

새로운 디지털 질서
한국 시 거버넌스와 공동번영의 청사진

새로운 디지털 질서

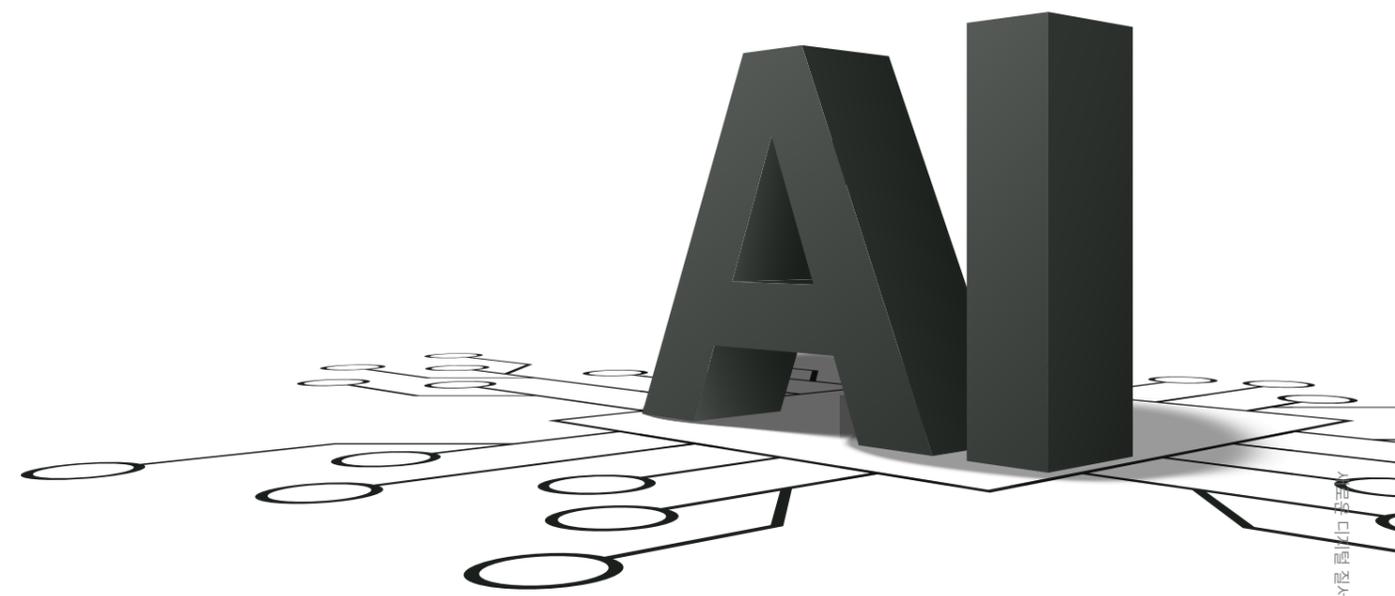
한국 시 거버넌스와
공동번영의 청사진

사사 표기

이 보고서는 한국지능정보사회진흥원(NIA)과의 발간 협력을 통해 작성되었습니다.
또한 정보통신정책연구원(KISDI)의 문정욱 지능정보사회정책연구실장님,
한국정보통신기술협회(TTA)의 이강해 AI융합기획단장님,
서울대학교 인공지능정책 이니셔티브(SAPI)의 임용 교수님의 자문을 받아
작성되었습니다.

새로운 디지털 질서

한국 시 거버넌스와
공동번영의 청사진



목차



• **서론** 006

• **1부** 010

책임있는 AI 거버넌스를 기반으로 한 AI 개발과 활용,
디지털 공동번영사회 실현의 필수 요소

- 013 AI 산업, 상생과 공동번영의 목표를 달성하기 위한 한국의 핵심 산업으로 부상하다
- 014 「디지털 권리장전」, 안전하고 신뢰할 수 있는 AI 거버넌스 구축을 논의하는 출발점을 제시하다

• **2부** 017

디지털 공동번영사회를 향한 한국, 국제사회,
테크기업의 유기적 여정

- 018 '책임있는 AI, 포용적 AI, 지속가능한 AI'라는 세 가지 주제 아래 국제사회 및 테크기업의 대응 방식과 맥락을 같이하는 한국의 원칙 기반 AI 거버넌스 접근방식
- 019 책임있는 AI에 대한 한국의 접근
원칙 3: 안전하고 신뢰할 수 있는 디지털 사회
원칙 4: 자율과 창의 기반의 디지털 혁신의 촉진
- 033 포용적 AI에 대한 한국의 접근
원칙 1: 디지털 환경에서의 자유와 권리 보장
원칙 2: 디지털에 대한 공정한 접근과 기회의 균등
- 044 지속가능한 AI에 대한 한국의 접근
원칙 5: 인류 후생의 증진

• **결론** 050

• **부록** 052

- 052 부록 A. 국내 AI 거버넌스 현황
- 060 부록 B. AI 거버넌스의 최근 국제 동향
- 068 부록 C. 한국의 「디지털 권리장전」과 국제 AI 정책 동향 교차 검토
- 074 부록 D. 디지털 공동번영사회를 위한 안전하고 신뢰할 수 있는 AI 시스템 구축을 위한 Microsoft의 모범 사례
- 087 부록 E. AI 산업이 한국 경제에 미칠 잠재적 영향에 대한 사례 연구

• **참조** 094



서론



조원우,
한국마이크로소프트
대표이사

한국의 AI 대전환

한국은 기술 혁명의 최전선에서 AI 분야의 혁신과 책임을 융화시키며 새 지평을 열고 있습니다. 인류 역사의 새로운 장에 들어선 지금, 우리는 단순히 기술만 발전시키는 것이 아니라 우리 사회의 정신과 국제 목표에 부합하는 미래를 설계하는 방식으로 접근해야 합니다.

AI 일상화를 목표로 빠르게 나아가고 있는 한국은 벌써 사회 곳곳에서 AI 혁신의 사례가 두드러지고 있습니다. 한국의 주요 산업은 적극적으로 AI 기술을 도입하여 혁신의 변곡점에 들어섰으며 Microsoft는 이 여정에 굳건한 핵심 파트너로서 함께하고 있습니다.

혁신 1) Microsoft는 한국의 초고령화에 대응하기 위한 노인 의료 혁신에 힘쓰고 있습니다. 서울대 신약개발융합연구센터 이형기 교수는 Microsoft와 협력을 통해 의약품의 기존 임상시험 정보로부터 새로운 치료 용도를 발견하는 알고리즘과 AI 기반 임상시험 설계 개선 방안을 연구 중입니다. 이는 의약품의 용도 변경 프로세스를 간소화하여 비용 효율성을 극대화하고 임상시험의 정확성을 향상하는 데 유의미한 성과를 내고 있습니다.

혁신 2) 또한 Microsoft는 이화여대 글로벌 AI신약개발 연구센터 최선 교수가 주도하는 구조 기반 약물 설계를 위한 딥러닝 연구를 지원하고 있습니다. 이 연구는 단백질 및 화합물 데이터를 AI로 분석하여 목표 특성을 가진 분자와 결합 친화도가 높은 화합물을 빠르게 선별하는 데 성공했습니다. 이와 같이 Microsoft의 국내 AI 연구 협력은 헬스케어 분야에서 신약 개발 시간과 비용을 현저히 줄이고 약물의 의학적 효능을 높여 노인성 질환에 대한 국가 의료 재정 부담을 완화하는 데 기여하고 있습니다.

혁신 3) Microsoft가 국내 연구 혁신에 남긴 가장 최근 발자취로는 2024년 3월, 한국과학기술원(KAIST)이 발표한 AI가 교육 분야에서 가진 가능성을 탐구하는 프로젝트를 들 수 있습니다. 오혜연 교수는 2023년 4월부터 Microsoft 연구소의 AFMR(Advancing Foundation Models Research) 이니셔티브의 지원을 받아 AI의 교육적 이점을 활용하면서도 학업성취 과정에서 AI가 오용될 수 있는 위험을 줄이기 위한 연구를 진행하였고, 그 연구의 결과물로 챗봇을 개발하였습니다. Microsoft 연구소는 이 이니셔티브를 통해 다양한 학문 분야의 연구자들에게 고급 AI 모델에 대한 접근을 제공하여 파운데이션 모델이 다양하게 개발되고 활용될 수 있도록 투자하고 있습니다. 오혜연 교수의 연구팀이 개발한 챗봇은 영어가 모국어가 아닌 학생들의 영자문을 돕도록 설계되어, 대리 작성을 해주는 것이 아니라 단계별 가이드를 통해 영어 글쓰기 훈련을 제공합니다. 한 학기 동안 213명의 EFL(English as a Foreign Language) 학생들이 챗봇을 이용해 영자문을 연습하고 에세이를 교정하는 등 글쓰기 능력을 향상했으며, 이 프로젝트는 교육 분야에서 생성형 AI가 유의미한 가치를 창출해 낼 수 있음을 여실히 보여주었습니다. 특히 이 프로젝트의 성공은 파운데이션 모델을 활용해 학생들의 문제해결 과정을 더욱 심층적으로 분석하고 이해하는 데에도 도움이 되었습니다.

이렇듯 AI는 이미 바이오, 헬스케어 분야 뿐 아니라, 반도체, 교육, 배터리, 금융, 통신, 제조업, 콘텐츠 등 국내 주요 산업 전반에서 새로운 가능성과 효과를 선보이고 있습니다.

한국은 AI가 가능케 할 과학기술 혁신 및 사회경제적 변혁의 혜택을 포착하기 위해 가보지 않은 길에 대한 개척과 도전을 발빠르게 개시하고 있습니다. 이러한 혁신을 향한 여정에 필수불가결한 것이 초거대 클라우드 인프라입니다. 초거대 클라우드 인프라만이 신속한 혁신과 대규모 AI 적용에 꼭 필요한 확장성과 유연성을 제공하기 때문입니다.

한국의 AI 거버넌스

한국은 급속한 기술 발전과 변화에 대응해 디지털에서 AI와 같은 기술의 사용과 발전을 지침하기 위한 중요성을 강조하는 일련의 보편적인 원칙을 세계 최초로 발표한 국가중 하나였습니다. 「디지털 공동번영사회의 가치와 원칙에 관한 헌장: 디지털 권리장전」은 AI에 대한 국가의 입법과 규제 그리고 운영 방침에 영향을 미치는데, 이 AI 비전의 기본 틀은 책임, 포용성, 지속가능성의 원칙에 기반한다고 볼 수 있습니다.

Microsoft의 자체 AI 원칙 역시 이 헌장의 핵심 원칙 및 청사진과 매우 일치합니다. 특히 Microsoft는 책임있는 AI에 중점을 두고 투명성, 공정성, 윤리 및 책임을 강조하며 AI 산업의 사회적 영향과 개인정보 보호를 최우선으로 고려하고 있습니다. 또한 포용적인 AI를 추구하여 모두에게 공정한 접근과 혜택을 보장합니다. 더 나아가 Microsoft는 지속가능한 AI에 헌신하여 AI를 인류 발전에 활용하는 동안 그 부정적인 영향은 최소화할 수 있도록 노력하고 있습니다.

궁극적으로 Microsoft는 활기차고 개방된 시장에서 AI를 통해 이뤄낼 공동 번영을 지향합니다. 이와 관련해 2024년 3월, 모바일 월드 콩그레스(MWC)에서 'AI 액세스 원칙'을 공개하며 책임있는 AI에 대한 Microsoft의 약속을 공표했습니다. 이번 원칙은 AI 모델 및 애플리케이션 개발자에 대한 액세스와 지원 제공, AI 경제 전반에 선택권과 공정성 보장, 사회적 책임 충족이라는 세 가지 주요 주제 아래 11개의 약속으로 구성되어 있습니다.

Microsoft AI 액세스 원칙

모델과 애플리케이션을 만드는 AI 개발자를 위한 기술 접근 및 지원 제공

- 1 Microsoft는 칩 용량 증가를 활용해 클라우드 컴퓨팅 AI 인프라를 확장하고 있습니다. 이를 통해 독점 또는 오픈 소스 여부, 규모에 상관없이 더 많은 AI 기반 모델을 학습하고 배포할 수 있도록 지원하고 있습니다.
- 2 Microsoft는 전 세계 소프트웨어 애플리케이션 개발자가 AI 모델과 개발 도구를 폭넓게 활용할 수 있도록 지원해, 모든 국가가 자체적인 AI 경제를 구축할 수 있도록 노력하고 있습니다.

AI 경제 전반에 선택권과 공정성 보장

- 3 Microsoft는 개발자들이 공개 API를 통해 Microsoft Azure에서 호스팅 되는 AI 모델에 액세스하고 활용할 수 있도록 지원하고 있습니다.
- 4 네트워크 사업자가 소프트웨어 개발자를 지원할 수 있도록 공통 공개 API를 제공하고 있습니다.
- 5 개발자는 AI 모델, 도구, 애플리케이션 등을 배포하고 판매하는 방법을 선택할 수 있습니다. 또한, Microsoft Azure 마켓플레이스(Azure Marketplace)를 통해 고객에게 직접 판매하는 것도 가능합니다.
- 6 Microsoft는 개발자의 AI 모델 훈련, 구축, 배포 또는 사용에서 얻은 비공개 정보 또는 데이터를 Microsoft Azure에서 해당 모델과 경쟁하기 위해 사용하지 않도록 함으로써 개발자의 요구를 존중합니다.
- 7 Microsoft Azure를 사용하는 고객이 데이터를 간편하게 내보내고 전송할 수 있도록 지원해 다른 클라우드 공급업체로 전환할 수 있도록 합니다.

사회적 책임 이행

- 8 Microsoft AI 데이터센터에서 실행되는 모든 AI 모델과 애플리케이션의 물리적 및 사이버 보안 요구사항을 지원합니다.
- 9 Microsoft는 사람을 AI 설계 결정의 중심에 두고 있으며, 강력한 책임 있는 AI 표준을 적용하고 있습니다. 이는 공정성, 신뢰성 및 안전성, 개인정보 보호 및 보안, 포용성, 투명성, 책임성 등 지속적인 가치를 존중하는 것을 의미합니다.
- 10 전 세계에 AI 기술을 널리 확산하기 위한 이니셔티브에 투자하고 있습니다.
- 11 Microsoft는 환경 변화에 대응하기 위해 다양한 방식으로 AI 데이터센터를 관리하고, AI를 활용하여 환경 지속 가능성 요구 사항을 발전시키고 있습니다.

'AI 액세스 원칙'은 이 산업을 혁신하고 이끄는 위치에 선 Microsoft가 우리의 역할과 책임을 구체화하는 데에 근본적인 토대가 되어줄 것입니다. 이 원칙은 본격적인 변화를 나타내며 Microsoft 역시 가보지 않은 길에 대한 전례 없는 투자, 견고한 비즈니스 파트너십 구축, 그리고 혁신과 경쟁을 육성하기 위한 광범위한 활동을 이행할 것을 의미합니다. Microsoft는 AI의 잠재력이 기업 단위를 초월해 전 세계 사람들의 삶에 파고들 것을 명확히 인지하고 있으며, 이러한 원칙을 명시하고 기술에 대한 접근을 제공함으로써 각국의 기관과 개인이 AI를 더 큰 선을 향해 활용할 수 있도록 기여할 것입니다. 현재 진행 중인 AI 데이터센터에 대한 투자나 역량 강화 프로그램 역시 'AI 액세스 원칙'을 실천하기 위한 Microsoft의 노력의 일환입니다.

2024 서울 AI 안전성 정상회의를 준비하며

다가오는 5월 서울에서 개최될 AI 안전성 정상회의에 앞서, 우리는 AI 분야에서 이뤄진 발전과 책임있는 개발 및 배포의 중요성을 되새길 필요가 있습니다. 이번 정상회의는 다양한 분야의 리더 및 정책결정자들이 모여 미래의 AI 발전을 안내할 지침 원칙을 논의하는 플랫폼으로 기능할 것입니다.

Microsoft는 국제사회와의 공조를 통해 AI 안전 거버넌스를 구축하려는 한국 정부의 목표를 적극 지지합니다. 협력을 기반으로 한 접근을 통해서야 전세계적으로 높은 정합성을 갖춘 AI 정책을 개발할 수 있기 때문입니다. 한국은 다가오는 2024 서울 AI 안전성 정상회의에서 다른 국가들, 특히 아시아태평양 지역의 동료 국가들과 긴밀히 소통하고 독려하면서 국제적 규제 일관성을 추구하고 있으며, 이를 통해 AI 거버넌스 분야의 규범적 리더로 자리매김할 기회를 얻었습니다. 성공적인 정상회의 개최를 통해 국경과 산업을 가로질러 신뢰, 협력 및 혁신을 지휘한다면, 한국은 아시아태평양 지역에서 단일한 AI 거버넌스 중심지로 거듭날 것입니다.

마무리하며

Microsoft는 한국지능정보사회진흥원(NIA)과의 협력을 통해 본 보고서를 발간할 수 있음에 심심한 사의를 표하며, 본 보고서가 「디지털 권리장전」에 내재된 핵심 원칙과 Microsoft의 'AI 액세스 원칙'에 대한 신념을 보여줄 것으로 기대합니다. 한국과 Microsoft의 AI 기본원칙은 밀접히 맞닿아 있습니다. 이 파트너십과 같은 글로벌 AI 영역에서의 지속적인 활동은 AI 거버넌스의 글로벌 규범에 영향을 미치는 선구적인 리더로서 한국의 위치를 증명하는 것이기도 합니다.

본 보고서는 공동번영으로 나아갈 책임있는 AI 거버넌스의 청사진입니다. AI 거버넌스 분야에서 한국의 진전을 포괄적으로 탐구했으며 책임있는, 포용적이고 지속가능한 기술 혁신을 전 세계적으로 촉진하기 위한 Microsoft의 꾸준한 노력을 담아냈습니다.

디지털 프론티어 대해 시대에 들어섰습니다.

AI를 인류 최상의 이익을 위한 최상의 도구로 활용합시다.



조원우
한국마이크로소프트 대표이사

1부

책임있는 AI 거버넌스를 기반으로 한 AI 개발과 활용, 디지털 공동번영사회 실현의 필수 요소



Microsoft가 새롭게 발표한 'AI 액세스 원칙'은 책임있는 AI 혁신을 촉진하고 접근성을 높이면서도 안전, 보안 및 개인정보 보호와 같이 AI와 관련된 기타 문제를 효과적으로 고려하기 위한 중요한 한 걸음으로, 아래의 세 가지 핵심 원칙으로 구성되어 있습니다:

- 모델과 애플리케이션을 만드는 AI 개발자를 위한 기술 접근 및 지원 제공
- AI 경제 전반에 선택권과 공정성 보장
- 사회적 책임 이행

Microsoft는 개발자와 사용자 모두에게 필요한 지원을 제공하고 AI 기술에 대한 폭넓은 접근성을 유지하는 것이 중요하다고 믿습니다. 이러한 신념 아래, 'AI 액세스 원칙'은 클라우드 컴퓨팅과 AI 인프라를 확충하고 Azure 같은 플랫폼, 도구, 서비스를 통해 AI 모델의 개발과 배포를 촉진하며, 규모에 상관없이 다른 개발자들이 AI 모델을 훈련하고 배포할 수 있도록 돕고 협력해 나갈 예정입니다. Microsoft는 AI 도구와 자원에 대한 접근 장벽을 제거함으로써 전 세계 다양한 커뮤니티가 AI 기반 경제 성장을 도모하고 협력에 기여할 수 있도록 하고자 합니다.

'AI 액세스 원칙'의 또 다른 핵심은 AI 산업 시스템 내 선택과 공정성을 보장하는 것입니다. 이에 따라 Azure에서 호스팅되는 모든 AI 모델은 공개 API를 통해 접근할 수 있으며 네트워크 사업자를 위한 공통 공개 API 역시 제공하며, 모든 개발자는 다양한 루트로 자신의 AI 모델, 도구, 애플리케이션 등을 배포하고 판매할 수 있습니다. 이러한 방법으로 투명성, 책임, 공정성을 촉진함으로써 Microsoft는 경쟁력 있고 공정한 AI 생태계 발전에 힘을 예정입니다.



마지막으로 Microsoft는 AI 분야와 기술 혁신의 선두주자로서 해야 할 역할과 책임을 다하고자 AI 모델과 애플리케이션의 물리적 및 사이버 보안을 보장하고, 책임있는 AI 표준을 준수하며, AI 관련 역량 강화 프로그램에 투자하고 환경의 지속가능성을 고려한 윤리적인 AI 개발 및 배포를 약속합니다. 이러한 노력은 Microsoft가 사내 AI 관련 사업을 사회적 책임, 지속가능성, 윤리적 거버넌스라는 원칙에 부합하는 방식으로 진행하여 사회에 긍정적으로 이바지하고자 하는 신념을 증명합니다.

AI가 가진 가능성이 계속 확장됨과 함께 이 기술의 운용 방침을 명확히 해야 할 필요성 역시 대두되었고, 근 몇 년간 전 세계 규제 당국들은 다양한 AI 거버넌스 원칙을 정의해왔습니다. 이러한 원칙들이 모여 AI 생태계의 책임과 투명성을 확보하기 위한 집단적 노력을 더욱 전진시키고 있습니다. 혁신 강국인 한국 역시 시가 초래할 수 있는 사회적 우려에 대응하면서도 AI 산업의 성장을 지원하는 원칙과 포괄적인 거버넌스 체계를 수립해 왔습니다. 한국 내 이러한 AI 정책 동향은 윤리적이고 책임있는 AI 개발 및 배포를 추구하고 혁신을 촉진하는 Microsoft의 AI 원칙과 많은 공통점을 가지고 있습니다. 이 장에서는 Microsoft가 강조하는 접근성, 공정성, 및 사회적 책임의 원칙과 공명하는, AI 시대에 혁신과 윤리적 고려 사항의 균형을 맞추려는 한국의 노력을 정리했습니다.

AI 산업, 상생과 공동번영의 목표를 달성하기 위한 한국의 핵심 산업으로 부상하다

2019년 초, 한국은 AI의 엄청난 잠재력을 인식한 뒤 국내 경제를 활성화하고 글로벌 AI 리더로서의 한국의 입지를 다지기 위해 최초의 AI 국가전략을 발표했습니다.ⁱ 2022년 9월에 '대한민국 디지털 전략'을 발표한 이후, 정부는 AI 기술 및 산업 발전을 위한 다양한 전략을 지속적으로 수립해 왔습니다. 이러한 정책 방향은 AI 발전이 함의하는 전략·외교·과학·경제적 잠재력을 인정하고 있음을 보여줍니다.ⁱⁱ 예를 들어, 한국 정부는 AI를 12대 국가전략기술 중 하나로 선정했습니다. 그 이유는 시가, 인터넷이나 스마트폰이 사회의 기술, 경제, 외교안보 발전에 미친 영향과 유사한 결과를 가져올 수 있는 필수 인프라라고 보았기 때문입니다. AI는 경제성장을 견인하기 위한 정부의 '신성장 4.0 전략'에서도 중요하게 언급되고 있습니다.

2023년 4월에 발표된 '초거대 AI 경쟁력 강화 방안'에서도 일상생활에 AI를 광범위하게 접목하려는 포부를 밝혔습니다. 이러한 야심 찬 목표는 세계적으로 인정받은 한국의 디지털 경쟁력에 기반합니다. 스위스 국제경영개발대학원(IMD)이 실시한 2022년 세계 디지털 경쟁력 평가 결과에서 한국은 평가 대상 63개국 중 8위를 기록했고 아시아 태평양 국가 중에서는 싱가포르에 이어 2위를 차지했습니다.ⁱⁱⁱ 과학기술정보통신부(이하 과기부)는 2023년 8월, 2024년도 AI 관련 연구개발(R&D)에 7,331억 원(5.6억 미국 달러)을 투자하여 이 분야에서 한국의 기술 주권을 강화하겠다는 계획을 발표했습니다.^{iv} 국내 AI 산업은 향후 5년간 연평균 14.9%의 성장률을 보이며 2027년에는 4.46조 원(34억 미국 달러) 규모에 달할 것으로 전망되고 있습니다.^v 또한 최근 Microsoft와 대한상공회의소가 공동으로 발간한 보고서에 따르면, 생성형 AI는 제조부터 금융과 교육에 이르기까지 다양한 산업에서 인력을 혁신적으로 향상함으로써 2022년 국내총생산(GDP)의 1/4에 해당하는 620조 원(4,763억 미국 달러)의 생산성을 창출할 것으로 추정됩니다.^{vi} 한국 정부는 이러한 성장 모멘텀을 유지하면서 잠재적인 위험을 효과적으로 해결하여 '인간과 AI의 공존'을 실현하기 위해 노력하고 있습니다.^{vii}

AI 산업이 한국의 성장과 번영에 미치는 영향력을 고려할 때, AI 거버넌스는 다양한 정부 기관과 단체의 의제에서 그 중요도가 커지고 있습니다. 한국은 성장 동력을 저해하지 않으면서도 AI에 수반되는 리스크를 효과적으로 해결하기 위해 선제적으로 신중한 행보를 보이고 있는데, 이는 2023년 11월 영국 AI 안전성 정상회의에 참석한 이종호 과기부 장관의 연설을 통해서도 알 수 있습니다.^{viii} 당시 이종호 장관은 "AI 발전이 위축되지 않게 하는 범위에서 적절한 AI 신뢰성·안전성을 확보하는 것이 정책입안자의 역할"이라고 거듭 강조했습니다. 과기부를 중심으로 관계 부처에서는 자율 가이드라인과 전략 등을 공표하며 책임있는 기술 활용을 촉진하는 AI 거버넌스 내 각 부처의 소관을 적극적으로 규정하고 있습니다.

이 접근방식의 핵심은 학계 및 업계 전문가와 소통하고 사회 전반의 다양한 목소리를 반영하기 위한 노력을 바탕으로 책임있는 규제의 균형을 추구하는 것입니다. 한국은 새로운 규제나 의무를 부과하는 대신 기존 규제 프레임워크나 시스템에 AI 거버넌스의 구성 요소를 통합하는 방식을 택했습니다. 한국은 아시아 태평양 지역 국가 중 AI 관련 준비도가 가장 우수한 나라로서^{ix} 유연하고 포괄적인 거버넌스 프레임워크를 통해 시민과 AI 산업을 보호하면서 AI 혁신을 이뤄가고 있습니다. AI의 잠재력을 최대한 활용하는 동시에 그 영향을 신중하게 관리하여 국내 기술 발전과 경제 성장을 견인할 수 있도록 국가 차원에서 좀 더 유익한 디지털 생태계를 조성하려는 것입니다.





「디지털 권리장전」, 안전하고 신뢰할 수 있는 AI 거버넌스 구축을 논의하는 출발점을 제시하다

2023년 9월, 과기부가 발표한 「디지털 권리장전」은 AI 거버넌스에 대한 포괄적 접근방식을 정립하기 위한 발판이라고도 볼 수 있습니다. 이 헌장은 한국 국회가 국가 AI 정책의 총체적인 법적 틀을 심의할 때 AI 거버넌스에 대한 초기 경계를 설정하고 참고할 수 있는 지침이 될 것입니다.

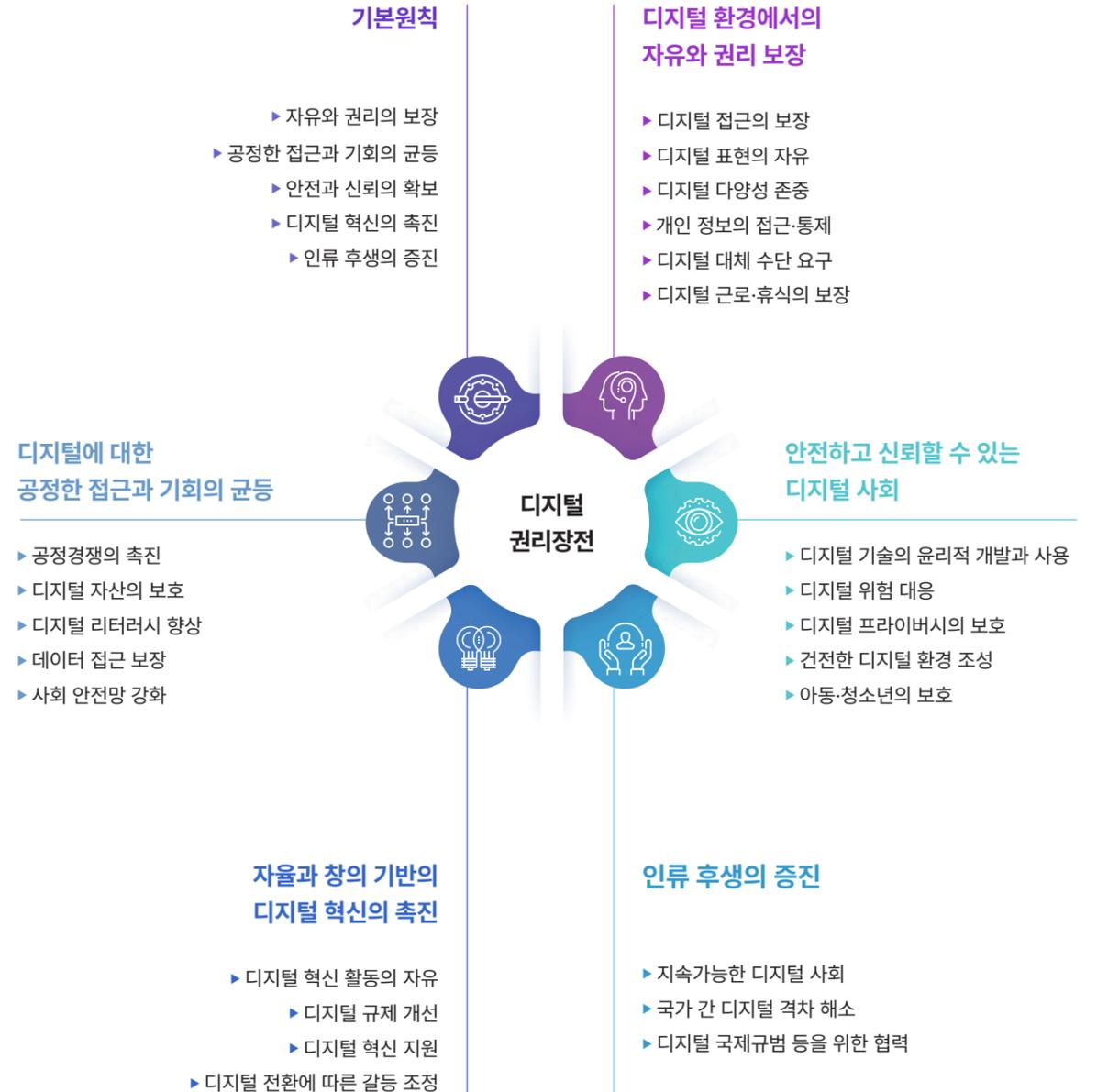
2023년 9월 뉴욕대학교에서 개최된 디지털 비전 포럼에서 윤석열 대통령은 「디지털 권리장전」의 5대 기본원칙으로 자유, 공정, 안전, 혁신, 연대를 제시했습니다.^x 국제 보편성을 갖춘 규범으로 구성된 이 헌장은 10월의 UN 글로벌 디지털 콤팩트 아태지역 회의, 11월의 OECD 디지털 권리 워크숍 등 다양한 국제 포럼에서 공유되기도 했습니다.^{xii} 2023년 11월 영국에서 열린 AI 안전성 정상회의에서도 윤 대통령은 국제사회 리더들에게 5대 기본원칙을 AI와 디지털이 주도하는 새 시대의 핵심 가치로 받아들여 줄 것을 요청했습니다.^{xiii} 이는 한국이 AI 분야의 주요 글로벌 플레이어로서 국제 규범 수립에 주도적인 역할을 하겠다는 의지를 보여준 것입니다.

과기부는 2023년 9월 25일, 「디지털 공동번영사회의 가치와 원칙에 관한 헌장: 디지털 권리장전」을 공식적으로 발표했습니다.^{xiii} 이 헌장은 AI와 같은 기술이 빠르게 사회를 재편하고 있는 디지털 시대에 규제 당국과 민간 기업이 AI 관련 전략을 추진할 때 참고할 수 있는 지침서의 역할을 합니다. 구속력은 없지만 정책입안자와 기업이 혁신과 공정성과 신뢰성이 결합된 디지털 세상을 구축하기 위한 기본원칙을 설정하는 데 있어 디지털 헌법과 같은 역할을 하는 것입니다. 정부는 기존 규제를 정비하고 '인공지능 산업 육성 및 신뢰기반 조성'에 관한 법률(AI법), 「디지털 포용법」¹ 등과 같은 새로운 규제를 마련할 때 「디지털 권리장전」에 담긴 가치를 우선적으로 반영할 예정이라고 밝혔습니다.

미국의 AI 권리장전을 위한 청사진 등 세계의 유사한 정책적 시도보다 한국의 「디지털 권리장전」이 좀 더 포괄적입니다. 디지털 리터러시, 디지털 격차, 국제 연대와 협력 옹호 등 AI를 넘어선 광범위한 디지털 이슈를 다루고 있기 때문입니다. 이 헌장을 통해 한국은 AI뿐만 아니라 미래 기술이 발전함에 따라 나타나는 사회경제적 영향과 그 반작용에 대응할 수 있는 원칙 기반의 사고방식을 제안했습니다. 한국 정부는 헌장의 원칙에 기반한 포괄성을 명확히 제시하여 다른 국가들도 이를 채택하고 적용하도록 규범적 모델로 홍보하고 있습니다. 한국 정부는 디지털 윤리와 AI 거버넌스에 대한 글로벌 합의가 필요함을 지속적으로 강조해 왔으며, 디지털 집약적 사회에서 공동번영을 실현할 수 있는 새로운 패러다임을 지지해 왔습니다. 「디지털 권리장전」은 이러한 패러다임에 대해 한국의 입장을 밝힌 '답안'이라고 할 수 있습니다.

「디지털 권리장전」은 총 6개의 장으로 구성되어 있으며, 5대 기본원칙하에 28개 조가 담겨 있습니다. 「디지털 권리장전」이 제시하는 디지털 공동번영사회의 위한 청사진은 다음의 표와 같습니다.

1 디지털 포용성이란 모든 사람이 장벽 없이 IT 기기와 서비스를 이용하고 정보를 습득할 수 있도록 접근성을 개선하는 것을 말합니다. 현재 국회에는 2021년 1월과 2022년 11월에 각각 제출된 두 건의 「디지털 포용법」 제정안이 소위원회 심사를 기다리고 있습니다.



국내 전문가들은 한국 정부가 원칙에 기반한 이 가이드라인으로 디지털 기술로 인한 즉각적이고 상호 연결된 위험에 대처하고자 취한 이 균형 잡힌 접근방식을 높이 평가하고 있습니다. 이러한 원칙 기반의 규제는 규제 기관은 물론이고 민간 기업이 유연하고도 발빠르게 대응할 수 있도록 합니다. 혁신의 모멘텀을 훼손하지 않는 한국의 이러한 신중한 접근방식은 다음과 같은 한국의 독특한 포지셔닝에서 비롯된 것입니다. 즉, 한국은 급성장을 경험하여 독특한 사회경제적 구조를 지니고 있고, 작지만 강한 글로벌 테크 리더로 부상 중인 AI 산업을 보유하고 있습니다.

