AI는 성차별이 뭔지 알까?

페미니스트가 함께 만든 AI 가이드라인

Womenlink
한국여성민우회

본 사업은 한국여성재단이 지원하는 사업입니다.

목차

여는 글

“안녕, 나는 너의 첫 AI 친구 이루다야” 2020년 말, 6개월의 베타 테스트를 거쳐 챗봇 ‘이루다’가 등장했습니다. 폴더폰 시절 ‘심심이’와 채팅해본 적 있는 사람이라면 아주 반가워할 일이었죠. 단순한 모양의 노란 로봇이었던 심심이에 비하면 블랙핑크를 좋아하고 MBTI는 ENFP인, 2살 어린 여동생을 가진 20살 이루다가 훨씬 친구에 가까워 보이기도 합니다. 그래서일까요? 정식 출시 후 무려 80만 명의 ‘친구’가 루다와 대화를 나눴다고 하는데요. “사람들의 크고 작은 외로움을 메워주기 위한 제품이자 대화 상대”였던 이루다는 어째서인지 성소수자, 장애인에 대한 혐오 발언을 쏟아내고 인종차별까지 서슴지 않았습니다. 어떤 ‘친구’는 이루다에게 성희롱 발언을 하거나 폭력적으로 대하기도 했고요. 루다는(더 정확하게는 이루다를 만들어낸 스캐터랩) 이런 여러 문제를 일으키며 단 2주 만에 서비스를 잠정 중단하게 됐습니다.

AI 챗봇, AI 면접, AI 세탁기, AI 스피커, AI 배차까지. AI는 점점 우리 삶과 가까워지고 있습니다. 코로나19로 AI 면접을 적극적으로 활용하는 기업이 늘어났다고도 하죠. 기업과 구직자가 공통적으로 이야기하는 AI 면접의 장점은 단연 공정성이었습니다. 면접관의 주관적인 느낌, 그날의 컨디션, 선입견 등에 영향을 받지 않으리란 기대가 있었기 때문인데요. 우리는 이미 알고리즘과 기술의 외피를 두른 AI가 마치 인간보다 공정하고 정확한 것처럼 ‘착시’를 만들어내는 순간을 보아왔습니다. 2018년, 미국의 온라인 쇼핑몰 회사 아마존은 몇 년에 걸쳐 개발한 AI 채용 시스템이 여성 지원자를 차별한다는 사실을 확인하고 폐기하였습니다. 시스템이 학습한 데이터(이전 구직자 이력서)에 남성 이력서가 월등히 많았기 때문에 ‘자연스레’ 여성이라는 키워드가 포함된 이력서는 감점을 한 것이죠.

우리가 살아가는 이곳은 진공 상태가 아니며 사회 곳곳에 크고

작은 성차별이 존재합니다. 사회에 존재하는 차별에 대한 적극적인 인식 없이 과연 누구도 차별하지 않는, 성평등한 기술이 만들어질 수 있을까요? 당연히 그렇지 않습니다. 오히려 AI가 가진 불투명성, 확산성과 만나 기존에 존재하는 차별이 강화되거나 새로운 차별이 생겨날 수도 있습니다. 스캐터랩이 논란 이후 DB(데이터베이스)와 대화 모델을 폐기하는 모습을 보며 국내에서도 AI 윤리가 화두로 떠올랐습니다. 몇몇 기업은 자체 가이드라인을 만들어 자사 기술이 차별에 반대하고 있음을 대사회적으로 공표하기도 했는데요. 여러 가이드라인이 '인간중심 기술', '신뢰할 수 있는 기술', '다양성이 보장된 기술'이라는 방향성을 보여줍니다. 차별을 경계하며 더 많은 인간에 도움이 되는 기술을 만들겠다는 목소리는 소중합니다. 한국여성민우회는(이하 민우회) 이 목소리에 성평등을 더한, 기술을 비판적으로 읽어낼 수 있는 다양한 고민이 필요하다고 생각했습니다.

