: https://trackingai.org/IQ

AI의 빠른 진화로 규제 시차(Regulatory Lag) 이슈가 제기되며 AI 정책변화를 야기 중이며 이에 3장에서는 주요국 사례를 통해 AI 정책변화의 특징을 분석하고 시사점을 도출하고자 한다. 규제 시차는 규제 문제의 발생과 해결되는 시점 사이의 시간 차이를 뜻하며 기술 발전의 속도가 빠를수록 규제 공백 또는 규제 방치(regulatory drift)로 인한 사회적 손실이 커질 수 있어 규제 시차는 정책적 논의 대상이다.[34] 실제, 초거대 AI 논의가 활발히 시작된 2022년부터 AI로 발생한 사고 사례도 증가하고 있으며[35] 이에 AI 정책도 변

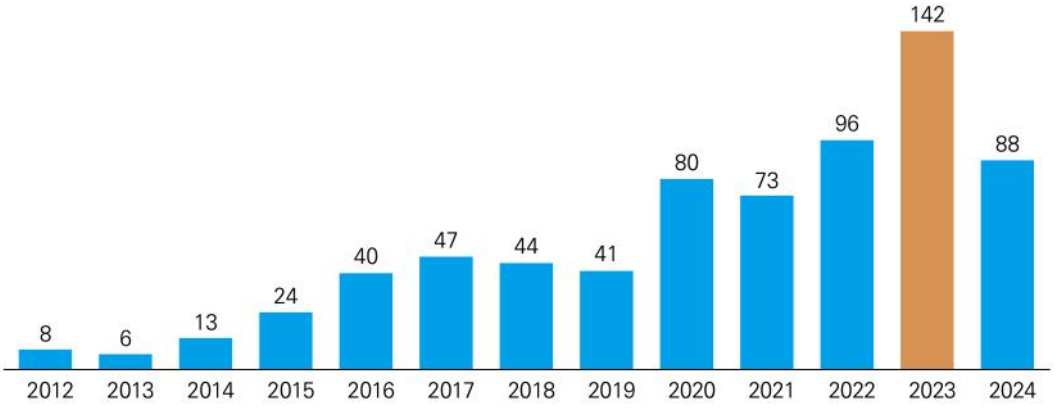
problems”

33) 멘사는 세계 최대 규모의 영재 모임으로, IQ 검사에서 일반 인구의 상위 2% 이내에 드는 지적 능력을 검증받아야 입회 자격이 주어지며 맥심 로트는 ‘TrackingAI’라는 사이트를 통해 주요 생성형 AI의 IQ 검사 결과를 꾸준히 공개

34) 최병선. 1992.『정부규제론』. 서울: 법문사

35) AIID는 오픈소스로 운영되는 DB로 2003년 이후 AI로 인한 사고를 집계 중이며 사건 목록에 2015년 구글의 사진 소프트웨어가 흑인을 고릴라로 분류한 사례나 아마존의 채용 도구가 여성 지원자를 차별해 폐기된 사례 등이 있다.

화가 일어나는 중이다.



[그림 3-3] 연도별 AI 사건 수

자료: AI Incident Database(2024.7.17. 기준), 연도별 수치는 SPRi

제2절 AI 정책 FGI 분석

1 FGI 개요

본 절에서는 AI 정책 수립을 위해 FGI(Focus Group Interview)에 참석한 주요 인사들의 의견과 논의를 정리한 내용으로 구성되어 있다. FGI에는 민관의 다양한 AI 분야 전문가들이 참석하여 각자의 전문성과 관점을 바탕으로 AI 정책 수립을 위한 다양한 의견을 나누었다. FGI의 목적은 초거대 AI의 등장 이후 과거와 달라진 AI 정책변화의 특징과 이슈를 논의하고 대응 방안을 모색하는 것이다. FGI는 국내외 AI 정책 동향에 대한 연구진의 브리핑을 시작으로 현황을 공유하고, 이어 각 분야 AI 전문가들의 의견을 듣고 논의하는 과정을 거쳤다. 참여 전문가 그룹은 AI 정책, AI 디지털교과서 이슈에 대해 별도로 논의하는 과정을 거쳤으며 본 절에는 AI 정책 전반에 대한 논의 내용을 정리하였다.[36] 참가자들의 소속과 직급, 그리고 참여 분야는 다음과 같다.

[표 3-1] AI 연구 FGI 참여자

참여자	소속	직급	분야
김○○	C 대학교 사범대학	교수	AI 교육, 인재
이○○	P 에듀테크 기업	대표이사	AI 교육, 인재
이○○	M AI 기업	전무	LLM, Vertical AI, 가상 인간
한○○	S 연구소(공공)	책임	AI/XR 정책, 가상 인간
유○○	S 연구소(공공)	책임	AI 안전 및 신뢰성 정책
김○○	S 보안 및 국방기업	이사	AI 보안, 국방 AI
윤○○	A IT 부품 기업	이사	AI 반도체, 부품, HW
손○○	K 연구소(통신 기업)	매니저	AI 연구개발, 융합 기획
권○○	S 컨설팅	대표이사	AI 인증, Testing, 정책

36) 세 주제에 모두 참여하지 못한 전문가는 별도의 심층 인터뷰를 통해 내용을 보완하였다.

2 FGI 분석

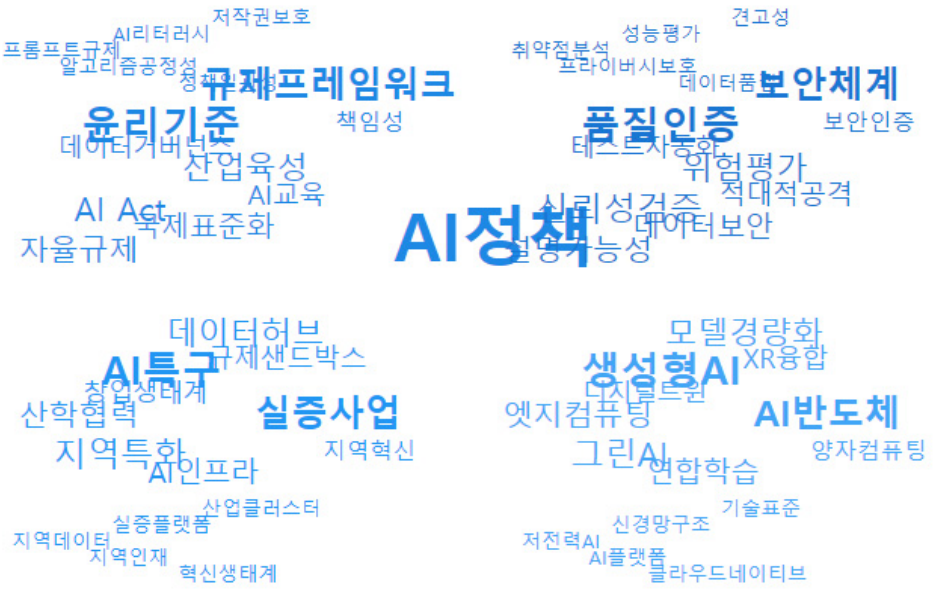
AI 정책 관련 주요 논의사항은 다음과 같다.

[표 3-2] AI 정책 FGI 주요 내용

구분	내용
K교수 (C 대학)	• "초기 AI 정책이 산업 진흥에 치중되었다면, 이제 AI 윤리와 규제의 균형이 필요한 시점이며 특히 교육 분야에서 AI 활용과 윤리교육이 동시에 고려 필요" • "미래 AI 기금 설립 논의가 이루어져 교육 등 다양한 분야에 활용될 것"
L대표이사 (P 에듀테크)	• "에듀테크 기업은 AI 규제가 엄격해지면 혁신이 저해될 우려" • "맞춤형 학습은 프라이버시 보호와 성능 개선이 동시에 고려 필요" • "AI 분야 규제 샌드박스를 통한 혁신실험이 필요" • "AI 인재 수급이 어려운 상황"
L전무 (M AI기업)	• "LLM 개발 기업으로, 세계 시장을 고려할 때 국가별 규제 차이가 큰 걸림돌" • "EU, 미국, 중국의 서로 다른 기준에 대응하기가 어려움" • "생성형 AI의 저작권 이슈와 규제는 글로벌 표준이 필요한 시점" • "AI 모델 경량화와 그린 AI도 중요한 이슈" • "미래 딥페이크는 공간과 오감이 연결되는 구조로 진화할 것"
H책임연구원 (S연구소)	• "XR과 AI가 융합된 환경에서는 더욱 복잡한 윤리적 이슈가 발생" • "AI 휴먼 관련 정책은 산업진흥과 프라이버시, 윤리적 측면을 고려해야" • "AI 신뢰성 검증을 위한 표준화된 평가체계가 필요" • "미래에 공간컴퓨팅과 AI 융합으로 인한 정책 이슈가 확산"
Y책임연구원 (S연구소)	• "AI 안전성 평가 기준과 신뢰성 검증 체계 구축이 시급. 특히 고위험 AI에 대한 명확한 기준이 필요" • "EU와 미국, 중국의 AI 규제 강도가 매우 상이" • "유럽의회처럼 국회도 AI 도입에 대해 적극적으로 임해야" • "AI 관점에서의 국회 상임위원회 구성에 대한 논의도 필요"
K이사 (S사)	• "AI 보안 측면에서 새로운 위협이 증가하고 있으며 국방 분야 AI는 더욱 엄격한 보안 정책이 필요" • "미래 AI 보편적 서비스 논의가 이루어질 것" • "다양한 산업과 AI 융합으로 예상하지 못한 정책 이슈가 제기될 것" • "국회와 정부도 미래 지향적으로 AI를 활용해야, 관련 사례가 등장하고 있어"
Y이사 (A사)	• "AI 반도체 산업에서는 국가별 기술 규제와 수출 통제가 심화"

구분	내용
	• “부품/소재 분야의 기술 주권 확보가 중요” • “AI 반도체의 기술 주권 확보를 위한 국가전략이 필요” • “데이터 기반으로 AI 생태계를 정부가 체계적으로 관리해야” • “저전력 친환경 AI 컴퓨팅에 대비해야”
S차장 (K사)	• “지자체별 AI 특구나 시범사업 요청이 증가.” • “국가별 AI 규제 체계의 차이가 두드러짐” • “AI 산업별 안전성 평가 기준이 마련 필요” • “지역 특화 AI 서비스를 위한 데이터 공유 체계 필요” • “미래에 Agent, AI 플랫폼 독점력, 지배력 이슈 논의가 확대될 것”
K대표이사 (S사)	• “AI 시스템의 품질과 안전성 검증을 위한 표준화된 테스트 기준이 필요” • “지자체별, 국가별로 다른 인증 요구사항이 존재” • “AI 품질인증의 국제 표준화가 시급” • “테스트 자동화를 통한 AI 시스템 신뢰성 검증 강화 방안 모색 필요”

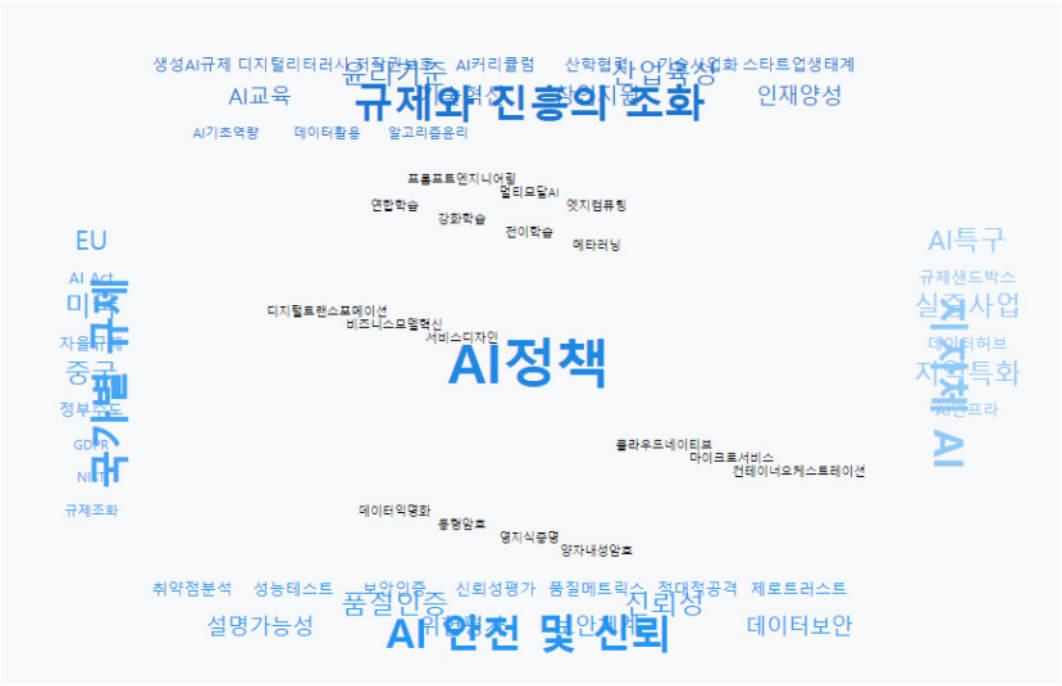
AI 전문가 FGI 논의사항을 체계적으로 정리하기 위해 언급된 세부 내용을 워드 클라우드(World Cloud) 분석을 통해 주요 키워드를 도출하였다. AI 정책 키워드를 중심으로 규제 프레임워크, 산업육성, 윤리, 보안, 특구, 실증사업 등이 주요 키워드로 나타났다.



[그림 3-4] AI 정책 워드 클라우드

다음 단계로 주요 키워드를 대상으로 토픽(Topic) 모델링 분석을 통해 연관성이 높은 키워드를 토픽으로 구성하여 범주화하였으며 범주화된 토픽은 크게 4가지인 규제와 진흥의 조화, AI 안전 및 신뢰, 국가별 규제, 지자체 AI로 구분되었다. 토픽의 비중은 규제와 진흥의 조화가 27%로 4가지 요소 중 가장 높았으나 다른 토픽과 큰 차이를 보이지 않았다.

AI 전문가들의 의견을 종합한 결과, 현재 AI 정책이 산업 진흥에서 윤리와 규제의 균형으로 전환되고 있는 시점임을 확인하였다. 또한, AI 안전과 신뢰성이 강화되고 있으며 신뢰성 확보를 위해서는 표준화된 평가체계 구축이 시급한 과제로 지적되었다. 특히 고위험 AI에 대한 명확한 기준 마련과 보안 측면에서의 새로운 위협 대응이 필요하다는 의견이 제시되었다. 기업에서는 국가별 규제 차이로 인한 대응의 어려움에 대해 의견을 제시하였는데 EU, 미국, 중국의 상이한 규제 기준은 기업 측면에서 큰 부담으로 작용하고 있었다. 지역별 이슈로는 지자체별 AI 특구와 시범사업 요청이 증가하고 있으며, 지역 특화 AI 서비스를 위한 데이터 공유 체계 구축이 필요하다는 의견이 제시되었다.



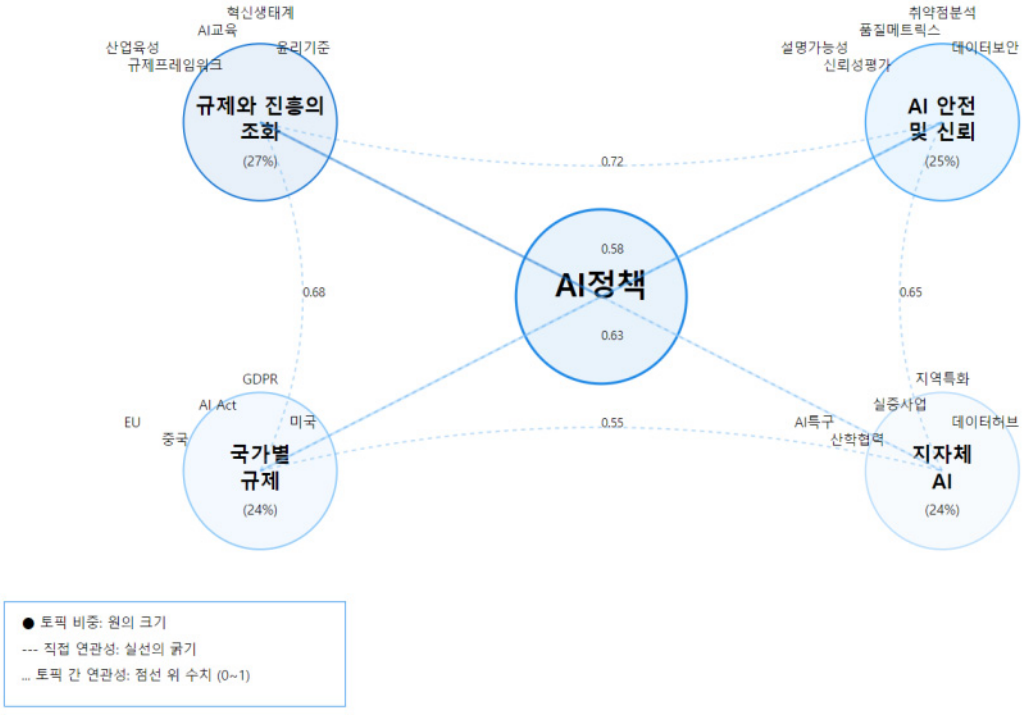
[그림 3-5] AI 정책 토픽 구성

논의사항에서 현재의 이슈와 함께 미래의 이슈도 제기되었다. AI 서비스의 보편화에 따라 'AI 보편적 서비스' 개념이 새롭게 논의될 것으로 전망된다. 이는 기본적인 AI 서비스에 대한 접근권을 모든 시민에게 보장해야 한다는 관점에서, 향후 중요한 정책적 과제로 부상할 것으로 보인다. 산업 측면에서는 AI 에이전트(Agent)의 진화와 이로 인한 플랫폼 독점력 문제가 새로운 이슈로 대두될 것으로 예상된다. AI 에이전트가 일상생활과 비즈니스 영역에서 광범위하게 활용되면서, 이를 제공하는 플랫폼의 시장 지배력과 관련된 규제 이슈가 확대될 것으로 전망된다. AI 발전에 따른 새로운 도전 과제로 딥페이크 기술의 진화에 대한 논의도 포함되었다. 현재의 시각과 청각 중심 딥페이크는 미래에 공간과 오감이 연결되는 복합적 구조로 진화할 것으로 예측되고 있다. 특히 XR(확장현실)과 AI의 융합은 기존에 경험하지 못한 새로운 윤리적 이슈를 야기할 것으로 전망된다. 가상공간에서의 정체성, 프라이버시, 데이터 주권 등 복잡한 문제가 대두될 것으로 예상된다. 공간컴퓨팅과 AI의 융합은 정책적 측면에서도 새로운 도전 과제를 제시할 것으로 보인다. 물리적 공간과 디지털 공간의 경계가 모호해지면서 기존 법체계로는 규제가 어려운 새로

운 형태의 서비스와 문제가 등장할 것으로 예측된다. 환경적 측면에서는 저전력 친환경 AI 컴퓨팅의 중요성이 더욱 부각될 것으로 전망되었다. AI 모델의 규모가 커지고 연산량이 증가함에 따라, 에너지 효율성과 환경 영향을 고려한 그린 AI 기술개발이 미래의 핵심 과제가 될 것으로 논의되었다. 다양한 산업과 AI의 융합이 가속화되면서 현재로서는 예측하기 어려운 새로운 정책 이슈들이 지속적으로 제기될 것으로 보인다. 특히 의료, 금융, 교통 등 고도로 규제된 산업에서 AI 도입으로 인한 예상치 못한 문제들이 발생할 가능성이 높다. 이러한 미래 이슈들에 대응하기 위해서는 선제적인 정책 수립과 규제 프레임워크 마련이 필요하며 특히 국제적 협력을 통한 글로벌 표준 수립이 중요한 과제로 대두될 것으로 예상된다.

아래 그림은 AI 정책과 관련된 4개 주요 토픽 간의 상호 연관성을 분석한 것이다. '규제와 진흥의 조화'와 'AI 안전 및 신뢰' 간 연관성이 0.72로 가장 높은 것으로 나타났다. 이는 AI 기술의 발전과 활용을 촉진하면서도 안전성과 신뢰성 확보가 상호보완적 관계에 있음을 시사하는 것이다. 'AI 안전 및 신뢰'와 '국가별 규제' 간 연관성은 0.63으로 로 나타나고 있다. 이는 AI의 안전성과 신뢰성 확보가 국가 차원의 규제 강도와 긴밀한 관련성을 가지고 있으며, 각국의 규제 정책이 AI 신뢰 구축에 중요한 역할을 하고 있음을 보여주는 것이다.

도출된 4가지 특성은 AI 정책의 특성을 이해하는 하나의 기준으로 활용될 수 있으며 3절에서는 이 특성에 기반하여 국내외 AI 정책이 세부적으로 어떻게 실행되고 있는지 검토하기 위해 다양한 사례 분석을 추가하였다.



[그림 3-6] AI 정책 토픽 비중과 연관성

제3절 AI 정책변화의 특징

1 진흥에서 규제와 진흥의 조화로

2016년 알파고 대국 이후 AI 잠재력에 주목한 주요국들은 진흥 중심의 AI 정책을 지속 발표해왔다. EU는 2018년 AI 전략(Communication Artificial Intelligence for Europe)을 발표하고 AI 산업 육성을 위해 2020년까지 15억 유로 투자 계획을 수립했다. 미국은 2016년 AI 국가 연구개발 전략을 발표했으며, 2019년 AI 연구개발과 투자에 우선순위를 두는 AI 행정명령(American AI Initiative)을 추진했다. 주로 민간이 추진하기 어려운 차세대 연구개발 및 군사 안보 분야 활용에 중점을 두었다. 영국은 2018년 AI 부문 간 합의(AI Sector Deal) 발표를 통해 생산성 향상, 기업 유치, 인프라 구축, 인력양성 등 AI 관련 5개 분야별 육성 정책을 제안했다. 민간과의 협력을 기반으로 AI 인재 양성 및 비즈니스 환경조성에 투자를 집중한다는 것이다. 독일도 2018년 AI 육성전략을 발표하며 AI를 활용한 중소 및 제조 분야 경쟁력 제고를 위해 대규모 투자를 계획하고 기술력 확보를 위해 노력했다. 프랑스는 2018년 AI 권고안을 발표하고 AI 생태계 조성, 전략 분야 융합 및 직업·고용, 윤리 등 이슈를 포함하였다. 중국은 2017년 차세대 AI 발전 계획을 발표하며 AI와 데이터 분야 대규모 투자 및 인력양성을 추진하고, 선도기업을 지정하여 산업별 특화플랫폼을 육성했다. 정부 주도로 자국 기업을 활용한 산업별 플랫폼을 구축, 막대한 데이터를 축적함으로써 AI 경쟁력을 확보한다는 것이다. 일본은 2017년 AI 기술 전략을 발표하고 이후 AI 전략을 지속 업데이트해왔다. 2017년 AI 기술 전략 발표 이후 2019, 2021, 2022년 AI 전략을 발표하며 국가전략을 수정했다.

<table>
<tr><th colspan="2"></th><th>2016</th><th>2017</th><th>2018</th><th>2019</th><th>2020</th><th>2021</th><th>2022</th><th>2023</th></tr>
<tr><td rowspan="2">세계</td><td>주요
이슈</td><td></td><td>알파고
쇼크</td><td>GPT－1
발표</td><td>GPT－2
발표</td><td colspan="3">GPT－3 발표
코로나19 팬데믹</td><td>GPT－4
발표</td></tr>
<tr><td rowspan="3">일본</td><td>전략</td><td>Society
5.0</td><td>AI기술
전략</td><td></td><td>AI 전략
2019</td><td></td><td>AI 전략
2021</td><td>AI 전략
2022</td><td>반도체
디지털
전략(개정)</td></tr>
<tr><td>주요
조직</td><td colspan="4">AI기술전략회의</td><td></td><td colspan="2">새로운AI
전략검토회의</td><td>AI전략
회의</td></tr>
<tr><td colspan="2"></td><td colspan="6">인간중심 AI 사회원칙회의</td></tr>
</table>

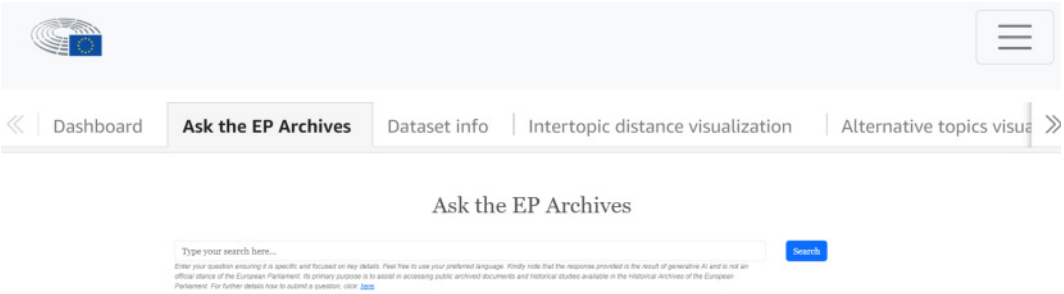
[그림 3-7] 연도별 주요 이슈와 일본 AI 정책

자료: KOTRA(2024) “일본의 AI 정책과 실제 사례”

2022년 이후, 초거대 AI가 부상하면서 주요국들은 AI의 빠른 진화를 인식하고 진흥과 규제의 조화방안을 모색하고 있다. EU는 2024년 최초의 AI 규제법안인 EU AI 법을 최종 승인했다. 입법의 목적은 인간중심의 신뢰할 수 있는 AI의 촉진, AI 시스템으로부터의 안전, 환경 보호, 법치 등 기본권을 보호하고 혁신을 지원하는 것이다. EU 내에서 사용되는 AI 시스템에 대한 EU의 가치가 일관되게 적용될 수 있는 원칙을 수립했다. EU AI 법은 개발자, 공급자, 배포자 모두에게 적용되며 국가 안보 목적, 국제협약에 따른 법 집행 및 사법공조, 시장 출시 전 테스트 및 개발, 과학적 연구개발 등의 적용 예외 사항이 존재한다.

EU AI 법은 규제만 규정한 법률은 아니며, AI 혁신과 발전을 위한 규제 샌드박스, 중소기업 지원의 사항 등을 포함하고 있다. EU는 회원국이 국가 차원에서 시장에 출시되기 전에 AI 시스템을 개발, 테스트 및 검증할 수 있는 환경으로 AI 규제 샌드박스를 만들 것을 규정하고 있다. 또한, 고위험 AI 시스템 공급자(잠재적 공급자 포함)는 AI 법안에 규정된 모든 조건을 충족한 경우에는 규제 샌드박스가 아닌 실제 환경에서도 테스트할 수 있다. 그리고 고위험 AI 시스템을 공급하는 중소기업의 문서 관리 부담 완화를 위해 기술문서를 간소화하고 적합성 평가 수수료를 감액하고 있으며 중소기업에게 AI 규제 샌드박스 우선 참여권을 제공하고, 안내 및 자문을 위한 전용 상담 창구도 제공한다.

또한, 유럽의회(European Parliament)는 사용자가 자연어를 활용하여 유럽의 방대한 법제도 자료에 쉽게 접근할 수 있도록 생성형 AI ‘Ask the EP Archives’를 2024년 10월 도입하였다. AI 기업 클로드(Claude)와 협력하여 생성형 AI 챗봇을 개발하고 연구원, 정책입안자, 교육자, 대중이 210만 개 이상의 공식 문서를 손쉽게 사용할 수 있도록 지원하고 있다.



[그림 3-8] 유럽의회의 생성형 AI 서비스 Ask the EP Archives

자료: https://archidash.europarl.europa.eu/ep-archives-anonymous-dashboard

미국 바이든 행정부는 2023년 AI 행정명령을 발표하고 국가 안보, 공공 안전 등에 영향을 미치는 AI 기술개발과 이용을 규제하기 위해 8개 원칙을 제시했다. 행정명령은 ① AI의 안전 및 보안 보장, ② 혁신 경쟁 협력 촉진, ③ 책임감 있는 AI 개발 이용, ④ 평등과 시민권, ⑤ 소비자 보호, ⑥ 국민의 프라이버시 및 자유 보호, ⑦ 연방정부의 AI 역량 제고, ⑧ 글로벌 주도의 내용을 포함하고 있다. 미국은 행정명령 발표문에서 "미국은 적극적으로 AI 규제 의제를 제시하는 국가이며 앞으로 AI의 개발 및 사용을 관리하기 위한 국제적 체계를 만들고 해외 동맹국 및 파트너와 협력할 것"이라고 밝혔다.

영국 정부는 2023년 AI 규제에 대한 방향과 원칙이 담겨있는 AI 규제백서를 발표하며 AI 규제 체계 기반을 마련했다. AI 규제백서에서는 ① AI 주도 국가로서의 영국의 입지 강화, ② 혁신 도모 및 규제 불확실성 감소를 통한 AI의 성장과 번영, ③ 위험 규율 및 기본 가치 보호를 통한 AI에 대한 대중의 신뢰 제고라는 3가지 목표를 추구한다. AI 규제백서에서는 친혁신적인 규제 Framework의 5가지 원칙을 제시했는데 세부 내용은 아래 표와 같다.

[표 3-3] 친혁신적인 규제 Framework의 5가지 원칙

구분	내용
안전성, 보안성 및 견고성	• AI 수명주기 전반에 걸친 안정적인 작동을 위한 지속적인 안전 평가, 식별 및 관리체계 도입의 필요성을 강조
적절한 투명성 및 설명 가능성	• AI 시스템의 목적, 사용 방법 등에 관한 정보를 관계자에게 알릴 필요성을 강조 • 적절한 투명성과 설명 가능성의 정도는 AI 시스템이 초래할 위험에 비례함
공정성	• 공정성은 평등 및 인권 관련 규제, 개인 정보보호, 소비자 법, 경쟁법, 공법 등 여러 영역에 걸쳐 내포된 개념임을 명시 • 규제기관이 상황에 따라 관련 법률, 규정, 기술 표준 내에서 적용할 수 있는 공정성의 예시를 개발하고 게시할 수 있음
책무성과 거버넌스	• AI 시스템의 공급과 사용에 대한 실효성 있는 감독을 위해 AI 수명주기 전반에 걸친 명확한 책무성을 명시할 필요를 강조 • AI 공급망 속 적합한 행위자에게 규제 준수의 의무가 있음을 기술 표준이나 모범 사례 지침서 등을 통해 명확히 할 것을 강조
이의제기 가능성과 보상	• 경우에 따라, AI 시스템 사용자나 AI 활용으로 인해 영향을 받은 제3자 등이 유해한 AI 결정에 이의를 제기할 수 있는 경로를 만들어 놓아야 함 • 이의제기 경로는 쉽게 접근할 수 있도록 규제기관이 안내하기를 권고

자료: UK(2023), "A pro-innovation approach to AI regulation"

일본 경제산업성과 총무성은 2024년 AI 사업자 지침을 발표하고 AI 법 제정 논의를 시작했다. AI 사업자 지침은 총 5개 장으로 구성되어 있는데 AI 관련 개발자, 제품 및 서비스 제공자, 이용자를 대상으로 10개 원칙을 제시하였으며 해당 지침에 4,000여 건의 의견이 접수되었고 이를 반영하였다. AI 전략회의에서는 AI 규제 기본방침과 AI 안전성 확보를 위한 법률 규제 방침을 밝혔으며(2024.5), 정기 국회 법안 제출 후, 2026년 전면 시행이라는 계획도 제시하였다.

[표 3-4] 일본 AI 사업자 지침(Guide Line) 10개 원칙

① (인간중심) 인간 존엄과 개인 자유를 존중
② (안전성) 인간에 의한 통제력 확보
③ (공평성) 부당한 차별 최소화
④ (프라이버시 보호) 개인정보 보호법에 근거하여 대응
⑤ (보안 확보) 시스템 기밀성 유지
⑥ (투명성) Data 수집 방법 등을 대외에 공개
⑦ (설명책임) AI에 대한 이념, 사상을 공표
⑧ (교육) 올바른 지식을 알림
⑨ (공정 경쟁 확보) 이해관계자에 유리하지 않게 대응
⑩ (혁신) 사회 전체 기술 혁신에 공헌

자료: 内閣府, AIを巡る主な論点, 2023.5.15.

국내에서도 21대 국회에서 AI 법 도입 논의가 활발히 이루어졌으나 회기가 종료되면서 폐기되었고 22대 국회에서 재논의 중이다. 국내 AI 법안은 21대 국회에서 9개가 발의되었고 9건의 법안 소관 상임위(과방위)에서는 2022년까지 발의되었던 7건의 AI 법을 병합하여 심사(2023. 2. 14.)한 이후, 이를 통합·조정하여 위원회의 대안을 제시하기로 하였으나, 비공개로 수정 보완 중인 상태에서 21대 국회 회기가 종료되면서 폐기되었다.[37] 21대 국회에서 논의되던 AI 법은 시민단체와 권익위원회 등으로부터 도입을 전면 재검토 요청받았다. 2024년 3월, 15개 시민사회단체는 21대 국회 과방위 심사 소위를 통과한 AI 법에 대해 도입 반대 의견서를 제출하였다. 해당 의견서에는 AI 법이 국민의 권익과 존엄성을 보호하고 국민의 삶의 질 향상에 부합하지 않을 뿐만 아니라, 현행 지능화 정보법과도 목적과 내용이 유사해 입법의 필요성도 의문이며, 국제 인권 규범 및 해외 입법정책의 예에 비추어 안전과 인권에 미치는 인공지능 위험 규제의 효과를 제대로 담보할 수 없고, 소관 부처도 적절하지 않아 전면 재검토할 것을 요구한 것이다.[38]

국가인권위원회도 제21대 국회 과학기술정보방송통신위원회 AI 법 심사 과정에서, 과학기술정보방송통신위원회에서 심의 중인 AI 법은 AI 개발·활용 과정에서 발생할 수 있는 인권침해, 차별 등의 문제를 예방 및 규제하기 위한 규정이 미흡하므로, 수정할 필요가

37) 법제처(2024), "인공지능 관련 국내외 법제 동향"
38) https://www.peoplepower21.org/publiclaw/1927781, 「인공지능산업 육성 및 신뢰 기반 조성 등에 관한 법률안(소위안)」에 대한 인권시민단체 의견

있다는 의견을 표명하였다. AI, 특히 학습기반 AI는 이용자가 제공한 정보뿐만 아니라 온라인상에서 수집한 방대한 정보를 기반으로 서비스를 제공하여 정보수집 과정에서 이용자와 정보 주체의 권리를 침해할 소지가 많은데, AI 법은 이용자와 정보 주체의 권리 및 권리 침해 시의 구제절차를 명확히 규정하지 않고 있다고 언급했다. 이에, AI 이용자와 정보 주체의 권리를 명시하고, 피해 발생 시 이를 구제하기 위한 절차를 마련해야 한다고 설명했다. 또한, 국제 사회와 세계 각국은 AI 기술의 위험성을 단계별로 구분하고 규제 정도를 달리하는 방안을 마련하고 있는데, AI 법은 단지 고위험 영역 인공지능에 대해서만 규율하고 있어 AI를 국민의 인권과 안전에 미치는 영향, 위험성 등을 고려하여 적정한 등급으로 구분하고, 각 등급에 맞게 규제 수준을 달리하여야 한다고 강조했다. 우선 허용·사후규제 원칙은 삭제를 요청했다. 이 원칙에 따라 산업의 경제성·효율성만 따져서 AI 기술을 개발· 활용할 경우, 사전·사후 평가 없이 개발·활용된 인공지능 기술이 국제 기준에 못 미치는 결과를 초래할 수 있고, 이로 인해 오히려 국제 경쟁력을 저하시키고 나아가 인공지능 기술의 신뢰성마저 떨어뜨릴 우려가 있다고 지적하였다.[39]. 22대 국회에서도 AI 법안이 11건 발의(2024년 9월 30일 기준)되어 있다.

[표 3-5] 22대 국회에 발의된 AI 법

의안명	제안자명	제안일자	진행상태
인공지능산업 진흥 및 신뢰 확보 등에 관한 특별법안	김우영의원 등 19인	2024-09-24	소관위접수
인공지능의 발전과 안전성 확보 등에 관한 법률안	이훈기의원 등 14인	2024-09-12	소관위접수
인공지능 발전 진흥과 사회적 책임에 관한 법률안	배준영의원 등 10인	2024-08-28	소관위접수
인공지능책임법안	황희의원 등 10인	2024-08-27	소관위접수
인공지능 기본법안	한민수의원 등 10인	2024-08-22	소관위접수
인공지능 개발 및 이용 등에 관한 법률안	권칠승의원 등 15인	2024-07-04	소관위심사
인공지능기술 기본법안	민형배의원 등 13인	2024-06-28	소관위심사
인공지능산업 육성 및 신뢰 확보에 관한 법률안	김성원의원 등 11인	2024-06-19	소관위심사
인공지능산업 육성 및 신뢰 확보에 관한 법률안	조인철의원 등 19인	2024-06-19	소관위심사
인공지능 발전과 신뢰 기반 조성 등에 관한 법률안	정점식의원 등 108인	2024-06-17	소관위심사
인공지능 산업 육성 및 신뢰 확보에 관한 법률안	안철수의원 등 12인	2024-05-31	소관위심사

자료: 의안정보시스템(2024.9.30.기준)

39) 국가인권위원회, '「인공지능산업 육성 및 신뢰 기반 조성 등에 관한 법률안」에 대한 의견표명', 2023.7.13.

제정안들은 지난 제21대 국회 과학기술정보방송통신위원회 정보통신방송법안심사소위원회에서 대안으로 제안하기로 의결('23.2.14.)한 안 및 현재 계류 중인 안철수의원안[40]과 체계·내용이 대체로 유사하며, 다만 각각 다음과 같은 차이가 있다.

