

목차

- I. 서론
 - 1. GPAI란?
 - 2. 워킹그룹에 대한 설명 및 연계
- II. AI와 데이터
 - 1. 둘 간의 관계
 - 2. 발생하는 문제: 편향성을 중심으로
 - 3. 해결 방안: 투명성을 중심으로
- III. GPAI의 정책적 권고사항의 틀
- IV. 한국의 대응방안: 정부차원과 민간차원에서
 - 1. 정부 차원
 - 1-1 한국의 현황
 - 1-2 각 국의 대응
 - 1-3 합의
 - 2. 민간 차원
 - 1-1 민간 역량 향상의 필요성
 - 1-2 미디어 리터러시 각 국 사례
 - 1-3 합의
- V. 결론

I. 서론

본 연구는 GPAI의 워킹그룹 중 책임있는 AI와 데이터 거버넌스 간의 연결성과 공통점에 집중하여 AI가 데이터를 활용하는 과정에서 발생하는 편향성 문제와 이를 해결하기 위한 투명성이라는 방안을 검토한다. 이를 바탕으로 GPAI가 권고하는 정책적 권고 사항 중 투명성의 측면에 집중하여 한국의 현황 및 각국의 대응을 살펴보고, 한국의 합의를 도출한다. 마지막으로 이러한 정부적 조치가 효과적일 수 있도록 민간의 역량을 증진하는 방법을 살펴본다.

1. GPAI란?

글로벌 파트너십 인공지능(GPAI)은 인공지능(AI)의 책임감 있는 개발과 사용을 지표로 삼는 국제적이고 다자간의 이니셔티브다. 이는 인권, 기본적인 자유, 그리고 OECD 인공지능 권고안에서 반영된 공유된 민주주의적 가치를 지향한다. 이 GPAI의 설립 회원에는 호주, 캐나다, 프랑스, 독일, 인도, 이탈리아, 일본, 멕시코, 뉴질랜드, 대한민국, 싱가포르, 슬로베니아, 영국, 미국, 그리고 유럽연합이 포함되어 있다.

GPAI는 AI와 관련된 이론과 실제 사이의 격차를 줄이는 것을 목표로 산업, 시민 사회, 정

부, 학계의 다중이해당사자들을 통합한다. 이를 위해 최첨단 연구와 응용 활동을 지원하여 국제 협력을 촉진하며, 네 가지 주요 워킹 그룹 주제에 걸쳐 전문가들이 협력하고 있다. 이는 1) 책임 있는 인공지능, 2) 데이터 거버넌스, 3) 미래의 일터, 4) 혁신 및 상업화를 포함한다.

GPAI는 AI 실무자들 사이에서 다학제 연구를 공유하고 핵심 문제를 식별하는 메커니즘을 제공하며, 이를 통해 국제 협력을 용이하게 하고, 중복을 줄이며, 특정 AI 문제에 대한 글로벌 참조점으로서의 역할을 하고, 결국 신뢰할 수 있는 AI의 신뢰와 채택을 촉진한다. 워킹 그룹 내에서의 협력을 통해, GPAI는 AI의 영향을 이해하는 데 관련된 과학적, 기술적, 사회경제적 정보를 포괄적이고 객관적이며 공개적이고 투명한 기준으로 평가한다. 이는 AI의 책임 있는 개발을 장려하며, 잠재적인 도전에 대응하고 이를 완화하는 옵션을 제시한다.

2. 워킹그룹에 대한 설명 및 연계

RAI(책임성 있는 AI):

- 인간 중심적이고 공정하며 공평하고 포용적이며 인권과 민주주의를 존중하고 공익에 긍정적으로 기여하는 것을 목표로 하는 AI의 비전에 기반을 두고 있음.
- RAI의 임무는 이러한 비전과 GPAI의 전반적인 사명, 즉 유엔의 지속 가능한 개발 목표에 따라 인간 중심의 AI 시스템의 책임 있는 개발, 사용 및 거버넌스를 육성하고 기여하기 위해 노력하는 것을 지향한다.

데이터 거버넌스:

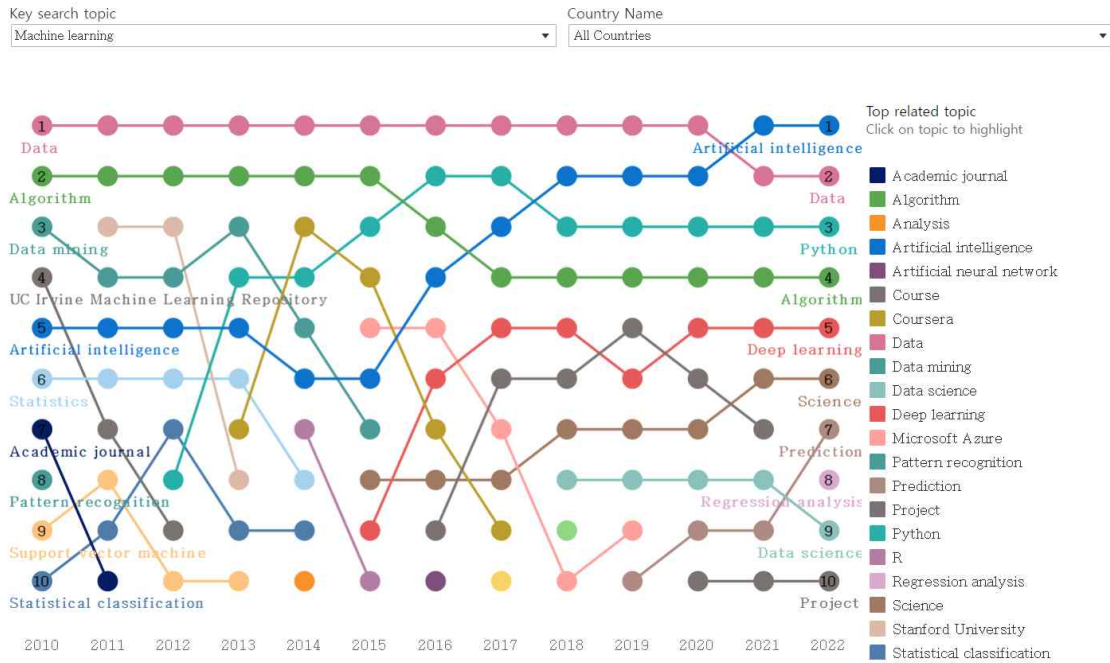
- GPAI 참조 약관에 따르면 '인권, 포용, 다양성, 혁신, 경제 성장 및 사회적 이익에 부합하는 방식으로 AI용 데이터가 수집, 사용, 공유, 보관 및 삭제되는 것을 촉진
- 동시에 UN 지속 가능한 개발 목표를 달성하기 위해 증거를 수집하고, 연구를 구체화하고, 응용 AI 프로젝트를 수행하고, 데이터 거버넌스에 대한 전문 지식을 제공하는 임무를 수행

RAI는 각 프로젝트가 공통된 차원을 공유하는 경우 데이터 거버넌스 워킹 그룹과 상호 작용한다.