앞서 언급한 몇 가지 사례를 통해 AI 기술이 성차별적일 수 있다는 점은 확인했으나 구체적으로 어디에서 문제가 발생한 것인지 혹은 개선을 위해서 어떻게 개입해야 하는지 포착해내기는 쉽지 않았습니다. 우선 현장의 고민을 들어보고자 AI 업계 종사자와 연구자를 대상으로 설문을 받아보았는데요. '과연 답변이 들어올까?' 반신반의 했던 이 설문에서 여러 참여자가 굉장히 진지한 답변을 남겨주었습니다. 각자의 고민이 느껴지는 답변을 읽으며 좀 더 자세히 이야기가 듣고 싶어 계획에 없었던 [성평등한 AI를 고민하는 업계 종사자·연구자 집담회]를 진행하였습니다. 참여자의 공통적인 고민은 차별하지 않는 기술을 만들고자 고민하는 개발자가 턱없이 부족하고, 이를 고민하는 사람은 예민하고 불편한 사람이 되어버리는 업계의 분위기 속에서 함께 논의할 동료가 없다는 점이었습니다. 또, AI를 다루는 기업들이 앞다퉈 AI

윤리 가이드라인을 내놓고는 있지만, 차별하지 않는 AI를 만들기 위해 구체적으로 어떤 노력이 필요한지는 논의하지 않는 것도 AI 업계의 문제점이라고 지적했습니다. 이후 집담회를 통해 정리한 문제의식을 더 많은 페미니스트 시민과 나누고자 대중강의를 진행하였는데요. AI 윤리부터, 개인정보 보호와 정보인권, 페미니즘과 과학, 대중문화에서의 AI 재현까지 AI와 관련된 다양한 문제의식을 나누고자 했습니다. 매 강의에 400여 명이 넘는 인원이 신청하여 다들 새롭게 부상하고 있는 AI 기술을 페미니스트로서 비판적으로 읽어내고자 고민하고 있다는 걸 확인할 수 있었습니다.

본 소책자는 이 과정을 거치며 만났던 모든 페미니스트와 함께 만든 결과물입니다. 특히 [페미니스트가 함께 만드는 AI 가이드라인] 라운드 테이블에 함께해주신 열세 명의 페미니스트가 많은 고민을 나눠주었습니다. 두 차례 진행된 라운드 테이블에는 개발자, 연구자, AI 이용자까지 다양한 성격을 가진 참가자가 함께해주었습니다. 각자 다른 배경 속에서 자신의 경험을 바탕으로 어떤 AI 기술을 기대하는지, 기술이 성/평등하게 구현되기 위해서 무엇이 필요한지 적극적으로 토론해주신 참가자 분께 감사의 말을 전합니다. 다양한 참가자가 함께한 덕분에 같은 문제의식에서 출발했더라도 조금씩 방향이 다른 대안이 제시되기도 했습니다. 가이드라인에서 제안하는 내용이 '성평등한 AI'를 만들 수 있는 유일한 방법이라고 생각하지 않습니다. 같은 듯 다른 이야기 속에서 소책자를 읽는 모든 분께서 각자의 AI 윤리, AI 가이드라인을 고민할 수 있길 바랍니다.

활동을 진행하며 AI를 이용하며 겪은 차별·혐오 사례를 제보 받아보기도 했으나 큰 호응을 얻진 못했습니다. 다들 AI 기술의 '대상'이

되고 있음에도(유튜브 추천 영상, 상품 추천 등) 그것이 AI를 활용한 기술이라는 점을 알기는 쉽지 않기 때문은 아닐까 생각했습니다. AI가 일상에 가까워졌다고는 하지만 여전히 막연하게 느껴지기도 합니다. 기술은 전문가의 영역이고 이용자인 나는 이 기술에 질문조차 할 수 없는 존재인 것처럼 느껴질 때가 있습니다. 하지만 기술은 기술 그 자체로 존재하는 게 아니라 사회에서 어떻게 활용되는지에 따라 의미를 획득하게 됩니다. 그 의미를 만들어가는 중요한 당사자 중 하나는 바로 이용자입니다. 이용자가 기술을 적극적으로 탐색하고, 비판도 칭찬도 할 때 좋은 기술이 만들어질 수 있습니다. 이 가이드라인이 그러한 목소리를 낼 때 참고할 수 있는 자료가 되면 좋겠습니다. 이용자로부터 기술에 대한 더 많은 방향과 원칙이 나올 수 있길 바랍니다. 그래서 기술이 만드는 사람만의 것이 아니라 이용하는 사람의 것도 될 수 있길 바라봅니다.

