The Next Move ‘Responsible AI’

최근 미국의 행정명령 발표에 따른 ‘Responsible AI’를 위한 대응

December 2023

pwc

Contents

들어가며

ChatGPT의 아버지라 불리는 OpenAI의 CEO 샘 올트먼이 최근 이사회로부터 해임된 뒤 5일 만에 복귀한 일련의 사태를 전 세계가 주목하고 있다. 새로 구성 중인 이사회가 AI의 안전성보다는 상업성을 지지하는 입장일 것이라는 전망도 있다. 뉴욕타임즈는 '이제 AI 유토피아를 꿈꾸는 이들이 운전석에 앉아 전속력으로 전진할 것'이라며 우려의 목소리를 내기도 했다. 하지만 OpenAI의 최대 주주인 마이크로소프트는 이사회에 의결권 없는 참관인으로 참여하기로 하여 업계의 예상은 빗나갔다. 결국 OpenAI는 이익 추구와 안전성 사이에서 절충점을 찾으려는 것으로 보인다.

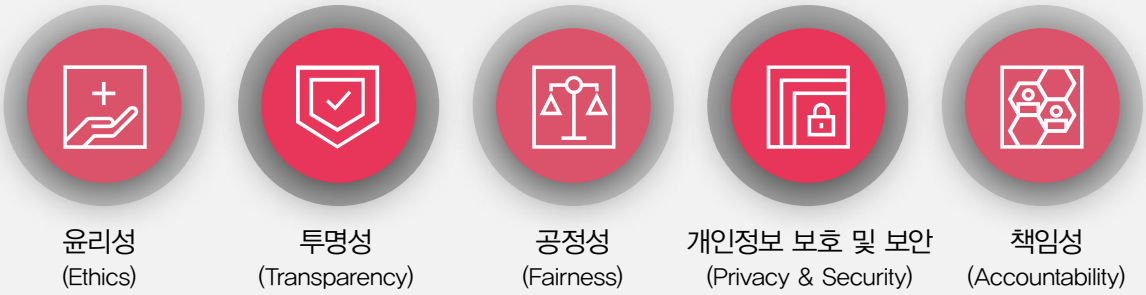
각국은 AI의 위험을 통제하기 위한 규제를 이미 발표하고 있다. 유럽 의회가 2021년 세계 최초로 AI 법 초안을 발의하였고 이달 첫 합의안에 이르렀다. 미국, 중국 등 주요국도 법제화 과정에서 가시적인 성과를 보이고 있는 상황이다. 특히, 미국 바이든 정부는 지난 10월 30일 행정명령을 발표하며 AI 서비스를 준비하는 과정부터 정부가 깊이 개입하겠다고 하여 가장 강력한(Most ambitious) 조치라는 평가를 받고 있다. 이에 따라 본 보고서에서는 행정명령의 세부 내용과 이에 대응하기 위한 PwC의 방법론을 살펴보고자 한다.

※ AI 행정명령에 서명하는 바이든 대통령 (뉴욕타임즈)

1. Responsible AI의 개념 및 배경

AI 기술이 발전하고 ChatGPT와 같은 생성형 AI가 우리 삶에 활용되면서 AI가 초래할 수 있는 잠재적 위험에 대한 두려움이 커지고 있다. Responsible AI(책임 있는 AI)는 AI 시스템 사용의 윤리적, 사회적, 법적 영향을 다루기 위해 도입된 개념이다. 즉, **AI 시스템을 개발, 운영, 배포하는 과정이 윤리적이고 투명하며 사용자 개인 정보와 사회적 가치를 존중하는 방식으로 이뤄지는 것**을 의미한다. Responsible AI는 직원과 비즈니스에 권한을 할당하고 고객과 사회에 공정한 영향을 미치려는 선한 의도로 설계, 개발 및 배포하는 과정에서 기업이 신뢰를 쌓아 AI를 확장할 수 있도록 하는 것이다. Responsible AI는 인간의 가치와 복지에 부합하는 방식으로 사용되도록 다음과 같은 원칙을 포함한다.