⇒ 데이터 거버넌스 워킹 그룹과 책임 있는 AI 워킹 그룹의 임무는 특히 밀접하게 관련되어 있고, 어느 정도 중복되며 GPAI도 이를 인식하고 워킹 그룹 간 연계를 도모하고 있다. AI와 데이터는 단절된 맥락으로 보아선 안 된다.

- 일반적으로 책임감 있는 AI 워킹 그룹이 집중하는 바는 AI 개발을 모델링하는 방법과 어떤 데이터 세트를 사용할 것인지에 대해 더 많이 검토하여 AI가 책임감 있는 방식으로(예: 불필요한 편견 없이). 형성되고 기능할 수 있도록 하는 것을 목표로 한다.
- 데이터 거버넌스 워킹 그룹은 데이터를 처음에 어떻게 책임감 있게 수집하고 관리할지에 초점을 맞춘다. 특히 데이터의 출처와 맥락에 어떤 식으로든 연관된 당사자나 데이터 사용으로 인해 영향을 받을 수 있는 당사자(예: 데이터 주체와 데이터가 수집되는 커뮤니티의 구성원들)의 상황을 고려한다.

II. AI와 데이터



Explanation: This visualization displays the top ten search terms per country related to Machine Learning topics from 2010 to 2022. Topics are represented by color, and ranked by popularity in a given year.

Note: Rankings are calculated by the 'most-searched' topics from January through December of a given year. Values are normalised per country and year. Terms are filtered to exclude certain searches that are unrelated to the original core concept. Please see [methodological note](#) for more information.

Source of data: Google Trends *Please cite as:* OECD.AI (2023), visualisations powered by Tableau using data from Google Trends, accessed on www.oecd.ai. Supported by the [Patrick J. McGovern foundation](#).

그림 1 OECD 제공. 머신러닝 토픽과 연계된 탑 토픽 1위, 2위에 위치한 AI와 Data

OECD가 제공하는 탑 토픽에 대한 그래픽에서 보여주듯이, AI와 데이터는 전 세계적으로 오늘날 뜨거운 이슈로 자리 잡고 있습니다. AI와 데이터를 독립적으로 보는 것보다 연계해서 바라보아야 하며, 둘 간의 관계에 집중하여 그 이유를 제시하고자 합니다. AI와 데이터 간의 관계의 중요성을 보여주기 위해 GPAI의 보고서를 참고하여 정리하였습니다.

1. 둘 간의 관계¹⁾

AI의 맥락에서의 데이터:

- 일반적으로 디지털 데이터 또는 디지털 데이터로 쉽게 변환될 수 있거나 기계에 의한 처리가 가능한 데이터를 의미한다.
- 일반적으로 디지털 데이터의 주요 기능은 **세상에 대한 정보를 구현하는 것**이며, 거버넌스 맥락에서 중요한 것은 바로 이러한 **정보 표현으로서의 데이터 차원**이다.

1) GPAI (2022). Data Governance Working Group - A Framework Paper for GPAI's Work on Data Governance 2.0, Report, November 2022, Global Partnership on AI, Paris.

1) AI의 학습, 검증 및 테스트를 위한 데이터

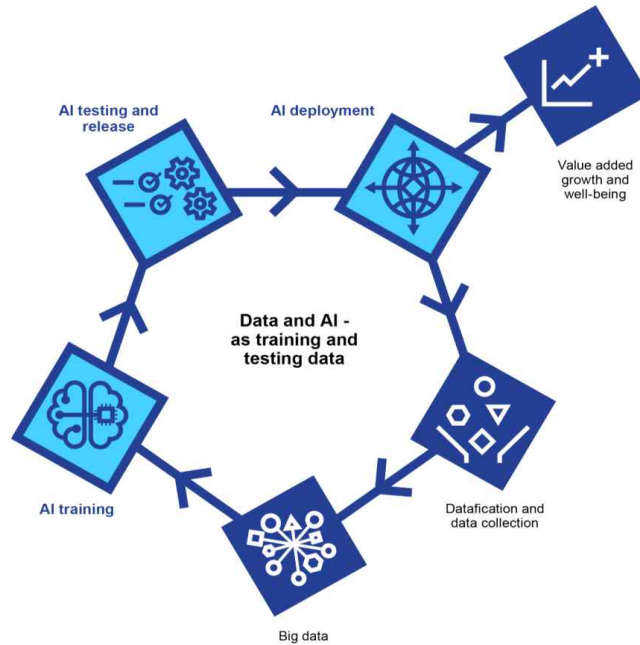


그림 2 Figure 2: Data and AI - as training and testing data (adapted from OECD 2015)

- 학습 데이터: AI 시스템이 결과를 생성하기 위해 사용하는 모든 기술을 적용하고 개선하는 방법을 이해하는 데 사용되는 데이터 집합
- 검증 데이터 및 테스트 데이터: 각각 AI 시스템의 작동(학습 데이터에서 학습한 내용)을 반복적으로 평가하고, 특히 '실제로 보고자 하는 데이터'를 도입했을 때 정확한 결과를 생성할 준비가 되어 있는지에 초점을 맞춰 AI 시스템이 목적을 얼마나 잘 수행하는지에 대한 최종 분석을 수행하기 위해 사용됨

그림에서 나타나듯이 AI가 데이터에 의해 학습 데이터에 의해 훈련되고, 이후 테스트를 받는 순서로 AI가 발전된다고 볼 수 있다.