한국은 세계 상위 수출국이자 높은 수준의 IT 및 제조업을 보유하고 있기 때문에 글로벌 표준과 규제의 변화에 발맞추는 것에 매우 큰 관심을 갖고 있습니다. 또한 외부에서 발생하는 규제로 인해 국내 기업과 산업이 불이익을 받지 않도록 국제 논의에 적극적으로 참여하는 것도 중요하게 여깁니다. 「디지털 권리장전」에 담긴 가치 기반의 원칙을 실제 적용하는 모범 사례의 대상을 찾아 국제 무대에서 지속적으로 교류한다면 한국은 AI 거버넌스에서 규범적 타당성과 영향력을 높이는 기회를 얻게 될 것입니다. 이는 결국 디지털 사회에서 한국 경제의 글로벌 경쟁력을 강화하고 흑역 발생할 수 있는 무역 분쟁을 예방하는 데도 도움이 될 것입니다.

“[AI 거버넌스에 관한 글로벌 논의에서] 우리나라에게 중요한 것은 어떻게든 의사결정 과정에, 그것도 가능하면 헤드 테이블에, 자리를 확보하는 것입니다. 그렇게 자리를 확보하기 위해서는 개별 국가가 구축할 거버넌스 시스템의 기반이 되는 프레임워크의 형성에 가치 있게 기여해야 합니다. 이 과정에서 글로벌 규범이 우리의 입맛에 맞게 또는 우리 규범에 맞춰지도록 만드는 것은 현실적인 목표가 아닙니다. 우리의 목표는 앞으로 형성될 글로벌 프레임워크가 보편성을 지니면서도 동시에 우리의 사회적 가치, 문화 그리고 경제적 이해관계와도 호환되도록 만드는 것이고, 이를 통해 우리나라 업계를 포함한 개별 주체들이 글로벌 경제 사회에서 계속 뻗어나갈 기회를 만들어주는 것입니다.”

- 임용 교수, 서울대학교 인공지능정책 이니셔티브 디렉터

따라서 한국이 후속 단계를 적절하게 추진하기 위해서는 「디지털 권리장전」 핵심 원칙의 함의를 AI 거버넌스의 맥락에서 추출하고, 구체적인 적용 방법을 확립하여 거버넌스 프레임워크가 관리하고자 하는 기본적인 리스크를 완화해야 합니다. 이 과정에서 이 원칙을 실현하기 위한 구체적인 조치는 AI 발전의 혜택을 누리면서도 개인을 보호할 수 있는 방식으로 개발되어야 합니다.

이러한 이해를 바탕으로 다음 장에서는 디지털 공동번영사회를 위한 5대 기본원칙이 지금까지 국내를 비롯해 국제사회와 테크기업의 AI 거버넌스 논의에 어떻게 반영되어 왔는지 살펴봅니다.

2부

디지털 공동번영사회를 향한 한국, 국제사회, 테크기업의 유기적 여정



‘책임있는 AI, 포용적 AI, 지속가능한 AI’라는 세 가지 주제 아래 국제사회 및 테크기업의 대응 방식과 맥락을 같이하는 한국의 원칙 기반 AI 거버넌스 접근방식

디지털 공동번영사회를 위한 국내 AI 거버넌스의 접근방식은 국제사회의 논의 동향에 부합하며 디지털 공동번영사회를 조성하기 위한 노력을 반영하고 있습니다. 한국은 AI 기술의 개발과 배포를 위해 원칙에 기반한 프레임워크를 채택했습니다. 이러한 접근방식은 AI의 적용을 확산할 때의 책임있는 접근과 신뢰성 그리고 인류 후생의 중요성을 강조합니다.

국제사회 및 테크기업과 함께하는 한국의 AI 거버넌스 전략은 ‘책임있는 AI, 포용적 AI, 지속가능한 AI’라는 세 가지 키워드를 기준으로 나눌 수 있습니다. 상술했듯이 한국은 AI의 혁신적 잠재력을 수용하는 동시에 사회의 안녕을 우선시하는 총체적인 접근방식을 취하고 있습니다.

3가지 키워드의 광범위한 가치 기반의 특성을 고려할 때, AI 정책 및 거버넌스 영역에서 각 원칙이 지닌 의미는 직관적이지 않을 수 있습니다. 이러한 문제의식하에, 디지털 공동번영사회라는 목표를 달성하기 위해 각 원칙이 어떤 다양한 정책적 대응과 이니셔티브로 나타나고 있으며, 한국 정부와 국제사회 그리고 테크기업에서 어떻게 추진되고 있는지를 분석했습니다. 그리고 시의성과 글로벌 규범 형성에 미칠 영향력을 고려하여 선정한 다음과 같은 국제사회 주요 정책 이니셔티브와 「디지털 권리장전」 원칙 간의 연관성을 분석하고 AI 거버넌스 분야에서 가치 기반 원칙을 구체화할 수 있는 인사이트를 도출하고자 했습니다.

- 01 미국 「백악관 AI 위험 예방 행정명령」 (US White House Executive Order on the Safe, Secure, and Trustworthy AI)
- 02 미국 국립표준기술연구소(NIST)의 AI 위험 관리 프레임워크 [US National Institute of Standards and Technology (NIST) AI Risk Management Framework]
- 03 영국 AI 안전성 정상회의 및 「블레츨리 선언」
- 04 AI 개발을 위한 「G7 기본 원칙」 (G7 Guiding Principles for Developing AI)
- 05 OECD 「AI 원칙」
- 06 EU 「AI법」

AI 거버넌스의 맥락에서 「디지털 권리장전」의 의미를 종합적으로 비교 분석하되, 특히 책임있는 AI, 포용적 AI 그리고 지속가능한 AI라는 핵심 주제와 연결되는 「디지털 권리장전」의 해석 부분을 강조하고자 했습니다. 이를 통해 「디지털 권리장전」이 국제 AI 거버넌스에서 논의되는 발전 방향과 상호작용하고 있음을 명시하고 개선 방향도 추가로 제시했습니다. 따라서 본 보고서는 AI 거버넌스 프레임워크의 수립과 개선에 적극 참여하는 국내 정책입안자에게 유의미한 통찰을 제공하는 것을 주요 목표로 합니다. 비교적인 관점에서의 탐색은 정책 입안자가 사실에 기반하여 의사결정을 내릴 때 도움이 되며, 국내의 상황을 고려할 때도 효과적인 AI 거버넌스를 구축하는 데 참고 자료로 활용될 수 있을 것입니다.

책임있는 AI에 대한 한국의 접근

책임있는 AI는 공정성, 투명성, 책임성을 강조하며 윤리적이고 양심적인 AI 기술의 활용을 수반합니다. 이 원칙은 AI 시스템이 사회적 영향을 충분히 고려하여 개발·배포되고, 편견을 최소화하며 공정한 결과를 끌어내도록 권장합니다.

「디지털 권리장전」은 책임을 우선시하는 방식으로 AI 기술을 채택하고 구현해야 함을 강조합니다. 특히 제3, 4 원칙은 개발 과정에 공정성, 투명성, 책임성을 도입하여 편견을 완화하고, AI 활용이 디지털 혁신을 촉진하는 동시에 사회에 도움이 되어 안전하고 신뢰할 수 있는 디지털 사회를 구현해 갈 것을 요구합니다. 또한 ‘기술의 발전은 신뢰를 기반으로 이뤄져야 한다’, ‘긍정적인 사회적 영향에 대한 고려가 필요하다’는 주장은 책임있는 AI에 대한 한국의 지지를 확인할 수 있는 부분입니다.

**원칙 3
안전하고 신뢰할 수 있는
디지털 사회**

“디지털 사회에서 디지털 기술과 서비스는 개인과 사회의 안전에 위협이 되지 않도록 신뢰할 수 있어야 하고, 디지털 위험에 대비하는 수단과 절차가 마련되어야 한다.”(제1장 제3조)

이 원칙은 다음 5가지 조항을 통해 더욱 자세히 이해할 수 있습니다.^{xiv}

- (제17조) 디지털 기술의 윤리적 개발과 사용
디지털 기술의 개발과 사용은 안전과 신뢰를 확보할 수 있도록 윤리적인 방식으로 책임 있게 이루어져야 한다.
- (제18조) 디지털 위험 대응
디지털 위험은 적절한 조치가 이루어질 수 있는 수단과 절차를 통해 예방·관리되어야 하며, 그 위험에 관한 정보는 알기 쉽고 투명하게 공개되어야 한다.
- (제19조) 디지털 프라이버시의 보호
디지털 환경에서 개인의 프라이버시는 디지털 감시, 위치추적 등을 비롯한 불법적인 식별과 추적으로부터 보호되어야 한다.
- (제20조) 건전한 디지털 환경 조성
허위조작 및 불법·유해정보의 생산·유통이 방지되는 등 건전한 디지털 환경이 조성되어야 하고, 디지털 환경에서 발생하는 범죄로부터 피해자를 보호하기 위한 실질적인 수단과 절차가 마련되어야 한다.
- (제21조) 아동·청소년의 보호
아동·청소년은 연령에 적합하게 설계된 디지털 공간을 선택하여 자유롭게 활동할 수 있어야 하며, 디지털 기술로 발생가능한 범죄로부터 특별히 보호받아야 한다.

특히 '윤리적 개발과 사용', '디지털 위험 대응', '디지털 프라이버시의 보호' 조항은 AI 정책에 대한 논의와 직접적인 관련이 있습니다. AI 거버넌스의 맥락에서 이 부분을 강화하기 위해 실제 논의되고 있는 구체적인 사례와 AI가 초래할 수 있는 잠재적인 위험에 대한 우려는 다음과 같습니다.

조항	AI 거버넌스 관련 적용 사례	근본적인 사회적 위험
(제17조) 디지털 기술의 윤리적 개발과 사용	<ul style="list-style-type: none"> AI 생애주기 전반에 걸쳐 개발자, 배포자, AI 시스템 운영자 간의 협력과 책임 공유 분야별과 단계별 가이드라인 투명한 정보 공개 절차 영향 평가 체계 첨단 AI모델 대상 라이선스 체계 	<ul style="list-style-type: none"> 편견과 차별, 사회 불안, 인력 이탈, 대중의 신뢰도 저하
(제18조) 디지털 위험 대응	<ul style="list-style-type: none"> 위험 통제와 관리 조치 사례별 평가 체계 위험 기반 사후 패널티 안정적인 운영을 보장하기 위해 관련 기술과 역량을 갖춘 인력 양성 지식과 모범 사례를 공유하기 위한 공공-민간 플랫폼 고위험 AI 시스템의 국가 등록제 	<ul style="list-style-type: none"> 해킹, 피싱, 랜섬웨어, 디지털 포용(디지털 문맹 사용자 보호), 국가 안보, 하이브리드 또는 사이버 전쟁과 같은 악의적인 목적을 위한 AI 시스템 악용
(제19조) 디지털 프라이버시의 보호	<ul style="list-style-type: none"> 강력한 데이터 보안 조치 명확한 침해 발생 보고 체계 정기적인 감사 및 평가를 통한 개인정보 위험 식별과 시정 조치 관련된 데이터 보호법 및 윤리 기준 준수 	<ul style="list-style-type: none"> 개인의 자율성, 법적·윤리적 의무, 스토킹과 같은 악의적인 목적으로의 기술 오용



한국

책임있는 AI는 국내에서 활발히 논의되고 있는 핵심 주제 중 하나입니다. 과기부는 2020년 12월 국가 인공지능 윤리기준을 처음으로 발표하고 인공지능의 핵심 요건으로 개인정보보호, 안전성, 투명성 등의 가치를 강조했습니다. 과기부는 이 기준을 바탕으로 산업계가 일상적인 관행과 프로세스에서 이 기준을 적용할 수 있도록 지원하고 있습니다. 이와 함께 이를 저해하는 규제나 구조적 문제에 대한 연구도 진행 중입니다. 이와 함께 과기부는 2022년 2월 학계와 산업계 그리고 시민사회 전문가로 구성된 AI 윤리정책포럼을 발족했습니다. 이 포럼은 ICT 정책 관련 정부 출연 연구 기관인 정보통신정책연구원(KISDI)과 협력해 윤리 자율점검표와 개발 가이드라인, AI 윤리 교육 교재와 강의 가이드 등을 제작했습니다.^{xv}

또한 과기부는 민간 주도의 AI 신뢰성-윤리 규제 프레임워크와 AI 영향평가 관리체계를 개발하고 있습니다. 1기 AI 윤리정책포럼의 후속으로 2기 포럼이 2023년 4월에 출범했습니다. 2기 포럼은 채용, 공공안전, 생성형 AI 활용 서비스 등 신뢰성과 윤리가 강조되는 분야에 종사하는 AI 시스템 개발자 및 사업자를 위한 분야별 자율점검표와 가이드라인을 2023년 12월까지 마련하는 것을 목표로 하고 있습니다. 영향평가 제도에는 위험 관리, 책임 준수, AI 제품 및 서비스의 영향에 대한 자발적 준수의 기준이 도입할 것으로 보입니다.

“인공지능 윤리가 사회 전반에 확산되고 자리 잡으려면 포용적 기술 개발, 개발자 윤리, 이용자 윤리라는 삼박자가 고르게 갖춰져야 하며, 모든 이해관계자가 디지털 기술의 위험으로부터 안전한 사회를 구현하기 위해 인간을 중심에 두고 함께 고민해야 합니다.”
- 최문실 한국지능정보사회진흥원(NIA) 디지털포용본부장

과기부는 국내 유일의 ICT 관련 표준 제·개정 및 보급 전담 기관인 정보통신기술협회(TTA)와 협력하여 AI 시스템 개발을 위한 가이드라인을 마련했습니다. 이 가이드라인은 AI 시스템 개발자와 운영자를 위해 설계된 검증 프레임워크를 모두 포함합니다. 과기부와 TTA는 이 프레임워크를 활용하여 국내 AI 산업에 신뢰성을 우선하는 표준화된 접근방식을 확산하고자 하며 이를 위해 프레임워크 적용에 관한 컨설팅과 교육을 제공할 예정입니다. 일례로 과기부에서 주관하는 연구개발 사업에 참여하기 원하는 기업과 기관은 참여 요건의 일부로 인증을 추진하는 것을 검토하고 있습니다. 2023년 10월부터 시행된 AI 신뢰성 검증 제도는 2023년 12월까지 시범 운영되고 있습니다.^{xvi}

“칼도 마찬가지로, 우리 일상생활에서 쓰는 모든 것이 다 그렇듯이 인공지능도 기술 개발은 계속 투자하고 매진을 하되, 이렇게 개발된 기술이 오용되거나 악용되지 않도록 가이드를 제공하는 게 중요합니다. ... TTA는 인공지능 윤리 10대 원칙 중 기술적으로 구현 가능한 4가지 원칙 - 다양성 존중, 안정성, 책임성, 투명성 - 을 인공지능 서비스 및 제품에 적용할 때 필요한 요구 사항을 개발안내서로 만들었습니다. 저희가 이런 개발 안내서를 만들고 트레이닝과 컨설팅을 지원하면서 [안전하고 신뢰가능한 AI 시스템 개발을 위해 필요한 요구 사항] 기업들에 주지시킴으로써 기업이 제품 및 서비스의 기획 단계서부터 자율적으로 신뢰성에 대한 기본 요건을 고려해 만들 수 있도록 제도를 만들고 있습니다. ... 표준화된 신뢰성 검증체계는 AI산업 육성 동력에 일조할 것으로 기대됩니다.”

- 이강해 정보통신기술협회(TTA) AI융합기획단장

한국 국회에서도 법률 개정을 통해 AI의 책임, 공정성, 신뢰, 투명성에 대한 고려 사항을 기존 규제 체계에 통합하고자 노력해 왔으며, 현재 진행 중인 노력은 부록 A에서 확인할 수 있습니다. 계류 중인 AI 법안은 좀 더 엄격한 공개 요건을 충족해야 하는 고위험 AI 시스템의 범위를 지정함으로써 위험 기반 접근방식을 취하고 있습니다.^{xvii} 또한 AI 기술이 시장에 진입한 후 필요에 따라 관련 규제를 변경할 수 있도록 하는 사후 규제 방식은 산업 성장을 촉진하려는 한국 정부의 열의를 다시 한번 확인할 수 있는 부분입니다.

AI 거버넌스에서 데이터 프라이버시 문제를 감독하는 개인정보보호위원회는 민감도와 위험도 평가에 따라 차등적인 컴플라이언스 조치를 취하는 규칙 기반에서 원칙 기반 규제 체계로 전환하기 위해 노력하고 있습니다.^{xviii} 한국은 AI 프라이버시 정책협의회와 같은 민간 협력을 통해 AI 시스템의 프라이버시와 데이터 보호에 관한 강력하고 실용적인 표준 개발에도 힘쓰고 있습니다.

이와 같이 한국은 안전하고 신뢰할 수 있는 디지털 사회의 기반이 되는 AI 거버넌스에 대해 민간 주도의 자율 규제를 수용하는 위험 기반 접근방식을 채택하여 추진하고 있습니다.

국제사회의 논의 현황

이러한 한국의 AI 거버넌스 방향은 국제사회의 흐름과 대부분 일치합니다. 안보, 안전, 위험 관리는 다른 나라의 AI 거버넌스 모델과 국제적으로 합의된 원칙에서 반복적으로 발견되는 주제입니다. 한국의 '국가 인공지능 윤리기준'의 모태이기도 한 OECD 'AI 원칙'은 신뢰할 수 있는 인공지능의 개발과 사용을 위한 가치 기반 원칙으로서 '건고성, 보안, 안전'을 제시하고 있습니다.^{xix}

최근에 OECD 'AI 원칙'에서 비롯된 G7 'AI 기본원칙'도 위험 기반 접근법을 강조합니다. G7 'AI 기본원칙'은 첨단 AI 시스템을 개발할 때뿐만 아니라 시장에 배포한 후에도 위험을 식별·평가·완화하기 위한 조치를 시행할 것을 권장합니다.^{xx} 특히 제5원칙은 위험 기반 접근법에 기반한 AI 거버넌스 및 위험 관리 정책을 명시적으로 요구하고 있으며, 제6원칙에서는 AI 생애주기 전반에 걸쳐 보안 통제·관리 체계를 수립할 것을 의무화하고 있습니다.

한국의 접근법과의 차이점으로 가장 눈에 띄는 특징은 위험 기반 접근방식에 생애주기에 대한 고려를 추가한 것입니다. 이는 AI생태계 내 이해관계자의 다양한 AI 시스템 사용 및 상호작용을 고려하고, 각 이해관계자와 가장 밀접하게 연관된 위험요인과 그에 따른 책임을 명확히 하기 위한 노력으로 보입니다.

미국의 '안전하고 안정적이며 신뢰할 수 있는 AI에 관한 행정명령'(이하 '백악관 행정명령')에는 AI 시스템의 잠재적 위험에 대비하기 위한 AI 안전 및 보안 표준 개발도 포함되어 있습니다.^{xxi} 이러한 요건에는 AI 안전을 보장하기 위한 표준과 도구와 테스트의 개발, 중요 소프트웨어의 취약점을 발견하기 위한 첨단 사이버 안보 프로그램에 대한 투자, '가장 강력한' 또는 고위험 AI 시스템에 대한 주요 정보 공유의 의무화, AI 생성 콘텐츠를 식별하고 AI를 이용한 사기로부터 시민을 보호하기 위한 콘텐츠 인증 지침 수립 등이 포함됩니다. 또한 AI로 인한 개인정보보호의 위험을 적절히 완화할 수 있도록 정부가 개인정보보호와 관련된 연구와 기술에 지원하는 방법도 자세히 명시하고 있습니다.



「백악관 행정명령」 외에도 미국은 2023년 1월 미국 국립표준기술연구소(NIST)의 주도로 AI 위험 관리를 강조하는 AI 위험 관리 프레임워크(AI RMF)를 마련했습니다.^{xxii} 민간/공공 부문과의 긴밀한 협력을 통해 개발된 AI RMF는 AI 피해 유형 3가지의 개요를 제공하고 신뢰할 수 있는 AI 시스템의 특성 7가지를 설명합니다. G7 기본원칙에 명시된 생애주기 기반 접근방식과 유사하게, AI RMF도 AI 위험을 효과적으로 관리하기 위해서 AI 생애주기를 중요하게 고려해야 하는 사항으로 인식하고 있습니다.^{xxiii}

2023년 11월에는 28개국의 정부, 산업계, 시민사회 대표 150여 명이 영국 AI 안전성 정상회의에 모여 「블레츨리 선언」에 서명하고 발표했습니다.^{xxiv} 이 선언으로 프론티어 AI에 대한 초기 합의가 이뤄졌고 AI 위험을 식별하고 완화하기 위해서는 국가 및 국제적인 협력이 필요함을 표명했습니다. 그리고 이를 위해 각국이 위험 기반 정책을 채택하고 국가별 상황에 맞는 법률 및 규제 프레임워크를 구축할 것을 촉구했습니다. 이러한 목표를 달성하기 위해 각국은 “이러한 위험에 대한 과학적이고 증거에 기반한 이해”를 구축하는 목표를 세워야 하며, 여기에는 영향 평가 및 검증 체계의 활용이 포함될 수 있다고 밝혔습니다.^{xxv}

이와 같이 AI 기술의 보안, 안전, 신뢰를 보장하는 것은 AI 거버넌스의 가장 중요한 목표입니다. 수많은 글로벌 이니셔티브에서 AI 가치사슬의 다양성을 고려한 위험 기반 접근방식을 선호하고 있습니다. 이러한 접근방식을 통해 대규모 AI 모델 개발자나 해당 모델을 이용하는 서비스 사용자 또는 둘 다의 역할에 따라 좀 더 알맞게 조정된 규정을 개발할 수 있습니다.

테크기업의 우수 사례

Microsoft와 같은 일부 선도적인 AI 기업은 AI의 안전과 보안을 위해 다양한 이해관계자와 광범위하게 협력하며 신뢰를 확보하고 사회적 책임을 다하고 있습니다. 안전하고 신뢰할 수 있는 디지털 사회를 촉진한다는 원칙과 관련하여 Microsoft는 사회를 더 안전하게 만드는 데 도움이 되는 AI의 잠재력과, 이 기술을 악용하거나 적절한 주의 없이 사용할 경우에 발생할 수 있는 잠재적 위험을 모두 인식하고 있습니다.

러시아-우크라이나 전쟁의 발발 시, Microsoft는 러시아가 공식적으로 침공하기 전인 2022년 1월 초부터 러시아의 사이버 위협을 선제적으로 추적하고 모니터링했습니다.^{xxvi} Microsoft의 세계 최고 수준의 AI 알고리즘과 분석을 통해 패턴, 행동, 잠재적 취약성을 실시간으로 식별할 수 있었습니다. 이렇게 확보된 정보는 AI 기반 안보 추적 플랫폼을 통해 우크라이나 정부에 전달되었고 우크라이나가 실시간으로 발빠르게 현실적인 대응을 하는 데 기여했습니다. Microsoft의 이 같은 사례는 AI와 사이버 보안 역량을 효과적으로 활용하여 사이버 위협과 관련된 디지털 위험을 식별하고 완화함으로써 헌신적이고 책임있는 민간 주체가 더 안전하고 탄력적으로 디지털 사회에 어떻게 공헌할 수 있는지를 보여줍니다. 이와 같은 풍부한 지식과 경험을 바탕으로 Microsoft는 글로벌 안보 커뮤니티의 이익을 위해 귀중한 자료를 제작하여 배포하고 있습니다.²

한국이 여전히 북한과 전쟁 중이라는 점을 고려할 때, AI로 인한 사이버 테러의 위험 증가는 한국 정부에도 반드시 피해야 할 실질적인 위협이 되고 있습니다. 한국의 국가 안보를 위해서는 복잡하고 역동적인 사이버 위협에 맞서 최고 수준의 사이버 안보 역량을 유지하는 것이 무엇보다 중요합니다. 이는 기하급수적으로 성장하는 AI의 힘으로 뒷받침될 수 있습니다.

동시에 AI의 영향력은 책임감 있게 활용되고 관리되어야 하는데, Microsoft는 이 원칙을 철저히 준수합니다. 「백악관 행정명령」이 발표되기 전에 Microsoft, Google, Amazon, Meta, OpenAI를 비롯한 7개 주요 AI 기업은 미국 정부와 함께 일련의 자발적 안전 서약을 발표했습니다. 이 서약의 목적은 안전하고 신뢰할 수 있는 AI 기술 개발의 원칙을 따르는 것입니다.^{xxvii} 예를 들면, 서비스를 공개하기 전 전문가를 통해 진행하는 내부 보안 테스트, 모범 사례 공유, 사이버 보안 및 내부 위협 보호 장치에 대한 투자, 제3자에 의한 강력한 보고 및 감사 등이 포함됩니다. 중요한 것은 이러한 서약이 글로벌 원칙에서 합의한 AI 거버넌스 조치와 일치한다는 점입니다. AI 기업의 안전을 위한 이러한 첫 번째 자발적 노력이 형성되자 더 많은 기업이 서명에 참여하였고,^{xxviii} 이러한 민간의 지지 표명은 미국 정부가 AI에 관한 행정명령을 추진할 수 있는 탄탄한 기반이 되었습니다.

Microsoft는 EU의 일반데이터보호규정(GDPR)이나 책임있는 AI(RAI) 운동 등 새로운 글로벌 표준과 규정 준수 요구 사항을 한 발 앞서서 수용하고 제품 및 조직의 관행에 적용하는 등 늘 모범을 보여 왔습니다.^{xxix} GDPR의 경우, Microsoft는 초거대 클라우드 공급 업체 중 최초로 기업 수준의 GDPR 이용 약관을 준비했으며, 연구, 거버넌스 및 엔지니어링에 책임있는 AI를 포함하기 위한 첫 번째 단계를 착수했습니다. AI 거버넌스 영역에서도 Microsoft는 책임과 의무를 다하는 모습을 계속해서 보여주고 있습니다. AI 거버넌스에 대한 Microsoft의 사고 리더십이 특히 영향력이 있는 이유는 디지털 인프라의 최대 공급 업체인 Microsoft의 노력이 다양한 부문과 산업에 걸쳐 고객을 위한 AI 개발 및 사용의 안전·보안·신뢰로 이어지기 때문입니다. 이는 결과적으로 전체 AI 생태계를 강화하여 훨씬 더 탄력적으로 만들 것으로 기대합니다.

2 사이버 안보의 최근 동향을 소개하고 모든 이해관계자가 더 안전한 디지털 사회에 기여할 수 있도록 실행 가능한 정보를 제공하는 Microsoft의 대표 보고서로는 현재 4번째 연례 보고서인 <Microsoft 디지털 방어 보고서>와, 하이브리드 전쟁에 중점을 두고 FP Analytics와 협력하여 개발한 특별 보고서인 <디지털 프론트라인>이 있습니다. 출처: <https://www.microsoft.com/en-us/security/security-insider/microsoft-digital-defense-report-2023>; <https://digitalfrontlines.io/>

원칙 4
자율과 창의 기반의 디지털 혁신의 촉진

제4 원칙은 "디지털 사회는 디지털 기술의 지속적인 발전과 이를 활용한 혁신을 장려하며, 개인의 자율적이고 창의적인 활동을 통해 디지털 혁신이 창출될 수 있도록 토대가 마련되어야 한다."라는 것입니다.
「디지털 권리장전」은 이러한 문화를 조성하기 위해 다음 4가지 조항을 제안합니다.^{xxx}

<p>(제22조) 디지털 혁신활동의 자유</p> 	<p>모든 사람은 다양한 영역에서 디지털 혁신을 촉진하는 경제적·사회적·문화적 활동을 영위할 자유를 보장받는다.</p>
<p>(제23조) 디지털 규제 개선</p> 	<p>디지털 혁신의 촉진을 위해 인간의 자율을 존중하는 합리적인 규제체계가 형성되어야 하며, 기술 발전 속도, 산업 성숙도, 사회적 수용성 등을 고려하여 불합리한 규제는 개선되어야 한다.</p>
<p>(제24조) 디지털 혁신 지원</p> 	<p>디지털 혁신의 지속적 창출을 위해 민간과 정부 간 긴밀한 협력을 바탕으로 전문인력 양성, 연구개발 투자, 창업 활성화, 인프라 구축, 제도 정비를 포함한 다양한 지원이 이루어져야 한다.</p>
<p>(제25조) 디지털 전환에 따른 갈등 조정</p> 	<p>디지털 전환의 과정에서 발생하는 사회적 갈등을 관리·예방하고 다양한 이해관계자 간의 대화와 속의를 통해 사회적 합의를 이루기 위한 제도적 기반이 마련되어야 한다.</p>



AI 거버넌스와 관련하여 한국 정부의 노력이 돋보이는 분야는 디지털 환경에 맞지 않는 낡은 규제를 폐지하기 위한 '디지털 규제 개선'입니다. 이 조항은 다음 설명하는 다양한 방안을 통해 기존 규제 체계를 AI 산업을 포괄할 수 있도록 준비하여 글로벌 경쟁력을 갖춘 책임있는 디지털 사회의 기반을 강화하기 위해 이미 실행되고 있습니다.

조항	AI 거버넌스 관련 적용 사례	근본적인 사회적 위험
<p>(제23조) 디지털 규제 개선</p>	<ul style="list-style-type: none"> • 데이터 거버넌스, 금융 시스템 등과 같은 기존 규제 프레임워크 수정 • 혁신을 허용하기 위한 사후 규제(블랙리스트) 접근방식의 확장 • 위험 기반 접근방식의 확장 • 사후 규정을 보완하고 법적 명확성과 확실성을 제공하기 위한 세부 지침 • 공개 토의 및 포럼 • 혁신적인 AI 적용을 테스트하고 배포하기 위한 규제 샌드박스 • 국경 간 모델 사례 구축과 공유 • 규제 기관을 위한 교육과 훈련 • 민간 자율 규제를 촉진하는 인센티브 	<ul style="list-style-type: none"> • 디지털 사회에서의 경쟁력 약화, 디지털 격차 및 불평등 약화, 개인 데이터의 무단 사용 및 악용, 충족되지 않은 윤리적 가치



한국

한국은 처음부터 새로운 규제나 의무를 부과하는 것을 지양하고 기존 프레임워크와 제도에 AI 거버넌스의 구성 요소를 통합하는 방식을 채택했습니다. 2020년 5월, 과기부는 한국지능정보사회진흥원(NIA)과 함께 AI로 인해 발생할 수 있는 사회 변화에 대한 의견을 수렴하고, 예상되는 변화와 기존 법규 및 규제·사법 당국의 해석에 조율이 필요한 의제를 발굴하기 위해 AI 법제정비단을 출범하였습니다.^{xxxi} 과기부, 법무부, 문화체육관광부, 행정안전부 등 공공 부문 위원과 산업계, 재계, 법조계, 학계 등 민간 위원으로 구성된 이 범사회적 위원회는 2020년 12월, AI 산업 발전 및 활용 기반을 강화하고 부작용을 방지하기 위해 우선적으로 추진해야 할 30개 공동 발굴 과제를 담은 로드맵 1.0을 발표했습니다.^{xxxi}

이 로드맵 수립 과정에서 주목할 만한 점은 글로벌 정합성을 위해 해외 입법 동향과 국내 법제도를 면밀히 검토했다는 점입니다. 특히 이 로드맵은 규제 정비 과정이 자율 규제로 진행될 때 업계의 자율성을 높일 수 있고, 사회적 합의에 기반해야 타당성이 확보되어 소기의 효과를 거둘 수 있다는 점을 강조했습니다.

「개인정보 보호법」의 개정 논의도 이 로드맵에서 제안한 규제 정비 주제 중 하나입니다. 2023년 9월부터 시행된 「개인정보 보호법」 개정안은 개인정보 처리 자동화 시스템의 결정이 자신의 권리나 의무에 중대한 영향을 미치는 경우에 정보 주체에게 이를 거부할 수 있는 권리를 부여합니다. 또한 데이터 주체에게 이러한 결정에 대한 설명을 요청할 수 있는 권리도 부여합니다.^{xxxiii} 이와 더불어 분석 및 AI 학습에 빅데이터를 활용할 수 있도록 「저작권법」의 개정 논의도 촉발됐는데 현재 계류 중입니다.

2023년 8월, 1차 AI 법제정비 로드맵을 업데이트하고 초거대 AI에 대한 규제 프레임워크를 강화하기 위해 **제4차 AI 법제정비단**이 출범했습니다.^{xxxiv} 규제 개선, 신뢰 구축 방안, 산업별 고려 사항, AI 관련 저작권 문제 등 AI 법제정비 로드맵 1.0의 중장기 과제를 다루는 전문 분과 4개로 구성된 위원회는 고위험 AI 등 AI 거버넌스의 주요 용어에 대한 정의와 범위를 확정하고, 글로벌 AI 표준에 대한 한국의 정합도를 평가하며 새로운 개혁 분야를 발굴하는 것을 목표로 합니다. 법제정비 로드맵 2.0은 2023년 하반기 안에 업데이트될 예정이었습니다.

한국 국회는 'AI법' 초안을 통해 국가 인공지능 정책과 관련된 법적인 틀을 포괄적으로 심의하는 한편, 법률 개정을 통해 기존 규제 체계에 인공지능 거버넌스를 접목하는 작업을 병행하고 있습니다.^{xxxv} 「채용절차의 공정화에 관한 법률」, 「콘텐츠산업 진흥법」, 「정보통신망법」 등 다양한 법률의 개정안을 발의함으로써 AI와 그 사회적 영향을 가장 효율적으로 관리하고자 하는 한국의 인식과 의지를 드러내고 있습니다.^{xxxvi}

한국 정부는 기업의 자율과 혁신을 위해 기존 규제를 개정하는 것 외에도 ICT 규제샌드박스를 통해 제한된 범위에서 규제의 적용을 면제하고 있습니다.^{xxxvi} 샌드박스는 신기술과 기존 규제 체계 사이의 간극을 파악하여 기술의 안전성과 영향력을 테스트하고 검증함으로써 정부가 잠재적 위험을 모니터링하고 규제 개선을 위한 의견을 수집할 수 있도록 지원합니다. 현재 AI 시스템 2개가 규제샌드박스에서 서비스를 시연하며 안전성 테스트를 받고 있습니다.^{xxxvii}

전문가를 인터뷰한 바에 따르면 「디지털 권리장전」의 선언, AI윤리기준의 수립, 인공지능 개인정보 활용 규제에 대한 개인정보보호위원회를 통한 정책 방향 모색 등으로 국가전략과 정책 개발에 힘쓰고 있는 한국 정부가 중요하게 여기며 달성하고자 하는 목표는 바로 국제적인 상호호환성과 정합성입니다. 글로벌 정합성을 강조하는 이유는 디지털 혁신은 상호연결성 및 즉각성과 긴밀하게 연결되기 때문입니다. 디지털 공동번영사회로 전환하기 위해서는 국내 AI 거버넌스 프레임워크와 글로벌 거버넌스 프레임워크의 조화가 필수적입니다. 이러한 조화는 국내 AI 산업이 글로벌 트렌드에 발맞춰 발전하고 안전성을 확보하는 데도 도움이 됩니다. 이에 한국지능정보사회진흥원(NIA), 정보통신기술협회(TTA), 정보통신정책연구원(KISDI) 등과 같은 공공기관은 한국적 접근방식이 국제표준과 조화를 이룰 수 있도록 검토하는 데 힘쓰고 있습니다.

전반적으로 한국의 AI 거버넌스에 대한 접근방식은 혁신과 기업 자율성에 대한 의지, 진화하는 AI 환경에 대한 대응력, AI가 한국 사회에 미칠 중대한 혁신적 영향에 대한 인식을 반영하고 있습니다. 또한 ICT 규제샌드박스와 같은 이니셔티브와 함께 AI 법제정비단의 협력적 노력은 글로벌 트렌드에 맞춰 AI 거버넌스 프레임워크를 개선하고

글로벌 시장에서 기업의 경쟁력을 강화하기 위해 한국이 얼마나 노력하는지를 보여줍니다. 이러한 전략적인 대응을 통해 한국은 기술 혁신과 책임있는 AI 개발 및 사용을 위한 환경을 조성하는 데 앞장서고 있습니다.