Responsible AI 주요 구성 요소



- **윤리성(Ethics):** 윤리적 의사결정, 공정성, 투명성, 책임성, 인권 같은 원칙을 준수한다. 또한 잠재적으로 미칠 사회·경제·환경적 영향을 고려하고 운영 과정이 윤리적 지침에 부합하는지 확인하는 과정을 거친다.
- **투명성(Transparency):** 모형의 알고리즘이 내린 결정을 설명하고(설명 가능성), 입력 데이터, 학습 방법 및 의사결정 프로세스에 대한 정보를 사용자와 이해 관계자와 공유한다.
- **공정성(Fairness):** 인종, 성별, 연령 등의 요인에 따라 차별적인 결과를 초래할 수 있는 데이터 수집과 알고리즘 모델의 편견을 없앤다.
- **개인정보 보호 및 보안(Privacy & Security):** AI를 개발하는데 데이터에 대한 의존도가 높아짐에 따라 데이터가 유출되거나 공개되지 않도록 한다.
- **책임성(Accountability):** AI 역할과 책임 정의, 정기적인 감사 실시, 윤리 지침 및 규제 요건 준수 등을 통해 AI 시스템의 결정과 행동에 대한 책임을 강조한다.

최근 미국 아마존(Amazon)의 채용추천 AI에서 불거진 성차별 논란, 유나이티드 헬스그룹(UnitedHealth)의 치료 우선순위 결정 알고리즘에서 인종차별 문제가 표면화되면서 논의의 범위가 인간의 판단 권리 보장에서 AI 자체의 윤리 기준 마련으로 확장됨에 따라, Responsible AI가 사회적인 이슈가 되고 있다. 따라서 개인과 사회를 위해 AI 시스템을 개발하고 운영하는 모든 국가, 사회, 기업에게 **원칙, 정책, 도구, 프로세스가 포함된 프레임워크의 중요성**이 강조되고 있으며, 이는 '**인간 중심의 접근 방식으로 AI의 연구, 개발 및 배포를 수행함으로써 AI를 안전하고 신뢰할 수 있으며, 윤리적인 방식으로 발전시킨다**'는 개념이기도 하다.

2. 미국 바이든 행정부의 행정명령

2023년 10월 30일, 미국 바이든 행정부에서 AI와 관련된 행정명령(EO: Executive Order)을 발표하였다. 여기에는 모든 연방기관들의 AI 사용을 위해 필요한 정부지침 및 규제, 자금조달, 위험 대처를 위한 훈련 방안이 포함되어 있으며, 이를 발판으로 국민들이 안전하게 활용하도록 AI 기술의 포괄적인 도입 방안을 제시하였다.

이번 행정명령은 AI의 안전, 보안, 신뢰에 대해 전 세계 정부가 취한 조치 중 가장 강력한 조치로 평가 받고 있으며, 인류에 위협이 될 수 있는 최첨단 모델의 규제에 초점을 맞추고 있다. 요약하면 다음과 같다.

1. 미국의 안보, 경제, 공중 보건 또는 안전에 위험을 초래하는 **AI 시스템 개발자는 국방물자생산법(Defense Production Act)에 따라 안전 테스트 결과를 대중에게 공개하기 전에 미국 정부와 공유**해야 한다.
2. 국립표준기술연구소(NIST, National Institute of Standards and Technology)는 **AI 공개 전 안전 보장을 위한 레드팀(Red Team)*의 테스트에 대한 엄격한 표준을 설정하여, 화학, 생물학, 방사선, 핵 및 사이버 보안 리스크를 해결**해야 한다.

   * 시스템의 안전성을 향상시키기 위해 취약점을 발견하고 해결하는 팀. 보통 모의 공격을 주도하는 역할을 수행

3. 상무부는 **AI가 생성한 파일과 콘텐츠를 가려내기 위한 인증 워터마크** 지침을 개발한다.

이번 행정명령은 '2022 AI 권리장전을 위한 청사진(2022 Blueprint for an AI Bill of Rights)'에 기반하여 인력, 핵심 기반시설, 국가안보 및 민주주의 등에 대한 리스크를 해결하고, AI를 통해 새로운 국가적 기회를 언급한 바이든 행정부의 의지이기도 하다. 이번 행정명령을 통해 AI 모델을 개발하거나 연방정부와 함께 일하는 기업들이 가장 직접적으로 영향을 받을 것으로 예상되며, 마이크로소프트, 구글, 오픈 AI, 엔비디아 등 15개의 대표적인 기업들도 백악관 'AI 안전 서약'에 참여하여 미국 정부의 입장에 동참하였다.

● 바이든 정부의 Responsible AI를 위한 8가지 가이드

바이든 행정부는 AI에 '가능성과 위험성 모두에 대한 혁신적인 잠재력'이 있음을 사전에 인지하고, AI 기술의 리스크 완화, 이로운 활용을 위한 정부, 산업, 학계, 시민사회 간의 파트너십의 필요성을 강조하였다. NIST AI 리스크 관리 프레임워크(NIST AI Risk Management Framework)의 영향을 받은 이번 행정명령에서는 제시된 8가지 가이드에 따라 AI를 개발하고 사용할 것을 강조하였다.