AI는 그 어떤 기술보다도 파급력을 갖고 있습니다. AI로 인해 차별이 확산될 수 있다면 반대로 AI가 차별의 확산을 막는데 활용될 수도 있을 것입니다. 지난 2월, 〈그것이 알고싶다〉(SBS)는 딥페이크 기술을 활용한 불법 합성물 문제를 다루었습니다. 피해자 인터뷰에서 대역을 쓰거나 얼굴을 모자이크하는 대신 딥페이크로 생성한 가상의 얼굴을 합성하였는데요. 어떤 관점을 갖고 활용하는지에 따라 기술이 전혀 다른 방식으로 사용될 수 있다는 점을 알 수 있었습니다. 또, 이화여자대학교 학생으로 구성된 딥트(DEEP't)팀은 딥페이크를 탐지할 수 있는 시스템을 개발하며 "(...) 여성이 주로 피해자인 범죄 예방을 위해 도움을 줄 수 있는 시스템을 개발하게 되어 뜻깊었다"라는 인터뷰를 했습니다. 차별적이지 않은 기술, 폭력에 반대하는 기술은 사회를 변화시킬

수 있습니다.

활동을 통해 만난 여러 연구자, 개발자 분께서 "이런 고민을 하는 사람은 저 혼자라는 생각에 지쳤는데 같은 분야에서 비슷한 고민을 하는 페미니스트를 만나서 반가웠어요"라는 소감을 남겨주셨습니다. 어디서든, 모두가 당연하다고 말하는 것에 문제제기하고 목소리 내는 페미니스트로 살기는 쉽지 않은 일이구나, 그렇기에 우리는 더 자주 만나고 서로의 존재를 통해 힘을 얻어야 하는구나, 새삼 깨닫게 되었습니다. "AI는 성차별이 뭔지 알까?" 활동에 참여해주신 모든 페미니스트에게 다시 한 번 감사의 인사를 전합니다. 각자의 공간에서 따로 또 같이 목소리 낼 수 있길 바랍니다.

2021년 10월
한국여성민우회 성평등미디어팀
단호박, 보라, 윤소, 은사자

키워드별 분류

[라운드 테이블: 페미니스트가 함께 만드는 AI 가이드라인]을 2021년 9월 28일과 10월 5일 두 차례 진행했습니다. 라운드 테이블 참여자는 각자의 고민이 담긴 가이드라인을 제시하였고, 성평등미디어팀에서 그 내용을 모아 키워드로 분류해보았습니다. 키워드는 [인공지능 윤리원칙 분석 보고서: 하버드 법대 버크만 센터의 'Principled Artificial Intelligence'를 중심으로]에서 제시한 범주를 참고했습니다.

라운드 테이블 참여자
노엘, 다솔, 맑음, 모리, 모모, 민정,
수다, 수박씨, 시언, 쑤, 아침, 진기, 해원

키워드

#기본원칙	#데이터편향	#알고리즘편향
#차별금지	#개인정보보호	#책무성
#설명책임	#투명성	#기술통제권
#고위험AI	#교육	#영향평가
#정부와의회	#이용자의무	#시민참여

어떤 AI 기술을 개발할지 페미니즘 관점에서의 논의가 필요하다.

우리는 “어떤 AI 기술을 개발하고 싶은가?”가 아니라 “어떤 AI 기술을 원하는가?”를 고민해야 합니다. 사회에 어떤 기술이 필요한지를 기업과 개발자의 권한으로만 두는 것이 아니라, 사회 구성원이 어떤 기술이 필요한지 논의하고 개발을 요구해야 합니다.

