2024.10.28.

국회미래연구원 | 국가미래전략 Insight | 112호

초거대 AI 등장 이후 AI 정책변화의 특징과 전망

이승환(혁신성장그룹 연구위원)

국회미래연구원
NATIONAL ASSEMBLY FUTURES INSTITUTE

국가미래전략 Insight | 112호

ISSN
2733-8258

발행일
2024년 10월 28일

발행처
국회미래연구원
서울시 영등포구 의사당대로 1
Tel 02-786-2190 Fax 02-786-3977

「국가미래전략 Insight」는 국회미래연구원이 정책고객을 대상으로 발행하는 단기 심층연구결과물로, 내부 연구진이 주요 미래이슈를 분석한 내용을 토대로 국가의 미래전략을 제시합니다.

Contents

요약

초거대 AI 등장 이후, AI의 빠른 진화로 규제 시차(Regulatory Lag) 이슈가 제기되면서 AI 정책에 변화가 일어나고 있어 이에 주요국 사례를 통해 AI 정책변화의 특징을 분석하고 시사점을 도출

(특징1) AI 진흥에서 규제와 진흥의 조화로

- 초거대 AI가 부상하면서 주요국은 AI의 빠른 진화를 인식하고 진흥과 규제의 조화방안을 모색
- EU는 2024년 최초의 AI 규제법안인 EU AI 법을 최종 승인
- 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

(특징2) AI의 빠른 진화로 야기될 위험에 대비하기 위한 안전 및 신뢰 정책을 강화

- EU AI 법은 위험 수준에 따라 AI 시스템에 대한 규제 수준을 차등화하는 위험 기반 접근
- 미국의 AI 행정명령은 AI 기술의 안전 및 보안(Safety and Security) 보장을 강조
- 미국, 영국, 일본 등은 AI 안전연구소 설립 및 운영을 통해 안전과 신뢰성 확보를 위해 노력

(특징3) 국가별 AI 정책 추진 방식 및 규제 강도가 상이

- EU의 AI 법은 AI 전반을 포괄하며 강도 높은 처벌 규정을 포함
- 미국은 포괄적인 규제 대신 행정명령을 통해 AI의 잠재성은 극대화하고 중국 견제 등 국가 안보와 거짓 정보 대응 등 우려 사항에 대비
- 영국은 EU의 강력한 규제와는 달리 친혁신적 AI 정책 접근 방식을 추진 중이며, 중국은 포괄 규제하기보다 새로운 AI 이슈에 대해 개별 규칙을 신속하게 제정

(특징4) AI에 주목하는 지자체

- 미국 행정부의 AI 행정명령 이행과 함께, 주(州)별 AI 법제 논의가 활발히 진행 중
- 중국은 지방정부 차원에서도 AI 정책을 추진하며 지역 경쟁력 확보를 위해 노력 중
- 주요국 지자체에서 AI 도입을 통해 행정서비스 제고 방안을 모색

(시사점) 초거대 AI로 변화하는 AI 정책의 특징에 주목하고 국내 AI 정책 수립 시 반영

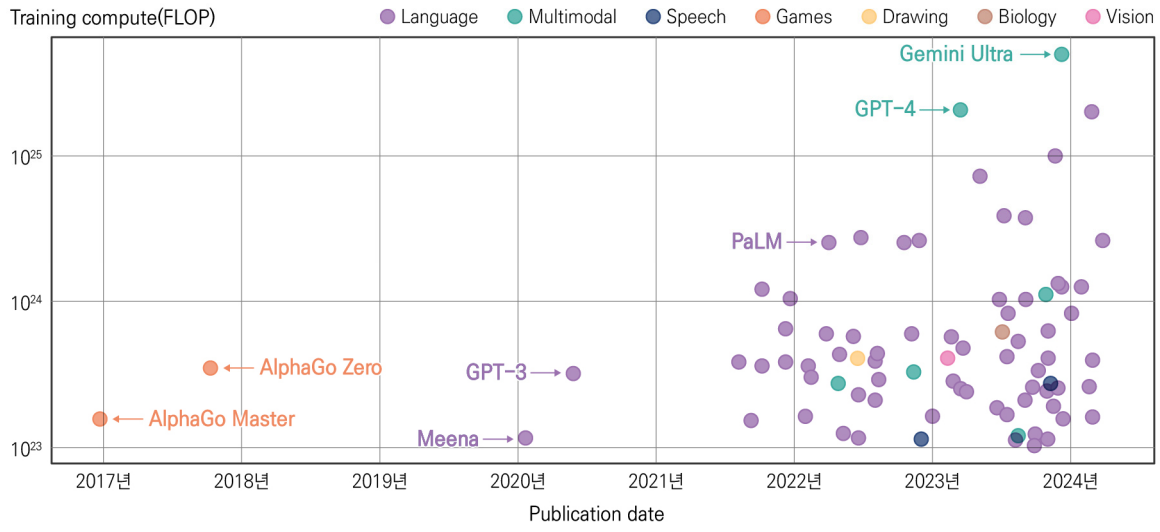
- 계류 중인 AI 법 도입 논의에 있어 변화하는 AI 환경과 정책 특성을 종합적으로 고려
- 국가별 초거대 AI 경쟁력을 고려한 AI 정책을 수립
- 국내 AI 안전연구소 설립 및 운영 관련 글로벌 사례를 참고하고 글로벌 정책 공조 체계를 강화
- 중앙정부와 지자체 AI 정책의 유기적 연계 방안을 모색
- 동태적 관점에서 빠르게 변화하는 AI 생태계를 관찰하고 미래를 준비