바이든 정부의 Responsible AI를 위한 8가지 가이드

① 안전과 보안을 위한 기준

- 강력한 AI 시스템들의 개발자들을 대상으로, 그들의 안정성 테스트 결과와 기타 중요한 정보를 미국 정부와 공유한다. 국방물자생산법에 따라, 국가안보, 국가 경제안보 또는 국민건강과 안전에 심각한 리스크를 발생시킬 수 있는 모델을 개발하는 기업은 모델 훈련 시 연방정부에 공지하고, 모든 레드팀의 안전성 테스트 결과를 공유한다.

- AI 시스템이 안전하고, 안심할 수 있으며, 신뢰할 수 있음을 보장하도록 지원하는 표준, 도구 그리고 테스트가 개발되어야 한다. 국립표준기술연구소는 공개배포 전 안전성을 보장하기 위한 광범위한 레드팀 테스트 기준을 엄격하게 설정한다. 국토안보부(Department of Homeland Security)는 그러한 기준들을 중요한 기반시설 영역에 적용할 것이며, AI 안전보안 위원회를 설립한다. 에너지부(Department of Energy)와 국토안보부는 화학, 생물, 방사능, 핵과 사이버 보안상의 리스크와 함께 중요한 기반시설에 대한 AI의 위협을 다룬다.

- 생명과학 프로젝트에 자금을 지원하는 기관은 생물학적 합성에 대한 강력한 새로운 표준을 마련하여 AI를 활용하여 위험한 생물학적 물질 개발과 관련된 위험으로부터 보호한다.

- AI가 생성한 콘텐츠를 감지하고 공식 콘텐츠임을 인증하기 위한 기준을 수립하여 미국인들을 사기와 속임수로부터 보호한다. 상무부는 AI가 생성한 콘텐츠를 명확히 식별하기 위해 콘텐츠 인증과 워터마킹에 대한 가이드라인을 수립할 것이며, 연방 기관들은 미국인들이 정부의 커뮤니케이션을 신뢰하고, 민간 영역과 타국 정부에 모범을 보이기 위해 이러한 도구들을 사용한다.

- 바이든–해리스 행정부에서 진행중인 AI 사이버 챌린지*에 기반하여, 주요 소프트웨어상의 취약점을 찾아내고 수정하기 위한 AI 도구들을 개발하도록 진보된 사이버보안 프로그램을 구축한다. 이러한 노력은 AI의 사이버 역량을 활용하여 소프트웨어와 네트워크를 더욱 안전하게 만들 것이다.

  * AI로 각종 소프트웨어의 취약점을 찾고 개선하는 경진대회. 미국 정부가 주관하여 총 상금은 2,000만 달러

- 국가안전보장회의 (NSC)와 백악관 비서실장은 AI와 보안에 대해 추가적인 행위를 지시하는 국가 안보 각서(National Security Memorandum)를 작성한다. 이 문서는 미군과 정보기관이 각자의 임무에서 AI를 안전하고, 윤리적이고, 효과적으로 사용하는 것을 보장하며, 적국이 군사적으로 AI를 활용하는 것에 대항하기 위한 대책을 지시한다.

② 미국인들의 프라이버시 보호

- AI가 학습 데이터의 개인정보를 보호하면서 학습할 수 있도록 하는 등 개인정보 보호 기술을 가속화하기 위한 연방 정부의 지원으로 미국인의 개인정보를 보호한다.
- 신속한 기술혁신을 이루도록 협력 연구 네트워크에 자금을 지원함으로써 프라이버시를 보호하는 암호화 도구 등에 대한 연구와 기술을 강화한다. 국립과학재단 (National Science Foundation)은 이러한 연구 네트워크와 협업하여 연방기관들의 선도적인 프라이버시 보호 기술 도입한다.
- 데이터 브로커를 통해 얻는 정보를 포함하여, 상업적으로 활용가능한 정보를 수집하고 사용하는 방법을 평가하고 AI 리스크에 대비하여 연방기관들의 프라이버시 가이드를 강화한다. 이 조치는 개인식별가능 정보를 포함한 상업적으로 이용가능한 정보에 초점을 맞출 것이다.
- AI에 사용되는 프라이버시 보호 기법을 비롯하여, 연방기관이 프라이버시 보존 기술의 효과성을 평가하기 위한 가이드라인을 개발한다. 이러한 가이드라인은 미국인들의 데이터를 보호하기 위한 노력을 배가할 것이다.