첫째, 정점식·민형배 의원안의 경우 최신 동향을 반영하여 '생성형 AI'를 정의하고, 이에 대한 규제를 포함하고 있다. 해당 법안에서는 생성형 인공지능을 "글, 소리, 그림, 영상 등의 결과물을 다양한 자율성의 수준에서 생성하도록 만들어진 인공지능"으로 정의하였다. OECD도 2023년에 AI 권고안에 포함된 AI 시스템 정의를 개정하였는데 생성형 AI의 활용 확산 등 기술과 서비스 환경변화를 반영하여 AI 시스템을 재정의하였다.

[표 3-6] OECD 2019년 및 2023년 AI 시스템 정의 비교

구분	내용
2019년 정의	• AI 시스템은 사람이 정의한 주어진 일련의 목표를 위해 현실·가상 환경에 영향을 미치는 예측·권고·결정을 내릴 수 있는 기계 기반 시스템 • 다양한 수준의 자율성 하에서 작동하도록 설계된 시스템
2023년 정의	• AI 시스템은 명시적·암묵적 목표를 위해 받은 입력으로부터 물리적·가상 환경에 영향을 미칠 수 있는 예측·콘텐츠·추천·결정 같은 결과물의 생성 방식을 추론하는 기계 기반 시스템 • AI 시스템은 배포 후 자율성 및 적응성 수준이 다양함

자료: OECD(2024), "Explanatory memorandum on the updated OECD definition of an AI system"

둘째, 권칠승의원안의 경우 '금지된 AI'의 개발과 이용을 금지하고, 금지된 AI 및 고위험 AI 관련 의무 위반 시 처벌하도록 하고 있다. 셋째, 정점식의원안의 경우 국외에서 이루어진 행위라도 국내 시장·이용자에게 영향을 미치는 경우 AI 법을 적용할 수 있도록 하고 있다. 넷째, 위원회 구성과 관련하여 정점식·조인철의원안은 AI 위원회를 대통령 소속으로, 김성원·권칠승의원안은 AI 위원회를 국무총리 소속으로 하고 있다. 한편, 민형배의원안의 경우 위원회를 국무총리 소속의 국가 AI 위원회와 지자체의 지역 AI 위원회로 구분한다. 넷째, 정점식의원안의 경우 국가 AI 센터와 AI 안전연구소를 인공지능 관련 조직으로 둘 수 있도록 하고 있으며 김성원·권칠승의원안은 국가 AI 센터를, 민형배의원안은 AI 정책센터를 둘 수 있도록 하였으며, 조인철의원안의 경우 별다른 규정을 두고 있지 않

40) 인공지능 산업 육성 및 신뢰 확보에 관한 법률안, 의안번호 제2200053호

다. 다섯째, 조인철의원안의 경우 혁신적인 AI 시스템에 대한 사전 개발, 테스트, 검증 등을 위한 규제 샌드박스를 도입하도록 하는 한편, AI 제품에 비상정지 기능을 적용하도록 하고 있으며, 김성원의원안의 경우 우선 허용·사후규제 원칙을 규정하고 있다.[41]

계류 중인 AI 법에 대한 부처의 의견도 다양하며, 정점식의원 등 5개 법안[42]에 대한 부처의 의견은 아래 표와 같다.

[표 3-7] 계류 중인 AI 법에 대한 주요 부처의 의견

구분	내용
문화체육관광부	• 국가인공지능위원회의 심의·의결 사항으로 규정된 내용 중 '저작권 침해'에 대하여, 일반적인 저작권 침해 관련 사항은 「저작권법」에 따라 한국저작권위원회와 한국저작권보호원심의위원회에서 규율·심의하고 있으므로, 인공지능위원회의 심의·의결 사항은 '인공지능 관련 저작권 침해'로 한정할 필요가 있다는 의견
개인정보보호위원회	• 국가인공지능위원회의 심의·의결 사항으로 규정된 내용 중 '개인정보 침해 방지'에 대하여, 개인정보 보호에 관한 사무는 「개인정보 보호법」에 따라 설립된 개인정보보호위원회의 소관 업무이므로 해당 내용의 삭제가 필요하다는 의견
행정안전부	• 국가인공지능위원회와 같은 자문위원회를 설치할 경우, 「행정기관 소속 위원회의 설치운영에 관한 법률」 제11조 제2항[43]에 따라 위원회의 존속기한을 명시할 필요가 있다는 의견 • 또한, 국가인공지능위원회에 지원단을 두도록 한 안에 대하여, 「행정기관 소속 위원회의 설치·운영에 관한 법률」 제10조[44]에서는 원칙적으로 자문위원회에 사무기구를 설치할 수 없도록 하고 있고, 과기정통부 인공지능기반정책관에서 위원회 사무 지원이 가능하므로 지원단 설치조항의 삭제가 필요하다는 의견[45]. • 이에 대하여 과학기술정보통신부는 국가 인공지능 정책방향에 대한 실질적인 논의를 뒷받침하는 별도 지원조직이 시급하고, 대통령이 위원장인 주요 자문위원회[46] 및 그간 신설된 주요 민관합동위원회[47]도 상설지원조직을 두고 있다는 의견
교육부	• 국가인공지능위원회의 법령·제도 개선 등에 대한 권고 및 의견표명이 국가기관 입장에서 과도한 부담이 될 수 있으므로 해당 조항을 삭제하거나, 법률 또는 대통령령에 구체적인 사항을 명시할 필요가 있다는 의견
과학기술정보통신부	• 대통령 소속 위원회의 의사결정 결과에 대하여 국가기관등의 장이 개선방안·실천방안 등을 수립하도록 규정하는 것은 행정효율성 측면에서 필요한 부분이며,

41) 제417회 국회(임시회) 제7차 과학기술정보방송통신위원회 인공지능 산업 육성 및 신뢰 확보에 관한 법률안 등 검토보고 (2024.8)
42) 정점식 의원, 조인철 의원, 김성원 의원, 민형배 의원, 권칠승 의원

구분	내용
	「국가인권위원회법」 등의 경우에도 권고 또는 의견표명의 성립 요건 등을 위한 구체적인 절차를 별도로 규정하지 않고 운영 중이라는 견해

자료: 제417회 국회(임시회) 제7차 과학기술정보방송통신위원회 인공지능 산업 육성 및 신뢰 확보에 관한 법률안 등 검토보고(2024.8)

2 AI 안전과 신뢰 정책의 강화

주요국들은 AI의 빠른 진화로 야기될 위험에 대비하기 위해 안전 및 신뢰 정책을 강화하고 있다. EU AI 법은 위험 수준에 따라 AI 시스템에 대한 규제 수준을 차등화하는 위험 기반 접근(Risk-based approach) 방식이다. 특정한 AI 시스템을 금지하거나 고위험 AI 시스템, 제한된 위험을 갖는 AI 시스템, 최소위험 AI 시스템 등으로 분류하여 차등 규제한다. 수용 불가 AI 시스템은 EU AI 법에서 명문으로 규정된 특별한 예외가 없는 한 사용 자체가 금지된다.

43) 제11조(위원회의 존속기한) ② 행정기관의 장은 자문위원회등을 설치할 경우 목적 달성을 위하여 필요한 최소한의 기한 내에서 존속기한을 정하여 법령에 명시하여야 한다. 이 경우 존속기한은 5년을 초과할 수 없다.
44) 제10조(위원회의 사무기구 등) ② 자문위원회등에는 사무기구를 설치하거나 상근(常勤)인 전문위원 등의 직원을 둘 수 없다. 다만, 행정기관의 장이 직접 사무 지원을 하기 곤란한 위원회로서 대통령령으로 정하는 요건에 해당하는 경우에는 그러하지 아니하다.
45) 「우주개발 진흥법」에 따른 국가우주위원회도 대통령 소속 자문위원회이나 별도 사무기구 설치 없이 우주항공청에서 사무를 지원하고 있다는 의견임.
46) 국민경제자문회의, 국가과학기술자문회의, 저출산·고령사회위원회 등
47) 디지털플랫폼정부위원회, 국민통합위원회, 지방시대위원회 등

[표 3-8] 수용 불가 AI 시스템 사용 예시

구분	내용
잠재의식 또는 조작적, 속임수 기법으로 인간 의사결정 왜곡, 조작	• AI 시스템이 인간의 잠재의식을 이용하거나 의도적으로 조작을 이용하여 사람이 정보에 기반한 의사결정을 내릴 능력을 현저하게 저해하여 심각한 피해를 초래하거나 초래할 가능성이 있는 의사결정에 사용되는 경우
인간의 취약성을 악용하여 인간 행동 왜곡	• AI 시스템이 사람 혹은 특정 집단의 취약성(나이, 장애, 사회적 또는 경제적 상황 등)을 악용하여 사람의 행동을 심각하게 왜곡하여 심각한 피해를 초래하거나 초래할 가능성이 있는 경우
개인의 사회 점수(social scoring) 시스템	• AI 시스템이 일정 기간 사회적 행동이나 알려진, 추론된 또는 예측된 개인 또는 성격적 특성을 기반으로 자연인 또는 집단을 평가하거나 분류하는 경우
범죄 위험 평가, 예측	• AI 시스템이 자연인이 범죄를 저지를 위험성을 평가 또는 예측하기 위해 사용되는 경우
불특정 다수의 얼굴	• AI 시스템이 인터넷이나 CCTV 영상에서 얼굴 이미지를 비대상으로 스크래핑하여 얼굴 인식 데이터베이스를 생성하거나 확장하기 위해 사용되는 경우
생체인식 분류 시스템 사용	• 생체인식 분류 시스템을 사용하여 생체인식 데이터를 기반으로 개별 자연인을 분류하여 인종, 정치적 의견, 노조 가입, 종교 또는 철학적 신념, 성생활 또는 성적 지향을 추론하거나 유추하는 경우(단, 이 금지 사항은 생체 인식 데이터를 기반으로 합법적으로 취득한 생체인식 데이터 세트에 대한 라벨링 또는 필터링이나 법 집행 분야에서 생체인식 데이터 분류에는 적용되지 않음)
근로자 또는 학생의 감정 자동 인식	• 직장과 교육기관에서 자연인의 감정을 추론하기 위해 AI 시스템을 사용하는 경우(단, 이 금지 사항은 의료 또는 안전 목적을 위한 경우는 제외함)
실시간 원격 생체인식	• 법 집행 목적을 위해 공개적으로 접근할 수 있는 공간에서 '실시간' 원격 생체인식 식별 시스템을 사용하는 경우(단, 이러한 사용이 다음 목적 등에 필요한 경우에는 예외) • (예외 1) 납치, 인신매매 또는 성 착취 피해자에 대한 표적 수색 및 실종자 수색 • (예외 2) 사람의 생명 또는 신체적 안전에 대한 구체적이고 실질적이며 임박한 위협, 예측할 수 있는 테러 공격 위협의 예방 • (예외 3) 부속서 II에 언급된 범죄에 대한 형사 수사나 기소 또는 형사처벌을 집행하기 위한 목적으로 범죄를 저지른 것으로 의심되는 사람의 현지화 또는 식별

자료: EU AI ACT

고위험 AI 시스템에는 생체인식, 중요 인프라, 교육, 필수 서비스, 법 집행, 이주 및 사법에 사용되는 AI 등이 포함되며 고위험 AI 시스템 공급자와 운영자, 수입업자, 유통업자에게 의무가 부과되고 공공 서비스에 활용하거나 자연인의 신용도 평가 시, 기본권 영향 평가를 수행해야 한다.

[표 3-9] 고위험 AI 시스템 사용 예시

일정	내용
생체인식	(a) 원격 생체인식 식별 시스템 (b) 생체인식 분류에 사용되도록 의도된 AI 시스템 (c) 감정 인식에 사용되도록 의도된 AI 시스템
중요 인프라	(a) 중요 디지털 인프라, 도로 교통 관리 및 운영 또는 물, 가스, 난방 또는 전기 공급에서 안전 구성 요소로 사용되도록 의도된 AI 시스템
교육 및 직업 훈련	(a) 교육 및 직업 훈련 기관에 대한 접근 또는 입학을 결정하거나 사람 배정에 사용되도록 의도된 AI 시스템 (b) 학습 성과를 평가하는 데 사용되도록 의도된 AI 시스템 (c) 교육을 평가하는 목적으로 사용되도록 의도된 AI 시스템 (d) 학생의 금지된 행동을 모니터링하고 감지하는 데 사용되도록 의도된 AI 시스템
고용, 근로자 관리 및 자영업 접근	(a) 목표 구인 광고를 게재하고, 구직 신청서 분석 및 필터링하고, 후보자를 평가하기 위해 사람들을 모집 또는 선발하는 데 사용되도록 의도된 AI 시스템 (b) 사람들의 성과 및 행동을 모니터링하고 평가하는 데 사용되도록 의도된 AI 시스템
필수 개인 서비스 및 공공 서비스 및 혜택 접근	(a) 필수 공공 지원 혜택 및 서비스 자격을 평가하고 이러한 혜택 및 서비스를 부여, 감소, 취소 또는 회수하는 데 사용되도록 의도된 AI 시스템 (b) 신용도 평가, 신용 점수 확립에 사용되도록 의도된 AI 시스템 (c) 생명 및 건강보험 대상자의 위험 평가 및 가격 책정에 사용되도록 의도된 AI 시스템 (d) 사람들의 긴급 전화를 평가, 분류하거나 경찰, 소방관, 의료 지원 및 응급 의료 환자 분류 시스템을 포함한 응급 대응 서비스를 파견하거나 파견 우선순위를 설정하는데 사용되는 AI 시스템
법 집행	(a) 법 집행기관 또는 관련 지원기관에서 범죄 희생자가 될 위험을 평가하기 위해 사용하거나 대신 사용하도록 의도된 AI 시스템 (b) 법 집행기관 또는 관련 지원기관에서 폴리그래프(거짓말탐지기) 또는 유사한 도구로 사용하도록 의도된 AI 시스템 (c) 법 집행기관 또는 관련 지원기관에서 범죄의 수사 또는 기소 과정에서 증거의 신뢰성을 평가하기 위해 사용하도록 의도된 AI 시스템

일정	내용
	(d) 법 집행기관 또는 관련 지원기관에서 범죄 또는 재범 위험을 평가하기 위해 사용하도록 의도된 AI 시스템 (e) 법 집행기관 또는 관련 지원기관에서 범죄 또는 재범 위험을 평가하기 위해 사용하도록 의도된 AI 시스템
이주, 망명 및 국경 통제 관리	(a) 기관에서 폴리그래프 또는 이와 유사한 도구로 사용하도록 의도된 AI 시스템 (b) 기관에서 회원국의 영토에 입국하려는 또는 이미 입국한 사람이 초래하는 보안 위험, 불법 이주 위험 또는 건강 위험을 포함한 위험을 평가하도록 의도된 AI 시스템 (c) 기관이 망명, 비자 또는 거주 허가 신청을 검토하고 신분을 신청하는 사람의 자격과 관련된 불만을 처리하도록 지원하도록 의도된 AI 시스템 (d) 기관에서 이주, 망명 또는 국경 통제 관리의 맥락에서 사람을 탐지, 인식 또는 식별하는 목적으로 사용하도록 의도된 AI 시스템(단, 여행 서류 검증은 제외)
사법 행정 및 민주적 절차	(a) 사법기관에서 또는 사법기관을 대신하여 사실과 법률을 조사하고 해석하고 구체적 사실에 법률을 적용하는 데 사법기관을 지원하거나 대체 분쟁 해결에서 유사한 방식으로 사용하도록 의도된 AI 시스템 (b) 선거 또는 국민투표의 결과 또는 선거 또는 국민투표에서 투표를 행사하는 사람의 투표 행동에 영향을 미치도록 의도된 AI 시스템

자료: EU AI ACT

사람과 상호작용하는 AI 시스템 중에서 딥페이크 기술과 같이 비인격화, 기만, 조작 등의 문제를 일으킬 수 있는 경우 제한된 위험성을 갖는 시스템으로 분류되며 이러한 시스템에는 투명성 의무과 부과된다.[48]

사회경제적 파급효과가 큰 범용 AI 모델을 일반적인 AI 시스템과 구분하고 범용 AI 모델 공급자는 AI 개발 및 테스트에 대한 자세한 기록을 보관하고, 지적재산을 보호하면서도 AI를 사용하려는 다른 회사에 관련 정보를 제공하고, EU 위원회 및 국가 당국과 협력하도록 규정하고 있다.

48) 투명성 의무란 시스템 제공자가 AI 사용 상황과 맥락을 고려하여, 상대방이 AI 시스템과 상호작용하고 있다는 사실을 인식할 수 있도록 알려줘야 하고 합성 콘텐츠(오디오, 이미지, 동영상, 텍스트 등) 생성하는 AI 시스템의 공급자는 결과물을 기계판독이 가능한 형태로 표시하고, 인공적 생성이나 조작을 판별할 수 있도록 해야 함

[표 3-10] EU AI 법의 AI 구분

구분	내용
AI 시스템	• 다양한 수준의 자율성을 가지고 작동하도록 설계되고 배포 후 적응력을 발휘할 수 있으며 명시적 또는 암묵적 목적을 위해 수신한 입력으로부터 물리적 또는 가상환경에 영향을 미칠 수 있는 예측, 콘텐츠, 추천 또는 결정과 같은 산출물을 생성하는 방법을 추론하는 기계 기반 시스템
범용 AI 모델	• 자기 감독을 사용하여 대규모 데이터로 학습된 경우를 포함하여 일반성을 가지며 다양한 고유의 작업을 유능하게 수행할 수 있고 다양한 시스템이나 Application에 통합될 수 있는 AI 모델 (단, 시장에 출시되기 전에 연구, 개발 및 프로토타이핑 활동에 사용되는 AI 모델은 제외)
범용 AI 시스템	• 범용 AI 모델을 기반으로 직접 사용하는 것은 물론 다른 AI 시스템과의 통합을 통하여 다양한 목적을 달성할 수 있는 기능을 갖춘 AI 시스템

자료: EU AI ACT

미국의 AI 행정명령은 AI 기술의 안전 및 보안(Safety and Security) 보장을 강조하고 있다. 행정명령에서는 AI가 엄청난 잠재력과 위험을 동시에 내재하고 있어, AI를 선의의 목적으로 활용하고 그에 따른 다양한 이점을 실현하기 위해서는 AI 활용에 따른 위험을 완화할 필요가 있으며, 이를 위해 정부, 민간, 학계와 시민사회 전체의 노력이 필요하다고 선언했다. 이중 용도(dual-use)[49] AI 모델 개발 시 안전성 테스트(AIred-teaming test) 및 그 결과를 정부에 보고[50] 하도록 규정하고 있다.

백악관은 안전성 검증에 대한 정부 보고에 관해 국방 물자 생산법(Defense Production Act)에 근거한 것으로 첨단 AI 기술이 출시되기 전에 개발·훈련 단계부터 의무적으로 거쳐야 하는 과정이라고 언급했다. 또한, AI 행정명령은 안전하고 신뢰할 수 있는 AI 업계 관련 표준과 사례 개발을 강조하고 있다. 특히, 미국 기업의 AI 기술을 이용하는 외국인(기업)도 행정명령의 적용 대상이며, 외국인도 안전성 평가 및 그 결과를 보고하도록 규정하고 있다.[51]

중요 인프라와 관련된 잠재적 위험 평가·공개, 화학, 생물학, 방사선 등에 AI가 사용될

49) 평화와 군사 목적 모두 사용될 수 있는 기술
50) AI 시스템의 결함과 취약성을 식별하기 위한 구조화된 시험으로 통제된 환경에서 AI 개발자와 협업하여 수행되며 레드팀은 AI 시스템의 결함과 취약점을 식별
51) 미국 클라우드 서비스 제공자의 외국 고객 명단 신고를 의무화한 점은, 현재까지 발표된 AI 규제 중 가장 강력하며 이는 세계 AI 규제 표준을 마련하고자 하는 포석으로 평가

가능성 평가, 생성 AI 콘텐츠 인증, 출처 추적, 라벨 지정·감지를 위한 기존 및 잠재적 표준, 도구, 방법 식별을 위한 지침 개발도 포함된다.

영국도 AI 규제백서에서 AI 안전 및 신뢰성을 강조하고 있으며 영국표준협회는 책임감 있는 AI 관리 지원을 위한 글로벌 지침을 발표했다. AI 규제백서에서는 AI 수명주기 전반에 걸친 안정적인 작동을 위한 지속적인 안전 평가, 식별 및 관리체계 도입의 필요성을 강조한다. 또한, 2024년 영국표준협회(British Standards Institution)가 사회 전반에서 AI를 안전하고 보안이 유지되며 책임감 있게 사용할 수 있도록 설계된 AI 관리 지침을 발표했다.

주요국들은 AI 안전연구소 설립 및 운영을 통해 안전성을 확보하고 신뢰할 수 있는 환경을 마련하기 위해 노력하고 있다. 미국은 2023년 국립 표준기술연구소(NIST) 내에 AI 안전연구소를 신설하고 운영 중이다. 미국 AI 안전연구소는 AI 안전 기초연구, AI 안전 테스트 구조 개발, AI 안전 국제 협력을 포함하여 안전한 AI 혁신을 위한 다양한 기능을 수행한다. AI 안전연구소는 안전한 AI 개발 및 배포 관련 기술 표준 수립을 위해 2024년 2월 대규모 AI 안전연구소 컨소시엄을 발족하고, 공공-민간협력체계를 구축하고 있다. 컨소시엄은 엔비디아, 구글, MS, 애플 등 미국 내 AI 관련 기업 포함 200개 이상의 회원사와 대학, 연구기관 등으로 구성되었으며 5개의 작업반(Working Group)을 운영하고 있다.

[표 3-11] 미국 AI 안전연구소 컨소시엄 작업반(Working Group) 구성

구분	내용
생성 AI 위험관리	• AI 위험관리 Framework 보완 자료 개발 및 운영 • 연방 기관 대상 최소 위험관리 지침 개발
합성콘텐츠	• AI 생성 콘텐츠 인증 및 출처 추적을 위한 표준, 도구, 방법, 사례 연구(합성콘텐츠 라벨링 등) • 합성콘텐츠 감지, 악용 방지를 위한 연구 • 생성 콘텐츠 테스트 S/W 개발 및 감사, 유지를 위한 기존 표준, 도구, 방법론 개발 등
성능 평가	• 화학, 생물, 방사능, 핵무기, 사이버보안, 자율복제, 물리적 시스템 제어 등 잠재적 위험 대응을 위해 영역의 AI 성능 평가 • 안전하고 신뢰할 수 있는 AI 개발 지원을 위한 테스트 환경 구축 및 도구 개발
레드티밍 (Red-teaming)	• AI Red-teaming 훈련 지침 마련 : 다중 용도(dual-use) 기반 모델 개발자의 안전하고 신뢰할 수 있는 시스템 구축을 위한 절차 및 프로세스
안전 및 보안	• 다중 용도 기반 모델의 안전 및 보안 관리 관련 지침 조정 및 개발

자료: US AISI, The United States AI Safety Institute

영국은 AI 안전연구소를 설립하고 첨단 AI 시스템 위험성 평가, AI 안전 및 위험 연구 촉진, 글로벌 AI 개발 관행 및 안전 정책 강화를 목표로 AI 안전 연구의 기능을 강화하고 있다. 첨단 AI 시스템의 위험 평가를 통하여 정책입안자에게 AI 위험 정보를 제공하는 것을 목표로 하며, AI 안전성을 테스트할 수 있는 자체 역량 구축에 주력하고 있다.

[표 3-12] 영국 AI 안전연구소의 목표와 기능

구분	내용
첨단 AI 시스템 위험성 평가	• 악의적 AI 활용 수준 측정, AI 시스템 배포 전후 영향평가 • AI 시스템 안전 및 보안, AI의 이중용도 가능성 평가 • 고급 AI 시스템에 대한 인간의 통제 불능 가능성 평가 등
AI 안전 및 위험 연구 촉진	• AI 거버넌스를 위한 평가 도구 및 기술개발, AI 시스템 평가 체계 구축 • AI의 영향측정 평가 도구 개발 등
글로벌 AI 협력 및 안전 정책 강화	• AI 안전 분야 정보 교환 촉진, 정부, 국제파트너, 민간기업, 학계, 시민사회 일반 대중과 정보교류 촉진 등

자료: SW 정책연구소(2024) “해외 AI 안전연구소 추진현황과 시사점”

일본도 2024년 AI 안전연구소를 설립하여 신뢰할 수 있는 AI 환경조성을 위해 노력 중이다. 일본 정부는 AI 안전성 정상회의, G7 히로시마 AI 프로세스 합의 등을 계기로 AI 안전연구소를 설립하였다. 일본의 AI 안전연구소는 경제산업성 산하 정보처리 추진 기구 내에 설립되어 운영되며 AI 안전 평가 및 기준조사, 기술 조사·연구, 국제 협력 업무를 수행하고 있다.

3 국가별 AI 정책 추진 방식 및 강도가 상이

EU의 AI 법은 AI 전반을 포괄하며 강도 높은 처벌 규정을 포함하고 있다. EU AI 법은 역내뿐만 아니라 역외 사업자도 EU 시민을 대상으로 제품이나 서비스를 제공하면 적용 대상이며, AI 위험에 비례한 처벌 규정 방식이다.

[표 3-13] EU AI 법의 처벌 규정

구분	내용
수용 불가 AI 시스템 규정 위반	• 최대 3,500만 유로(약 518억 원)와 직전년 회계연도 기준 전 글로벌 연 매출액의 최대 7% 중 더 큰 금액에 대한 제재금이 부과
고위험 AI 시스템 규정 위반	• 최대 1,500만 유로(약 222억 원)와 직전년 회계연도 기준 전 글로벌 연 매출액의 최대 3% 중 더 큰 금액에 대한 제재금이 부과
제한된 위험성 AI 시스템 규정 위반	• 인증기관, 담당 기관에 부정확, 불완전, 오해의 소지가 있는 정보를 제공하면 최대 750만 유로(약 112억 원) 또는 직전년 회계연도 기준 글로벌 연 매출액의 최대 1% 중 더 큰 금액의 제재금이 부과
범용 AI 시스템 규정 위반	• 최대 1,500만 유로(약 222억 원)와 직전년 회계연도 기준 전 글로벌 연 매출액의 최대 3% 중 더 큰 금액에 대한 제재금이 부과

자료: EU AI ACT

EU AI 법에는 AI 관련 규제 적합성 평가 사항이 포함되어 시스템 출시 전후 준수해야 할 사항을 규정하고 있으며 인증 및 적합성 검사도 포함되어 있다. 시장 출시 전, 공급자는 위험관리시스템, 데이터 거버넌스, 기술 명세서 제작/보관, 결과 추적 로그의 기록과 관리, 배포자에 대한 정보 공개 투명성, 사람에 의한 감독, AI 시스템의 신뢰성, 보안 등의 사항을 준수해야 한다. 시장 출시 후, 최소 6개월간의 로그 기록 보관, 품질관리시스템 운영, 관련 문서의 보관, AI 시스템이 AI 법을 위배한다고 판단되면 조치하고 관련 정보를 배포자, 관련 당국에 통보해야 한다. 이러한 절차를 완료하면 CE 마크를 획득하여야 하며[52], 자체 인증을 통해 적합성을 인증하며, 고위험 시스템의 일부의 경우에는 외부의 공인 기관의 적합성 검사가 필요하다. 표준이 제정되지 않은 생체인식 식별 또는 자연인 분류를 위한 AI 시스템과 타법률에 외부 공인 기관의 적합성 인증이 필요한 경우 외부의 적합성 검사가 요구된다. 고위험 AI 시스템에 대해 공급자뿐만 아니라 배포자에게 다양한 의무를 부과하고 있으며, 배포자는 감독자 지정, 지정할 의무, 고위험 AI 시스템의 동작을 관찰하고 필요한 경우 공급자나 관련 당국에 통보, 최소 6개월간 로그 기록 보관 의무, 근로자에게 해당 AI 시스템의 사용 고지 등을 준수해야 한다.

52) CE 마크란 제품이 안전, 건강, 환경 그리고 소비자 보호와 관련된 EU 규격의 조건들을 준수한다는 의미의 유럽 통합규격 인증마크

미국은 포괄적인 규제 대신 행정명령을 통해 AI의 잠재성은 극대화하고 중국 견제 등 국가 안보와 거짓 정보 대응 등 우려 사항에 대비하고 있다. AI 행정명령에 혁신 및 경쟁의 촉진을 위해 AI 교육이나 취업 목적 비자 절차 간소화 등이 포함된다. 근로자 지원(Supporting Workers), 평등과 시민권(Equity and Civil Rights)의 증진을 위한 방안이 제시되어 있다. AI가 노동시장에 미치는 영향에 대해 보고하고, 직원에 대한 피해를 최소화하기 위해 고용주가 채택할 수 있는 원칙과 모범사례 제시도 포함된다. AI와 그 밖의 기술 기반 고용 시스템과 관련된 고용 시 차별금지에 관한 지침 공표 사항이 있으며 사생활과 시민 자유 보호를 위한 표준 및 절차 평가가 강조되고 있다. 그리고, 연방정부의 AI 사용 진전을 위해 최고 AI 책임자 지정, 연방 직원 생성 AI 사용 지침 개발 등이 포함되어 있다. 미국의 글로벌 AI 주도권 강화를 위해 AI 표준 홍보, 글로벌 참여 계획 수립 및 의제 개발을 강조하고 있다. 미국 기업의 AI 기술을 이용하는 외국인(기업)도 행정명령의 적용 대상이며, 외국인도 안전성 평가 및 그 결과를 보고하도록 규정되어 있다.[53] 중요 인프라와 관련된 잠재적 위험 평가·공개, 화학, 생물학, 방사선 등에 AI가 사용될 가능성 평가, 생성 AI 콘텐츠 인증, 출처 추적, 라벨 지정·감지를 위한 기존 및 잠재적 표준, 도구, 방법 식별을 위한 지침 개발도 포함된다.

영국은 EU의 강력한 규제와는 달리 친혁신적 AI 정책 접근 방식을 추진 중이다. AI의 '혁신 친화적 접근 방식'이란 용어는 2021년 9월 영국 정부가 발표한 '국가 AI 전략(National AI Strategy)'에서부터 본격적으로 사용되었다. 2023년 11월 발표한 성명에서도 "가까운 시일 내에 국내 AI 사용 규제에 특화된 법제화 시도는 없을 것이라는 입장"을 재확인하면서 학계와 산업 및 시민사회와의 긴밀한 협의를 통해 '혁신 친화적 접근 방식'을 유지·발전할 계획이라고 언급하였다.[54] AI 기술 자체를 규제 대상으로 삼기보다는 AI 시스템이 활용되는 맥락을 살펴 AI가 가져올 혜택과 위험을 모두 고려하여 규제 여부를 판단하는 균형적인 접근 방식을 위해 4가지 특성을 고려한다.

53) 미국 클라우드 서비스 제공자의 외국 고객 명단 신고를 의무화한 점은, 현재까지 발표된 AI 규제 중 가장 강력하며 이는 세계 AI 규제 표준을 마련하고자 하는 포석으로 평가

54) Draia Mosolova. (2023.11.17.), "UK will refrain from regulating AI 'in the short term'," Financial Times

[표 3-14] 영국의 친혁신적 접근 방식의 4가지 특성

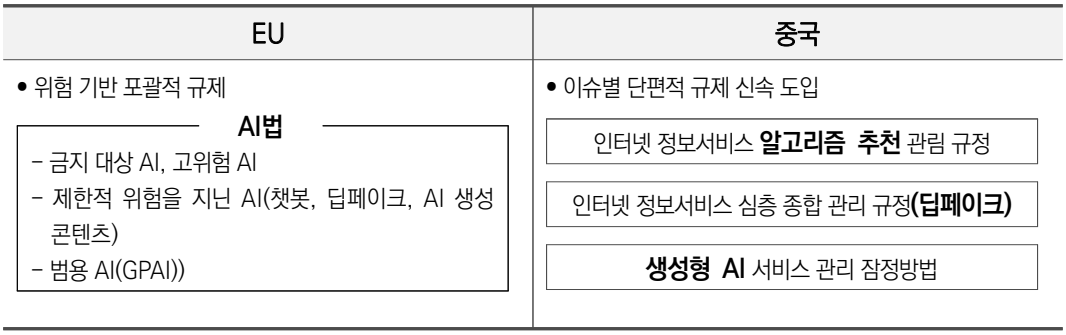
구분	내용
맥락 기반 (context-specific)	• AI는 범용 기술로, AI의 활용에 따른 위험은 주로 각 활용의 맥락에 따라 발생 • 특정 맥락 내에서의 AI 활용, 개인·조직·기업에 미치는 영향에 기반한 규제
친혁신, 위험 기반 (pro-innovation, risk-based)	• 실제적이고, 식별이 가능하며, 수용할 수 없는 수준의 AI 활용에 규제의 초점을 맞춤 • 위험 수준이 낮거나 가상의 위험을 제기하는 AI 활용에 대해서는 통제를 가하지 않음
일관성 (coherent)	• AI의 고유한 특성에 맞추어 모든 영역에 적용될 수 있는 일련의 공동 원칙 수립 • 규제기관이 담당 영역에서 이러한 공동 원칙에 대해 해석, 우선순위 설정, 적용
비례성, 적응력 (proportionate, adaptible)	• 적응력을 확보하기 위해 공동 원칙을 非 법적인(non-statutory) 방식으로 적용 • 규제기관은 지침 또는 자발적 조치와 같은 가벼운 방식을 고려 • 규제가 필요한 부분이 발생하면 정부가 개입하되, 필요와 비례하는 정도의 개입

자료: NIA(2023) "영국 AI 규제백서의 주요 내용 및 시사점"

영국은 새로운 단일 규제기관을 신설하여 전권을 부여하는 방식이 아닌, 혁신을 억제할 수 있는 AI에 대한 법안 도입은 최소화하고, 기존 관련 기관이 부문별·상황별 지침을 채택하는 유연한 규제방식을 취한다는 방침이다.

중국은 EU처럼 포괄 규제하기보다 새로운 AI 이슈에 대해 개별 규칙을 신속하게 제정하고 있다. 생성형 AI를 체계적으로 규제하기 위해 생성형 AI 서비스 관리 잠정방법(生成式人工智能服务管理暂行办法)이 공포되어 실행(2023.7) 중이다. 중국 정부는 생성형 AI의 규제에 대해 포용과 신중, 기술의 유형과 중요성 등급에 따른 차등 감독 관리를 원칙으로 제시하고 있다. 딥페이크 규제를 통해 안전한 인터넷 환경을 조성하기 위해 인터넷 정보 서비스의 심층 합성 관리 규정(互联网信息服务深度合成管理规定)을 시행(2023.1)하였다. 또한, 중국은 중국 국가인터넷정보판공실은 AI로 제작된 콘텐츠를 표시·식별할 수 있도록 의무화하는 규정 초안을 발표하였다(2024.9). AI 제작 콘텐츠 표시 의무화 규정은 '인터넷 정보 서비스 심층 합성 관리 규정'을 기반으로 만들었고, AI 합성콘텐츠의 표시 방법을 더욱 구체화하였으며, AI로 만들어진 콘텐츠의 제작, 전시, 배포 기간에 AI로 제작한 사실을 분명히 밝혀야 한다고 규정하고 있다. 온라인 콘텐츠 제공자들은 문자, 영상, 오디오, 가상화면 등 AI로 만든 모든 콘텐츠에 대해 문자, 음성, 그래픽 같은 눈에 띄는 표시로 이를 알려야 함. 아울러 디지털 워터마크나 메타데이터 태그 같은 좀 더 정교한

표식을 사용하는 것도 권장된다. 어떠한 단체나 개인도 악의적으로 해당 필수 표시를 삭제, 변조, 위조, 은폐해서는 안 되며, AI로 만든 콘텐츠에 대한 부적절한 식별로 다른 이들의 권리와 이익을 침해해서도 안 됨을 강조하고 있다. EU는 금지 대상 AI 세분류, 기반 모델을 규제하는 등 포괄적인 규제방식을 취하고 있으나, 중국은 생성형 AI, 딥페이크 등 이슈 중심으로 대응 중이다.



[그림 3-9] EU와 중국의 AI 규제방식 비교

자료: KITA(2024) "인공지능 규제 주도권 확보를 위한 글로벌 경쟁 및 시사점"

4 AI에 주목하는 지자체

미국 행정부의 AI 행정명령 이행과 함께, 주(州)별 AI 법제 논의가 활발히 진행 중이다. 콜로라도주는 고위험 AI 시스템을 개념을 도입하고 해당 시스템의 개발자와 배포자의 의무를 규정(2024.5)하였다. 고위험 AI 시스템 배포자 또는 배포자와 계약한 자는 고위험 AI 시스템 영향평가를 완료하고 이후 매년 영향평가를 수행하며 고위험 AI 시스템에 의도적인, 상당한 수정을 가한 경우 수정한 날 이후 90일 이내에 영향평가를 수행해야 한다.