01 AI의 빠른 진화와 정책의 고민

■ **AI 기술의 빠른 진화로 다수의 초거대 AI 모델이 출시되며 경쟁 중이고, 이에 따라 AI의 성능도 급격히 증가 추세**

○ 2020년~2023년까지 4년간 총 144개의 초거대 AI 모델이 출시되었으며 2024년에도 지속 출시 중

그림 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)[1]에서 은메달 수준의 성적을 기록했다고 발표[2]

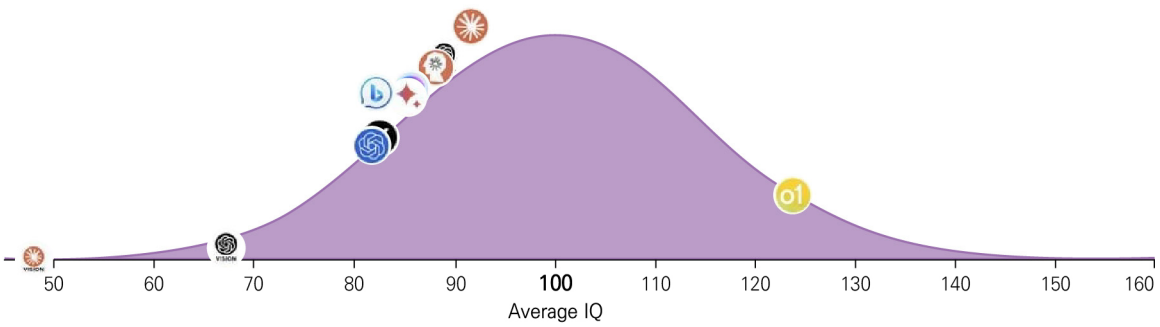
- AI가 국제수학올림피아드의 메달권에 해당하는 성능을 보인 건 처음

1) 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

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

- 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 가능성
  - 미국의 데이터 분석가인 맥심 로트(Maxim Lott)는 Open AI가 공개한 AI인 o1이 노르웨이 멘사의 IQ 테스트에서 120을 기록했다고 밝힘[3]

그림 2 주요 초거대 AI 모델의 IQ 테스트 결과



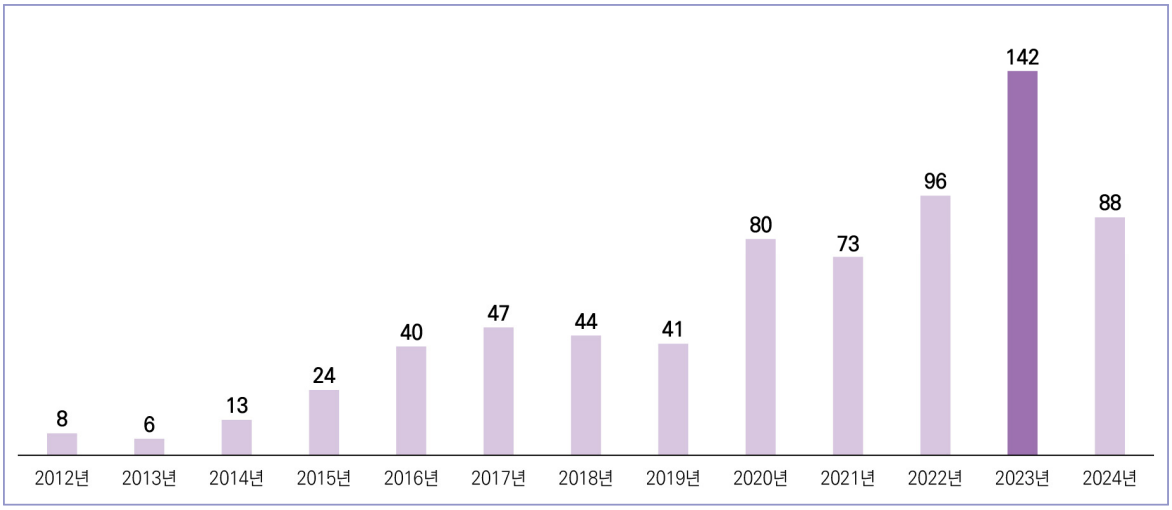
자료: https://trackingai.org/IQ

■ **AI의 빠른 진화로 규제 시차(Regulatory Lag) 이슈가 제기되며 AI 정책변화를 야기 중이며 이에 주요국 사례를 통해 AI 정책변화의 특징을 분석하고 시사점을 도출**

- 규제 시차는 규제문제의 발생과 해결되는 시점 사이의 시간 차이를 의미하며 기술 발전의 속도가 빠를수록 규제 공백 또는 규제 방치(regulatory drift)로 인한 사회적 손실이 커질 수 있어 규제 시차는 정책적 논의의 대상[4]
- 실제, 초거대 AI 논의가 활발히 시작된 2022년부터 AI로 발생한 사고 사례도 증가하고 있으며[5] 이에 AI 정책도 변화가 일어나는 중

3) 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
4) 최병선. 1992.『정부규제론』. 서울: 법문사
5) 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.