- 널리 알려진 AI의 대부분은 인간 코더가 (완전히) 프로그래밍한 것이 아니며 반복적인 경험을 통해 성능을 개선하고 대량의 데이터를 필요로 하는 프로세스를 통해 작업을 수행하는 방법을 학습한 것이다.
- 예를 들어, 특정 그룹에 대해서만 데이터를 수집하거나 불안정하거나 누락된 데이터 단위에 대해 데이터를 제거할 경우, 두 가지 모두 모델의 전체 출력에 상당한 영향을 미칠 수 있음

이처럼 AI는 데이터를 통해 학습하고 반복하여 그 성능을 개선해 나가며, 어떤 데이터를 통하여 학습하냐에 따라 다른 결과가 나올 수 있습니다.

2) AI 기반 시스템의 입력 및 출력으로서의 데이터

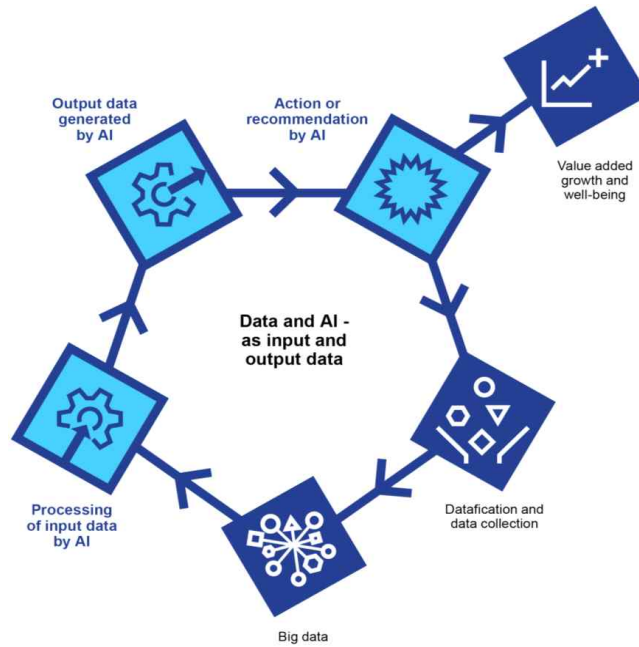


그림 3 Figure 3: Data and AI - as input and output data (adapted from OECD 2015).

- 일단 개발 및 배포된 AI는 알고리즘 시스템으로도 볼 수 있다.
- 따라서 분류, 예측 또는 추천과 같은 특정 출력 데이터를 얻기 위해 입력 데이터를 처리하는 데 사용됨(인풋과 아웃풋)
- 출력 데이터는 다시 동일 또는 다른 시스템에 대한 입력 또는 학습 데이터가 될 수 있음

⇒ 데이터는 AI 시스템의 입력 데이터로서의 역할뿐만 아니라 AI 결과물을 학습, 검증 및 테스트하는 데에도 중요한 역할을 함

- 학습 데이터에 결함이 있으면 신뢰할 수 없는 결과, 의사 결정 또는 출력 데이터로 이어질 수 있으므로 학습 데이터의 품질과 양이 중요하다
- 거버넌스 관점에서 특히 집중되어야 할 점은 훈련 데이터에 존재하는 어떠한 편향이든 AI 시스템의 개발에 영향을 미쳐 그 프로세스나 결과물이 해당 편향을 강화시키는 결과를 만들 수 있다는 것이다

2. 발생하는 문제: 편향성을 중심으로²⁾

앞에서 살펴보았듯이 AI는 데이터를 활용하여 발전하고, 검증하며, 결과물을 산출하므로 둘 간의 관계는 밀접하다는 것을 알 수 있습니다

이러한 관계에 있어 최근 많은 관심을 받고 있는 문제 중 하나는 데이터와 알고리즘의 편향성임.

아래 사례는 AI 알고리즘과 데이터의 편향성의 이해를 돕기 위해 간략한 예시를 제시하겠습니다.

1) 편향성에 관한 연구 및 사례:

① Noble, S. U.의 *Algorithms of Oppression: How Search Engines Reinforce Racism*: 검색 엔진에 소수자에 대한 편견이 어떻게 내재되어 있는지에 대해 설명한 종합적인 연구



뉴스 비디오 라디오 다운로드 TOP 뉴스

성차별: 아마존, '여성차별' 논란 인공지능 채용 프로그램 폐기

2018년 10월 11일

- ② 아마존의 채용 시스템(Reuters, 2018): Amazon.com Inc의 머신러닝 전문가들이 새로운 채용 엔진이 여성을 선호하지 않는다는 큰 문제를 발견함.
- 해당 팀은 2014년부터 인재를 찾는 과정을 기계화하기 위해 입사 지원자의 이력서를 검토하는 컴퓨터 프로그램을 개발
 - “말 그대로 이력서 100개를 주면 상위 5개의 이력서를 산출하고, 그 중에서 채용할 수 있는 엔진이 되길 원했습니다.”
 - 2015년, 아마존은 새로운 시스템이 소프트웨어 개발자 및 기타 기술직 지원자를 성 중립적인 방식으로 평가하지 못한다는 사실을 파악함.
 - 그 이유는 아마존의 컴퓨터 모델이 10년 동안 회사에 제출된 이력서의 패턴을 관찰하여 지원자를 심사하도록 훈련되었기 때문.
 - 대부분은 남성이었는데, 이는 기존 기술 업계 전반에 남성 기술자가 많은 것을 반영하는 결과였음.
 - 사실상 아마존의 시스템은 기존의 편향된 데이터를 통해 남성 지원자가 더 유리하다고 스스로 학습한 것임.
- ⇒ 앞에서 검토하였듯이 AI가 어떤 데이터를 학습하냐에 따라 편향된 결과가 산출될 수 있음.

2) GPAI (2022). Data Governance Working Group - The Role of Data in AI, November 2022, Global Partnership on AI, Paris.