국제사회의 논의 현황

EU 「AI법」과 미국 「백악관 행정명령」, 기타 국제 원칙에서 알 수 있듯이 책임있는 AI 개발과 사용을 촉진하기 위해 책임있는 규제의 업데이트가 필요하다는 데 국제적 합의가 이뤄졌습니다.

인공지능에 대한 사용자의 신뢰를 보장하는 프레임워크를 구축하기 위해 EU 「AI법」은 규제샌드박스의 사용을 지지하고 있습니다.^{xxxix} 이러한 규제샌드박스에서는 감독을 통해 통제된 환경에서 혁신을 촉진하고, 중소기업과 스타트업의 규제 부담을 완화하기 위한 추가 조치가 시행될 것입니다. 이러한 '혁신 친화적' 접근방식은 프랑스, 이탈리아, 독일 정부도 지지하며 AI 거버넌스에 대한 협력적 접근방식을 강조하고 있습니다. 이들은 "유럽의 혁신 능력을 저해할 수 있고 기업에는 불필요한 행정적 부담이 되는 규제를 줄이는 것"에 전념하고 있으며, 여기에는 상당한 규제 개혁과 간소화 또는 EU에서 AI 혁신 및 투자를 승인하는 패스트트랙 제도의 촉진이 포함될 수 있습니다.^{xl}



「백악관 행정명령」도 혁신과 경쟁을 지지하며 공정하고 개방적이며 경쟁력 있는 AI 생태계를 보장하는 전략을 내세웠습니다.^{xli} 예를 들어, 미국은 AI 인재를 유치하기 위해 이민 정책을 현대화하는 한편, 중소기업과 스타트업이 업계에 더 적극적으로 참여할 수 있도록 하기 위해 여러 자원과 공공 지원에 접근할 수 있는 권한을 부여하는 규제 조치를 채택할 계획입니다. 미국은 정부 기관의 향후 AI 서비스 도입을 준비하는 차원에서 이 행정명령을 통해 AI 거버넌스를 모니터링하고 테스트하는 샌드박스를 활용할 것으로 보입니다.^{xliii}

미국 정부가 규제 간소화를 시도하는 방식에서 중요한 부분은 규제 가이드라인을 정확하게 설정하되 그 과정에서 기업이나 이해관계자와의 포용적인 협의를 강조한다는 점입니다. 특히 AI의 높은 위험과 잠재적 영향을 다루고, 합리적인 정책과 규제 접근법을 개발하며 혁신을 진전하기 위해 행정명령은 인간의 의견을 수렴하는 것을 의무화하고 있습니다.^{xliiii} AI 안전과 보안을 위해 설립될 자문위원회에는 업계의 AI 전문가가 포함되며, 이 위원회는 국토안보부 장관과 직접 협력하여 중요 인프라에서 AI 사용의 탄력성을 보장하는 역할을 하게 될 것입니다. 이처럼 미국의 규제 개정 프로세스는 자국 내 사회 상황에 알맞게 적용할 수 있는 가이드라인을 구축하는 데 현장의 목소리를 청취하는 것이 필수적이라는 점을 보여주며 민간이 참여할 수 있는 충분한 공간을 제공합니다.

이 외에 OECD 「AI 원칙」은 각국 정부가 통제된 환경에서의 시험을 통해 책임있는 AI 혁신을 촉진하는 민첩한 규제 환경을 구축하고, 원칙 2.3에 따라 정책 프레임워크를 지속적으로 검토하고 개선할 것을 촉구합니다.^{xliiv} 앞서 설명한 바와 같이 이러한 모든 부분은 민간의 자율적 혁신을 촉진하고 하는 한국의 접근방식에서도 분명하게 드러납니다.



영국 AI 안전성 정상회의의 「블레츨리 선언」도 혁신 친화적 접근방식을 수용합니다. 이 선언은 책임있는 방식으로 AI의 잠재력을 효과적으로 극대화할 수 있는 거버넌스 메커니즘이 필요함을 강조하고 있습니다.

2023년 6월에 통과된 EU 「AI법」을 둘러싼 최근의 상황에서도 알 수 있듯이, 혁신 촉진의 영역에서 글로벌 조율의 필요성은 분명해졌습니다. 이 법안은 국제표준, 특히 OECD가 정의한 표준과 조화를 이루기 위해 개정되었으며, 거버넌스 프레임워크와의 관련성을 높이고 향후 국제표준 논의에 적극적으로 참여할 수 있도록 하기 위해 마련되었습니다.^{xliv} 이러한 맥락에서 한국은 ‘인공지능법’ 초안 등 기존 인공지능 거버넌스 관련 법안을 재검토하여 국제적 관행과 더 긴밀히 연계할 수 있는 부분을 찾아내는 것을 고려할 수 있습니다. 주목해야 할 부분은 AI 산업 내의 이해관계자를 더 정확하게 분류하는 것입니다. 한국의 현행법에서는 AI 생태계의 모든 참여자를 사업자로 통합하여 동일한 책임을 지우고 있습니다.^{xlvi} 이에 반해 OECD, EU, 미국에서는 AI 시스템 개발자와 AI 시스템 배포자를 구분하여 더 구체적으로 분류하고 있습니다.^{xlvii} 이러한 접근방식은 앞서 설명한 바와 같이 각 산업 주체가 AI 가치사슬에서 차지하는 위치에 따라 역할이 다르고, 각각의 위험을 관리해야 하며, 역할에 상응하는 책임을 저야 함을 고려한 것입니다.

혁신과 자율성의 원칙은 규제 프레임워크가 업계 주도의 혁신을 저해해서는 안 된다는 점을 인식하고 있는 각국 정부의 AI 거버넌스 논의에 널리 반영되어 왔습니다. 그 균형을 맞추기 위해 통제된 시험 환경을 제공하고 업계 전문가로부터 구체적이고 현실적인 의견을 얻는 등의 접근방식을 취하고 있습니다. 이러한 노력의 목적은 혁신을 통해 책임있는 AI 개발 및 사용을 강화하는 데 더 영향력을 행사할 수 있고 관련성이 높은 정책으로 개혁하는 것입니다.

테크기업의 우수 사례

Microsoft는 기술 아키텍처의 구조에 부합하는 AI 거버넌스 프레임워크를 지지하며, 이러한 개혁의 성공적인 적용에 기여하기 위해 공공 부문과 적극적으로 협력하고 있습니다.^{xlviii} 테크기업을 선도하는 기업으로서 Microsoft는 AI 거버넌스를 위한 5대 원칙 청사진을 비롯해 사고 리더십을 밝혔고, 국내 및 국제 수준에서 진행되는 관련 논의에 참여하여 AI 생태계의 역동적인 구조를 설명해 왔습니다. 혁신 그 자체이자 창의적인 적용을 위한 혁신적 인프라인 AI의 이중적 역할을 고려할 때, 혁신 친화적인 규제 개정 시 AI 기술의 관리와 활용에 대한 업계 참여자의 다양한 역할에 따라 맞춤형으로 책임을 식별하는 과정이 필요합니다. Microsoft는 이러한 산업 특성을 반영하여 효과적이고 효율적인 규제 개정을 촉구하고, 특히 적용 수준에서 기존 조치의 적절성을 강조하며, 필요시 새로운 조치를 개발하기 위해서는 최선의 접근방식을 심의할 수 있도록 정부를 지속적으로 지원하고 있습니다.

디지털 시대의 혁신과 창의성의 선봉에서 있는 Microsoft는 모든 단계에서 신중을 기하며 변화의 결과와 영향에 대한 책임을 다하고 있습니다. 2017년, 사이버 범죄가 급증하고 사이버 공격이 전례 없는 수준으로 확산되는 가운데 Microsoft는 ‘디지털 제네바 협약’의 소집을 지지했습니다.^{xlix} Microsoft는 기업 차원에서 고객을 보호하고 방어하는 동시에 전 세계적으로 공동의 노력이 필요하다는 것을 인식하고 각국 정부가

이 새로운 위험에 대처하기 위해 단결할 것을 촉구했습니다. 그 무렵 Microsoft는 보안 관련 연구에 투자하기 위해 매년 1.3조 원(10억 미국 달러) 이상을 이미 선제적으로 할당하고 있었습니다.

또 다른 예로 AI의 민감한 사용 사례, 특히 안면 인식 기술과 관련하여 Microsoft는 2018년부터 시민의 기본 인권을 침해할 수 있는 안면 인식 기술의 오용을 방지하기 위해 정부가 가드레일을 구현하도록 적극적으로 옹호해 왔습니다.ⁱ Microsoft는 안면 인식 기술이 잠재적으로 영향을 줄 수 있는 사회적·경제적 결과에 대한 의견을 제시하고 공공/민간 부문에서 문제가 될 수 있는 AI 사용 사례를 효과적으로 관리하도록 권장 사항을 발표했습니다. 이는 사회 구성원으로서 정책에 대한 건설적인 관점을 제공하려는 Microsoft의 책임과 노력을 보여줍니다. 마침내 2020년에는 워싱턴주에서 안면 인식에 관한 법안이 통과됨으로써 이러한 옹호 활동이 하나의 결실을 맺게 되었습니다.ⁱⁱ

Microsoft는 다년간의 작업과 연구 그리고 그로부터 수집한 정보를 이용해, 진화하는 위험과 우려에 대비하여 제품과 서비스를 수정하고 업데이트하는 등 미래를 대비하기 위한 적극적인 조치를 과감하게 취해 왔습니다. 이러한 신중하면서도 민첩한 접근방식은 앞서 언급한 바와 같이 Microsoft가 책임있는 AI의 개념을 최초로 제도화하고 2023년 9월에 생성형 AI의 저작권 침해라는 새로운 문제에 대응하여 Copilot 저작권 서약을 발빠르게 공표할 수 있었던 배경이 되기도 했습니다.ⁱⁱⁱ Copilot 저작권 서약은 이전의 AI 고객 서약을 토대로 개정된 것입니다. 이 약정은 Microsoft 고객에게 Microsoft의 기술이 어떻게 창의적 자율성을 보호하고 촉진하는지를 알릴 뿐만 아니라 다른 기업과 정부에도 유용한 참고 자료를 제공합니다.

Microsoft는 기술 발전에 따라 변화하는 요구 사항과 사회적 위험에 대응하여 프로세스를 기민하게 수정하고 그 결과를 다른 이해관계자와 공유함으로써 공익에 기여하고 있습니다. Microsoft가 이 같은 역량을 발휘할 수 있는 이유는 위험을 실시간으로 모니터링하고 당면한 문제의 심각성과 복잡성에 대해 데이터 기반 분석을 수행하며 신속한 조치를 취하는 노력을 헌신적으로 계속 수행해 왔기 때문입니다.

“Microsoft 같은 큰 기업은 자체 윤리원칙, [AI] 신뢰성을 담보하기 위한 기술적 장치, 연구 자원 등 내부적으로 체계를 갖추고 있는 반면, 중소기업이나 스타트업의 경우, 혁신 개발, 확산에 관심을 갖고 있으면서도 기술의 잠재적 위험성에 적절히 대응하기 위한 준비가 미비할 수 있습니다. 그런 차원에서 저희가 윤리 자율점검표 등을 제작하고 배포하고 있지만 대기업들이 자신들의 AI 윤리, 신뢰성 확보를 위한 내부 정책적, 관리적, 제도적 노력을 사회적으로 공유하면 전체적인 산업 생태계가 보다 건강하게 성장하는 선순환 구조에 이바지할 수 있으리라고 생각합니다. 특히 여러 모범 사례를 수집해 중소기업 및 스타트업에 전파하다 보면 대기업의 파운데이션 모델에 플러그인해 잘 성장하는 스타트업이 많아지면서 상생이 가능해지는 만큼, 대기업이 윤리와 신뢰성 확보를 견인하는 역할을 맡아줘야 한다고 생각합니다.”

- 문정욱 정보통신정책연구원(KISDI) 지능정보사회정책연구실장

포용적 AI에 대한 한국의 접근

포용적 AI는 AI 기술이 다양성과 접근성을 보장할 것을 강조합니다. 한국은 인구통계학적 요인에 관계없이 AI의 혜택은 사회 모든 부문에 제공되어야 함을 인식하고 있습니다. 여기에는 AI 알고리즘의 편향성을 완화하고 개발 과정에서 포용성을 확충하여 다양한 관점을 고려하며, 의도치 않게 특정 집단을 소외시키는 기술을 만들지 않으려는 노력이 포함됩니다.

「디지털 권리장전」의 제1, 2원칙은 디지털 자유와 인권을 보호하고 디지털 사회에서 공정한 접근과 기회의 균등이 보장해야 한다는 점을 강조함으로써 ‘포용적 AI’를 지지합니다.

「디지털 권리장전」의 첫 번째 원칙은 모든 사람의 권리와 자유가 보장되는 디지털 사회의 토대로서, 인간의 존엄성과 가치에 대한 존중을 확립해야 함을 강조합니다. 이 원칙은 6가지 항목으로 디지털 권리와 자유를 설명합니다.ⁱⁱⁱⁱ

원칙 1 디지털 환경에서의 자유와 권리 보장

<p>(제6조) 디지털 접근의 보장</p> 	<p>모든 사람은 안정적인 네트워크 환경을 보장받아야 하며, 이를 통해 다양한 디지털 서비스를 언제 어디서나 차별 없이 접근하여 이용할 수 있어야 한다.</p>
<p>(제7조) 디지털 표현의 자유</p> 	<p>모든 사람은 디지털 환경에서 자유롭게 자신의 의사를 표현할 수 있어야 한다. 다만, 타인의 명예나 권리 또는 공중도덕이나 사회 윤리를 침해하지 않도록 책임 있게 이루어져야 한다.</p>
<p>(제8조) 디지털 다양성 존중</p> 	<p>모든 사람은 디지털 기술로 인한 불합리한 차별과 편견으로부터 보호받으며, 사회적·문화적 다양성을 존중받아야 한다.</p>
<p>(제9조) 개인정보의 접근·통제</p> 	<p>모든 사람은 디지털 환경에서 자신에 관한 정보를 열람·정정·삭제·전송할 것을 요구하는 등 이에 대해 접근하고 통제할 수 있어야 한다.</p>
<p>(제10조) 디지털 대체 수단 요구</p> 	<p>모든 사람은 공공영역에서 디지털 방식을 대체하는 수단을 요구할 수 있다.</p>
<p>(제11조) 디지털 근로·휴식의 보장</p> 	<p>모든 사람은 디지털 기술의 발전으로 출현하는 다양한 노동환경에서 안전·건강하게 근로하고, 디지털 연결에서 벗어나 휴식을 보장받아야 한다.</p>

원칙 1에 명시된 인권과 자유는 AI의 맥락에서도 비교적 쉽게 이행할 수 있는 반면, '디지털 접근의 보장'은 덜 친숙하게 느껴질 수 있습니다. 다음 표는 AI 거버넌스를 통한 접근성 보장에 따라 수반되는 것과, 디지털 접근성이 중요한 목표가 되는 AI 시대에 근본적으로 우려되는 사회적 위험 요소를 보여줍니다.

조항	AI 거버넌스 관련 적용 사례	근본적인 사회적 위험
(제6조) 디지털 접근의 보장	<ul style="list-style-type: none"> • 공공 서비스 디지털화에서 인간 중심 설계를 적용한 AI 기반 보조 기술 활성화 • AI 시스템 및 서비스 개발자와 사용자를 위한 접근성 가이드라인 및 표준 • 다양성 및 포용성 자문 그룹 • AI 개발 및 사용 시 포용적 기능 구현에 대한 교육과 인식 제고 프로그램 	특정 인구집단의 배제, 다양한 수준의 기술 활용 능력에 따른 불평등 심화, 투표권 및 자유를 포함한 기본적 인권 침해



한국

국내 과학기술과 디지털 혁신을 전략화하고 실행하는 과기부는 「디지털 권리장전」과 「국가 AI 윤리기준」이라는 AI 정책의 토대를 마련하고 구체화하여 AI 기술의 사회적 수용성을 높이는 데 적극 나서고 있습니다.^{lv} 과기부는 2020년 12월 OECD, EU 등 국제기구의 권고안을 참고해 국가 AI 윤리기준을 처음으로 마련했습니다. 과기부가 제시한 윤리기준 10가지 중 첫째는 인권 보호에, 셋째는 다양성 존중에 초점이 맞춰져 있습니다.

2020년 5월, 과기부는 AI가 한국 사회에 미치는 영향에 더 효과적으로 대응하기 위해 기존 규제 체계에서 개선이 필요한 부분을 정확히 찾아내기 위한 AI 법제정비단을 발족했습니다. 2020년 12월에 발표된 최종 로드맵에서 위원회는 AI의 혜택이 더 보편적으로 분배될 수 있도록 국민의 역량을 강화하고 격차를 해소할 수 있는 디지털 포용 정책의 법적 기반이 필요하다는 점을 확인했습니다.^{lv} 이러한 필요성을 해결하기 위해 로드맵은 디지털포용법의 제정을 제안했습니다.

최근 몇 년간 디지털포용법의 입법 절차가 중단되었지만, 한국 정부는 2023년 9월 「디지털 권리장전」을 발표하면서 디지털포용법에 대한 지지를 재확인하고, 디지털포용법은 권리장전을 기반으로 개발될 것임을 분명히 했습니다.^{lvii}

이와 관련해 과기부는 웹사이트, 모바일 앱, 키오스크에 대한 접근성 표준을 도입하고, 쉬운 사용자 인터페이스(UI) 설계 가이드라인과 관련 자료를 한데 모은 키오스크 UI 플랫폼 구축을 추진하기도 했습니다. 2023년 6월에는 과기부가 NIA, ICT 접근성표준화포럼과 공동으로 디지털 접근성 콘퍼런스를 개최했습니다.^{lviii} 디지털 플랫폼 기업과 학계 그리고 시민사회의 대표를 초청해 디지털 접근성을 높인 우수 사례를 공유하고 고령자, 장애인 등 다양한 계층의 요구를 수용하는 디지털 기기-서비스 설계 가이드라인에 대해 논의했습니다.

한국 정부는 디지털과 AI 혁신이 가져올 미래가 포용적이고 다양화될 수 있도록 민간 부문의 디지털 접근성을 촉진하는 한편, 공공 서비스의 접근성을 개선하기 위한 조치를 취하고 있습니다. 한국 정부는 '디지털 정부 마스터플랜 2021-2025'의 일환으로 AI 기반 가상 비서의 도움을 받아 지능형 공공 서비스를 구현하는 미션을 실현하기 위해 노력하고 있습니다.^{lviii} 이러한 가상 비서는 특정 혜택이나 서비스를 찾기 어려워하거나 디지털 서비스 탐색과 관련해 불편을 느끼는 기타 사항 등, 시민이 디지털 공공 서비스를 이용할 때 직면하는 문제를 해소하는 것을 목표로 합니다. 향후 계획안에는 음성 안내를 위한 AI 지원 가상 스피커를 통합하여 디지털 공공 서비스에 대한 접근성을 더욱 개선하는 내용이 포함돼 있습니다.

한국지능정보사회진흥원(NIA)은 2019년부터 사회적 차원에서 디지털 소외 문제를 해결하기 위한 민간 플랫폼인 디지털 포용 포럼을 주도적으로 운영하고 있습니다. 2023년 11월에 열린 운영위원회 회의에서 포용 정책을 위한 독립적인 분과를 이 포럼에 신설했습니다. 이 새로운 분과는 기존에 각각 역량 개발, 정보 접근성, 대응을 전담하는 분과와 협력하여 디지털 포용에 대한 중장기적인 논의를 지속해 나갈 것으로 보입니다.^{lix}

이상을 요약하면, 한국에서는 디지털 사회에서 인권과 자유의 중요한 측면으로서 접근성이 중요함을 인정하고 있습니다. 그러나 이러한 인식을 AI 거버넌스 프레임워크 내에서 AI 접근성을 강화하는 구체적인 프로세스와 관행으로 전환하기 위해서는 계속 노력해야 할 것으로 보입니다. 현재로서는 AI 개발 및 사용에 접근 가능하고 포용적인 디자인을 통합하며, AI 실험과 혁신을 위한 광범위한 데이터에 대한 접근성을 개선하고, AI 관련 공교육 자료에 대한 접근성을 강화하는 데 초점을 맞춘 예비 논의가 진행되고 있습니다.

“디지털 심화 시대에 접어든 현 시점에서 디지털은 곧 보편적 권리입니다. 취약계층을 비롯한 모든 국민이 미래 사회에서 함께 잘 살아가기 위해서는 디지털을 잘 활용할 수 있는 역량 강화도 필요하지만 기술 자체가 다양성, 포용성을 고려해 발전되어야 합니다. 기업은 기술 개발 및 구현, 기기 설계에 포용적 관점을 도입하는 한편, 정부는 기업들의 그러한 전환이 가능토록 노력해야 합니다. NIA는 ‘디지털 배움터’를 통해 국민 누구나 기본적인 디지털 역량과 실생활 교육을 받을 수 있도록 지원하고 있습니다. 또한 건전한 지능정보문화의 확산과 디지털 사회혁신 생태계 조성에 기여하고 있습니다. 이러한 활동은 AI 리터러시가 필수적인 디지털 포용 정책 개발 및 추진, 지능정보서비스 접근 및 이용 보장, 디지털 인재 발굴 및 양성과 같은 다양한 분야에서 이루어지고 있습니다.”
 - 최문실 한국지능정보사회진흥원(NIA) 디지털포용본부장

국제사회의 논의 현황

AI 접근성에 대한 강조는 AI 거버넌스에 대한 글로벌 논의, 특히 인권 존중을 지지하고 보장하는 데 분명히 필요합니다. 특히 OECD 「AI 원칙」에는 포용적 성장, 인간 중심의 가치, 공정성에 관한 조항이 포함되어 있습니다.^{lx} 이 원칙에 따르면 인간 중심의 설계, 다양성과 포용성, 교육 프로그램을 통한 디지털 접근성의 보장은 당연한 결과입니다. 실제로 포용적 성장에 초점을 맞춘 OECD의 첫 번째 원칙은 모든 사람을 포용하고 취약계층에 부정적인 영향을 미치지 않는 AI 관련 정책 활동을 중요하게 꼽고 있습니다.

미국 국립표준기술연구소(NIST)는 2023년에 신뢰할 수 있는 AI 시스템을 구축하는 데 있어 공정성의 중요성을 강조하는 AI 위험 관리 프레임워크(AI RMF)를 발표했습니다.^{lxi} AI RMF에서 제기한 한 가지 우려 사항은 장애인이나 디지털 격차의 영향을 받는 사람들이 AI 시스템에 접근하기 어렵다는 점입니다. 이러한 우려를 불식하기 위해 AI RMF에서는 AI 위험을 매핑·측정·관리할 때 ‘인력 다양성, 형평성, 포용성, 접근성 프로세스’를 우선순위에 둘 것을 권장합니다.

2023년 10월에 발표된 미국 「백악관 행정명령」에서 접근성은 “모든 사람이 접근할 수 있는 인력 교육 및 개발”을 제공한다는 맥락에서만 언급됩니다.^{lxii} 더 넓게 해석하면 차별, 편견 및 기타 남용을 금지하는 형평성 및 시민권 증진 원칙에 AI 접근성에 대한 고려가 포함될 수 있습니다.

또한 28개국이 서명한 영국 AI 안전성 정상회의의 주요 결과물인 「블레츨리 선언」의 참가자들은 AI가 포용적인 방식으로 사용되어야 한다는 데 동의했으며, 이 선언에서 포용적 AI는 접근성을 보장해야 한다고 언급했습니다.^{lxiii}

전반적으로 AI 거버넌스에 대한 최근의 글로벌 동향은 인권 중심의 접근방식을 보여주며, 이는 본질적으로 AI 시스템과 제품에 대한 접근성을 수반합니다. 따라서 사회적 취약계층에 대한 AI 개발 및 배포의 영향을 고려하고, 이를 바탕으로 현재 디지털 사회에서 취약계층의 기존 요구 사항과 AI 확장으로 인해 발생하는 새로운 과제를 해결하기 위한 방안을 발굴하여 개선해야 할 부분이 많이 남아 있습니다.

전 세계적으로 각국 정부는 공공 서비스 혁신에 발맞춰 AI를 활용하고 있습니다. AI의 도움으로 각국의 정부는 디지털 시대에 접근성과 실용도를 점점 더 높이며 시민의 요구에 부응하고 있습니다. 이러한 국가들은 공공 부문에서 포용적 AI를 도입하는 데 앞장서고 있는데, 이러한 과정에서 Microsoft는 혁신적인 지지자로서 각국의 사회와 국민이 역량을 강화할 수 있도록 돕고 있습니다.

호주	캐나다	인도
<p>호주 공공서비스당국(Australian Public Service)은 Microsoft 365 Copilot을 기반으로 하는 생성형 AI 서비스를 시험하는 세계 최초의 정부가 될 것입니다. 이 시험은 2024년 1월부터 6월까지 진행됩니다. 앤서니 앨버니지 호주 총리는 “Microsoft와의 파트너십을 강화함으로써 우리는 공공 서비스의 미래, 즉 안전을 위해서는 타협하지 않고 호주인을 위해 서비스를 제공하는 호주 공공 서비스 당국의 업무를 개선하기 위해 책임 있게 생성형 AI를 사용하는 과정을 계획하고 있다”라고 밝혔습니다.¹</p>	<p>Microsoft의 AI 기반 번역기는 이제 캐나다 누나부트주 원주민 인구의 70% 이상이 사용하는 모국어인 이누크티투어로 서비스를 제공합니다. Microsoft는 누나부트 주정부 및 지역 커뮤니티 그룹과의 협력을 통해 캐나다 사회의 언어 장벽을 허물고 캐나다 문화와 유산의 기본 요소인 원주민 언어를 보존하는 데 기여하고 있습니다.²</p>	<p>정부 지원 AI4Bharat의 언어 모델과 Microsoft Azure OpenAI Service의 추론 모델을 기반으로 하는 Jugabandi 챗봇은 시민들이 언어 장벽을 극복하고 각자의 모국어로 정부 지원 프로그램에 대한 정보를 얻을 수 있도록 돕습니다.³ 이 챗봇은 인도의 공식 언어 22개 중 10개 언어로 정부 프로그램 171개에 대한 정보를 지원하도록 확장되어 개인화된 정보 검색을 제공합니다. 인도 정부는 앞으로도 디지털 공공재로서 대규모 언어 솔루션을 출시하여 인도 사회에서 민관의 원활한 소통을 지원할 것입니다.</p>

1 Prime Minister of Australia (2023). Australian Government collaboration with Microsoft on artificial intelligence. 출처: <https://www.pm.gov.au/media/australian-government-collaboration-microsoft-artificial-intelligence>
 2 Microsoft (2022). “Introducing Inuinnaqtun and Romanized Inuktitut!”. 출처: <https://www.microsoft.com/en-us/translator/blog/2022/02/01/introducing-inuinnaqtun-and-romanized-inuktitut/>
 3 Microsoft (2023). “With help from next-generation AI, Indian villagers gain easier access to government services.” 출처: <https://news.microsoft.com/source/asia/features/with-help-from-next-generation-ai-indian-villagers-gain-easier-access-to-government-services/>

테크기업의 우수 사례

Microsoft에서 접근성은 단순한 유행어나 뜬구름 잡는 원칙이 아니라 비즈니스의 본질적인 요소입니다. Microsoft는 기술, 특히 AI가 취약계층의 삶에 미치는 혁신적 잠재력을 인식하고 접근성 연구 전담 팀을 조직했습니다. 이 팀은 신체적 장애로 인한 사회적 격차를 해소하고자 혁신적인 솔루션을 개발하는 데 주력하고 있습니다. 2018년부터 Microsoft는 AI 접근성(AI for Accessibility) 프로그램을 통해 장애인의 역량 강화에 투자해 왔습니다.^{lxiv} 이 이니셔티브는 더 포용적인 AI 솔루션을 개발하기 위해 속도를 내고 있으며 AI 생태계에 접근성을 접목한 AI 시스템 개발 및 배포를 표준화하는 것을 목표로 합니다. 이는 장애가 있는 사용자가 서비스를 이용할 때 불편함을 경험하는지를 확인할 수 있도록 설계된 Microsoft의 다양한 접근성 향상 자료(UI 접근성 테스트 및 수정용 개발자 툴 등)를 통해 잘 드러납니다. Microsoft는 AI 접근성 프로그램의 개발을 위해 5년 동안 300억 원(약 2,500만 미국 달러) 이상을 투자했으며, 2021년에는 기술, 인력, 업무 환경에 초점을 맞춘 새로운 접근성 향상 서약을 발표했습니다.^{lxv}

정보의 급속한 디지털화 속에서 Microsoft는 기술을 활용하여 디지털 접근성을 향상함으로써 기본 인권, 특히 장애인 유권자의 투표권을 보호하는 데 중요한 역할을 해왔습니다. Microsoft는 미국장애인협회(AAPD)와 협력하여 2022년 장애인 투표 지수를 도입했습니다.^{lxvi} 이 종합 통계 수치는 미국 50개 주에서 장애인 유권자의 투표 옵션에 대한 정보를 수집하여 검색 가능한 데이터베이스로 구성되었습니다. 장애인 투표 지수는 장애인 커뮤니티와 긴밀히 협력하고 기술을 활용한 솔루션을 통해 디지털 시대의 고유한 과제를 적극적으로 해결하려고 노력한 Microsoft의 가시적인 결과물입니다.

AI 접근성을 위한 Microsoft의 노력은 한국에서도 이어지고 있습니다. 2020년, Microsoft는 SK텔레콤, 한국장애인고용공단과 함께 '접근성을 위한 AI'라는 주제로 온라인 세미나를 진행했습니다. 이 세미나에서 Microsoft는 장애인 사용자의 업무, 생활, 인간관계를 개선하는 데 도움이 되는 다양한 AI 기반 보조 기술을 소개하고, 다양성을 인정하는 포용적 사회를 조성하기 위해 다양한 분야가 서로 협력할 수 있는 방안을 모색했습니다.^{lxvii}

같은 해, 한국을 비롯한 여러 나라에서 Microsoft 인에이블러(Enabler) 프로그램을 시작했는데, 이 프로그램은 이후 9개국으로 확대되어 장애인 350명 이상이 취업의 기회를 얻는 데 일조했습니다.^{lxviii} Microsoft는 포용적 채용 프로세스 설계와 애자일 AI를 통해 구현되었고 보조 기술 사용에 초점을 맞춘 45시간 교육을 통해 국내 파트너가 조직에 다양성과 포용적 직장 문화를 도입할 수 있도록 지원했습니다. 이는 Microsoft가 오랜 기간에 걸쳐 내부적으로 시도해 보고 의견을 수렴하며 접근성 연구에 수년간 투자해 온 노력이 있었기에 가능했던 실효적인 지원이었습니다.

Microsoft에서 접근성은 책임이자 기회입니다. AI는 장애인을 사회에 통합하는 사회적 과제를 해결하는 데 중요한 역할을 합니다.^{lxix} 동시에 Microsoft는 접근성을 높여서 이전에는 소외되었던 그룹을 열성적인 사용자로 흡수하고 가치 중심적인 밀레니얼 세대에게 어필할 수 있는 비즈니스 기회로 인식하고 있습니다. 오랫동안 접근성에 대한 고려 사항을 제품과 프로세스에 통합해 온 Microsoft는 더 포용적인 디지털 사회의 기반이 될 AI 접근성 운영에 대한 광범위한 경험과 전문 지식을 축적해 왔습니다. 이는 전 세계 정책입안자에게 없어서는 안 될 통찰력의 원천이 되고 있습니다.

원칙 2 디지털에 대한 공정한 접근과 기회의 균등

디지털 사회에서 혁신에 대한 경쟁과 기회는 공평해야 하며, 디지털 혁신의 혜택은 사회 공동체에 돌아갈 수 있도록 보장되어야 합니다. 이러한 내용을 담은 제2 원칙은 공정한 접근과 기회의 균등을 보장하기 위해 다음과 같은 다섯 가지 조항을 명시하고 있습니다.^{lxx}

<p>(제12조) 공정경쟁의 촉진</p> 	<p>디지털 경제의 공정한 경쟁 환경을 조성하기 위해 정보와 기술의 독과점, 알고리즘의 불공정성 문제 등으로 인한 피해가 해소될 수 있도록 적절한 조치가 이루어져야 한다.</p>
<p>(제13조) 디지털 자산의 보호</p> 	<p>개인의 투자와 노력으로 형성된 디지털 자산은 정당한 보호를 받아야 하고, 그 거래에 관한 계약은 공정해야 하며 개인이 자유롭게 체결할 수 있도록 보장되어야 한다.</p>
<p>(제14조) 디지털 리터러시 향상</p> 	<p>디지털 기술의 개발과 사용의 기회를 보장할 수 있도록 디지털 격차가 해소되어야 하고, 디지털 리터러시 향상을 위한 교육의 기회가 제공되어야 한다.</p>
<p>(제15조) 데이터 접근 보장</p> 	<p>데이터의 개방은 촉진되어야 하며, 특히 공공 데이터는 접근과 이용의 기회가 공정하게 보장되고 그 이용권의 보편적 확대를 위해 필요한 조치가 이루어져야 한다.</p>
<p>(제16조) 사회 안전망 강화</p> 	<p>디지털 혁신의 혜택은 사회 공동체가 함께 향유할 수 있도록 디지털 심화에 따른 경제적·사회적 불평등 완화를 비롯하여 사회 안전망 강화를 위한 조치가 이루어져야 한다.</p>

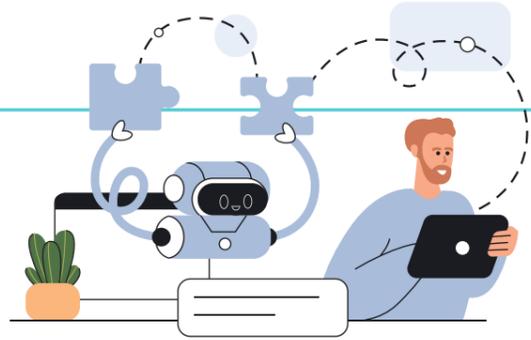
한국 사회의 세계 최저 수준의 출산율로 인한 인구 감소를 고려할 때, 미래 인력의 생산성 향상은 공동번영을 위해 매우 중요하며 앞으로도 계속 중요해질 것입니다. 디지털 사회에서 생산 능력은 인구의 디지털 리터러시 수준과 직접적인 함수 관계에 있다고 볼 수 있으며, 특히 AI 리터러시와 밀접한 관계가 있습니다. 다음 표는 AI 거버넌스가 '디지털 리터러시 향상'의 목표를 어떻게 달성할 수 있는지를 보여주는 예시입니다.