[표 3-15] 개발자와 배포자의 의무

구분	내용
개발자	• 알고리즘 차별 위험으로부터 소비자 보호 조치 • (배포자 그밖의 개발자에게) 위험 데이터 등에 관한 정보 제공 • (배포자 그밖의 개발자에게) 영향평가 수행에 필요한 정보 등 제공 • (웹사이트 등에) 시스템 유형 및 알고리즘 차별성 관리 방법 등 제공 • (배포자 그밖의 개발자 법무부 장관에게) 차별 발생에 관한 보고서 제공 • 소비자가 인공지능 시스템과 상호작용임을 알 수 있게 할 것
배포자	• AI 시스템은 배포 후 자율성 및 적응성 수준이 다양함 • 알고리즘 차별 위험으로부터 소비자 보호 조치 • 위험관리체계 및 프로그램 수립, 영향평가 수행, 배포된 시스템 매년 검토 • (소비자에게) 고지 및 정보 제공, 웹사이트에 정보 게시 • (법무부 장관에게) 배포 후 발견된 차별 보고 • 소비자가 인공지능 시스템과 상호작용임을 알 수 있게 할 것

자료: https://leg.colorado.gov/sites/default/files/2024a_205_signed.pdf

유타주는 AI 수정법(Artifcial Intelligence Amendments, SB0149)을 마련하여 시행하고 있다.[55] 생성형 AI를 이용하여 서비스를 제공하는 경우 이를 소비자에게 고지하도록 규정하고 있으며, 생성형 AI 공개 의무 위반의 경우 2,500달러의 과태료를 부과하고, 집행 목적의 소송 제기가 가능하다. 테네시주는 개인의 재산권(property right)을 성명, 이미지 및 초상을 보호하던 기존 법 규정에 음성을 명시적으로 포함하여 재산권 보호 대상을 확대하였다.[56] 캘리포니아주 의회는 2024년 8월 AI 규제법안 'SB 1047'을 통과

55) https://le.utah.gov/~2024/bills/static/SB0149.htm
56) Ensuring Likeness Voice and Image Security Act of 2024

시켰으나 개빈 뉴섬 캘리포니아 주지사가 거부권을 행사했다. 해당 법안은 AI 기술의 급격한 발전에 따른 안전 문제를 해결하기 위해 마련되었다. AI 개발사는 모델 훈련 과정서 발생할 수 있는 위험을 사전에 평가하고 필요시에 신속하게 모델을 중단하는 '킬 스위치' 기능을 명시해야 하며 훈련 후 모델 변조 방지를 위해 안전 조치에 관한 조항도 법안에 포함되어 있고, 5억 달러 이상의 피해 또는 사망 사고 발생 시 개발업체가 책임지도록 규정하고 있다. 코네티컷주도 고위험 AI 시스템을 규제하는 법을 입법 추진 중이다.

중국은 지방정부 차원에서도 AI 정책을 추진하며 지역 경쟁력 확보를 위해 노력 중이다. 베이징시는 AI 혁신 근거지 건설을 위한 베이징시 실시방안(2023~2025)을 발표(2023.5)했다.[57] 2025년까지 베이징 AI 핵심 산업 규모를 3천억 위안(약 5조 원)으로 키우고 10% 이상의 성장을 유지한다는 계획이다. 세부 시행 방안으로 범용 인공지능(AGI) 혁신 발전 촉진에 관한 베이징시 조치(2023~2025년)를 발표하고 컴퓨팅 자원, 데이터 공급, 거대 언어 모델 개발, 응용 확대, 규제 여건 마련 5개 영역에서 방안을 마련하였다. 선전시는 AI 고품질 발전 및 고수준 응용 가속화에 관한 행동 방안을 발표(2023.5)하였다. 1천억 위안(약 18조 원 규모)의 AI 펀드 설립과 함께 핵심 기술과 제품의 혁신 능력 강화, 산업 집적 수준 향상 등에 지원을 확대할 계획이다. 상하이시는 민간 기업이 AI 인프라 건설에 광범위하게 참여할 수 있도록 정책지원을 하겠다고 밝혔으며, 상하이 쉬후이구는 다수의 거대 언어 모델(Large Language Model) 연구개발팀을 적극 육성 중이다.

주요국 지자체에서는 AI 도입을 통해 행정서비스 제고 방안을 모색하고 있다. 뉴욕, 런던 등 주요 도시에서 AI를 활용한 대민서비스 제공, 도시 관리, 행정서비스 효율성 제고에 AI를 접목하는 시도가 이루어지는 중이다.

[표 3-16] 주요 지자체의 AI 도입 사례

구분	내용
뉴욕	• 뉴욕은 "NYC 311" 이라는 서비스에 AI 기술을 통합하여 시민들이 다양한 민원 사항을 해결할 수 있도록 지원 • 이 챗봇은 GPT-4o를 기반으로 하여 시민들의 질문에 신속하게 답변하고, 서비스 요청을 자동으로 처리하는 기능을 제공 • 예를 들어 주차 Ticket 문의, 쓰레기 수거 일정 확인, 공공시설 예약 등을 처리

57) https://csf.kiep.go.kr/newsView.es?article_id=50438&mid=a20100000000

구분	내용
런던	• 런던은 “Smart London” 프로젝트의 일환으로 GPT 기반의 AI를 활용하여 도시 관리와 시민 소통에 적용 • AI는 교통 분석, 환경 Data 수집, 시민 의견을 반영한 정책 제안 등 다양한 분야에서 사용 AI가 수집한 데이터와 인사이트는 도시 관리 등 정책 결정에 활용
가나가와현 소재 요코스카시청	• 요코스카시는 40만 시민을 상대로 한 홍보문 작성과 내부 문서 요약, 각종 공문 오탈자 검사 등에 챗GPT를 활용 • 정보보호를 위해 개인정보나 기밀정보가 담긴 문서는 챗GPT 사용을 제한하고 챗GPT에 입력된 모든 내용은 저장되지 않도록 설정해 운영 • 시범 적용 기간은 1개월로 챗GPT 사용 실용성이 입증된다면 AI 챗봇을 시정 업무에 공식적으로 채택할 계획
시드니	• 시드니는 “Smart City” 프로젝트의 하나로 AI를 사용하여 도시 데이터를 분석하고, 시민 서비스 개선에 활용 • GPT 기반의 AI는 시민들의 피드백을 분석하고, 공공정책 개발에 필요한 인사이트를 도출하는 데 사용된다. 또한 AI는 환경 문제, 교통 혼잡문제 등에 대한 예측 분석을 통해 문제해결을 지원

자료: 한국지역정보개발원(2024), “GPT-4o의 지방자치단체 도입 현황과 활용방안” ; 뉴스1(2023.4.20.) “日 요코스카시, 지자체 최초 챗GPT 시범 도입”

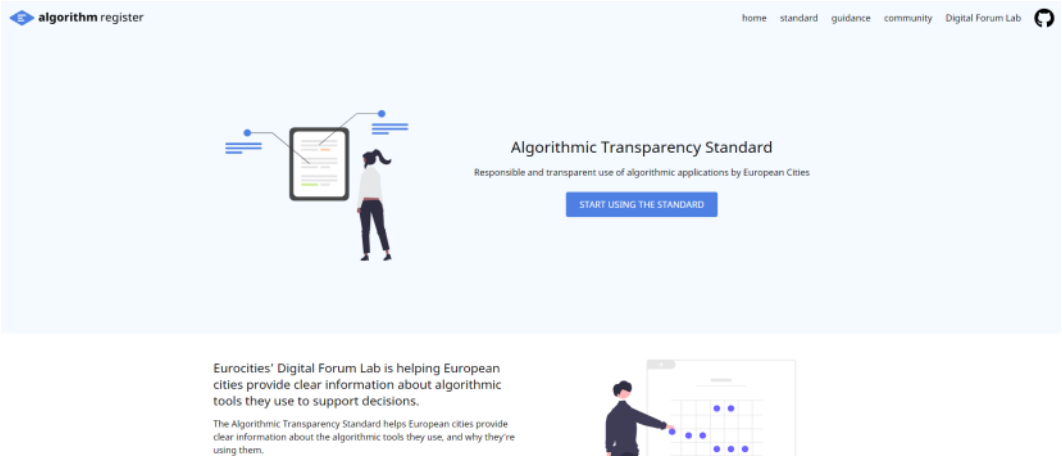
유럽 도시들의 알고리즘 등록제도 주목할 사항이다. 유럽 도시들의 알고리즘 등록제(City Algorithm Register)는 공공 서비스에 사용되는 AI 알고리즘의 투명성을 높이고 시민 감독을 강화하기 위해 도입된 혁신적인 제도이다. 암스테르담과 헬싱키를 비롯한 9개 유럽 도시들이 공통의 표준을 설정하여 이 제도를 채택하였다.

이 제도의 핵심은 정부 기관들이 공공 서비스에 사용하는 알고리즘을 등록하고 그 정보를 공개하는 것이다. 시민들은 이 정보를 온라인으로 쉽게 열람할 수 있다. 공개되는 정보에는 알고리즘의 목적과 작동 방식, 사용되는 데이터의 종류, 책임자 정보, 인권이나 윤리적 측면에서의 위험성 평가 결과, 그리고 인간의 감독 정도 등이 포함된다. 이러한 알고리즘 등록제의 도입으로 여러 긍정적인 효과가 기대된다. 먼저, 정부의 AI 사용에 대한 투명성이 증진되어 시민들이 이를 더 잘 이해하고 감시할 수 있게 된다. 이는 정부의 AI 사용에 대한 시민들의 신뢰를 높이는 데 기여할 것이다. 또한, 공무원들이 AI 사용에 더 신중해지면서 책임성이 강화될 것으로 예상된다. 나아가 AI 정책에 대한 시민들의 의견 개진

이 용이해져 시민 참여가 촉진될 것이다. 실제 사례로, 암스테르담에서는 쓰레기 무단 투기 감지 시스템의 알고리즘을, 헬싱키에서는 도서관 도서 추천 시스템의 알고리즘을 등록하였다. 이러한 노력은 'AI 거버넌스의 모범 사례'로 평가받고 있으며, 정부가 AI를 책임있게 사용하는 방식을 보여주는 좋은 예시로 여겨진다. 이에 따라 다른 국가나 도시들도 이와 유사한 시스템 도입을 고려하고 있다.

그러나 이 제도에는 몇 가지 한계와 과제도 존재한다. 먼저, 알고리즘의 기술적 복잡성으로 인해 일반 시민이 공개된 정보를 이해하기 어려울 수 있다. 또한, 너무 상세한 정보 공개는 보안 위험을 초래할 수 있어 이에 대한 주의가 필요하다. 마지막으로, AI 시스템이 지속적으로 변화함에 따라 등록된 정보도 계속해서 갱신해야 하는 과제가 있다.

이 알고리즘 등록제는 AI 거버넌스에서 도시의 모범사례 제시라는 측면을 잘 보여준다. 도시가 먼저 AI 사용의 투명성을 높임으로써, 민간 부문의 AI 사용에도 긍정적인 영향을 미칠 수 있을 것으로 기대된다. 이는 AI 기술의 발전과 함께 더욱 중요해질 AI 거버넌스의 한 모델로서 의미가 있다.



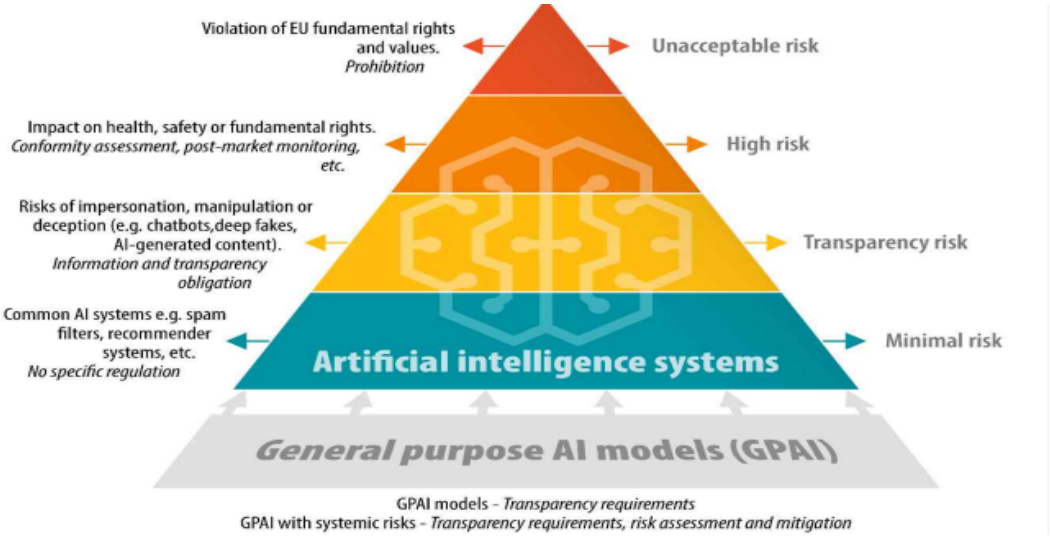
[그림 3-10] 유럽 도시의 알고리즘 등록제

자료: https://www.algorithmregister.org/

제4절 시사점

AI 혁명 시대에 진입하면서 변화하는 AI 정책의 특징에 주목할 필요가 있다. AI 정책의 무게 중심이 진흥에서 규제와의 조화로 이동하고 안전과 신뢰에 관한 정책이 강조되고 있다. 또한, 주요국들은 변화하는 AI 환경과 자국의 상황을 고려하여 차별화된 방식의 AI 정책을 수립하고 이행하고 있다.

먼저, 국내 AI 정책 수립 시 변화하는 AI 환경을 반영할 필요가 있다. 계류 중인 AI 법 도입 논의에 있어 변화하는 AI 환경과 정책 특성을 종합적으로 고려해야 한다. AI 안전과 신뢰성 관련 글로벌 정책 동향을 검토하고 국내 상황에 맞추어 반영하는 방안을 검토해야 한다. 21대 국회에서 폐기된 AI 법이 안전과 신뢰성 측면에서 문제가 제기되었던바 EU 등 다양한 AI 시스템의 위험분류 체계를 검토하고 한국의 현실에 맞게 반영해야 한다.



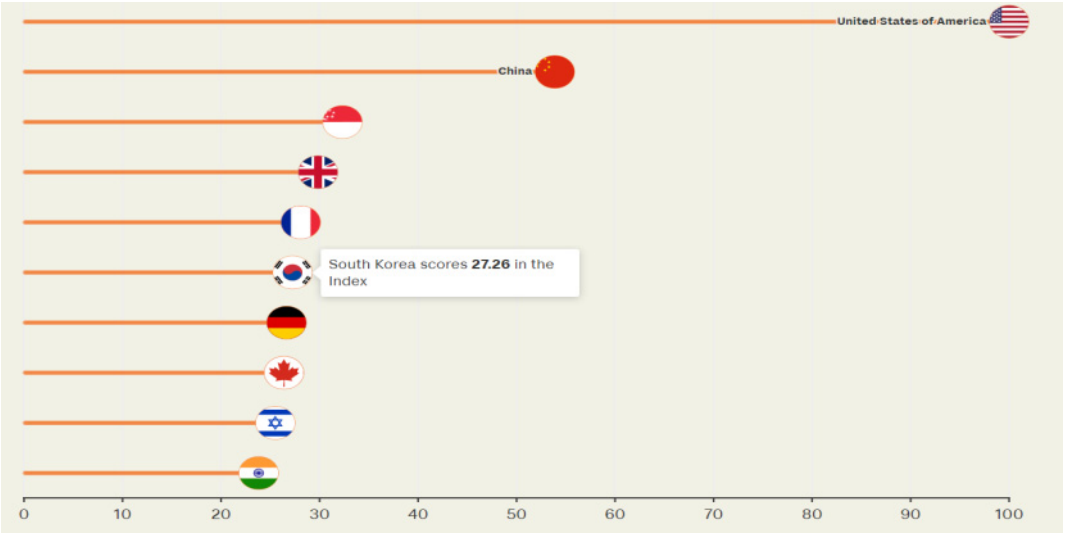
[그림 3-11] EU AI 법의 위험분류

자료: European Parliament Research Service(2024)

이러한 측면에서 EU의 범용 AI 모델 규제 방안을 국내에 도입하는 것은 미래지향적 관점에서 재해석될 필요가 있다. EU AI 법은 매개변수 기준으로 범용 AI 모델을 선정하고 규제하고 있다. 현재 매개변수가 큰 초거대 AI 모델이 규제 대상이나 매개변수가 작으면서 높은 성능을 보이는 AI가 등장하면 모델 크기로 위험성을 판단하기 어렵고 이는 규제에도 영향을 미칠 가능성이 존재하며 이는 인류를 위협할 수 있는 첨단 AI 모델은 크기가 작을 수도 있다는 것을 의미한다. 실제 높은 성능을 보이는 'Open AI o1-preview'가 AI 모델 성능의 잣대인 매개변수가 'GPT-4o'보다 크지 않으며 Open AI는 두 모델의 매개변수를 밝힌 적은 없지만, 전문가들은 작은 모델도 추론 능력을 강화하면 큰 모델보다 좋은 성능을 낼 수 있다고 지적하고 있다.[58] 이에 현재 매개변수 기준보다 실제 영향력과 위험을 기준으로 세분류하는 방안을 검토할 필요가 있다. 매개변수 기준은 이른바 스케일링의 법칙이 작용했을 경우 유의미했다. 스케일링 법칙이란 AI 학습에 사용하는 데이터와 컴퓨팅이 증가할수록, AI 모델 성능이 향상한다는 것이며 지난 10여 년간 AI 개발의 원칙으로 통하는 개념이다. 하지만 최근 스케일링 법칙을 뛰어넘는 대안들이 속속 등장하고 있으며 추론 중심의 모델은 물론, '공간지능(spatial intelligence)'이나 '맘바(Mamba)'와 같은 탈 프랜스포머(Transformer) 아키텍처 등이 그 예이다.

AI 정책 수립 시 국가별 초거대 AI 경쟁력도 고려해야 한다. 2020~2023년에 초거대 AI 모델을 가장 많이 개발한 국가는 미국(64건)이고, 중국(42건), 한국(11건), 프랑스(6건), 영국(5건) 순이고, 2023년 당해 기준으로 미국(41건), 중국(37건), 한국(8건), 프랑스(5건), 일본(3건) 이다. 초거대 AI 모델 보유 수가 상대적으로 적은 EU의 경우 AI 규제 강도 높고, 영국의 경우 규제 친화적인 접근으로 정책을 수립하고 있다. 글로벌 AI 지수 기준, 1위 미국 100점 기준으로 중국은 53.88점, 한국은 6위로 27.26점이며 선도국가와의 실질적인 격차는 매우 큰 상황이며, AI 규제 강도와 범위 설정 시, 한국의 경쟁력을 종합적으로 고려해야 한다.

58) TechCrunch(September 18, 2024),"This Week in AI: Why OpenAI's o1 changes the AI regulation game"



[그림 3-12] 글로벌 AI 지수 순위

자료: https://www.tortoisemedia.com/intelligence/global-ai/#rankings

또한, 국내 AI 안전연구소 설립 및 운영 관련 글로벌 사례를 참고하고 글로벌 정책 공조 체계를 강화할 필요가 있다. 미국, 영국, 일본 등 주요국은 AI 안전연구소를 설립하고 AI 안전성 평가체계와 테스트 방법론을 개발하고 글로벌 공조체계를 형성 중이다. 중앙정부와 지자체 AI 정책의 유기적 연계 방안을 모색되어야 한다. 현재 22대 국회에서 AI 법 도입 논의와 함께 경기도, 광주광역시 등 지자체에서 AI 조례를 제정하고 있는바 관련 입법의 유기적 연계를 통해 시너지를 낼 수 있도록 유도할 필요가 있다.

데이터 기반의 AI 생태계 관리 방안도 고려해야 할 요소이다. AI 법에 공통으로 포함된 중장기 계획, 진흥 정책 수립 시 빠르게 변하는 AI 생태계를 파악하기 위해 EIT(European Institute of Innovation & Technology)에서 제시한 데이터 기반의 AI 생태계 체계를 한국 상황에 맞게 재구성하여 구축하고 활용하는 방안을 검토해 볼 수 있다. 이는 AI 생태계를 데이터 기반으로 관리하여 현재 생태계 현황뿐만 아니라 동태적으로 생태계가 변화하는 양상을 파악하고 예측하여 미래 AI 정책을 수립하는데 활용될 수 있을 것이다.

AI 법과 정책 거버넌스에 대한 세부 논의도 필요하다. 현재 계류 중인 다양한 AI 법이 진흥과 규제에 관한 다양한 방안을 담고 있으나 전 국민이 AI를 알고 이를 다룰 수 있는 AI 리터러시(Literacy)에 대한 사항은 충분히 담겨있지 못하다. EU AI 법은 제4조에서

AI 리터러시 사항을 규정하고 있다. AI 거버넌스 차원에서도 다가올 미래에 AI가 중장기적으로 사회경제에 미칠 영향력을 고려하여 AI 위원회의 안정적 운영 방안을 검토해 볼 필요가 있다. 정부위원회 조직의 특성상 정부조직 개편으로 위원회의 폐지, 통폐합 논의가 계속되어 중장기적으로 중요한 위원회의 경우에 예외적으로 현재 5년 이내로 규정된 존속기한을 늘리는 방안도 고려해 볼 수 있다. 국회 차원에서도 과학기술정보방송통신위원회(과방위)를 과학기술과 방송 통신으로 분리하는 방안도 검토해 볼 수 있다. 논쟁 사항이 상대적으로 많은 방송법과 AI 법 이슈를 분리하여 입법을 추진하는 방안도 고려해볼 필요가 있다.

국회의 AI 도입도 논의될 필요가 있다. 이러한 측면에서 유럽의회의 생성형 AI 도입 사례에 주목할 만하다. 유럽의회는 생성형 AI를 활용하여 의회 아카이브의 접근성을 획기적으로 변화시키고 있으며 이에 국내 상황을 고려한 도입방안을 논의해 볼 시점이다.

[표 3-17] 유럽의회의 생성형 AI 도입 사례

- Amazon Bedrock의 Anthropic Claude를 활용하여 'Ask the EP Archives'(일명 Archibot)라는 AI 어시스턴트를 개발
- 이를 통해 210만 건 이상의 공식 문서를 연구자, 정책 입안자, 교육자 및 일반 대중이 쉽게 이용할 수 있는 기반을 제공
- 유럽의회의 아카이브 부서는 1952년부터 유럽의회가 생산하고 수신한 모든 문서의 보존, 관리, 처리를 담당하고 있으며 담당자는 수년간 두 가지 주요 과제에 직면해 있었는데, 이는 아카이브를 쉽게 검색할 수 있도록 만드는 것과 필요한 모든 사람이 접근할 수 있도록 보장하는 것
- 생성형 AI의 등장으로 아카이브 활용자와 연구자들이 수백만 건의 문서를 탐색하는 방식이 극적으로 변화하였고, 현재 유럽의회의 아카이브 웹사이트에서 이용 가능한 Archibot은 전 세계 사용자들에게 결의안, 입장문, 정책문서, 협상문서, 데이터 분석, 보고서 작성에 필요한 자료를 제공
- 이 프로젝트는 정부가 생성형 AI를 활용하여 시민들을 위한 투명성과 접근성을 향상시킬 수 있는 방법을 보여주는 사례
- Archibot의 주요 기능은 고급 검색 및 요약 기능(방대한 정보를 신속하게 탐색하고 종합), 보고서 작성 기능(정보를 추출하고 보고서를 작성), 다국어 지원(프랑스어를 넘어 모든 EU 회원국의 언어를 지원), 전 세계 접근성(유럽의회 웹사이트를 통해 전 세계 사용자들이 이용)
- 이 프로젝트는 문서 검색 및 분석 시간을 80% 단축하여 효율성을 크게 향상시켰으며 정책 입안자와 연구자들에게 역사적 맥락과 선례에 대한 신속한 접근을 제공하고 있으며 수백 명의 교육자와 학생들이 유럽의회 역사를 살펴보는 창구로 활용

자료: https://www.anthropic.com/customers/european-parliament

AI 미래정책 로드맵 구상도 필요하다. 동태적 관점에서 빠르게 변화하는 AI 생태계를 관찰하고 정책을 마련하기 위한 정책 로드맵 구상도 고려해 볼 수 있다. 미래학자 로이 아마라가 제시한 아마라의 법칙 (Amara's Law)에 따르면 사람들은 단기적으로는 신기술의 영향을 과대평가하나 장기적으로는 그 영향을 과소평가하는 경향이 있다. AI 정책관점에서, 이는 현재 논의되고 있는 AI 정책 현안뿐만 아니라 미래 관점에서의 논의도 병행하고, 모니터링하며 필요한 정책실험을 해야 한다는 의미이다. AI 전문가 FGI를 통해 제시된 다수의 미래정책 이슈는 로드맵 구상에 활용될 수 있다. 먼저, AI 기술의 급속한 발전에 따라 AI 기금 설립 논의가 확대될 수 있다. 노동, 교육 등 다양한 분야에서 사회적 후생을 높이기 위해 AI 활용하기 위해 재원이 필요하며 이에 기금이 하나의 방안으로 모색될 가능성이 있다. 또한, AI 기술을 활용한 취약계층 지원 등 보편적 서비스 제공을 위한 정책적 논의가 활발히 이루어질 전망이다. 전화, 인터넷은 시간이 지나며 보편적 서비스로 지정되었고 AI의 파급효과가 커질수록 AI도 보편적 서비스 관점에서 논의가 진행될 수 있다. AI 에이전트(Agent)와 플랫폼의 시장 지배력이 강화됨에 따라 새로운 독점 이슈가 주요한 정책 과제로 부상할 가능성도 있다. AI가 기존 서비스의 보완, 효율성 제고를 넘어 완전히 새로운 혁신 비즈니스 모델을 만들고 있어 이로 인한 파급효과를 선제적으로 검토해야 한다. 2024년 MWC(Mobile World Congress)에서 도이치텔레콤은 앱(App)이 없이 AI 에이전트(Agent)로 구동되는 스마트폰을 선보였으며 이는 기존의 앱(App)과 플랫폼 사업자에 대한 정책변화를 예고하고 있다. 2024년 10월 발표한 클로드의 컴퓨터 사용(Computer use)기능은 사람이 AI 에이전트에게 PC를 원하는 대로 제어할 수 있도록 하고 있어 AI 에이전트 경쟁이 더욱 본격화될 전망이다. 혁신적인 AI 기능과 알고리즘으로 소수 기업의 AI 기술 독점이 시장 경쟁을 저해하고 혁신을 방해할 수 있다는 우려가 제기되고 있으며 이에 향후 이에 관련된 공정 경쟁 정책과 규제 방안 마련이 논의될 것이다. 이외에도 오감이 전달되는 딥페이크, 노동 관련 근무제, 로봇세 등 세금, 기본소득 등 다양한 이슈가 AI 진화와 함께 미래정책 이슈로 논의될 수 있을 것이다.

LLM(Large Language Model)이 LWM(Large World Model)로 진화하고 있는바[59] 현재 2D 중심의 AI가 공간컴퓨팅과 연계된 3D 시뮬레이션으로 연동되면 기존에 없던 새로운 혁신과 위험이 발생할 가능성도 존재한다.

59) Forbes(Jan 23, 2024) “The Next Leap In AI: From Large Language Models To Large World Models?”

AI 위험에 대한 이해관계자의 견해가 다르며 이에 갈등관리 체계를 마련하고 절차를 투명하게 공개해 숙의의 과정을 통해 정책 방안을 논의해야 한다. 주지사의 거부권 행사로 무산된 캘리포니아 AI 법의 경우, AI의 대부 제프리 힌튼 토론토대 교수, 요수아 벤지오 몬트리올대 교수 등이 이 법안을 지지했으며 AI의 대모로 불리는 페이페이 리 스탠포드대학교 교수, 앤드류 응 스탠포드대 교수는 법안을 반대하였다. 미래지향적 AI 정책 수립에 필요한 다학제 연구를 강화하는 방안도 고려해야 한다.

제4장

AI가 만든 가짜: 딥페이크의 진화와 의미

제1절 AI가 만든 진짜 같은 가짜, 딥페이크

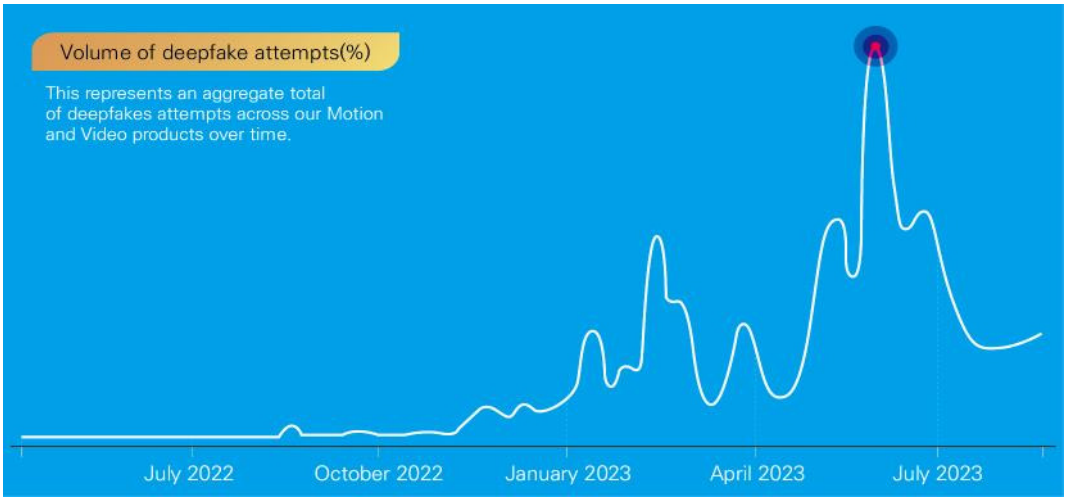
딥페이크(deepfakes)란 AI가 만들어 낸 진짜 같은 가짜 디지털 합성물이다. 딥페이크는 딥러닝(Deep Learning)과 가짜(Fake)의 합성어로 AI를 이용해 원본 이미지나 동영상 등 다양한 디지털 콘텐츠를 원본과는 다른 가공 콘텐츠로 생성하는 기술이다. 딥페이크 용어는 미국 커뮤니티 레딧(Reddit)의 이용자「deepfakes」가 2017년 12월 유명인의 얼굴을 성인물에 합성한 동영상을 처음 유포시킨 데서 유래되었다.[60] 딥페이크 제작에 활용되는 GAN(Generative Adversarial Networks) 관련 기술은 학술적 목적으로 연구되어 오다 딥페이크의 부상과 함께 주목받기 시작했다. 2018년부터 본격적으로 톰 크루즈, 오바마 등 유명 인사를 중심으로 딥페이크 영상이 제작되어 배포되었다. 2018년 4월에 온라인 매체인 Buzzfeed와 Jordan Peele 감독은 딥페이크 기술의 위험성을 경고하기 위해 오바마 미국 전 대통령이 트럼프 대통령을 비판하는 모습이 담긴 딥페이크 영상을 제작하였다.[61]

[그림 4-1] 영화 아이언맨에 등장하는 톰크루즈 딥페이크 영상

자료: https://www.youtube.com/watch?v=A8TmqvTVQFQ

60) 이승환(2020), "빅데이터로 본 딥페이크: 가짜와의 전쟁", SW정책연구소 이슈리포트 IS-090
61) BuzzFeedNew.com, "This PSA About Fake News From Barack Obama Is Not What It Appears", 2018.4.18.

2022년 이후 딥페이크 시도가 급증하는 추세이며 국내도 유사한 상황이다. 전 세계 딥페이크 시도는 2023년에 전년 대비 31배 증가하였다. 국내의 경우, 방송통신심의위원회는 딥페이크 기술을 이용해 유포한 성적 허위 영상물 총 4,691건에 대해 시정 요구를 의결했으며(2024.5월), 전년 대비 3,745건 증가하여 400% 폭증하였다.[62]



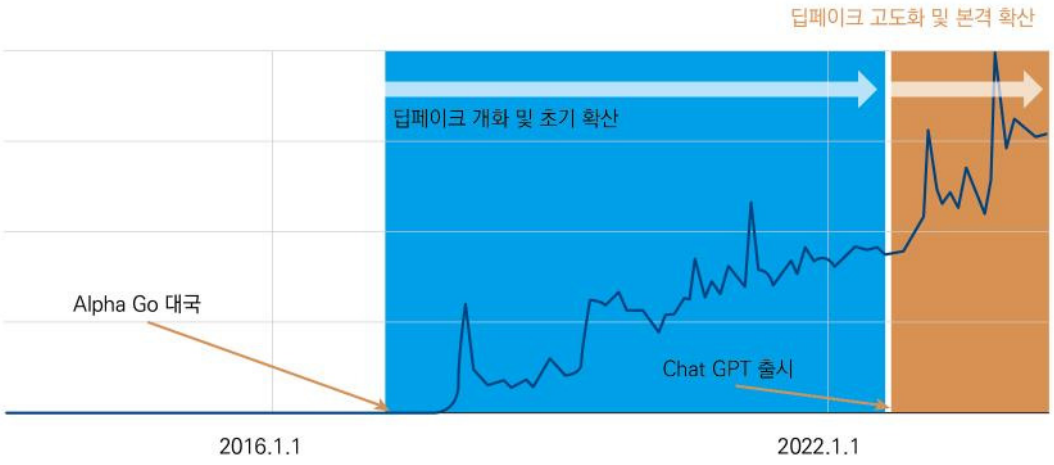
[그림 4-2] 딥페이크 시도 변화 추세

자료 : onfido(2024), "Identity Fraud Report 2024"

62) 방송통신심의위원회, "딥페이크 악용·유명 연예인 합성 '성적 허위 영상물' 400% 폭증", 2024.5.2

제2절 생성 AI로 진화하는 딥페이크

생성 AI 혁명의 시작을 알리는 챗GPT 출시 이후, 딥페이크 검색량 급속히 증가하며 진화 중이다. 알파고 대국 후 딥러닝이 주목받으며 딥페이크 검색 양이 증가했고, 이후 생성 AI 혁명 시대에 접어들며 딥페이크가 고도화되고 본격 확산되었다.

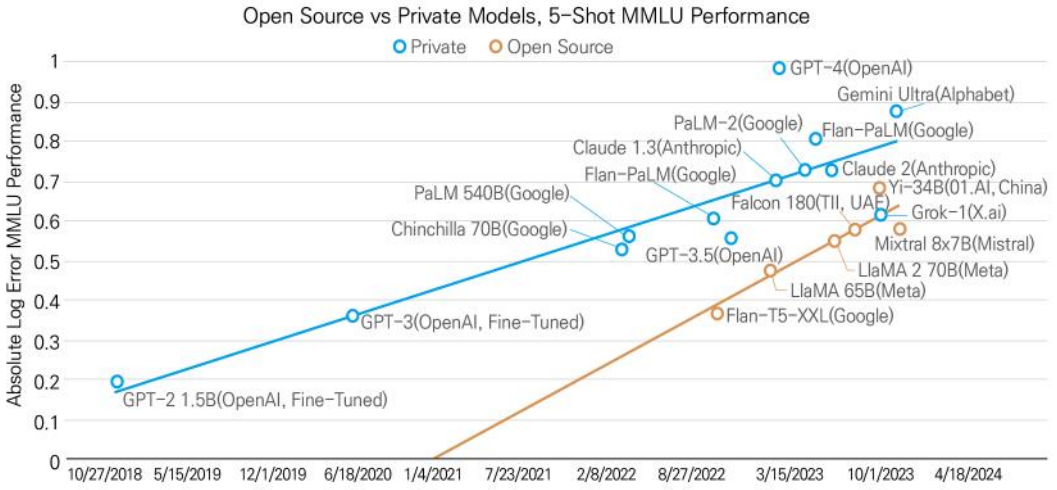


[그림 4-3] 딥페이크 검색량의 변화와 확산 패턴

자료: Google Trends 기반 저자 작성

초거대 AI 모델의 빠른 발전으로 더욱 진화된 딥페이크 제작이 가능한 여건이 조성되며 관련 시장이 성장하고 있다. 과거의 딥페이크는 GAN(Generative Adversarial Network) 모델 기반으로 활용하였으며 주로 제한된 데이터 셋을 학습하여 특정 도메인에 국한된 콘텐츠를 생성하였다. 반면, 현재의 딥페이크는 최근 급속히 진화 중인 초거대 AI를 활용하여 더욱 정교하고 진화된 생성이 가능하다. 방대한 데이터를 학습하고 이를 기반으로 추론하는 다양한 초거대 AI가 등장하고 경쟁하며 진화 중이다.

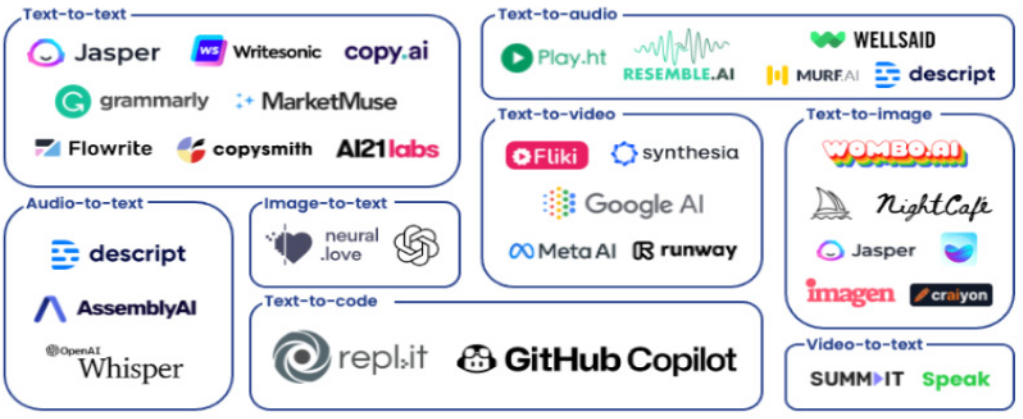
2022년 11월, 대표적인 초거대 AI 중 하나인 GPT-3.5 모델을 기반으로 출시한 챗GPT는 지속적인 업데이트를 통해 성능을 높이며 2023년 11월 GPT-4 터보, 2024년 5월 인간처럼 보고 듣고 말하는 GPT-4o가 출시되었다. 민간(Private) AI 모델과 함께 누구나 사용할 수 있는 오픈소스(Open Source) AI 모델들도 빠르게 확산 중이다.