2) AI 개발 과정 및 결과에 영향을 미치는 데이터 품질 중 대표성의 특성

- 개발 단계에서 사용되는 데이터의 품질은 AI가 정확하고 유효하며 편견 없는 실제 결정을 내리는 지점에 도달하기 위해 매우 중요
- 아직까지 데이터 품질에 대한 단일 정의는 없으며, 데이터 품질에 대한 벤치마크를 정의하는 데 있어 다양한 접근 방식이 사용되고 있음
- 그러나 품질 데이터의 공통 분모는 '사용하기에 적합하고 사양, 요구 사항 및 기대치를 충족하는 데이터'라는 데에는 의견이 일치함(Fürber, 2015).

데이터 품질 중 편향성과 관계가 깊은 특성

대표성: 데이터 수집의 편향(선택 편향, 제외 편향 등)으로 인해 과대 또는 과소 대표성이 발생할 수 있음

알고리즘의 결과를 신뢰하려면 학습 데이터가 알고리즘이 사용될 실제 인구 통계를 대표해야 합니다.

첫 번째: 과소대표성

- 얼굴 인식 및 성별 분류 도구의 편향성과 관련하여 많은 논의가 이루어졌으며, 다양한 인구통계학적 그룹에 대한 알고리즘 정확도가 다양하다는 문헌이 있음
- 미국 내 다양한 출처(사법 시스템, 이민 및 국경 통제)의 네 가지 데이터 세트를 사용한 NIST의 연구에서는 성별, 연령, 인종 또는 출생 국가에 따라 정의된 얼굴 인식 알고리즘의 정확도를 정량화하려고 시도(Grother, et al., 2019).
- 그 결과 여성, 원주민, 아시아인, 아프리카계 미국인 피험자에서 높은 오탐지율(즉, 피험자의 얼굴이 비슷한 경우 샘플을 잘못 연관시키는 것)이 나타남
- 반면, 중국에서 알고리즘을 개발한 경우 아시아 피사체 이미지에 대한 오탐지율이 현저히 낮았음.
- 개발 지역에 따른 알고리즘 성능의 이러한 차이는 사용된 학습 데이터의 대표성의 중요성을 더욱 강조
- 간단히 말해, 학습 및 검증 데이터가 실제 인구를 대표하지 못하면 AI가 실제 적용에서 이를 인식하지 못할 가능성이 높으며, 이는 기존의 불평등을 증폭시킬 위험이 있습니다.

두 번째: 과대대표성

- 데이터 수집 및 선택 과정을 통해 편향이 도입될 수 있는 민감한 특성의 과대표현
- 범죄 예측 알고리즘은 편향된 데이터의 사용을 보여주는 빈번하게 사용되는 예시
- 이는 특정 높은 인종 상관관계를 가진 이웃지역이 알고리즘 훈련용 샘플링에 선택되기 때문입니다 - 이미 경찰에게 알려진 지역 - 이는 그러한 지역에서의 더욱 증가된 경찰 활동을 초래하고, 새로운 범죄가 관찰되고, 따라서 더 높은 범죄 통계와 더욱 높은 경찰 수준을 초래합니다(Silberg and Manyik, 2019; Lum and Isaac, 2016).
- 따라서 경찰의 편향성이 데이터에 스며들어 이 분야의 AI 개발에서 추가적으로 복제되고 확대됩니다.
- 유럽 수준에서 유럽 커미션은 성별과 민족성 면에서 보다 폭넓게 대표되는 데이터 세트의 사용에 대한 정책 권고를 내놓았음에 주목할 가치가 있습니다(European Commission, 2020).

3) 정부 서비스 맥락에서 데이터로 인한 알고리즘의 편향 사례

- 유엔 극빈 및 인권 특별보고관은 디지털화된 복지 국가가 취약한 사람들에게 막대한 영향을 미칠 것이라고 밝힘(Alston A/74/48037 2019).
- 이러한 문제는 다양한 목적의 AI 개발에서 드러날 수 있음
- 일부 사람, 그룹 및 지역에 대한 데이터를 다른 사람보다 쉽게 얻거나 사용할 수 있으며, 이는 또한 배제 문제로 이어질 수 있음.

ex) 자동화된 의사 결정을 위한 사회 보장 지급 및 알고리즘 설계

호주에서는 부분적으로 자동화된 시스템이 한 주 동안의 실제 수입이 아닌 여러 주 동안의 평균 수입을 계산하여 특히 취약한 인구 비율에 영향을 미치는 오류로 인해 약 50만 건의 잘못된 채무 통지서가 사회보장 수급자에게 발송됨(Carney 2019).

⇒ 데이터 피드에 부정확하거나 오래되었거나 기한이 지난 정보가 포함되어 자동으로 의사 결정이 내려지는 취약한 사람들에게 불공평하게 영향을 미칠 수 있음을 보여주는 예시

- 편견과 관련된 사회윤리적 문제는 AI 개발 과정에서 드러날 수 있음.
- 따라서 이러한 시스템은 포용을 실현하지 못하고 사회 구성원을 부당하게 배제할 수 있음.

(편향성 부분을 마치며)

보건, 정부 등 중요한 분야는 물론 경제 전반에서 효율성과 인사이트의 심도 있는 향상을 통해 AI가 가져다주는 상당한 이점 존재할 수 있음. 또한, 새로운 제품과 솔루션(예: 보건 분야, 전문, 과학 또는 기술 서비스)의 개발로 인해 혁신에서 AI의 역할이 중요해졌음. 그러나 근본적인 구조적 문제로 인해 불평등이 심화되고 장기적으로 경제적 피해가 발생하며 혁신이 억제될 가능성이 존재. 또한 데이터의 출처, 데이터 선택, 데이터 세트에 내재된 편향성, 데이터 접근성 부족, 데이터의 품질 및 특성 등이 개인에게 해를 끼치고 취약 계층을 소외시키며 불평등을 증폭시킬 수 있음.

⇒ 이러한 문제를 해결하기 위해서는 투명성이 요구됨

투명성이 포함해야 하는 요소들인데, 설명 가능성에 집중해도 무방함.