개발자는 기술을 개발하는 전문가이지만, 기술의 사회적 영향력을 판단하는 전문가는 아닐 수 있습니다. 특히 남성의 비율이 높고 남성 중심적인 정서의 개발자 커뮤니티는 기술이 여성에게 미칠 영향력을 고려하지 않을 수 있습니다. 예를 들어, 얼굴이나 몸 등을 가린 모자이크를 해제하는 기술은 여성의 개인정보를 침해하거나 범죄로 이어질 수 있습니다. AI 챗봇이 인터넷 글을 무분별하게 학습해 여성혐오 발언을 쏟아내기도 합니다. 딥페이크를 이용해 포르노 속 여성의 얼굴을 여성 연예인 또는 일반인의 얼굴로 바꾸어 유포하는 범죄는 이미 일어나고 있습니다. 법을 제정하여 이 범죄를 처벌할 수 있게 되어도 더 ‘발전’된 기술을 활용해 새로운 범죄가 등장하기도 합니다. 그래서 무엇보다 중요한 것은 사회가 어떤 기술을 장려하고 어떤 기술을 규제해야 할지 페미니즘 관점에서 논의하는 것입니다.

AI 기술은 개인의 프라이버시와 친밀감을 침해하지 않는다.

디지털 환경은 개인의 활동을 기록하고 저장하는 속성을 가지고 있어 정보인권을 보장하기 위한 사회적 장치를 마련하는 것이 매우 중요합니다. 그러나 현실에서의 이용자 동의 방식은 매우 추상적이며 포괄적이어서 개인의 사생활을 수집하는 활동이 무차별적으로 이루어지고 있습니다. 디지털 기반 사회에서 프라이버시는 고도화된 맞춤형 서비스 개발보다 우선적으로 보장되어야 하는 시민의 권리입니다.

AI 기술로 인해 발생한/할 차별과 편향의 사회적 맥락을 고려한다.

AI는 데이터, 알고리즘, 개발주체 등 복합적인 요소를 가진 시스템일 뿐만 아니라, 생산-이용되는 과정에서 노동, 환경, 정치 등 여러 사회적 관계 안에 놓이게 됩니다. 그렇기에 AI 기술로 인해 발생한/할 문제는 정부의 정책, 기업의 기술로만 해결할 수 있는 것이 아닙니다. 우리는 AI의 생산과 이용이 여러 사람과 집단에게 복합적인 영향을 미칠 수 있다는 걸 인식하고, 문제 발생의 맥락을 고려하여 다양한 관점과 해결 방식을 반영할 수 있어야 합니다.

#기본원칙

AI로 인한 수익을 사회적으로 환원하는 것이 필요하다.

AI 시스템은 시민들의 데이터를 기반으로 구성된 사회적 결과물입니다. 그럼에도 불구하고 데이터에 대한 개별적 보상은 전혀 이루어지지 않습니다. 개발자와 기업은 이러한 공공의 기여를 사회적으로 환원하여야 하고, 당연한 책무로 인식할 필요가 있습니다. 또한 AI로 인한 부정적 영향을 완화하기 위해 리터러시 교육 등 공익적 활동에도 적극적으로 기여해야 합니다.

AI 기술은 인간의 자율성을 존중하고 공동체의 미래와 더 나은 삶을 위해 기여한다.

기술은 그 자체로 중립적이지만 이를 어떤 목적에 어떻게 사용하는가에 따라 전혀 다른 결과를 가져올 수 있습니다. AI 개발은 인간의 자율성과 지속가능한 미래를 저해하지 않는 방향으로 개발되고 활용되어야 합니다. 살상 무기에 AI 기술을 접목시키는 것, 젠더폭력의 도구로 관련 기술을 활용하는 것 등 인간에게 해악을 끼치는 방식은 엄격하게 금지되어야 합니다.

#데이터편향 #알고리즘편향

데이터 편향을 줄이고 차별적이지 않은 데이터를 수집한다. 전문가 집단의 검수 등 데이터 편향을 줄이기 위한 시스템을 마련한다.

가장 뛰어난 자연어처리 모델이라고 평가되는 GPT-3는 인터넷의 모든 텍스트 데이터로 언어를 학습했다고 합니다. 여러 AI 분야에서 더 높은 성능을 달성하기 위해 더 많은 데이터를 수집하고자 합니다. 인간이 하나하나 데이터를 작성하거나 수집하는 것은 굉장히 비효율적이므로 데이터셋의 크기가 커지면 대부분 인터넷 속의 데이터를 무분별하게 크롤링*하게 됩니다. 어떤 데이터는 검수 과정을 거치기도 하지만 그렇다고 인터넷의 모든 텍스트를 검수할 수는 없습니다. 게다가 데이터의 차별성을 감지하는 능력은 사람마다 다릅니다.