그림 3 연도별 AI 사건 수



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

02 AI 정책변화의 4대 특징

1. '진흥'에서 '규제와 진흥'의 조화로

■ **2016년 알파고 대국 이후 AI 잠재력에 주목한 주요국들은 진흥 중심의 AI 정책을 지속 발표**

- EU는 2018년 AI 전략(Communication Artificial Intelligence for Europe)을 발표하고 AI 산업 육성을 위해 2020년까지 15억 유로 투자 계획을 수립
- 미국은 2016년 AI 국가 연구개발 전략을 발표했으며, 2019년 AI 연구개발과 투자에 우선순위를 두는 AI 행정명령(American AI Initiative)을 추진
  - 민간이 추진하기 어려운 차세대 연구개발 및 군사 안보 분야 활용에 중점
- 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 제안
  - 민간과의 협력을 기반으로 AI 인재 양성 및 비즈니스 환경조성에 투자 집중
- 독일도 2018년 AI 육성전략을 발표하며 AI를 활용한 중소 및 제조 분야 경쟁력 제고를 위해 대규모 투자를 계획하고 기술력 확보를 위해 노력
- 프랑스는 2018년 AI 권고안을 발표하고 AI 생태계 조성, 전략 분야 융합 및 직업·고용, 윤리 등 문제 해결을 위해 노력
- 중국은 2017년 차세대 AI 발전 계획을 발표하며 AI와 데이터 분야 대규모 투자 및 인력양성을 추진하고, 선도기업을 지정하여 산업별 특화플랫폼을 육성
  - 정부 주도로 자국 기업을 활용한 산업별 플랫폼을 구축, 막대한 데이터를 축적함으로써 AI 경쟁력 확보
- 일본은 2017년 AI 기술 전략을 발표하고 이후 AI 전략을 지속 업데이트
  - 2017년 AI 기술 전략 발표 이후 2019, 2021, 2022년 AI 전략을 발표하며 국가전략을 수정

표 1 연도별 주요 이슈와 일본 AI 정책

<table>
<tr><th></th><th></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>세계</td><td>주요 이슈</td><td></td><td>알파고 쇼크</td><td>GPT-1 발표</td><td>GPT-2 발표</td><td colspan="2">GPT-3 발표
코로나19 팬데믹</td><td></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 rowspan="2">주요 조직</td><td colspan="3">AI기술전략회의</td><td></td><td></td><td>새로운AI 전략검토회의</td><td></td><td>AI전략 회의</td></tr>
<tr><td colspan="2"></td><td colspan="7">인간중심 AI 사회원칙회의</td></tr>
</table>

자료: KOTRA(2024) “일본의 AI 정책과 실제 사례”

- 한국도 2018년 AI 연구개발 전략, 2019년 AI 국가전략 등을 발표하며 미래를 준비
- 이외에도 EU 신뢰할 수 있는 AI 지침(‘18), OECD AI 권고안(’19), 영국의 설명 가능한 AI 지침(’20) 등이 발표되었으나 구속력이 없는 지침으로 대부분 진흥 중심으로 추진

■ **이후, 초거대 AI가 부상하면서 AI의 빠른 진화를 인식하고 진흥과 규제의 조화방안을 모색**

- EU는 2024년 최초의 AI 규제법안인 EU AI 법을 최종 승인
  - 입법의 목적은 인간중심의 신뢰할 수 있는 AI의 촉진, AI 시스템으로부터의 안전, 환경 보호, 법치 등 기본권을 보호하고 혁신을 지원하는 것
  - EU 내에서 사용되는 AI 시스템에 EU의 가치가 일관되게 적용될 수 있는 원칙을 수립
  - 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 존재
- 유럽의회(European Parliament)는 사용자가 자연어를 활용하여 유럽의 방대한 법제도 자료에 쉽게 접근할 수 있도록 생성형 AI ‘Ask the EP Archives’를 도입(‘24.10)
  - AI 기업 클로드(Claude)와 협력하여 생성형 AI 챗봇을 개발하고 연구원, 정책입안자, 교육자, 대중이 210만 개 이상의 공식 문서를 손쉽게 사용할 수 있도록 지원

그림 4 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 제시

- 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 포함
- 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 밝힘

○ 영국 정부는 2023년 AI 규제에 대한 방향과 원칙이 담겨있는 AI 규제백서를 발표하며 AI 규제 체계 기반을 마련

- 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 추구
- AI 규제백서에서는 친혁신적인 규제 Framework의 5가지 원칙을 제시

표 2 친혁신적인 규제 Framework의 5가지 원칙

구분	내용
안전, 보안,견고성	• AI 수명주기 전반에 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 명시하였으며 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 필요성과 AI 공급망 속 적합한 행위자에게 규제 준수의 의무가 있음을 기술 표준이나 모범사례 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 시작

- AI 사업자 지침은 총 5개 장으로 구성되어 있으며, AI 관련 개발자, 제품 및 서비스 제공자, 이용자를 대상으로 10개 원칙을 제시하였으며 해당 지침에 4,000여 건의 의견이 접수되었으며 이를 반영
- AI 전략회의에서 AI 규제 기본방침과 AI 안전성 확보를 위한 법률 규제 방침을 밝혔으며(2024.5), 정기 국회 법안 제출 후, 2026년 전면 시행이라는 계획도 제시

표 3 **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: 総務省, 経済産業省(2024) "AI 事業者ガイドライン案"

○ 국내에서도 21대 국회에서 AI 법 도입 논의가 활발히 이루어졌으나 회기가 종료되면서 폐기되었고 22대 국회에서 재논의 중

- 국내 AI 법안은 21대 국회에서 9개가 발의되었고 9건의 법안 소관 상임위(과방위)에서는 2022년까지 발의되었던 7건의 인공지능 관련 법률안을 병합하여 심사(2023. 2. 14.)한 이후, 이를 통합·조정하여 위원회의 대안을 제시하기로 하였으나, 비공개로 수정 보완 중인 상태에서 21대 국회 회기가 종료되면서 폐기[6)]