3. 해결 방안: 투명성을 중심으로

1) 투명성이 필요한 이유

- 사용자가 데이터 사용의 잠재적인 유익한 영향과 유해한 영향, 데이터를 통해 도출된 인사이트, AI를 포함한 해당 데이터를 사용하는 시스템이 내린 결정을 더 잘 이해할 수 있도록 지원
- 법률을 위반한 경우를 식별하고 해결할 수 있음
- 이를 통해 데이터와 AI가 사용하고자 하는 대상과 존중하고 품위 있고 신뢰할 수 있는 상호작용을 할 수 있는 기회를 늘릴 수 있음

거버넌스와 더불어 민간의 역할도 필요

- 투명성은 전통적으로 거버넌스와 관련이 있지만, 데이터와 관련하여 민간 부문의 역할을 강조할 필요가 있음

- 산업계의 경우, 투명성은 정부와 사회를 설득하고 대중의 정서, 도덕은 물론 개인정보 및 데이터 보호법에 위배되는 행동을 하지 않는다는 것을 보여주기 위한 노력에 필수적임
- 또한 정부 및 데이터 보호 기관의 경우 사전 모니터링 활동과 조사 활동을 원활하게 수행하기 위해 데이터 투명성이 요구됨

2) AI와 데이터 간의 관계에서 투명성

- 투명성 노력에는 AI를 포함한 모든 데이터 기반 프로젝트를 시작하기 전에 데이터 품질 평가가 포함되어야 하며, 이를 통해 프로젝트의 목표 또는 AI가 사용 중인 데이터와 일치하는지 여부를 결정해야 함
- 평가에는 어떤 데이터가 사용되었는지, 왜 그 데이터가 사용되었는지, 데이터가 AI가 해결하고자 하는 질문에 답하는 데 적절하고 수용 가능한 증거로 간주되는지 등의 고려 사항이 포함되어야 함
- 이러한 종류의 평가는 AI의 시스템이 블랙박스(인공지능이 내린 결론이 어떤 과정과 방법을 통한 것인지 설명이 불가능한 문제)로 점점 더 많이 사용하기 때문에 특히 중요한데, 신경망은 때때로 어떤 결정이 어떻게, 왜 내려졌는지 이해하는 것을 거의 불가능하게 만들기도 합니다(PFST 2020).
- 운영 목표, 입력, 출력, 특히 결과 측면에서 이러한 초기 질문, 즉 데이터와 알고리즘의 인간적 맥락에 초점을 맞추는 것은 투명성의 중요한 형태(Kroll 2018). 여기에는 데이터를 수집할 때 적용한 가정과 데이터가 세상과 AI의 대상을 충분히 그리고 타당하게 반영하는 것으로 간주된 방법을 설명할 수 있어야 함

투명성이 포함해야 하는 요소들인데, 설명 가능성에 집중해도 무방함.

3) 투명성이 포함해야 하는 요소

① 설명 가능성:

- 이는 사용자와 대중이 AI 시스템을 해석할 수 있는 능력을 의미하며, 이를 통해 AI에 의한 의사 결정이 어떻게 이루어지고, 무엇이 고려되며, 왜 그런 결정을 내렸는지 이해할 수 있음
- 사용자가 특정 상황에서 알고리즘이 어떻게 작동하고 왜 그런 결과를 생성했는지 알 수 있는 능력(PFST 2019).
- 중요한 것은 설명가능성이란 사용자 또는 대중이 AI의 근거가 알고리즘 결과와 자신의 관행 및 결정에 어떻게 관련되고 영향을 미치는지 이해할 수 있는 능력을 의미
- 따라서 설명 가능성은 설명으로 제공되는 내용과 적절한 교육 및 디지털 리터러시 기술을 결합하여 사용자의 역량을 강화해야 함(유럽의회 2019).
- 이는 특히 대중에게 중요: 데이터 관계의 사슬을 대중에게 설명할 수는 있지만, 대부분의 사람들이 가지고 있지 않은 리소스와 기술이 필요(오스왈드 외. 2018).

② 정당성:

- 이는 사용자가 a) 도구가 어떻게 설계되었는지, b) 광범위한 의사 결정 프로세스의 일부로 AI가 어떻게 구현되었는지, c) 결과가 특정 의사 결정에 어떻게 사용되는지를 정당화할 수

있을 정도로 도구에 대해 충분히 이해하는 것을 의미(AIHLEG 2019).

- 여기에는 AI가 윤리적 문제에 어떻게 접근하는지(예: 공정성에 어떻게 접근하는지, 편향성 확인을 어떻게 지원하는지, 설계자와 사용자가 AI가 차별하지 않도록 지속적으로 보③장하는지),
- 맥락에서, 그리고 사용자나 대중이 AI의 결과를 신뢰해야 하는 이유가 포함되어야 함.
- AI의 근거가 명확하게 제시되지 않으면 데이터 처리에 대한 의미 있는 동의가 불가능하며, 결정에 대한 이의 제기가 정당화되기 어려워짐(Mittelstadt 외 2016).

③책임성:

- 알고리즘의 작동 방식을 이해하는 능력, AI 시스템, 제품 또는 서비스의 다양한 속성과 사용 방법을 설명할 수 있는 능력을 의미
- 불투명한 AI와 불투명한 데이터 사용은 인권법을 포함한 법률을 위반했는지, 위반한 법률이 어디에 있는지, 책임이나 의무를 어떻게 귀속시킬지 파악하기 어렵게 만듦(유럽위원회 2020).
- 투명성 메커니즘은 사용 맥락과 책임 조치를 지원하기 위해 존재하는(또는 존재해야 하는) 규정과 신중하게 균형을 맞춰야 함

④추적 가능성:

- 특정 기능이나 인사이트가 어떻게 도출되었는지, 어떤 논리를 통해 생성되었는지 이해하기 위해 역추적할 수 있는 기능
- 여기에는 어떤 데이터 세트가 어떤 의사 결정에 사용되었는지, 그리고 AI를 학습시키기 전에 어떻게 수집, 레이블 지정, 정리되었는지 명확히 파악하는 것이 포함(AIHLEG 2019).
- 책임 귀속을 위해 중요하며, "시스템이 남긴 흔적을 통해 시스템 작동을 설명할 수 있으므로 설명 가능성의 품질을 보장"(Cerna 2018, 18쪽).
- 취해진 조치를 정당화할 뿐만 아니라 광범위한 데이터 관행에 대한 시사점으로서도 중요
- 예를 들어, 데이터를 사용하여 어떤 결정을 내렸는지 역추적할 수 없다면 잘못된 결정에 대한 책임이 어디에 있는지 알기 어려움, 이러한 기능이 없으면 올바른 수정 경로를 파악하는 것이 거의 불가능