③ 형평성과 시민권 증진

- AI 알고리즘이 차별을 악화시키는 것을 방지하기 위해 연방 지원 프로그램 그리고 연방 계약 담당자 등에게 명확한 가이드를 제공한다.
- 알고리즘에 의해 발생하는 차별은 훈련, 기술 지원, 그리고 법무부와 연방 시민권 사무소 간 조율을 통해 대응한다.
- 선고, 가석방과 보호관찰, 보석과 구금, 리스크 평가, 감시, 범죄 예측과 예측적 경찰력 운용, 포렌식 분석에 대해 AI 선진 사례를 개발하여 형사 사법 시스템의 공정성을 보장한다.

④ 소비자, 환자, 학생 보호

- 국민의 생존과 직결된 의약품 개발분야에서 AI가 책임 있게 사용되도록 해야 한다. 보건복지부(Department of Health and Human Services)는 또한 AI를 포함한 안전하지 않은 헬스케어 행위들을 방지 및 대응하기 위한 프로그램을 개발한다.
- 개인별 맞춤형 교육과 같이 AI 기반의 교육 도구들을 학교에서 활용할 수 있도록 자원을 마련한다.

⑤ 근로자 보호

- 근로자에 대한 AI의 피해를 완화하고 효익을 극대화하기 위한 원칙과 선진 사례를 만든다. 이러한 원칙과 선진 사례는 고용주가 근로자들이 받아야 할 급여보다 적게 지급하거나, 지원서를 불공평하게 평가하거나, 근로자의 단체 조직을 방해하는 행위를 방지하여 근로자에게 도움이 될 것이다.
- AI가 노동시장에 미치는 잠재적 영향에 대한 보고서를 작성하고, AI로 인해 노동 시장의 불안정성에 대한 연방 지원 강화 방안을 연구한다.

⑥ 혁신과 경쟁의 촉진

- AI 연구자와 학생들이 핵심 AI 자원과 데이터에 접근성을 높이는 국가 AI 연구 자원(National AI Research Resource)의 파일럿 프로그램과 헬스케어나 기후변화와 같은 중요한 영역에 보조금을 확대하여 AI 연구를 촉진한다.
- 소규모 개발자와 사업가에게 기술 지원과 자원을 제공하고 소규모 기업이 AI를 통해 사업화 할 수 있도록 지원하며 연방거래위원회 (Federal Trade Commission)의 권한 행사를 장려함으로써 공정하고 개방적이며 경쟁력 있는 AI 생태계를 증진한다.
- 소비자 기준, 인터뷰 및 심사를 현대화하고 간소화하여 고도로 숙련된 이민자와 중요 분야에 전문성을 갖춘 비이민자가 미국에서 학업을 하고, 체류 및 취업할 수 있는 기회를 확대한다.

⑦ 미국 리더십의 해외 확장

- AI에 대한 협력을 위해 양자, 다자, 그리고 다수의 이해관계자들의 참여를 확장한다. 국무부는 상무부와 함께 AI의 효익을 취하고, 안전을 보장하며, 리스크를 관리하기 위한 탄탄한 국제적인 프레임워크를 수립하기 위해 노력한다.
- 해외의 파트너 기관 및 표준 제정 기관과 함께 중요한 AI 표준의 개발과 구현을 가속화하여 기술의 안전성, 보안성, 신뢰성, 상호 운용성을 보장한다.
- 안전하고 책임감 있으며 권리를 보장하는 AI의 개발과 배포를 촉진하여 지속 가능한 개발을 추진하고 중요한 인프라에 대한 위험을 완화하는 등 글로벌 과제를 해결한다.

⑧ 정부의 책임 있고 효과적인 AI 사용 보장

- 권리와 안전을 보호하고 AI 배포를 강화하기 위한 명확한 표준을 포함하여 기관의 AI 사용에 대한 지침을 제시한다.
- 보다 신속하고 효율적인 계약을 통하여 기관들이 특정 AI 제품과 서비스를 더 빠르고, 더 저렴하고, 더 효과적으로 확보할 수 있도록 지원한다.
- 연방 인사관리처, 미국 디지털 서비스*1, 미국 디지털 부대*2, 대통령 혁신 펠로우십*3이 주도하는 정부 차원의 AI 인재 수요 급증에 대응하여 AI 전문가의 신속한 채용을 가속화한다. 각 기관은 관련 분야의 모든 직급을 대상으로 AI 교육을 제공한다.