- 22대 국회에서는 AI 법안이 11건 발의(2024년 9월 30일 기준)

표 4 **제22대 국회 인공지능 관련 법률안 발의 현황('24.9.30 기준)**

의안명	제안자명	제안일자	진행상태
인공지능산업 진흥 및 신뢰 확보 등에 관한 특별법안	김우영 의원 등 19인	2024-09-24	소관위접수
인공지능의 발전과 안전성 확보 등에 관한 법률안	이훈기 의원 등 14인	2024-09-12	소관위접수
인공지능 발전 진흥과 사회적 책임에 관한 법률안	배준영 의원 등 10인	2024-08-28	소관위접수
인공지능책임법안	황 희 의원 등 10인	2024-08-27	소관위접수
인공지능 기본법안	한민수 의원 등 10인	2024-08-22	소관위접수
인공지능 개발 및 이용 등에 관한 법률안	권칠승 의원 등 15인	2024-07-04	소관위심사
인공지능기술 기본법안	민형배 의원 등 13인	2024-06-28	소관위심사
인공지능산업 육성 및 신뢰 확보에 관한 법률안	김성원 의원 등 11인	2024-06-19	소관위심사
인공지능산업 육성 및 신뢰 확보에 관한 법률안	조인철 의원 등 19인	2024-06-19	소관위심사
인공지능 발전과 신뢰 기반 조성 등에 관한 법률안	정점식 의원 등 108인	2024-06-17	소관위심사
인공지능 산업 육성 및 신뢰 확보에 관한 법률안	안철수 의원 등 12인	2024-05-31	소관위심사

자료: 의안정보시스템(2024.9.30.기준)

6) 법제처(2024), "인공지능 관련 국내외 법제 동향"

2. AI 안전과 신뢰 정책의 강화

■ **AI의 빠른 진화로 야기될 위험에 대비하기 위한 안전 및 신뢰 정책을 강화**

○ EU AI 법은 위험 수준에 따라 AI 시스템에 대한 규제 수준을 차등화하는 위험 기반 접근(Risk-based approach) 방식

- 특정한 AI 시스템을 금지하거나 고위험 AI 시스템, 제한된 위험을 갖는 AI 시스템, 최소위험 AI 시스템 등으로 분류하여 차등 규제

- 수용 불가 AI 시스템은 EU AI 법에서 규정된 특별한 예외가 없는 한 사용 자체가 금지

표 5 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 사회점수(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: https://artificialintelligenceact.eu/

- 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 영향평가를 수행

표 6 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 교통의 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 신청서를 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 서비스에 대한 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 평가하고 분류하거나 경찰, 소방관, 의료 지원 및 응급 의료 환자 분류 시스템을 포함한 응급 대응 서비스를 파견하거나 파견 우선순위를 설정하는 데 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 사용되는 경우 • (c) 법 집행기관 또는 관련 지원기관에서 범죄의 수사 또는 기소 과정에서 증거의 신뢰성을 평가하기 위해 사용하도록 의도된 AI 시스템 • (d) 법 집행기관 또는 관련 지원기관에서 범죄 또는 재범 위험을 평가하기 위해 사용되는 경우
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 사용되는 경우 • (b) 선거 또는 국민투표의 결과 또는 선거 또는 국민투표에서 투표를 행사하는 사람의 투표 행동에 영향을 미치도록 의도된 AI 시스템

자료: https://artificialintelligenceact.eu/

- 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 의무가 부과[7)]
- 사회경제적 파급효과가 큰 범용 AI 모델을 일반 AI 시스템과 구분하고 범용 AI 모델 공급자는 AI 개발 및 테스트에 대한 자세한 기록 보관, 지적재산 보호, AI를 사용하려는 다른 회사에 관련 정보 제공, EU 위원회 및 국가 당국과 협력하도록 규정

표 7 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: https://artificialintelligenceact.eu/

○ 미국의 AI 행정명령은 AI 기술의 안전 및 보안(Safety and Security) 보장을 강조

- 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 선언
- 이중 용도(dual-use)[8)] AI 모델 개발 시 안전성 테스트(AI red-teaming test) 및 그 결과를 정부에 보고[9)] 하도록 규정
- 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 언급
- 안전하고 신뢰할 수 있는 AI 업계 관련 표준과 사례 개발을 강조하며 특히, 미국 기업의 AI 기술을 이용하는 외국인(기업)도 행정명령의 적용 대상이며, 외국인도 안전성 평가 및 그 결과를 보고하도록 규정[10)]

7) 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

8) 평화와 군사 목적 모두 사용될 수 있는 기술

9) AI 시스템의 결함과 취약성을 식별하기 위한 구조화된 시험으로 통제된 환경에서 AI 개발자와 협업하여 수행되며 레드팀은 AI 시스템의 결함과 취약점을 식별

10) 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

- 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 포함

○ 영국도 AI 규제백서에서 AI 안전 및 신뢰성을 강조하고 있으며 영국표준협회는 책임감 있는 AI 관리 지원을 위한 글로벌 지침을 발표

- AI 규제백서에서는 AI 수명주기 전반에 걸친 안정적인 작동을 위한 지속적인 안전 평가, 식별 및 관리체계 도입의 필요성을 강조
- 영국표준협회(British Standards Institution)가 사회 전반에서 AI를 안전하고 보안이 유지되며 책임감 있게 사용할 수 있도록 설계된 AI 관리 지침을 발표(2024)