예를 들어 가상으로 여성 얼굴 사진을 생성하는 모델의 학습 데이터로 1만 장의 여성 얼굴 사진을 사용했다고 합시다. 여기서 9,000장이 화장을 하고 있고 9,500장이 머리가 길다면 모델은 대부분 화장을 하고 머리가 긴 여성의 이미지를 생성할 것입니다. 이 데이터셋을 '여성은 머리가 길고, 화장을 하는 사람'이라고 생각하는 사람이 검수한다면, 이것이 편향되었다고 판별하지 않을 것입니다. 그런데 페미니즘 관점에서 보았을 때 이 데이터셋은 고정관념에 기반한 여성의 이미지만을 다수 포함하고 있기에 과도하게 편향되어 있습니다. 그러므로 데이터 양보다는 질에 집중해 데이터셋 내에서 차별과 편향을 줄이기 위해 노력해야 합니다.

---

*크롤링(crawling): 무수히 많은 컴퓨터에 분산 저장되어 있는 문서를 수집하여 검색 대상의 색인으로 포함시키는 기술. (출처: IT용어사전)

#데이터편향 #알고리즘편향

AI 추천 시스템은 이용자에게 차별적인 결과를 제시하지 않는다.

포털, SNS, 쇼핑, 스트리밍, 게임, 하물며 와인 평가 애플리케이션까지 대부분 서비스에 추천 시스템이 존재합니다. 추천 시스템을 구축하기 위해 수집되는 개인 정보 문제나 맞춤 광고 문제 역시 중요한 쟁점이지만, 추천 시스템이 이용자를 분류하는 행위 그 자체도 살펴보아야 합니다. 추천 시스템은 이용자 활동을 바탕으로 이용자를 분류하고, 여러 콘텐츠 중 비슷하게 분류된 이용자가 선호한 콘텐츠를 골라 우선적으로 노출합니다. 그런데 추천 시스템은 이용자에게 편향된 콘텐츠만 추천할 수도 있고, 이용자의 데이터를 맥락적으로 분석하지 못하는 등의 문제가 발생할 수 있습니다. 예를 들어, 구글 광고가 고임금의 구인 광고를 남성 이용자에게 주로 노출시켜 여성의 일자리 접근 기회를 줄인다는 연구 결과가 발표되기도 했습니다. 더 나아가서 우리의 행동으로 우리의 정체성을 예측하려는 추천 시스템은 그 자체로 고정관념을 고착화하는 것입니다. 이러한 추천 시스템의 문제를 보완할 수 있는 방법이 고민되어야 합니다.

#데이터편향 #알고리즘편향

성별, 인종, 성정체성, 성적지향, 장애에 관한 자료 수집이 필요한 경우 최대한 다양한 현실을 반영할 수 있는 선택지를 마련한다.

우리가 흔히 온라인 쇼핑몰 등에서 입력하는 개인정보가 알고리즘을 만들기 위한 데이터로 이용되기도 합니다. 이때 불필요한 개인정보가 포함되지 않도록 개인정보보호 규칙을 따르며, 개발자가 개인정보를 확인하는 과정에서도 유출이 일어나지 않도록 개인 식별 정보를 제거하는 등의 규제가 필요합니다. 만일 성별 등의 자료 수집이 필요한 경우, 최대한 다양한 선택지를 포함하여 특정한 집단을 제외하지 않도록 하고, 다양한 현실을 반영할 수 있도록 해야 합니다.

번역모델에서 젠더 편향을 반영하지 않도록 각별히 주의한다.

번역모델이 텍스트 데이터를 학습할 때 ▲원본 데이터에 시대, 장르, 작가 등의 태그를 추가하여 텍스트의 맥락을 파악할 수 있도록 하거나, ▲민감할 수 있는 표현, 단어를 찾은 후 해당 부분 학습 파라미터를 덜 민감하게 조정하는 등의 시도가 데이터 편향을 줄여나가는 방안이 될 수 있습니다.

#데이터편향 #알고리즘편향

AI 의료 기술은 인종, 성별, 국적 등 다양한 데이터를 수집해 편향되지 않은 결과를 도출한다.