*1 US Digital Service: 미국연방정부에 IT 서비스를 제공하는 조직
*2 US Digital Corps: 바이든 행정부가 디지털 인력 집중 양성하기 위해 설립한 교육 프로그램
*3 Presidential Innovation Fellowship: 미국 정부가 디지털 전문가들을 유치, 특정 프로젝트와 기관에 투입하여 혁신적인 솔루션을 개발하고 정부의 서비스를 개선하는 데 참여하는 프로그램

3. Responsible AI를 위한 국내의 준비

바이든 행정부는 AI의 개발과 활용을 통제하기 위한 강력한 국제 프레임워크에 대해 동맹국, 파트너와 협력할 것을 강조하고 있다. 또한 G7국가들과 함께 AI 국제 행동 강령(Code of Conduct)에 합의하여 각국 정부가 준비 중인 규제안의 기준을 마련했다. 사실 Responsible AI와 관련된 최초의 국제적 법안은 2021년 4월 유럽연합 의회에서 채택한 EU AI 법안(AI Act)이다. 이는 AI 기술에 관한 포괄적인 규제를 담은 최초의 법안으로서, 2023년 6월 14일 유럽의회 본회의 표결에서 가결되었다. 2026년부터 발효되어 한국을 비롯한 여러 나라의 AI 규제에 영향을 미칠 것으로 예상된다. EU 법안 역시 국제적으로 금지되어 있는 AI와 고위험 AI 시스템을 사전에 정의 및 분류하고, 사람과 교류하는 AI에 대한 정의와 이에 따른 포괄적인 규제 법안을 담고 있다.

이 법안은 AI를 '허용할 수 없는 위험', '고위험', '제한된 위험', '저위험 또는 최소 위험' 등 4단계의 위험으로 분류하고, 이 중 '저위험 또는 최소 위험' 외 나머지 세 단계에 대해 규제를 부과한다. 특히 제5조는 사람의 안전, 생명, 권리에 명백한 위협이 되는 유해한 AI를 '허용할 수 없는 위험'으로 분류해 유럽연합 내에서 사용하는 것을 금지한다. 유럽연합은 '허용할 수 없는 위험'을 가진 AI의 예시로 특정 취약군에 대한 인지 행동 조작(예를 들어, 아이들에게 위험한 행동을 유도하는 음성 인식 장난감), 개인의 사회·경제적 지위 등을 기반으로 한 사회적 점수 평가, 안면 인식과 같은 실시간 원격 생체 인식 시스템을 제시하였다. 이 법안에 따르면 유럽연합 회원국은 국가 감독 기구를 포함한 하나 이상의 관할 당국을 지정해 이 법의 적용을 감독하고, 회원국 대표와 유럽연합 집행위원회로 구성된 유럽 AI 이사회를 설립하여야 한다. 이를 위반할 경우에는 최대 3,500만 유로(약 490억 원), 글로벌 매출의 7%까지 과징금이 부과된다. 이에 더해 유럽연합 회원국은 국내에 관련 벌칙 규정을 도입하여 이 법이 적절하고 효과적으로 시행될 수 있도록 보장하여야 한다.

미국, 유럽의 법안과 더불어 한국도 AI 윤리 기준을 수립하고(2020년 12월), 자율적으로 준수 및 점검할 수 있는 가이드라인을 발간하였다(2022년 2월). 하지만, 국제적 기준에 대한 대응 방안과 구체적 가이드라인은 미흡한 실정이며, 현재 추진 중에 있다.

Responsible AI의 현실화는 피할 수 없는 흐름이다. 향후 마련될 AI 신뢰체계 및 인증체계를 사전에 대비하는 것이 국내 기업에게 있어 무엇보다 중요하다. 그러나 민간 기업이 당장 눈에 보이지 않는 AI학습 결과상의 잠재적 편향을 사전 검토하고 대비하기란 사실상 어렵다. 이는 비단 내수시장에 국한된 문제가 아니며 차후 글로벌한 AI 신뢰성 표준이 정립될 시 국내 기업의 수출 경쟁력에 큰 영향을 미칠 수 있다. 따라서, 신뢰확보 체계의 정밀한 설계와 민간 기업의 노력을 효과적으로 지원할 수 있는 자원의 활용, 그리고 공공 중심의 원천기술 개발 및 표준화된 기준 배포까지 이 모든 부문이 유기적으로 작동해야만 AI신뢰 거버넌스의 성공적 안착을 도모할 수 있다는 점에서, 향후 구체적 수행방안에 대한 면밀한 검토가 무엇보다 중요하다. PwC는 최근 수년 간 AI Advisory 영역에서 축적한 풍부한 경험을 바탕으로 기업이 Responsible AI를 구축하기 위한 방법론을 제안한다.