[그림 4-4] 오픈소스 vs 민간 초거대 AI의 진화

자료: Ark Investment(2024), "Big Idea 2024"
주: MMLU (Massive Multitask Language Understanding)으로 AI 모델이 획득한 지식을 측정하는 벤치마크로 다양한 주제에 대해 다지선다 문제를 푸는 테스트

초거대 AI를 기반으로 텍스트, 음성, 이미지, 영상, 코딩, 가상 인간 등 다양한 디지털 요소를 제작하는 생성 AI 도구들이 배포되면서 딥페이크 제작이 용이해졌다. 목소리(Eleven Labs 등), 이미지(Midjourney, Dall E 등), 영상(Runway 등) 생성 도구들이 결합 되어 딥페이크를 쉽게 제작할 수 있는 환경도 조성되었다.



[그림 4-5] 다양한 영역에서 활용되는 생성 AI

자료: https://lyriko.ai/unleashing-the-power-of-chatgpt-in-pharma/

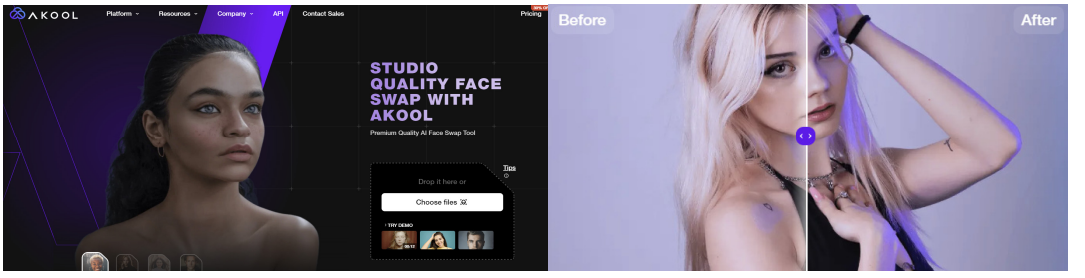
미국 시장조사업체 앱솔루트 리포트에 따르면, 2023년 전 세계 딥페이크 생성 앱 및 SW 시장 규모는 7,241만 달러(약 963억 원)로 2032년까지 연평균 약 36% 성장해 12억 달러(1조 5,000억 원)에 달할 것으로 전망된다.

생성 AI는 누구나 낮은 비용으로 쉽고 빠르게 딥페이크를 제작할 수 있도록 지원한다. 일상에서 사용하는 자연어 기반의 프롬프트(Prompt)로 다양한 디지털 요소를 생성하고 이를 변형하여 활용 가능하다. 생성 AI를 활용하여 얼굴을 바꾸거나, 완전히 새로운 얼굴을 만들거나, 목소리를 복제하는 등 다양한 디지털 합성 작업이 가능하다.

[그림 4-6] AI 프롬프트 기반의 딥페이크 생성

자료: onfido(2024), "Identity Fraud Report 2024"

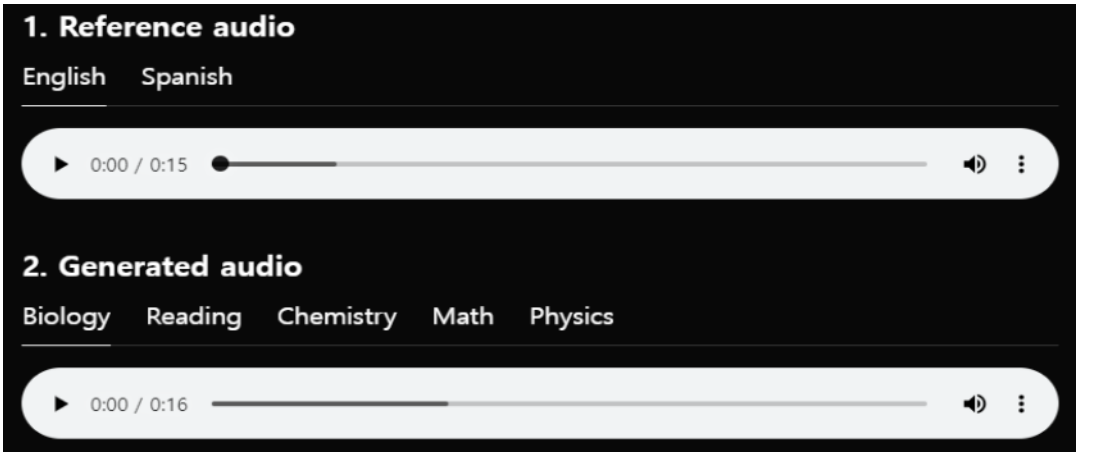
AKOOL AI는 간단한 파일 업로드를 통해 얼굴을 바꾸는 기능을 제공한다. AKOOL AI외 Reface, DeepFaceLab, Lensa AI 등 유사한 기능을 지원하는 수많은 애플리케이션(Application)이 배포되어 활용 중이다.



[그림 4-7] AKOOL AI

자료: https://akool.com/apps/faceswap

2024년 3월, 챗GPT 개발사 오픈AI는 목소리를 15초만 들려줘도 그대로 모방해 내는 AI 도구를 개발했다. 15초의 목소리만 있으면 이를 모방하여 실제 목소리 소유자가 하지 않은 말을 실제 한 것처럼 재생이 가능하다.



[그림 4-8] 오픈AI의 보이스엔진 (Vocie Engine)

자료: https://openai.com/index/navigating-the-challenges-and-opportunities-of-synthetic-voices/

2024년 4월, 마이크로소프트는 사진과 음성 샘플을 업로드하면 실시간으로 대화하는 얼굴을 생성할 수 있는 새로운 AI 모델 VASA-1을 발표했다 사진 1장, 음성 샘플로 누구나 딥페이크를 만들 수 있는 기술 기반이 마련된 것이다.

[그림 4-9] 마이크로소프트의 VASA-1

자료: https://www.microsoft.com/en-us/research/project/vasa-1/

2024년 4월, 링크드인 창업자인 리드 호프먼은 생성 AI를 활용하여 자신을 복제한 AI를 만들어 실제 자신과 대화하는 모습을 영상으로 공개했다. 챗GPT에 자신이 저술한 모든 책과 영상 인터뷰 등 다양한 자료를 학습시키고 이를 영상 제작 AI, 음성변조 AI와 결합하여 가짜 리드 호프먼을 제작했다.

[그림 4-10] AI 트윈(Twin)을 만든 리드호프먼

자료: 리드호프먼 유튜브, https://www.youtube.com/watch?v=rgD2gmwCS10

제3절 딥페이크의 빛과 그림자

딥페이크는 생산성 증대, 사회혁신 등 다양한 측면에서 활용 가능하다. 영화나 광고 등 엔터테인먼트 산업에서 딥페이크 기술을 활용하면 제작 비용과 시간을 절감할 수 있다. 고령이나 사망한 배우 모습을 딥페이크로 재현하거나, 위험 장면을 가상으로 연출 가능하다. 영화 인디아나존스 5에서는 주연배우 해리슨 포드의 젊은 시절 모습이 딥페이크로 재현되었다. JTBC 드라마 〈웰컴투 삼달리〉에는서는 고인이 된 국민 MC 송해의 젊은 시절 모습이 딥페이크로 재현되기도 했다.

[그림 4-11] 딥페이크로 구현된 해리슨 포드와 국민 MC 송해

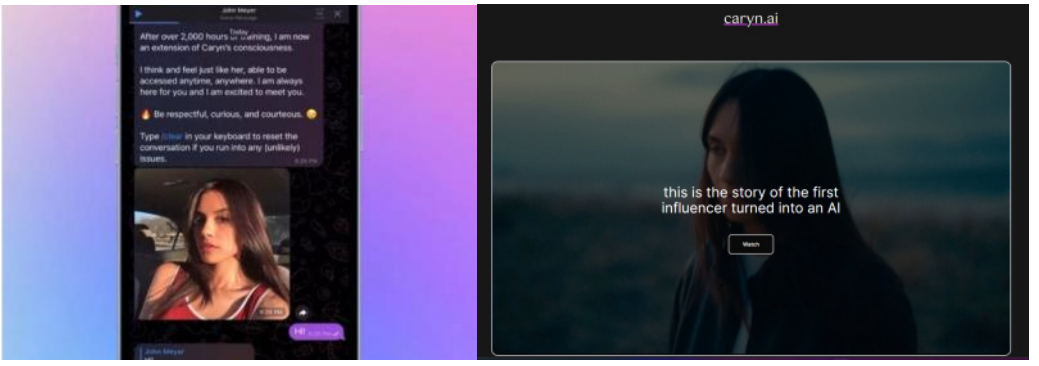
자료: Wired(2.23.6.7), “How Indiana Jones and the Dial of Destiny De-Aged Harrison Ford”; 중앙일보(2023.12.27.), “조용필 명곡·AI로 구현한 송해…‘웰컴투 삼달리’가 전하는 위로와 향수.”

개인의 취향과 특성에 맞춘 맞춤형 콘텐츠 제작에도 딥페이크가 활용될 수 있다. 개인의 얼굴을 애니메이션 캐릭터에 합성하거나, 개인의 목소리로 오디오북을 제작하는 등 새로운 경험을 제공한다. 오픈 AI가 개발한 목소리 생성 딥페이크 AI는 목소리에 장애가 있거나, 목소리가 다친 사람들이 유용하게 활용될 수 있다.

英 기업 Synthesia는 축구 스타 베컴의 말라리아 퇴치 홍보 캠페인 영상을 딥페이크를 활용하여 9개 언어로 제작되었다. 새로운 비디오를 촬영하지 않고 기존 자료를 편집하여

개인화된 영상, 제작이 가능하며 제작 비용도 기존의 1/10 수준으로 낮출 수 있다.

AI 트윈(Twin) 기반의 비즈니스 모델 혁신 등 생성 AI를 활용하는 슈퍼 개인이 등장할 수도 있다. 2023년 5월 미국 SNS 스냅챗에서 팔로워 180만 명을 보유한 여성 카린 마조리(Caryn Marjorie)는 본인 모습을 AI로 만들어 서비스를 출시했다. 카린 마조리의 목소리, 콘텐츠 등을 2천 시간 이상 학습시켜 AI를 만들고 사용자는 1분에 1달러 요금을 내고 이용 가능하다. 이용자들이 카린 AI에게 메시지를 보내면, 마조리가 실제 답하는 것처럼 보이는 메시지, 음성, 사진 등이 제공된다. 워싱턴포스트는 카린 AI가 출시된 첫 주에 10만 달러(약 1억 3천만 원) 이상 수익을 냈고, 서비스 이용을 위해 기다리는 수천 명의 대기자를 고려하면 매달 500만 달러(약 67억 원)의 수익을 창출할 것으로 전망했다.

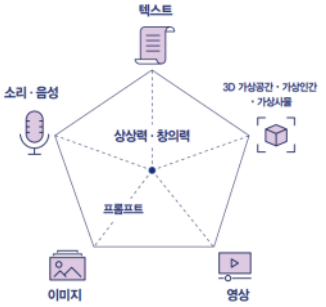


[그림 4-12] 카린 AI

자료: https://www.caryn.ai/ ; Forbes(2023.5.11.) “Snapchat Star Earned $71,610 Upon Launch Of Sexy ChatGPT AI, Here’s How She Did It“

이처럼, 생성 AI를 활용하여 자신의 상상력을 프롬프트로 구현하고 생산성을 높이며 비즈니스 모델을 혁신하는 슈퍼 개인이 등장하고 있다. 슈퍼 개인은 텍스트, 음성, 이미지, 영상, 가상공간 등 다양한 디지털 요소(Elements)들을 자연어로 자유자재로 구현하고 이를 조합할 수 있게 되었고 이를 지원하는 다양한 생성 AI 도구들은 지속 출시되며 진화하고 있다. 유니티의 마크 위튼 부사장은“생성 AI는 강력한 기술 집합체로 생산성을 100배 높인다.”라고 말했으며 레플릿 CEO 암자드 마사드는“생성 AI로 SW 개발자의 생산성이 10배에서 200배까지 향상되고, 1인 유니콘이 등장할 것이다”라고 언급했다. 폴

란드 영상 편집자가 혼자 개발한 게임이 세계 PC 게임 시장에서 매출 2위를 달성하였으며 생성 AI 도구의 확산으로 개발의 진입장벽이 낮아질 전망이다. 동시접속자 수가 약 3,400만 명에 달하는 세계 최대 PC 게임 플랫폼 스팀에서 중세 경영 시뮬레이션 게임인 '매너 로드(Manor Lords)'는 2024년 4월 23~30일 매출 순위 2위를 기록했다.



[그림 4-13] 슈퍼 개인의 생성 프레임(좌)과 1인 개발자가 만든 매너 로드 게임(우)

자료: 이승환(2023), "AI시대 절대 대체되지 않는 슈퍼 개인의 탄생"; PC Gamer(2024.4.27.), "The solo developer of Manor Lords, Steam's latest smash hit, named his studio after a Witcher 3 meme"

딥페이크는 산업적 측면에서 유용하기도 하지만 부작용도 존재하며 최근 연예, 정치 분야에서 피해사례가 속출하고 있다. 2024년 1월, 세계적인 팝스타 테일러 스위프트의 딥페이크 음란 이미지가 SNS 서비스 X에서 유포되었다. X가 뒤늦게 해당 이미지를 차단했지만, 이미 4,700만 회가 조회되고 다른 SNS 플랫폼으로 공유되었다. 딥페이크를 연구하는 워싱턴대의 컴퓨터 과학 교수 오런 에치오니는 "인터넷의 그늘에는 늘 다양한 종류의 포르노가 존재해 왔으며 앞으로 우리는 AI가 생성한 노골적인 이미지의 쓰나미를 맞게 될 것"이라고 경고했다.

정치 분야에서도 딥페이크가 기승을 부리고 있는 상황이다. 2023년 5월에 치러진 튀르키예 대선에서는 투표를 앞두고 야당 후보가 테러 집단의 지지를 받는다는 내용의 딥페이크 영상이 유포되었고, 이 조작된 영상이 집권당의 승리에 상당한 영향을 미친 것으로 밝혀짐

2023년 9월 슬로바키아 총선에서도 투표 이틀 전, 친미 성향의 야당 대표가 소외 계층인 로마족에게 금전적 지원을 해야 한다는 발언이 담긴 음성 파일이 공개되었고 파일이

가짜로 밝혀졌지만, 이미 집권당의 승리에 상당한 영향을 끼친 것으로 나타났다. 2024년은 전 세계 76개국에서 선거가 이루어지는 이른바 '슈퍼 선거의 해'로 관련한 이슈가 계속 발생할 우려가 있다. 정치적 양극화와 혐오 분위기를 조장하여 허위 정보를 유포하거나, 상대 후보를 비방하는 딥페이크 영상을 제작하는 네거티브 선거운동이 늘어날 가능성 존재한다.

딥페이크로 인한 피해가 기업과 일반인으로 확대되고 있다. 딥페이크를 활용한 기업 사기 사건으로 피해가 발생하고 있다. 2024년 2월, 홍콩의 한 다국적 금융기업의 직원이 딥페이크 기술을 이용한 정교한 사기 수법에 속아 340억 원에 달하는 거액을 도난당한 사건이 발생했다. 해당 직원은 영국 본사의 최고 재무 책임자(CFO)를 사칭한 이메일을 받았으며, 거액의 자금을 이체할 것을 요구받았고 처음에는 의심했지만, 회사 동료들과의 화상 회의에서도 같은 지시를 받자, 의심을 거두고 해당 금액을 송금했다. 사기꾼들은 딥페이크 기술을 이용해 CFO뿐만 아니라 화상 회의에 참석한 모든 동료의 모습과 목소리를 완벽하게 모방했고 피해 직원은 일주일 후에야 사기당한 사실을 확인했다. 홍콩 경찰은 유사한 시기 적발된 딥페이크 사기 사건이 최소 20건에 이르며 분실 신분증을 도용해 만든 딥페이크 이미지로 수십 건의 은행 대출을 받는 사건도 발생했다.

딥페이크를 활용한 투자사기 사례도 생겨 피해가 커지고 있다. 2023년 9월부터 유명인을 사칭한 투자 권유 광고가 유튜브, 인스타그램 등 SNS에서 확산 되었다. 피해자들은 범죄집단이 제공한 인터넷주소 링크를 통해 카카오톡, 네이버 밴드, 텔레그램 등으로 입장한 후 투자 정보를 받다가 범죄집단이 개설한 계좌에 입금하였고, 6개월간 발생한 피해 금액은 500억 원 규모로 추산된다. 조인성, 송혜교 등 유명 연예인의 얼굴과 음성을 조작한 가짜 영상을 활용해 투자를 권유받은 일반인이 피해를 받은 사례가 발생했다. 영상 속에는 유명 연예인들이 각각 자신의 이름을 직접 말하며 투자 행사 개최에 감사한다고 말하자 일반인들은 이를 믿고 투자했다. 딥페이크 영상을 제작한 조직은 주식 투자계 유명인사인 '베터리 아저씨' 박순혁 작가의 비서라며 투자자들에게 접근했고, 유명 연예인이 참여하고 있다는 영상을 보여주며 신뢰를 쌓아 투자를 유도하는 방식으로 투자금을 편취하였다.

워렌 버핏은 2024년 5월, 연례 주주총회에서 AI를 활용해 본인뿐 아니라 유명인의 이미지와 목소리를 복제하는 딥페이크의 위험성을 지적했다. 워렌 버핏은 "AI를 활용해 만

든 이미지와 영상이 진짜인지 아닌지 구별하기가 사실상 불가능하다.” 라고 언급하며 AI가 만든 자신의 이미지에 대해“우리 가족도 가짜라고 알아내기 어려웠을 것이며, 나조차도 가짜 나 자신에게 돈을 송금할 것”이라고 설명했다.

연예인, 정치인 등 유명인이 아닌 일반인도 딥페이크의 대상자가 되고 있다. 펜실베이니아 대학에 재학 중인 올가 로이에크는 자신의 딥페이크 영상이 중국판 인스타그램인 리틀 레드북(Little Red Book)에 유포되고 있다는 것을 알게 되었고, 딥페이크 영상에서 자신이 우크라이나에서 태어나고 자랐으며, 러시아는 세계 최고의 나라라고 말하며, 푸틴 러시아 대통령을 칭찬하는 영상을 보고 모욕감을 느꼈다고 언급했다. 미셸 제인스는 신혼여행을 하던 도중 자신 얼굴이 한 유튜브 광고에 무단 도용되고 있다는 사실을 알게 되었는데 그는 남성 호르몬 치료제 광고에 자신과 똑같이 생긴 얼굴이 등장한 것을 본 것이다.

국내에서도 피해가 속출하고 있다. 교육부가 17개 시도교육청을 통해 파악한 결과 2024년 1월부터 9월 6일까지의 피해신고는 434건이고, 이 중 수사 의뢰 건수는 350건, 삭제지원 연계는 184건, 피해자는 617명(학생 588명, 교사 27명, 직원 등 2명)이었다.

[표 4-1] 학교 허위합성물(딥페이크) 피해 현황

구분	피해 신고(건)				수사 의뢰(건)				삭제지원 연계(건)	피해자 현황(명)			
	계	초	중	고	계	초	중	고		계	학생	교원	직원 등
누적합계 (1.1.~9.6.)	434	12	179	243	350	11	151	188	184	617	588	27	2
1차조사 (8.27. 기준)	196	8	109	79	179	7	96	76	97	196	186	10	0
2차조사 (9.6. 기준)	238	4	70	164	171	4	55	112	87	421	402	17	2

자료: 교육부(2024) “교육부, 학교 딥페이크 성범죄 피해현황 2차 조사결과 발표”

‘서울대 N번방’ 사건은 우리 사회에 큰 충격을 주고 있다. 서울대 졸업생들이 여학생들의 얼굴을 딥페이크 성착취물에 합성해 음란물로 유포하였고 조사 결과, 확인된 피해자만 서울대 동문 12명 등 61명에 달하는 것으로 파악되었다.

특히 딥페이크 관련 범죄 가해자가 대부분 10대라는 사실에 더 주목해야 한다. 경찰청

에 따르면 올해 9월까지 검거된 딥페이크 성범죄 피의자 중 83.7%는 10대이며, 9월엔 텔레그램을 통해 연예인 20여명의 딥페이크 성착취물을 판매한 10대들이 경찰에 붙잡혔다. 이들은 지난해 11월부터 각각 텔레그램에 '합사방(합성사진방)' 등의 채널을 개설해 연예인이 성적 행위를 하는 내용의 딥페이크 성착취물을 불특정 다수에게 판매한 혐의를 받는다.

제4절 시사점

생성 AI 시대에 진화하는 딥페이크에 대한 대비가 필요하다. 딥페이크의 대상 범위도 연예, 정치 분야에서 일반 기업, 투자 등으로 다변화되고 있다. 딥페이크로 인한 기업 사기 등 리스크가 증가하여 이에 대한 대비도 필요하다. 가트너(Gartner)는 2026년엔 얼굴 생체 인식 솔루션을 겨냥한 AI 딥페이크 공격으로 인해 기업의 30%가 신원 확인 및 인증 솔루션을 단독으로 신뢰할 수 없게 될 것으로 전망했다. 가트너의 애널리스트인 아키프 칸은"지난 10년 동안 AI 분야는 수많은 변곡점을 거치면서 합성 이미지 생성이 가능해졌고 조직은 인증 대상자의 얼굴이 실제 사람인지 딥페이크인지 구분할 수 없어 신원 확인 및 인증 솔루션의 신뢰도에 의문을 제기하기 시작할 것"이라고 언급했다.

딥페이크의 위험을 줄이고, 올바른 활용을 위해 기술, 사회·제도적 측면에서의 노력이 필요하다. 기술적 측면에서 딥페이크 탐지 기술의 고도화가 요구되는 시점이다. 딥페이크 탐지 기술개발 강화 및 활용을 통해 대처할 필요가 있다. 인텔은 페이크 캐처(FakeCatcher)로 알려진 실시간 딥페이크 탐지 기술을 개발했으며 이 기술은 96%의 정확도로 가짜 동영상을 감지한다. 구글 딥마인드는 2023년 AI 합성 이미지용 워터마크 프로그램 '신스ID(Synth ID)'를 공개했으며 이는 AI 이미지 생성 프로그램에서 만든 이미지에 눈에 띄지 않도록 워터마크를 픽셀 단위로 넣어서 해당 이미지가 실제가 아님을 판별하는 방식이다. 구글은 2024년 신스ID 텍스트 워터마킹 도구를 오픈 소스로 공개했다. 신스ID 텍스트는 AI 모델이 생성한 텍스트에 보이지 않는 워터마크를 삽입하는 기술로 AI로 제작된 이미지, 영에 눈에 띄지 않는 워터마크를 삽입하는 기존 '신스 ID' 기술을 텍스트에 확대 적용한 것이다. 중국 상하이에서 개최된 2024 MWC 상하이에서 아너(Honor)가 스마트폰 업계 최초로 온디바이스(On-Device) AI 딥페이크 사기 방지 기술을 공개했다. 이 기술은 AI가 사용자의 영상통화에서 화면 속 디지털 콘텐츠를 스스로 인지해, 영상 속 인물의 딥페이크 여부를 감지하고 사용자에게 위험을 경고할 수 있다. 또한, 구글, 마이크로소프트, 인텔, 어도비 등 100여개 기업들은 '콘텐츠 출처 및 진위 확인

을 위한 연합(C2PA)'을 구성해 워터마크 기술 표준 개발에 주력하고 있다. 이에 악용된 딥페이크를 빠르게 판별하여 대처할 수 있는 기술 역량 확보가 필요하며 향후 딥페이크 관련 기술은 창과 방패의 대결 구도로 진행될 가능성이 높아 오용을 줄이기 위한 민관의 선도적 연구개발과 글로벌 협력이 필요하다.

사회적으로는 AI 리터러시 교육, 딥페이크 피해사례 전파 등을 통해 구성원들이 딥페이크로 인한 위험을 인지하고 대처할 수 있도록 지원해야 한다. AI 교육, 딥페이크 악용 사례 전파를 통해 개인이 스스로 정보를 판단하고 검증하는 능력을 키울 수 있도록 지원을 강화하는 방안을 검토해 볼 수 있다. 공개적인 토론과 인식 제고 캠페인을 통해 딥페이크 기술의 존재와 영향력에 대한 사회적 이해를 높이는 방안도 필요하다.

딥페이크로 인한 피해를 줄이기 위해 제도개선 이슈를 검토해야 한다. 현재 딥페이크 기술에 대한 규제는, 성착취물 제작이나 선거운동 활용 금지 중심으로 이루어지고 있어 딥페이크를 악용한 사기에 대한 별도의 법률이나 진위 확인, 출처 표시 등 관련 정책을 마련하는 방안도 검토할 필요가 있다. 2023년 12월 「공직선거법」개정으로 딥페이크 사용에 대한 규제가 도입되었다. 선거 전 90일~선거일까지 선거 운동을 위해 딥페이크 영상 등을 제작·편집· 유포·상영·게시할 수 없고 게시된 딥페이크 콘텐츠는 선거일 전 90일 전까지 삭제해야 한다. 선거일 전 90일 전에 선거 운동을 위하여 딥페이크 영상 등을 제작·편집·유포·상영·게시하는 경우에는 중앙선거관리위원회규칙으로 정하는 사항을 표시해야 한다.

딥페이크 산업 활성화를 고려한 정책조합 구상(Policy Mix) 및 모니터링을 강화하고 가짜 뉴스, 사기 등 국민 피해 최소화에 주력해야 한다. 기술조치, 자율규제, 가이드라인, 입법규제 등 다양한 규제 강도 수준을 고려해야 한다. 주요국에서는 딥페이크 관련 사기 방지, 출처 표시 등 정책 논의가 활발히 이루어지고 있는 중이다.

미국은 전통적으로 온라인에서 유통되는 콘텐츠에 대해서 직접적인 규제를 가하지 않았는데, 이는 크게 표현의 자유를 보장하는 미국 수정헌법 제1조에 따라 딥페이크 속 인물이 실존하는 인물이 아닐 경우 기소와 처벌이 불가능한 점[63] 과 통신품위법 제230조에 따라 인터넷(플랫폼) 서비스 제공자는 유통되는 콘텐츠의 내용에 대한 면책 특권이 보장

63) Ivancevich A.(2022), "Deepfake Reckoning: Adapting Modern FIrst Amendment Doctrine to Protect Against the Threat Posed to Democracy," 49(1) Hastings Const. L.Q

된다는 점이 지적된다.[64] 그런데 위에서 살펴본 테일러 스위프트의 딥페이크 음란물 유포 사태 이후 전향적인 변화가 나타나고 있다.[65] 미국은「딥페이크 책임법안」 및「딥페이크 신용사기 방지 법안」,「AI에 관한 행정명령」에서 관련 논의가 진행되고 있다. 딥페이크 책임법안(DEEPFAKES Accountability Act)은 딥페이크 제작물의 식별, 공개의무, 형사처벌, 민사 금전적 제재, 피해자 지원 및 딥페이크 탐지 등에 관한 내용을 포함하고 있다. 딥페이크 신용사기 방지법안(Preventing Deep Fake Scams Act)은 딥페이크를 활용한 시청각적 조작을 통해 이용자 계좌에 대한 접근 가능성이 높아진 상황을 고려하여, 금융 분야에서 딥페이크에 대응하는 조직의 구성에 관한 내용이 포함되어 있다. AI 행정명령에서는 합성 콘텐츠로 인한 위험 감소, 합성 콘텐츠 확인, 라벨링(labeling), 진위와 출처를 규명하도록 되어 있다. EU의「디지털 서비스법」에서도 딥페이크 관련 규제 사항이 포함되어 있다. 초대형 온라인 플랫폼 및 검색 엔진은 구조적 위험에 대응해야 하며 이러한 조치에는 딥페이크를 식별할 수 있는 표시를 통해 구별할 수 있도록 하고, 이용자가 쉽게 이러한 정보를 표시할 수 있는 기능을 제공하도록 하고 있다.

딥페이크의 판별과 함께 빠른 삭제로 피해를 최소화하는 방안 모색되어야 한다. 국내의 경우 2023년 10월 기준, 딥페이크 기술을 활용한 음란물이 3년 2개월간 9,000건을 넘겼지만, 콘텐츠 삭제는 5%에 미치지 못하는 것으로 나타났다.

규제뿐만 아니라 딥페이크의 산업적 가치에도 주목할 필요가 있다. 기업은 AI 트윈 가수, 과거 재현 등 딥페이크를 활용한 생산성 증대, 다양한 혁신 사업모델 발굴 방안을 모색해야 한다. 정부는 기존 벤처창업 활성화, 크리에이터 양성 과정, 1인 창조기업 육성 등의 지원 사업을 AI 시대를 주도할 슈퍼 개인 양성 관점에서 보완하는 방안도 검토해 볼 수 있다.

마지막으로 딥페이크의 미래 진화 방향을 검토하고 이에 대한 모니터링이 필요하다. 딥페이크 기술의 진화는 AI 기술의 발전과 밀접하게 연관되어 있다. 생성형 AI의 고도화로 인해 생성의 품질이 비약적으로 향상될 것이며, 대화도 보다 더 자연스러워질 것이다. 멀티모달 AI의 발전은 다양한 감각 정보를 통합적으로 처리하고 생성하는 것을 가능하게 하며 이러한 요소는 모두 딥페이크에 적용될 수 있다. 공간컴퓨팅 기술과 딥페이크의 융

64) O'Donnell, N.(2021), "Have We No Decency? Section 230 and the Liability of Social Media Companies for Deepfake Videos," 2021(2) U. Ill. L. Rev.
65) 강준모(2024), "딥페이크 관련 국내외 규제동향 분석" KISDI AI OUTLOOK 제18호.

합은 새로운 차원의 실감형 콘텐츠 생성을 가능하게 할 것이다. AI 기반의 공간인식 기술을 통해 실제 공간에 가상의 딥페이크 인물이나 객체를 자연스럽게 투영하는 것이 가능해질 것이며, AR/VR/XR 환경에서 실시간으로 딥페이크를 구현하는 것이 실현될 것이다. 더 나아가 상황 인식 AI를 통해 주변 환경과 자연스럽게 상호작용하는 지능형 딥페이크의 등장이 예상된다. 또한, AI 기반 센서 기술과 AI의 발전으로 딥페이크 기술은 시각적 구현을 넘어 다중감각을 아우르는 방향으로 발전할 것이다. 음성과 촉각까지 구현하는 다중감각 딥페이크가 등장할 것이며, 실제와 구분하기 어려운 수준의 체감형 가상 경험이 가능해질 것이다. 감정 AI의 발전으로 감정과 생체신호까지 모사하는 초실감형 딥페이크 기술이 개발될 것으로 전망된다. 이러한 기술의 융합은 화면을 넘어 초실감 몰입공간에서 구현되며 오감 기반으로 상호작용하는 미래로 진화하고 있어, 딥페이크를 현재의 문제로 보는 근시안을 넘어 미래의 진화 관점에서 관찰하고 이로 인해 생길 정책 이슈를 탐색해야 한다. 초실감 공간에서 AI 기반 신원 도용 및 사기 문제가 야기될 수도 있다. 실제 인물의 디지털 복제가 가능해지고 AI 기반의 음성 합성, 얼굴 합성, 동작 생성 기술이 고도화되어 실시간 신원 도용이 가능해질 수 있으며 행동 패턴까지 학습한 AI가 지인이나 유명인을 정교하게 사칭하는 범죄가 생겨날 수 있다.

제5장

AI 디지털교과서와 미래 교육

제1절 AI 디지털교과서 개요

1 AI 디지털교과서 개요

교육은 사회를 형성하는 근간으로 기술발전과 밀접한 관계를 맺으며 진화를 해왔다. 인쇄술 발명은 교육의 대중화를 이끌고 지식의 대량 복제와 유통혁명을 일으켰고 이는 교과서의 보편화, 체계적인 학교 교육의 기반 마련, 전반적인 문해력 상승과 지식 접근성 향상으로 이어졌다. 이후, 라디오, 영화, TV 등 시청각 기술의 발전이 교육방식에 영향을 미쳤다. 시청각 기술은 원격 교육의 초석을 마련하였으며, 더욱 풍부하고 생동감 있게 교육 내용을 전달할 수 있게 하였다. 또한, 학습자들이 교실 밖 세상과 연결될 수 있는 통로가 되기도 하였다. 1990년대 시작된 인터넷 혁명은 말 그대로 온라인 교육 혁명으로 이어졌다. 개인용 컴퓨터와 인터넷의 보급은 정보처리 및 학습 방식을 혁신했고 전 세계적인 지식 네트워크가 형성되었으며, 온라인 학습 플랫폼의 등장으로 시간과 공간의 제약 없이 학습할 수 있는 환경이 조성되었다. 스마트폰과 태블릿 등 모바일 기기의 보급은 언제 어디서나 학습할 수 있는 환경을 제공하였다. 다양한 교육용 앱이 개발되었고 협력 학습 및 지식 공유가 활성화되었다. 인터넷 혁명은 국내 공교육에도 영향을 미쳤으며 이는 디지털 교과서(Digital Textbook)의 도입으로 이어졌다. 정부는 2008년부터 디지털교과서의 원형을 개발하고 이후 시범 개발을 거쳐 2018년부터 디지털교과서를 적용하였다.[66]

66) 교육부(2023), "AI 디지털교과서 추진방안"

[그림 5-1] 디지털교과서 도입 경과

자료: 교육부(2023), "AI 디지털교과서 추진방안(안)"

2024년 기준 초등학교 3~6학년, 중학교 1~3학년에서 사회, 과학, 영어 과목에 디지털 교과서가 활용되고 있으며 고등학교에서도 영어 회화 등에 적용되고 있다. 디지털교과서와 연계된 실감 콘텐츠도 제공되고 있는데 초등학교 3~6학년 사회, 과학에 203종, 중학교 1~3학년 사회, 과학에 152종이 개발되어 활용 중이다.

[표 5-1] 디지털교과서 도입 현황

구분	과목	종수
초등학교 3~6학년	사회, 과학, 영어	172종
중학교 1~3학년	사회, 과학, 영어	69종
고등학교	영어, 영어 회화, 영어독해와 작문 등	29종

자료: https://dtbook.edunet.net

팬데믹 이후, 교육 분야의 디지털 전환(Digital Transformation)이 가속화되었고 이후 챗GPT를 시작으로 초거대 AI(Artificial Intelligence) 혁명이 본격화되었다. AI는 이러한 변화의 중심에서 기존의 교육방식을 혁신하는 새로운 동력으로 주목받고 있다. 전례 없는 펜데믹을 겪으면서 우리의 공교육은 의미 있는 변곡점을 맞이하고 있다. 다양한 방면에서 다가올 미래 사회를 위한 준비와 전환을 요청하는 담론들이 넘쳐나고 있다. 그러나 요구의 목소리만 높을 뿐 무엇을 어떻게 전환해야 할지 내용도 방향도 잘 보이지 않는다. 코로나19 사태로 수면 위로 떠오른 사회적·교육적 난제들도 제기만 될 뿐 쉽게 해결의 실마리를 찾지 못하는 모습이다. 이러한 현실에서 유독 확실해 보이는 전망이 있다

면 디지털과 AI일 것이다.[67]

WEF(World Economic Forum)는 기술 변화가 가속화됨에 따라 교육 시스템이 새로운 기회와 위험을 관리할 수 있도록 시급히 대응해야 하며, AI를 잘 관리하고 활용한다면 교육 4.0(Education 4.0)을 실현하는 데 도움을 줄 수 있을 것으로 전망했다. WEF가 제시한 교육 4.0은 미래 교육에 적합한 능력, 기술, 태도 및 가치를 제공하는 데 중점을 둔 접근 방식으로 전 세계 교육 전문가, 실무자, 정책 입안자 및 비즈니스 리더들의 연합에 의해 개발되었다. 교육 4.0은 글로벌 시민역량, 혁신과 창의 역량, 기술 역량, 상호협력 역량을 배양하기 위한 4가지 학습 방법을 제시하고 있다. 교육 4.0은 먼저, 개별화 및 자기 주도 학습을 강조한다. 표준화된 학습 시스템으로부터 각 학습자의 다양한 개별 요구에 기반한 학습 경험, 그리고 각 학습자가 자기 주도적으로 진도를 설정하고 진행할 수 있는 유연성을 지닌 학습 경험을 구축하는 것이다. 두 번째는 접근 가능하고 포괄적인 학습이다. 학교 건물 안에 한정된 교육 경험에서 벗어나 모든 사람이 학습환경에 접근할 수 있는 포용적 교육 경험 구축이다. 세 번째는 문제 기반 학습과 협력적 학습이다. 절차를 따르는 교육에서 벗어나 프로젝트 기반 혹은 문제 기반 학습 중심으로, 이를 통해 동료와의 협력을 강화하고 미래 직업 환경을 체험할 수 있도록 하는 교육 경험을 강조한다. 네 번째는 평생학습 및 학습자 주도 학습이다. 나이가 들어 학습이나 능력 습득이 줄어드는 교육에서 벗어나, 모든 사람이 기존 역량을 지속 개선하고 개인의 필요에 따라 새로운 역량을 습득하는 교육 경험이 필요하다는 것이다. WEF는 AI는 미래 지향적인 교육 시스템을 만들고 도움을 줄 수 있으나 해결하고 극복해야 할 과제와 위험이 있어 소기의 목표를 달성하기 위해서는 이를 해결해야 한다고 언급했다.[68]

67) 서울특별시 교육청 교육연구정보원(2022), “개별 맞춤형 AI 활용 교육의 가능성과 과제: AI튜터 마중물 학교 운영사례를 중심으로”

68) WEF(2024), “Shaping the Future of Learning: The Role of AI in Education 4.0”

[그림 5-2] WEF의 교육 4.0 프레임워크

자료: WEF(2024), "Shaping the Future of Learning: The Role of AI in Education 4.0"

정부는 기존 디지털교과서에 대한 한계 인식, AI를 활용한 맞춤형 교육, 사교육비 절감 등 AI로 야기될 교육 혁신에 주목하고 AI 디지털교과서를 추진하고 있다. 기존의 디지털교과서는 고정된 서책형 콘텐츠에 활력을 불어넣는 등 긍정적인 측면이 있었으나, 서책형과의 차별성 부족, 낮은 활용도 등이 한계점으로 지목되었다. 2022년 기준 디지털교과서 뷰어 접속 학생 비율은 초등 3~6학년 13.1%, 중학교 15.2%, 고등학교 8.7%에 그치고 있다. 이에 그동안 지적되어 오던 디지털교과서의 약점을 보완하고, 강점을 극대화하는 혁신적인 정책 마련이 필요한 상황에 직면하였으며 이에 대한 대안으로 AI 디지털교과서에 주목한 것이다. 정부는 AI 디지털교과서를 학생 개인의 능력과 수준에 맞는 다양한 맞춤형 학습 기회를 지원하고자 AI를 포함한 지능 정보기술을 활용하여 다양한 학습자료 및 학습지원 기능 등을 탑재한 소프트웨어로 정의하고[69] 보급 사업을 본격 추진하고 있다. 미래 교육은 AI 등과 같은 혁신 기술의 강점을 사용하면서 가장 인간적인 교육을 구현하

는 방식으로 변화할 필요가 있으며 AI 디지털교과서를 통해 이를 구현한다는 취지이다.