■ AI 안전연구소 설립 및 운영을 통해 안전하고 신뢰할 수 있는 환경을 마련하기 위해 노력

○ 미국은 2023년 국립 표준기술연구소(NIST) 내에 AI 안전연구소를 신설하고 운영

- 미국 AI 안전연구소는 AI 안전 기초연구, AI 안전 테스트 구조 개발, AI 안전 국제 협력을 포함하여 안전한 AI 혁신을 위한 다양한 기능을 수행
- AI 안전연구소는 안전한 AI 개발 및 배포 관련 기술 표준 수립을 위해 2024년 2월 대규모 AI 안전연구소 컨소시엄을 발족하고, 공공-민간협력체계를 구축
- 컨소시엄은 엔비디아, 구글, MS, 애플 등 미국 내 AI 관련 기업 포함 200개 이상의 회원사와 대학, 연구기관 등으로 구성되었으며 5개의 작업반(Working Group)을 운영

표 8 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(dual-use) 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 수행

- 첨단 AI 시스템의 위험 평가를 통하여 정책입안자에게 AI 위험 정보를 제공하는 것을 목표로 하며, AI 안전성을 테스트할 수 있는 자체 역량 구축에 주력

표 9 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 시스템의 인간 통제 불능 가능성 평가 등
AI 안전 연구 촉진	• AI 거버넌스를 위한 평가 도구 및 기술개발, AI 시스템 평가 체계 구축, AI 영향측정 평가 도구 개발 등
글로벌 AI 협력 및 안전 정책 강화	• AI 안전 분야 정보 교환, 정부, 국제파트너, 민간기업, 학계, 시민사회 및 일반 대중과 교류 확대 등

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

○ 일본도 2024년 AI 안전연구소를 설립하여 신뢰할 수 있는 AI 환경조성을 위해 노력

- AI 안전성 정상회의, G7 히로시마 AI 프로세스 합의 등을 계기로 AI 안전연구소를 설립
- 일본의 AI 안전연구소는 경제산업성 산하 정보처리 추진 기구 내에 설립되어 운영되며 AI 안전 평가 및 기준조사, 기술 조사·연구, 국제 협력 업무를 수행

3. 국가별 AI 정책 추진 방식 및 규제 강도가 상이

■ EU의 AI 법은 AI 전반을 포괄하며 강도 높은 처벌 규정을 포함

○ EU AI 법은 역내뿐만 아니라 역외 사업자도 EU 시민을 대상으로 제품이나 서비스를 제공하면 적용 대상이며, AI 위험에 비례한 처벌 규정이 포함

표 10 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: https://artificialintelligenceact.eu/

- EU AI 법에는 AI 관련 규제 적합성 평가 사항이 포함되어 시스템 출시 전후 준수해야 할 사항을 규정하고 있으며 인증 및 적합성 검사도 포함
  - 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 준수
  - 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 통보
  - 이러한 절차를 완료하면 CE 마크를 획득하여야 하며[11], 자체 인증을 통해 적합성을 인증하며, 고위험 시스템의 일부의 경우에는 외부의 공인 기관의 적합성 검사가 필요
  - 표준이 제정되지 않은 생체인식 식별 또는 자연인 분류를 위한 AI 시스템과 타법률에 외부 공인 기관의 적합성 인증이 필요한 경우 외부의 적합성 검사가 필요
  - 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 함

■ **미국은 포괄적인 규제 대신 행정명령을 통해 AI의 잠재성은 극대화하고 중국 견제 등 국가 안보와 거짓 정보 대응 등 우려 사항에 대비**

- AI 행정명령에 혁신 및 경쟁의 촉진을 위해 AI 교육이나 취업 목적 비자 절차 간소화 등이 포함
- 근로자 지원(Supporting Workers), 평등과 시민권(Equity and Civil Rights)의 증진을 위한 방안을 제시
  - AI가 노동시장에 미치는 영향에 대해 보고하고, 직원에 대한 피해를 최소화하기 위해 고용주가 채택할 수 있는 원칙과 모범사례 제시
  - AI와 그 밖의 기술 기반 고용 시스템과 관련된 고용 시 차별금지에 관한 지침 공표
- 사생활과 시민 자유 보호를 위한 표준 및 절차 평가 강조
- 연방정부의 AI 사용 진전을 위해 최고 AI 책임자 지정, 직원 AI 사용 지침 개발 등이 포함
- 글로벌 AI 주도권 강화를 위해 AI 표준 홍보, 글로벌 참여 계획 수립 및 의제 개발을 강조
- 미국 기업의 AI 기술을 이용하는 외국인(기업)도 행정명령의 적용 대상이며, 외국인도 안전성 평가 및 그 결과를 보고하도록 규정[12]

11) CE 마크란 제품이 안전, 건강, 환경 그리고 소비자 보호와 관련된 EU 규격의 조건들을 준수한다는 의미의 유럽 통합규격 인증마크
12) 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

○ 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 포함

■ 영국은 EU의 강력한 규제와는 달리 친혁신적 AI 정책 접근 방식을 추진

○ AI의 '혁신 친화적 접근 방식'이란 용어는 2021년 9월 영국 정부가 발표한 '국가 AI 전략(National AI Strategy)'에서부터 본격적으로 사용

- 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 접근방식'을 유지·발전할 계획이라고 언급[13)]