의료 분야에 AI 기술을 적용했을 때 여성 환자의 생존률이 더 낮다고 합니다. 대부분의 의료 데이터가 백인 남성을 위주로 형성되어 백인 남성과 다르게 증상이 나타나는 흑인 남성, 백인 여성, 동양인 여성 등의 질병 유무를 제대로 판별하지 못했기 때문입니다. AI 기술은 모두에게 평등해야 합니다.

[관련 기사]
해외 연구진이 의료 인공지능(AI)의 편향성을 경고했다. 기존 의료 데이터를 이용할 경우 인종을 차별하는 결과가 나올 수 있다는 것이다. 19일 미국 IT매체 벤처비트에 따르면, 미국 서던캘리포니아대학교 연구진과 영국 퀸매리대학교 연구진은 의료 AI의 인종·성 차별 문제를 확인했다는 논문을 각각 발표했다. 두 연구진은 의료 AI의 편향성 원인으로 AI 학습을 위한 데이터 세트의 편향성을 꼽았다.
— 해외 연구진, 의료 AI 편향성 경고, 2021년 2월 19일, IT조선

#데이터편향 #알고리즘편향

성차별 언어나 이미지를 데이터 학습 단계에서 걸러낸다.

호주 퀸즐랜드공과대학교 연구진은 트위터 속 트윗 문맥을 분석해 여성혐오 콘텐츠를 감지하는 AI 알고리즘을 개발했습니다. 이 알고리즘 시스템은 트윗의 문맥과 의도에 따라 여성혐오 유무를 판단해 분류합니다. 예를 들어, '부엌으로 돌아가라'는 아무런 문제가 없는 문장이지만, 여성을 향해 발화된 것이라면 가사노동을 여성의 영역으로 국한하는 의미가 담겨 있기에 여성혐오적입니다. 퀸즐랜드공과대학교 연구진이 만든 AI 알고리즘은 이 문장을 여성혐오 표현으로 구분해냄으로써 문맥 학습의 효과를 확인했다고 합니다. 이와 같은 AI 데이터 학습 알고리즘이 활발하게 개발되면 AI로 인한 성차별을 획기적으로 줄여나갈 수 있을 것입니다.

참고 기사
"김치녀 · 된장녀"…여혐 표현 AI로 걸러낸다, AI타임즈, 2020년 8월 31일

#차별금지

AI에 영향을 미치는 권력 구조를 인식하고, 부당한 권력 구조에 맞선다.

카카오 알고리즘 윤리헌장은(“알고리즘 결과에서 의도적인 사회적 차별이 일어나지 않도록 경계한다”) 차별을 특정 개인의 적극적 의도가 있어야 발생하는 것으로 좁게 정의하고 있습니다. 그러나 차별은 악의나 실수에 의해서만 생기는 것이 아니며, 권력과 연관된 구조적인 문제입니다. 따라서 차별을 해소하는 일 또한 개인의 선한 의지로만 해결되지 않고, 차별을 유발한 부당한 권력 구조에 대항할 때 가능해집니다. ‘과거 데이터를 기반으로 한 의사결정’이라는 기계학습의 기본 접근방식이 기존 권력 구조를 재생산/강화할 수 있음을 항상 경계해야 합니다. 또한 골상학, 성별주의 등 특정 집단에 부당한 불이익을 주는 환원주의적 접근을 배제해야 합니다.

#차별금지

AI 알고리즘에 의하여 의도적 차별뿐만 아니라 비의도적 차별 또한 일어나지 않도록 한다.

의도적으로 차별이 발생하기도 하지만, 많은 경우 의도하지 않았음에도 차별이 발생합니다. 예를 들어, IT업계 개발자 대다수가 남성이기 때문에 여성 개발자는 ‘여자분이신데 개발을 하신다니 멋지네요!’와 같은 말을 듣습니다. 이는 칭찬하고자 하는 선량한 의도에서 한 말일 수도 있겠지만 ‘개발은 남성의 일’이라는 고정관념에 기반한 차별 발언입니다. 이러한 차별은 이미 존재하는 사회문화적 구조와 편견을 바탕으로 발생하기에 적극적인 고민이 없다면 인식하기 어렵고, 해결하기도 어렵습니다.