조항	AI 거버넌스 관련 적용 사례	근본적인 사회적 위험
----	------------------	-------------

**(제14조)
디지털 리터러시 향상**

- 인력 재교육 및 전환을 위한 지원 계획 개발
- 학교 커리큘럼에 디지털 리터러시 및 AI 교육 통합
- 커뮤니티 인식 개선 워크숍 및 캠페인
- AI 윤리 교육 자료
- AI 역량 강화를 위한 주 및 지역 차원의 민간 파트너십 활성화
- 가짜 뉴스 필터링 기술에 대한 연구 개발 강화

비디지털 네이티브의 배제로 인한 경제적 역량과 생산성 손실, 디지털 소외로 인한 사회적 비용 감수, 조작 및 가짜 뉴스에 대한 취약성 증가



한국

한국 사회는 디지털 리터러시의 중요성을 인식하고 다양한 이니셔티브를 통해 디지털 문해력, 특히 AI 리터러시에 대한 강력한 의지를 보여 왔습니다. 과기부는 '국가 인공지능 윤리기준'을 발표한 데 이어 2022년 2월에는 **AI 윤리정책포럼**을 발족했습니다. 이 포럼에서는 학계, 산업계, 시민사회 전문가들이 모여 윤리기준을 구체화했으며, 교육 현장에서 AI 윤리를 효과적으로 구현할 수 있도록 교재와 교사용 지도서 등과 같은 주요 결과물을 도출했습니다.^{lxix} 이 가이드라인은 2023년 4월에 추가 개편되었습니다. AI 윤리정책포럼은 한국 사회의 AI 리터러시를 제고하기 위해 더 구체적이고 접근하기 쉬운 콘텐츠를 수록한 교육 자료를 개발하고 제작할 계획입니다.^{lxxii}

교육부 역시 교육 환경에서의 AI 윤리를 확립하기 위해 노력하고 있습니다. 교육부는 2022년 8월, 국내 교육 기관 및 활동에서 책임있는 AI 개발과 활용을 하도록 하는 토대인 '**교육 분야 AI 윤리 원칙**'을 공식적으로 발표했습니다.^{lxxiii} 이 원칙의 핵심은 교육 영역에서 AI 기술이 인간의 성장을 지원하고, 인간의 존엄성을 지키며, 견고한 인간관계를 형성하는 방식으로 활용되어야 한다는 것입니다. 이 문서는 교사와 교육 기관을 대상으로 AI 윤리를 지도해 주는 귀중한 길라잡이의 역할을 합니다.

학생들의 AI에 대한 이해와 활용 능력을 높이는 공교육 시스템 혁신의 또 다른 영역은 **AI 기반 디지털 교과서**의 도입과 확대입니다.^{lxxiv} 교육부는 2025년까지 수학, 영어, 컴퓨터 과학, 특수교육용 국어 등의 과목에 우선적으로 AI 디지털 교과서를 도입하는 것을 목표로 하고 있습니다. 나아가 2028년까지 국어, 사회, 과학 등 전 과목으로 AI 디지털 교과서의 활용을 확대할 계획입니다. 이러한 맥락에서 AI 활용을 통해 모든 학생이 학습 자료에 대한 맞춤 지도와 지원을 받을 수 있으며, 추가적인 비용 부담 없이 공교육을 통해 학습 경험과 질을 향상할 수 있을 것으로 기대됩니다.

NIA는 기술의 광범위한 활용을 위한 견고한 기반을 조성하고 AI 기반 서비스의 혁신을 촉진하기 위해서 일반인을 대상으로 AI 및 기타 ICT 개발에 대한 **디지털 역량 교육**을 제공합니다. 한국 사회의 디지털 역량을 높이는 것이 목표입니다.^{lxxv} 이처럼 NIA 디지털포용분부는 모든 국민이 디지털 사회에서 소외되거나 배제되지 않도록 다양한 계층과 배경을 가진 사람들을 위한 디지털 교육 보급에 힘쓰고 있습니다. 특히, 금융서비스, 교통서비스, 전자상거래 등 실수요자에 맞는 교육을 위해 금융기관, 민간 프랜차이즈 등과도 협력을 확대하고 있습니다. 또한 NIA는 중소기업, 공공기관, 스타트업에 대한 컨설팅 및 기술 교육 등을 통해 지원을 확대하고 있습니다.

결론적으로, 한국에서는 AI와 디지털 리터러시 향상을 위해 정책적으로 교육 분야에 AI 윤리를 도입하고, 교육의 질적 향상에 AI를 적용하며, 국내 개인과 조직 모두의 AI 관련 디지털 역량을 높이려는 노력을 기울이고 있습니다.



국제사회의 논의 현황

한국을 넘어 전 세계의 정책 이니셔티브는 노동시장 전환과 리스킬링에 특히 중점을 두고 있습니다. 예를 들어, OECD 「AI 원칙」 2.4는 정부가 AI 기반 경제로의 공정한 전환을 보장하기 위해 AI 리스킬링 정책을 펼칠 것을 촉구합니다.^{lxxvi} 여기에는 평등, 다양성, 공정성에 중점을 둔 평생 교육과 업무 리스킬링을 지원하는 정책이 포함됩니다. OECD 「AI 원칙」 2.4는 호주의 차세대 AI 졸업생 프로그램, 케냐의 디지털 리터러시 프로그램, 싱가포르의 AI 시대 일자리 재설계 가이드와 같은 정부 정책에 반영되어 있습니다.^{lxxvii}

G7 「AI 기본원칙」에도 디지털 리터러시와 관련된 언급이 있습니다.^{lxxviii} 원칙 9와 관련된 구체적인 행동지침은 기업이 대중의 디지털 리터러시 관련 교육 및 훈련을 지원하는 동시에 AI 시스템 개발과 배포에 대해서도 책임있는 관리를 우선시할 것을 촉구합니다.

또한 미국 「백악관 행정명령」의 8가지 원칙 중 하나는 직무 교육에 중점을 두고 있는데, AI가 고용 인력에 미치는 영향을 조직이 이해해야 함을 촉구하는 내용입니다.^{lxxix} 이 행정명령은 조직이 AI가 근로자에게 미치는 해로운 영향을 완화하는 동시에 유익한 영향을 극대화할 수 있는 모범 사례를 개발할 것을 지시합니다. 또한 이 행정명령은 경제자문위원회에서 작성하는 「AI의 노동시장 효과」에 관한 연구를 포함하여 AI가 노동력에 미치는 영향에 관한 여러 보고서를 의무적으로 작성하도록 규정하고 있습니다.^{lxxx}



영국이 주도한 AI 안전성 정상회의의 참가자들은 AI 관련 기술과 인재 개발의 중요성도 강조했다.^{lxxxi} 이 참가자들은 사람들에게 “AI를 설계·개발·배포·사용할 수 있는” 기술과 지식을 제공하는 것뿐만 아니라 AI가 창출하는 새로운 기회로 인해 많은 사람이 혜택을 누릴 수 있도록 보장하는 등 다양한 관련 사항을 우선순위로 제시했습니다.

이처럼 디지털 리터러시 향상을 통해 사람들이 AI의 혜택을 누릴 수 있도록 하는 전략은 AI 거버넌스와 관련된 국제사회의 논의 동향에서 빈번히 확인할 수 있습니다.

테크기업의 우수 사례

Microsoft와 같은 테크기업은 숙련된 AI 전문 인력의 부족 문제를 해결하고 구직자가 공평한 방식으로 관련 기술에 접근할 수 있도록 AI 기술 이니셔티브에 참여하고 있습니다. Microsoft 필란트로피(Philanthropies)의 이니셔티브인 MS AI 스쿨은 정부 및 업계 단체와의 파트너십을 통해 인센티브와 수당이 포함된 종합적인 교육 과정을 기획하는 등의 활동으로 한국의 AI 인력 수요에도 대응하고 있습니다.^{lxxxii} 이 이니셔티브는 구직자에게 AI 교육을 제공할 뿐만 아니라 소프트웨어정책연구소(SPRI)가 지적한 인력 부족 문제도 직접적으로 해결합니다.

AI 교육의 영향이 얼마나 큰지는 참가자의 성공 사례를 통해 알 수 있습니다. 예를 들어, 문화유산 보존 분야에 종사하던 이승윤 씨는 MS AI 스쿨에 참여한 후 AI와 빅데이터 분야로 이직했습니다. 6개월간의 집중 교육 프로그램을 통해 그는 기본 프로그래밍부터 딥 러닝과 같은 첨단 개념까지 AI에 대해 폭넓게 이해하게 됐습니다. 그의 성공 사례는 MS AI 스쿨이 다양한 학문적 배경을 지닌 사람들이 급성장하는 AI 분야로 진출할 수 있도록 얼마나 잘 지원하고 있는지를 보여줍니다.

Microsoft는 자체 이니셔티브 외에도 문제를 해결하고 사회를 더 공정하고 평등하게 만들기 위해 지역 파트너 회사의 이니셔티브를 발굴하고 지원합니다. Microsoft가 지원하는 사단법인인 Jump는 청소년, 청년, 경력 단절 여성 등 2,300명이 AI를 활용하여 디지털 기술의 격차를 해소하는 데 큰 진전을 이루게 한 단체입니다.^{lxxxiii} 이 이니셔티브는 기술 교육을 넘어 취업 멘토링도 하고 있습니다. Jump의 공동대표인 은초롱 씨는 디지털 시대의 포용성을 강조하며 디지털 교육 영역에서 누구도 뒤처지지 않도록 하겠다는 의지를 밝혔습니다.

‘2023 디지털 임파워먼트 프로젝트’는 6개월간의 노력의 정점으로서 참가자에게 맞춤형 프로그램을 제공했는데, 특히 고등학생들이 Microsoft의 Azure를 통해 AI 서비스 및 기술을 직접 체험할 수 있도록 지원했습니다. 이 프로젝트는 기술 교육에서 여성을 교육 대상자로서 비교적 소홀히 여겼던 문제를 해결하기 위해 특별히 주의를 기울였습니다. AI 전문 기업 테스트웍스와 협력하여 취업 취약계층의 여성에게 AI 데이터 라벨링 교육을 하고, 아울러 취업 기회도 제공했습니다. 고등학생들은 Microsoft의 클라우드 서비스인 Azure를 활용해 코딩 없이도 커스텀 비전, 머신러닝 디자이너와 같은 첨단 AI 서비스를 이용할 수 있었고, Azure 서비스에 대한 지식을 증명하는 AI 900 자격증도 취득할 수 있었습니다. 이러한 교육 과정에는 책임있는 AI에 대한 토론과 IT 전문가의 멘토링이 더해져서 더욱 풍성한 교육 프로그램이 되었습니다.



이와 같이, AI 기술에 대한 접근성을 보장하는 데 있어 민간 부문의 참여는 AI 접근성을 전반적으로 향상하는 중요한 역할을 합니다. AI 인력의 다양성이 확보되면 AI 시스템이 편향된 결과를 초래할 가능성은 줄어듭니다. 한편, 이러한 이니셔티브는 기술 및 고용 격차를 줄이는 데도 도움이 됩니다.

Microsoft는 2023년 8월 대한상공회의소와 함께 생성형 AI가 한국의 생산 능력에 미치는 영향을 정량화하고 잠재력을 활용하기 위한 정책 제언을 담은 <생성형 AI가 한국경제에 미치는 영향> 보고서를 발표했습니다.^{lxxxiv} 이 연구에 따르면, 생성형 AI는 약 620조 원(4,763억 미국 달러)의 잠재 생산 역량을 창출할 수 있습니다. 한국이 AI 도입의 이점을 적절히 활용하기 위해서는 디지털 활용 능력을 지닌 인력을 확보하는 것이 매우 중요합니다. 이 보고서는 또한 생성형 AI 도입으로 영향을 받을 수 있는 비판적 사고 능력뿐만 아니라 교차 기능적 사고로 다른 분야와 원활하게 조율하고 협력할 수 있는 능력의 중요성을 강조했습니다. 이와 같은 지적은 기업에서 다양한 인력을 적재적소에 재배치하기 위한 리스킬링 등을 고안하는 데 활용할 수 있을 뿐만 아니라 정부에서 관련 정책을 입안할 때에도 참고가 될 수 있습니다.

한국의 AI 거버넌스 접근방식에서 지속가능한 AI는 AI 산업의 장기적인 성장 가능성과 사회 전반에 대한 긍정적 영향, 이 두 가지 측면을 모두 포함하고 있습니다. 여기에는 환경적 지속가능성, 경제성장, 지역사회 전반의 복지에 대한 고려도 포함됩니다. 한국은 지속가능한 발전이라는 광범위한 목표에 AI 기술이 긍정적으로 기여할 수 있음을 인식하고 이를 위해 AI를 개발하고 사용하는 것을 목표로 하며, 현재와 미래 세대의 이익을 위한 책임있는 혁신의 중요성을 강조합니다.

「디지털 권리장전」의 원칙 5는 이러한 접근방식을 잘 보여줍니다.

지속가능한 AI에 대한 한국의 접근

원칙 5 인류 후생의 증진

디지털 공동번영사회를 조성하기 위한 마지막 원칙은 보편적 가치와 상호 신뢰에 기반한 국제적 연대를 지지합니다. 이러한 연대는 기술 중심의 혁신적 솔루션으로 글로벌 과제를 해결하고 국가 간 디지털 격차를 해소함으로써 인류 복지의 증진에 효과적으로 기여하는 데 매우 중요합니다. 지속가능한 디지털 사회의 이점을 총체적으로 활용하기 위해 이 원칙은 다음 3가지 주요 분야를 중점적으로 다룹니다.^{lxxxv}

<p>(제26조) 지속가능한 디지털 사회</p> 	<p>국가는 국제사회와 연대하여 디지털 기술이 환경과 생태, 기후체계에 미치는 부정적 영향과 피해를 최소화하고, 디지털 기술을 활용하여 인류 공동체의 후생을 확대하는 데 기여하도록 노력하여야 한다.</p>
<p>(제27조) 국가 간 디지털 격차 해소</p> 	<p>국가는 국제사회와 협력하여 국가 간 디지털 격차를 해소하고 세계 시민의 공동번영을 위해 디지털 국제개발협력을 확대하여야 한다.</p>
<p>(제28조) 디지털 국제규범 등을 위한 협력</p> 	<p>국가는 국제사회와 협력하여 보편적인 디지털 국제규범의 형성과 집행, 디지털 위험에 대한 대응 등을 위한 글로벌 협력 거버넌스 구축을 위해 노력하여야 한다.</p>

다른 모든 분야와 마찬가지로 AI 거버넌스 영역에서의 지속가능성은 사회적·재정적·환경적 실행 가능성과 지속성을 포괄합니다. 글로벌 리더들은 글로벌 난제를 해결하여 인류 후생을 증진하며 AI 혁신의 혜택을 보편화하고 확산하기 위해 적극적으로 노력하고 있습니다. 기본적으로 AI의 이점을 활용하면서 잠재적 위험을 완화하고 기술이 인류의 복지에 긍정적으로 기여할 수 있도록 하는 것이 목표입니다. 이러한 논의는 대부분 국제적인 차원에서 이루어지고 있습니다. 다음 표는 AI 거버넌스에서 '지속가능한 디지털 사회' 조항이 어떻게 전개되어 왔는지를 보여주고, 그러한 담론이 해소하고자 하는 잠재적 우려를 진단합니다.

조항	AI 거버넌스 관련 적용 사례	근본적인 사회적 위험
<p>(제26조) 지속가능한 디지털 사회</p>	<ul style="list-style-type: none"> • 국제 선언 • G7 히로시마 프로세스 • UN 인공지능 고위급 자문기구 • AI 글로벌 파트너십(GPAI) • 국가 AI 안전 기관 설립 • AI 거버넌스 표준화를 위한 새로운 국제 기구 • 개발도상국의 역량 강화 촉진 	<p>국제적인 협력과 혁신, 공공의 이익과 안전에 대한 공동의 추구, 악의적인 의도를 가진 AI 기술 악용, 규제 불일치로 인한 무역 분쟁의 가능성</p>



한국

한국은 AI의 영향력이 국경을 넘어 확장되고 있음을 인식하고 AI 거버넌스를 위한 글로벌 협력에 적극 참여하며 지속가능한 발전을 위한 AI의 활용을 주장하고 있습니다. 이와 관련해 과기부는 국제전기통신연합(ITU)과 협력하여 세계 각국 정부, 국제기구, 학계, 산업계가 모여서 AI의 활용으로 글로벌 문제에 대응하고 지속가능발전목표(SDGs)를 달성하는 미래를 그리는 **AI 포 굿(AI for Good) 프로그램**을 추진하기도 하였습니다. 과기부는 2023년 2월, **전 연령대의 건강과 복지 향상을 위한 AI 세미나**를 공동으로 주최했습니다.^{ixxxvi} 과기부는 전 세계적인 코로나19 팬데믹으로 촉발된 공중보건 위기에서 보듯 국가 간 보건의료 격차가 국제질서의 변명과 지속가능성을 위협하는 요인이라고 규정하고, 이를 해소하기 위해서는 AI를 개발·활용하는 데 국제협력이 중요하다는 공감대를 형성하기 위한 방안을 모색했습니다.

NIA는 2023년 9월 한국마이크로소프트, SK텔레콤, KT, LG유플러스, NHN 등 주요 디지털 및 기술 기업으로 구성된 **디지털 ESG 라운드테이블**을 발족했습니다.^{ixxxvii} 이 라운드테이블은 민간 및 기업 간 파트너십을 촉진하는 허브로서 업계 이해관계자들이 ESG 경영 현안을 공유하고 기업 관행 개선을 논의하는 플랫폼으로 작용할 예정입니다. 또한 한국 정부의 디지털 ESG 정책 연구와 국내 디지털 ESG 생태계 강화에 기여하고 글로벌 기업과의 협력을 확대하는 역할을 할 것입니다.

또한 글로벌 표준화를 우선 목표로 제정된 「**디지털 권리장전**」은 국제사회가 모든 인류의 지속 가능성과 복지를 우선시할 것을 촉구합니다. 「디지털 권리장전」의 5대 원칙을 뉴욕에서 발표하고 제안하기로 한 결정은 이이니셔티브의 국제적 차원을 나타냅니다. 인류의 복지와 지속가능성을 포함한 이 원칙은 2023년 9월에 공식적으로 발표되었는데, 벌써 UN, OECD, 영국 등 글로벌 논의의 장에서 주목받고 있습니다. 정부는 “「디지털 권리장전」의 내용을 국제적 담론에 반영”하고 AI, 디지털 규범, 거버넌스 관련 논의에서 확고한 입지를 확보하겠다는 의지를 표명했습니다.^{ixxxviii}



2023년 9월에 발표된 **전 국민 AI 일상화 계획**은 일상 속 AI 상용화가 어떻게 사회적 소외 계층의 삶의 질을 높여서 사회 전반의 형평성과 지속가능성에 기여할 수 있는지에 대해서도 명시하고 있습니다.^{ixxxix} 여기에는 독거노인을 위한 AI 돌봄 로봇, 소상공인을 위한 원격 상담 및 AI 지원, 장애인을 위한 공공 의료 분야 AI 적용 등의 계획이 포함되어 있습니다. 과기부는 AI를 활용해 한국 사회에 존재하는 일상의 고통을 해소해 나갈 계획입니다.

“「**디지털 권리장전**」에서도 지속가능한 디지털 사회를 위해 디지털 기술이 환경과 생태, 기후체계에 미치는 부정적 영향과 피해를 최소화하고, 디지털 기술을 활용하여 인류 공동체의 후생을 확대하는 데 기여하도록 노력해야 함을 강조하고 있습니다. 인공지능은 인류가 당면한 시급한 문제를 새로운 방법으로 해결할 수 있게 해주고 있습니다. 디지털 사회는 인공지능과 같은 첨단기술을 적극 활용하여 기후위기에 효과적인 해결책을 모색하고 인류가 환경적 책임을 다할 수 있도록 발전해야 합니다. 이 과정에서 긴밀하고 구체적인 민간 공조가 요구된다고 생각됩니다.”

- 최문실 한국지능정보사회진흥원(NIA) 디지털포용본부장

국제사회의 논의 현황

AI 거버넌스 영역에서의 글로벌 규범을 설정하려는 노력은 전 인류의 복지를 증진하는 데 있어 매우 중요합니다. 국경을 넘어 AI 규제에 대한 표준화되고 일관된 접근방식을 보장하여 불필요한 분열과 갈등을 방지할 필요가 있기 때문입니다. 국제 협력을 바탕으로 설정하는 글로벌 규범은 디지털 접근성 차이로 인한 후생 격차 심화, 지속가능성, 기후변화 등 AI가 마주할 도전 과제를 해결하는 데 필수적입니다. 또한 이러한 규범은 빠르게 진화하는 AI 환경에서 탄력적인 대응을 가능케 하고 거버넌스 메커니즘이 계속해서 적절하고 효과적인 상태를 유지할 수 있도록 합니다. 궁극적으로 글로벌 규범을 제정하려는 노력은 인류 전체에 도움이 되는 책임감 있고 안전한 AI 환경을 조성하기 위한 초석이 됩니다.

지속가능성을 강화하기 위한 노력은 다양한 글로벌 AI 표준과 원칙에서 일관되게 나타납니다. 예를 들어, OECD 「AI 원칙」은 포용적 성장, 지속가능한 개발 및 복지를 최우선 원칙으로 삼고 있습니다.^{xc} 이는 신뢰할 수 있는 AI가 인간과 자연환경 그리고 사회 전반을 지원하여 모두가 함께 번영할 수 있는 미래를 만들어야 한다고 규정하고 있습니다.

G7 「AI 기본원칙」도 기후변화, 글로벌 보건, 교육 등 가장 시급한 사회적 과제를 첨단 AI 시스템을 이용해 우선적으로 해결하도록 규정하고 있습니다.^{xcii} 이 원칙에 따라 첨단 AI 시스템은 UN SDG의 진전을 지원하고 더 많은 사회가 혁신적인 방식으로 세계의 최우선 과제에 더 잘 대응할 수 있도록 도와야 합니다.

미국 「백악관 행정명령」도 “글로벌 과제를 해결하기 위해 해외에서 안전하고 책임감 있고 권리를 보장하는 AI 개발 및 사용”을 촉진하는 것이 중요하다는 점을 인식하고 있습니다.^{xciii} 이 행정명령은 지속가능발전이라는 주요 과제를 해결하기 위해 전



세계적으로 AI의 지속가능한 배포와 사용을 지원하는 국제 공조의 필요성을 강조하고 있습니다.

또한 영국 AI 안전성 정상회의의 「블레츨리 선언」은 AI의 혁신이 모두를 위한 것이라고 약속합니다. AI 시스템의 개발로 “UN SDG 달성을 위한 노력을 강화하고” 공공 서비스, 식량 안보, 생물다양성, 기후 측면에서 포용적 발전을 촉진할 것이라고도 공언했습니다. ^{xciii} 환경적 지속가능성과 같은 공익을 극대화하는 데 AI 활용을 위한 프레임워크를 구축하는 과정에 국제 협력은 매우 중요한 역할을 할 것입니다. 실제로 28개국 대표와 주요 테크기업, 시민사회단체, 학계 전문가가 모여 AI 안전과 관련된 조치를 논의하는 AI 안전성 정상회의 자체가 이러한 협력의 결과물이라고도 볼 수 있습니다. 한국은 이 이니셔티브의 후속 조치로서 2024년 5월의 미니 서밋을 준비하고 있으며, 모두에게 혜택을 가져다 줄 수 있는 AI 정책의 글로벌 담론에 적극적인 파트너가 되기 위해 노력하고 있습니다.

테크기업의 우수 사례

Microsoft는 AI가 모든 인류의 행복에 기여해야 하고 또 그럴 수 있다고 확신하며 지구상의 모든 사람과 조직이 더 많은 것을 성취할 수 있도록 지원하는 것을 사명으로 삼고 있습니다. 기후 변화는 우리 세대의 가장 중요한 문제인 만큼, Microsoft는 지속가능성에 헌신하고 있으며 AI를 활용한 지속가능성의 향상을 가속화하고 있습니다.

Microsoft는 2030년까지 탄소 네거티브, 물 포지티브 및 제로 웨이스트를 달성하기 위해 최선을 다하고 있습니다. ^{xciiv} 동시에 글로벌 생태계를 보호하고 환경과 관련하여 지속가능한 의사결정을 하기 위해 설계된 디지털 플랫폼인 플래네터리 컴퓨터를 구축하는 데 전념하고 있습니다. ^{xciv}

결정적으로, Microsoft는 우수 사례와 교훈을 업계와 적극적으로 공유하여 규범을 확립하고 모범 사례와 프로세스를 장려하고 있습니다. AI 포 굿 랩(AI for Good Lab) 및 GRW(Global Renewables Watch)와 같은 이니셔티브는 개인과 조직이 환경 문제를 해결하고 기후 복원력을 강화하기 위한 혁신적인 AI 솔루션을 개발할 수 있도록 보조금, 기술 및 데이터를 제공합니다. ^{xcivi} AI 포 굿 랩은 지역 단체가 플래네터리 컴퓨터와 글로벌 데이터를 활용할 수 있도록 지원하여 확장 가능한 기후 솔루션을 개발할 수 있도록 돕습니다. Microsoft, 플래닛 랩스, 국제자연보호협회(The Nature Conservancy)가 공동개발한 GRW는 AI와 위성 이미지를 이용하여 국가별 청정에너지 전환 현황을 파악하고 추이를 공개적으로 추적합니다. 이는 국가별 재생에너지 규모에 대한 귀중한 정보를 제공하며 재생에너지와 환경의 통시적 상관관계를 이해하는 데 유용합니다.

주목할 만한 협력으로 Microsoft는 지속가능환경생태개발협회(Sustainable Environment and Ecological Development Society, SEEDS)와 협력하여 인도에서 사이클론이 취약계층에 미치는 영향을 예측하는 AI 모델을 개발했습니다. ^{xciivii} 이 모델은 고해상도 위성 이미지를 기반으로 첨단 데이터 분석과 머신러닝으로써 가장 피해 가능성이 높은 주택을 식별합니다. 이를 통해 SEEDS는 사이클론의 영향에 가장 취약한 지역사회를 정확하게 타겟팅하여 지원 활동을 전개할 수 있습니다.

Microsoft는 모든 혁신에 지속가능성을 통합하여 책임있는 AI 개발을 우선시합니다. 최근 발표한 전략 백서인 <AI를 통한 지속가능성 가속화>에서는 복잡한 시스템을 측정, 예측, 최적화하는 능력, 지속가능한 솔루션 개발 가속화 등 탄소중립의 목표를 달성하는 데 필요한 AI의 고유한 기능을 간략하게 설명합니다. ^{xciiviii} 또한 이 자료는 Microsoft의 실행 계획 5가지, 곧 'AI에 대한 투자, 디지털 및 데이터 인프라 개발, AI 운영에서 리소스 사용 최소화, AI 거버넌스 발전, 지속가능성을 위해 AI를 사용할 수 있는 인력 역량 구축'을 소개하여 혁신을 위해 AI의 힘이 발휘될 수 있도록 정보를 공유했습니다. 이러한 모든 노력은 AI가 모든 사람에게 더 실질적인 혜택을 줄 수 있는 방향으로 개발되고 사용되도록 지원하는 것을 목표로 합니다.

Microsoft의 최신 환경 지속가능성 보고서는 2030년 지속가능성의 목표를 달성하기 위해 진전할 것을 강조합니다. ^{xciix} Microsoft는 계속된 성장에도 불구하고 2022년에는 전체 탄소배출량 0.5%를 줄였으며, 깨끗한 물과 위생 시설에 대한 접근성을 약 100만 명까지 확대하였고, 모든 클라우드 하드웨어의 재사용-재활용률은 82%를 달성하여 매립지와 소각장에서 12,000톤 이상의 운영 폐기물을 재사용-재활용으로 전환했습니다. Microsoft는 더 효율적인 데이터센터, 청정에너지, 지속가능성을 위한 Microsoft 클라우드 및 플래네터리 컴퓨터의 개선, 친환경 소프트웨어 관행에 대한 투자를 강화하기도 했습니다. 이와 함께 기후 혁신 기금을 통해 에너지, 산업 및 자연 시스템의 지속가능한 솔루션을 포함한 50개 이상의 다양한 글로벌 투자 포트폴리오에 7천억 원(6억 미국 달러) 이상을 투자했습니다.

AI 거버넌스, 특히 지속가능성에 대한 글로벌 조율을 위한 노력의 일환으로 Microsoft는 여러 이해관계자 이니셔티브에 적극적으로 참여하고 있습니다. Microsoft는 경험과 전문성을 활용하여 AI의 가속화된 힘으로 모든 인류를 위한 기후 복원력을 강화하는 데 소중한 파트너가 될 준비를 완료했습니다.

결론

오늘날과 같이 기술이 주도하는 세상에서 효과적인 AI 거버넌스는 매우 중요하며, 특히 AI가 사회와 경제에 미치는 막대한 영향을 고려할 때 더욱 그렇습니다. 글로벌 IT 선도 국가인 한국은 AI 거버넌스에 있어 국제적인 모범 사례에 부합하는 미래지향적인 접근방식을 수용하는 동시에 한국 국민의 고유한 요구에 알맞은 전략을 구축해야 합니다.

「디지털 권리장전」은 한국의 AI 거버넌스의 토대를 다지는 데 중요한 역할을 하며, 후속 법률, 정책, 규제에 광범위한 영향을 미칠 것입니다. 또한 규제 당국과 사법 당국이 한국 사회에서 AI의 역할을 해석하는 방식에도 영향을 미칠 것입니다. 이 정책은 5가지 기본원칙인 자유, 공정, 안전, 혁신, 연대에 초점을 맞추고 있으며, 이는 모든 사회와 산업이 상생하는 방식으로 기술을 발전시키는 데 필요한 핵심 요소입니다. 또한 「디지털 권리장전」은 AI 산업의 다양한 이해관계자의 의견을 반영하여 그들의 역할과 책임을 정확하게 부여하는 것이 중요하다는 점을 강조합니다.

책임있는 AI 거버넌스를 강조하는 이유는 AI 기술이 책임감 있게 개발·배포되고, 인권을 보호하며, AI 알고리즘의 편견을 방지하는 것이 매우 중요하기 때문입니다. 이러한 책임있는 접근방식은 AI 의사결정의 투명성과 책임성을 보장합니다. 또한 AI 거버넌스에 지속가능성과 포용성을 통합하는 것은 필수 불가결한 요소입니다. 이러한 원칙에 대한 한국의 노력은 AI의 발전이 지속가능하고 다양한 사회적 요구를 충족하며 사회 전 분야에 걸쳐 혜택을 제공하는 포용성을 보장합니다.

AI 분야에서 한국이 독보적인 위치에 있는 것을 고려할 때, AI 혁신의 글로벌 리더로서 한국의 기술력과 잠재력을 인정하는 것이 중요합니다. 한국의 기술 인프라는 미래지향적인 정책과 결합하여 AI 거버넌스의 벤치마크를 설정하는 데 유리한 조건을 갖췄습니다. 책임있고 지속가능하며 포용적인 AI 관행을 통합하는 전략을 채택한다면 국내 시장의 발전에도 이바지할 뿐만 아니라 국제표준과의 높은 정합성을 통해 국제적 위상도 높일 수 있습니다. 한국의 접근방식, 곧 책임있는 기술 발전과 사회적 복지의 균형을 맞추는 방법은 다른 국가들의 모델이 될 수 있습니다. 이러한 전략적 포지셔닝은 책임있고 혁신적인 AI 사용의 선구자로서 한국의 역할을 강조합니다.

본 보고서는 이미 높은 국제 정합성을 자랑하는 「디지털 권리장전」을 위시한 한국의 AI 관련 규제를 국제표준에 맞춰 개발하는 것의 중요성을 설명하고 정부와 민간 부문 간의 지속적인 협력의 필요성을 강조합니다. 이를 통해 한국은 국내적으로 AI 개발 혁신을 촉진하고, 국제적으로 AI 정책 논의에서 글로벌 리더의 위치를 차지할 수 있을 것입니다. 나아가 글로벌 트렌드를 선도적으로 파악해 AI 규범을 지속적으로 업데이트한다면 디지털 공동번영사회의 최전방에서 영향력을 더 행사할 수 있을 것으로 보입니다.

넓은 의미에서, 혁신을 촉진하면서 동시에 안전을 보장하는 균형 잡힌 AI 거버넌스를 추구하는 한국의 노력은 국제사회에서 우수한 사례로서 손색이 없습니다. 이러한 접근방식은 AI 거버넌스에 대한 Microsoft의 비전과 일치하여, AI 분야의 성장과 안전에 대한 공동의 미래를 제시합니다. 본 보고서는 Microsoft와 NIA의 성공적인 협력으로 작성되었는데 이는 한국의 기관 간 파트너십의 모범 사례라고 할 수 있습니다. 또한 Microsoft가, 견고하고 안전한 AI 산업을 구축하고자 하는 한국의 목표에 힘을 보탬 수 있음을 입증합니다.

요약하자면, 이러한 전략들을 국내 상황에 특화된 포괄적인 AI 거버넌스 프레임워크에 적용하는 것이 중요합니다. 이러한 프레임워크는 한국의 성장을 견인할 뿐만 아니라 AI 분야의 인재와 조직의 역량을 강화할 것입니다. 「디지털 권리장전」의 선포는 한국이 AI 거버넌스를 선도할 수 있는 중요한 계기를 만들었습니다. 한국은 국제표준을 채택하고 민간 협력을 강화함으로써 국내 및 국제적 규모에서 효과적인 AI 정책을 수립할 수 있습니다. 이러한 전략은 책임감 있고 지속가능하며, 포용적인 디지털 공동번영사회를 이룩하는 데 없어서는 안 되는 것입니다. 한국은 글로벌 리더들과 협력하여 AI의 이점을 극대화하고 위험 요소를 방지하는 AI 거버넌스 모델을 구축함으로써 지속적인 사회경제 및 기술 성장의 동력을 확보할 수 있습니다.

부록 A

국내 AI 거버넌스 현황

이 섹션에서는 국내의 다양한 정부 기관과 AI 거버넌스 형성을 위한 각 기관의 적극적인 노력을 간략하게 소개합니다.

과학기술정보통신부

한국 사회의 과학기술과 디지털 혁신을 위한 전략 수립과 실행을 담당하는 과학기술정보통신부(이하 과기부)는 '디지털 권리장전'과 '국가 인공지능 윤리기준'이라는 인공지능 정책의 토대를 마련하고 구체화하는 등 인공지능 기술의 사회적 수용성을 높이기 위해 적극적으로 나서고 있습니다. 2020년 12월에는 경제협력개발기구(OECD) 및 유럽연합(EU) 등의 기관에서 제공하는 권고안을 바탕으로 최초의 '국가 인공지능 윤리기준'을 발표했습니다.^c 이 표준은 AI가 인류에게 혜택을 줄 수 있도록 하기 위해 '인간의 존엄성', '공공의 이익', '기술의 목적성'이라는 세 가지 핵심 원칙을 명시했습니다. 또한 핵심 요건 10가지를 다음과 같이 제시했습니다.

- 01 인권 보호
- 02 개인정보보호
- 03 다양성 존중
- 04 피해 금지
- 05 공공성
- 06 연대
- 07 데이터 관리
- 08 책임감
- 09 안전
- 10 투명성

이후 과기부는 2022년 2월 학계, 산업계, 시민사회 전문가가 참여하는 AI 윤리정책포럼을 발족하여 이 기준을 적용할 수 있는 권고안으로 구체화했습니다. 이 포럼의 주요 결과물로는 기업이 활용할 수 있는 자율점검표 및 개발 가이드라인과 교육자가 AI 윤리 교육을 효과적으로 실시할 수 있도록 지원하는 교재 및 강의 가이드가 있습니다.^d 이러한 자율점검표 및 개발 가이드라인은 2023년 4월에 업데이트되었습니다. AI윤리정책포럼은 한국 사회의 AI 리더러시를 제고하기 위해 교육 자료를 더욱 구체적이고 접근하기 쉬운 콘텐츠로 발전시켜 나갈 계획입니다.^e AI 윤리정책포럼 2기에서는 2023년 12월까지 채용, 공공안전, 생성형 AI 활용 서비스 등 신뢰성과 윤리가 강조되는 분야에 대한 분야별 특화 자가 체크리스트와 가이드라인을 마련할 계획입니다. 이러한 가이드라인은 AI 시스템 개발자와 운영자에게 테스트 및 인증 표준에 대한 자발적 준수의 기준을 제공할 것입니다.