⑤투명성이 해를 끼치지 않는지 확인하기:

- 투명성 확보에는 때때로 비용이 수반되거나 피해가 발생할 수 있다는 점도 유의해야함
- 이러한 피해는 비례성의 원칙에 따라 측정: 즉, 피해가 이익보다 더 큰지 평가
- 투명성을 위해 정보를 공유하는 것이 득보다 실이 더 클 수 있기 때문에 정보를 누구에게 투명하게 공개해야 하는지가 항상 명확하지는 않음(Floridi and Taddeo 2018).
- 예를 들어, 데이터에 의료 데이터와 같이 개인 정보나 민감한 정보가 포함되어 있는 경우, 이를 대중에게 투명하게 공개하는 것은 책임감보다 더 큰 해를 끼칠 수 있음.
- 마찬가지로 데이터와 프로세스를 투명하게 공개하면 데이터가 어떻게 사용되었는지에 대한 대중의 평가가 가능하지만, 공개 대상과 대상에 따라 특정 집단의 취약성이 증가할 가능성도 존재(AIHLEG 2019).
- 여기에는 설명이 의미 있고 목적에 맞는지 확인하는 데 필요한 추가 직원 교육 및 리소스도 포함

- 마찬가지로 규정 준수를 감독할 수 있는 규제 기관이 있어야 함
- 또한 투명성은 영업 비밀과 균형을 이루어야 하며, 기업은 실행 가능한 감독 프로토콜이 필요요함.
- 더욱이 "데이터에 대한 대중의 접근은 무분별하거나 부도덕한 데이터 오용을 통해 이해관계에 따른 논란에 불을 붙일 수 있으며, 알고리즘을 더 쉽게 해킹할 수 있게 만들거나 의도하지 않은 개인정보 침해로 이어질 수 있습니다."(PFST 2019).

(투명성 파트를 마치며)

- 공공 및 민간 주체 간의 데이터 접근과 공유를 규율하는 구체적인 규칙을 수립하는 것이 AI 개발에 도움이 되고 데이터의 흐름과 가용성을 높일 수 있음
- 이러한 문제에 대한 대응은 정책과 관행을 통해 적절하게 다루어지고 인식되어야 함
- 궁극적으로 데이터와 AI를 사용할 때는 투명성의 원칙을 준수하여 긍정적인 결과를 극대화하고 부정적인 사회윤리적 영향을 최소화해야 함
- 투명성은 현재 관행에 대한 의문을 제기하고 앞서 언급한 문제를 해결할 수 있는 기회를 제공하며 정책 권고 및 모범 사례로 이어질 수 있는 개념임

III. GPAI의 정책적 권고사항의 틀³⁾

GPAI는 이러한 앞서 논한 편향성과 같은 AI와 데이터에서 발생하는 위험을 최소화하기 위한 10가지 정책적 권고사항을 제시함. 그 중 투명성의 요소도 포함되어 있음.

“데이터 관행 및 시스템의 투명성”:

데이터 수집 사용 및 데이터 기반 혁신 프로세스에서 권한을 가진 사람들은 AI 입력 및 알고리즘에 대한 정보를 포함하여 어떤 데이터가 수집되고 어떻게 사용되는지에 대한 정보를 공개하고 이 정보를 영향을 받는 개인과 커뮤니티에 직접 제공해야 할 의무가 있습니다.

⇒ 이러한 GPAI의 권고사항 중 투명성에 측면에서 한국의 현황과 함의를 제시하고자 함.

3. 해결 방안: 투명성을 중심으로

<p>개관</p> <p>1.기업의 알고리즘 투명성 공개에 대한 법규 마련</p> <p>유럽의 경우 알고리즘 투명성을 법제화 하고 있음</p> <p>한국은 1. 알고리즘에 관한 가이드라인이나 권고사항 밖에 없으며, 법제화는 아직임. 2. 더불어 네이버와 카카오 같은 기업들은 알고리즘의 투명성이 기업의 기밀 사항을 까발리는 행위라고 알고리즘의 투명성 공개를 반대함.</p> <p>따라서, 한국은 알고리즘을 중점적으로 다룰 수 있는 기관이 필요하며(알고리즘투명성위원회) 이 기관은 GDPR을 참고하여 투명성 공개를 찬성하는 정부와 투명성 공개를 반대하는 기업 간의 협의점을 찾은 법안을 발표해야 할 것임.</p>

3) GPAI (2022). Data Governance Working Group - Data Justice Policy Brief: Putting data justice into practice, November 2022, Global Partnership on AI, Paris.

2. 디지털 리터러시 향상

> 정부가 소비자 권리 보호를 위해 알고리즘의 투명성을 공개하는 법안을 만들더라도, 결국 소비자가 투명성 법안에 관심이 없거나, 혹은 이를 해석할 능력이 되지 않는다면 정부의 법안은 큰 효력을 발휘 할 수 없음.

따라서 시민들의 디지털 리터러시가 필요한것

영국의 경우 정부 총괄하에 디지털 리터러시 관련 교육 자료 배포 및 방송사와 연관지어 민간이 직접 디지털 리터러시 관련 활동 참여 확대를 위해 노력함

따라서 한국도 영국의 사례를 참고하여, EBS를 통한 관련 자료 배포 및 민간들의 참여 확대를 위한 프로그램 기획이 필요할것

1. 정부 지원 - 기업의 알고리즘 투명성 공개에 대한 법규 마련

1-1 한국의 현황

- 과학기술정보통신부가 「인공지능 윤리 가이드라인」과 「신뢰할 수 있는 인공지능 실현 전략」 등 현재 한국에서는 가이드 라인만 존재하며,
- 알고리즘의 투명성과 관련해서 법제화 된 것은 없음

한국의 경우, 네이버와 카카오톡의 경우 알고리즘의 메커니즘을 공개하는 것은 기업 기밀을 공개하는 것과 마찬가지로 알고리즘의 메커니즘을 공개하는 것을 거부하고 있음.