4. Responsible AI를 위한 PwC 방법론

기업 내부에서 Responsible AI에 대한 공감대와 가이드라인을 지속적이고 일관성 있게 가져가려면 AI 모델 개발 및 전 라이프사이클에 걸쳐 심도 있는 고민과 전사적인 거버넌스를 갖춰야 한다. 기업은 본 행정명령의 직접적인 영향과 이차적인 영향을 검토해야 하며, 그에 따른 갭과 기회를 식별하여 대응 계획을 세워야 한다. PwC는 Responsible AI를 적용하기 위해 중시해야 할 7가지 요소를 다음과 같이 제안한다.

① 행정명령이 조직에 미치는 영향을 확인 및 검토하라.

▫ 연방 차원의 여러 조치와 비즈니스 부문 간의 모든 잠재적 교차점을 파악하여 발생가능한 이슈를 인지하고 대응 계획을 마련한다.

② 규제가 제정되는 과정을 모니터링하고 의견을 제시하라.

▫ 향후 몇 달간 유관 부서들은 행정명령의 준수 정도를 측정하기 위해 보다 구체적인 지침을 발표할 예정이다. 기업의 규제 준수 팀은 이러한 지침이 발표될 때를 대비해야 한다. 추가적으로, 의견 수렴 기간에 참여하여 의견을 제시하는 것도 고려해야 한다.

③ AI 기술들의 리스크를 평가하고 관리하라.

- AI 리스크를 측정하고 관리하고 필요 시 보고할 수 있는 가이드를 반영하는 리스크 분류체계를 정의한다.
- 행정명령에서 설명된 의무와 Responsible AI 원칙을 보다 넓게 고려하는 기업 거버넌스 모델을 개발한다. 거버넌스 모델 개발에서 중요한 단계는 기존 팀과 새로운 팀의 R&R을 정의하는 것이다.
- AI의 리스크에 따라 AI를 평가하고 테스트할 수 있는 프로세스를 통합하라. 여기에는 행정명령에서 언급된 바와 같이 레드팀과 관련된 프로세스도 포함될 수 있다.
- AI를 책임 있게 사용하고 AI 도입에 늘어남에 따라 달라질 업무 상의 변화에 대비할 수 있도록 직원들의 역량을 강화한다.
- 개인 정보 보호 강화 기술(PETs) 프레임워크를 포함하는 포괄적인 프라이버시 프로그램의 도입을 고려한다.

④ Trust-by-design* 접근법을 AI에 적용하라.

▫ 행정명령은 표준을 개발할 것을 요구하지만, AI와 그 결과물의 신뢰를 입증하기 위한 요건은 정의하지 않으며 이는 개별 기관의 몫일 가능성이 높다. 이 요건은 경영진이 먼저 자체 규정을 준수하고 효과적인 절차임을 입증해야 한다. 많은 이해관계자가 AI 시스템과 관련 데이터 및 테스트에 대해 문서화를 요구할 가능성이 높다.

* 상품이나 서비스 설계 단계부터 신뢰 향상을 지향하는 업무 방식

⑤ AI 모델의 세분화를 고려하라.

- 국가안보, 국가 경제안보, 공중 보건을 위해 사용되는 AI 모델과 보다 광범위하고 일반적인 용도의 모델을 구분하는 것이 바람직하다. 예를 들어, 경우에 따라서는 전사적으로 하나의 LLM(Large Language Model, 대규모 언어모델) 보다는 다중 모델 방식이 더 적절할 수 있다.

⑥ AI를 방어에 활용할 방법을 모색하라.

- 소프트웨어의 취약점을 찾아 수정하거나 사이버 사고를 자동으로 감지하는 등 위험을 완화하는 데 AI를 활용하는 사례를 참고한다.

⑦ 투명성을 강화하라.

- 기업이 위험 평가, 레드팀의 시뮬레이션 결과 공유와 같은 의무를 이행하기 위해서는 프로세스를 문서화하고 외부 보고를 위한 준비 상태를 평가한다. 규제 당국, 고객, 언론의 감시가 강화되고 있는 상황에서 공개적인 발표와 내부 관행이 일치하는지 확인한다.

이와 관련하여 PwC는 책임있는 AI 적용을 위하여 기업이 갖추어야 할 프레임워크를 다음과 같이 제시한다.