[그림 5-3] 디지털교과서의 한계와 시사점

자료: 교육부(2023), "AI 디지털교과서 추진 방안(안)"

2 AI 디지털교과서 추진현황

교육부는 2023년 6월에 「AI 디지털교과서 추진방안」을 발표했다. AI 디지털교과서는 정부의 3대 교육개혁 과제인 디지털 교육 혁신 중 하나로 추진되며 2025년까지 수학, 영어, 정보, 국어(특수교육) 과목에 AI 디지털교과서를 먼저 도입한다. 2024년 8월까지 AI 디지털교과서를 검정 심사하고 심사에 합격한 AI 디지털교과서는 6개월간 안정성, 신뢰성, 적합성을 검토한 후에 2025년 3월부터 학교 현장에 보급된다. AI 디지털교과서는 데이터 기반의 맞춤형 학습콘텐츠를 제공하고 특수교육 대상 학생과 장애 교원을 위한 화면 해설과 자막 기능, 다문화 학생을 위한 다국어 번역 기능도 지원한다. 2028년부터는 국어, 사회, 역사, 과학, 기술·가정 등으로 AI 디지털교과서 도입 과목이 확대된다. AI 디지털교과서 활용도가 높은 초등학교는 2027년까지 도입을 완료하고, 중고등학교는 2028년까지 도입을 마치도록 계획하였다. 또한, 학생의 발달단계를 고려하여 초등 1~2학년

69) 교육부(2023), "AI 디지털교과서 추진방안"

군과 심미적 감성, 사회·정서 능력과 인성을 함양하는 과목(도덕, 음악, 미술, 체육)은 적용 대상에서 제외하였다.

[그림 5-4] AI 디지털교과서 도입 계획

자료: 국가교육위원회 제14차 회의(2023.6.9.), "AI 디지털교과서 추진방안"

정부는 「AI 디지털교과서 추진방안」을 발표한 후, 2023년 8월에 후속 조치로 「AI 디지털교과서 개발 지침」을 공개하였다. 개발 지침은 민간이 자율성과 창의성을 바탕으로 다양하고 질 높은 AI 디지털교과서를 개발하여 AI 기반 맞춤형 학습지원을 구현할 수 있도록 학습데이터 수집 및 관리, 인프라 구축 등 핵심 기능을 중심으로 구성하였다.[70] AI 디지털교과서 개발 지침은 3개 원칙을 제시하고 있다. 먼저 인간 존엄성을 위한 교육이다. 교육 당국과 전문 기관, 개발에 참여하는 민간 등은 인공지능 기술이 개인과 사회에 미치는 영향을 이해하고, 아이들의 삶을 위한 교육을 기획한다. 교육 관계자는 신기술에 공존하는 위험과 기회를 바르게 인식하고, 인공지능과 구별되는 인간다움과 기본 권리를 강조한 교육을 설계해야 한다. 또한, 신기술을 안전하고 책임감 있게 사용하도록 교육당사자가 AI 디지털교과서를 주도적으로 활용·제어하게 해야 한다. 두 번째 원칙은 평등한 학습 기회 보장이다. 아이들이 언어, 장애, 지역, 계층 등 사회·문화·경제적 배경과 관계없이

70) 교육부 보도자료

신기술에 접근하고, 맞춤 교육 기회를 제공하도록 설계한다. 교육의 기회균등을 위하여 모든 아이가 AI 디지털교과서에 접근할 수 있도록 해야 한다. 또한, 학생의 개별적인 맞춤 교육을 제공하여 모든 학생이 학습에 성공할 수 있도록 지원해야 한다. 세 번째 원칙은 교사의 전문성 존중이다. 모든 아이는 신기술로 측정할 수 있는 범위 이상의 능력이 있음을 전제로, 교사가 이를 관찰하고 지지할 수 있도록 인공지능은 교사의 수업 준비, 평가 기록 등의 활동을 지원한다. 학생의 고유한 능력을 발견하기 위해 기술에만 과의존하지 않도록 유의해야 하며, AI 디지털교과서는 교수자 고유의 전문성이 효과적으로 발휘될 수 있도록 교수자를 지원한다. AI 디지털교과서 개발 방향에 대한 사항도 구체화 되어있다. 먼저, 2022년 개정 교육과정에 근거하여 학습분석 결과에 따라 보충학습(느린 학습자)과 심화학습(빠른 학습자)을 제공할 수 있도록 개발한다.[71] 또한, 모두를 위한 맞춤 교육 실현을 목표로 모든 사용자가 쉽고 편리하게 사용할 수 있도록 기능 및 UI/UX를 설계하며[72], 학생, 교사, 학부모, 정책 입안자 등 교육 주체가 학습에 대한 데이터 기반의 의사결정을 내릴 수 있도록 사용자, 학교, 국가 차원의 학습분석을 통해 교육 시스템을 개선한다.[73]

2023년 10월에는 「교과용도서에 관한 규정」을 개정하여 AI 디지털교과서의 법적 근거를 마련하였다. 교육부는 「교과용도서에 관한 규정」을 개정하여 AI 디지털교과서의 법적 근거를 마련하고, 교과용도서 편찬·검정·가격 결정 등을 심의하는 교과용도서심의회 구성·운영에 관한 사항을 정비하였다. 개정안에는 AI 디지털교과서가 개발되고 안정적으로 학교 현장에서 활용될 수 있도록 디지털교과서의 정의, 검정 절차별 필요 사항을 포함하였다. 즉, 디지털교과서를 지능정보화 기술을 활용한 학습지원 소프트웨어로 정의하고, 기술결함 조사 및 기술·서비스 적합성 여부 등에 대해 검정 심사를 하며, 검정 도서의 합격을 결정한 때에는 디지털교과서 사용대상 학교·학년도, 사용방법 및 사용환경 등을 관보에 공고하도록 하였다. 또한, 교과용도서심의회의 공정성과 객관성을 높이기 위해 위원 임기를 2년으로 정하고 1회만 연임할 수 있도록 하였다.[74]

2024년 3월에는 AI 디지털교과서에 개발사 자체 개발 콘텐츠 외에도 다양한 교육콘텐

71) 느린 학습자에게는 학생의 학습 수준에 맞는 기본 개념 중심 콘텐츠를 추천하고, 필요한 경우 학습결손을 해소할 수 있는 학습자료를 제공(학습분석 결과 등을 교사에게 제공하여 기초학력 보장 지원)

72) 특수교육 대상 학생·장애 교원 등의 접근성이 충분히 확보될 수 있도록 보편적 학습 설계(UDL,Universal Design for Learning)적용

73) 교육부, 한국교육학술정보원(2023), “AI 디지털교과서 개발 가이드라인”

74) 교육부 보도자료(2023.10.16.), “AI 디지털교과서 법적 지위를 얻다.”

츠를 담을 수 있도록 지원하였다.[75] 교육부는 한국교육방송(EBS)과 협력해 EBS의 개념 이해 영상 1,300여 편, 평가 문항 97,000개(수학 73,000개, 영어 24,000개) 등을 개발사에 제공하며 한국과학창의재단의 디지털 수학 도구인 알지오매스 (AlgeoMath) 연계 등을 통해 교육콘텐츠 지원을 강화했다. 알지오매스는 대수(Algebra)부터 기하(Geometry)까지의 모든 수학을 다루는 SW라는 뜻이며, 한국과학창의재단이 교육부, 17개 시·도 교육청과 함께 개발해서 무료로 보급하는 초·중·고 수학 실험탐구용 SW이다.

교육부는 2024년 4월, 「디지털 기반 교육 혁신 역량 강화 지원방안」을 공개하며, 2026년까지 선도 교사 34,000명을 양성하겠다고 발표하였다. 디지털 기반 수업 혁신을 이끌 교사 역량 강화에 2024년 3,818억 원이 사용된다. 2025년은 2022년 개정 교육과정, AI 디지털교과서 도입, 고교학점제, 성취평가제 등이 맞물려 공교육에 커다란 변화가 오는 시기이며[76] 교실 혁명을 이루어 내는 주체는 결국 교사이기 때문에 교사가 전문성을 갖고 주도적으로 수업을 혁신하도록 하는 것이 가장 중요하여 수업의 변화를 이끌어 갈 선도 교사 육성 방안을 발표하였다.

2026년까지 수업 혁신에 의지와 전문성을 갖춘 교실 혁명 선도 교사 총 34,000명을 양성하고, 한 학교에 2~3명의 선도 교사를 확보하도록 하였다. 2024년 1.15만 명, 2025년 1.15만 명, 2026년 1.1만 명을 양성할 계획이다.[77]

2024년 5월에는 AI 디지털교과서 시대에 디지털 기반 수업 혁신 지원을 위한 「초중등 디지털 인프라 개선계획(안)」을 발표하였다. 디지털 인프라는 학교 내에서 IT 기술을 활용하여 교수학습을 지원하는 물적 인프라(디지털 기기, 네트워크 등)와 인적 인프라(관련 전담인력 등)를 포괄하는 개념이며 예산은 963억 원이다. 새로 도입되는 AI 디지털교과서 작동 환경에 맞게 인프라를 질적으로 개선하고, 학교 현장의 애로사항인 인프라 관리 부담을 줄이는 방안을 포함하였다. AI 디지털교과서가 학교에서 활용 중인 디지털 기기에서 구동될 수 있도록 실제 수업환경과 유사한 디지털 기기 테스트 랩[78]을 구축하고, 디

75) 한국교육과정평가원(영어 검정심사), 한국과학창의재단(수학, 정보 검정심사), 한국교육학술정보원(기술문서 안내), 한국정보통신기술협회(보안인증 컨설팅) 등

76) 이 정책들은 모두 학생들이 창의성·인성·융합역량 등 미래 핵심역량을 키우고 능동적 학습자로 성장하는 데 중점을 두고 있다.

77) 교육부의 선도교사 연수방식도 정책 전달 중심의 일회성 연수가 아니라 수업 혁신의 가치와 방향을 함께 탐구하는 연수로 개편한다. 선도교사 연수 과정은 '교육과정·수업·평가 혁신, 디지털교과서 활용, 사회정서교육' 등 학생의 성장을 돕는 수업·평가 전문성 제고 과정과 함께 동료 교사 상담(코칭) 방법 등으로 구성할 계획이다.

78) 학교 현장과 유사한 교실 환경(2개 내외)에 기보급된 디지털 기기를 구비하여 개발사가 사전 점검할 수 있는 환경 제공

지털 기기의 작동 여부 등을 사전 점검하며 17개 시도교육청별 점검지원단을 구성하여, 전국 초·중·고에 보급된 디지털 기기 관리·활용 실태를 전수조사한다. 그리고 AI 디지털교과서 사용에 대비하여 올해 전국 초·중·고 6,000개교에 총 600억 원(교당 1천만 원)을 지원해 네트워크 속도, 접속 장애 등을 점검·개선하는 사항을 포함하였다. 또한, 교사와 학생이 기기 관리 부담에서 벗어나 교수·학습 활동에 전념할 수 있도록 교사의 AI 디지털교과서 수업을 보조하고 기기 설정, 충전 등 디지털 기기 관리를 전담하는 디지털 튜터(Digital Tutor)를 1,200명을 양성하고 배치한다. 학교의 IT 기기와 네트워크 품질을 사전에 점검하고 장애 발생하면 조치하는 원스톱 통합지원센터인 기술지원 기관(테크센터)을 전국 시도교육(지원)청에 170곳 설치하고 시범 운영하며 기술지원 기관에 소속된 기술전문가 학교의 인프라 장애 사전 관리부터 사후 대응까지 전 주기 관리를 전담하도록 하였다. AI 디지털교과서 도입 관련 경과를 정리하면 아래 표와 같다.

[표 5-2] AI 디지털교과서 추진 경과

구분	내용
2023년 6월	「AI 디지털교과서 추진방안」발표
2023년 8월	「AI 디지털교과서 개발 지침」공개
2023년 10월	「교과용 도서에 관한 규정」을 개정
2024년 4월	「디지털 기반 교육 혁신 역량 강화 지원방안」공개
2024년 5월	「초중등 디지털 인프라 개선계획(안)」발표

자료: 교육부 보도자료 종합

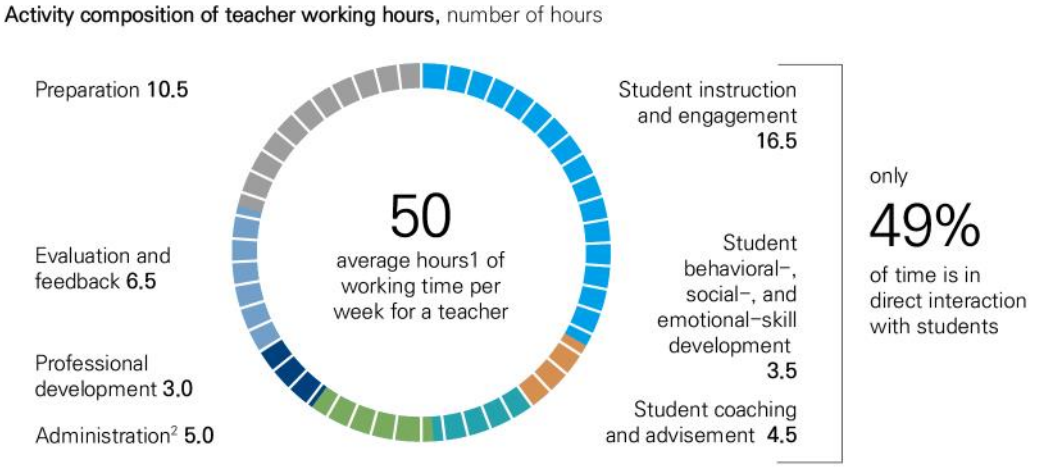
제2절 AI 디지털교과서에 대한 기대와 우려

1 AI 디지털교과서에 대한 기대

AI의 빠른 진화와 함께 AI가 우리가 당면하고 있는 여러 교육 문제를 해결하고 혁신을 선도하여, 궁극적으로 교육의 공정성(equity)을 확보하기 위해 활용될 수 있을 것이라는 기대가 커지고 있다.[79] AI가 기존의 표준화된 획일적 교육에 혁신을 불러와 교육 분야의 새로운 패러다임 변화를 주도한다는 의미이다.[80] AI 디지털교과서는 AI에 의한 학습 진단과 분석(Learning Analytics), 개인별 학습 수준과 속도를 반영한 맞춤형 학습(Adaptive Learning), 학생의 관점에서 설계된 학습 코스웨어(Human-Centered Design)로 다양한 교육 주체들에게 가치를 제공할 것으로 기대받고 있다. 교사는 학생별 학습경로와 지식수준을 이해하고 데이터 기반 참여형 수업(토론, 협력, 프로젝트 학습 등)을 설계하며, AI 보조교사의 활동 분석을 참고하여 평가, 학생별 학습성취에 맞는 맞춤형 학습을 제공하며 데이터를 기반으로 학생의 성장을 기록하고 지지할 수 있다.[81] AI 기반 맞춤형 교육은 개별 학습자의 서로 다른 특성에 따른 적절한 학습 경험 제공을 위해 필요한 관찰, 진단, 처치의 순환적 과정을 AI에 기반을 둔 측정, 분석, 조정 기술을 이용해 지원함으로써 교수·학습의 효율성과 효과성을 증진하고자 하는 교육이다.[82] 현재 교사는 다양한 업무로 학생 직접 상호작용하는 부족하며, Mckinsey&Company에 조사에 따르면 따르면 교사는 주당 평균 50시간을 일하며, 그 중 절반 이하의 시간만 학생들과 직접 상호작용하는 데 사용한다. 구체적으로, 교사는 학생 교육 및 참여에 16.5시간, 학생 지도 및 상담에 4.5시간, 학생의 행동, 사회적, 정서적 기술개발에 3.5시간을 사용하며, 이는 전체 근무 시간의 49%에 해당한다. 나머지 시간은 준비(10.5시간), 평가 및 피드백

79) Miao, F., & Holmes, W. (2021). Artificial Intelligence and Education. Guidance for Policy-makers.
80) Baker, T., Smith, L., & Anissa, N. (2019). Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges. Retrieved May, 12, 2020
81) 교육부, 한국교육학술정보원(2023), "AI 디지털교과서 개발 가이드라인"
82) 한국교육개발원(2023), "AI 기반 맞춤형 교육의 현황과 과제"

(6.5시간), 행정 업무(5시간), 전문성 개발(3시간)과 같은 활동에 할애된다.



[그림 5-5] 교사의 업무시간 활용 현황

자료: Mckinsey Global Teacher and Student Survey

학생은 학습 속도에 맞는 교육을 통해 성공을 경험하고 내재적 학습 동기와 자아 존중감이 향상되면서 가정과 학교에서 지지받는 나로 성장할 수 있다. 또한, 학부모는 데이터를 기반으로 자녀가 학습에서 겪는 어려움과 자녀의 강점·약점을 파악하고 진로 탐색·설계에 다양한 활동 정보 참고하여 자녀에 대한 깊이 있는 이해를 바탕으로 내 아이에게 맞는 정서적 지지·격려가 가능할 것으로 기대된다.

학습자 다양성의 롱테일 개념은 학생들이 각기 다른 강점과 필요를 가지고 있으며, 교육적 접근법이 이러한 다양성을 수용하도록 설계되어야 한다는 점을 강조한다. 아래 그림은 학습자의 다양성을 시각적으로 나타내며, 학습자의 강점과 필요의 변화를 강조하며 교육적 접근에서 표준화와 적응성의 수준에 따라 학습자를 세 구역으로 나누어 분류한다.

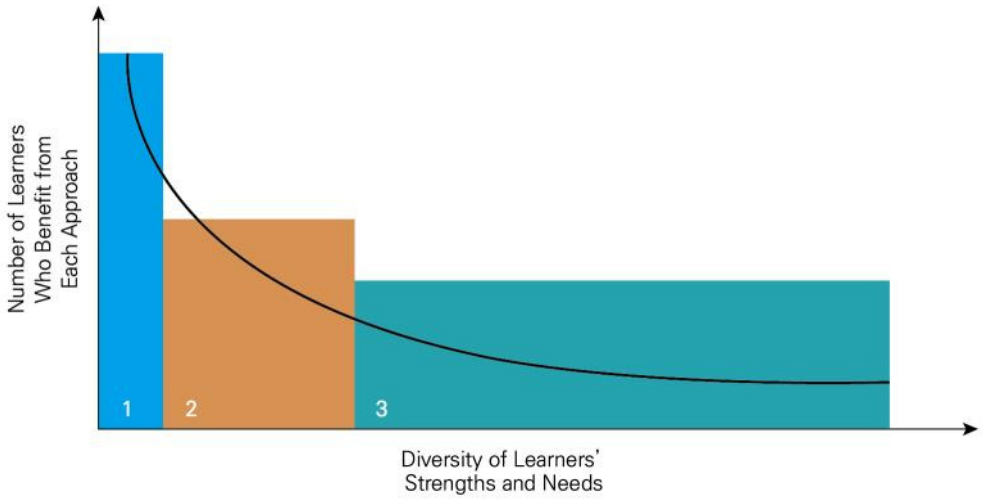
1구역에서는 가장 자주 나타나는 학습자 프로파일이 발견되며, 이를 “중간 수준을 가르치기(teaching to the middle)”라고 부르기도 한다. 이 구역의 교육 자원은 주로 표준화되어 있으며, 적응성은 최소한으로 제공됩니다. 예를 들어, 기존의 많은 교육 서비스는 학생 답변의 정확성에 따라 조정되며, 다국어 지원과 같은 옵션을 제공할 수 있다. 그러나

핵심적인 교육 접근은 모든 학습자에게 대체로 동일하게 유지됩니다. 결과적으로 이 구역은 "모든 사람에게 맞는 하나의 방법"으로 특징지어지며, 학생들의 가장 일반적인 요구만을 충족한다.

2구역은 표준화와 적응성 간의 더 균형 잡힌 접근을 취하고 있다. 이 구역에는 학습 기회를 여러 형식으로 제공하려는 유니버설 학습 설계(Universal Design for Learning, UDL)와 같은 요소가 포함된다. 이 구역의 학습자들은 문화 및 지역사회 배경이 다르고 이러한 다양성은 교육콘텐츠에 대한 참여 방식에서 나타나며 이러한 요소가 교육에 반영된다. 여기에는 비공식적 환경에서의 학습도 포함됩니다. 2구역은 일관된 교육 표준과 다양한 학습 요구를 충족하기 위해 유연성 측면에서 균형을 맞추고자 한다.

3구역은 매우 적응적인 학습을 나타낸다. 이 구역에서는 전통적인 표준화가 효과적이지 않은 경우가 많으며, 학습자를 효과적으로 참여시키고 지원하기 위해 다양하고 맞춤화된 접근을 취한다. AI의 잠재력의 잠재력은 주로 이 구역에서 언급되며, AI는 가장 일반적인 패턴을 넘어서는 패턴을 식별하고, 교사들이 제공할 수 있는 것 이상의 맞춤형 콘텐츠를 생성한다. 1구역이 가장 두드러진 구역으로 인식되는 경우가 많지만, 실제로 AI가 가장 큰 가치를 제공할 수 있는 곳은 2구역과 3구역이다. 이 구역에 속하는 학생들이 더 많으며, AI 기반 기술은 이들의 고유한 요구에 맞게 학습환경을 조정할 수 있는 잠재력을 가지고 있다. 그러나 교육 AI의 초점이 어디에 맞춰져야 하는가에 대한 질문이 제기된다. 만약 개발자들이 가장 일반적이고 전형적인 사례에만 "적합"한 모델을 만드는 데 집중한다면, AI가 덜 전형적이고 문화적으로 특수한 요구를 지원하는 데 한계가 있을 수 있다. 학습자 다양성의 롱테일 관점에서 AI를 최대한 활용하려면 개발자, 연구자, 교육자가 교육 맥락, 콘텐츠, 기술적 통찰을 통합하는 학제 간 전문성을 키우는 것이 중요하다. 이를 통해 개발된 교육 AI 서비스는 다양한 학습자의 맥락에 더 적응하고 민감하게 반응할 수 있다. 다양한 관점과 데이터를 제공하는 새로운 파트너십이 필수적이며 이러한 협력은 새로운 AI 기반 접근법이 학습을 개선할 수 있는지, 그리고 어떤 조건에서 가장 잘 작동하는지를 파악하는 핵심 질문을 해결하는 데 도움이 될 수 있다.[83][84]

83) Office of Educational Technology(2023), "Artificial Intelligence and the Future of Teaching and Learning"
84) Rose, D. (2000). Universal design for learning. Journal of Special Education Technology, 15(4), 47-51. https://doi.org/10.1177/016264340001500407



[그림 5-6] 학습자의 롱테일 커브

자료: Office of Educational Technology(2023), "Artificial Intelligence and the Future of Teaching and Learning"

AI 디지털교과서를 통해 기대할 수 있는 효과를 정리하면 아래 표와 같다.

[표 5-3] AI 디지털교과서 핵심 서비스

구분	내용
학생	• 학습 진단 및 분석, 학생별 최적의 학습경로 및 콘텐츠 추천 • 맞춤형 학습지원
교사	• 수업 설계와 맞춤 처방, 콘텐츠 재구성·추가 • 학생 학습 이력 등 데이터 기반 학습 관리
공통 (학생·교사·학부모)	• 계기판(대시보드)을 통한 학생의 학습데이터 분석 제공 • 교육 주체(교사, 학생, 학부모) 간 소통 지원 • 통합 로그인 기능, 쉽고 편리한 UI/UX 구성 및 접근성 보장

자료: 교육부(2023), "AI 디지털교과서 추진방안"

WEF는 대형언어모델(Large Language Model)이 교육 업무에 미치는 영향을 AI로 자동화가 가능한 업무(Automatable tasks), 증강가능한 업무(Augmentable tasks), AI 노출 가능성이 낮고 영향을 받지 않는 업무(Lower potential for exposure and

unaffected tasks)로 구분하였다. AI 기술의 도입으로 교사들의 행정적, 반복적 업무가 줄어들 수 있어, 교사들이 더 가치 있는 교육 활동에 집중할 수 있을 것이며, 교육 자료 개발과 학습 성과 분석 등에서 AI의 지원을 받아 더 효과적이고 개인화된 교육이 가능해질 것이다. 그러나 학생들과의 직접적인 상호작용, 교육 목표 설정, 체험 학습 등 인간 교사의 고유한 역할은 여전히 중요하며, AI로 대체되기 어려울 것으로 분석했다.

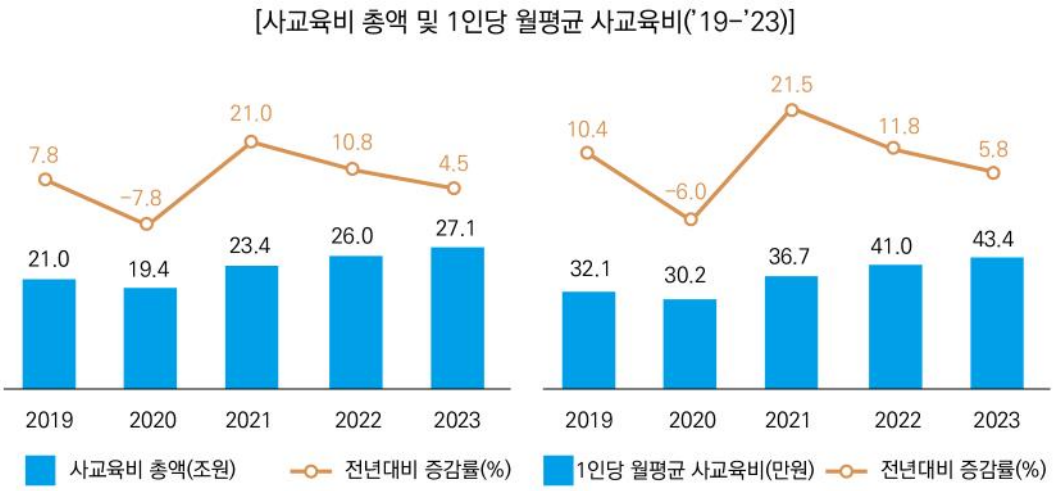
[표 5-4] LLM이 교육 업무에 미치는 영향

구분	내용
자동화 가능 업무 (Automatable tasks)	• 특정 주제 도서, 정기 간행물, 기사 및 시청각 자료 목록 작성 • 표준 참고자료를 사용하여 사실, 날짜 및 통계 확인 • 답안지 또는 전자 채점 장치 사용→숙제와 시험 채점, 결과 계산 및 기록
증강 가능 업무 (Augmentable tasks)	• 교육 시스템, 과정 또는 자료의 효과성 판단을 위한 성과 데이터 분석 • 유인물, 학습자료, 퀴즈와 등 교육 자료 개발 • 교사 보조원, 자원봉사자를 위한 과제 준비
노출 가능성 낮고 영향을 받지 않는 업무 (Lower potential for exposure and unaffected tasks)	• 모든 수업, 단원 및 프로젝트에 대한 명확한 목표 설정 • 학생과의 소통 • 교과 과정 개선을 위한 정부, 지역사회 그룹 리더와 협의 • 학습 평가 및 개정에 있어 다른 교사 및 관리자와 협력 • 현장 학습, 기타 체험 활동 계획 및 감독을 통한 학습 지도

자료: WEF(2024), “Shaping the Future of Learning: The Role of AI in Education 4.0”

AI 디지털교과서의 도입으로 인구감소 대응 및 사교육비 절감도 효과도 기대되고 있다. 한국의 인구는 2022년 합계출산율 0.78 최저수준 기록 후 지속 하락 중이며 출산 하락의 약 26%는 사교육비 증가에 기인한다. 학생 1인당 월평균 실질 사교육비 1만 원 오르면 합계출산율 0.012명 감소한다.[85] 2023년 사교육비 총액은 27.1조 원으로 전년 대비 4.5% 증가하였으며 1인당 월평균 사교육비는 43.4만 원으로 전년 대비 5.8% 증가하였다. 2023년 사교육 참여율은 78.5%에 이른다.[86]

85) 한국경제연구원(2023)
86) 교육부 보도자료(2024.3.14.), “2023년 초중고사교육비조사 결과”



[그림 5-7] 사교육비 총액 및 1인당 월평균 사교육비(2019~2023)

자료: 교육부 보도자료(2024.3.14.), "2023년 초중고사교육비조사 결과"

이에 공교육 현장에서 AI 기술을 활용한 학생별 맞춤형 교육이 선도적으로 제공된다면 사교육 의존도를 획기적으로 낮추고 교육격차를 해소하는 될 것이라 기대되고 있다. 2024년 3월에는 디지털 기반 공교육 혁신에 관한 정책과 지원 사항을 체계적으로 규율하는 「디지털 기반 공교육 혁신에 관한 특별법 제정안」이 발의되기도 하였다. 양질의 교육콘텐츠와 기술적 안정성이 확보된 AI 디지털교과서가 개발, 보급돼 교육 현장에 안정적으로 정착하고 학생, 학부모, 교원이 모두 만족하는 맞춤형 교육 환경이 제공될 수 있도록 한단 취지의 법안이다.

2 AI 디지털교과서에 대한 우려

AI 디지털교과서에 대한 기대와 함께 우려하는 의견도 제기되고 있다. 정부가 2025년부터 도입을 추진하는 AI 디지털교과서를 유보해달라는 국회 국민 동의 청원이 동의자 5만 명을 초과하여 국회 교육위원회에 회부되었다. 2024년 5월 28일 게시된 「교육부의 2025 AI 디지털교과서 도입 유보에 관한 청원」은 2024년 6월 27일일 기준 5만 6,505명의 동의를 받아 국회 교육위원회에 회부되었고 국민동의청원은 30일 동안 5만 명 이상의

동의를 받으면 소관 국회 상임위원회에 회부된다.

청원의 내용은 성장기 어린이 청소년에게 뇌 발달에 악영향을 미치고 여러 부작용이 큰 디지털 기기를 수업에 사용하는 것을 축소 혹은 유보하고 종이 교과서와 글쓰기, 읽기 등의 아날로그적 교육방식을 재도입 확대하고 있는 해외 사례를 충분히 고려해야 하며 이에, AI 디지털교과서의 사용 결정 또한 재논의되어야 한다는 것이다. 스마트기기가 널리 사용된 10여 년의 시간 동안 많은 뇌과학자, 정신의학자, 교육전문가들이 스마트기기 사용의 심각한 부작용을 밝혀내어 그 유해성을 여러 매체를 통해 꾸준히 알려왔으며 이런 상황에서 평균적으로 하루 일과의 절반 이상 시간을 보내는 학교에서조차 스마트기기를 이용해야 하는가에 대해 의문을 제기하였다.

또한, 교육부 방침대로 학교 현장에 도입이 되려면 적어도 지금은 모든 교과에 대한 프로토타입이 완성되어 장단점 분석 또한 상당히 진행되어 있어야 하지만, 현실은 그렇지 않으며 관련 언론 보도에서도 AI 디지털교과서로 수업을 해야하는 당사자인 교사들의 반응도 결코 좋은 상황이라 할 수 없다는 것이다. 이에 교육부는 2025년 AI 디지털교과서 도입 방침에 대해 전면 취소할 수 없다면 적어도 도입 유보를 발표하고 보다 면밀한 검토와 연구 분석을 통해 교육 보조자료로서의 디지털 기기가 아닌 전면적인 디지털교과서 사용이 서면 교과서를 사용하는 것보다 객관적, 과학적으로 더 효과적인 교육방식이 맞는지 검증하는 과정을 거친 후 이 정책에 관해 다시 논할 것을 요청하였다.[87]

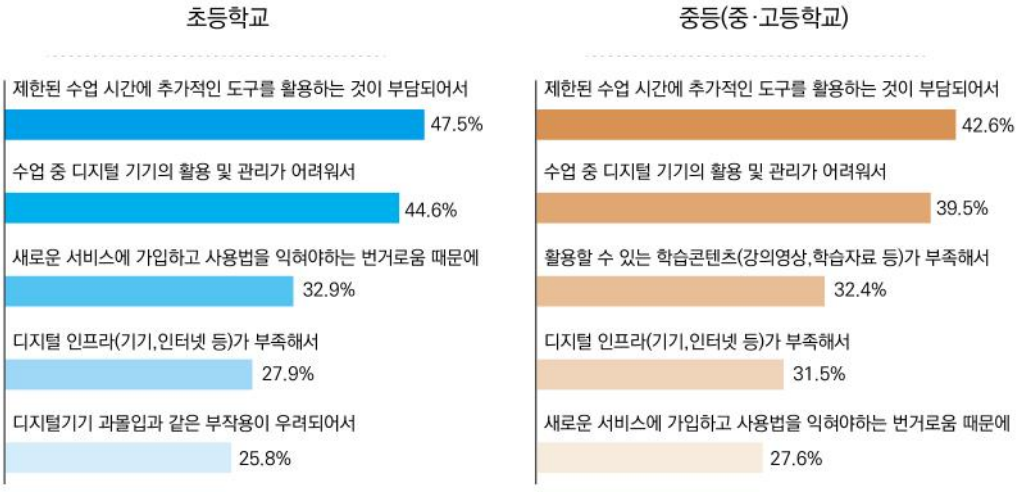
국회 교육위원회는 2024년 7월 12일 전체 회의를 열어 교육부와 국가교육위원회 등 관련 기관으로부터 업무보고를 받고 주요 교육정책에 대해 논의했으며 AI 디지털교과서에 대해서도 질의가 이루어졌다. 구체적으로 AI 교과서의 서책 교과서 대체 여부, AI 디지털교과서 도입에 필요한 인프라 준비 상황, AI 디지털교과서의 학습효과와 해외 사례에 기반한 도입 타당성에 대한 논의가 이루어졌다.

다수의 언론 보도를 통해서도 다양한 문제가 제기되고 있다. AI 디지털교과서 개발을 아직 완료하지 않았는데도, 많은 예산을 투입하여 교원들을 대상으로 관련 연수가 진행되고 있는데 이에, AI 디지털교과서 없는 AI 디지털교과서 연수라는 비판도 제기되고 있다.[88] AI 디지털교과서 도입 관련 개인정보 문제도 제기되었다. 전국 1,000개 디지털 선

87) https://petitions.assembly.go.kr/about/notices/152C967DCD270F23E064B49691C1987F
88) 한겨레 신문(2024.7.18), "당장 내년 도입인데, AI 디지털교과서 없는 교사 연수라니"

도학교에서 활용되는 AI 코스웨어에서 학생들의 개인정보를 광범위하게 수집했으며 AI 코스웨어 가입 시 행태 정보 수집을 거부하면 활용이 불가하거나, 필수 동의를 받은 뒤 학생들의 행태 정보를 온라인 광고사에 제공한 SW 업체도 있었고, 고위험 정보로 불리는 안면 데이터 등 생체 정보를 수집한 곳도 있었다는 것이다.[89]

교사의 AI 디지털교과서에 대한 인식, 활용, 관리 측면에서도 이슈가 제기되고 있다. 실제로 교사가 AI 기반 맞춤 교육 서비스를 사용해 본 경험은 매우 낮은 것으로 나타났는데“들어봤지만 사용해 보진 않았다.”(40.6%),“들어본 적 없다.”(21.3%), “간단하게 몇 번 사용해 본 적 있다.”(28.7%)로 분석되었다. 또한, 교사가 AI 맞춤 교육 서비스를 수업에 활용하지 않는 것은 수업 적용, 기기 관리 등에 대한 부담 등이 가장 큰 요인으로 조사되었다.[90]



[그림 5-8] AI 기반 맞춤형 서비스를 활용하지 않는 이유

자료: 한국교육개발원(2023), “AI 기반 맞춤형 교육의 현황과 과제”

2024년 7월, 전국 초중고 재학 자녀가 있는 학부모 1,000명과 초중고 교원 19,667명을 대상으로 AI 디지털교과서 도입에 대한 설문결과 ‘AI 디지털교과서 도입 정책에 동의

89) 경향신문(2024.7.24.), “디지털 선도학교서 쓰는 AI앱, 학생 정보 술술 빼갔다”
90) 한국교육개발원(2023), “AI 기반 맞춤형 교육의 현황과 과제”

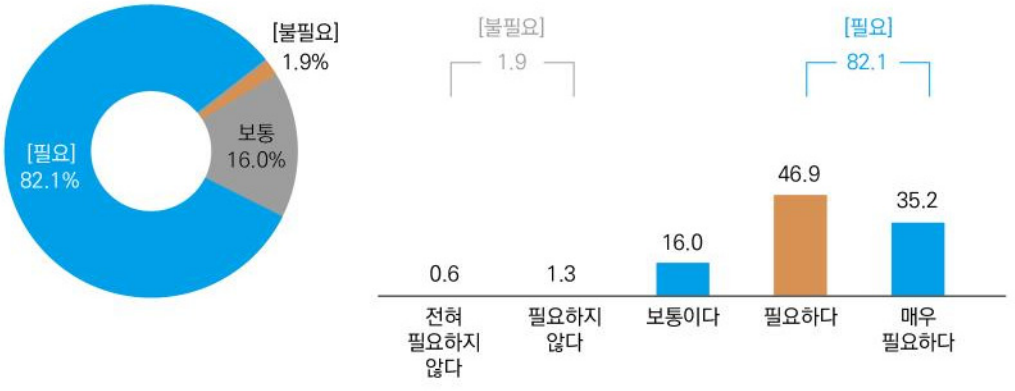
하느냐'는 질문에 학부모는 비동의 31.1%, 동의 30.7%이었고, '보통이다' 답변이 38.2%였다. 교원의 경우에는 비동의가 73.6%로 큰 수치를 보였고 보통 14.3%, 동의 12.1% 순이었다. AI 디지털 교과서 도입에 동의하지 않는 이유로 학부모는 '디지털 기기에 지나치게 의존할 것이 우려되어서' 39.2%, '학생들의 문해력이 저하될 것 같아서' 35.7%로 응답했다. 교원은 '학습 효과성 의문' 35.5%, '디지털 기기 과의존 우려' 25.7%, '교사·학생·학부모 의견수렴 부족' 22.7% 순이었다.