이에 대해 방송통신위원회가 네이버와 다음 등 대형 포털의 뉴스와 동영상 추천 과정을 투명화하기 위해 '알고리즘투명성위원회'를 법적 기구화하는 방안을 추진하고 있음. 이는 포털뉴스와 동영상 추천 알고리즘 투명화를 위해 기사 배열과 노출 기준을 검증하는 알고리즘투명성위를 법적 기구로 설치하는 방안을 검토하고 추진하고 있음. 4)

4) <http://www.newsroad.co.kr/news/articleView.html?idxno=15166>

뉴스홈 | 최신기사

방통위, 포털뉴스 알고리즘투명성위원회 법정화 추진

송고시간 | 2023-02-03 14:02



이정현 기자
기자 페이지

새해 업무계획 발표...방송사 심사시 공익성 평가 강화
EBS 중학생 인터넷강의 프로그램 무료서비스 전환

<https://www.yna.co.kr/view/AKR20230203063800017>

1-2 각 국의 대응

알고리즘의 투명성과 관련한 다른 국가의 법제화 사례 (EU연합)

: EU는 기업이 알고리즘의 투명성을 밝힐 의무를 GDPR이라는 법안 안에 명시하고 있음.

- (13) In order to ensure a consistent level of protection for natural persons throughout the Union and to prevent divergences hampering the free movement of personal data within the internal market, a Regulation is necessary to provide legal certainty and transparency for economic operators, including micro, small and medium-sized enterprises, and to provide natural persons in all Member States with the same level of legally enforceable rights and obligations and responsibilities for controllers and processors, to ensure consistent monitoring of the processing of personal data, and equivalent sanctions in all Member States as well as effective cooperation between the supervisory authorities of different Member States. The proper functioning of the internal market requires that the free movement of personal data within the Union is not restricted or prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data. To take account of the specific situation of micro, small and medium-sized enterprises, this Regulation includes a derogation for organisations with fewer than 250 employees with regard to record-keeping. In addition, the Union institutions and bodies, and Member States and their supervisory authorities, are encouraged to take account of the specific needs of micro, small and medium-sized enterprises in the application of this Regulation. The notion of micro, small and medium-sized enterprises should draw from Article 2 of the Annex to Commission Recommendation 2003/361/EC⁽⁵⁾.
- (14) The protection afforded by this Regulation should apply to natural persons, whatever their nationality or place of residence, in relation to the processing of their personal data. This Regulation does not cover the processing of personal data which concerns legal persons and in particular undertakings established as legal persons, including the name and the form of the legal person and the contact details of the legal person.
- (15) In order to prevent creating a serious risk of circumvention, the protection of natural persons should be technologically neutral and should not depend on the techniques used. The protection of natural persons should apply to the processing of personal data by automated means, as well as to manual processing, if the personal data are contained or are intended to be contained in a filing system. Files or sets of files, as well as their cover pages, which are not structured according to specific criteria should not fall within the scope of this Regulation.

<https://gdpr-info.eu/> (GDPR 법률 일부)

< GDPR 13조 ~15조>

- 알고리즘에 의해 행해진 결정에 대해 질문하고 반박할 수 있다.
- 데이터를 다루는 사람은 개인에게 언제, 왜 데이터를 모으고 처리하는지 알려야 한다.
- 개인의 권리행사를 위해 투명하게 공개된 정보, 쉬운 의사소통 요구.

*** GDPR에 따라 다국적 기업들은 알고리즘의 메커니즘을 공개하고 있음.

GDPR

The General Data Protection Regulation (EU GDPR) went into effect on May 25, 2018, replacing the 1995 EU Data Protection Directive. On January 1, 2021, the retained EU law version of the EU GDPR went into effect in the UK (UK GDPR). The EU GDPR and the UK GDPR (together, the GDPR) lay out specific requirements for businesses and organizations who under the EU GDPR are established in Europe or who serve users in Europe and who under the UK GDPR are established in the UK or who serve users in the UK. The GDPR regulates how businesses can collect, use, and store personal data.

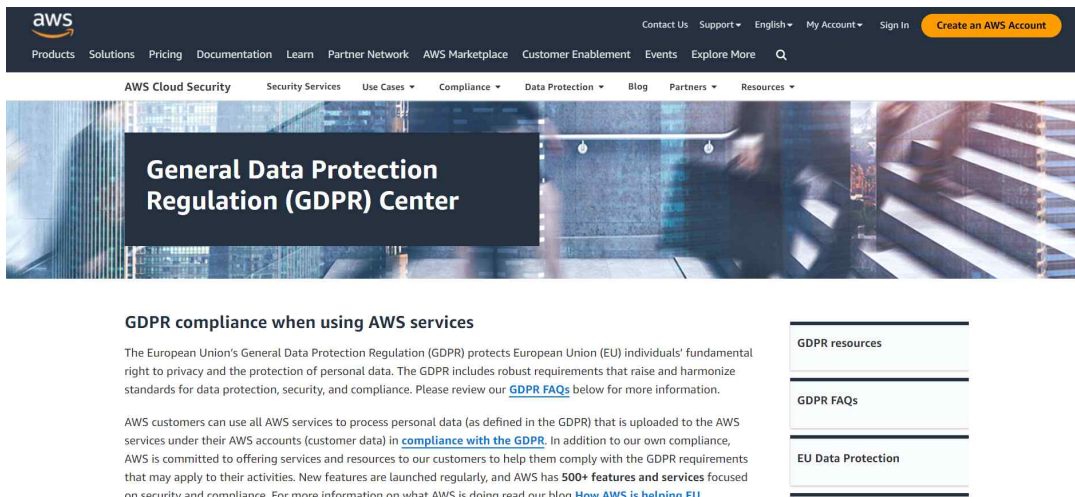
Terms & contractual protections

Where we act as a processor of personal data, we make available data processing terms, reflecting the controller-processor relationship, where required. Products where Google may act as a processor include:

- Ads Data Hub
- Google Ads Customer Match
- Google Ads Store sales (direct upload)
- Google Marketing Platform (including Display & Video 360, Campaign Manager, Search Ads 360, Google Analytics, Tag Manager, Optimize, Data Studio, and Google Analytics for Firebase)
- Google Cloud Platform services
- Google Workspace offerings

The data processing terms that we offer for the Ads products listed above are available [here](#). More information about the types of personal data in scope for those terms for each Ads product can be found [here](#). Information about Google Cloud Platform and Google Workspace commitments to the GDPR, including data processing terms, can be found [here](#).