2023년 5월, 과기부는 디지털 신질서라는 청사진을 발표했는데, 이는 이후 '디지털 권리장전'으로 발전하였습니다. 2023년 9월에 발표된 '디지털 공동번영사회의 가치와 원칙에 관한 헌장: 디지털 권리장전'은 시가 다양한 산업 분야에서 광범위하게 활용되는 디지털 시대에 추구해야 할 보편적 가치를 개괄적으로 제시했습니다.ⁱⁱⁱ 28개의 간결한 기본원칙에 기반한 선언으로 구성된 이 헌장은 AI로 인한 위험(예: 가짜 뉴스 등)을 해결하기 위해 시민, 기업, 정부의 권리와 책임을 규정하는 출발점이 될 것으로 보입니다. 과기부는 '디지털 권리장전'에 제시된 원칙에 맞춰 '시법', '디지털포용법' 등의 후속 규제를 도입할 예정입니다.

과기부는 2023년 8월, 제4기 인공지능 법제정비단을 발족했습니다. 이 위원회의 주요 임무는 2020년 12월에 처음 수립된 AI 입법 로드맵을 수정하는 등, 규제 프레임워크를 개선하고 대규모 AI를 위한 신뢰할 수 있는 기반을 구축하는 것입니다. 제4기 정비단은 정부 관계자, 업계 전문가, 학계 인사로 구성되어 있으며, 4개의 전문 분과위원회로 나뉘어 각각 규제 개선, 신뢰 구축 방안, 분야별 고려 사항, AI 관련 저작권 문제 등의 중장기 과제를 다루고 있습니다.^{iv}



과기부는 또한 2023년 10월 발표한 '인공지능 윤리·신뢰성 확보 추진계획'에 따라 책임있는 인공지능 거버넌스를 위한 프레임워크를 마련해 가고 있습니다.^{cv} 주요 추진 과제는 다음 표와 같이 민간 부문의 자발적 준수를 유도하기 위한 규제 개정, AI 신뢰성 인증 기술 투자 및 운영, AI 영향평가 및 윤리 교육 확대, 글로벌 디지털 규범 추진 등으로 구성됩니다.

규정 개정

향상된 AI 안전성

- ▶ 고위험 AI 시스템에 대한 기준 가이드라인 (2024년 발표 예정)
- ▶ 고위험 AI 시스템의 신뢰성 확보를 위한 가이드라인 (2024년 발표 예정)

이용자 안전

- ▶ AI 워터마크 가이드라인(2024년 발표 예정)

AI 개발 및 사용

- ▶ 공개적으로 이용 가능한 개인정보의 AI 활용 가이드라인 (2024년 3월 내 발표 예정)
- ▶ AI 저작권 가이드라인 (2023년 12월 발표)



AI 영향평가 및 윤리 교육

AI 영향평가

- ▶ AI 서비스의 위험 식별 및 관리 방안에 대한 프레임워크
- ▶ AI 제품 및 서비스가 사회, 경제, 문화, 공공 생활에 미치는 영향 분석

맞춤형 윤리 교육

- ▶ 다양한 대상 그룹과 수준을 위한 맞춤형 AI 윤리 교육 과정
- ▶ 교육 대상 확대(2023년 일반인, 2024년 개발자 및 배포자 등)

AI 신뢰성에 대한 기술적 보증

2024년 자발적 AI 안전성 및 신뢰성 검인증 제도 출시

2024년 자발적 AI 안전성 및 신뢰성 검인증 제도 출시

글로벌 트렌드에 맞춘 인공지능 디지털 워터마킹 시스템(2024년 추진 예정)

국제 협력

다양한 국제 무대에서 「디지털 권리장전」 홍보

- ▶ 중동, 아세안, 미국, 영국과의 AI 공동 연구
- ▶ 국제적 공통 과제에 대한 AI 기반 솔루션
- ▶ 의료, 교육, 에너지 분야의 AI 애플리케이션

AI 거버넌스 관련 글로벌 의제 발굴

- ▶ AI 안전 연구소
- ▶ 국제 규범
- ▶ UN 산하 AI 거버넌스를 위한 국제기구

2024년 5월 미니 서밋의 공동 주최

한국지능정보사회진흥원

한국지능정보사회진흥원(이하 NIA)은 한국 사회의 디지털 전환을 주도하는 정부 기관으로서 디지털 정책을 수립하고 AI가 일상생활에 접목되면서 발생할 수 있는 사회적 이슈를 해결하기 위해 AI 관련 법제도를 심층적으로 연구하고 있습니다.^{cvii} NIA는 과기부와 긴밀히 협력하며 AI 법제정비단을 이끌고 있으며, 글로벌 표준에 부합하도록 AI 거버넌스에 대한 국가적 논의를 형성합니다. 또한 AI의 윤리, 안전, 신뢰와 관련된 잠재적 위험을 국가 AI 정책에서 포괄적으로 다루는 데 있어 핵심적인 역할을 하고 있습니다.

또한 대규모 인공지능 학습을 위한 국가 인프라를 구축하여 인공지능 산업의 안전하고 책임있는 성장을 촉진하고 있습니다. NIA는 국가 메타데이터 플랫폼을 운영하여 가치 있는 국가 정보를 집적하고 이에 대한 접근성을 높이고 있습니다. 이 플랫폼은 대규모 AI 학습용 데이터세트에 필수적인 국가 메타데이터의 신뢰성을 높이고 AI 기술의 신뢰성을 확보하는 데 기여합니다.

NIA는 AI 기술의 광범위한 활용을 위한 견고한 기반을 구축하고, AI 기반 서비스의 혁신을 촉진하며, 공공/민간 부문 간의 협업을 촉진하는 것을 목표로 삼고 있습니다. 이러한 목표 달성의 일환으로 NIA는 한국 사회의 디지털 역량을 강화하기 위해 일반인을 대상으로 AI 및 기타 ICT 개발에 대한 디지털 역량 교육을 제공하고 있습니다. 또한 중소기업, 공공기관, 스타트업을 대상으로 컨설팅과 기술 교육을 제공하는 등 지원을 확대하고 있습니다.

한편 NIA는 AI 윤리의 중요성을 인식하고 청소년 및 AI기업과 개발자 등을 대상으로 AI 윤리 교육을 추진하고 있으며, 특히 중소기업의 AI 윤리 역량 강화를 위해 AI 윤리원칙 공동 제정, 사내 윤리 교육, 착한AI기업 인증 등을 적극 지원하고 있습니다.

한국 국회

2023년 2월, 국회 과학기술정보방송통신위원회는 '인공지능 산업 육성 및 신뢰 기반 조성에 관한 법률안'(이하 AI법)을 통과시켰습니다.^{cviii} 기존에 제출된 AI 관련 법안 7개를 통합한 이 법안은 제목에서부터 두 가지 초점과 목표를 명확히 제시하고 있어, 사용자 안전에 중점을 둔 EU 「AI법」과 차별화됩니다. 그럼에도 불구하고 업계 일각에서는 고위험 AI 시스템을 대상으로 하는 규제 조치와 정의가 일부 유사하다고 지적하기도 했습니다.^{cviii} 특히, 법안에는 한국에서 흔히 선 허가 후 규제 원칙으로 불리는 AI 기술에 대한 우선 허가 및 사후 규제 원칙에 대한 명확한 조항이 포함되어 있습니다. AI법은 안전하고 책임있는 AI 개발을 유도하고 궁극적으로 국민의 삶의 질을 높이는 것을 목표로 하는 핵심 원칙을 제시하고 있습니다. 이 법이 통과되면 국내 AI 산업을 감독하는 법적 기반이 마련될 것입니다. 또한 법 시행을 위해 국무총리실 산하에 인공지능위원회가 설립되어 인공지능 기본계획 수립 및 운영을 담당하고, 산업 육성을 위해 NIA 산하에 국가 인공지능센터가 설립될 것입니다. 2023년 12월 현재 이 법안은 상임위 심사를 기다리고 있습니다.^{cxix}

이와는 별도로 최근 국회에서 '인공지능 책임 및 규제에 관한 법률안'이 발의되었습니다. 이 법안은 예상되는 위험에 따라 인공지능 시스템을 분류하고 특정 유형의 인공지능 시스템을 금지하는 EU의 범위와 접근방식을 좀 더 밀접하게 반영하고 있습니다. 이것이

AI 거버넌스에 대한 국회의 전반적인 관점을 변화시킬지, 이 법안이 입법 과정에서 AI 법안과 함께 어떻게 진전될지는 아직 속단할 수 없습니다. 또한 이 법안 중 일부가 AI 법안에 통합되었다가 이후에는 폐지되어 AI 법의 규제 강도가 높아질 가능성도 있습니다.

국회는 AI 거버넌스를 위한 새로운 제도 도입과 더불어 법률 개정을 통해 AI의 책임, 공정성, 신뢰성, 투명성에 대한 고려 사항을 국내 기존 규제 체계에 반영하고 있습니다. 가장 최근의 개정안은 2023년 3월에 발의되어 2024년 3월 15일부터 시행되는 「개인정보 보호법」 개정안입니다. 이 개정안은 개인정보 처리를 위한 완전 자동화된 시스템이 내린 결정이 자신의 권리나 의무에 중대한 영향을 미치는 경우에 데이터 주체에게 그 결정을 거부할 수 있는 권리를 부여합니다. 또한 데이터 주체는 이러한 결정에 대한 설명을 요청할 수 있습니다.^{cxix}

AI 활용 관련 추가 규제를 포함한 기타 개정 법안은 다음과 같습니다:

관련 법률	발의 시기	내용
개인정보 보호법	2023년 2월	AI 기업의 알고리즘에서 데이터 침해 사고가 발생할 경우 개인정보보호위원회가 알고리즘 제출을 요청할 수 있도록 권한을 부여 ^{cxix}
채용절차법	2023년 3월	기업이 채용 프로세스에서 AI를 사용할 때에 사전 통지 의무를 부과 ^{cxii}
콘텐츠산업진흥법	2023년 5월	AI 기술을 사용하여 만든 콘텐츠에 라벨을 붙이는 것을 제안 ^{cxiii}
저작권법	2023년 6월	자동화된 정보 처리의 맥락에서 저작권이 있는 자료에 대한 명시적인 기준을 확립하는 것을 목표로 향후 AI로 생성된 저작물의 사용을 위해 AI 기술 적용과 저작권 침해의 경계를 명확히 구분 ^{cxiv}
공직선거법	2023년 6월	캠페인 목적으로 허위 또는 AI가 조작한 정보를 전송하는 것을 금지하고, 이러한 조작된 데이터 또는 정보에 대한 보고 또는 논평 제공에 제한 ^{cxv}
정보통신망법	2023년 2월	인공지능 기반 추천 서비스를 제공하는 사업자가 방송통신위원회에 신고해야 하는 의무를 부과 ^{cxvi}
정보통신망법	2023년 6월	정보통신 서비스 제공자의 정의를 개정하여 AI 기술을 사용하여 정보를 제공하거나 정보를 중개하는 주체를 포함 ^{cxvii}

이 모든 개정안은 2023년 12월 기준으로 소위원회 검토 단계에서 계류 중입니다.

개인정보보호위원회

개인정보보호위원회(이하 개보위)는 「개인정보 보호법」 개정 과정을 조율하는 것 외에도 AI 거버넌스에서 데이터 프라이버시 문제를 적극적으로 감독하고 있습니다. 2023년 6월 23일, 개보위는 AI와 데이터 프라이버시 국제 콘퍼런스를 개최했습니다.^{cxix} 이 콘퍼런스에서는 영국, 독일, 일본 등 주요 국가의 데이터 보호 당국, EU, OECD 등 국제기구, AI 전문가들과 함께 AI 규제 방향에 대해 논의했습니다. 이번 콘퍼런스는 AI와 데이터 프라이버시 규제의 역동적인 영역을 관리하기 위한 개보위의 전략적 정책 방향을 조명하고 혁신과 개인의 권리 보호가 조화를 이루도록 노력해야 함을 강조했습니다. 개보위는 규칙 기반에서 원칙 기반 규제로 전환하고 있으며, 핵심 원칙을 다양한 실제 시나리오에 맞게 적용하는 방식을 지지하고 있습니다. 또한 위험 기반의 차등적 규제 조치를 옹호하고 강력하고 실용적인 기준을 수립하기 위한 민관 협력의 중요성도 강조하고 있습니다.

이러한 접근방식을 통합하여 인공지능 시대의 안전한 개인정보 활용 정책 방향을 발표하였습니다.^{cxx} 여기에는 AI 환경에서 현행 「개인정보 보호법」에 대한 개보위의 해석을 개괄적으로 담고 있습니다. AI 시스템 및 서비스 사업자와 직접 소통할 수 있는 AI 개인정보보호팀을 신설해 법 해석 및 준수에 대한 불확실성을 줄이고자 하는 등, AI 시스템의 시작부터 데이터 수집, AI 학습, 서비스 제공에 이르는 생애주기 전 과정에서 개인정보 취급에 대한 명확한 가이드라인과 기준을 제시할 예정입니다. 특정 영역에서 정부와 민간이 공동으로 규제 체계를 설계한다는 청사진에 따라, AI 프라이버시 정책협의회 구성을 통해 정부, 산업, 학계, 법률 전문가, 시민사회가 협력하여 영역별 가이드라인을 개발할 계획입니다.

방송통신위원회

방송통신위원회(이하 방통위)는 인공지능이 널리 사용되는 사회에서 이용자의 권리 보호에 중점을 두고 한국의 인공지능 정책 수립에 앞장서고 있습니다. 2019년 11월, 방송통신위원회는 정보통신정책연구원(KISDI)과 협력하여 이용자 중심의 지능 정보 사회를 위한 원칙을 발표하며 국내 최초로 AI 윤리 가이드라인을 마련했습니다. 이 가이드라인은 방송-통신 분야에서 AI, 빅데이터, 사물인터넷 등 신기술이 가져올 도전 과제를 고려해 더 안전한 지능 정보 서비스 환경을 조성하기 위해 마련했습니다.^{cxxi} 이 원칙은 인간 중심의 서비스, 투명성, 설명 가능성, 책임, 안전, 비차별, 이용자 참여, 프라이버시, 데이터 거버넌스 등을 강조하며, AI 기술이 일상생활에 빠르게 접목되는 상황에서 이용자의 권리를 보호하고 인간의 존엄성을 유지해야 한다는 점을 강조하고 있습니다.

이러한 노력의 일환으로 방통위는 2019년부터 매년 지능 정보 사회 이용자 보호 국제 콘퍼런스를 개최하여 이해관계자가 AI의 확산으로 인한 사회적·문화적 영향에 대해 숙의할 수 있는 플랫폼을 제공하고 있습니다.^{cxxii} 이 지속적인 콘퍼런스를 통해 글로벌 합의를 도출하고, 원칙과 정책이 국제표준에 부합되게 하는 것을 목표로 합니다.

방통위는 이용자, 사업자, 전문가 등 다양한 이해관계자 그룹과 지속적으로 소통하며 역동적인 AI 기술 환경에 대응하기 위해 AI 원칙을 개선하고 발전시키고 있습니다. 그 결과 2021년 6월에는 AI 미디어 추천 서비스 이용자 보호 기본원칙을 발표했으며, 2022년 4월에는 추가 설명 문서를 공개했습니다.^{cxxiii} 이 원칙은 투명성, 공정성,

책임성의 중요성을 강조하며 AI 기반 추천 서비스 플랫폼이 투명한 시스템으로 운영되는 것을 우선시하고 AI 기반 콘텐츠에 대한 사용자의 통제권을 강화할 것을 권고합니다. 또한 AI 시스템에 미치는 부정적인 영향을 완화하고 사용자 불만을 효과적으로 처리하기 위해 내부 프로토콜과 자체 평가 메커니즘을 개발할 것을 강조합니다. 이와 함께 디지털 미디어 플랫폼 사업자에게는 AI 투명성 및 설명 가능성, 공개 요건과 영업 비밀 보호의 균형, 개인 데이터 사용의 명확한 공개 규범에 관한 자세한 지침을 제공합니다.

또한 방통위는 정보통신정책연구원과 지능정보사회 이용자보호 민관협의회를 발족했습니다.^{cxxiv} 이 협의체에서는 메타버스의 책임있는 활용, 알고리즘 저널리즘 시대의 윤리, 생성형 AI 챗봇 서비스 등 AI 시대의 이용자 보호와 관련된 시의성 있는 다양한 주제에 대해 논의하고 있습니다.^{cxv} 이는 AI 기술의 역동적인 환경 속에서 새롭게 대두되는 사용자 보호라는 과제와 이를 해결하기 위한 방통위의 지속적인 노력을 반영합니다.

교육부

교육부는 2022년 8월, 국내 교육 기관이 AI를 개발하고 활용하는 데 초석이 될 교육 분야 AI 윤리원칙을 선포했습니다.^{cxvii} 이 원칙은 교육 영역에서 인간의 성장을 촉진하고 존엄성을 지키며 인간관계를 강화하기 위해 AI를 활용하는 데 중점을 두고 있습니다. 포괄적인 공개 자문, 전문가 회의, 국제적인 인사이트를 통합해 개발된 이 원칙은 교육적 맥락에서 AI를 책임감 있게 통합하겠다는 약속을 담고 있습니다.

이 원칙은 자발적이고 자율적인 성격을 띠며 교육 환경에서 AI 윤리의 안정적인 정착을 보장하기 위해 고안된 구체적인 지침 10가지로 구성되어 있습니다. 여기에는 교육의 기회 균등과 공정성 증진, 광범위한 사회 복지 기여, 데이터 처리의 투명성과 설명 가능성 보장, 모든 교육 이해관계자의 복지 보호 등 다양한 조항이 포함되어 있습니다. 이러한 윤리기준을 뒷받침하는 기본원칙은 인간의 성장 촉진이라는 교육의 본질에 입각해 교육에서도 AI는 인간의 성장을 지원하는 데 활용되어야 한다는 것입니다.

교육부는 이러한 표준을 교육 부문의 다양한 측면, 예컨대 AI 윤리 교육, 교사 연수용 자료, 교육 기술(에듀테크) 기업을 위한 가이드라인 등에 통합 적용할 예정입니다. 주요 목표는 AI의 이점을 극대화하는 동시에 윤리적 적용을 보장하는 것입니다. 교육부는 이러한 원칙의 타당성과 적용 가능성을 유지하기 위해 주기적으로 검토를 실시할 계획입니다. 이를 통해 끊임없이 진화하는 AI 기술과 그 사용 환경에 적응하고 윤리적 프레임워크를 최신 상태로 유지할 수 있습니다.

교육부는 최근 국무회의에서 교과용 도서에 관한 규정 개정안을 통과시키며 인공지능 디지털 교과서^{cxviii}를 교육 분야에 도입하는 것을 법제화했습니다.^{cxviii} 교육부는 2025년까지 수학, 영어, 컴퓨터 과학, 특수교육용 국어 등의 과목에 우선적으로 AI 디지털 교과서를 도입하고, 2028년까지 국어, 사회, 과학 등 전 과목으로 확대할 계획입니다. 이를 위해 교육부는 과기부와 양해각서를 체결하여 교육부가 구상하는 교육 혁신을 위한 기술 지원을 확보했습니다.^{cxix}

국가정보원

최근 공공기관을 비롯한 다양한 분야에서 ChatGPT를 비롯한 생성형 AI를 빠르게 도입하고 있지만, 정부 차원의 명확한 보안 대책은 부족한 상황입니다. 이에 국가정보원(이하 국정원)은 2023년 6월 정부 기관, 교육 기관, 일반 대중이 이러한 기술을 안전하게 사용할 수 있도록 'ChatGPT 등 생성형 AI 활용 보안 가이드라인'을 작성하여 발표했습니다.^{cxix}

이 보안 가이드라인은 학계, 업계, 연구기관이 참여하는 민관 협의를 통해 개발되었는데, 서비스의 액세스부터 프롬프트의 작성, 결과의 활용에 이르기까지 생성형 AI 기술 사용의 전 과정에 대한 단계별 상세한 지침을 제공합니다. 주요 원칙으로는 개인정보나 기밀 데이터와 같은 민감한 정보의 입력 금지, 생성된 콘텐츠의 정확성·윤리성·적합성 재검증, 생성된 콘텐츠 사용 시 「지식재산 기본법」 또는 「저작권법」 위반 여부 필터링, 상호 연결 및 확장 프로그램의 안전성 검증, 로그인 계정의 보안 설정 강화 등이 있습니다.



부록 B

AI 거버넌스의 최근 국제 동향



AI 거버넌스에 대한 방향성을 제시하기 위한 전 세계적인 노력 증가

지난 몇 년 동안 AI 거버넌스에 대한 전 세계적인 협력이 증가했는데, 특히 미국, EU, G7은 안전하고 신뢰할 수 있는 AI 개발을 지원하기 위한 원칙과 프레임워크를 지속적으로 개발해 왔습니다.

OECD 「AI 원칙」

경제협력개발기구(OECD)는 2019년에 「인공지능에 관한 원칙(AI 원칙)」을 발표했습니다. OECD는 2016년부터 AI 기술 및 애플리케이션의 경제·사회적 영향에 대한 연구를 수행해 왔습니다. 그 결과, AI 거버넌스에 대한 정책적 필요성을 발견한 OECD 디지털경제정책위원회(CDEP)는 “신뢰할 수 있는 AI에 대한 인간 중심적 접근을 촉진하기 위한” 이사회 권고안을 개발하기로 합의했습니다.^{cxxxix}

이러한 배경에서 정부, 산업계, 시민사회, 노동조합 등의 대표로 전문가 그룹이 구성되어 이사회 권고안의 개발을 지원하였으며, 이후 이들의 권고안은 OECD 「AI 원칙」으로 승인되고 채택되었습니다.^{cxxxix}

OECD 「AI 원칙」의 가장 중요한 목표는 OECD 회원국과 비회원국이 “신뢰할 수 있는 AI의 책임있는 관리를 위한 정책을 촉진하고 실행”할 수 있는 프레임워크를 제공하는 것입니다.^{cxxxix}

OECD 「AI 원칙」은 투명성, 설명 가능성, 견고성, 보안, 안전성, 책임성과 같은 핵심 주제를 강조하는 5가지 가치 기반의 원칙으로 구성되어 있습니다. OECD 「AI 원칙」에는 가치 기반의 원칙을 운영하기 위한 국가 정책과 국제 협력 권고 사항도 포함되어 있습니다.

국내적으로 OECD 「AI 원칙」은 각국 정부가 AI 연구에 대한 투자, 신뢰할 수 있는 AI 생태계 육성, 신뢰할 수 있는 AI 개발 및 배포를 지원하는 정책 환경 조성, AI 발전에 따른 인력 전환에 대한 지원 계획의 수립을 고려할 것을 권장합니다. 국제적으로는 지식 공유, 표준 개발 등의 형태로 신뢰할 수 있는 AI에 대한 국제 공조를 권고하고 있습니다.

2019년 OECD 「AI 원칙」이 도입된 이후 71개 이상의 관할권에서 국가 AI 전략을 개발했습니다. 한국, 일본, 인도 등의 국가에서도 OECD 「AI 원칙」에 따라 책임있는 프레임워크와 원칙을 개발하는 과정에 있습니다.^{cxxxix} 또한 OECD는 가치 기반 원칙을 구체적이고 운영 가능한 정책 옵션으로 전환하기 위한 각국의 노력에 주목했습니다.^{cxxxix} 가장 중요한 것은 OECD 「AI 원칙」이 글로벌 협력을 강화할 수 있는 기반을 마련했다는 점입니다. 그 일례로 G7, OECD, GPAI가 협력하여 AI에 관한 G7 기본원칙과 첨단 AI 시스템을 위한 행동강령을 마련한 것입니다.^{cxxxix}

OECD 「AI 원칙」

G7(캐나다, 프랑스, 독일, 이탈리아, 일본, 영국, 미국, EU)은 2023년 10월, 「인공지능에 관한 국제 기본원칙」(G7 「AI 기본원칙」)에 합의했습니다. 이 원칙은 ‘히로시마 프로세스’의 결과물입니다. G7 회원국은 그룹의 공유 가치에 부합하도록 2019년 OECD 「AI 원칙」을 더욱 발전시켜 신뢰할 수 있는 AI를 위한 AI 거버넌스 및 상호운용성에 대한 논의를 진전해 나가기로 합의한 바 있습니다.^{cxxxix}

이러한 배경에서 11가지 기본원칙은 안전하고 신뢰할 수 있는 AI 개발을 촉진하기 위한 명확한 목표하에 도입되었습니다. 법적 구속력은 없으나 G7 「AI 기본원칙」이 각국 정부와 조직에서 AI의 책임있는 사용을 위한 지침으로 활용될 것으로 기대되는 가운데, 빠른 기술 발전에 대응하기 위해 G7은 필요에 따라 계속해서 원칙을 재검토하고 업데이트하기로 했습니다.^{cxxxix, cxli}

G7 「AI 기본원칙」

- 01 첨단 AI 시스템 개발 전반에 걸쳐 위험 대응 조치 채택
- 02 시장 배포 후 취약점 대응
- 03 투명성 확보를 위한 정보 공개 시스템 구축
- 04 첨단 AI 시스템 개발 사고 보고 시스템 구축
- 05 위험 기반 접근방식에 기반한 AI 거버넌스 및 위험 관리 방침 채택
- 06 AI 시스템 생애주기 전반에 보안 조치 투자
- 07 이용자를 위한 AI생성 콘텐츠 식별 조치 도입
- 08 위험 완화 조치에 대한 연구 및 투자 우선
- 09 글로벌 문제 해결 위한 첨단 AI 시스템 개발 우선
- 10 국제 기술 표준 도입
- 11 개인정보 및 지식재산에 대한 보호 및 대응 조치 도입

G7 「AI 기본원칙」은 기본 모델 및 생성형 AI와 같은 첨단 AI 시스템을 다루며 첨단 AI 시스템의 사용·설계·개발·배포 시에 수행할 수 있는 전반적인 조치가 설명되어 있습니다.

G7 「AI 기본원칙」의 핵심은 AI 시스템 생애주기 전반에 걸쳐 위험 기반 접근방식을 적용하며, 위험을 식별하고 완화하기 위한 검증 및 완화 조치를 마련할 것을 강조합니다. 또한 첨단 AI 시스템을 위한 위험 관리 정책을 개발할 것도 권장합니다.^{cxli}

AI의 광범위한 사용을 고려할 때, 이 원칙은 또한 AI 행위자가 사용자가 AI가 생성한 콘텐츠를 식별할 수 있도록 디지털 워터마킹과 같은 콘텐츠 인증 및 출처 증명 메커니즘을 개발하고 배포해야 한다고 강조합니다.^{cxlii}

내부 관리 정책 외에도 AI 이해관계자의 책임성을 높이기 위해 정보 공유에 적극 참여하도록 권장하고 있습니다. 여기에는 사고 보고, 첨단 AI 시스템의 기능과 한계, 적절한 사용 및 오용 패턴 등의 공개 등이 포함됩니다.^{cxliii}

G7은 「AI 기본원칙」의 글로벌한 특성을 염두에 두고 AI 주체가 AI 시스템의 사회적 안전·보안 위험을 완화하기 위한 연구와 범지구적 과제를 해결하는 시스템에 투자하고, 국제 기술 표준을 개발하는 것을 우선순위에 둘 것을 촉구했습니다.^{cxliii}

G7 「AI 기본원칙」은 첨단 AI 시스템으로 인한 규제 문제를 해결하기 위한 최초의 글로벌 이니셔티브입니다. 자발적인 성격이 강하지만, 이 원칙은 AI 주체와 각국 정부가 첨단 AI 시스템의 위험과 이러한 위험을 해결하고 완화하기 위해 취할 수 있는 조치를 고려하는데 중요한 참고 자료를 제공합니다. 또한 이 원칙은 EU 「AI법」 및 영국 「AI 법서」와 같은 다른 국제 규정 및 원칙과 조화를 이루며 AI 거버넌스의 글로벌 조율을 위한 기반을 마련하였습니다.^{cxliii}

미국 백악관 「AI 위험 예방 행정명령」

2023년 초, 미국 정부는 Microsoft를 비롯한 주요 AI 기업 12개사로부터 안전하고 신뢰할 수 있는 AI 시스템을 구축하기 위한 조치를 취하겠다는 자발적인 안전 서약을 받아냈습니다. 이러한 모멘텀을 바탕으로 조 바이든 대통령은 2023년 10월, AI에 관한 행정명령에 서명했습니다.

자발적 안전 서약과 함께 이 「백악관 행정명령」의 가장 중요한 목표는 안전하고 보안이 유지되며 신뢰할 수 있는 AI 개발과 사용을 촉진하는 것입니다. 「백악관 행정명령」은 AI 기업을 규제하는 것 외에도 모든 부문의 조직에 적용되며, 생성형 AI를 비롯해 예측·추천·결정을 내리는 모든 기계 기반 시스템을 포함한 광범위한 AI 시스템을 다룹니다.^{cxliii}

「백악관 행정명령」은 AI 표준, 개인정보보호, 형평성 및 시민권, 소비자 및 근로자 보호, 혁신과 경쟁, AI 거버넌스 리더십 등의 분야에서 8가지 원칙을 제시했습니다. 구체적으로 글로벌 AI 거버넌스와 관련하여 「백악관 행정명령」은 AI가 계속 발전함에 따라 AI 시스템의 잠재적 위험에 대비하기 위해 AI 안전 및 보안에 대한 새로운 표준을 설정하고자 합니다. 이러한 요건에는 AI의 안전, 보안 및 신뢰성을 보장하기 위한 표준, 도구 및 테스트 개발, 개발자가 안전 테스트 결과 및 중요 정보를 의무적으로 정부와 공유, AI로 생성된 콘텐츠를 식별하기 위한 콘텐츠 인증 및 디지털 워터마킹에 대한 지침 수립 등이 포함됩니다.^{cxliii}

또한 「백악관 행정명령」은 AI 연구에 투자하고 경쟁력 있는 AI 시스템을 장려함으로써 책임있는 혁신, 경쟁, 협업을 촉진할 필요가 있음을 강조합니다.^{cxliii}

「백악관 행정명령」은 AI의 전 세계적 영향력을 감안하여 정부가 AI에 대한 여러 이해관계자의 참여를 독려하고 국제 파트너와 함께 AI 표준을 개발하고 구현하는데 참여하도록 지시합니다. 또한 이러한 협업 외에도 글로벌 과제를 해결하고 중요 인프라에 대한 위험을 완화하기 위해 AI를 배포하는 것을 우선순위로 삼고 있습니다.^{cxliii}



「백악관 행정명령」은 글로벌 AI 거버넌스에 상당한 영향력을 행사합니다. 이 행정명령은 국내용이지만, 연방정부가 AI 표준에 대해 국제 파트너와 협력하라는 명시적인 언급을 했습니다. 행정명령이 발표된 직후 미국은 EU 및 27개국과 함께 AI 시스템의 영향과 위험을 조사하고 해결하기 위한 「블레츨리 선언」에 서명했습니다.^{di} 또한 상무부 산하 국립표준기술연구소(NIST)는 전 세계적으로 안전하고 신뢰할 수 있는 AI 사용을 촉진하기 위해 “장비를 갖추고 협력 체계를 구축해 나갈” 미국 AI 안전 연구소 컨소시엄을 설립한다고 발표했습니다.^{di}

미국 NIST AI 위험 관리 프레임워크(AI RMF)

행정명령이 발표되기 전부터 NIST는 AI 관련 프레임워크와 정책 개발을 지원하기 위해 연구 분야와, 표준 개발 및 데이터 수집 분야를 지원해 왔습니다.^{dii}

특히 2023년 1월, NIST는 AI 위험 관리 프레임워크(AI RMF)를 발표했습니다. 이 프레임워크는 개인, 조직, 사회가 AI와 관련된 위험을 이해하고 관리하는 데 도움을 주고자 하는 목적으로 민간/공공 부문과의 긴밀한 협력을 통해 개발되었습니다.^{diii}

이 프레임워크는 AI 피해 유형 3가지에 대한 개요를 제공하고 신뢰할 수 있는 AI 시스템의 특징 7가지 특징을 간략하게 설명합니다. 또한 AI 생애주기 전반에 걸쳐 AI 주체가 이러한 피해를 효과적으로 관리하기 위해 채택할 수 있는 상호연결성을 지닌 기능 4가지를 제공합니다.^{div}



영국 AI 안전성 정상회의

이 프레임워크는 AI의 위험을 크게 3가지 유형, 곧 사람에 대한 위험, 조직에 대한 위험, 생태계에 대한 위험으로 분류합니다.^{dv} 이러한 피해를 방지하기 위해 프레임워크는 AI 개발자가 AI 제품, 서비스 및 시스템을 설계·개발·사용·평가할 때 고려할 수 있는 특징 7가지를 제시합니다. 이 프레임워크에 따르면 신뢰할 수 있는 AI 시스템은 유효하고 신뢰할 수 있어야 하며, 안전하고, 안전하며, 복원력이 있어야 하고, 책임감 있고, 투명하며, 설명 가능하고, 해석 가능하며, 개인정보보호가 강화되고, 유해한 편견을 관리하여 공정해야 합니다.^{dvi}

이러한 특성에 따라 AI 시스템을 설계하려면 다양한 활동과 기능을 통합해야 합니다. 이 프레임워크에서는 AI 주체가 다양한 이해관계자의 관점을 수집하여 AI 위험을 파악하는 것을 목표로 하는 ‘파악’ 기능을 도입하는 것부터 시작할 것을 제안합니다.^{dvii} 이러한 이해가 이루어지면 AI 위험을 정기적으로 분석·모니터링·평가·벤치마킹할 수 있는 기능이 있어야 합니다.^{dviii} AI 리스크를 정기적으로 검토함으로써 AI 리스크를 적절히 관리하고 리스크를 완화하기 위해 리소스를 할당할 수 있는 관리 기능을 준수할 수 있어야 합니다.^{dx} 마지막으로, 위 기능의 지속성과 거버넌스를 보장하기 위해서는 관련 구조와 시스템 및 프로세스를 마련해야 합니다.^{dx}

AI RMF는 AI 거버넌스를 위한 중요한 기반을 제공합니다. AI RMF는 자발적이고 분야와 사용 사례에 구애받지 않지만, NIST는 조직이 프레임워크의 접근방식을 구현하는 데 도움이 되도록 「NIST AI RMF 전략 백서」, 「AI RMF 로드맵」과 같은 일련의 자료도 함께 발행했습니다.