○ 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 고려

표 11 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 개입이 되어야 함

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

○ 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 방침

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

■ **중국은 EU처럼 포괄 규제하기보다 새로운 AI 이슈에 대해 개별 규칙을 신속하게 제정**

- 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)
  - 중국 정부는 생성형 AI의 규제에 대해 포용과 신중, 기술의 유형과 중요성 등급에 따른 차등 감독 관리를 원칙으로 제시
- 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)
- 중국 국가인터넷정보판공실은 AI로 제작된 콘텐츠를 표시·식별할 수 있도록 의무화하는 규정 초안을 발표(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 규정
  - 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 권장됨
  - 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 강조
- EU는 금지 대상 AI 세분류, 기반 모델을 규제하는 등 포괄적인 규제방식을 취하고 있으나, 중국은 생성형 AI, 딥페이크 등 이슈 중심으로 대응

그림 5 **EU와 중국의 AI 규제방식 비교**

EU	중국
– 위험 기반 포괄적 규제 AI법 – 금지 대상 AI, 고위험 AI – 제한적 위험을 지닌 AI(챗봇, 딥페이크, AI 생성 콘텐츠) – 범용 AI(GPAI)	– 이슈별 단편적 규제 신속 도입 인터넷 정보서비스 **알고리즘 추천** 관리 규정 인터넷 정보서비스 심층 종합 관리 규정**(딥페이크)** **생성형 AI** 서비스 관리 잠정방법

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 중

○ 콜로라도주는 고위험 AI 시스템을 개념을 도입하고 해당 시스템의 개발자와 배포자의 의무를 규정(2024.5)

- 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 수행

표 12 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 시행[14]

- 생성형 AI를 이용하여 서비스를 제공하는 경우 이를 소비자에게 알려야 하며 생성형 AI 공개 의무 위반의 경우 2,500달러의 과태료를 부과하고, 집행 목적의 소송 제기 가능

○ 테네시주는 개인의 재산권(property right)을 성명, 이미지 및 초상을 보호하던 기존 법규정에 음성을 명시적으로 포함하여 재산권 보호 대상을 확대[15]

○ 캘리포니아주 의회는 2024년 8월 AI 규제법안 'SB 1047'을 통과시켰으나 개빈 뉴섬 캘리포니아 주지사가 거부권을 행사

- 해당 법안은 AI 기술의 급격한 발전에 따른 안전 문제를 해결하기 위해 마련

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

- 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 포함
- 5억 달러 이상의 피해 또는 사망 사고 발생 시 개발업체가 책임지도록 규정

○ 코네티컷주도 고위험 AI 시스템을 규제하는 법을 입법 추진 중

■ **중국은 지방정부 차원에서도 AI 정책을 추진하며 지역 경쟁력 확보를 위해 노력 중**

○ 베이징시는 AI 혁신 근거지 건설을 위한 베이징시 실시방안(2023~2025)을 발표(2023.5)[16)]

- 2025년까지 베이징 AI 핵심 산업 규모를 3천억 위안(약 5조 원)으로 키우고 10% 이상의 성장을 유지한다는 계획
- 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 마련

○ 선전시는 AI 고품질 발전 및 고수준 응용 가속화에 관한 행동 방안을 발표(2023.5)

- 1천억 위안(약 18조 원 규모)의 AI 펀드 설립과 함께 핵심 기술과 제품의 혁신 능력 강화, 산업 집적 수준 향상 등에 지원을 확대할 계획

○ 상하이시는 민간기업이 AI 인프라 건설에 광범위하게 참여할 수 있도록 정책지원을 하겠다고 밝혔으며, 상하이 쉬후이구는 다수의 거대 언어 모델(LLM) 연구개발팀을 적극 육성

■ **유럽 도시들은 공공 서비스에 사용되는 AI 알고리즘의 투명성을 높이고 시민 감독을 강화하기 위해 알고리즘 등록제(City Algorithm Register)를 도입**

○ 정부 기관들이 공공 서비스에 사용하는 알고리즘을 등록하고 그 정보를 공개하는 것이며 시민들은 이 정보를 온라인으로 쉽게 열람 가능

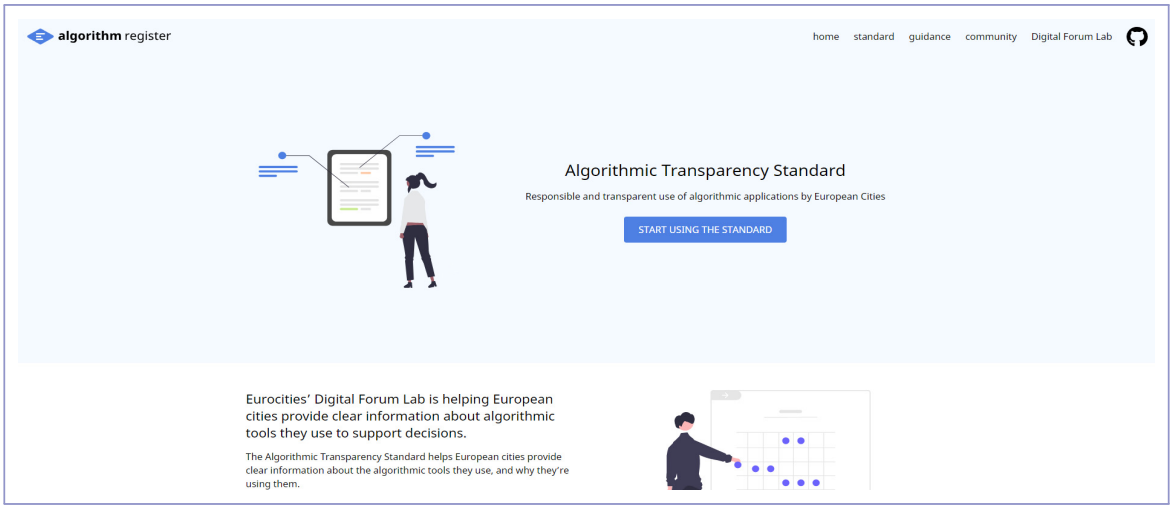
○ 공개되는 정보에는 알고리즘의 목적과 작동 방식, 사용되는 데이터의 종류, 책임자 정보, 인권이나 윤리적 측면에서의 위험성 평가 결과, 그리고 인간의 감독 정도 등이 포함