(Google)



(amazon)

1-3 합의

알고리즘투명성위원회의 법정화가 무산되더라도, 알고리즘투명성 관리를 위한 단독적인 기관이 필요하기에, 다른 알고리즘투명성 관리를 위한 기관의 설립은 필수적이다. 이러한 기관은 영국의 GDPR을 참고하여 기업의 입장인 기업 기밀의 침해로 인한 공개 거부와 소비자 권리 보호 및 알고리즘의 투명성의 두가지 입장을 모두 보호하고 반영하기 위한 협의점을 찾은 법안을 마련해야 할 것임.

2. 민간 차원- 디지털 리터러시 향상을 위한 법안

2-1 민간 역량 향상의 필요성

- 설명 가능성은 설명으로 제공되는 내용과 적절한 교육 및 디지털 리터러시 기술을 결합하여 사용자의 역량을 강화해야 한다는 유럽의회의 입장을 재 상기하며, 현재 한국은 코딩과목을 필수 과목으로 인정하는 등 AI 교육을 의무화 하고 있음.
- 소프트웨어를 어떻게 작동시키고 사용하는지는 알지만, 정작 디지털 기술과 콘텐츠에 대한 비판적 해석 및 올바른 활용 능력을 기를 수 있는 교육은 제대로 이루어지지 못하고 있음.
- 즉, 아무리 정부가 알고리즘의 투명성을 위한 법안을 노력하여도, 소비자들은 자신들의 권리에 관심이 없고, 디지털 콘텐츠에 대한 비판적 시각을 갖지 못한다면, 정부의 노력은 큰 효과를 보지 못한다는 이야기임.

2-2 미디어 리터러시 각국의 사례

a. 영국은 방송통신 규제 기관인 오프컴이라는 기관이 있음. 오프컴은 디지털 사회에서 국민들의 참여도를 제고하기 위한 필수과제로 미디어 리터러시 증진을 명시하고, 이를 위해 이해 당사자와 대중, 민간 부문이 모두 참여하는 컴소시엄을 구축해 재정적 기여와 디지털 참여 확대를 추진해야 한다고 밝힘.

7. 오프컴은 초등학교 때부터 ‘미디어 스마트’라는 국가 주도의 수준 높은 교육 프로그램을 만들어 학교에서 사용 될 수 있게 하고 있음. 이는 많은 미디어 자료들 중 광고의 타깃 등 광고 콘텐츠를 있는 그대로가 아닌 비판적으로 이해하기 위한 구체적인 교육자료와 프로그램을 포함하고 있음.⁵⁾

About Us



Advertising can be a creative and inspiring world, where ideas come alive with drama, excitement and humour to entertain and engage us. But today, young people grow up in a world of commercial messaging that touches most areas of their lives – so it is more important than ever that they understand exactly what is being suggested, promised and sold to them.

Our innovative educational resources are designed to provide 7–17 year olds with the understanding and tools they need to be critical consumers of the media, looking at important subjects like social media and digital advertising, influencer marketing and body image.

In developing the key skills of media literacy – resilience, empathy, creativity, communication and critical thinking – young people are also building life skills which will help them on their journey towards future employability and well-being, and equip them to be successful global citizens.

Please take a look around, immerse yourself in our fascinating world and, above all, have fun teaching and talking to the young people in your lives.

For the latest news on our campaigns and teaching materials, sign up to our social feeds.

Useful Info

Terms & Conditions

Privacy Policy

Cookie Policy

Site Map

Expert Panel

Supporters

<<https://mediasmart.uk.com/> Media Smart 페이지 소개 글 - 미디어 스마트의 디지털 리터러시 증진을 위한 교육에 대한 내용을 명시하고 있음>

8. 오프컴은 공영방송 BBC와 협력하여 BBC Teach채널을 통해 미디어 리터러시 교육 콘텐츠를 제공하고 있음. 정확성과 편향성, 비판적으로 사고하는 방법 등 다양한 주제를 교육에서 다루고 있음.

또한, BBC Young Reporter을 통해 청소년들이 AI을 주제로 하여 직접 자신의 목소리를 담아 뉴스를 제작하고 이를 송출 할 수 있는 새로운 방송을 구상함. BBC Young Reporter는 1000여개 이상의 학교에서 3만명이 넘는 청소년들이 참여하고 있으며, 영국의 총리 , 영국 중앙은행 총재, 교육부 장관 등을 인터뷰하는 등 전문 기자들에 뒤떨어지지 않는 활약을 하고 있음.

6)

2-3 합의

a. 따라서 EBS와 같은 교육 사이트를 통한 시민들의 디지털 리터러시 능력 증진을 위한 교육 영상의 제작과 더 나아가 학교 교육에서는 디지털 리터러시를 다루는 과목을 지정해 선택 과목 혹은 필수 과목으로 선정하는 등 디지털 리터러시 증진에 대한 과제를 최우선시 해야 할 것임.

b. KBS와 SBS와 협업하여 공동체 구성원이 직접 관련 자료들을 만들고 배우는 참여 형태의 프로그램들을 통해 시민들의 디지털 리터러시 능력의 제고가 필요하다고 생각됨.

5) <http://www.newsroad.co.kr/news/articleView.html?idxno=15166>

6) <http://www.newsroad.co.kr/news/articleView.html?idxno=15166>

<참고문헌>

GPAI (2022). Data Governance Working Group - A Framework Paper for GPAI's Work on Data Governance 2.0, Report, November 2022, Global Partnership on AI, Paris.

GPAI (2022). Data Governance Working Group - Data Governance Working Group Report, November 2022, Global Partnership on AI, Paris.

GPAI (2022). Data Governance Working Group - Data Justice Policy Brief: Putting data justice into practice, November 2022, Global Partnership on AI, Paris.

GPAI (2022). Data Governance Working Group - The Role of Data in AI, November 2022, Global Partnership on AI, Paris.