PwC의 Responsible AI 적용을 위한 프레임워크

전략 — CEO·이사회

데이터 & AI 윤리
데이터 및 AI 활용에 따른 윤리적 영향을 인지하고, 이를 기업 가치에 녹여낼 것

정책 & 규제
주요 공공 정책과 규제 트렌드를 예측하고 이해하여 이를 성실히 준수·이행할 것

관리·통제 — 위험 및 준법 관리 최고 책임자

거버넌스
3차 방어선까지 관리할 수 있는 시스템을 구축할 것

규정 준수
규정, 조직 정책, 산업 표준을 준수할 것

위기 관리
리스크 감지 및 경감책을 확장하여 AI에 생소한 위험마저 관리할 수 있도록 할 것

책임있는 관리 제도 실행 — 정보 보안 최고 책임자

해석가능성 & 설명가능성
투명한 의사결정 모델을 구축할 것

지속 가능성
환경에 미치는 악역향을 최소화하고 사람의 권리와 권한을 강화할 것

강건성
고성능의 신뢰할 수 있는 시스템을 구축할 것

편향성 & 공정성
공정성을 정의, 측정하고 관련 표준과 비교해볼 것

보안
시스템의 사이버 보안 기능을 향상시킬 것

개인정보보호
데이터 보안을 위한 시스템을 개발할 것

안전
물리적 피해를 방지하기 위한 시스템을 설계하고 테스트할 것

핵심 관리 제도 실행 — 데이터 전문가·비즈니스 도메인 전문가

문제 구체화
해결해야 하는 문제에 대해 정확히 이해하여 AI/ML 솔루션이 요구되는 사항인지 파악할 것

표준 규격화
산업 표준 및 모범 사례들을 본보기 삼을 것

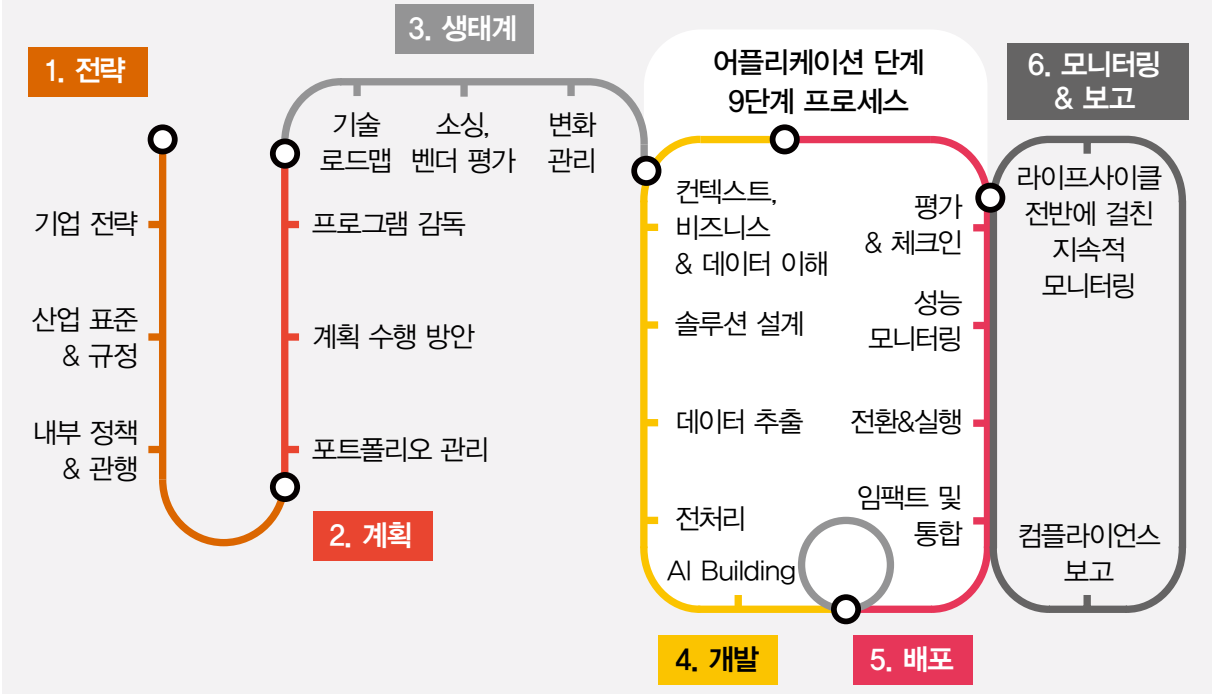
입증
모델 성과를 평가하고 모델 설계 및 개발 과정을 반복 수행하여 성과치를 높일 수 있도록 할 것

모니터링
잠재적 리스크를 발견하기 위해 지속적인 모니터링 시스템을 구축할 것

Responsible AI를 구현하기 위해 기업의 추진 전략, 관리 및 통제, 책임 있는 관리 제도의 실행, 실행조직이 정의해야 할 핵심 관리 제도의 실행으로 구분한다. 네 가지 영역별 CEO와 이사회, 위험 및 준법관리 책임자, 정보보호·보안 책임자, 데이터·비즈니스 도메인 전문가별 활동 기준과 해당 규제 시행여부에 대해 다음과 같은 체크리스트를 통해 모니터링한다.