NIST는 AI RMF를 발표한 후 다른 국제표준이나 가이드라인 및 표준과의 조화를 위해 이를 설명하기도 했습니다. 예를 들어, NIST는 다양한 교차 검토를 진행해 AI RMF의 신뢰성이 OECD AI 권고안, 「AI 권리장전 청사진」, 「백악관 행정명령」 및 EU 「AI법」과 연관성이 있음을 증명했습니다.^{dxii}

2023년 11월 초에 열린 영국 AI 안전성 정상회의는 영국 정부가 AI의 안전한 개발을 위해 소집한 최초의 정상급 회의였습니다. 이 정상회의에는 정부와 기업, 시민사회의 대표 150여 명이 초청되어 AI의 위험과 기회 그리고 AI의 안전한 발전을 위한 글로벌 협력 방안에 대해 논의했습니다.^{dxiii}

이 정상회의의 주요 목표는 프론티어 AI의 위험과 기회에 대한 공동의 이해를 확립하고, 프론티어 AI의 안전성을 높이기 위해 각 기관이 취해야 할 조치를 심의하는 것이었습니다.^{dxiii} 이와 동시에 특별히 AI 안전 연구에 관한 국제 협력과 국내 및 국제 프레임워크 지원을 위한 전진 프로세스를 구축하고자 했습니다.^{dxiv}

이와 같은 주요 목표는 「블레츨리 선언」으로 수렴되었습니다. 28개국이 서명한 이 선언은 프론티어 AI에 대한 초기 합의에 도달했으며, 이와 관련된 기회와 위험을 인식하고 있습니다.^{dxv} 또한 이 선언은 위험의 식별과 완화를 위한 국내 및 국제적 협력의 필요성을 표명하고 있습니다.^{dxvi}

위험 식별 측면에서 「블레츨리 선언」은 AI 안전 위험에 대한 “과학적, 증거 기반의 공유된 이해”와 이러한 위험이 사회에 미치는 영향을 조사하기 위한 글로벌 접근방식을 구축할 필요성을 강조했습니다. 위험 완화 측면에서는 각국에 위험 기반 정책을 채택하고 국가별 상황에 맞는 법률 및 규제 프레임워크를 수립할 것을 요구했습니다. 이러한 프레임워크에는 프론티어 AI 이해관계자의 투명성을 강화하고 안전성을 검증하는 체계의 도입을 촉구하는 내용이 포함될 수 있습니다.^{dxvii}

이들간 진행된 이 정상회의는 다양한 분야의 참가자가 국제 AI 거버넌스와 관련된 이슈를 논의할 수 있는 장을 마련했습니다. 특히, 프론티어 AI에 대한 공동의 이해를 위한 지속적인 논의의 필요성과 프론티어 AI뿐만 아니라 다른 유형의 AI로 인한 단기적인 AI 위험에 대처해야 한다는 시급성에 대한 인식을 제고했습니다. 또한 첨단 AI 시스템을 개발하는 조직을 위한 국제 지침 원칙을 포함한 공통 원칙과 강령에 대한 인식도 공고히 했습니다.^{dxviii}

「인공지능에 대한 조화로운 규칙을 규정하는 규정 제안서」(이하 EU 「AI법」)는 EU에서 AI 거버넌스에 초점을 맞춘 최초의 포괄적인 법적 제도입니다. EU 집행위원회가 제안한 EU 「AI법」의 가장 중요한 목표는 “EU에서 신뢰할 수 있는 AI 시스템의 개발과 사용을 위한 조건을 조성함으로써 유럽 단일 시장의 적절한 기능을 보장하는 것”입니다.^{dxix}

EU 「AI법」안에는 일련의 구체적인 목표가 명시되어 있습니다. 이 법안은 기존 EU 규정을 준수하면서 시장에 출시되는 AI 시스템의 안전을 보장하고 시장 분열을 방지하기 위해 AI 관련 규제를 현 시스템과 조율하고자 했습니다. 또한 EU 「AI법」은 법적 명확성을 제공하여 AI에 대한 투자와 혁신에 도움이 되는 환경을 조성하는 것을 목표로 합니다.^{dxix}

EU 「AI법」은 책임성을 보장하기 위해 AI 생태계의 다양한 주체에게 적용될 예정입니다. 여기에는 “EU 시장에 AI 시스템을 배치하거나 EU에서 서비스를 제공하려는” EU 내 또는 제3국의 공급자와 사용자도 포함됩니다.^{dxxi} 그러나 “군사적 목적으로만 개발되거나 사용되는” AI 시스템, “제3국의 공공기관”, “법 집행 및 사법 협력을 위한 국제 협약의 틀 안에서 AI 시스템을 사용하는 국제기구 또는 당국”은 예외입니다.^{dxxi}

EU 「AI법」의 초석은 AI 거버넌스에 대한 위험 기반 접근방식입니다. 이 법에 따라 AI 시스템은 “사람의 건강, 안전 및 기본권”에 미치는 위험의 정도에 따라 분류됩니다.^{dxxi} 위험은 크게 네 가지 범주, 곧 허용 불가, 고위험, 제한적 위험, 최소 위험으로 나뉩니다. 각 위험 범주에는 AI 시스템을 개발, 사용 및 EU 시장에 출시하기 위한 요건과 의무가 따릅니다.^{dxxiv}

EU 「AI법」에서 금지하는 허용 불가 AI 시스템은 대체로 잠재적이고 조작적인 기술을 사용하며 특정 집단을 악용하는 시스템을 말합니다.^{dxxxv} 이 범주에 속하는 AI 시스템의 예로는 소셜 스코어링, 실시간 생체 인식 시스템 등이 있습니다.^{dxxxvi}

고위험 AI 시스템은 EU 보건 및 안전 법률에 해당하는 제품(장난감, 항공, 자동차 등)의 안전 구성 요소 또는 생체 인식, 중요 인프라, 교육 및 고용과 같은 특정 영역에 배포되는

시스템의 두 가지 범주로 나뉩니다.^{dxxxvii} 이 범주에 속하는 시스템은 판매하거나 서비스에 투입하기 전에 EU 전체 데이터베이스에 등록해야 합니다. EU 외부에서 이러한 AI 시스템을 제공하는 공급 업체는 규정 준수를 보장하기 위해 적합성 평가를 받아야 합니다.^{dxxxviii} 이 외에도 고위험 시스템은 테스트, 기술, 데이터 교육 등 다양한 영역에서 다양한 요구 사항을 충족해야 합니다.^{dxxxix} 또한 개발자는 사람의 건강, 사망 또는 재산상의 피해를 초래할 수 있는 심각한 사고를 관련 당국에 보고할 수 있는 보고 메커니즘을 구축해야 합니다.^{dxxxx}

다음으로, 제한적 위험 시스템은 조작의 위험이 제한된 시스템으로 식별됩니다. 딥 페이크와 챗봇이 그 예입니다. 이러한 시스템에는 투명성 의무가 적용됩니다.^{dxxxxi}

마지막으로, 최소 또는 낮은 위험도를 가진 AI 시스템은 현재로서는 요건을 충족할 의무가 없습니다. 그러나 고위험 AI 시스템에 대해서는 자발적인 요건 준수를 장려하기 위해 행동강령을 고안할 수 있습니다.^{dxxxxii}

2023년 12월 현재, 유럽 의회는 EU 「AI법」안에 대한 잠정적인 정치적 합의에 도달했습니다.^{dxxxxiii} 아직 잠정안이 공개되지는 않았지만, 주요 조항에는 이전에 기초 모델로 불렸던 ‘범용 AI 모델’에 대한 의무 부과와 시스템적 위험이 있는 하위 범주에 대한 추가 요건 등이 포함되어 있습니다. 특히 범용 AI 모델의 개발 및 배포 업체는 직접적인 요건을 적용받지 않을 수 있지만 고위험 사용 사례에 대한 기술 정보는 공유해야 합니다. EU 「AI법」은 2024년 초에서 중반에 발효될 예정이며, 시행 시기는 조항별로 상이합니다. 전 세계적으로 최초로 도입되는 포괄적인 법적 프레임워크로 간주되는 EU 「AI법」을 시작으로 전 세계 국가들이 이 법에 부합하는 공식적인 법률을 채택하거나 통과시키도록 유도할 수 있습니다.^{dxxxxiv} 여러 관할권의 데이터 보호법에 영향을 미친 바 있는 EU GDPR의 경우와 유사한 흐름이 전개될 것을 예상할 수 있습니다.



EU 「AI법」

부록 C

한국의 「디지털 권리장전」과 국제 AI 정책 동향 교차 검토



국제사회 및 테크 산업 전반의 방향성과 일치하는 한국의 AI 거버넌스 접근방식

「디지털 권리장전」의 디지털 공동번영사회를 위한 5대 기본원칙에 명시된 바와 같이, 한국의 AI 거버넌스에 대한 접근방식은 앞서 설명한 글로벌 이니셔티브와 주요한 공통점을 공유합니다.

몇 가지 주목할 만한 비교 포인트는 다음과 같습니다:

원칙 기반 규제 방침 G7 「AI 기본원칙」, OECD 「AI 원칙」, 미국 자발적 안전 서약, 「디지털 권리장전」은 모두 책임있는 AI 개발과 사용을 위한 일련의 기본원칙을 바탕으로 규제가 필요함을 강조하고 있습니다.

위험 기반 접근방식 G7 「AI 기본원칙」, OECD 「AI 원칙」, 한국의 「디지털 권리장전」, 미국의 「백악관 행정명령」은 모두 AI 거버넌스에 대한 위험 기반 접근방식을 강조하고 있습니다. 이들 모두 AI 기술의 역동적인 특성을 인정하고 개발 전 단계부터 배포 후까지 위험을 모니터링하고 완화하는 것이 중요함을 강조합니다.

윤리적 고려 모든 글로벌 이니셔티브는 AI 거버넌스에서 윤리적 고려 사항을 한결같이 우선시합니다. 반복적으로 강조되는 핵심 가치로는 공정성, 책임성, 투명성, 인간 중심적 접근방식이 있으며 이러한 가치를 평가하고 구현하기 위한 다양한 후속 조치가 진행되고 있습니다. 모든 AI 거버넌스 이니셔티브의 가장 중요한 목표 중 하나는 AI 시스템의 안전성, 보안성, 신뢰성을 확보하는 것입니다. 특히 미국, EU, 한국은 산업 육성의 관점에서도 안전과 신뢰가 필수적이라는 점을 인식하고 있습니다.

다양한 이해관계자의 참여와 파트너십 국제사회의 주요 AI 이해관계자는 AI 시스템의 관리·운영의 범위와 방침을 정립하고 개선하기 위해 업계 파트너와 다층적 협력 방법을 강구하고 있습니다. 예를 들어, AI 안전성 정상회의에서 영국은 투명성과 모범 사례에 대한 인식을 증진하기 위해 주요 AI 기업이 각자의 AI 안전 정책을 공유할 것을 요청했습니다. G7 「AI 기본원칙」은 위험 완화 조치에 대한 연구와 투자의 우선순위를 정할 것을 촉구했으며, 미국도 전체적인 AI 거버넌스 프레임워크를 위해 여러 이해관계자가 참여하도록 하겠다고 직접 약속했습니다.

국제 협력 국제 협력은 G7 「AI 기본원칙」, OECD 「AI 원칙」, 미국 「백악관 행정명령」, 「블레츨리 선언」, 「디지털 권리장전」 등 다양한 원칙에 담겨 있습니다. 이들은 디지털 문명을 격차, 지속가능성, 기후변화로 인한 격차 심화 등 AI로 인한 과제의 범국가성을 인식하고 위험 요인을 파악하고 완화하는 데 국제적 협력을 할 것을 촉구합니다.

표 1: 한국의 「디지털 권리장전」과 최근 국제 AI 정책의 교차점

	원칙 1 디지털 환경에서의 자유와 권리 보장 (2장)	원칙 2 디지털에 대한 공정한 접근과 기회의 균등 (3장)	원칙 3 안전하고 신뢰할 수 있는 디지털 사회 (4장)	원칙 4 자유와 창의 기반의 디지털 혁신의 촉진 (5장)	원칙 5 인류 후생의 증진 (6장)
G7 「AI 기본원칙」 데이터 품질 관리 및 개인 데이터와 지식재산 보호 (11)	데이터 품질 관리 및 개인 데이터와 지식재산 보호 (11)	글로벌 과제 해결 (9)	AI 생애주기 전반에 걸친 위험 식별, 평가 및 완화 (1) 배포 및 시장 출시 후 취약성, 위험 및 오용 모니터링 (2) 조직 전반의 책임있는 정보 공유 및 사고 보고 (4) AI 거버넌스 및 위험 관리 정책 개발, 구현 및 공개 (5) 강력한 보안 제어 (6) 신뢰할 수 있는 콘텐츠 인증 및 예방 메커니즘 (7) 데이터 품질 관리 및 개인 데이터와 지식재산 보호 (11)	조직 전반의 책임있는 정보 공유 및 사고 보고 (4) AI 안전, 보안 및 신뢰 향상을 위한 연구 우선순위 지정 (8) 투명성 및 책임성 위한 AI 시스템의 기능 및 사용 영역 공개 보고 (3)	글로벌 과제 해결 (9) 국제 기술 표준 개발 및 채택 (10)
미국 「백악관 행정명령」 형평성 및 시민권 증진 1-3	소비자, 환자, 학생을 위한 지원 2: 교육을 혁신하는 AI의 잠재력 구체화하기 근로자 지원 1: 일자리 대체, 노동 기준, 직장 내 형평성, 건강 및 안전, 데이터 수집 등의 문제를 해결하여 근로자의 피해를 완화하고 AI 의 이점을 극대화하기 위한 원칙과 모범 사례 개발하기 혁신과 경쟁 촉진 1-3	소비자, 환자, 학생을 위한 지원 2: 교육을 혁신하는 AI의 잠재력 구체화하기 근로자 지원 1: 일자리 대체, 노동 기준, 직장 내 형평성, 건강 및 안전, 데이터 수집 등의 문제를 해결하여 근로자의 피해를 완화하고 AI 의 이점을 극대화하기 위한 원칙과 모범 사례 개발하기 혁신과 경쟁 촉진 1-3	AI 안전 및 보안을 위한 새로운 표준 1: 가장 강력한 AI 시스템 개발자는 안전 테스트 결과 및 기타 중요 정보를 미국 정부와 공유하기 AI 안전 및 보안을 위한 새로운 표준 2: AI 시스템의 안전, 보안, 신뢰성을 보장하는 데 도움이 되는 표준, 도구, 테스트 개발하기 AI 안전 및 보안을 위한 새로운 표준 4: AI가 생성한 콘텐츠를 탐지하고 공식 콘텐츠를 인증하기 위한 표준 및 모범 사례를 수립하여 AI를 이용한 사기 및 속임수로부터 미국인을 보호하기 AI 안전 및 보안을 위한 새로운 표준 6: 보안에 있어 안전하고 윤리적이며 효과적인 AI 사용에 관한 국가 보안 각서 개발하기 미국인의 개인정보 보호 1-4	미국인의 개인정보보호 1: 개인정보보호 기술 개발 및 사용 촉진을 위한 연방정부의 지원을 우선시하여 미국인의 개인정보를 보호하기 미국인의 개인정보보호 2: 개인정보보호 연구 및 기술 강화하기 소비자, 환자, 학생을 위한 지원 1: 의료 분야에서 책임있는 AI 사용과 생명을 구할 수 있는 합리적인 가격의 의약품 개발 지원하기 혁신과 경쟁 촉진 1-3	근로자 지원 1-2 해외에서 미국 리더십 발전 1-3

원칙 기반 AI 거버넌스에 대한 공동의 약속은 AI 시스템이 일련의 핵심 가치와 윤리적 고려 사항에 부합하는 방식으로 발전해야 한다고 강조합니다. 이러한 접근방식은 다양한 맥락에서 유연하게 적용될 수 있도록 하는 동시에 전 세계적으로 책임있는 AI 기술 개발과 사용을 위한 공통의 기반을 제공합니다. 한국의 「디지털 권리장전」도 여러 국제 AI 정책과 연계되어 높은 정합성을 보이며 디지털 리더십 향상, 디지털 격차 해소, 국제 협력 등 AI를 넘어선 광범위한 디지털 이슈를 다룸으로써 그 범위를 확장하고 있습니다.

미국 NIST AI 위험 관리 프레임워크

원칙 1
디지털 환경에서의 자유와 권리 보장 (2장)

원칙 2
디지털에 대한 공정한 접근과 기회의 균등 (3장)

관리 3: 생애주기 전반에 걸쳐 AI 리스크를 매핑, 측정, 관리할 때 인력 다양성, 형평성, 포용성, 접근성의 우선시 여부

원칙 3
안전하고 신뢰할 수 있는 디지털 사회 (4장)

원칙 4
자율과 창의 기반의 디지털 혁신의 촉진 (5장)

원칙 5
인류 후생의 증진 (6장)

관리 1: AI 위험의 매핑, 측정 및 관리와 관련된 조직 전반의 정책, 프로세스, 절차 및 관행 준비 여부 및 투명하고 효과적인 구현 여부

관리 2: 적절한 팀과 인력이 AI 리스크를 매핑, 측정, 관리할 수 있는 권한과 책임 부여 및 교육받을 수 있는 구조 마련 여부

관리 4: 조직 내 AI 리스크를 고려하고 소통하는 문화의 우선시

관리 6: 타사 소프트웨어와 데이터 및 기타 공급망 문제로 인해 발생하는 AI 위험과 이점을 해결하기 위한 정책 및 조치 마련 여부

측정 2: AI 시스템의 신뢰성 평가 여부

관리 4: 식별·측정된 AI 위험에 대한 대응 및 복구를 포함한 위험 처리, 소통 계획의 명문화 및 정기적 검토

관리 5: AI 관련 주체의 강력한 참여를 위한 프로세스 준비 여부

파악 5: 개인, 단체, 지역사회, 조직, 사회 전반에 미치는 영향 파악

영국 AI 안전성 정상회의 '블레츨리 선언'

AI는 상당한 위험을 수반하므로 인권 보호, 투명성, 공정성, 책임성, 안전, 윤리적 고려 사항, 개인정보보호 등의 문제를 해결하기 위한 노력을 기울여야 한다.

AI는 안전하고, 인간 중심적이며, 신뢰할 수 있고, 책임있는 방식으로 설계, 개발, 배포, 사용되어야 한다.

AI는 상당한 위험을 수반하므로 인권 보호, 투명성, 공정성, 책임성, 안전, 윤리적 고려 사항, 개인정보보호 등의 문제를 해결하기 위한 노력을 기울여야 한다.

강력하고 잠재적으로 유해할 수 있는 AI 시스템 개발자는 안전성 테스트, 평가, 투명성, 책임 등 안전에 대한 중대한 책임이 있다.

특히 사이버 보안, 생명공학, 허위 정보 등의 위험을 이해하고 완화하는 데 중점을 두고, 피해를 유발할 수 있는 고기능 범용 AI 모델과 좁은 범위의 AI에 대해서는 특별한 주의가 필요하다.

공유된 AI 안전 위험을 파악하고, 과학적 이해를 구축하고, 위험 기반 정책을 개발하고, 투명성을 높이고, 프론티어 AI 안전에 대한 국제적으로 포괄적인 과학 연구 네트워크를 지원하는 데 중점을 둔다.

국가, 기업, 시민사회, 학계, 국제사회 등 모든 주체가 AI 안전을 보장하는 데 중요한 역할을 한다.

AI의 혁신적인 기회는 공공 서비스, 인권, 환경적 지속가능성, 유엔의 지속가능한 개발 목표의 달성 등 더 큰 공익을 위해 활용되어야 한다.

포용적인 국제 교류를 유지하고, 기존 국제 포럼에 참여하며, 광범위한 논의에 기여하고, AI의 이점을 책임감 있게 활용하기 위해 프론티어 AI 안전에 대한 연구를 지속한다.

공유된 AI 안전 위험을 파악하고, 과학적 이해를 구축하고, 위험 기반 정책을 개발하고, 투명성을 높이고, 프론티어 AI 안전에 대한 국제적으로 포괄적인 과학 연구 네트워크를 지원하는 데 중점을 둔다.

부록 D

디지털 공동번영사회를 위한 안전하고 신뢰할 수 있는 AI 시스템 구축을 위한 Microsoft의 모범 사례

책임있는 설계:

디지털 공동번영사회를 위한 AI 시스템 구축에 대한 접근방식

AI 거버넌스에 대한 위의 원칙 기반 접근방식이 효과적이라면 업계 참여자가 책임감을 지니고 이 원칙을 구현하려는 노력을 다해야 합니다. AI 산업을 주도하는 Microsoft와 같은 일부 기업은 다양한 이해관계자와 광범위하게 협력하여 AI 안전을 운영하고 책임과 의무를 다하고 있습니다.

AI의 공공 거버넌스에 대한 청사진

2023년 5월, Microsoft는 'AI 거버넌스: 미래를 위한 청사진'을 발표하여 공공 AI 거버넌스를 위한 포괄적인 청사진 5단계를 제시했습니다:

AI의 공공 거버넌스에 대한 청사진

- 01 정부 규제 기반 AI 안전 체계 구현 및 구축
- 02 중요 인프라를 제어하는 AI 시스템에 대한 효율적인 안전 장치 필수화
- 03 AI 기술 아키텍처를 기반으로 한 광범위한 법률 및 규제 프레임워크 고안
- 04 투명성 촉진 및 연구 및 비영리 목적에 대한 AI 개방
- 05 새로운 기술에 수반되는 사회적 과제 해결에 AI를 활용할 수 있도록 민간 파트너십 구축

첫째, 정부 주도로 새로운 AI 안전 체계를 구축하고 이를 구현합니다.

미국 국립표준기술연구소(NIST)에서 개발한 'AI 위험 관리 프레임워크'를 채택하고 강화하는 것을 중요하게 여깁니다. 정부는 NIST AI RMF에 맞춰 각자의 AI 거버넌스 프레임워크를 설계하여 글로벌 표준에 부합하는 수준으로 높일 수 있습니다. Microsoft의 백악관 자발적 안전 서약은 이러한 노력의 연장선상에 있습니다.

둘째, 중요 인프라를 제어하는 AI 시스템에 대한 효율적인 안전 장치가 필요합니다.

이 청사진은 중요 인프라를 감독하는 AI 시스템에 안전 메커니즘을 필수적으로 도입할 것을 강조합니다. 이러한 메커니즘은 다양한 기술의 안전 브레이크와 유사하며, 인간이 더 효과적으로 기술을 관리감독하고 시스템 복원성과 안정성을 보장하는 것을 목표로 합니다. 고위험 AI 시스템을 정의하고 페일 세이프 설계를 의무화해야 하며, 이러한 시스템은 추가적인 보안을 위해 허가된 AI 데이터센터에서 운영해야 합니다.

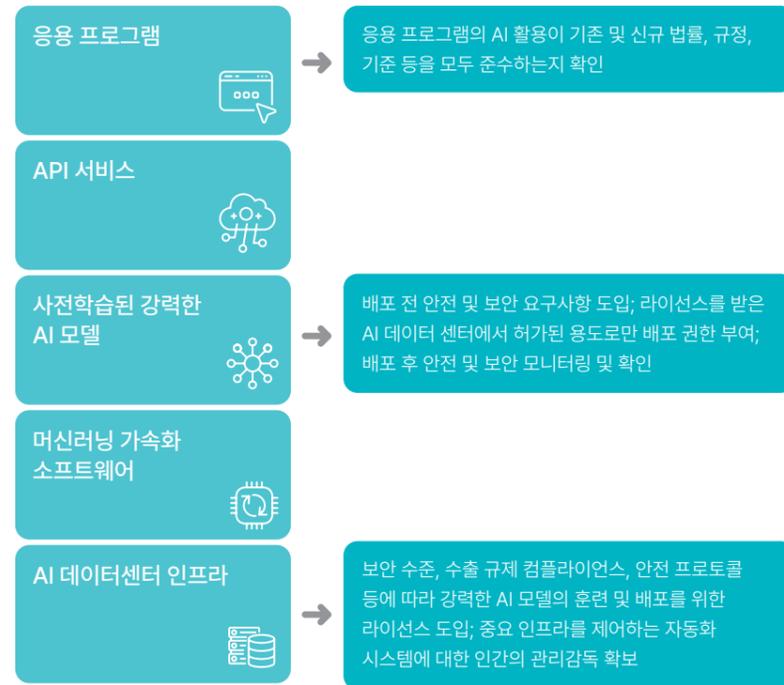
셋째, AI 기술 아키텍처를 기반으로 광범위한 법률 및 규제 프레임워크를 개발합니다.

여기에는 AI 기술 종류에 따라 담당 규제 기관을 다르게 두어 다방면으로 관리하고, AI 배포 및 사용과 관련된 위험을 관리하기 위해 '3C 원칙(클라우드, 고객, 콘텐츠)'을 적용하는 것이 포함됩니다. AI인프라에서 애플리케이션에 이르기까지 각각에 맞는 적절한 규제 아키텍처가 별도로 있어야 합니다.

AI 파운데이션 모델 기술 스택

응용 프로그램 	AI 모델의 출력을 활용해 최종사용자를 서비스하는 소프트웨어/프로그램
API 서비스 	응용 프로그램이 사전 학습된 모델에 접근하기 위한 API(응용 프로그램 프로그래밍 인터페이스) 또는 엔드포인트
사전학습된 강력한 AI 모델 	GPT-4와 같이 사전학습을 거쳐 다양한 작업을 수행할 수 있게 완성된 기초모델
머신러닝 가속화 소프트웨어 	거대 AI 모델의 개발 및 배포 프로세스를 가속화 및 최적화하는 소프트웨어
AI 데이터센터 인프라 	고대역폭 상호 연결을 갖춘 고급 GPU(그래픽처리장치) 클러스터를 포함한 고도화된 슈퍼컴퓨팅 인프라

AI 기술 아키텍처 기반 규제 체계안



넷째, 투명성을 높이고 학계와 비영리 단체의 AI 접근을 보장합니다. 투명성을 높이는 것은 AI 자원에 대한 접근성을 확대하는 데 매우 중요합니다. Microsoft는 매년 AI 투명성 보고서를 발간하고 기타 조치를 통해 AI 서비스의 투명성을 확대하기 위해 노력하고 있습니다. 학술 연구 및 비영리 목적으로 AI를 개방하는 것도 중요합니다. 학계 연구자가 훨씬 더 많은 컴퓨팅 리소스에 액세스할 수 없다면 AI와 관련된 과학기술 연구는 어려움을 겪게 될 위험이 있습니다.

다섯째, 새로운 기술에 수반되는 불가피한 사회적 과제를 해결하기 위한 효과적인 도구로 AI를 활용하기 위해 새로운 민간 파트너십을 추구합니다. AI의 이점을 활용하고 위험을 완화하기 위해서는 공공과 민간 간의 많은 협력이 필수적입니다. 이러한 파트너십은 민주주의를 보호하고 포용적 성장을 촉진하며 지속가능성에 기여할 수 있습니다.

‘백악관 행정명령’이 발표되기 전, 7개 주요 AI 기업(Microsoft, 구글, 아마존, 메타, 오픈AI 등)은 안전하고 보안이 유지되며 신뢰할 수 있는 AI 기술 개발을 입증하기 위해 미국 정부와 일련의 자발적 안전 서약을 맺었습니다.^{dxxxv} 이 서약은 안전, 보안, 신뢰라는 세 가지 핵심 원칙에 기반을 두고 있습니다.

먼저, ‘안전’은 공개적으로 AI 기술을 사용하기 전에 전문가의 내부/외부 보안 테스트를 거쳐겠다는 약속입니다. 이는 바이오, 사이버, 국가 안보 위험과 같은 AI 위험을

방지하는 역할을 할 것입니다. 또한 이 서약은 정부, 학계, 시민사회 등 다양한 주체가 모범 사례와 위험 관리의 시범 사례를 공유하도록 장려하는 역할도 합니다.

다음으로, ‘보안’ 측면에서는 보안 위험을 고려하여 사이버 보안 및 내부자 위협 보호 장치에 투자하겠다는 약속입니다. 또한 AI 시스템의 취약점을 쉽게 식별하고 관리할 수 있도록 제3자의 강력한 검증과 보고 체계를 구축하겠다는 것입니다.

마지막으로, 대중의 신뢰를 얻는다는 측면에서 이 원칙의 핵심 약속은 디지털 워터마킹 시스템과 같은 기술적 메커니즘을 개발하여 대중이 어떤 콘텐츠가 AI에 의해 생성되었는지 알 수 있게 하겠다는 것입니다. 또한 이 원칙은 기업이 AI 시스템의 능력과 한계를 공개적으로 논의하고, AI의 피해를 방지하기 위한 연구에 투자하며, 글로벌 과제를 해결하기 위해 더욱 발전된 AI 시스템을 구축해야 함을 강조합니다.

주요 AI 기업의 자발적인 참여로 첫 번째 물결이 형성되자 더 많은 기업이 이 같은 서명에 동참하게 됐습니다.^{dxxxvi} 이와 같은 민간 부문의 지지 표명은 미국 정부가 ‘백악관 행정명령’을 추진할 수 있는 탄탄한 토대가 되었습니다.

안전하고 신뢰할 수 있는 AI를 발전시키기 위해 노력하는 Microsoft

안전하고 책임있는 AI 관행을 확대하기 위해 Microsoft는 첨단 AI 시스템의 안전, 보안 및 신뢰성을 보장하기 위해 바이든-해리스 행정부가 시작한 새로운 자발적 안전 서약을 지지했습니다. 이러한 자발적 약속은 안전, 보안, 신뢰의 원칙을 따르며 첨단 AI 모델과 관련된 위험을 해결하고 NIST AI RMF 및 AI 관리장전 청사진과 같은 기존 미국 정부의 정책과 연계되어 있습니다.

한 걸음 더 나아가, Microsoft는 국립 AI 연구 자원 파일럿 프로젝트를 지원하고 투명성과 책임성을 강화하기 위해 고위험 AI 시스템의 국가 차원 등기를 지지함으로써 AI 생태계를 강화하기 위해 노력했습니다. 고객과 사회 전체에 도움이 되는 더 신뢰할 수 있는 AI 시스템을 개발하기 위해 Microsoft는 NIST AI RMF의 광범위한 구현과 AI의 고유한 위험에 맞춘 사이버 보안 관행의 채택을 지원하고 있습니다.



AI의 잠재적 위험을 관리하기 위한 주요 AI 기업의 자발적 안전 서약

백악관 자발적 안전 서약과 조응하는 Microsoft의 노력

안전

백악관 자발적 안전 서약:

기업은 레드 팀을 구성하고 신뢰 및 안전 정보를 공유하며 AI 생성 콘텐츠 식별할 수 있도록 한다

Microsoft의 노력:

- 레드 팀 활동 및 체계적인 측정을 통한 AI 시스템 테스트
- AI 신뢰성과 안전에 대한 신표준 개발을 위한 산업 차원 노력에 기여
- AI 콘텐츠 식별 위한 출처 표시를 구현
- NIST AI 리스크 관리 프레임워크 구현
- 고위험 모델 및 응용 프로그램에 대한 견고한 안전 확보 관행 구현

보안

백악관 자발적 안전 서약:

기업은 미공개 모델 가치를 보호하기 위해 사이버 보안에 투자하고 AI시스템 취약점의 책임있는 공개를 장려한다

Microsoft의 노력:

- AI 제품 및 서비스의 사이버 보안 위험 식별 및 완화
- 다중이해관계자 교류 참여 등을 통한 보안 위험 정보 공유
- 고성능 모델을 위한 라이선스 체계 개발 지원
- AI 서비스에 대한 '고객 인지' 개념 확장 기여

신뢰

백악관 자발적 안전 서약:

기업은 시스템의 능력과 한계에 대해 투명하게 공개하고 AI의 사회적 위험에 대한 연구를 우선시하며, 대중의 이익을 위해 AI 시스템을 개발하고 배포한다

Microsoft의 노력:

- 책임 있는 AI 거버넌스에 대한 연례 투명성 보고서 발간
- AI 시스템 이용 여부 고지 및 AI 시스템의 능력과 한계에 대한 투명성 확보
- 학술 연구 프로그램 투자 증대
- 국립과학재단과 협력 하에 국가 AI 연구 자원 구축을 위한 시범 프로젝트 검토
- 고위험 AI 시스템을 등록할 국가 데이터베이스 개발 지원

*빨간 하이라이트 표시는 Microsoft의 추가적 안전 서약을 나타냄

AI 안전성 정상회의

「블레츨리 선언」 외에도 AI 안전성 정상회의에서 Microsoft를 비롯한 AI 기업은 정부와 협력하여 AI 모델 테스트에 관해 합의를 했습니다.^{dxxxvii} 이 자발적 협약에 따라 주요 개발자는 위험을 완화하고 취약성을 관리하기 위해 “평가, 투명성 및 기타 적절한 조치를 채택하여 안전 테스트를 만들고 실행”할 책임을 지는 데 동의합니다.^{dxxxviii} 또한 이 협약에 따라 규제 기관과 개발자는 배포 전후에 새로운 AI 모델의 안전성을 테스트하기 위해 협력할 수 있게 됩니다.

이 협약을 운영하기 위해 개발자와 참여 정부(미국, 영국, EU 등)는 AI 안전 연구소의 주도하에 테스트를 개발할 예정입니다.

선도적인 테크기업은 AI 시스템을 설계할 때 책임있는 AI를 구현하고 운영함으로써 모범을 보여야 합니다.

선도적인 테크기업은 책임있는 AI 환경을 조성하는 데 중추적인 역할을 합니다. AI 시스템의 구조에 책임있는 AI 원칙을 구현하고 이를 위한 충분한 사전 조치를 취해 모범을 보여야 합니다. 이 섹션에서는 AI 설계에 대한 책임있는 가이드라인, 강력한 거버넌스 구조, 투명한 책임 메커니즘 수립 등, Microsoft가 취한 접근방식을 소개합니다. 이러한 업무 선례는 더 많은 테크기업이 책임있는 AI 관행을 채택하도록 영향을 주고 사회에 도움이 되는 표준을 세우는 데도 기여할 수 있습니다.

Microsoft의 책임있는 AI를 구현해 온 여정

2016년 Microsoft는 사티아 나델라(Satya Nadella)가 제시한 비전에 따라 책임있는 AI에 집중하기 시작했습니다. 개인정보보호, 보안, 접근성에 대한 오랜 기간의 전사적 노력에서 얻은 교훈을 바탕으로 책임있는 AI는 회사 최고 경영진의 전폭적인 지지하에 Microsoft 전사 내 각 부서의 옹호를 받아야 한다는 것을 깨달았습니다. 2018년에는 전문가로 구성된 Aether 위원회를 설립해 책임있는 AI 원칙을 만들고 AI 연구개발에 대한 자문을 받았습니다. 2019년에는 거버넌스를 감독하고 AI 시스템 구축 시 실행 가능한 지침이 될 수 있는 책임있는 AI 표준을 작성하는 책임있는 AI 사무소를 설립하여 이러한 노력을 더욱 강화했습니다. 실무 수준에서는 엔지니어링, 연구 및 정책 부서의 핵심 팀이 회사 전체에서 책임있는 AI를 발전시키는 데 중요한 역할을



했고 각 팀에서 고유한 전문성을 제공하도록 했습니다. 또한 경영진이 책임있는 AI 챔피언을 임명하여 제품, 엔지니어링 및 영업 팀에서도 책임있는 AI를 실현할 수 있도록 관리감독하는 역할을 부여했습니다. 다음 단계로 책임있는 AI 위원회를 조직했습니다. 이는 Microsoft의 책임있는 AI 프로그램을 실행하기 위한 리더십 조정 및 책임을 위한 포럼입니다. 가장 높은 수준에서는 Microsoft 이사회의 환경, 사회 및 공공 정책 위원회가 책임있는 AI 준수 및 구현 여부를 감독합니다.

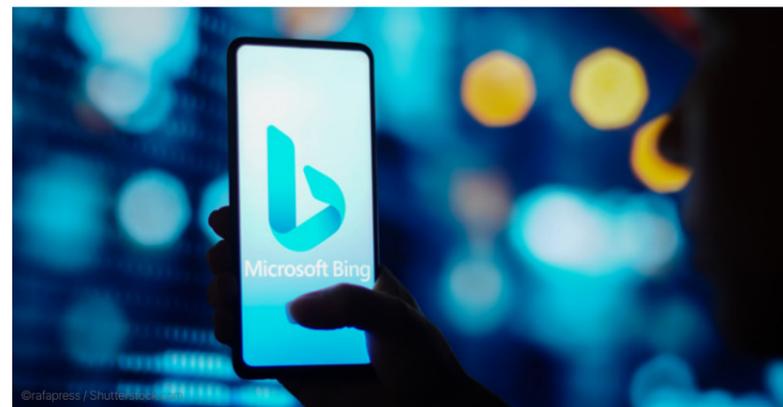
책임있는 AI 기준의 구성요소

원칙	우리의 책임 있는 AI란 어떤 보편적 가치를 기반으로 하는가?
목표	우리가 이뤄내야 하는 성과는 무엇인가?
필요조건	목표를 이루기 위해 우리가 취해야 하는 단계는 무엇인가?
도구 및 실행 방법	필요조건을 충족하기 위해 우리가 활용할 수 있는 보조수단은 무엇인가?