○ 암스테르담과 헬싱키를 비롯한 9개 유럽 도시들이 공통의 표준을 설정하여 이 제도를 채택

○ 암스테르담에서는 쓰레기 무단 투기 감지 시스템의 알고리즘을, 헬싱키에서는 도서관 도서 추천 시스템의 알고리즘을 등록

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

그림 6 유럽 도시의 알고리즘 등록제



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

■ 주요국 지자체에서 AI 도입을 통해 행정서비스 제고 방안을 모색

○ 뉴욕, 런던 등 주요 도시에서 AI를 활용한 대민서비스 제공, 도시 관리, 행정서비스 효율성 제고에 AI를 접목하는 시도가 이루어지는 중

표 13 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
런던	• 런던은 "Smart London" 프로젝트의 일환으로 GPT 기반의 AI를 활용하여 도시 관리와 시민 소통에 적용 • AI는 교통 분석, 환경 Data 수집, 시민들의 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 일환으로 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 문제 해결을 지원

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

03 시사점

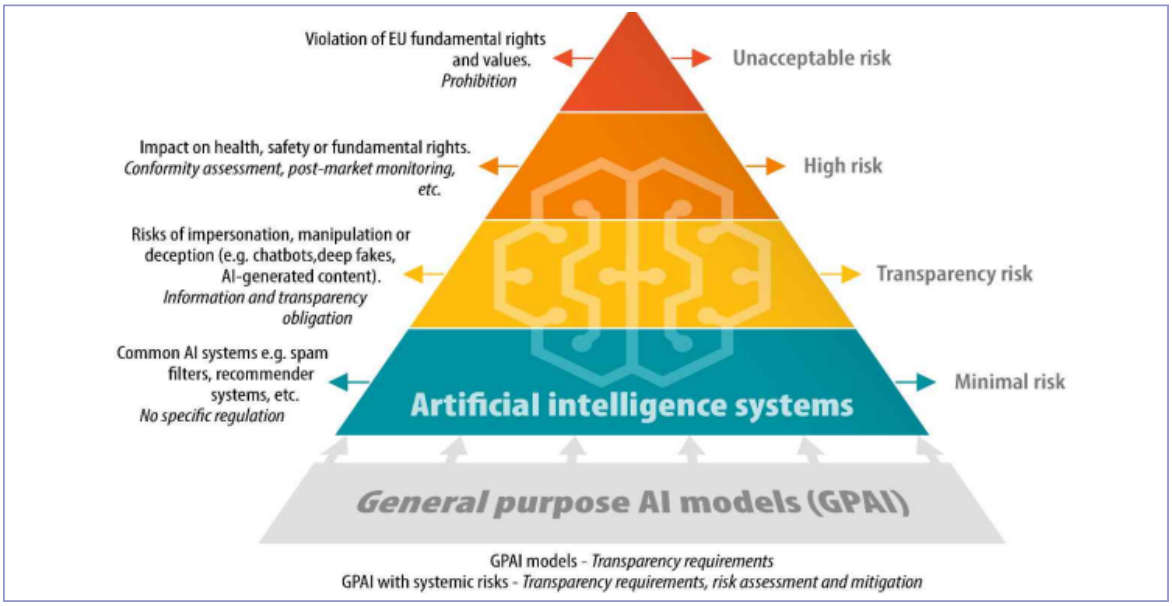
■ **초거대 AI 시대에 변화하는 AI 정책의 특징에 주목할 필요**

- AI 정책의 무게 중심이 진흥에서 규제와의 조화로 이동하고 안전과 신뢰에 관한 정책이 강조
- 주요국들은 변화하는 AI 환경과 자국의 상황을 고려하여 차별화된 방식의 AI 정책을 수립하고 이행

■ **국내 AI 정책 수립 시 변화하는 AI 환경을 반영**

- 계류 중인 AI 법 도입 논의에 있어 변화하는 AI 환경과 정책 특성을 종합적으로 고려
  - AI 안전과 신뢰성 관련 글로벌 정책 동향을 검토하고 국내 상황에 맞추어 반영
  - 21대 국회에서 폐기된 AI 법이 안전과 신뢰성 측면에서 문제가 제기되었던바 EU 등 다양한 AI 시스템의 위험분류 체계를 검토하고 한국의 현실에 맞게 반영