- **비즈니스 적합성 검증: 해당 비즈니스에 AI 솔루션을 적용하는 것이 적합한가?**
- **성능 적합도 검증: AI 모델은 기대한 수준의 성능을 보여주는가?**
- **개발과 유지 보수 환경 하에서의 배포: AI 모델을 개발 및 유지보수 시 포함시켜 배포할 것인가?**
- **AI 모델의 운영 방식 결정: AI 모델을 지속적으로 운영할 준비가 되어 있는가?**
- **모델 개선 및 중단 여부 결정: AI 모델은 현재 상태로 지속해야 하는가? 재학습 또는 재설계하거나 중단해야 하는가?**

PwC의 Responsible AI 도입을 위한 End-to-end 프로세스



PwC의 Responsible AI 거버넌스 프로세스는 윤리 원칙을 적용하는 것 이상으로, 조직 내에서 Responsible AI를 도입하고, 운영하는 과정에서 발생하는 리스크와 제어 방법에 초점을 맞춘 End-to-end 프로세스를 제안하고 있다. 이 거버넌스 프로세스는 다음과 같이 6단계 프로세스로 이루어진다. 규제 시행여부에 대해 다음과 같은 체크리스트를 통해 모니터링한다.

1. 전략(Strategy)
2. 계획(Planning)
3. 생태계(Ecosystem)
4. 개발(Development)
5. 배포(Deployment)
6. 모니터링 및 보고(Monitor & Report)

이러한 여섯 단계의 업무 플로우에서 명확한 역할과 책임, 추적 가능성과 지속적인 평가를 적용해야 한다. AI 모델의 라이프사이클 관리를 위해서는 일반적인 제품 개발 프로세스보다 더욱 세분화되고 통제된 프로세스가 필요하다. 전략을 이해한 뒤, AI 적용 범위를 결정하고 단계별 질문을 제시하여 AI 모델 배포의 타당성을 검증한다. AI 모델의 적용 범위를 결정한 뒤, 배포 이후의 사용자 반응에 따라 지속적으로 모델의 개선 및 중단 여부를 판단한다. 기능을 배포하는 것 이상으로 모델의 상태와 반응을 지속적으로 관찰하며, 모델의 라이프사이클을 관리하여 품질을 높이는 것이 핵심이다.

5. 결어

'5시간 걸리던 데이터 업무 1분 만에' 국내 언론사의 최근 기사 제목이다. 불과 몇 개월 만에 현시점에서는 현실화되었을 뿐만 아니라, 단순 업무 외에도 복잡하고, 해결하기 힘든 의사결정도 생성형 AI를 활용하면 어떻게 될지 초미의 관심사가 되었다. 이러한 기술 발전에 국내 기업도 박차를 가하고 있다. 최근 몇 년 동안 다양한 산업의 국내 기업들이 AI 기술에 많은 투자를 하고 성과를 거두고 있는 동시에 법적, 윤리적 위험을 불러일으키기도 한다. 특히 생성형 AI의 발전과 잠재력은 의심할 여지없이 광범위할 것으로 예상되므로 국가 사회 전반에 미치는 영향력을 주의 깊게 모니터링할 필요가 있다. 이에 최근 바이든 정부의 행정명령과 EU의 AI 법안은 국제 협력을 통해 법제 개선 방향의 지향점을 맞추어 가면서도 신속성을 담보할 수 있도록 국내 법제 개선에 대한 노력을 필요하다고 시사하고 있다. 이에 정부와 기업은 Responsible AI구현을 통해 윤리적 가이드라인을 마련하고, 이를 준수하도록 하는 법제화가 반드시 필요하다.

참고 문헌

- White House mobilizes bold push for responsible adoption of AI - PwC (2023. 11)
- FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence - The white house (2023. 10)
- PwC's Responsible AI - AI you can trust - PwC (2023. 5)
- What is responsible AI and how can it help harness trusted generative AI? - PwC (2023. 4)

Contacts

PwC Consulting

구 본 재, Partner

+82 2 3781 1435
bon-jae.koo@pwc.com

양 대 일, Partner

+82 2 3781 9979
dae-il.yang@pwc.com

www.pwcconsulting.co.kr

2312C-RP-045