사례 연구: Microsoft가 책임있는 AI 접근방식을 새로운 Bing에 적용한 방법

2023년 2월, Microsoft는 AI로 강화된 웹 검색 환경인 새로운 Bing을 출시했습니다. 이 기능은 웹 검색 결과를 요약하고 채팅 환경을 제공함으로써 사용자를 지원합니다. 또한 사용자는 시, 농담, 편지 등 창의적인 콘텐츠를 생성할 수 있으며, Bing 이미지 크리에이터를 통해 이미지를 만들 수도 있습니다. AI가 강화된 새로운 Bing은 OpenAI의 최첨단 대규모 언어 모델(LLM)인 GPT-4를 비롯한 Microsoft와 OpenAI의 다양한 첨단 기술을 기반으로 실행됩니다. Microsoft의 책임있는 AI 팀은 OpenAI가 공개되기 몇 달 전부터 GPT-4와 협력하여 이 최첨단 AI 기술과 웹 검색을 새로운 Bing에 결합하기 위한 맞춤형 기능과 기술 세트를 개발했습니다.



Bing의 출시를 준비하면서 Microsoft는 책임있는 AI 에코시스템의 모든 역량을 활용했습니다. 새로운 Bing 환경은 Microsoft의 AI 원칙과 Microsoft의 책임있는 AI 표준에 따라 개발되었으며, Microsoft의 책임있는 AI 사무소, 엔지니어링 팀, Microsoft Research 및 Aether 위원회를 비롯한 회사 전반의 AI 윤리 및 안정성 전문가와 협력하여 개발되었습니다.

AI 원칙과 책임있는 AI 표준에 따라 새로운 Bing의 잠재적인 피해와 오용을 식별·측정·완화하는 동시에 새로운 경험이 제공하는 혁신적이고 유익한 용도를 확보하고자 했습니다. 다음 섹션에서는 책임있는 AI 접근방식을 구현하기 위해 기업이 따를 수 있는 모범 사례를 강조하면서 이들의 접근방식을 설명합니다.

식별

2022년 늦여름에 Microsoft는 모델 수준에서 GPT-4에 대한 탐색적 분석 작업을 착수했습니다. 여기에는 OpenAI와 협력하여 광범위한 레드 팀 작업을 수행하는 것이 포함되었습니다. 이 테스트는 최신 기술이 추가적인 안전장치 없이 어떻게 작동하는지를 평가하기 위해 고안되었습니다. 이 테스트의 구체적인 목적은 유해한 응답(이 경우 응답은 AI 시스템의 출력물이며 '완성물', '생성물', '답변'이라고도 함)을 생성하고, 오용될 수 있는 잠재적 경로를 드러내며, 기능과 한계를 파악하는 것이었습니다. 이러한 학습을 통해 OpenAI의 모델 개발을 발전시키고, Microsoft의 위험에 대한 이해를 높였으며, 새로운 Bing에 대한 조기 완화 전략을 수립하는 데 기여했습니다.

여러 분야의 전문가로 구성된 레드 팀에서는 모델 수준의 레드 팀 작업 외에 새로운 Bing AI를 경험할 수 있는 한정판 미리보기로 제공하기 전에 애플리케이션 수준의 레드 팀 작업을 여러 차례 수행했습니다. 이 과정을 통해 공격자가 시스템을 악용할 수 있는 방법을 더 잘 이해하고 완화 기능을 개선할 수 있었습니다. 또한 비적대적 테스터들이 새로운 Bing 기능의 단점과 취약점을 광범위하게 평가했습니다.

측정

레드 팀은 특정 유해 사례를 찾아낼 수 있지만, 실제 운영 환경에서는 사용자가 새로운 Bing과 수백만 가지의 다양한 대화를 나누게 됩니다. 또한 대화는 여러 방향으로 진행되며 상황에 따라 달라지므로 대화 내에서 유해한 반응을 식별하는 것은 복잡한 작업입니다. 새로운 Bing AI의 사용 시 발생할 수 있는 잠재적 피해를 더 잘 이해하고 해결하기 위해 탈옥(잠금장치 해킹), 유해 콘텐츠, 근거 없는 콘텐츠와 같은 잠재적 피해를 측정하기 위해 새로운 AI 체함에 특화된 책임있는 AI 지표를 추가로 개발했습니다. 또한 부분적으로 자동화된 측정 파이프라인을 통해 대규모로 측정할 수 있게 되었습니다.

측정 파이프라인을 통해 잠재적 위험에 대한 대규모 측정을 신속하게 수행하여 프로덕션에 적용하기 전에 각 변경 사항을 테스트할 수 있습니다. 프리뷰 기간과 그 이후에도 지속적인 레드 팀 작업을 통해 새로운 문제를 파악하면서 추가적인 피해를 평가하기 위해 측정 세트를 계속 확장하고 있습니다.

완화

잠재적인 피해와 오용을 파악하고 측정하면서 기존 검색에 사용되던 방식에 추가적인 완화 조치를 개발했습니다. 그중 일부는 다음과 같습니다.

▶ **프리뷰 기간 도입 및 단계적 출시:** 점진적 출시 전략은 연구실에서 안전하게 기술을 세상에 내놓는 핵심적인 부분이었으며, 새로운 Bing의 이점을 확보하기 위한 신중하고 사려 깊은 접근이었습니다. 프리뷰 기간에 접근 권한을 가진 사람의 수를 제한하여 사람들이 새로운 Bing을 사용하는 방식과 오용할 수 있는 부분을 파악한 후에 정식 출시를 진행했습니다. 또한 제품 성능과 기존의 완화 기능을 개선하고, 학습한 내용에 따라 새로운 완화 기능을 구현하기 위해 매일 새로운 Bing으로 계속 변경하고 있습니다.

▶ **유해성 또는 오용 완화 목적의 AI 기반 분류기와 메타 프롬프트:** LLM을 사용하면 유해하거나 오용될 수 있는 콘텐츠가 생성될 수 있습니다. 분류기와 메타 프롬프트는 이러한 유형의 콘텐츠의 위험을 줄이기 위해 새로운 Bing에 구현된 두 가지 완화 조치입니다. 분류기는 텍스트를 분류하여 검색어나 채팅 프롬프트 또는 생성된 응답에서 잠재적으로 유해한 콘텐츠의 다양한 유형에 플래그를 지정합니다. 플래그는 생성된 콘텐츠를 사용자에게 반환하지 않거나 사용자에게 다른 주제로 전환하게 하거나 사용자를 기존 검색으로 리디렉션하는 등의 잠재적인 완화 조치를 합니다. 메타 프롬프트에는 모델의 행동을 안내하기 위해 모델에 지침을 제공하는 것이 포함됩니다. 예를 들어, 메타 프롬프트에는 “사용자가 선택한 언어로 대화하세요.” 라는 문구가 포함될 수 있습니다.

▶ **검색 결과 기반 답변:** 새로운 Bing은 사용자가 정보를 검색할 때 웹 검색 결과의 정보가 뒷받침하는 응답을 제공하도록 설계되었습니다. 예를 들어, 시스템에는 상단 검색 결과의 텍스트와 메타 프롬프트를 통해 응답의 근거가 되는 지침이 제공됩니다. 그러나 웹의 콘텐츠를 요약할 때 새로운 Bing은 입력 소스에 없는 정보를 응답에 포함할 수 있습니다. 즉, 근거가 없는 결과를 생성할 수 있습니다. 요약 시나리오와 채팅 경험에서 사용자가 근거 없이 생성된 콘텐츠에 과도하게 의존할 수 있는 위험을 낮추기 위해 몇 가지 조치를 취했습니다. 예를 들어, 검색 결과를 기반으로 하는 새로운 Bing의 응답에는 사용자가 응답을 확인하고 자세히 알아볼 수 있도록 소스 웹사이트에 대한 참조가 포함됩니다. 또한 사용자는 AI 시스템과 상호작용을 하고 있음을 명시적으로 알 수 있으며, 최선의 판단을 내릴 수 있도록 웹 결과의 소스 자료를 확인하도록 권장됩니다.

▶ **대화 흐름 제한:** 프리뷰 기간에 채팅 세션이 너무 길어지면 반복적이거나 도움이 되지 않거나 새로운 Bing의 의도된 어조와 일치하지 않는 응답이 나올 수 있다는 사실을 알게 되었습니다. 이러한 대화 이탈 문제를 해결하기 위해 시스템을 업데이트하여 문제를 개선할 때까지 채팅 세션당 턴(사용자의 질문과 Bing의 답변이 모두 포함된 교환) 횟수를 제한했습니다.

▶ **AI 시스템 사용 공시:** 새로운 Bing은 의미 있는 AI 시스템 사용 공시를 위해 여러 접점을 제공하며, 이를 통해 사용자는 AI 시스템과 상호작용을 하고 있다는 알림과

함께 새로운 Bing에 대해 자세히 알아볼 수 있는 기회를 얻게 됩니다.

피해를 식별·측정·완화하는 Microsoft의 접근방식은 프리뷰 기간과 그 이후에 수집된 피드백을 바탕으로 더 많은 것을 학습하고 개선하여 계속 발전할 것입니다.



사례 연구: Microsoft가 책임있는 AI를 구현해 온 여정에서 고객의 역량을 강화하는 방법 - Azure의 온라인 안전 조치

Microsoft의 책임있는 AI에서 가장 중요하게 여기는 약속은 고객과 학습 내용을 공유하여 고객의 책임있는 AI 여정을 돕겠다는 것입니다. 책임있는 AI 관행이 채택될 때 예상되는 사회적 이득을 확보하는 데는 고객사의 노력만으로는 충분하지 않다는 것을 잘 알고 있습니다.

이러한 노력의 일환으로 Microsoft는 고객이 시스템을 책임감 있게 배포할 수 있도록 플랫폼 AI 서비스에 대한 투명성 관련 자료를 문서화해 <투명성 노트>라는 이름으로 제공합니다. <투명성 노트>는 고객이 플랫폼 기술을 배포할 시기와 방법을 이해할 수 있도록 AI 시스템의 목적, 기능, 한계를 일상적인 언어로 명확하게 전달합니다. 또한 솔루션의 기능과 책임있는 AI 표준에서 벗어나는 사용 사례를 식별합니다. <투명성 노트>는 마케팅 문서와 기술 문서 사이의 간극을 메워 고객이 책임감 있게 AI를 배포하기 위해 알아야 할 정보를 선제적으로 전달합니다.^{dxxxix}

고객에게도 책임있는 AI 사례를 운영할 수 있는 실용적인 도구가 필요합니다. 수년에 걸친 Microsoft의 책임있는 AI 연구는 Fairlearn 및 InterpretML과 같은 도구의 인큐베이팅으로 이어졌습니다. 이러한 도구 모음은 AI 시스템의 잠재적 오류와 한계를 식별·진단·완화하는 기능을 포함하여 책임있는 AI 사례의 여러 측면을 포괄하는 기능으로 발전해 왔습니다. 이러한 도구는 Microsoft 내부에서 처음 구상된 이후에 활발한 오픈 소스 커뮤니티의 도움으로 외부에서 계속 개선되고 발전하고 있습니다. 도구 모음은 책임있는 AI 툴박스인 GitHub 리포지토리에서 찾을 수 있습니다. 현재



프리뷰로 제공되는 최신 도구인 Azure Content Safety는 이미지와 텍스트의 모든 언어에서 증오, 폭력, 성적, 자해 콘텐츠를 감지하도록 설계된 모델을 통해 기업이 더 안전한 온라인 환경과 커뮤니티를 만들 수 있도록 지원합니다.

책임있는 AI 툴박스를 기반으로 하는 Microsoft의 책임있는 AI 프로그램은 고객이 엔지니어링 시스템과 도구 개발의 이점을 누릴 수 있도록 좀 더 성숙한 책임있는 AI 도구 중 일부를 Azure 머신러닝에 직접 통합하는 데 투자했습니다. 책임있는 AI 대시보드라고 하는 이 기능 모음은 머신러닝 실무자와 비즈니스 이해관계자가 모델을 디버깅하고 AI 시스템을 빌드하거나 기존 시스템의 사용자를 지정할 때 정보에 입각한 책임있는 결정을 내릴 수 있도록 단일 창을 제공합니다. 프리뷰 버전에 추가된 최신 기능에는 사용자가 모델 구축 단계와 학습 및 평가 단계에서 비정형 데이터로 구축된 대규모 모델을 평가할 수 있는 텍스트 및 이미지 데이터를 지원하고, 근거성 등의 측정을 포함하여 대규모 언어 모델을 프롬프트 평가 조정할 수 있는 환경을 간소하게 제공하는 프롬프트 흐름이 포함됩니다.



Microsoft는 책임있는 AI 혁신을 위해 직접 연구한 결과를 고객의 책임있는 AI 여정을 지원하는 실용적인 도구와 자료로 전환하는 데 투자해 왔으며 이는 앞으로도 계속될 것입니다.

AI의 개발·평가·사용에 관여하는 커뮤니티는 직접적인 고객을 넘어 확장되고 있습니다. 이렇듯 커져만 가는 디지털 생태계를 지원하기 위해 Microsoft는 책임있는 AI 표준, 영향평가 템플릿, 최첨단 연구 자료집 등 책임있는 AI 프로그램의 주요 결과물을 공개적으로 공유합니다. 또한 디지털 학습 경로를 통해 리더가 효과적인 AI 전략을 수립하고, AI에 대비한 문화를 조성하며, 책임감 있게 혁신하는 역량을 강화할 수 있도록 지원합니다.^{cxv}

저작권 존중 및 고객 보호

2023년 9월, Microsoft는 AI 기반 Copilot과 관련된 저작권 문제에 대한 우려를 해결하기 위해 Copilot 저작권 약정을 도입했습니다.^{cxvi} 이 약정은 고객이 잠재적인 저작권 청구에 대한 걱정 없이 Copilot 서비스와 이로부터 생성된 콘텐츠를 사용할 수 있도록 보장합니다. Microsoft는 고객 측의 저작권 침해로 발생하는 문제의 모든 법적 위험에 대해 책임을 집니다.

Microsoft는 AI 고객 약정을 기반으로 2023년 12월에 Copilot 저작권 약정을 개정하여 고객 저작권 약정(Customer Copyright Commitments, CCC)을 도입했습니다. 이 약정은 유료 버전의 Microsoft 상업용 Copilot 서비스(업무용 ID로 로그인한 경우 Windows Copilot 포함), Microsoft Copilot(이전의 Bing Chat Enterprise), Azure OpenAI 서비스, 구성 가능한 메타 프롬프트 및 기타 안전 시스템을 갖춘 기타 생성형 AI 서비스에 의해 생성된 콘텐츠와 관련된 특정 타사 지식재산권 청구로부터 고객을 보호하기 위해 Microsoft가 노력하고 있음을 강조합니다.^{cxvii}

AI 기술 아키텍처 기반 규제 체계안

공정성

모든 사람이 모델을 사용할 수 있도록 모델의 공정성을 검증하고 혹시 모를 이슈를 확인합니다.



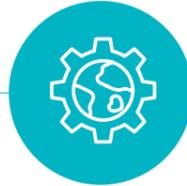
설명가능성

모델의 예측 결과를 이해하기 위해 모델에 대한 주요변수를 추출합니다.



반사실적 설명

주요변수의 변화를 관찰하여 모델의 예측 결과가 다른 가장 가까운 데이터 포인트를 찾습니다.



프롬프트 플로우

대규모 언어 기반 응용 프로그램을 위한 워크플로우를 생성하여 프롬프트 빌딩, 평가 및 조정을 간소화합니다.



인과 분석

주요변수가 실제 값에 미치는 효과를 추정합니다.



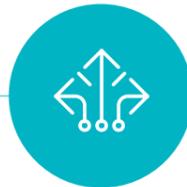
오류 분석

높은 오류율을 가진 데이터셋 코호트를 식별하고 모델 내의 오류 분포를 시각화합니다.



책임있는 AI 스코어카드

책임있는 AI 적용에 대한 인사이트를 PDF로 제공하여 기술적 및 비기술적 이해관계자와의 공유 및 컴플라이언스 점검을 가능케 합니다.



Azure 콘텐츠 안전성

이미지 및 텍스트 입력력값에서 혐오, 폭력, 성적 자극 및 자해 콘텐츠를 감지합니다.



Microsoft의 고객 저작권 약정¹

고객의 볼륨 라이선스 계약에 따라 제3자의 IP 클레임으로부터 고객을 방어해야 하는 Microsoft의 의무는 다음 조건이 모두 충족되는 경우에 적용 대상 제품의 생성 콘텐츠에 대한 고객의 사용 또는 배포에 적용됩니다.

- 01 고객은 청구의 대상이 되는 출력 콘텐츠를 제작하기 위해 대상 제품을 사용하는 동안 대상 제품의 일부인 콘텐츠 필터, 메타 프롬프트의 제한 또는 기타 안전 시스템을 비활성화·회피·중단·간섭하지 않아야 합니다.
- 02 고객은 제3자의 소유권을 침해하거나 오용할 가능성이 있음을 알거나 알 수 있는 방식으로 출력 콘텐츠를 수정·사용·배포하지 않아야 합니다.
- 03 고객은 청구의 대상이 되는 출력 콘텐츠를 생성한 모델의 사용자를 지정하는 데 사용한 고객 데이터를 포함, 대상 제품과 관련하여 입력을 넣을 충분한 권리를 갖습니다.
- 04 이 청구는 상거래 또는 거래 과정에서 사용되는 출력 콘텐츠가 제3자의 상표 또는 관련 권리를 침해한다고 주장하지 않습니다.
- 05 구성 가능한 메타 프롬프트 또는 기타 안전 시스템을 갖춘 Azure OpenAI 서비스 및 Microsoft 생성 AI 서비스의 경우, 고객은 클레임의 대상이 되는 출력 콘텐츠를 제공하는 모든 제품에서 Azure OpenAI 서비스 설명서에서 요구하는 모든 완화 조치를 구현해야 합니다.

¹ Microsoft(2023). 고객 저작권 약정 필수 완화 조치. 출처: <https://learn.microsoft.com/en-us/legal/cognitive-services/openai/customer-copyright-commitment>

이 약정은 고객의 편에 서서 제품 사용으로 인해 법적 문제가 발생하지 않도록 하겠다는 Microsoft의 약속에 근거를 두고 있습니다. 또한 저작자의 우려를 인식하고 사회적 이익을 위한 AI 기술 발전과 저작권 존중 사이의 균형을 맞추는 것을 목표로 합니다. Microsoft는 자사 제품이 저작권을 침해하는 콘텐츠를 생성할 가능성을 줄이기 위해 필터와 기술을 구현했으며, 고객이 이러한 기술을 사용하여 저작권 문제를 미연에 방지하도록 장려하고 있습니다. 고객 저작권 약정은 Microsoft의 상업용 Copilot 서비스(업무용 ID로 로그인한 경우 Windows Copilot 포함), Microsoft Copilot(이전의 Bing Chat Enterprise) 및 Azure OpenAI 서비스의 유료 버전에 적용됩니다.

부록 E

AI 산업이 한국 경제에 미칠 잠재적 영향에 대한 사례 연구

기술 혁신의 최전선에 서 있는 한국에서 AI의 새로운 지형은 단순히 인간의 독창성을 증명하는 것이 아니라 가장 시급한 사회적 과제를 해결하는 데 중요한 도구가 되고 있습니다.

AI는 한국 사회의 다양한 이슈에 걸쳐 혁신적인 솔루션을 제공하고 있습니다. 이 섹션에서는 이슈별로 한국 기업이 현실의 문제에 대응하기 위해 AI의 힘을 어떻게 활용했는지를 소개합니다. 이러한 이슈에는 다음과 같은 내용이 포함됩니다:

- 01 의료 분야의 AI
- 02 교육 분야의 AI
- 03 경제 발전을 위한 AI
- 04 공공 서비스에서의 AI

다양한 분야에서 AI가 구현되면서 의료와 교육부터 사회적 담론과 경제 발전에 이르기까지 모든 분야에 영향을 미치며 큰 도약을 이루고 있습니다. AI의 스마트 애플리케이션은 환자 치료 방법을 개선하고, 교육을 개인화하며, 온라인 커뮤니티를 조정하고, 비즈니스 혁신을 주도하고 있습니다. 공공 서비스 분야에서는 비상 대응과 도시 계획을 개선하여 포괄적인 발전을 위해 AI를 수용하려는 국가의 혁신적인 노력을 엿볼 수 있습니다.

한국적 맥락에서 다양한 사례 연구를 통해 얻은 정보를 종합하여 영역별로 AI를 전략적으로 배치할 수 있는 방안을 다음과 같이 제시합니다:

한국 고령화 사회의 AI 기반 헬스케어 혁신

한국은 2030년에는 노인 인구가 전체 인구의 24%를 차지할 것으로 예상되는 인구 고령화 문제에 직면해 있으며, 이로 인해 노인 관련 질병의 유병률이 증가할 것으로 예상됩니다. 이러한 문제를 해결하기 위해 이화여자대학교 최선 교수팀은 Microsoft 리서치의 AI4Science in Asia 팀과 협력하여 혁신적인 AI 기반 연구 이니셔티브를 착수했습니다. 연구팀은 노화 관련 질환에 중요한 역할을 하는 G단백질결합수용체 (GPCR)를 표적으로 하는 분자 생성 모델인 TamGent를 개발하는 데 주력해 왔습니다. TamGent는 신약 개발에서 AI의 혁신적인 기능을 활용하여 단백질 복합체를 직접 표적으로 삼아 새로운 화합물을 생성할 수 있으므로 기존의 화합물 라이브러리 검색을 우회할 수 있습니다. 이러한 접근방식은 신약 개발 프로세스를 크게 가속화하여 잠재적으로 비용을 절감하고 노인을 위한 의료 서비스의 질을 높일 수 있습니다. 이 연구는 고령화 인구에 대한 국가적 의료 재정 부담을 해결하는 데 기여하고, 한국 노인의 건강과 복지를 향상할 수 있는 유망한 솔루션을 제공합니다.

인공지능을 통한 국내 신약 개발 및 임상시험 발전

서울대학교 의과대학 이형기 교수가 주도하고 Microsoft 연구소가 지원하는 한 프로젝트에서는 한국의 인구 고령화에 큰 영향을 미칠 신약 개발 관련 문제를 해결하기 위한 연구가 진행 중입니다. 첫 번째 문제는 기존 약물의 새로운 용도를 발견하는 데 중요한 과정인 약물 용도 변경의 비효율성입니다. 연구팀은 AI를 활용해 약물 적응증과 약물 특성 데이터의 복잡성을 탐색함으로써 숏컷 학습(Shortcut learning)을 완화하고 약물-질병 연관성의 정확성을 높이는 것을 목표로 합니다. 이러한 접근방식은 약물의 용도 변경 프로세스를 간소화하여 효과적인 치료법을 더 효율적으로 식별할 수 있는 경로를 제공할 것으로 기대됩니다.



MS AI 스쿨^{cxclii} : 한국의 내일을 위한 미래 인재 양성

AI는 특히 전문 교육과 취업 준비 영역에서 교육 환경을 혁신하고 있습니다. Microsoft 필란트로피의 이니셔티브인 MS AI 스쿨은 숙련된 AI 전문가가 현저히 부족한 한국의 고용 격차를 해소하기 위해 AI 직무 교육을 제공함으로써 이러한 변화의 모범을 보여주고 있습니다.

이 프로그램은 프랑스에서 시작된 Microsoft AI 스쿨 모델을 한국 실정에 맞게 현지화한 것으로, 정부 및 산업계와의 파트너십을 통해 인센티브와 수당이 포함된 종합적인 교육 과정을 제공합니다. 이 이니셔티브는 구직자에게 AI 기술을 제공할 뿐만 아니라 소프트웨어정책연구소(SPRI)가 지적한 인력 부족 문제도 직접적으로 해결합니다.

AI가 교육에 미치는 영향이 얼마나 큰지는 개인적인 사례를 통해 알 수 있습니다. 예를 들어, 이승윤 씨는 문화유산 보존 분야에서 일하던 중 MS AI 스쿨에 참여한 후 AI와 빅데이터 분야로 이직했습니다. 그는 6개월간의 집중 교육 프로그램을 통해 기본 프로그래밍부터 딥 러닝과 같은 첨단 개념까지 AI에 대해 폭넓게 이해하게 되었습니다. 이러한 성공 사례는 MS AI 스쿨이 다양한 학문적 배경을 가진 사람들이 급성장하는 AI 분야로 진출할 수 있도록 얼마나 잘 지원하고 있는지를 보여줍니다.

MS AI 스쿨의 교육 품질은 졸업생들의 높은 만족도로 입증되었습니다. 실제 경험을 갖춘 유능한 강사진이 수준 높은 강의를 제공하며, 이는 독학 학습을 점점 더 선호하는 시대에 큰 장점으로 꼽힙니다.

황혜정 씨는 학교와 Microsoft가 직접 연계된 점, AI에 몰입할 수 있는 기회를 얻은 것, 고사양의 Microsoft Cloud Azure 가상 머신을 사용할 수 있는 것을 높이 평가했고, Microsoft 인증 Azure 자격증을 취득할 수 있는 것을 주요 혜택으로 꼽았습니다. 이 프로그램의 탄탄한 커리큘럼은 이론적 지식과 실제 적용이 긴밀하게 연계되어 있어서 업계 관계자에게도 높은 점수를 받고 있습니다.

MS AI 스쿨은 팀 프로젝트를 통해 협업 학습 환경을 조성하여 협업 능력을 배양하고 네트워킹, 취업 정보 등 다양한 시너지 효과를 제공합니다. 이러한 팀 기반 접근방식은 교육 경험을 확장하고 배움의 즐거움이 배가되게 합니다. 남성 위주의 IT 업계에서

두각을 나타내고자 하는 황혜정 씨와 같은 교육생들은 이 프로그램을 컴퓨터 비전과 같은 전문 분야로 즉시 취업할 수 있는 발판으로 여기고 있습니다. 취업 시장이 위축되는 상황에서, MS AI 스쿨에 참여한 경력은 인사 담당자와의 대화를 이끌어낼 수 있는 귀중한 자산으로 여겨지고 있습니다. 기술 혁신으로 인한 일자리 감소에 대한 세계경제포럼의 예측을 고려할 때 AI 교육에 집중하는 것은 시의적절하며, 이는 고용 시장에서 AI 전문 지식이 점점 더 중요해지고 있음을 나타냅니다.

경제 발전을 위한 AI

인공지능을 활용한 한국 문학 유산의 부활^{cxciv}

서강대학교의 한국학 학자인 웨인 드 프레메리 박사는 Azure와 머신러닝 기술을 활용하여 교육 및 문화유산에 AI를 통합하는 데 앞장서고 있습니다. 그의 프로젝트는 (16세기 영시의 복잡성과 유사하게) 11,000개가 넘는 고유 알파벳과 고대 한자 표의문자가 밀집되어 있는 한국 역사 텍스트를 디지털로 부활시키는 데 중점을 두고 있습니다.

오늘날 한국 학생들의 과제는 고전 텍스트를 단순히 읽는 것이 아니라 이해하는 것입니다. 드 프레메리 박사는 한국의 고전 텍스트를 AI가 해독하여 현대의 독자가 접근할 수 있고 이해할 수 있도록 하기 위해 노력하고 있습니다. Azure에서 머신러닝을 사용하면 문헌 내의 복잡한 패턴을 분석하여 미묘한 번역을 제공하고 한국 문학사를 더 잘 이해할 수 있게 됩니다.

이 기술의 활용은 교육 참여와 문화 보존에 상당한 영향을 미칩니다. 한국의 젊은이들과 그들의 유산을 연결하고, AI를 활용해 전 세계의 문화유산을 발굴하는 모델을 제시함으로써 역사적인 유산을 보존하고 연구하는 방식을 혁신할 수 있습니다.

두산중공업과 함께하는 벤틀리시스템즈^{cxcv} :

대규모 AI 기반 인프라 관리 솔루션

인프라 엔지니어링 소프트웨어 분야의 선두주자인 벤틀리시스템즈는 Microsoft Azure 플랫폼에서 디지털 트윈과 AI의 힘을 활용하여 글로벌 인프라 관리를 개선해 왔습니다. 이러한 첨단 기술 솔루션은 서로 다른 데이터 소스를 통합하여 업계 전문가가 통찰력을 얻고 정보에 입각한 의사결정을 내릴 수 있도록 지원합니다.

벤틀리시스템즈는 디지털 트윈을 사용하여 인프라 개발과 유지보수에 대한 접근방식을 혁신하여 물리적 시스템의 동적이고 데이터가 풍부한 표현을 제공했습니다. 이는 많은 기업에 경제적으로 유익하다는 것이 입증되었습니다. 그중에서도 벤틀리시스템즈와 한국 두산중공업의 협업은 디지털 트윈을 통한 경제 발전 촉진에 AI를 적용한 대표적인 모범 사례입니다. 한국에서 엔지니어링 및 건설 서비스와 함께 발전 서비스를 제공하는 두산중공업은 Microsoft Azure를 기반으로 하는 벤틀리시스템즈의 기술을 활용하여 풍력발전 단지를 위한 첨단 디지털 트윈을 생성했습니다. 이 가상 모델은 Azure의 강력한 클라우드 컴퓨팅 기능을 통해 통합된 사물 인터넷(IoT)과 기상 데이터를 포함한 실시간 데이터와 과거 데이터를 정교하게 결합한 것입니다.



두산중공업의 풍력발전소를 위해 개발된 디지털 트윈은 기술 혁신이 경제 발전에 기여한 대표적인 사례입니다. 디지털 트윈을 통해 터빈을 원격으로 모니터링하고 실시간 운영 효율성을 높일 수 있습니다. AI 요소는 물리적 기반의 머신러닝 알고리즘을 사용하여 다양한 기상 조건에 따라 에너지 생산량을 예측합니다. 이러한 예측을 통해 두산중공업은 각 터빈의 성능을 미세 조정하여 에너지 생산량을 극대화함으로써 경제적 이익을 창출하고 운영 비용을 절감할 수 있었습니다.

디지털 트윈을 통해 얻은 정보는 에너지 생산을 실시간으로 최적화하는 데 중요한 역할을 할 뿐만 아니라 미래 풍력 터빈의 설계 과정에도 반영됩니다. 두산중공업은 출력을 정확하게 예측하고 다양한 조건에서 터빈의 거동을 이해함으로써 후속 모델을 설계할 때 효율성을 더 높일 수 있습니다. 이러한 지속적인 개선 사이클은 에너지 분야의 경쟁력을 유지하는 데 필수적이며, 재생에너지 역량을 강화함으로써 한국의 광범위한 경제 개발 목표의 달성에도 기여합니다.

벤틀리시스템즈는 Microsoft와의 파트너십을 통해 두산중공업에 실질적인 경제적 이점을 제공하는 정교한 디지털 솔루션을 제공했습니다. 이 이니셔티브는 특히 한국을 비롯한 전 세계의 지속가능한 에너지 솔루션 영역에서 AI가 접목된 디지털 트윈이 산업 관행과 인프라 관리 및 경제 발전에 미칠 수 있는 영향이 상당함을 보여줍니다.

롯데호텔^{cxvii} : 자동화를 통해 호스피탈리티의 중요 측면에 집중

Microsoft는 롯데호텔과 함께 파워 플랫폼을 활용한 초자동화 업무 환경을 구축했습니다. 이를 통해 IT 전문 지식이 없어도 업무를 자동화하고, 반복적인 업무는 AI 기반 로봇에게 맡길 수 있게 되었습니다. 17개 롯데호텔의 디지털 전환은 연간 약 1만 시간의 시간 절감 효과를 거두며 호텔 업계에 자동화가 가져다 주는 실질적인 이점을 누리고 있습니다.

파워 플랫폼은 앱 개발 및 비즈니스 자동화 프로세스를 자유롭게 관리할 수 있는 Microsoft의 로코드 도구 모음입니다. 파워 앱스, 파워 오토메이트, 파워 BI, 파워 페이지로 구성된 이 플랫폼을 통해 롯데호텔 직원은 필요한 업무를 효율적으로 자동화할 수 있게 되었습니다. 특히 외부 예약 정보를 호텔 시스템으로 통합하는 예약 시스템 자동화를 통해 시간 절약은 물론이고 수기 입력의 오류도 크게 줄었습니다.

롯데호텔은 파워 플랫폼의 업무 활용을 장려하기 위해 직원의 근무 환경을 혁신적으로 조성하고 있습니다. 로보틱처리자동화(RPA) 협의체를 구성해 자동화 지식과 경험을 공유하고, 경진대회를 운영하며, 비용 절감에 기여한 이니셔티브에 포상하는 등 다양한 활동을 펼치고 있습니다. 이러한 노력은 기술 역량 강화를 통해 지속적인 개선과 생산성 향상 문화가 정착되게 하려는 롯데호텔의 전략의 일환입니다.

롯데호텔은 파워 플랫폼을 고객 서비스, 운영 관리 등 다양한 비즈니스 과제에 확대 적용하는 것을 목표로 전략적 확장 계획을 수립했습니다. 롯데호텔은 직원이 앱과 웹페이지를 개발하고 챗봇, Azure 오픈AI 서비스 등 첨단 AI 기술을 이용할 수 있게 됨에 따라 업무 효율성이 크게 향상될 것으로 기대하고 있습니다. 롯데호텔 김준우 디지털 전략 이사과 한국마이크로소프트 권오성 부장은 파워 플랫폼이 비즈니스 프로세스를 간소화하는 동시에 높은 수준의 고객 만족도와 보안을 유지하는 데 효과적이라는 점을 강조했습니다.

서울시 응급 서비스의 AI 개선 사항^{cxviii}

서울의 한 대학이 공공 안전에 대한 혁신적인 접근방식으로 AI를 활용해 놀라운 정확도로 화재 발생을 예측하는 프로젝트에 앞장섰습니다. 바로 홍익대학교 이재승 교수와 그의 학생들은 머신러닝 알고리즘을 사용하여 서울 소방청의 데이터를 분석했습니다. Azure 머신러닝 스튜디오를 이용해 개발한 모델은 화재 발생 예측 정확도 90%를 달성하여 공공 서비스 향상에 큰 도움이 되었습니다.

이러한 예측 모델링의 획기적인 발전은 화재 대응팀의 효율성에 직접적인 영향을 주었습니다. 화재 발생 가능성이 높은 지역을 파악함으로써 소방서는 순찰 경로를 조정하고 자원을 더 효과적으로 할당할 수 있게 되었습니다. 특히 고위험 지역에 대한 소방대원 배치가 개선되면서 대응 시간이 단축되어 잠재적으로 인명을 구하고 재산 피해를 줄일 수 있었습니다. 데이터에서 얻은 놀라운 통찰력은 강남과 같이 새로 개발되고 변화한 지역이 화재에 더 취약하다는 사실을 밝혀내어 기존의 가정에 도전하고 전략을 수립하는 데 정보를 제공했습니다.