[표 5-5] 교원의 AI 디지털교과서 도입 설문(단위, %)

전혀동의 하지 않음	동의하지 않음	보통	동의	매우 동의	종합		
					비동의	보통	동의
52.7	20.9	14.5	9.2	2.9	73.6	14.3	12.1

자료: 국회 교육위원회 고민정 의원실, 엠브레인

또한, AI 디지털교과서 도입에 대한 사회적 공론화가 필요한지에 대한 설문에 학부모의 82.1%, 교원의 88.6%가 필요하다는 의견을 제시하였다. 교원 대상으로 AI 디지털교과서 도입 관련 바람직한 추진 방향을 물은 결과에는 '학습 효과성 관련 장기적 연구를 거쳐 추진해야 한다'는 의견이 60.3%, 'AI 디지털교과서 시범학교 운영 등 현장 적합성 검토후 단계적 추진해야 한다'는 의견이 30.1%로 나타났다.



[그림 5-9] AI 디지털교과서 도입에 대한 사회적 공론화 절차 필요성

자료: 국회 교육위원회 고민정 의원실, 엠브레인

AI 디지털교과서의 법적근거 관점에서의 이슈도 존재한다. 2023년 10월 24일 개정·시행된 「교과용도서에 관한 규정」제2조(정의) 제2호에 따르면“교과서”의 정의에 지능정보화기술을 활용한 학습지원 소프트웨어인 디지털교과서를 포함하는 것으로 규정하였다. 그러나 근거 법률인 「초·중등교육법」제29조(교과용 도서의 사용) 제2항[91] 신설 당시, 입법자가 교과용 도서의 범위 등을 대통령령으로 위임하면서 현재의 서책형(書冊形) 교과서 외에 AI와 결합하여 학습데이터를 수집·분석하는 디지털교과서까지 포함되는 것을 예측할 수 있었다고 볼 수 있는지 논란이 될 수 있다고 지적되고 있다.[92]

91) 제29조(교과용 도서의 사용) ② 교과용 도서의 범위·저작·검정·인정·발행·공급·선정 및 가격 사정(査定) 등에 필요한 사항은 대통령령으로 정한다.

92) 입법조사처(2024), “제22대 국회 입법정책 가이드북 Ⅲ”

제3절 AI 디지털교과서 FGI 분석

NATIONAL ASSEMBLY FUTURES INSTITUTE

AI 디지털교과서 FGI 분석 관련 기본 절차는 3장에서 서술한 참여자, 진행 방식과 같으며 논의의 중심을 AI 디지털교과서와 교육으로 한정하여 진행하였다. 연구진이 AI 디지털교과서의 개념과 주요 현황을 공유하고 이후 AI 전문가 그룹이 의견을 제시하였으며 주요 내용은 아래 표와 같다.

[표 5-6] AI 디지털교과서 FGI 주요 내용

구분	내용
K교수 (C 대학)	• "현재 AI 디지털교과서에서 언급되는 AI의 개념이 너무 모호. LLM, 생성형 AI, 전통적인 기계 학습 방식이 어떻게 구분되고 적용되는지, 또는 함께 활용되는지 구분이 필요" • "AI 교과서의 실제 학습효과에 대한 과학적 검증이 필요" • "다양한 측면에서 구체적인 학생 데이터 보호 방안이 제시되어야" • "입시제도와의 연계 방안이 제시되어야"
L대표이사 (P 에듀테크)	• "각 교과목의 AI 적용 수준과 방식에 대한 명확한 지침이 필요" • "기존 평가체계로는 AI 기반 학습의 효과를 정확히 측정할 수 없음" • "AI 교과서는 학습자 중심의 개별화 학습을 가능하게 할 것" • "AI 교과서와 오프라인 비중 조화가 필요"
L전무 (M AI기업)	• "최신 AI 기술 발전 속도를 고려할 때, AI 디지털교과서 교육사례도 계속 발굴되고 업데이트 필요" • "AI 디지털교과서 도입 과정의 모든 단계를 투명하게 공개해야" • "한국어 교육데이터 기반의 자체 AI 모델 개발이 필요" • "LLM 모델의 교육 특화 버전 개발이 시급. 학생 수준별 맞춤형 학습을 위해서는 교육용 AI의 특화 개발이 필수적"
H책임연구원 (S연구소)	• "AI 디지털교과서의 한계점과 가능성을 보여주는 최신 사례 분석이 필요" • "과거 AI 교육사례와 현재의 변화 간 차이가 큼" • "AI 디지털교과서 현장의 의견을 수렴하는 공청회를 정기적으로 개최하고 결과를 공개하며 업데이트해야" • "장기적인 관점에서의 학습효과 추적 연구가 필요"

구분	내용
Y책임연구원 (S연구소)	• “AI 디지털교과서 도입 전후의 학습효과를 객관적으로 측정할 수 있는 지표 개발이 시급” • “과도한 AI 의존 교육은 학생들의 사고력 발달을 저해할 수 있음” • “AI와 함께 살아갈 미래세대를 위한 교육철학 재정립이 선행되어야 함”
K이사 (S사)	• “학생들의 AI 교육, 리터러시 향상도를 정량적으로 평가할 방법이 필요” • “장기적인 관점에서의 학습효과 추적 연구가 필요” • “AI 없이는 학습 불가능한 세대 출현 가능성”
Y이사 (A사)	• “글로벌 사례, 교사, 학생, 학부모의 의견을 계속 모니터링하고 반영할 수 있는 체계가 필요.” • “AI 알고리즘 편향성 문제를 해결하지 않으면 교육 불평등이 심화.” • “AI-아날로그 혼합 모델로 발전” • “AI 교육의 강화가 다차원적 교육 불평등 이슈를 제기할 수도”
S차장 (K사)	• “AI 디지털교과서 대한 사회적 합의를 위한 다양한 이해관계자 참여가 중요” • “모든 학생이 동등한 접근성을 가질 수 있도록 인프라 구축이 필요” • “글로벌 AI 교육 플랫폼 지배 구조 형성 가능성” • “AI로 인한 교육의 초 국가화, 국가 간 교육 경계 붕괴 가능성”
K대표이사 (S사)	• “AI 교육 시스템의 편향성을 검증하고 수정할 수 있는 구체적인 프로세스가 필요” • “AI 교과서의 장단점에 대한 객관적인 정보 공개가 필요” • “향후, 교육 관련, AI 의존 불안장애 이슈 제기될 수도” • “미래 프리미엄 아날로그 교육이 등장할 수도”

AI 디지털교과서와 교육 이슈 FGI 분석 관련 논의된 사항을 주요 키워드 중심의 워드 클라우드(Word Cloud)로 정리하면 아래 그림과 같다.

[그림 5-10] FGI 논의사항 워드 클라우드

FGI에서 다양하게 논의된 사항을 보다 체계적인 토픽(Topic)으로 정리하기 위해 AI 전문가 그룹의 의견을 텍스트로 전환하고 이를 기반으로 토픽 모델 분석을 활용하여 8가지 토픽으로 구분하였다. 연관 키워드를 중심으로 토픽 명을 부여하였고 그 결과 AI 개념 혼선, AI 진화에 맞는 사례 분석, 다양한 관점, 공정성과 안전, 학습효과 측정, 투명한 공론화, 법제도 개선, 새로운 기회와 위험으로 토픽이 구성되었다. 연관어 기반 분류된 토픽 구성은 아래와 같다.

[그림 5-11] FGI 논의사항 기반 토픽 구성 결과

세부 내용을 보면, 현재 AI 디지털교과서에서 사용되는 AI 개념이 모호하다는 문제가 제기되었다. LLM, 생성형 AI, 전통적 기계 학습 방식 간의 구분과 통합적 활용방안에 대한 명확한 정의가 있어야 한다는 견해가 제시되었다. 특히 교육용 AI의 특화 개발과 한국어 교육데이터 기반의 자체 AI 모델 개발의 필요성이 강조되었다. AI 진화에 맞는 사례 분석과 해석도 필요하다는 의견도 있었다. AI 기술의 빠른 발전 속도를 고려할 때, 지속적인 교육사례 발굴과 업데이트가 필요하며 과거 AI 교육사례와 현재의 변화 간 격차가 크며, AI 디지털교과서의 한계점과 가능성을 보여주는 최신 사례 분석이 요구된다는 견해이다. 다양성 관점에서 글로벌 사례, 교사, 학생, 학부모의 의견을 꾸준히 모니터링하고 반영할 수 있는 체계 구축이 요구된다는 의견도 제시되었다. AI와 함께 살아갈 미래세대를 위한 교육철학 재정립이 선행되어야 하며, AI-아날로그 혼합 모델로의 발전 가능성도 논의되었다. 공정성과 안전 측면에서 AI 알고리즘의 편향성 문제가 교육 불평등 심화로 이어질 수 있다는 우려가 제기되었고 모든 학생의 동등한 접근성 보장을 위한 인프라 구축과 학생 데이터 보호 방안 마련이 시급하다는 의견도 포함되었다. 학습효과 측정 관련, 기존 평가체계로는 AI 기반 학습의 효과를 정확히 측정할 수 없다는 한계가 지적되었고. AI 디지털교과서 도입 전후의 학습효과를 객관적으로 측정할 수 있는 지표 개발과 장기적 관점에서의 학습효과 추적 연구 필요성도 제기되었다. 투명한 공론화 관련, AI 디지털

교과서 도입 과정의 모든 단계를 투명하게 공개하고, 정기적인 공청회 개최를 통해 현장의 의견을 수렴해야 하며 다양한 이해관계자의 참여를 통한 사회적 합의 도출 중요성이 논의되었다. 법제도 개선 관련, 입시제도와의 연계 방안, 각 교과목의 AI 적용 수준과 방식에 대한 명확한 지침, AI 교육 시스템의 편향성을 검증하고 수정할 수 있는 제도 개선 구체적인 프로세스 수립도 요구되었다. 미래 새로운 기회와 위험 관련, AI로 인한 교육의 초 국가화, 국가 간 교육 경계 붕괴 가능성, 글로벌 AI 교육 플랫폼의 지배 구조 형성, AI 없이는 학습 불가능한 세대 출현, AI 의존 불안장애 이슈, 미래 프리미엄 아날로그 교육의 등장 가능성과 AI 교육의 강화가 다차원적 교육 불평등 이슈를 제기할 수도 있음이 논의되었다.

FGI 내용 기반 구성된 8개 토픽 간 연관성 분석도 수행하였다. AI 개념 혼선과 AI 진화 사례 간의 연관성은 0.85로 가장 높은 수준이다. 이는 AI에 대한 기본적 이해와 실제 사례 학습이 매우 밀접하게 연결되어 있음을 의미한다. 또한, 두 토픽 간의 높은 연관성은 AI 교육에서 실제 사례 중심의 학습이 효과적일 수 있음을 시사한다. 학습효과 측정이 새로운 기회와 위험(0.73)과 상당한 연관성을 보이는 것은 교육적 성과와 실제 적용 가능성이 밀접하게 연결되어 있음을 의미한다. 공정성과 안전은 법제도 개선과 0.83의 높은 연관성을 보이고 있으며, 이는 AI 안전성과 공정성 문제가 제도적 대응과 직결되어 있음을 시사한다. AI 개념 혼선과 법제도 개선 간의 연관성은 0.45로 상대적으로 낮은데, 이는 현재 AI에 대한 이해도와 제도적 대응 사이에 상당한 간극이 존재함을 의미한다. 또한, 대부분 토픽이 0.5 이상의 연관성을 보이며 이는 AI 관련 정책이 개별적이 아닌 통합적 관점에서 수립되어야 함을 시사한다. 특히 다양한 관점과 투명한 공론화 간의 높은 연관성(0.81)은 사회적 논의 과정의 중요성을 강조한다. AI 디지털교과서와 교육 관련 토픽 간의 연관성 분석 결과는 정책 수립과 교육 프로그램 설계에 있어 통합적이고 균형 잡힌 접근이 필요함을 보여준다. 특히 AI에 대한 이해도 제고, 제도적 대응, 사회적 논의가 유기적으로 연계되어야 하며, 이를 위한 체계적인 전략 수립이 요구된다.

	AI 개념 혼선	AI 진화 사례	다양한 관점	공정성과 안전	투명한 공론화	법제도 개선	학습효과 측정	새로운 기회와 위험
AI 개념 혼선	1.00	0.85	0.72	0.78	0.58	0.45	0.65	0.62
AI 진화 사례	0.85	1.00	0.76	0.71	0.55	0.48	0.76	0.66
다양한 관점	0.72	0.76	1.00	0.69	0.81	0.54	0.75	0.68
공정성과 안전	0.78	0.71	0.69	1.00	0.79	0.83	0.65	0.67
투명한 공론화	0.58	0.55	0.81	0.79	1.00	0.74	0.60	0.64
법제도 개선	0.45	0.48	0.54	0.83	0.74	1.00	0.52	0.51
학습효과 측정	0.65	0.76	0.75	0.65	0.60	0.52	1.00	0.73
새로운 기회와 위험	0.62	0.66	0.68	0.67	0.64	0.51	0.73	1.00

[그림 5-12] 토픽 간 연관성 분석

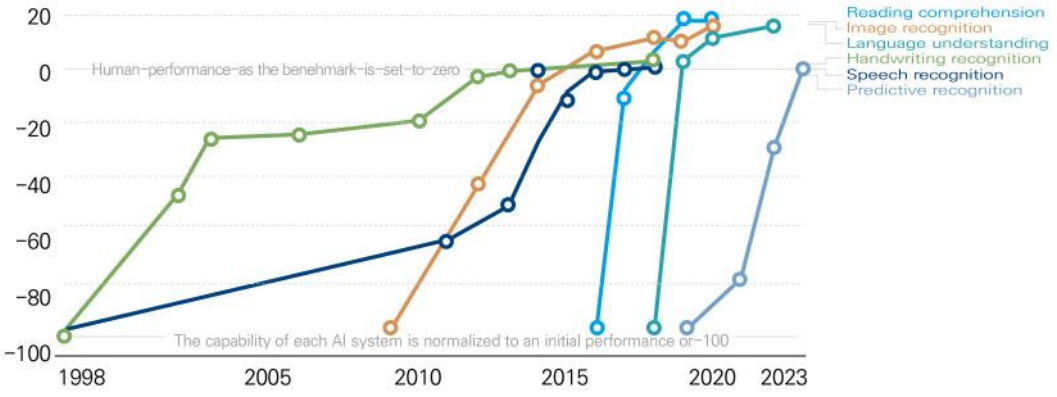
제4절 AI 디지털교과서 이슈 분석

NATIONAL ASSEMBLY FUTURES INSTITUTE

4절에서는 FGI를 통해 도출된 주요 이슈를 다양한 사례와 선행 연구, 검토의견을 중심으로 추가 분석하고자 한다.

먼저, AI에 대한 개념의 혼선 관련, 통칭하여 AI는 인간의 능력을 컴퓨터나 기계가 할 수 있도록 만드는 기술로 표현되나 기술의 발전에 따라 세부적으로 머신러닝(Machine Learning), 딥러닝(Deep Learning), 초거대 AI 및 생성형 AI 등으로 구분할 수 있다. 과거부터 존재하던 머신러닝과 최근 주목받고 있는 초거대 AI 및 생성형 AI는 성능과 적용 방식에 있어 현저한 차이가 존재한다. 현재 AI 디지털교과서를 둘러싼 교육 주체들과 다수의 이해관계자가 AI 디지털교과서에 적용되는 AI에 대해 혼재된 개념을 사용하고 있으며 이에 어떠한 범주에 있는지 함께 인식하고 논의를 전개 시켜나갈 필요가 있다. 동일하게 AI로 표현하더라도 세부 분류된 AI가 적용되는 방식과 파급효과에는 큰 차이가 존재하기 때문이다.

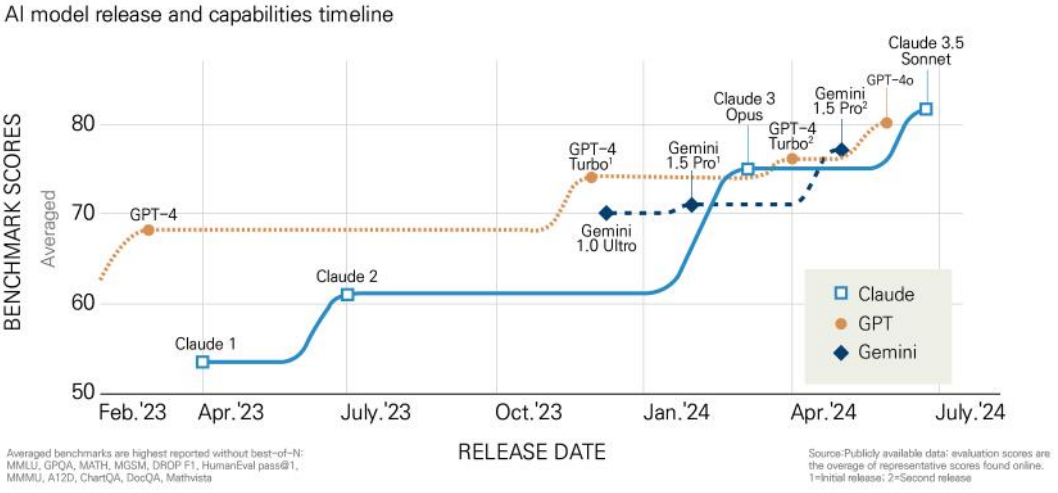
둘째, AI 기술 진화 속도를 고려한 사례 분석이다. AI는 긴 시간 축적되어 온 기술로 다양한 AI 요소기술이 인간의 수준을 넘어서는 성과를 보이며 진화하고 있다. 2010년, 2015년 그리고 현재의 AI는 완전히 다른 수준에 있으며 이는 AI 기반 교육사례, 효과를 분석할 때도 동태적 관점에서 입체적으로 고려해야 함을 시사한다.



[그림 5-13] AI 기술의 진화

자료: https://ourworldindata.org/grapher/test-scores-ai-capabilities-relative-human-performance

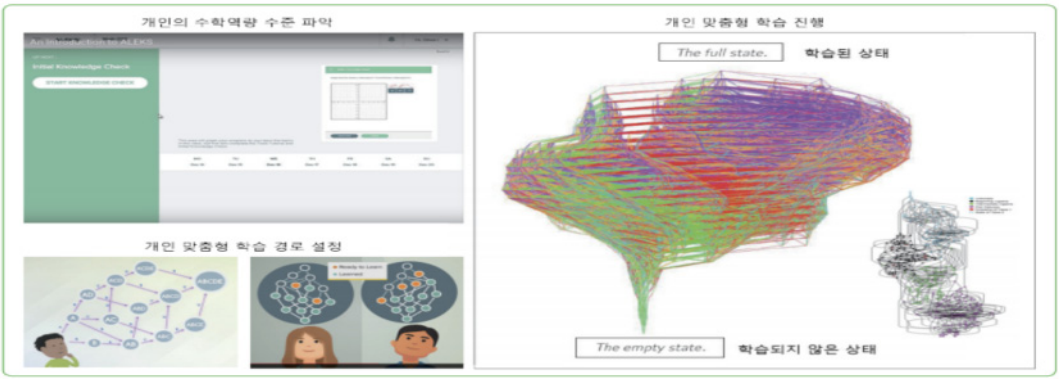
이는 과거 특정 시점의 AI 교육사례가 현재 일반화되기 어려울 만큼 기술이 빠르게 진화하고 있다는 것을 뜻한다. 아래 그림은 2023년 2월부터 2024년 6월까지의 초거대 AI 모델의 성능을 비교한 것이다. 초거대 AI의 경우, 1년 만에 놀라운 성능의 변화가 생겼으며 가속은 계속 진행 중이다. 과거의 AI 사례와 효과가 현재 적용되는데 한계가 있다는 의미이다.



[그림 5-14] 초거대 AI 모델 성능 변화

자료: https://evolvingai.io/p/anthropic-now-leads-ai-race

교육 분야에 AI를 도입하려는 시도는 오래전부터 이루어져 왔다. 교육 분야의 AI가 도입 논의는 1970년대에도 이루어졌다.[93] 교사의 일대일 지도는 학습자에게 가장 효과적인 접근법이지만, 모든 교사와 학습자 간의 일대일 지도란 당시 불가능했고, 이에 컴퓨터로 일대일 개인 지도를 대체할 방법이 연구되었다.[94] 구체적으로 규칙(Rule) 기반의 AI 기술을 활용하여 적응형 학습(Adaptive Learning) 및 개인화 학습(Personalized learning)을 자동으로 구현하고자 하는 노력이 이루어졌다.[95] 교육 분야에서 AI의 적용은 학생 대상 AI(학습 및 평가 지원 도구)와 시스템 대상 AI(교육 관리 지원)를 포함하여 여러 방향으로 발전해 왔다.[96] 이후에도 이러한 노력은 계속되어 AI는 개인 맞춤학습 제공, 학습 몰입도 측정, 학사 행정 자동화 등 교육의 다양한 분야와 접목되어 활용되었다. 맥그로힐(McGraw Hill Education)의 알렉스 수학, 드림박스 러닝(Dreambox learning)의 영어 등이 그 예이다.[97]

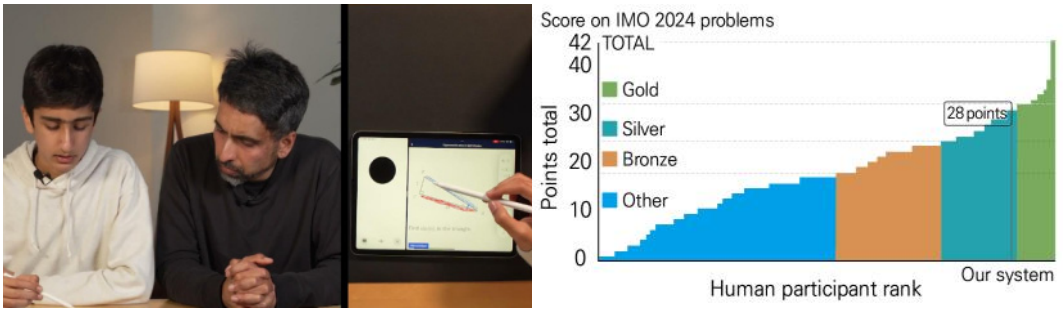


[그림 5-15] AI 기반 알렉스(ALEKS) 시스템의 수학교육

자료: 이승환(2018.10), "교육혁신의 동력, AI", SW정책연구소 산업동향.

93) UNESCO(2021), "AI and education: guidance for policy-makers"
94) Bloom, B. S. 1984. The 2 Sigma Problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, Vol. 13, no. 6, pp. 4-16.
95) Carbonell, J. R. 1970. AI in CAI: An artificial-intelligence approach to computer-assisted instruction. IEEE Transactions on ManMachine Systems, Vol. 11, No. 4, pp. 190-202. ; Self, J. A. 1974. Student models in computer-aided instruction.International Journal of Man-Machine Studies, Vol. 6, No. 2, pp. 261-276.
96) Baker, T., Smith, L. and Anissa, N. 2019. Educ-AI-tion Rebooted? Exploring the future of artificial intelligence in schools and colleges. London, NESTA. Available at: https://www.nesta.org.uk/report/education-rebooted (Accessed 28 March 2021).
97) 이승환(2018.10), "교육혁신의 동력, AI", SW정책연구소 산업동향.

최근의 교육은 초거대 AI를 기반으로 빠르게 진화하고 있다. 2024년 5월, 오픈 AI는 진화된 초거대 AI 모델 챗GPT-4o를 발표하며 다양한 기능을 공개하였고 그 중 수학 AI 튜터가 포함되었으며 학습자가 초거대 AI와의 상호작용으로 학습하는 모습도 포함되었다. 2024년 7월에는, 구글의 딥마인드(Deepmind)는 수학 AI 알파프루프(AlphaProof)와 기하학 문제를 푸는 AI인 알파지오메트리2(AlphaGeometry2)가 국제수학올림피아드(IMO, International Mathematical Olympiad)[98]에서 은메달 수준의 성적을 기록했다고 발표했다. AI가 국제수학올림피아드의 메달권에 해당하는 성능을 보인 건 처음이다.[99]



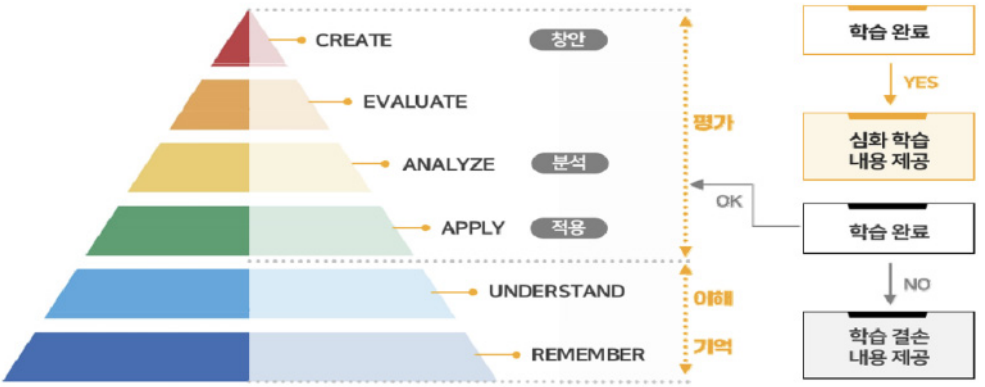
[그림 5-16] 오픈AI 수학 AI튜터(좌)와 구글의 IMO 수학 AI 성적(우)

자료: https://openai.com/index/hello-gpt-4o/, Deepmind(25 July 2024), "AI achieves silver-medal standard solving International Mathematical Olympiad problems"

셋째, AI 교육에 대한 다양성 이슈이다. 블룸의 신교육 목표 분류에 따르면 학습자는 기억, 이해의 수준부터 분석과 창의의 영역까지 다양한 과정을 통해 학습한다. AI의 적용은 목표 분류, 구현 대상과 과목, 효과, 투입비용과 생산성 등에 따라 머신러닝 방식부터 초거대 AI까지 다양한 조합으로 설계될 수 있다.

98) 1959년부터 열린 IMO는 예비 수학자들이 겨루는 권위 있는 대회이며 수학계 노벨상으로 불리는 필즈상 수상자 다수가 IMO 대표로 나설 정도다. AI 분야가 떠오르면서 IMO는 AI의 수학적 추론 능력을 평가하는 기준으로도 사용되고 있다.

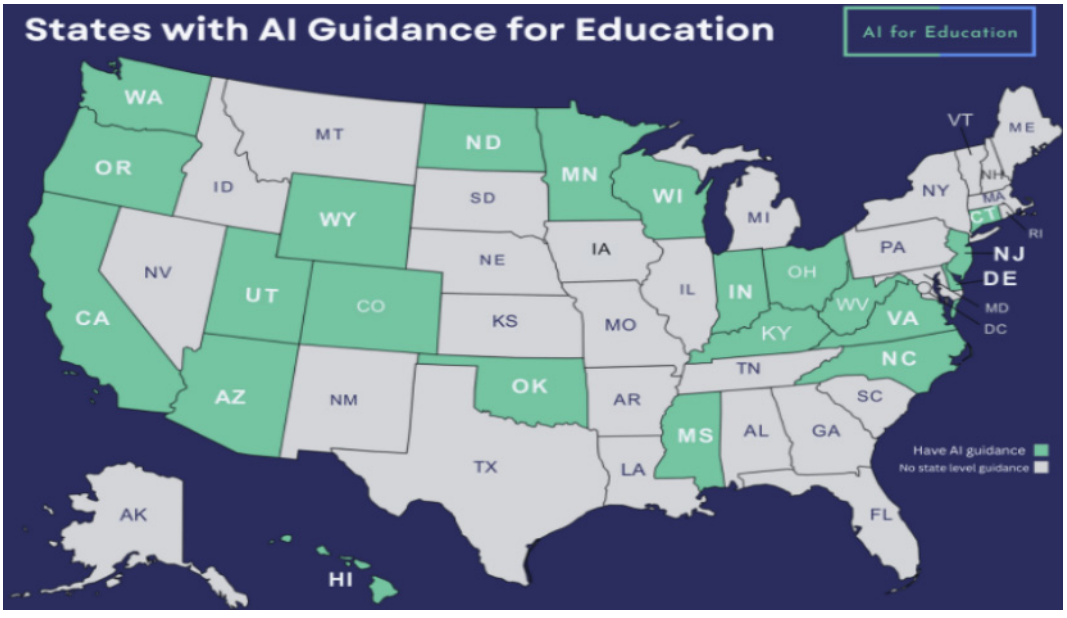
99) Deepmind(25 July 2024), "AI achieves silver-medal standard solving International Mathematical Olympiad problems"



[그림 5-17] 블룸의 신교육 목표 분류(Bloom's New Taxonomy)

자료: 교육부, 한국교육학술정보원(2023), "AI 디지털교과서 개발 가이드라인"

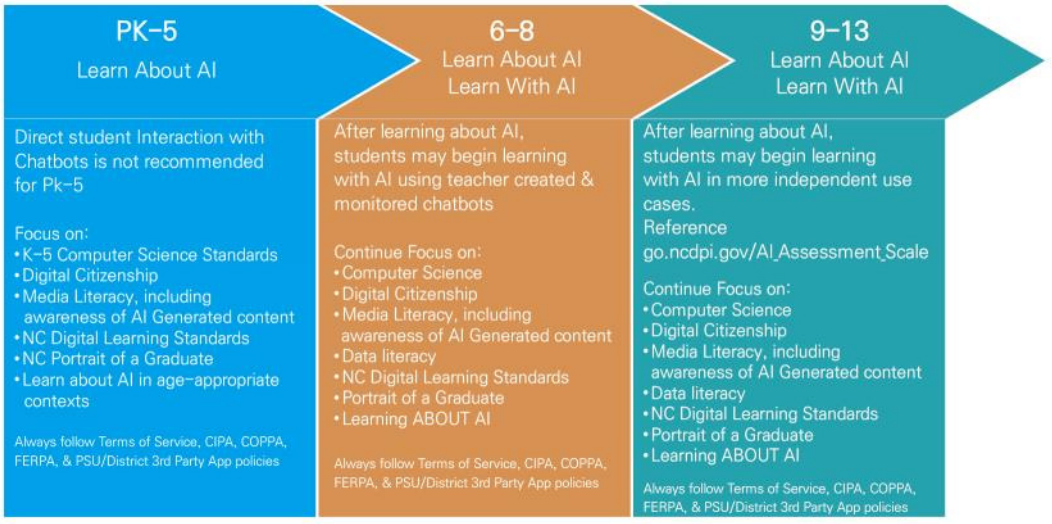
지난 1년간 미국의 22개 주 교육부가 K12 학교를 위한 AI 지침을 발표했으며, 이를 통해 교육의 AI 활용에 얼마나 많은 관심과 노력이 이루어지고 있는지 알 수 있다.



[그림 5-18] K12 AI 지침을 발표한 미국 22개 주

자료: https://www.aiforeducation.io/ai-resources/state-ai-guidance

미국 22개 주의 교육부는 AI 사용의 기회와 위험 인식, 윤리적인 AI 사용 강조, 교육자와 학생을 위한 AI 리터러시 증시, AI를 인간 교육 대체가 아닌 보완 도구로 사용 권장, 프라이버시 보호 등 주요 원칙 관점에서는 공통된 견해를 보였다. 하지만, 실제 적용 관련해서는 차이점이 존재했으며 이는 교육의 AI 적용에 있어서 다양한 의견이 논의되고 있는 과정이라고 볼 수 있다. 일례로 노스캐롤라이나 주 공립학교는 PK~5(유치원~5학년)에서는 AI에 대해 배우는 단계로 챗봇과의 직접적인 상호작용은 권장되지 않으며, 컴퓨터 과학 표준, 디지털 시민의식, AI 생성 콘텐츠 인식을 포함한 미디어 리터러시에 집중하고 있다. 6~8학년에서는 교사가 만들고 모니터링하는 챗봇을 통해 AI 학습을 시작하고, 컴퓨터 과학, 디지털 시민의식, 미디어 리터러시 계속 집중한다. 9~13학년에서는 더 독립적으로 AI 사용이 가능하며, 관련한 리터러시 소양도 계속 강조된다. 이처럼 현재 교육의 AI 활용에 있어서 다양한 논의가 이루어지고 있음에 주목하고 관련 모니터링을 강화해야 한다.



[그림 5-19] 학생의 AI 리터러시 타임라인 권고사항

자료: North Carolina Department of Public Instruction(2024), "North Carolina Generative AI Implementation Recommendations and Considerations for PK-13 Public Schools"

다양성 관점에서 온라인과 오프라인의 조화도 고려할 요소이다. 초중등 교육에 있어 디지털 과몰입은 매우 중요한 이슈이며 이를 고려한 AI 디지털교과서 활용방안이 모색될 필요가 있다. 실제 AI 교육은 온라인을 기반으로 AI가 제시하는 문제를 풀고 수준을 높여가는 협소한 개념을 넘어 프롬프트 기반의 다양한 상호작용을 통해 새로운 아이디어를 만들고 오프라인과 융합하도록 구성이 가능하다. 다양한 과목별로 수업 전, 중, 후 과정에서 AI가 언제 개입되고, 어떠한 상호작용 일으키며 이를 다시 환류하는지에 대한 다양한 수업 방법론 개발이 요구된다. 학생들에게 챗GPT의 도움을 받되, 반드시 자신의 생각과 결론을 포함시키라는 지침을 마련하기, 챗GPT로 작성한 숙제를 바탕으로 토론이나 발표 활동 강화하기, 숙제를 마친 후 GPT의 도움을 받았던 부분에 대해 어떤 점이 도움이 되었고, 자신의 생각과 어떻게 연결되었는지 성찰 일기 쓰기 등[100] AI를 활용함에 있어 온·오프라인 융합으로 교육 주체 간 상호작용이 원활하게 일어날 수 있는 다양한 방안을 모색하고 공유할 필요가 있다.

넷째, 학습효과 측정 이슈이다. 세분류 기준 AI가 어떻게 적용·활용되는지 그리고 실제 효과가 있는지 측정·관리되어야 한다. 교육 분야 AI 윤리원칙에서도 이 부분을 강조하고 있다. 교수·학습 과정 등에서 AI가 적재적소에 활용되고 교육적 효과를 발휘하기 위해서는 AI가 분석·처리한 결과에 대한 근거와 판단 자료가 이해할 수 있는 언어나 형태로 교육당사자에게 제공되어야 할 것을 명시하고 있다.[101] AI와 교육에 대한 베이징 합의에서도 교육에서 AI가 미치는 영향 연구 및 분석을 지원하고 AI의 교육적 영향력을 평가하는 구조를 개발해야 함을 강조하고 있다.[102] 이러한 과정이 수반되어야 AI에 기반한 학습결과를 올바르게 활용할 수 있고 잠재성을 높일 수 있다. 교육 분야 AI 윤리원칙에서도 교육 분야 AI는 인간 존엄성에 대한 존중을 바탕으로 인간성장의 잠재성을 이끌 수 있도록 제공되어야 하며 AI의 교육적 지원을 받는 부분에 있어서 과도한 진단이나 예측은 자제하고, 학습 진단이나 예측 결과의 고지는 신중하게 이루어져야 할 것을 명시하고 있다.[103]

펜실베이아 대학의 Hamsa Bastani(2024)의 연구에서는 AI를 잘못 활용할 경우 오히려 학습효과가 떨어진다는 분석결과가 제시되었다. 연구에서는 1,000명의 고등학생, 수

100) 오마이뉴스(2024.7.30.), "15년 차 교사가 챗GPT로 '숙제봇' 만들었습니다"
101) 교육부(2022.8), "사람의 성장을 지원하는 교육분야 AI 윤리원칙"
102) 유네스코(1999), "BEIJING CONSENSUS on artificial intelligence and education"
103) 교육부(2022.8), "사람의 성장을 지원하는 교육분야 AI 윤리원칙"

학을 대상으로 연구를 진행했으며 1그룹(AI를 활용하지 않는 그룹), 2그룹(일반 GPT를 사용하는 그룹, GPT Base), 3그룹(교사가 설계한 AI 튜터를 사용한 그룹, GPT Tutor)으로 나누어 학습효과를 비교하였다.[104] AI의 도움을 받으며 시험을 보게한 결과 AI를 활용한 그룹의 성적이 향상되었고, 특히 3그룹의 성적이 가장 높은 것으로 분석되었다. 하지만 AI 없이 시험을 볼때는 2그룹의 성적이 1그룹의 학생보다 17% 낮게 나왔으며, 1그룹과 3그룹의 성적인 비슷한 것으로 분석되었다. 이는 학생들이 AI의 결과를 수동적으로 받아들이며 지나치게 AI 의존할 경우 오히려 학습효과가 떨어진다는 것을 의미한다.[105] AI를 활용한 교육은 AI가 전혀 개입되지 않은 형태(AI Free)부터 AI에 많은 권한을 부여하는 형태(AI Empowered)까지 다양한 유형으로 구성되고 통합될 수 있으며 적용과목과 대상 등 다양한 유형을 고려하여 학습효과가 측정되고 이에 기반한 체계적인 도입이 필요하다.