그림 7 EU AI 법의 위험분류

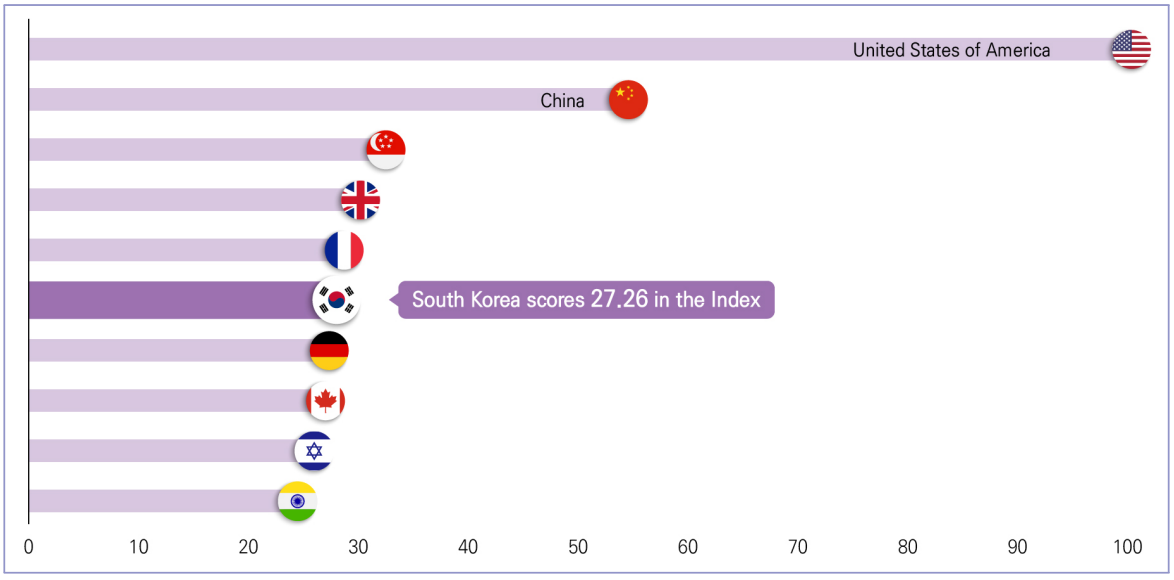


자료: European Parliament Research Service(2024)

○ 국가별 초거대 AI 경쟁력을 고려한 AI 정책을 수립

- 2020~2023년에 초거대 AI 모델을 가장 많이 개발한 국가는 미국(64건)이며, 중국(42건), 한국(11건), 프랑스(6건), 영국(5건) 순이고, 2023년 당해 기준으로 미국(41건), 중국(37건), 한국(8건), 프랑스(5건), 일본(3건) 임
- 초거대 AI 모델 보유 수가 상대적으로 적은 EU의 경우 AI 규제 강도 높고, 영국의 경우 규제 친화적인 접근으로 정책을 수립
- 글로벌 AI 지수 기준, 1위 미국 100점 기준으로 중국은 53.88점, 한국은 6위로 27.26점이며 선도국가와의 실질적인 격차는 매우 큰 상황
- AI 규제 강도와 범위 설정 시, 한국의 경쟁력을 종합적으로 고려하고 국회, 중앙정부, 지자체의 AI 활용방안 모색

그림 8 글로벌 AI 지수 순위



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

○ 국내 AI 안전연구소 설립 및 운영 관련 글로벌 사례를 참고하고 글로벌 정책 공조체계를 강화

- 미국, 영국, 일본 등 주요국은 AI 안전연구소를 설립하고 AI 안전성 평가 체계와 테스트 방법론을 개발하고 글로벌 공조체계를 형성 중

○ 중앙정부와 지자체 AI 정책의 유기적 연계 방안을 모색

- 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 유도

■ **동태적 관점에서 빠르게 변화하는 AI 생태계를 관찰하고 미래를 준비**

- AI 기술이 빠른 진화로 AI 정책의 필요성과 방향성도 변할 수 있어 관련 생태계 모니터링을 강화
  - 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 의미
  - 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 지적[17)]
- AI가 기존 서비스의 보완, 효율성 제고를 넘어 완전히 새로운 혁신 비즈니스 모델을 만들고 있어 이로 인한 파급효과를 선제적으로 검토
  - 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 예고
- LLM(Large Language Model)이 LWM(Large World Model)로 진화하고 있는바[18)] 현재 2D 중심의 AI가 공간컴퓨팅과 연계된 3D 시뮬레이션으로 연동되면 기존에 없던 새로운 혁신과 위험이 발생할 가능성
- AI 위험에 대한 이해관계자의 견해가 다르며 이에 갈등관리 체계를 마련하고 절차를 투명하게 공개해 숙의의 과정을 통해 정책 방안을 논의
  - 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 반대
- 미래지향적 AI 정책 수립에 필요한 다학제 연구를 강화

17) TechCrunch(September 18, 2024), "This Week in AI: Why OpenAI's o1 changes the AI regulation game"
18) Forbes(Jan 23, 2024) "The Next Leap In AI: From Large Language Models To Large World Models?"

참고문헌

법제처(2024), "인공지능 관련 국내외 법제 동향"

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

최병선(1992)『정부규제론』. 서울: 법문사

KOTRA(2024) "일본의 AI 정책과 실제 사례"

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

NIA(2023) "영국 AI 규제백서의 주요 내용 및 시사점"

SW 정책연구소(2024) "해외 AI 안전연구소 추진현황과 시사점"

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

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

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

Executive Order 14110: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

European Parliament Research Service(2024.3.), "Artificial intelligence act", Briefing document of EU legislation.

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

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

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

総務省, **経済産業省**(2024) "AI 事業者ガイド ライン案"

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

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

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

https://trackingai.org/IQ

국민을 지키는 미래로 나아가는 국회

ISSN | 2733-8258

이 자료는 **국회미래연구원 홈페이지**(www.nafi.re.kr) 및
열린국회정보(open.assembly.go.kr)에서 확인하실 수 있습니다.

국회미래연구원
NATIONAL ASSEMBLY FUTURES INSTITUTE