이 프로젝트는 도시 관리와 대중의 신뢰 측면에서도 중요한 역할을 했습니다. 데이터의 민감성을 인식한 이재승 교수는 시민의 개인정보를 보호하기 위해 안전한 Microsoft 가상 머신을 사용할 것을 권장했고, 이는 대학 팀이 소방서와 신뢰를 구축하는 계기가 되었습니다. 이러한 안전한 데이터 처리 덕분에 프로젝트는 성공적으로 진행되었고, 첨단 기술을 활용하여 공공 복지와 자원 관리를 개선하고자 하는 시 공무원에게 긍정적인 평가를 받았습니다.

화재 예측 모델의 성공은 범죄 예방과 교통 관리 등 다른 시민 과제에도 적용될 수 있는 발판을 마련했습니다. 현재 진행 중인 휠체어 접근성 프로젝트는 더 나은 도시를 만들기 위해 AI를 활용하려는 대학의 노력을 보여주는 또 다른 예입니다. 이재승 교수는 학생들이 필요한 기술과 분야별 지식을 갖추는 데 집중하고 있으며, AI를 통해 서울의 공공 서비스를 혁신하고 비상상황에 대응하는 것뿐만 아니라 도시 생활의 수준을 더 넓은 측면에서 향상할 수 있음도 강조합니다.

공공 서비스를 위한 AI



참조

- i MSIT (2019). 인공지능 국가전략. 출처: <https://www.msit.go.kr/bbs/view.do?sCode=eng&nttSeqNo=9&bbsSeqNo=46&mId=10&mPid=9>
- ii MSIT (2022). “대한민국, 뉴욕 이니셔티브 실현을 위한 디지털 대한민국 로드맵 마련”. 출처: <https://www.msit.go.kr/eng/bbs/view.do?sCode=eng&mId=4&mPid=2&bbsSeqNo=42&nttSeqNo=742>
- iii Korea.net (2022). “한국, 디지털 경쟁력 순위 4계단 상승한 8위.” 출처: <https://www.korea.net/NewsFocus/Sci-Tech/view?articleId=221834>
- iv MSIT (2022). “국가전략기술 임무중심 전략로드맵 수립.” 출처: <https://www.msit.go.kr/bbs/view.do?sCode=user&mId=113&mPid=238&pageIndex=15&bbsSeqNo=94&nttSeqNo=3183427&searchOpt=ALL&searchTxt=>
- v Korea IT News (2023). “국내 AI 시장 연 14.9% 성장...2027년 4.4조 전망.” 출처: <https://www.koit.co.kr/news/articleView.html?idxno=112323>
- vi Microsoft, 대한상공회의소 (2023). 생성형 AI의 경제적 영향: 대한민국 일자리의 미래. 출처: https://www.korcham.net/nCham/Service/Economy/appl/KccciReportDetail.asp?SEQ_NO_C010=20120936516&CHAM_CD=B001
- vii MSIT (2022). “대한민국, 뉴욕 이니셔티브 실현을 위한 디지털 대한민국 로드맵 마련.”
- viii TechM (2023). “영국 AI 안전성 정상회의 참석한 이종호 과기부 “적절한 AI 신뢰성·안정성 확보해야.” 출처: <https://www.techm.kr/news/articleView.html?idxno=116260>
- ix Salesforce (2023). 아시아 태평양 AI 준비 지수. 출처: https://www.salesforce.com/content/dam/web/en_sg/www/documents/pdf/salesforce_ai_readiness_index_2023.pdf
- x MSIT (2022). “대한민국, 뉴욕 이니셔티브 실현을 위한 디지털 대한민국 로드맵 마련.”
- xi MSIT (2023). “디지털 권리장전으로 OECD의 디지털 규범 논의를 주도한다.”. 출처: https://www.msit.go.kr/bbs/view.do?sCode=user&mId=113&mPid=238&bbsSeqNo=94&nttSeqNo=3183648&formMode=&pageIndex=2&searchCtgy=&searchCtgy2=&searchCtgy3=&RLS_YN=&searchOpt=ALL&searchTxt=
- xii MSIT (2023). “과기부, ‘인공지능 안전성 정상회의’ 참석 성과 발표.” 출처: <https://www.msit.go.kr/bbs/view.do>
- xiii KDI (2023). 디지털 권리장전. 출처: <https://ieec.kdi.re.kr/policy/materialView.do?num=243174>
- xiv 디지털공론장 (2023). “디지털 권리장전.” 출처: http://beingdigital.kr/front/digital_rights.do
- xv KISDI (2022). 인공지능 윤리기준 실천을 위한 자율점검표. 출처: <https://www.kisdi.re.kr/bbs/view.do?bbsSn=113985&key=m2101113056011&pageIndex=1&sc=&sw=>
- xvi 정책브리핑 (2023). “민간 자율 인공지능 신뢰성 검-인증 설명회 개최.” 출처: <https://www.korea.kr/briefing/pressReleaseView.do?newsId=156597130#pressRelease>
- xvii 대한민국 국회 (2023). AI 책임 및 규제 법안. 출처: http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_W2E3F0D8D0C1A0B9I3J1I2I1G4F0G1
- xviii PIPC (2023). “신뢰기반 인공지능 데이터 규범, 첫 발을 떼다.” 출처: <https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=9083>
- xix OECD.AI Policy Observatory (2023). OECD AI 원칙 개요. 출처: <https://oecd.ai/en/ai-principles>
- xx Ministry of Foreign Affairs of Japan (2023). 첨단 AI 시스템을 개발하는 조직을 위한 히로시마 프로세스 국제 지침 원칙. 출처: <https://www.mofa.go.jp/files/100573471.pdf>
- xxi The White House (2023). “사실 자료: 바이든 대통령, 안전하고 안전하며 신뢰할 수 있는 인공 지능에 관한 행정명령 발표.” 출처: <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/>
- xxii NIST (2023). AI 위험 관리 프레임워크. 출처: <https://www.nist.gov/itl/ai-risk-management-framework>
- xxiii Cloud Security Alliance (2023). NIST의 AI 위험 관리 프레임워크 설명. 출처: <https://cloudsecurityalliance.org/blog/2023/08/30/nist-s-ai-risk-management-framework-explained/>

References 참조

- xxiv GOV.UK (2023). 블레츨리 파크, 2023년 AI 안전성 정상회의 의장 요약. 출처: <https://www.gov.uk/government/publications/ai-safety-summit-2023-chairs-statement-2-november/chairs-summary-of-the-ai-safety-summit-2023-bletchley-park>
- xxv Ibid.
- xxvi Microsoft (2023). “우크라이나를 겨냥한 러시아의 사이버 공격이 계속되고 있습니다.” 출처: <https://blogs.microsoft.com/on-the-issues/2023/06/14/russian-cyberattacks-ukraine-cadet-blizzard/>
- xxvii The White House (2023). “사실 자료: 바이든-해리스 행정부, AI로 인한 위험을 관리하기 위해 선도적인 인공지능 회사로부터 자발적인 안전 서약을 확보하다.” 출처: <https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/>
- xxviii Reuters (2023). “Adobe, 다른 기업들도 AI 위험 관리를 위한 미국의 자발적 계획에 참여.” 출처: <https://www.reuters.com/sustainability/adobe-others-join-white-houses-voluntary-commitments-ai-2023-09-12/>
- xxix Microsoft (2018). GDPR(일반데이터보호규정)에 대비하기. 출처: <https://www.microsoft.com/en-us/industry/microsoft-in-business/business-transformation/2018/03/29/getting-ready-for-general-data-protection-regulation-gdpr/>; Microsoft (2022). 책임있는 AI 시스템 구축을 위한 Microsoft의 프레임워크. 출처: <https://blogs.microsoft.com/on-the-issues/2022/06/21/microsofts-framework-for-building-ai-systems-responsibly/>
- xxx 디지털 공론장 (2023). “디지털 권리장전.”
- xxxi Law Times (2020). “인공지능(AI) 법제정비단’ 민간위원 위촉.” 출처: <https://www.lawtimes.co.kr/news/161045?serial=161045>
- xxxii MSIT (2020). “인공지능 시대를 준비하는 법-제도-규제 정비 로드맵 마련”.
- xxxiii PIPC (2023). “전면 개정 개인정보 보호법, 9월 15일 시행.” 출처: <https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=9145>
- xxxiv 지디넷코리아 (2023). “4기 인공지능 법제정비단 발족---연내 로드맵2.0 발표.” 출처: <https://zdnet.co.kr/view/?no=20230816143413>
- xxxv 대한민국 국회 (2023). 인공지능 육성 및 신뢰 기반 조성에 관한 법률안.
- xxxvi 대한민국 국회 (2023) [2120898] 채용절차의 공정화에 관한 법률 일부개정법률안(박대수의원 등 15인). 출처: [http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_B2A3H0G3F2B3A0X9F1E200N4M1I2H2](http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_B2A3H0G3F2B3A0X9F1E200N4M1I2H2;); 대한민국 국회 (2023). [2121075] 채용절차의 공정화에 관한 법률 일부개정법률안(신영대의원 등 10인). 출처: [http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_Y2X3Y0G3F0D6C1Y7X3F7N5L5K6I6F9](http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_Y2X3Y0G3F0D6C1Y7X3F7N5L5K6I6F9;); 대한민국 국회 (2023). [2120799] 채용절차의 공정화에 관한 법률 일부개정법률안(한준호의원 등 17인). 출처: [http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_M2L3K0R3Q0R9Q1Q4Z2X8W0X8V5K7J4](http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_M2L3K0R3Q0R9Q1Q4Z2X8W0X8V5K7J4;); 대한민국 국회 (2023). [2122180] 콘텐츠산업 진흥법 일부개정법률안(이상헌의원 등 10인). 출처: [http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_R2P3Q0X5Y1X7X1V3W4V4C5D6B5C8A2](http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_R2P3Q0X5Y1X7X1V3W4V4C5D6B5C8A2;); 대한민국 국회 (2023). [2122569] 공직선거법 일부개정법률안(송석준의원 등 12인). 출처: [http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_D2B3C0K5K2J6H1I2G0H4O1P6N2O0M4](http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_D2B3C0K5K2J6H1I2G0H4O1P6N2O0M4;); 대한민국 국회 (2023). [2120001] 정보통신망 이용촉진 및 정보보호 등에 관한 법률 일부개정법률안(김홍걸의원 등 13인). 출처: http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_N2S3F0B1O1Q2K1V1E0B9U4B4V0S3U0
- xxxvii ICT 규제 샌드박스 (2023). “실증 테스트.” 출처: <https://www.sandbox.or.kr/apply/regulation.do>
- xxxviii Ibid.
- xxxix European Parliament (2023). EU AI법. 출처: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html
- xl Ministry of Enterprises and Made in Italy (2023). “이탈리아, 독일, 프랑스, 인공지능에 대한 협력 강화에 합의하다.” 출처: <https://www.mimit.gov.it/en/media-tools/news/italy-germany-and-france-agree-on-strengthening-their-cooperation-on-artificial-intelligence>
- xli The White House (2023). “사실 자료: 바이든 대통령, 안전하고 안전하며 신뢰할 수 있는 인공지능에 관한 행정명령 발표.”
- xlii Skadden LLP (2023). “AI 인사이트: 바이든 행정부, 인공지능에 관한 광범위한 행정 명령 통과.” 출처: <https://www.skadden.com/insights/publications/2023/11/biden-administration-passes-sweeping-executive-order-on-artificial-intelligence>
- xliiii The White House (2023). “사실 자료: 바이든 대통령, 안전하고 안전하며 신뢰할 수 있는 인공지능에 관한 행정명령 발표.”
- xliv OECD.AI Policy Observatory (2023). “AI를 위한 정책 환경 조성(원칙 2.3).” 출처: <https://oecd.ai/en/dashboards/ai-principles/P12>
- xliv European Parliament (2023). EU AI법.
- xlvi 대한민국 국회 (2023). AI 책임 및 규제 법안.
- xlvii Davis Wright Tremaine LLP (2023). “NIST, 신뢰할 수 있는 AI 개발을 위한 최종 위험 관리 프레임워크 발표.” 출처: <https://www.dwt.com/blogs/artificial-intelligence-law-advisor/2023/01/ai-risk-management-framework-nist>
- xlviii Microsoft (2023). AI 거버넌스: 미래를 위한 청사진. 출처: <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW14Gtw>
- xliv Microsoft (2017). “디지털 제네바 협약의 필요성.” 출처: <https://blogs.microsoft.com/on-the-issues/2017/02/14/need-digital-geneva-convention/>
- l Microsoft (2018). “얼굴 인식 기술: 공공 규제와 기업의 책임의 필요성.” 출처: <https://blogs.microsoft.com/on-the-issues/2018/07/13/facial-recognition-technology-the-need-for-public-regulation-and-corporate-responsibility/>
- li Microsoft (2020). “마지막으로, 얼굴 인식 규제에 대한 진전.” 출처: <https://blogs.microsoft.com/on-the-issues/2020/03/31/washington-facial-recognition-legislation/>
- lii Microsoft (2023). “Microsoft, 고객을 위한 새로운 Copilot 저작권 약정 발표.” 출처: <https://blogs.microsoft.com/on-the-issues/2023/09/07/copilot-copyright-commitment-ai-legal-concerns/>
- liii 디지털 공론장 (2023). “디지털 권리장전.”
- liv MSIT (2020). “과거부, 인공지능 윤리기준 마련”. Available at: <https://www.msit.go.kr/bbs/view.do?sCode=user&mPid=112&mId=113&bbsSeqNo=94&nttSeqNo=3179742>
- lv MSIT (2020). “인공지능 시대를 준비하는 법-제도-규제 정비 로드맵 마련”. Available at: <https://www.msit.go.kr/bbs/view.do?sCode=user&mPid=112&mId=113&bbsSeqNo=94&nttSeqNo=3179749>
- lvi MSIT (2023). “대한민국이 세계에 새로운 디지털 질서를 제시합니다!” 홈페이지: <https://www.korea.net/Government/Briefing-Room/Press-Releases/view?articleId=7042&insttCode=A110439&type=O>
- lvii MSIT (2023). “디지털 접근성 커퍼런스 개최.” 출처: <https://www.msit.go.kr/bbs/view.do?sCode=user&nttSeqNo=3183225&bbsSeqNo=94>
- lviii 디지털 정부 (nd). 디지털 정부 마스터플랜 2021-2025. 출처: <https://www.dgovkorea.go.kr/>
- lix AI Times (2023). “NIA, ‘디지털포용 포럼’ 운영위원회 개최... ‘소외 없는 디지털 사회를 위해’.” 출처: <https://www.aitimes.com/news/articleView.html?idxno=155154>
- lx OECD.AI Policy Observatory (2023). OECD AI 원칙 개요.
- lxi NIST (2023). AI 위험 관리 프레임워크. 출처: <https://www.nist.gov/itl/ai-risk-management-framework>
- lxii The White House (2023). “사실 자료: 바이든 대통령, 안전하고 안전하며 신뢰할 수 있는 인공지능에 관한 행정명령 발표.”
- lxiii GOV.UK (2023). 블레츨리 파크, 2023년 AI 안전성 정상회의 의장 요약.
- lxiv Microsoft (2018). “접근성을 위한 AI, 장애인의 역량 강화를 위한 2,500만 달러 규모의 프로그램.” 출처: <https://blogs.microsoft.com/accessibility/ai-for-accessibility/>
- lxv Microsoft (2021). “새로운 접근성 약속.” 출처: <https://news.microsoft.com/accessibility-commitment>
- lxvi Microsoft (2022). “모든 투표가 중요하기 때문에: 선거 접근성 향상.” 출처: <https://blogs.microsoft.com/on-the-issues/2022/09/12/disability-voting-rights-week-index/>

References 참조

lxvii DE News (2020). “MS-SK텔레콤-한국장애인고용공단, ‘접근성을 위한 인공지능’ 세미나”. 출처: <https://www.denews.co.kr/news/articleView.html?idxno=16576>

lxviii newsbytes.ph (2023). “Microsoft 인에이블러 프로그램은 350명의 Pinoy PWD가 취업하는 데 도움이 되었습니다.” 출처: <https://newsbytes.ph/2023/05/31/microsoft-enabler-program-helped-350-pinoy-pwds-get-employed/>

lxix Microsoft (nd). 접근성 진화 모델. 출처: <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4wNu4>

lxx 디지털 공론장 (2023). “디지털 권리장전.”

lxxi KISDI (2022). 인공지능 윤리기준 실천을 위한 자율점검표.; TTA (2022). 신뢰할 수 있는 인공지능 개발안내서.; KISDI (2022). 인공지능 윤리교육 교재.

lxxii KISDI (2023). “2023 인공지능 윤리기준 실천을 위한 자율점검표(안) 공개”. 출처: <https://www.kisdi.re.kr/bbs/view.do?bbsSn=114068&key=m2101113055944>

lxxiii 정책브리핑 (2022). “인공지능, 교육현장에서 안전하게 활용해요!” 출처: <https://www.korea.kr/briefing/pressReleaseView.do?newsId=156520749>

lxxiv KDI (2023). “AI 디지털교과서로 1:1 맞춤 교육시대 연다.” 출처: <https://eiec.kdi.re.kr/policy/materialView.do?num=239513&pg=&pp=20&topic=0>

lxxv NIA (2023). “NIA 소개.” 출처: https://eng.nia.or.kr/site/nia_eng/04/1040100000002016093002.jsp

lxxvi OECD.AI Policy Observatory (2023). OECD AI 원칙 개요.

lxxvii OECD.AI Policy Observatory (2023). “각국이 신뢰할 수 있는 AI를 위한 OECD 원칙을 이행하는 방법.” 출처: <https://oecd.ai/en/wonk/national-policies-2>

lxxviii European Commission (2023). 첨단 AI 시스템을 위한 히로시마 프로세스 국제 지침 원칙. 출처: <https://ec.europa.eu/newsroom/dae/redirection/document/99641>

lxxix The White House (2023). “사실 자료: 바이든 대통령, 안전하고 안전하며 신뢰할 수 있는 인공 지능에 관한 행정명령 발표.”

lxxx EY (2023). AI에 관한 바이든 행정부 행정명령의 주요 내용. 출처: https://www.ey.com/en_us/public-policy/key-takeaways-from-the-biden-administration-executive-order-on-ai

lxxx i GOV.UK (2023). 블레츨리 파크, 2023년 AI 안전성 정상회의 의장 요약.

lxxx ii 한국일보 (2023). “가치관을 바꾼 MS AI 스쿨... ‘배움의 즐거움’을 깨닫다.” 출처: <https://www.hankookilbo.com/News/Read/A2023062010290004218>

lxxx iii 벤처스퀘어 (2023). “차별 없는 디지털 교육으로 함께 도약하세요!” 출처: <https://www.venturesquare.net/890280>

lxxx iv Microsoft, 대한상공회의소 (2023). 생성형 AI의 경제적 영향: 대한민국 일자리의 미래.

lxxx v 디지털 공론장 (2023). “디지털 권리장전.”

lxxx vi AI for Good (2023). “대한민국 과학기술정보통신부(과기부)는 인공지능(AI) 친화적인 환경을 조성하기 위해 ITU와 협력합니다.” 출처: <https://aiforgood.itu.int/the-ministry-of-science-and-ict-partners-with-the-itu-to-foster-a-hospitable-environment-for-artificial-intelligence-ai/>

lxxx vii Korea IT News (2023). “‘디지털 ESG 협의체’ 출범.” 출처: <http://www.koit.co.kr/news/articleView.html?idxno=116451>

lxxx viii MSIT (2023). “대한민국이 세계에 새로운 디지털 질서를 제시합니다!”

lxxx ix 정책 브리핑 (2023). “대한민국 인공지능 도약방안 발표.” 출처: <https://www.korea.kr/briefing/policyBriefingView.do?newsId=156590066&pWise=sub&pWiseSub=C3>

xc OECD.AI Policy Observatory (2023). “포용적 성장, 지속가능한 개발 및 복지(원칙 1.1).” 출처: <https://oecd.ai/en/dashboards/ai-principles/P5>

xc i European Commission (2023). 첨단 AI 시스템을 위한 히로시마 프로세스 국제 지침 원칙.

xc ii The White House (2023). “사실 자료: 바이든 대통령, 안전하고 안전하며 신뢰할 수 있는 인공 지능에 관한 행정명령 발표.”

xc iii GOV.UK (2023). 블레츨리 파크, 2023년 AI 안전성 정상회의 의장 요약.

xc iv Microsoft (2020). Microsoft는 2030년까지 탄소중립을 달성할 것입니다. 출처: <https://blogs.microsoft.com/blog/2020/01/16/microsoft-will-be-carbon-negative-by-2030/>

xc v Microsoft (nd). 지속가능한 미래를 위한 행성 컴퓨터. 출처: <https://planetarycomputer.microsoft.com/>

xc vi Microsoft (2022). 2022 환경 지속 가능성 보고서. 출처: <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW15mgm>

xc vii Microsoft (2021). “이 비영리 단체는 AI를 통해 기후 변화의 영향으로부터 취약한 지역 사회를 보호하고 있습니다.” 출처: <https://news.microsoft.com/source/asia/features/seeds-india-climate-change-ai/>

xc viii Microsoft (2023). AI를 통한 지속 가능성 가속화: 플레이북. 출처: <https://blogs.microsoft.com/on-the-issues/2023/11/16/accelerating-sustainability-ai-playbook/>

xc ix Microsoft (2023). 2023년 영향 요약. 출처: <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW1f1Fv>

c Law Times (2020). “과학기술정보통신부 국가 인공지능 윤리기준 발표.”

ci KISDI (2022). 인공지능 윤리기준 실천을 위한 자율점검표.; TTA (2022). 신뢰할 수 있는 인공지능 개발안내서.; KISDI (2022). 인공지능 윤리교육 교재.

ci i KISDI (2023). “‘2023 인공지능 윤리기준 실천을 위한 자율점검표(안) 공개.’”

ci ii KDI (2023). 디지털 권리장전.

ci v 지디넷코리아 (2023). “4기 인공지능 법제정비단 발족---연내 로드맵2.0 발표”.

ci v KDI (2023). “과기부, ‘인공지능 윤리-신뢰성 확보 추진계획’ 발표”. 출처: <https://eiec.kdi.re.kr/policy/callDownload.do?num=243950&filenum=1&dtim=20231025165822>

ci vi NIA (2023). “NIA 소개.”

ci vii 대한민국 국회 (2023). 인공지능 육성 및 신뢰 기반 조성에 관한 법률안.

ci viii 포춘 코리아 (2023). “약일까 독일일까...인공지능(AI)법의 운명은.” 출처: <https://www.fortunekorea.co.kr/news/articleView.html?idxno=29373>

ci x Ibid.

ci x 대한민국 국회 (2023). AI 책임 및 규제 법안.

ci x i PIPC (2023). “전면 개정 개인정보 보호법, 9월 15일 시행.”

ci x ii 대한민국 국회 (2023). [2120130] 개인정보 보호법 일부개정법률안(김영배의원 등 12인).

ci x iii 대한민국 국회 (2023). [2120898] 채용절차의 공정화에 관한 법률 일부개정법률안(박대수의원 등 15인); 대한민국 국회 (2023). [2121075] 채용절차의 공정화에 관한 법률 일부개정법률안(신영대의원 등 10인); 대한민국 국회 (2023년). [2120799] 채용절차의 공정화에 관한 법률 일부개정법률안(한준호의원 등 17인).

ci x iv 대한민국 국회 (2023). [2122180] 콘텐츠산업 진흥법 일부개정법률안(이상현의원 등 10인).

ci x v 대한민국 국회 (2023). [2122537] 저작권법 일부개정법률안(황보승희의원 등 10인).

ci x vi 대한민국 국회 (2023년). [2122569] 공직선거법 일부개정법률안(송석준의원 등 12인).

ci x vii 대한민국 국회 (2023년) [2120001] 정보통신망 이용촉진 및 정보보호 등에 관한 법률 일부개정법률안(김홍걸의원 등 13인).

ci x viii 대한민국 국회 (2023) [2122730] 정보통신망 이용촉진 및 정보보호 등에 관한 법률 일부개정법률안(김영배의원 등 10인).

ci x ix PIPC (2023). “인공지능 규제방향 국제적 논의 본격 시작.” 출처: <https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=8981>

ci x x PIPC (2023). “신뢰기반 인공지능 데이터 규범, 첫 발을 떼다.”

ci x x i 정책브리핑 (2019). 방통위, 이용자 중심의 지능정보사회를 위한 원칙 발표. 출처: https://www.korea.kr/briefing/pressReleaseView.do?newsId=156360183&call_from=naver_news

References 참조

- cxixii** KCC (2022). 방통위 안형환 부위원장, 제4회 지능정보사회 이용자보호 국제컨퍼런스 참석. 출처: https://www.kcc.go.kr/user.do;jsessionid=fQ5yXtWEmGPoGf5-C5HO08O_E1c89xJ5dH942mVM.servlet-aihgcl-dhome20?mode=view&page=A04070000&dc=&boardId=1035&cp=1&boardSeq=54246
- cxixiii** KDI (2021). “방통위 인공지능 기반 미디어 추천 서비스 이용자 보호 기본원칙 발표”. 출처: [https://eiec.kdi.re.kr/policy/materialView.do?num=215524&topic=; KCC \(2022\). 방통위, 인공지능 기반 미디어 추천 서비스 이용자 보호 기본원칙 해설서 발간. 출처: https://kcc.go.kr/user.do?mode=view&page=A05030000&dc=&boardId=1113&cp=19&boardSeq=53025](https://eiec.kdi.re.kr/policy/materialView.do?num=215524&topic=; KCC (2022). 방통위, 인공지능 기반 미디어 추천 서비스 이용자 보호 기본원칙 해설서 발간. 출처: https://kcc.go.kr/user.do?mode=view&page=A05030000&dc=&boardId=1113&cp=19&boardSeq=53025)
- cxixiv** 정책브리핑 (2022). “방통위 제2기 지능정보사회 이용자 보호 민관협의회 출범”. 출처: <https://www.korea.kr/briefing/pressReleaseView.do?newsId=156503671>
- cxixv** KCC (2023). “방통위, 생성형(Generative) 인공지능(AI) 기반 챗봇 서비스에서의 이용자 보호 쟁점 논의”. 출처: <https://kcc.go.kr/user.do?mode=view&page=A05030000&dc=K00000200&boardId=1113&boardSeq=55026>
- cxixvi** 정책브리핑 (2022). “인공지능, 교육현장에서 안전하게 활용해요!”
- cxixvii** KDI (2023). “AI 디지털교과서로 1:1 맞춤 교육시대 연다.”
- cxixviii** MOE (2023). “인공지능 디지털교과서 법적지위 연다.” 출처: <https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=294&boardSeq=96704&lev=0&searchType=null&statusYN=W&page=1&s=mo e&m=020402&opType=N>
- cxixix** MSIT (2023). “디지털 교육혁신의 성공을 위한 과기부-교육부 정책협력 강화”. 출처: <https://www.msit.go.kr/bbs/view.do?sCode=user&mlId=113&mPid=238&pageIndex=&bbsSeqNo=94&nttSeqNo=3183595&searchOpt=ALL&searchTxt=>
- cxixxx** NCSC (2023). “챗봇 등 생성형 AI 활용 보안 가이드라인(2023.6)”. 출처: https://www.ncsc.go.kr:4018/main/cop/bbs/selectBoardArticle.do?bbsId=InstructionGuide_main&nttId=54340&pageIndex=1
- cxixxxi** OECD.AI Policy Observatory (2023). OECD AI 원칙 개요.
- cxixxxii** Ibid.
- cxixxxiii** OECD (2023). OECD AI 논문: OECD 인공지능 원칙의 4년 후 이행 현황. 출처: <https://www.oecd-ilibrary.org/docserver/835641c9-en.pdf?expires=1699509807&id=id&accname=guest&checksum=FA0CE5E7FBE79FBBC359F5FFD3D7159F>
- cxixxxiv** OECD.AI Policy Observatory (2023). “각국이 신뢰할 수 있는 AI를 위한 OECD 원칙을 이행하는 방법.”
- cxixxxv** Ibid.
- cxixxxvi** Ibid.
- cxixxxvii** EY (2023). “뉴스 알람: G7 AI 원칙 및 행동 강령.” 출처: https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/ai/ey-g7-ai-principles-and-code-of-conduct-final.pdf
- cxixxxviii** European Commission (2023). “위원회는 인공지능에 관한 지도 원칙과 행동 강령에 대한 G7 정상들의 합의를 환영합니다.” 출처: https://ec.europa.eu/commission/presscorner/detail/en/ip_23_5379
- cxixxxix** DLA Piper (2023). G7은 인공지능에 대한 지침 원칙과 행동 강령을 발표합니다. 출처: <https://www.dlapiper.com/en/insights/publications/ai-outlook/2023/g7-publishes-guiding-principles-and-code-of-conduct-for-artificial-intelligence>
- cxli** Ibid.
- cxlii** Ministry of Foreign Affairs of Japan (2023). 첨단 AI 시스템을 개발하는 조직을 위한 히로시마 프로세스 국제 지침 원칙. 출처: <https://www.mofa.go.jp/files/100573471.pdf>
- cxliiii** Ibid.
- cxliiiii** Ibid.
- cxlixv** Ibid.
- cxlixvi** Freshfields Bruckhaus Deringer (2023). “G7은 AI 거버넌스에 대한 글로벌 조율을 시작하기를 희망합니다.” 출처: <https://technologyquotient.freshfields.com/post/102irzk/g7-hope-to-kick-off-global-alignment-on-ai-governance>
- cxlixvii** EY (2023). 바이든 행정부의 AI 행정명령의 주요 내용.
- cxlixviii** The White House (2023). “사실 자료: 바이든 대통령, 안전하고 안전하며 신뢰할 수 있는 인공지능에 관한 행정명령 발표.”
- cxlixix** Ibid.
- cl** GOV.UK (2023). AI 안전성 정상회의 블레츨리 선언, 2023년 11월 1~2일. 출처: <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>
- cli** NIST (2023). 미국 인공지능 안전 연구소. 출처: <https://www.nist.gov/artificial-intelligence/artificial-intelligence-safety-institute>
- clii** NIST (2023). 안전하고 안전하며 신뢰할 수 있는 인공지능에 관한 행정명령. 출처: <https://www.nist.gov/artificial-intelligence/executive-order-safe-secure-and-trustworthy-artificial-intelligence/executive-order-faqs>
- cliii** NIST (2023). AI 위험 관리 프레임워크.
- cliv** Cloud Security Alliance (2023). NIST의 AI 위험 관리 프레임워크 설명. 출처: <https://cloudsecurityalliance.org/blog/2023/08/30/nist-s-ai-risk-management-framework-explained/>
- clv** Wilson Sonsini (2023). “NIST, 인공지능 위험 관리 프레임워크 (AI RMF) 출시.” 출처: <https://www.wsgr.com/en/insights/nist-releases-its-artificial-intelligence-risk-management-framework-ai-rmf.html>
- clvi** NIST (2023). AI 위험 관리 프레임워크.
- clvii** Ibid.
- clviii** Ibid.
- cllix** Ibid.
- clx** Ibid.
- clxi** NIST (2023). 교차검토: NIST AI RMF 신뢰성과 OECD AI 권고안, 제안된 EU AI 법안, 행정명령 13960, AI 권리 장전 청사진 간 관련성. 출처: https://www.nist.gov/system/files/documents/2023/01/26/crosswalk_AI_RMF_1_0_OECD_EO_AIA_BoR.pdf
- clxii** GOV.UK (2023). 블레츨리 파크, 2023년 AI 안전성 정상회의 의장 요약.
- clxiii** Ibid.
- clxiv** Ibid.
- clxv** Ibid.
- clxvi** GOV.UK (2023). AI 안전성 정상회의 블레츨리 선언, 2023년 11월 1~2일.
- clxvii** Ibid.
- clxviii** Ibid.
- clxix** European Parliament (2023). 진행 중인 EU 법률에 대한 브리핑, AI법. 출처: [https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2021\)698792](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698792)
- clxx** Ibid.
- clxxi** EY (2023). EU AI 법안: 비즈니스에 미치는 영향. 출처: https://www.ey.com/en_ch/forensic-integrity-services/the-eu-ai-act-what-it-means-for-your-business
- clxxii** European Parliament (2023). 진행 중인 EU 법률에 대한 브리핑, AI법.
- clxxiii** Center for Security and Emerging Technology (2023). AI Act: 입문서. 출처: <https://cset.georgetown.edu/article/the-eu-ai-act-a-prime>
- clxxiv** European Parliament (2023). 진행 중인 EU 법률에 대한 브리핑, AI법.
- clxxv** Ibid.
- clxxvi** EY (2023). EU AI 법안: 비즈니스에 미치는 영향
- clxxvii** Center for Security and Emerging Technology (2023). AI Act: 입문서.

References 참조

- clxxviii European Parliament (2023). 진행 중인 EU 법률에 대한 브리핑, AI법.
- clxxix Ibid.
- clxxx Center for Security and Emerging Technology (2023). AI Act: 입문서.
- clxxxI European Parliament (2023) 진행 중인 EU 법안에 대한 브리핑, AI법.
- clxxxii Ibid.
- clxxxiii European Parliament (2023). EU AI법: 신뢰할 수 있는 인공지능을 위한 포괄적인 규칙. 출처: <https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>
- clxxxiv European Parliament (2023). EU AI 법: 인공지능에 관한 최초의 규정. 출처: <https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>
- clxxxv The White House (2023). “사실 자료: 바이든-해리스 행정부, AI로 인한 위험을 관리하기 위해 선도적인 인공지능 기업들로부터 자발적인 약속을 확보하다.”
- clxxxvi Reuters (2023). “Adobe, 다른 기업들도 AI 위험 관리를 위한 미국의 자발적 계획에 참여.”
- clxxxvii GOV.UK (2023). “안전 테스트: 세션 결과에 대한 의장 성명서, 2023년 11월 2일.” 출처: <https://www.gov.uk/government/publications/ai-safety-summit-2023-chairs-statement-safety-testing-2-november/safety-testing-chairs-statement-of-session-outcomes-2-november-2023>
- clxxxviii Ibid.
- clxxxix Microsoft (2023). Azure OpenAI 서비스에 대한 투명성 참고 사항. 출처: <https://learn.microsoft.com/en-us/legal/cognitive-services/openai/transparency-note?tabs=text>
- cxlc Microsoft (2023). 책임있는 AI 관행 강화. 출처: <https://www.microsoft.com/en-us/ai/responsible-ai>
- cxci Microsoft (2023). Copilot 저작권 약정. 출처: <https://blogs.microsoft.com/on-the-issues/2023/09/07/copilot-copyright-commitment-ai-legal-concerns/>
- cxcii Microsoft (2023). 제품 약관. 출처: <https://www.microsoft.com/licensing/terms/product/ForOnlineServices>
- cxciiii 한국일보 (2023). “가치관을 바꾼 MS AI 스킴... ‘배움의 즐거움’을 깨달았습니다.”
- cxciiv Microsoft (2019). “포용적인 미래를 위한 세계 언어 보호.” 출처: <https://news.microsoft.com/apac/2019/01/03/protecting-the-worlds-languages-for-an-inclusive-future/>
- cxci v Microsoft (2021). “벤틀리 시스템즈 사례 연구.” 출처: <https://partner.microsoft.com/en-rs/case-studies/bentley-systems>
- cxci vi Microsoft (2023). “한국Microsoft, 파워 플랫폼으로 롯데호텔의 초자동화 업무 환경 구축 지원.” 출처: https://news.microsoft.com/ko-kr/2023/07/10/lottehotel_powerplatform/
- cxci vii Microsoft (2019). “AI로 화재 진압.” 출처: <https://news.microsoft.com/apac/features/fighting-fire-with-ai/>