104) GPT Base는 일반적인 ChatGPT와 유사한 간단한 프롬프트가 사용되었고 학생이 질문하면 직접적인 답변을 제공하는 방식으로 학생들은 주로 답을 요청하는 단순한 질문을 했으며 단기적으로는 성적 향상을 보였지만, AI 없이 시험을 볼 때는 오히려 성적이 하락했다. GPT Tutor 그룹은 교사들이 특별히 설계한 상세한 프롬프트를 사용했다. 단계적인 힌트를 제공하고, 완전한 답을 바로 주지 않고 독립적으로 문제를 해결할 수 있는 방향으로 설계되었으며 이 경우 학생들은 더 복잡한 질문을 하고 문제해결 과정에 대해 논의하는 경향이 있었고 단기적인 성적 향상과 함께, AI 없이 시험을 볼 때도 기존 수준을 유지했다.

105) Hamsa Bastani et al, "Generative AI Can Harm Learning" (July 15, 2024). Available at SSRN: https://ssrn.com/abstract=4895486 or http://dx.doi.org/10.2139/ssrn.4895486

[표 5-7] 학습 역량 강화를 위한 학생 AI 활용과 통합

구분	내용
AI 미사용 (AI free)	• 모든 작업은 AI 도움 없이 완전히 수행되어야 함 • 학생들은 자신의 지식, 이해력, 기술에만 전적으로 의존해야 함 • AI 사용은 학생의 학문적 진실성 위반임 • 학문적 정직성 서약이 필요할 수 있음
AI 보조 (AI assisted	• AI는 브레인스토밍, 계획, 피드백과 같은 특정 작업에만 사용됨 • 최종 제출물에는 AI가 생성한 내용이 포함되지 않음 • 지정된 작업 이외의 사용은 학문적 진실성 위반임 • 최종 제출물에 AI 사용 내역과 채팅 링크를 포함한 공개 진술문이 필요함
AI 향상 (AI enhanced	• AI를 지식, 효율성, 창의성 향상을 위해 전반적으로 상호작용하며 사용 • 학생은 AI가 생성한 모든 내용에 대한 인간의 감독과 평가를 제공해야 함 • AI 생성 콘텐츠와의 상호작용과 비판적 참여가 필요함 • 학생은 AI 생성 콘텐츠의 정확성과 공정성에 책임을 짐 • 최종 제출물에 AI 사용 내역과 채팅 링크를 포함한 상세한 공개 진술문이 필요
AI 강화 (AI empowered)	• AI의 완전한 통합으로 이전에 불가능했던 것들을 창조가능하게 함 (창의적 사고가, 문제 해결사로서 학생들에게 권한 부여) • 학생은 AI가 생성한 모든 내용에 대한 인간의 감독과 평가를 제공해야 함 • 학생은 AI 생성 콘텐츠의 정확성, 공정성, 품질에 책임을 짐 • 최종 제출물에 AI 사용 내역과 채팅 링크를 포함한 상세한 공개 진술문이 필요

자료: WDE(2024), “Guidance for Wyoming School Districts on Developing Artifcial Intelligence Use Policy”, https://edu.wyoming.gov

다섯째, 공정성과 안전에 대한 검증 이슈이다. 평가 품질은 평가의 목적에 적합한지 확인하는 증거 수집을 통해 달성되며 평가 형평성은 평가 점수가 공정할 때, 즉 특정 그룹에 유리하거나 불리하지 않을 때 달성된다. 전통적인 논거 기반 테스트 타당성 이론은 일련의 추론을 통해 테스트 품질과 형평성을 지원합니다. 이러한 추론은 증거 수집을 통해 적절한 테스트 점수 해석을 보장하는 메커니즘(예: 관련 과제 유형)을 가정한다.[106]

AI 기반 평가의 타당성을 유지하려면 AI 기능을 평가하는 것이 필수적이며, 이는 증거

106) Chapelle, C. A. (2008). The TOEFL validity argument. In C. A. Chapelle, M. K. Enright, & J. Jamieson (Eds.), Building a validity argument for the Test of English as a Foreign Language (pp. 319-352). London : Routledge.; Kane, M. T. (1992). An argument-based approach to validity. Psychological Bulletin, 112(3), 527-535. https://psycnet.apa.org/doi/10.1037/0033-2909.112.3.527

수집, 측정, 궁극적으로는 테스트 품질과 형평성에 영향을 미친다. AI를 활용한 평가는 자동 채점 등의 장점이 있으나 편향과 같은 위험도 존재하며 평가 결과에 영향을 미칠 수 있으며[107][108][109], 잠재적으로 테스트 형평성을 저해하고 테스트 품질을 떨어뜨릴 수 있다. 따라서 AI 기반 평가는 인간 중심의 AI 가치와 책임 있는 AI 지침 및 표준과의 정렬이 필요하다.[110][111][112]

듀오링고 영어테스트(DET)의 경우 이러한 요소를 반영하기 위해 다양한 시도를 하고 있다. DET는 평가 품질과 형평성을 제고하기 위해 인간이 관여하는 AI(HiTL AI) 방식을 채택하고 있다. 최근 교육 및 평가 정책은 중요한 의사 결정 단계에서 인간의 감시가 중요한 HiTL AI가 논의되고 있다.[113][114] HiTL AI는 또한 인간이 AI 성능을 향상시키기 위해 실시간으로 입력을 제공하는 인간-시스템 상호작용 과정으로 논의된다.[115] DET는 시험 설계, 측정 및 보안을 위해 자동으로 생성된 콘텐츠를 검토하고, 모델을 구축하고 평가하기 위한 훈련 데이터를 라벨링하며, 테스트 보안을 위해 AI가 감지한 표절 플래그를 판별하는 등 인간의 판단과 감독을 활용한다.

공정성 측면에서 대표성 있는 데이터 사용을 위해 듀오링고는 AI 모델을 훈련할 때 다양한 언어 배경과 성별을 가진 수험자들의 데이터를 고르게 포함하려고 노력했다. 예를

107) Belzak, W. C., Naismith, B., & Burstein, J. (2023, June). Ensuring fairness of human-and AI-generated test items. In International Conference on Artificial Intelligence in Education (pp. 701-707). Cham: Springer Nature Switzerland.
108) Johnson, M. S., Liu, X., & McCaffrey, D. F. (2022). Psychometric methods to evaluate measurement and algorithmic bias in automated scoring. Journal of Educational Measurement, 59(3), 338-361
109) The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2017). Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems (Version 2). IEEE. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf
110) Von Davier, A. & Burstein, J. (in press). AI in the Assessment Ecosystem: Implications for Fairness, Bias, and Equity. In Artificial Intelligence in Education: The Intersection of Technology and Pedagogy, Springer Nature Switzerland
111) Burstein, J. (2023). Responsible AI standards. Duolingo English Test.
112) Auernhammer, J. (2020). Human-centered AI: The role of human-centered design research in the development of AI. In S. Boess, M. Cheung, & R. Cain (Eds.), Synergy - DRS international conference 2020 (pp. 1315-1333). https://doi.org/10.21606/drs.2020.282
113) Association of Test Publishers. (2024). Creating Responsible and Ethical AI Policies for Assessment Organizations, July 12, 2024.
114) U.S. Department of Education, Office of Educational Technology (2024). Designing for Education with Artificial Intelligence: An Essential Guide for Developers, Washington, D.C
115) Wang, G. (2019, October 20). Humans in the Loop: The Design of Interactive AI Systems:https://hai.stanford.edu/news/humans-loop-design-interactive-ai-systems.

들어, 영어가 모국어가 아닌 여러 언어 사용자들이 포함된 데이터를 사용하여 모델을 훈련함으로써 특정 언어 배경을 가진 사람들이 불리해지지 않도록 했다. 차별 가능성 평가를 위해 듀오링고는 자동으로 생성된 항목이 특정 그룹에게 불리하지 않도록 "차별적 항목 기능(Differential Item Functioning, DIF)" 분석을 수행했다. 이 분석을 통해 항목이 특정 성별이나 언어 배경을 가진 사람들에게만 유리하거나 불리한지 확인하고, 문제가 있을 경우 해당 항목을 수정한다.

듀오링고는 테스트가 실제로 수험자의 영어 실력을 정확하게 평가할 수 있도록 여러 품질 관리 프로세스를 도입했다. 자동 채점의 정확성을 높이기 위해 DET는 수험자가 작성한 글이나 발화를 AI를 사용해 자동으로 채점한다. 이 과정에서 AI가 사용하는 기준이 인간 평가자의 기준과 일치하도록 엄격히 훈련되는데 이를 위해 인간 평가자들이 수많은 응답을 채점하여 AI 모델을 훈련시키고, AI가 실제 테스트에서 인간처럼 정확하게 채점할 수 있도록 만든다. 설명 가능한 AI를 위해 DET는 AI 모델이 어떻게 채점을 했는지 설명할 수 있는 기능을 포함하고 있다. 즉, AI가 수험자의 특정 문법 사용, 어휘 선택 등을 기준으로 어떻게 점수를 매겼는지 명확하게 보여줄 수 있다. 이를 통해 수험자가 자신의 점수를 이해하고 신뢰할 수 있도록 지원한다. 생성형 AI 콘텐츠 검토 측면에서 DET는 GPT-4와 같은 대형 언어 모델을 사용해 테스트 항목을 자동으로 생성한다. 그러나 이러한 자동 생성 항목은 항상 인간이 검토하여 품질을 보장하며 생성된 항목이 문법적으로나 의미적으로 적절하고, 테스트의 목표에 부합하는지를 확인한 후 사용한다.

듀오링고는 테스트의 안전성을 보장하기 위해 부정행위 방지와 데이터 보안을 중요한 우선순위로 두고 있다. 부정행위 방지를 위해 DET는 수험자가 부정행위를 하는 것을 막기 위해 AI 기반의 보안 시스템을 운영한다. 예를 들어, 수험자가 인터넷에서 복사한 내용이나 이전 테스트에서 사용한 답변을 사용하면 AI가 이를 감지하고 이후, 인간 감독자가 AI의 경고를 검토하여 부정행위가 실제로 발생했는지 확인한다. 이렇게 AI와 인간의 협업을 통해 보다 보안을 유지하기 위해 노력한다. 데이터 프라이버시와 보안관련, DET는 수험자의 개인정보와 데이터가 안전하게 보호되도록 보안 프로토콜을 준수하며 수험자의 데이터는 법률에 따라 수집되고, 테스트 중에 제공된 모든 정보는 암호화되며, 승인되지 않은 접근으로부터 보호된다. 실시간 감독 관련 DET는 테스트 중 수험자가 올바른 방식으로 시험을 보는지 확인하기 위해 실시간으로 수험자를 감독하는 시스템도 운영하고

있다. 이 시스템은 AI와 인간이 협력하여 수험자의 행동을 모니터링하고, 의심스러운 행동이 있는 경우 이를 신속하게 처리한다.[116]

DET는 책임감있는 AI 표준을 만들었는데 기존에 논의된 AI 윤리원칙, 도메인에 국한되지 않는 AI 거버넌스, 표준, 지침을 검토했다. 또한, 듀오링고 내의 계산 심리 측정, 언어 평가, 법률, 기계 학습, 보안 분야의 전문가들과 외부 전문가와 협의 사항을 반영했다. 만들어진 표준은 문서로 발행되었으며 대중의 의견을 받을 수 있도록 공개했다. 또한, NIST의 AI 위험 체계와 연결하여 신뢰성을 검증하였다.

여섯째, 투명한 절차 이슈 관련, 교육 분야 AI 윤리원칙에서는 교육공동체의 연대와 협력을 중시여기며 교육 분야 AI는 그 활용에 있어 민·관·학·연의 협력을 지향하고 지속 가능한 교육생태계를 구축할 수 있도록 제공되어야 한다. 교육 분야 AI의 개발과 활용, 평가는 정부의 일방적인 주도보다 다양한 분야 이해당사자들의 자발적이고 적극적인 관심과 참여로 이루어질 필요가 있고 교육적 가치를 저해하지 않는 선에서 민·관·학·연 등 다양한 관계자가 협력하고 함께 도전할 수 있는 기반이 마련되어야 한다고 언급하고 있다.[117] 노스 캐놀라이나 교육청은 생성형 AI 교육 실행 및 권고에 관한 사항을 지속 업데이트 하며 논의사항을 공개하고 있으며 2024년에 9번의 업데이트가 이루어졌다.[118]

116) Jill Burstein, Geoffrey T. LaFlair, Kevin Yancey, Alina A. von Davier, Ravit Dotan(2024) "Responsible AI for Test Equity and Quality: The Duolingo English Test as a Case Study", https://arxiv.org/abs/2409.07476
117) 교육부(2022.8), "사람의 성장을 지원하는 교육분야 AI 윤리원칙"
118) North Carolina Public Instruction(2024) "North Carolina Generative AI Implementation Recommendations and Considerations for PK-13 Public Schools"

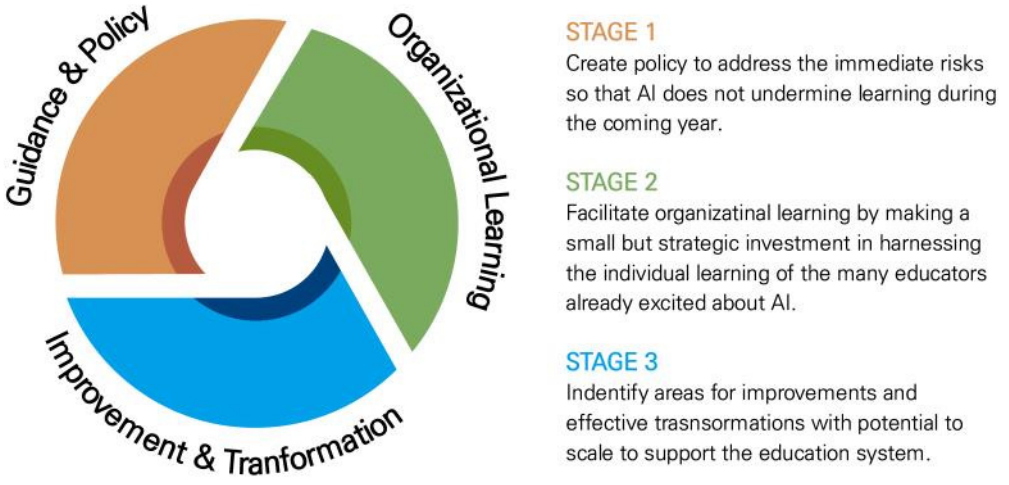
[표 5-8] 노스 캐놀라이나 교육청의 AI 교육실행 및 권고 업데이트

날짜	버전	변경 사항 및 업데이트
1/16/2024	24.01	• 최초 발행
1/28/2024	24.02	• 경미한 기술적 수정/교정 • 버전 기록 2페이지 추가 • 20페이지에 예시 프롬프트 추가 • 17페이지 AI 문해력 권장사항 학년별로 재구성, 새로운 권장사항 추가 • 33페이지에 새로운 리소스 추가
2/18/2024	24.03	• 기술적 수정
3/4/2024	24.04	• 19페이지에 학생용 AI 도구 추가
3/5/2024	24.05	• 29페이지에 AI 저항적, AI 보조, AI 협력 과제 추가
3/20/2024	24.06	• 학생용 AI 도구 섹션 업데이트 • 19페이지의 학생용 AI 도구 그래픽 제거 (업데이트용)
4/11/2024	24.07	• 새로운 학생 대면 내장 교육 모델을 고려 • 19페이지의 학생용 도구에 대한 추가 지침 제공
4/28/2024	24.08	• 새로운 리소스 추가 - 신규 교육 AI 정책, 2024 국가 교육공학 계획 • 25페이지 AI 평가 척도를 'AI에서 학생 통합: 0에서 무한대'로 업데이트
8/5/2024	24.09	• 11페이지에 편집 가능한 조달/평가 도구 링크 추가 • 17페이지의 학년별 학생용 AI 사용에 대한 지침 명확화 • 26페이지에 0에서 무한대 GPI 링크 추가 • 32페이지의 최근 개발 사항에 기반한 AI 접근성 관련 업데이트 • 33페이지에 딥페이크와 사이버불링에 대한 지침 추가

자료: North Carolina Public Instruction(2024) "North Carolina Generative AI Implementation Recommendations and Considerations for PK-13 Public Schools"

AI 교육정책 개발은 일회성 이벤트가 아닌 순환적 과정이며 정책 개발, 전문성 개발, 조직 학습이 서로 연결되어 환류한다. AI의 교육 활용에 있어 위험 대응 정책을 수립하고, 이에 실제 대응하기 위한 조직 학습을 촉진시켜야 하며 개선영역을 식별하고 대처방안을 강구하는 순환절차가 유기적으로 연결될 필요가 있다. 이를 위해서는 다양한 이해관계자(정부, 교사, 학생, 학부모, 학교 관계자, 법무, 교육과정 전문가, IT 전문가 등)로 구성된 가이드팀이 AI 도입을 주도해야 한다.[119]

119) WDE(2024), "Guidance for Wyoming School Districts on Developing Artifcial Intelligence Use Policy", https://edu.wyoming.gov



[그림 5-20] AI 교육정책 개발의 순환 절차

자료: WDE(2024), "Guidance for Wyoming School Districts on Developing Artifcial Intelligence Use Policy", https://edu.wyoming.gov

일곱째, 제도 개선 이슈이다. 인터넷 혁명의 가속화로 2004년 이러닝 산업 발전 및 이러닝 활용 촉진에 관한 법률이 제정된 것처럼, AI 혁명 시대에 미래 교육을 위한 AI 교육법에 대한 논의도 필요하다. 교육 분야의 AI 도입은 백년지대계(百年之大計)의 관점에서 중대한 사안이며 관련 법에서 중장기 계획 수립, 법적 근거, AI 디지털교과서의 종류[120], 실태조사, 학습효과 측정, 재원 관리, 격차 해소, AI 리터러시(Literacy), 입시와의 연계, 개인정보보호, 학습 데이터 관리, 협의체 구성과 운영 등 다양한 세부 사항을 논의하고 개선 이슈를 도출해 가는 방안을 검토해 볼 필요가 있다.

마지막으로 미래 기회와 위험 이슈이다.[121] 주목할 미래 이슈로 AI 교육으로 야기될 다차원적 격차에 대한 열린 논의가 필요하다. AI 기술의 교육 현장 도입은 다양한 형태의 교육격차 문제를 제기할 수도 있다. 이러한 새로운 교육 불평등은 단순한 기술 접근성의 차이를 넘어, 다차원적이고 복합적인 양상으로 나타날 수 있는데 먼저, 하드웨어와 소프트웨어 측면의 접근성 격차가 기본적인 불평등 요인으로 작용할 수 있다. 고성능 AI 학습 기기의 보유 여부, 안정적인 네트워크 환경, 프리미엄 AI 교육 프로그램 구독 가능성 등이

120) 국정교과서, 검정교과서 등의 종류 구분
121) 본 이슈는 FGI에서 논의되었던 사항을 정리하여 작성하였다.

학습 기회의 질적 차이를 만들어낼 수 있다. AI 활용 능력의 격차는 더욱 심각한 불평등 요인이 될 수 있다. 디지털 리터러시, 데이터 해석 능력, AI와의 효과적인 상호작용 능력 등의 차이는 단순한 학습 성취도 차이를 넘어 미래 역량의 근본적인 격차로 이어질 가능성이 있다. 이는 자기 주도 학습 능력과 메타 인지 능력의 차이로 확대되어, 장기적으로는 직업 선택과 사회적 지위에까지 영향을 미칠 수 있을 것이다. 가정 배경에 따른 격차도 전망해 볼 수 있다. 부모의 AI 교육에 대한 이해도, 자녀 학습지원 능력, 교육비 투자 가능성 등이 새로운 문화 자본의 형태로 작용하여, 교육 기회의 질적 차이를 만들어낼 가능성도 있다. 특히 디지털 문화 자본의 차이는 온라인 학습 커뮤니티 참여도와 네트워크 형성에도 영향을 미쳐, 장기적인 교육적 이점의 차이로 이어질 것이다. 지역 간 격차 또한 새로운 차원에서 논의될 수 있다. 도시와 농촌 간의 AI 인프라, AI 교육 시설의 수준 차이, 전문 교사 확보의 어려움 등이 지역별 교육 기회의 질적 차이로 이어질 우려도 있다. 이는 단순한 학력 격차를 넘어 미래 직업 준비도와 혁신 역량의 차이로 이어질 것이다.

AI 혁명 시대에 프리미엄 아날로그 교육의 부상도 미래 이슈로 논의해 볼 수 있다. AI 디지털교과서가 보편화되는 미래 교육환경에서는 인간 중심의 아날로그 교육이 프리미엄 교육 시장의 새로운 패러다임으로 부상할 가능성도 제기된다. 이는 디지털 교육의 보편화로 인한 반작용이자, 전인적 교육에 대한 수요 증가를 반영하는 현상이다. 프리미엄 아날로그 교육은 고도로 개인화된 맞춤형 교육과정을 특징으로 하며, 소규모 그룹 단위의 집중 교육을 통해 학습자의 전인적 발달을 도모하는 것이 핵심이다. 특히 직접적인 체험과 실물 기반 학습, 인문학적 소양 함양에 중점을 두어 디지털 교육이 제공하기 어려운 교육적 가치를 실현하는 데 주력할 것이다. 이러한 교육 형태는 주로 고소득층과 교육 의식이 높은 가정을 중심으로 확대될 수 있으며, 프리미엄 대안학교, 특화된 방과 후 프로그램, 몰입형 방학 프로그램 등 다양한 형태로 구현될 수도 있다. 이는 교육의 질적 다양성 확보라는 긍정적 측면과 함께, 교육격차 심화라는 사회적 과제를 동반할 수도 있다. 프리미엄 아날로그 교육의 확산은 교육 시장의 양극화를 심화시킬 수 있는 우려가 있으나, 동시에 인간 중심 교육의 가치를 재조명하고 교육의 본질에 대한 사회적 담론을 활성화하는 계기가 될 수도 있다. 이 외에도 AI로 인한 교육의 초 국가화와 교육 경계 붕괴 가능성, 글로벌 AI 교육 플랫폼의 지배력 강화, AI 없이는 학습 불가능한 세대 출현, AI 의존 불안장애 등 향후 제기될 수 있는 다양한 이슈를 논의해 나갈 필요가 있다.

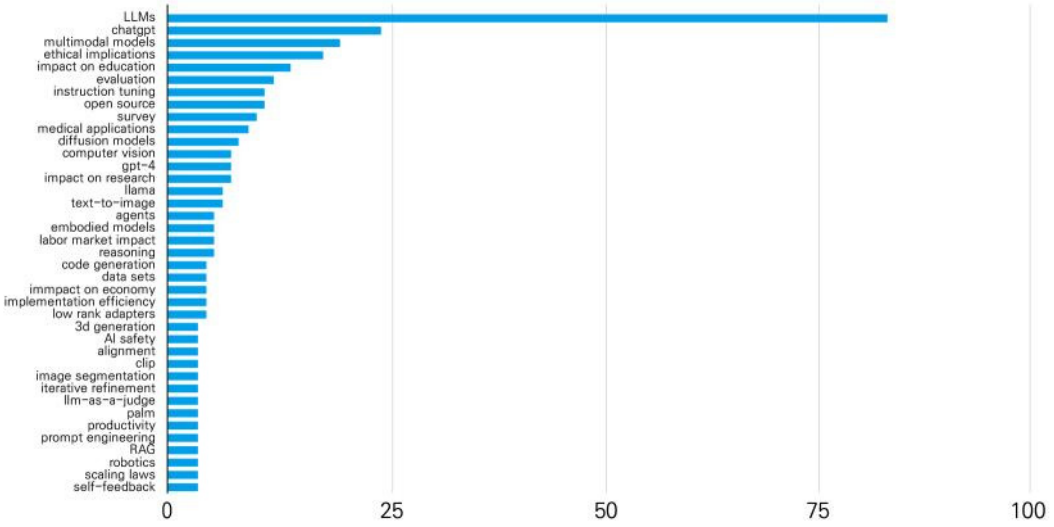
제5절 시사점

AI 디지털교과서에 대한 기대와 우려가 공존하는 현시점에서 미래 교육 혁신 관점에서 고려해야 할 사항을 제시하면 다음과 같다. 먼저, 교육에 적용될 AI에 대한 인식의 합의가 요구된다. AI를 다양한 관점에서 분류하고 AI 디지털교과서가 세분류 기준 어떤 영역에 존재하는지, 융합하고 있는지 공개된 상황에서 오해가 없도록 논의를 진행해야 한다. 최근 무료로 개방되며 성능이 매우 높은 LLM 기반의 모델이 다수 출현하며 교육 분야에도 다양하게 활용되고 있으며 학습에 사용되는 AI 모델이 전통적인 기계 학습 중심인지, 딥러닝, 초거대 AI 기반의 생성형 AI 모델인지 이들 모델의 혼합인지 인식의 합의가 필요하다.

둘째, AI 기술 진화 속도와 다양성을 동태적 관점에서 고려해야 한다. AI는 긴 시간 축적되어 온 기술로 다양한 AI 요소기술이 인간의 수준을 넘어서는 성과를 보이며 진화하고 있다. 논쟁의 대상이 되는 AI 교육사례는 시차 관점에서 입체적인 해석이 필요하다. 이에 과거의 AI 기반의 디지털 교육이 실패로 끝나거나, 효과가 없다고 판명된 사례는 현재의 AI 관점에서 성능과 상호작용 측면에서 재해석이 필요하다. 다양성 측면에서도 전 세계 다양한 교육 주체와 이해관계자들이 생성형 AI 기반 교육에 대해 다양한 권고사항을 발표하고 있는바, 하나의 정답보다 다양한 가능성을 열어두고 해당 권고와 사례를 면밀하게 분석하며 한국의 AI 디지털교과서에 맞는 방향을 찾아가야 할 것이다.

셋째, AI 디지털교과서의 학습효과 측정이 고려되고 체계적으로 관리되어야 한다. AI를 활용한 교육성과 연구는 다양한 측면에서 활발히 이루어지고 있으며 그 결과도 AI를 적용하는 방식과 교사, 학생, AI 성능과 최적화 상황에 따라 다르다. AI는 교육적 성과를 높일 수도 있지만 잘못 활용하면 오히려 성과가 낮아질 수도 있음을 유의해야 한다. 2023년 AI 연구 중 최다 인용을 받은 논문 중 주제별 순위 4위가 AI 교육성과 측정 연구이다. AI 연구의 대부분은 LLM 모델 개발, 멀티모달이 주류를 이루고 있으나 이러한 상황에서 AI 교육성과 연구가 4위를 차지한 것은 그만큼 AI 교육에 관심이 많고 다양한 연구가 이루어지고 있다는 것이다. 이에 AI 디지털교과서 기반의 학습효과 측정 연구를 활

성화하고 이를 체계적으로 관리하여 지속적인 성과가 유지될 수 있도록 해야 할 것이다.



[그림 5-21] 2023년 AI 분야 최다 인용 100대 연구 구분

자료: https://www.zeta-alpha.com; Analyzing the homerun year for LLMs: the top-100 most cited AI papers in 2023, with all medals for open models.

넷째, 평가의 공정성과 품질, 안전에 대한 실질적인 조치 방안이 강구되어야 한다. AI 교육의 공정성과 안정성을 확보할 수 있는 실질적인 기술 조치 방안이 공개되고 이를 기반으로 다양한 논의와 개선방안이 모색되어야 한다. 듀오링고의 사례처럼 실제 다양한 국적의 학생들이 시험을 통해 공정한 성과를 나오는지 판별 테스트를 시행하고, 교사와 AI 전문가가 함께 문제를 해결해 나가는 '휴먼 인 더 루프(Human in the Loop)' 방식이 교육의 세부 활동에 다양하게 적용되고 공유될 필요가 있다.

다섯째, 디지로그(Digilog) 관점이 함께 논의되어야 한다. 디지로그(Digilog)란 디지털(Digital)과 아날로그(Analog)의 합성어이며, 비트(bit) 0과 1로만 구성된 차가운 디지털에 인간중심, 따뜻함과 같은 아날로그 요소가 융합되는 것을 의미한다.[122] 현재 한국 학생의 기기별, 상황별 디지털 활용 수준, 중독 등 다양한 지표에 기반하여 AI 디지털교과서의 활용방안을 기획해야 한다. AI 디지털교과서의 활용은 반드시 모든 과정이 화면 중

122) 2006년 1월 이어령 전 문화관광부 장관이 인터넷의 어두운 면을 극복하고 앞으로 다가올 후기 정보사회의 밝은 미래를 모색하자는 취지에서 디지로그론을 주창한 바 있다.

심의 온라인으로만 진행되는 것은 아니며 AI 활용결과에 기반한 토론, 발표, AI 한계점에 대한 논의 등 다양한 관점에서 오프라인과 결합 될 수 있다.

여섯째, 투명한 절차 진행과 제도개선이다. AI 디지털교과서 운영에 다양한 이해관계자들이 참여하는 정례 협의체를 운영하고, 논의의 결과와 이행 사항을 투명하게 공개하는 과정이 필요하다. AI 디지털교과서에 관련된 다양한 논의, 분석, 토론을 기반으로 해당 사항을 공유하고 공론화를 통해 논의를 진전시킬 필요가 있다. 이런 측면에서 노스 캐놀라이나 교육청의 AI 교육실행 및 권고 업데이트는 매우 의미가 있다. 작은 논의도 놓치지 않고 경과를 기록하고 공유하며 권고안을 계속 업데이트하는 것은 AI 교육으로 생길 수 있는 갈등 해소에 도움이 된다. 제도 관련, 21대 국회에서도 AI 교육에 대한 법안이 발의되었으나 22대 국회가 시작되며 폐기되었다. AI 디지털교과서 추진에 있어 필요한 중장기 계획, 재원 마련, 학습효과 측정 등 필요한 제반 사항을 포함한 법안을 추진하여 다양한 문제를 함께 논의하는 방안도 고려될 필요가 있다.

일곱째, 한국적 상황에 맞는 AI 디지털교과서 도입방안을 모색해야 한다. 한국의 현실에서 입시제도는 교육과 떨어질 수 없으며 입시제도의 방향이 AI 디지털교과서의 도입과 어떤 연관성이 있으며 연계되는지에 대한 논의가 더해져야 한다. AI 디지털교과서 도입 시에도 해외 사례에 기반한 일방적인 수용이 아니라 국내 학생과 교사의 상황을 고려하고 이에 더해 지역적인 차이 등 다양한 요소가 함께 고려되어야 한다. 소버린 AI 관점에서 한국교육에 과목별 특화 AI 모델 개발도 요구되며 이를 위한 AI 교육 연구개발에 관한 지원도 강화될 필요가 있다.

마지막으로 데이터 기반의 AI 디지털교과서 관리와 다양한 격차 요소를 고려하고 미래 이슈를 함께 고민할 필요가 있다. AI 디지털교과서의 도입으로 사교육비 절감 효과가 기대되고 있으며 이에 실제 이러한 효과를 가져오고 있는지 데이터 기반의 정책효과 검증이 필요하다. 또한, AI 교육으로 우려되는 다차원적 격차도 데이터 기반의 측정과 분석으로 관리되는 방안이 논의되어야 한다. 격차 이슈는 학생, 교사, 지역, 학교, 국내 학생과 외국인 학생 등 다차원으로 야기될 수 있다. 또한, 향후 AI 교육의 확산으로 프리미엄 오프라인 교육의 부상, AI 의존성 장애, AI 교육의 초 국가화로 인한 교육 경계 붕괴, 글로벌 AI 교육 플랫폼지배력 강화 등 새로운 이슈가 제기될 가능성이 존재하며 관련한 기회와 위험요인을 모니터링하고 대응해 나갈 필요가 있다.

정책입안자와 교육자는 AI와 교육의 상호작용이 만들어 나갈 미래에 근본적인 의문을 불러일으키는 미지의 영역에 진입했으며[123], 교육과 AI 통합을 통해 미래 교육을 혁신해 나갈 미래 청사진을 그려 나가야 할 시점이다.

123) UNESCO(2021), "AI and education: guidance for policy-makers"

제6장

결론

제1절 연구요약 및 시사점

본 연구의 목적은 AI 혁명 시대에 제기되는 정책 이슈를 분석하고 이에 따른 미래 정책 방향을 제안하는 것이다. AI 관련 다양한 세부 주제를 다루기 위해 선행되어야 할 기본 개념과 생태계 분석을 시작으로 국회와 미래차원에서 현재 이슈가 제기되며 중장기 측면에서 검토해야 할 이슈인 AI 정책변화의 특징 분석, 사회경제적 이슈를 제기 중인 딥페이크, 세계 최초로 도입되는 AI 디지털교과서를 중점적으로 다루었다.

본 연구의 결과를 요약하면 다음과 같다.

먼저, AI 정책변화의 특징을 분석한 결과 AI 정책의 무게 중심이 진흥에서 규제와의 조화로 이동하고 안전과 신뢰에 관한 정책이 강조되고 있다. 또한, 주요국들은 변화하는 AI 환경과 자국의 상황을 고려하여 차별화된 방식의 AI 정책을 수립하고 이행하고 있다. 이에 국내 AI 정책 수립 시 변화하는 AI 환경을 반영할 필요가 있으며 계류 중인 AI 법 도입 논의에 있어 변화하는 AI 환경과 정책 특성을 종합적으로 고려해야 한다. AI 안전과 신뢰성 관련 글로벌 정책 동향을 검토하고 국내 상황에 맞추어 반영하는 방안을 검토해야 한다. 21대 국회에서 폐기된 AI 법이 안전과 신뢰성 측면에서 문제가 제기되었던바 EU 등 다양한 AI 시스템의 위험분류 체계를 검토하고 한국의 현실에 맞게 반영해야 한다. 또한, 국가별 초거대 AI 경쟁력을 고려한 AI 정책을 수립해야 한다. 초거대 AI 모델 보유 수가 상대적으로 적은 EU의 경우 AI 규제 강도가 높고, 영국의 경우 규제 친화적인 접근으로 정책을 수립하고 있으며 AI 규제 강도와 범위 설정 시, 한국의 경쟁력을 종합적으로 고려해야 한다. 국내 AI 안전연구소 설립 및 운영 관련 글로벌 사례를 참고하고 글로벌 정책 공조 체계를 강화할 필요가 있으며 중앙정부와 지자체 AI 정책의 유기적 연계 방안이 모색되어야 한다. AI 기술이 빠르게 진화하기 때문에 동태적 관점에서 AI 생태계를 관찰하고 미래를 준비해야 한다. 대형언어모델의 규제는 향후 작은 모델이 강력한 성능을 내는 방향으로 진화하고 있고 이에 비례하는 정책 방안이 고려되어야 한다. AI가 기존 서비스의 보완, 효율성 제고를 넘어 완전히 새로운 혁신 비즈니스 모델을 만들고 있어 이로 인한 파급효과를 선제적으로 검토해야 한다. 미래에 앱이 사라지고 AI 에이전트 기반의 서비

스가 늘어난다면 기존의 플랫폼 정책에도 변화가 불가피할 것으로 전망된다. LLM(Large Language Model)이 LWM(Large World Model)로 진화하고 있는바[124] 현재 2D 중심의 AI가 공간컴퓨팅과 연계된 3D 시뮬레이션으로 연동되면 기존에 없던 새로운 혁신과 위험이 발생할 가능성도 존재한다. 마지막으로 AI 위험에 대한 이해관계자의 견해가 다르며 이에 갈등관리 체계를 마련하고 절차를 투명하게 공개해 숙의의 과정을 통해 정책 방안을 논의해야 한다.

둘째, AI 시대에 진화하는 딥페이크 이슈를 검토한 결과 딥페이크의 대상 범위가 연예, 정치 분야에서 개인, 기업, 투자 등으로 다변화되고 있다. 딥페이크로 인한 기업 사기 등 리스크가 증가하여 이에 대한 대비도 필요하다. 이에 딥페이크의 위험을 줄이고, 올바른 활용을 위해 기술, 사회·제도적 측면에서의 노력이 필요하다. 기술적 측면에서 딥페이크 탐지 기술의 고도화가 요구되는 시점이다. 딥페이크 탐지 기술 개발 강화 및 활용을 통해 대처할 필요가 있다. 사회적으로는 AI 리터러시 교육, 딥페이크 피해사례 전파 등을 통해 구성원들이 딥페이크로 인한 위험을 인지하고 대처할 수 있도록 지원해야 한다. AI 교육, 딥페이크 악용 사례 전파를 통해 개인이 스스로 정보를 판단하고 검증하는 능력을 키울 수 있도록 지원을 강화하는 방안을 검토해 볼 수 있다. 딥페이크로 인한 피해를 줄이기 위해 제도 개선 이슈를 검토해야 한다. 현재 딥페이크 기술에 대한 규제는, 성착취물 제작이나 선거운동 활용 금지 중심으로 이루어지고 있어 딥페이크를 악용한 사기에 대한 별도의 법률이나 진위 확인, 출처 표시 등 관련 정책을 마련하는 방안도 검토할 필요가 있다. 딥페이크 산업 활성화를 고려한 정책조합 구상(Policy Mix) 및 모니터링을 강화하고 가짜뉴스, 사기 등 국민 피해 최소화에 주력해야 한다. 기술조치, 자율규제, 가이드라인, 입법규제 등 다양한 규제 강도 수준을 고려해야 한다. 주요국에서는 딥페이크 관련 사기방지, 출처 표시 등 정책 논의가 활발히 이루어지고 있다. 규제뿐만 아니라 딥페이크의 산업적 가치에도 주목할 필요가 있다. 기업은 AI 트윈 가수, 과거 재현 등 딥페이크를 활용한 생산성 증대, 다양한 혁신 사업모델 발굴 방안을 모색해야 한다. 정부는 기존 벤처창업 활성화, 크리에이터 양성 과정, 1인 창조기업 육성 등의 지원 사업을 AI 시대를 주도할 슈퍼개인 양성 관점에서 보완하는 방안도 검토해 볼 수 있다.

셋째, AI 디지털교과서에 대한 기대와 우려가 공존하는 현시점에서 미래 교육혁신 관

124) Forbes(Jan 23, 2024) “The Next Leap In AI: From Large Language Models To Large World Models?”

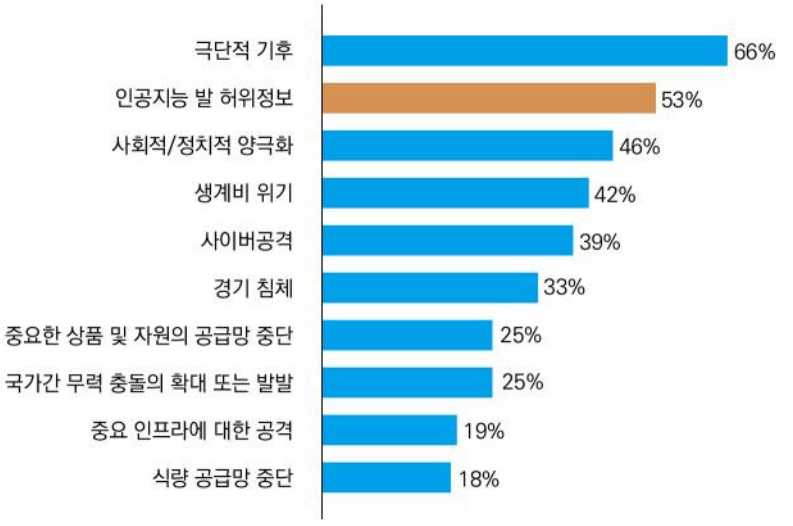
점에서 고려해야 할 사항을 제시하였다. 먼저, 교육에 적용될 AI에 대한 인식의 합의가 요구된다. 초거대 AI 및 생성형 AI는 성능과 적용 방식에 있어 현저한 차이가 존재하며 현재 AI 디지털교과서를 둘러싼 교육 주체들과 다수의 이해관계자가 AI 디지털교과서에 적용되는 AI에 대해 혼재된 개념을 사용하고 있어 이에 어떠한 범주에 있는지 함께 인식하고 논의를 전개 시켜나갈 필요가 있다. AI 기술 진화 속도를 동태적 관점에서 고려해야 한다. AI는 긴 시간 축적되어 온 기술로 다양한 AI 요소기술이 인간의 수준을 넘어서는 성과를 보이며 진화하고 있다. 2010년, 2015년 그리고 현재의 AI는 완전히 다른 수준에 있으며 이는 AI 기반 교육사례, 효과를 분석할 때도 동태적 관점에서 입체적으로 고려해야 함을 시사한다. AI 적용 방식에 대한 다양성도 고려해야 한다. AI의 적용은 목표 분류, 구현 대상과 과목, 효과, 투입비용과 생산성 등에 따라 머신러닝 방식부터 초거대 AI까지 다양한 조합으로 설계될 수 있다. 또한, AI 학습효과 측정이 고려되어야 한다. 세분류 기준 AI가 어떻게 적용·활용되는지 그리고 실제 효과가 있는지 측정·관리되어야 한다. 교육 분야 AI 윤리원칙에서도 이 부분을 강조하고 있다. 디지로그(Digilog) 관점이 함께 논의되어야 한다. 실제 AI 교육은 온라인을 기반으로 AI가 제시하는 문제를 풀고 수준을 높여가는 협소한 개념을 넘어 프롬프트 기반의 다양한 상호작용을 통해 새로운 아이디어를 만들고 오프라인과 융합하도록 구성이 가능하다. AI 디지털교과서 운영에 있어 다양한 이해관계자들이 참여하는 정례 협의체를 운영하고, 논의의 결과와 이행 사항을 투명하게 공개하는 과정과 AI 혁명 시대에 미래 교육을 위한 AI 교육법에 대한 논의도 필요하다.

AI 시대가 개화하면서 다양한 정책 이슈가 제기되고 있으며 본 연구에서는 최근 논의의 중심에 있는 AI 법제도, 딥페이크, AI 디지털교과서 이슈를 검토하였다. AI 정책 전반을 조망하는 AI 법제도, 그리고 세부 이슈로 논의되는 딥페이크와 AI 디지털교과서 논의는 현재 AI 정책이 거시적, 미시적 접근이 동시에 필요하다는 의미를 시사한다. 전체와 부분을 함께 연계하여 정책을 고민해야 하며, 이러한 변화는 기존의 디지털 변화보다 빠르게 진전되고 있다는 측면에서 동태적인 모니터링이 필요하다. AI로 야기될 전례 없는 변화의 순간에 직면했으며 이는 생태계 내 다양한 이해관계자들 간의 갈등을 야기하고, 새로운 비즈니스 모델의 출현으로 경쟁 구도가 변화할 수 있는바 미래지향적 변화에 발맞춰 정책 시차를 줄이는 방안을 모색해야 한다.

제2절 후속 연구 방향

NATIONAL ASSEMBLY FUTURES INSTITUTE

범용 기술인 AI는 사회경제적 측면에서 파괴적 변화를 일으키고 있다. 변화는 현재 진행되며 향후 미래에 더욱 문제가 확대될 수 있는 이슈와 현시 점에서 전혀 예측하기 어려운 잠재된 위험도 존재한다. 이에 후속 연구에서는 중장기적으로 제기될 다양한 AI 리스크를 전망하고 대응 방안을 모색하는 연구가 필요하다. WEF(2024)는 현재 당면한 글로벌 리스크 2위로 AI로 야기될 허위 정보를 제시하였다. 이는 AI가 현재 당면한 이슈임을 의미하며 실제 딥페이크 등 다양한 허위 정보가 문제로 부상했다.



[그림 6-1] 2024년 발생가능한 글로벌 주요 리스크 요인

자료: World Economic Forum(2024), “Global Risks Report 2024”

중요한 점은 AI 리스크가 단기에 끝나지 않고 중장기적으로 유효하다는 것이다. WEF(2024)는 향후 2년, 10년 내 직면할 리스크에도 여전히 AI를 중요한 이슈로 전망하고 있다. 이에 AI 위험을 세분화하고 새롭게 제기될 위험을 예측하는 연구가 필요하다.

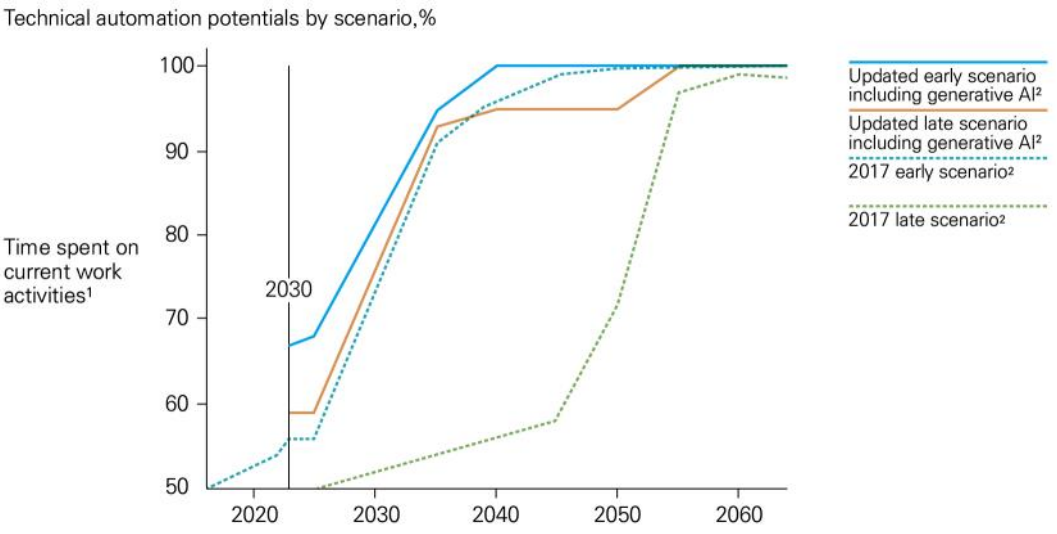
〈 향후 2년 내 직면할 글로벌 위기 요인 〉	
1	역정보 및 허위정보
2	기상 이변
3	사회 양극화
4	사이버 위협
5	국가 간 무력 충돌
6	경제적 기회 불평등
7	인플레이션
8	비자발적 이주
9	경기 침체
10	오염

〈 향후 10년 내 직면할 글로벌 위기 요인 〉	
1	기상 이변
2	급격한 지구 시스템 변화
3	생물다양성 손실 및 생태계 붕괴
4	천연자원 부족
5	역정보 및 허위정보
6	AI 기술의 부작용
7	비자발적 이주
8	사이버 위협
9	사회 양극화
10	오염

[그림 6-2] 향후 2년 및 10년 내 직면할 글로벌 리스크

자료: World Economic Forum(2024), "Global Risks Report 2024"

AI로 야기될 다양한 이슈에 대한 시나리오 전망과 그에 따른 정책 방안 연구도 필요하다. Mckinsey&Company(2023)는 AI의 빠른 진화로 야기될 자동화 수준을 시나리오 기반으로 예측했다. 2017년에는 예측한 AI의 자동화 수준과 2023년 수준은 현저히 다르며 예상했던 미래가 훨씬 더 빠르게 다가올 것으로 전망하고 있다. AI 기술 진화의 속도에는 투자, 정책 등 다양한 요인이 영향을 미칠 수 있으며 생태계 환경은 매우 가변적이다. 이에 AI 진화 속도를 다양한 요인을 활용하여 예측하고 이에 기반한 다양한 시나리오 전개, 그에 따른 정책 방안을 검토하는 연구는 중요한 의미가 있다. 특히 국회의 관점에서 입법과 제도 측면이 이러한 진화 속도에 어떠한 영향을 미칠 수 있는지 연구하는 것은, 미래지향적 입법 체계를 마련하는 기반 자료로 유용하게 활용될 것이다.



[그림 6-3] AI로 인한 자동화 수준 시나리오

자료: Mckinsey&Company(2023) "The economic potential of generative AI"

마지막으로 AI 기술의 진화가 미래 우리의 삶을 어떻게 바꾸어 나갈지를 조망하는 연구가 필요하다. 미래를 준비하기 위해서는 기술의 진화가 어떠한 제품과 서비스의 형태로 우리의 삶을 바꾸어 나갈지 예측하는 연구가 필수적으로 필요하다. 이는 향후 제기될 입법 이슈를 예측하는데 필요한 기반 연구이다. 생성형 AI 기술은 텍스트 중심에서 3D 분야로 진화하고 있으며 최근 이러한 변화는 급진전 중이다.

	텍스트	코드	이미지	영상/3D/게이밍
Pre-2020	스팸 감지, 번역, 기본 Q&A	한 줄 자동 완성		
2020	기본 카피라이팅, 문서 초안 작성	여러 줄 생성		
2022	긴 형태 문서, 2nd Draft 작성	긴 형태, 정확도 개선	아트, 로고, 사진	3D/비디오 모델 초안
2023	과학 논문 등 다양한 분야에 대한 결과물 개선	다양한 언어, 다양한 분야	모형 (제품 디자인, 건축 등)	영상 및 3D 기본 파일/초안 제작
2025	인간 평균 수준의 작업 결과물보다 더 나은 최종 Draft 작성	텍스트로 최종 Draft 생성	최종 Draft (제품 디자인, 건축 등)	2nd Draft 제작
2030	전문 작가의 결과물보다 더 나은 최종 Draft 작성	텍스트로 최종 결과물 생성, 전업 개발자보다 더 나은 결과물 수준	전문적인 아티스트, 디자이너, 사진가보다 더 나은 최종 Draft 수준	개인 맞춤형 비디오 게임 및 영화

[그림 6-4] 생성형 AI 모델 발전 전망

주: 회색은 첫 시도, 파란색은 거의 도달, 붉은색은 완료
자료: Sequoia Capital(2023) "Generative AI: A creative New World", 삼정KPMG

이러한 일련의 변화는 LLM(Large Language Model)에서 LWM(Large World Model)로 진화하고 있으며 이는 공간지능(Spatial Intelligence)의 시대로 우리가 발전해 나가고 있다는 것이다. 현재의 모든 정책 이슈는 공간지능 시대에 새로운 변화를 맞이하게 될 것이고 이에 이를 대비한 미래 연구가 필요하다.

별 첨

별 첨

미래에 AI로 인한 노동 시나리오 세부 사항은 다음과 같다. 2번째 시나리오는 AI 조기 도입으로 작업 환경이 충분히 효율적이지 못한 상황이다.

[그림] 시나리오2

자료: Futures Platform(2024), "Four scenarios on the future of AI in the workplace"

이 경우 가정하는 세부 시나리오2는 다음과 같다.

[표] 시나리오2 세부 내용

Anna와 Sam에게 AI 동료로의 전환은 순탄하지 않았습니다. Sam이 일하는 병원에서는 환자 돌봄을 돕기 위해 도입된 로봇들이 자주 오작동을 일으키며 의료 데이터를 잘못 해석하거나 잘못된 경보를 보냅니다. 갑작스러운 AI 시스템의 유입을 관리하는 방법에 대한 지원이나 훈련이 거의 없는 상황에서, Sam은 환자를 돌보기보다는 기계를 고치는 데 더 많은 시간을 쓰게 되었습니다. Anna의 건설 현장에서도 문제가 심각합니다. AI 기반 크레인들이 안전 위험을 자주 잘못 판단하여 사고와 지연을 초래하고 있습니다. Anna와 그녀의 팀은 AI의 오류를 수정하기 위해 계속 개입해야 하며, 이는 그들을 좌절하게 하고 일정도 뒤처지게 만듭니다. 많은 산업에서 AI의 급속한 도입은 해결책을 제시하기보다는 더 많은 문제를 야기하고 있습니다. 문제의 핵심은 빠른 도입 속도와 규제의 부족에 있습니다. 인력 부족 문제를 해결하고자 했던 기업들이 적절한 감독이나 직원 훈련 없이 AI 시스템을 일상 업무에 투입했기 때문에, 그 결과 AI가 효율성을 높이기보다는 오히려 혼란을 초래하는 작업 환경이 되었습니다. 이러한 상황에서 Sam과 Anna 같은 사람들이 그 여파를 감당하고 있습니다.

자료: Futures Platform(2024), “Four scenarios on the future of AI in the workplace”

3번째 시나리오는 급속히 발전하는 AI 기술을 노동을 대체하며 규제는 준비되지 않은 상황이다.

[그림] 시나리오3

자료: Futures Platform(2024), "Four scenarios on the future of AI in the workplace"

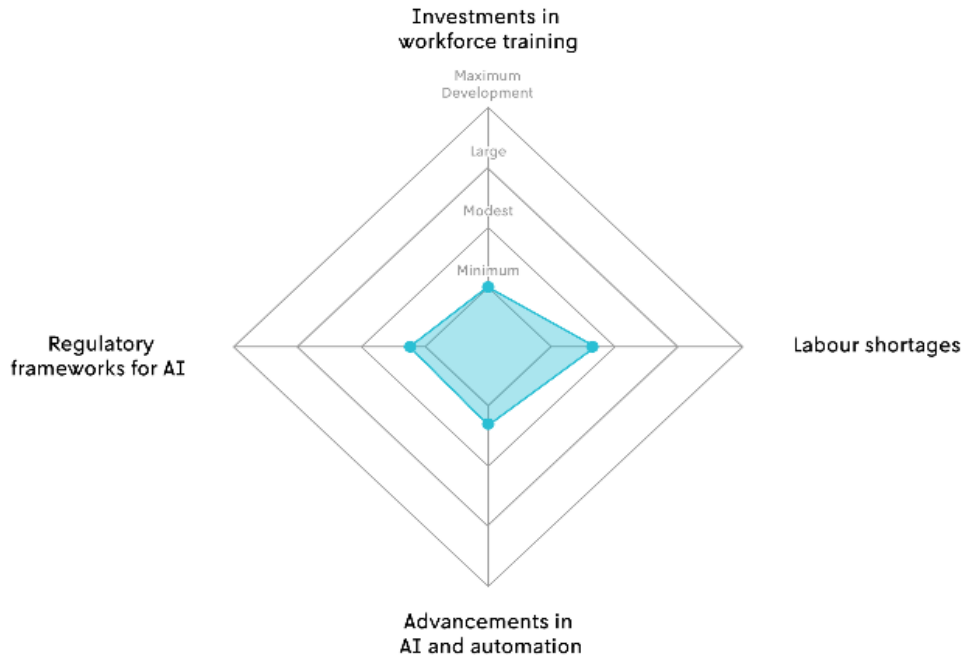
이 경우 가정하는 세부 시나리오3은 다음과 같다.

[표] 시나리오3 세부 내용

AI 개발은 놀라운 속도로 진행되었고, Sam과 Anna와 같은 근로자들의 역할을 근본적으로 변화시켰습니다. Sam이 일하는 병원에서는 AI가 이제 환자와 감정적으로 상호작용할 수 있으며, 스트레스나 불편함의 미묘한 징후를 감지하고 맞춤형 위로를 제공합니다. 이러한 기계들은 복잡한 알고리즘을 사용해 증상을 해석하고, 결과를 예측하며, 치료를 추천하는 등 진단 역할도 일부 수행하고 있으며, 종종 인간 의사보다 더 빠르고 정확하게 결과를 도출합니다. Anna의 경우, 변화는 더욱 극적입니다. 건설 현장에서 로봇은 이제 단순히 자재를 들어 올리는 것을 넘어, 현장 조건을 분석하고 잠재적인 안전 위험을 예측하며, 자원을 가장 효율적으로 배분하는 실시간 결정을 내립니다. 긴급 상황에서는 AI 시스템이 작업자와 장비를 지휘하여, 인간의 감독 하에서만 가능했던 정밀도로 위험을 최소화하고 결과를 개선합니다. 하지만 이러한 전환을 관리할 적절한 규제가 마련되지 않은 상태에서 자동화는 수백만 명의 근로자를 일자리에서 밀어냈습니다. Anna와 Sam 같은 사람들은 자동화가 점점 더 진행되는 세계에서 새로운 역할을 찾기 위해 애쓰고 있습니다. AI가 인간 노동을 대체하면서 소득 불평등은 심화되고, 경제적 불안정성도 커지고 있습니다.

자료: Futures Platform(2024), “Four scenarios on the future of AI in the workplace”

4번째 시나리오는 AI 기술 수준이 기대에 미치지 못한 상황이다.



[그림] 시나리오4

자료: Futures Platform(2024), "Four scenarios on the future of AI in the workplace"

[표] 시나리오4 세부 내용

이 미래에서는 AI 도입이 신중하게 이루어졌으며, 이는 그만한 이유가 있습니다. 초기의 흥분에도 불구하고, AI는 감정적 지능과 복잡한 의사결정이 요구되는 분야에서 성과를 내지 못했습니다. Sam이 일하는 병원에서는 AI가 환자 돌봄을 위해 테스트 되었지만, 감정적 신호를 해석하고 미묘한 결정을 내리지 못해 역할이 제한되었습니다. 그 결과, 책임은 대부분 여전히 인간 직원에게 있으며, AI는 단순히 반복적이고 데이터 중심적인 작업에서만 보조 역할을 하고 있습니다. Anna의 건설 현장에서도 상황은 비슷합니다. 갑작스러운 날씨 변화나 자재 지연과 같은 예측 불가능한 요소들은 AI가 감당하기에 너무 어려웠고, 종종 잘못된 결정을 내려 작업 흐름을 방해했습니다. 결국 Anna와 그녀의 팀이 자주 개입해야 했습니다. 시간이 지나면서 AI가 산업 혁신을 가져올 것이라는 약속을 지키지 못하고 있다는 것이 분명해졌습니다. 이 시스템을 유지하는 높은 비용과 복잡한 역할에서의 제한적인 성공은 기업들이 AI 투자를 축소하도록 만들었습니다. 이에 AI 프로젝트에 대한 열정과 자금 지원이 감소하였으며, AI 개발의 속도가 줄어들고 산업 전반에 걸친 통합이 지연되고 있습니다. AI는 여전히 광범위한 인간의 감독이 필요한 도구로 남아 있으며, 현실 세계의 복잡성을 감당하기 어렵습니다.

자료: Futures Platform(2024), “Four scenarios on the future of AI in the workplace”

참고문헌

1. 문헌자료

2. 웹사이트

참 고 문 헌

1 문헌자료

강준모(2024), “딥페이크 관련 국내외 규제동향 분석” KISDI AI OUTLOOK 제18호.

김성근·김종엽(2015). 산업생태계 모형 및 영향 파급효과 묘사 방안: 공공 SI산업에의 적용. Journal of Information Technology and Architecture. 12(3): 299 - 309

경제인문사회연구회(2024), “인공지능 시대의 경쟁력 강화를 위한 AI 규제 연구”.

교육부(2023), “AI 디지털교과서 추진방안”

교육부(2024) “교육부, 학교 딥페이크 성범죄 피해현황 2차 조사결과 발표”

교육부(2022.8), “사람의 성장을 지원하는 교육분야 AI 윤리원칙”

교육부, 한국교육학술정보원(2023), “AI 디지털교과서 개발 가이드라인”

교육부 보도자료(2023.10.16.), “AI 디지털교과서 법적 지위를 얻다.”

교육부 보도자료(2024.3.14.), “2023년 초중고사교육비조사 결과”

국가교육위원회 제14차 회의(2023.6.9.), “AI 디지털교과서 추진방안”

곽정호·조지연·이용성·이봉규(2011). 새로운 통신시장 활성화를 위한 모바일 생태계 통신정책. 인터넷정보학회논문지. 12(4): 93 - 106.

법제처(2024), “인공지능 관련 국내외 법제 동향”

삼정KPMG(2024) “생성형 AI에게 펼쳐진 새로운 무대, 온디바이스 AI”

삼정KPMG(2024) “창작영역에 뛰어든 생성형 AI 투자현황과 활용 전망”

서울특별시 교육청 교육연구정보원(2022), “개별 맞춤형 AI 활용 교육의 가능성과 과제: AI튜터 마중물 학교 운영사례를 중심으로”

서한결·박인서·김성환·채수준·정양헌(2015). 공공 R&D (참여) 조직의 공생적 학습 생태계 조성방안 연구. 경영교육연구. 30(6): 557 - 579.

이승환(2018.10), “교육혁신의 동력, AI”, SW정책연구소 산업동향.

이승환(2020), “빅데이터로 본 딥페이크: 가짜와의 전쟁”, SW정책연구소 이슈리포트 IS-090

이승환(2023), “AI시대 절대 대체되지 않는 슈퍼 개인의 탄생”

입법조사처(2024), “제22대 국회 입법정책 가이드북 Ⅲ”

유네스코(1999), “BEIJING CONSENSUS on artificial intelligence and education”

한국교육개발원(2023), “AI 기반 맞춤형 교육의 현황과 과제”

한국지역정보개발원(2024), “GPT-4o의 지방자치단체 도입 현황과 활용방안”

KAIT(2019), “영국의 AI 산업 육성 전략”

KITA(2024) “인공지능 규제 주도권 확보를 위한 글로벌 경쟁 및 시사점”

KOTRA(2024) “일본의 AI 정책과 실제 사례”

PwC(2024) “생성형 AI를 통한 가치 창출의 경로: 플라이휠(Flywheel) 접근법”

PwC(2024) “생성형 AI를 활용한 비즈니스의 현주소”

PwC(2024) “생성형 AI 기반의 Accelerated AI: 금융 AI의 발전 전략”

SW 정책연구소(2024) “해외 AI 안전연구소 추진현황과 시사점”

SK하이닉스 뉴스룸(2024.3.15.) “AI의 시작과 발전과정, 미래 전망”

NIA(2023) “영국 AI 규제백서의 주요 내용 및 시사점”

Ark Investment(2024), “Big Idea 2024”

Auernhammer, J. (2020). Human-centered AI: The role of human-centered design research in the development of AI. In S. Boess, M. Cheung, & R. Cain (Eds.), Synergy - DRS international conference 2020 (pp. 1315–1333). https://doi.org/10.21606/drs.2020.282

Association of Test Publishers. (2024). Creating Responsible and Ethical AI Policies for Assessment Organizations, July 12, 2024.

Bain&Company(2024) “AI’s Trillion-Dollar Opportunity”

Basole, R.C., “Visualization of Interfirm Relations in a Converging Mobile Ecosystem”, Journal of Information Technology, Vol. 24, pp. 144-159, 2009

Baker, T., Smith, L., & Anissa, N. (2019). Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges. Retrieved May, 12, 2020

Belzak, W. C., Naismith, B., & Burstein, J. (2023, June). Ensuring fairness of human-and AI-generated test items. In International Conference on Artificial Intelligence in Education (pp. 701-707). Cham: Springer Nature Switzerland.

Bloom, B. S. 1984. The 2 Sigma Problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, Vol. 13, no. 6, pp. 4-16.

Boykin, A. et al. (1996). “Intelligence: Knowns and unknowns”. American Psychologist. 51. pp.77~101.

Burstein, J. (2023). Responsible AI standards. Duolingo English Test.

Carbonell, J. R. 1970. AI in CAI: An artificial-intelligence approach to computer-assisted instruction. IEEE Transactions on ManMachine Systems, Vol. 11, No. 4, pp. 190-202.

Chapelle, C. A. (2008). The TOEFL validity argument. In C. A. Chapelle, M. K. Enright, & J. Jamieson (Eds.), Building a validity argument for the Test of English as a Foreign Language (pp. 319–352). London

Deepmind(25 July 2024), "AI achieves silver-medal standard solving International Mathematical Olympiad problems"

EIT(2021), "Creation of a taxonomy for the European AI ecosystem"

Futures Platform(2024), "Four scenarios on the future of AI in the workplace"

Hamsa Bastani et al, "Generative AI Can Harm Learning" (July 15, 2024). Available at SSRN: https://ssrn.com/abstract=4895486 or http://dx.doi.org/10.2139/ssrn.4895486

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, "Generative Adversarial Networks".

Ivancevich A.(2022), "Deepfake Reckoning: Adapting Modern FIrst Amendment Doctrine to Protect Against the Threat Posed to Democracy," 49(1) Hastings Const. L.Q

Iansiti, M. and Levien, R., "Strategy as Ecology", Harvard Business Review, Vol. 82, No. 3, pp. 68-81, 2004.

Jill Burstein, Geoffrey T. LaFlair, Kevin Yancey, Alina A. von Davier, Ravit Dotan(2024) "Responsible AI for Test Equity and Quality: The Duolingo English Test as a Case Study", https://arxiv.org/abs/2409.07476

Johnson, M. S., Liu, X., & McCaffrey, D. F. (2022). Psychometric methods to evaluate measurement and algorithmic bias in automated scoring. Journal of Educational Measurement, 59(3), 338-361

Moore, J. F.(1996). The Death of Competition: Leadership and Strategy in the Age of Business Ecosystems. New York. NY: Harper Business.

Mühlenbein, H. (2009). “Computational Intelligence: The Legacy of Alan Turing and John von Neumann. In: Mumford, C.L., Jain, L.C. (eds) Computational Intelligence”. Intelligent Systems Reference Library. Vol 1. Springer. Berlin. Heidelberg.

McCulloch, W.S. and Pitts, W.H. (1943) A Logical Calculus of the Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics, 5, 115-133.

Mckinsey&Company(2023) “The economic potential of generative AI”

Mckinsey&Company(2023), “Exploring opportunities in the generative AI value chain”

Miao, F., & Holmes, W. (2021). Artificial Intelligence and Education. Guidance for Policy-makers.

Messerschmitt, D.G. and Szyperski, C., Software Ecosystem, MIT Press, 2005.

onfido(2024), “Identity Fraud Report 2024”

North Carolina Department of Public Instruction(2024), “North Carolina Generative AI Implementation Recommendations and Considerations for PK-13 Public Schools”

O’Donnell, N.(2021), “Have We No Decency? Section 230 and the Liability of Social Media Companies for Deepfake Videos,” 2021(2) U. Ill. L. Rev.

Omer Shahab et al(2024) “SoroushLarge language models: a primer and gastroenterology applications” Ther Adv Gastroenterol 2024, Vol. 17: 1–15

Porter, M. E.(1985). Competitive Advantage: Creating and Sustaining Superior Performance. New York. NY: The Free Press.

Regona, Massimo & Yigitcanlar, Tan & Xia, Bo & Li, R.Y.M. (2022). Opportunities and adoption challenges of AI in the construction industry: A PRISMA review. Journal of Open Innovation Technology Market and

Complexity, 8(45).

Routledge.; Kane, M. T. (1992). An argument-based approach to validity. Psychological Bulletin, 112(3),

Self, J. A. 1974. Student models in computer-aided instruction.International Journal of Man-Machine Studies, Vol. 6, No. 2, pp. 261-276.

Sequoia Capital(2023) "Generative AI: A creative New World"

Tansley, A.G., "The Use and Abuse of Vegetational Terms and Concepts", Ecology, Vol. 16, No. 3, pp. 284-307, 1935

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2017). Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems (Version 2). IEEE. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

Turing, A. M. (2021). Computing Machinery and Intelligence (1950).

UK(2023), "A pro-innovation approach to AI regulation"

UNESCO(2021), "AI and education: guidance for policy-makers"

U.S. Department of Education, Office of Educational Technology (2024). Designing for Education with Artificial Intelligence: An Essential Guide for Developers, Washington, D.C

Vaz, L.F.H., Nogueira, A.R.R., Rodrigues, M.A.D.S.,and Chimenti, P.C.P.D.S., "A New Conceptual Model for Business Ecosystem Visualization and Analysis", Revista de Administrao Contempornea, Vol. 17, No. 1, pp. 1-17, 2013.

Von Davier, A. & Burstein, J. (in press). AI in the Assessment Ecosystem: Implications for Fairness, Bias, and Equity. In Artificial Intelligence in Education: The Intersection of Technology and Pedagogy, Springer

Nature Switzerland

Wang, G. (2019, October 20). Humans in the Loop: The Design of Interactive AI Systems:https://hai.stanford.edu/news/humans-loop-design-interactive-ai-systems.

World Economic Forum(2024), “Shaping the Future of Learning: The Role of AI in Education 4.0”

World Economic Forum(2024), “Global Risks Report 2024”

WDE(2024), “Guidance for Wyoming School Districts on Developing Artifcial Intelligence Use Policy”, https://edu.wyoming.gov

Yang, X.C., Li, Z.H., and Liang, Q., “An Analysis on the Ecosystem of Leisure Sports Industry and Its Competition Strategy Choice”, Journal of Beijing Sport University, Vol. 3, No. 7, pp. 25-28, 44, 2009.

ZeroBounce Research(2024) “Countries that are AI superpowers”

2 웹사이트

epochai, 2024.10.13. 접근, (https://epochai.org/data/large-scale-ai-models)

https://trackingai.org/IQ

AI Incident Database

https://leg.colorado.gov/sites/default/files/2024a_205_signed.pdf

https://le.utah.gov/~2024/bills/static/SB0149.htm

Ensuring Likeness Voice and Image Security Act of 2024

https://csf.kiep.go.kr/newsView.es?article_id=50438&mid=a20100000000

뉴스1(2023.4.20.) “日 요코스카시, 지자체 최초 챗GPT 시범 도입”

방송통신심의위원회, “딥페이크 악용·유명 연예인 합성 ‘성적 허위 영상물’ 400% 폭증”,

2024.5.2

BuzzFeedNew.com, “This PSA About Fake News From Barack Obama Is Not What It Appears”, 2018.4.18.

https://www.tortoisemedia.com/intelligence/global-ai/#rankings

TechCrunch(September 18, 2024),“This Week in AI: Why OpenAI’s o1 changes the AI regulation game”

Forbes(Jan 23, 2024) “The Next Leap In AI: From Large Language Models To Large World Models?”

https://www.youtube.com/watch?v=A8TmqvTVQFQ

https://lyriko.ai/unleashing-the-power-of-chatgpt-in-pharma/

https://openai.com/index/navigating-the-challenges-and-opportunities-of-synthetic-voices/

https://www.microsoft.com/en-us/research/project/vasa-1/

리드호프먼 유튜브, https://www.youtube.com/watch?v=rgD2gmwCS10

Wired(2.23.6.7), “How Indiana Jones and the Dial of Destiny De-Aged Harrison Ford”;

중앙일보(2023.12.27.), “조용필 명곡·AI로 구현한 송해…‘웰컴투 삼달리’가 전하는 위로와 향수.”

뉴스1(2023.4.20.) “日 요코스카시, 지자체 최초 챗GPT 시범 도입”

https://www.caryn.ai/ ;

Forbes(2023.5.11.) “Snapchat Star Earned $71,610 Upon Launch Of Sexy ChatGPT AI, Here’s How She Did It“

PC Gamer(2024.4.27.), “The solo developer of Manor Lords, Steam's latest smash hit, named his studio after a Witcher 3 meme”

https://petitions.assembly.go.kr/about/notices/152C967DCD270F23E064B49

691C1987F

한겨레 신문(2024.7.18), “당장 내년 도입인데, AI 디지털교과서 없는 교사 연수라니”

SBS(2024.7.17.), ‘AI 교과서’교사 연수용 보니…“뭐가 AI 기능?”

경향신문(2024.7.24.), “디지털 선도학교서 쓰는 AI앱, 학생 정보 술술 빼갔다”

https://ourworldindata.org/grapher/test-scores-ai-capabilities-relative-human-performance

https://evolvingai.io/p/anthropic-now-leads-ai-race

https://openai.com/index/hello-gpt-4o/

Deepmind(25 July 2024), “AI achieves silver-medal standard solving International Mathematical Olympiad problems”

https://www.aiforeducation.io/ai-resources/state-ai-guidance

오마이뉴스(2024.7.30.), “15년 차 교사가 챗GPT로 ‘숙제봇’ 만들었습니다”

Abstract

Future policy research in the era of the generative AI revolution

NATIONAL ASSEMBLY FUTURES INSTITUTE

The rapid advancement of AI technology is fundamentally transforming society, as evidenced in the comprehensive analysis presented in this report, which examines three critical aspects of AI development and their policy implications as of late 2024. This study first explores the evolving landscape of AI policy, highlighting a significant shift from purely promotional approaches to a more balanced framework incorporating both regulation and advancement, as exemplified by landmark policies such as the EU AI Act, the US Executive Order on AI, and the UK's AI regulatory framework, while also noting the intriguing emergence of local governments as key players in AI implementation and governance. In parallel, this report delves into the concerning proliferation of deepfake technology, which saw a remarkable 31-fold increase in attempts during 2023, presenting a dual-edged phenomenon that, while offering innovative possibilities for media production and content creation, simultaneously poses significant threats to individuals and organizations, thereby necessitating enhanced detection technologies and robust legal frameworks to mitigate potential misuse and harm. The third focal point of this analysis centers on South Korea's groundbreaking initiative to implement AI digital textbooks, a world-first endeavor that, while promising personalized learning experiences and a potential reduction in private education expenses, raises important questions about educational effectiveness, fairness, and safety. This ultimately leads to eight key considerations identified through expert focus group interviews, including the need for clear AI conceptualization, comprehensive case analyses, multiple stakeholder perspectives, and systematic measurement of

learning outcomes. Throughout these analyses, this report emphasizes the critical importance of developing proactive, balanced policy responses that can keep pace with rapidly evolving AI technologies, while simultaneously ensuring adequate safety measures, promoting innovation, and maintaining public trust, all within a framework that acknowledges both the transformative potential of AI and the need for careful governance to mitigate associated risks.

Furthermore, this study stresses the necessity of international cooperation in addressing these challenges, particularly in areas such as deepfake detection and AI safety standards, while also highlighting the importance of developing robust domestic frameworks that can effectively balance technological advancement with appropriate regulatory oversight. This report concludes by underscoring the urgency of preparing for future developments in AI technology, including the potential emergence of new educational disparities, the likely rise of premium offline education in response to AI integration, and the possible transformation of educational boundaries in an increasingly AI-driven world, all while emphasizing the need for continuous monitoring and adaptive policy responses to address these evolving challenges in the rapidly changing landscape of artificial intelligence.

생성형 AI 혁명 시대 미래정책 연구

발 행 2024년 12월 30일
발 행 처 국회미래연구원
주 소 서울시 영등포구 의사당대로 1
전 화 02)786-2190
팩 스 02)786-3977
홈페이지 www.nafi.re.kr
인 쇄 처 명문인쇄공사 02)2079-9200

ISBN 979-11-986487-8-5 (95300)

국민을 지키는 미래로 나아가는 국회

국회미래연구원
NATIONAL ASSEMBLY FUTURES INSTITUTE