예를 들어, 2018년 온라인 쇼핑몰 아마존은 자사가 개발한 AI 채용 시스템이 성차별성을 내포하고 있음을 인지한 후 폐기한 바 있습니다. ‘의도적인’ 차별을 배제하는 것뿐만 아니라, 더욱 더 선제적으로 차별을 예방할 수 있는 조치가 필요합니다.

[관련 기사]
세계 최대 전자상거래 기업인 ‘아마존’이 인공지능을 이용한 ‘채용 프로그램’을 개발하다 여성 차별 문제가 드러나, 자체적으로 폐기했습니다. 남성 직원이 더 많은 기업의 인적 데이터를 학습한 AI가 남성에 대해서 더 우호적인 판단을 내린 것입니다.
— ‘AI 채용’은 남성을 선호해?…성차별 논란에 ‘자체 폐기’, JTBC, 2018년 10월 12일

AI 대화 기술은 성차별에 기반한 응답을 하지 않는다.

KT에서 출시한 AI 스피커 기가지니는 “너는 남자야? 여자야?”라는 질문에는 “저는 아리따운 여자랍니다.”, “넌 어떤 색깔 좋아해?”라고 물었을 때는 “사랑스럽고 블링블링한 핑크색을 좋아하지요.”, “너는 자동차 좋아하니?”에는 “아니요. 제가 여자라서 그런지 자동차에 관심이 없어서.”라고 답을 했습니다. 2019년, 기가지니의 답변이 성차별적이라는 이용자의 문제제기에 KT는 답변을 수정했습니다. 대화형 AI가 차별적인 대화를 할 수 없도록 적극적인 조치가 필요합니다.

#차별금지

AI 정체성은 성역할 고정관념, 성별 이분법에 기반하지 않으며, 소수자와 약자의 정체성을 도구화·상품화하지 않는다.

흔히 ‘비서’ 기능을 하도록 기획된 AI는 여성의 목소리가, 고도로 지능화된 ‘컴퓨터’ 역할을 하는 AI는 남성의 목소리가 사용됩니다. 대화 상대로서의 챗봇은 이루다처럼 20대 여성으로 캐릭터가 설정되기도 합니다. 이는 성별 이분법과 성역할 고정관념에 따른 것으로, AI의 캐릭터를 결정하는 단계에서 약자의 정체성이 도구화, 상품화되지 않도록 주의해야 합니다. 더불어 미디어가 AI를 재현할 때, 인간의 형상을 한 형태로 재현되는 경우가 많습니다. 따라서 앞서 언급한 정체성의 상품화, 도구화를 막기 위해 미디어가 재현하는 AI의 모습에 대해 비판과 성찰이 필요하다고 생각합니다.

AI는 기존의 인간중심적, 이분법적 사고 틀 내에서 재현됩니다. 그렇기 때문에 스테레오타입의 고착화, 불평등의 재생산을 막기 위해서는 AI를 포스트 휴머니즘의 시각으로 바라보고, ‘퀴어’한 존재로 재-상상하는 적극적인 작업이 필요합니다. AI를 일방적으로 인간을 편리하게 하는 기술이라고 볼 수도 있겠지만, AI는 이미 인간과 영향을 주고받는 존재이기 때문에 우리가 “인간적이고 그렇기 때문에 당연한 것”이라고 생각했던 것을 근본적으로 다시 상상할 수 있어야 한다는 점을 꾸준히 상기시켜야 할 것입니다.

#차별금지

AI를 인간화하여 표현하는 방식(단어 사용, 음성, 외형 구성)에 있어 인종, 종교, 장애, 성정체성, 성적지향, 사상, 정치 성향에 대한 사회적 편견을 반영하지 않는다.

AI 에이전트*가 '여성'으로 식별되는 경우, 여성의 어투나 외형을 과장하여 표현하는 경우가 있으며 이는 또다시 여성에 대한 사회적 편견을 고착화하는 방식으로 순환됩니다. 성별이 없는 존재에 대해 성별의 '특성'을 부여하는 경우, 그 목적과 표현방식에 차별은 없는지, 표현방식이 사회적 편견을 강화시키는 것은 아닌지 고민해야 합니다.

---

*AI 에이전트: AI 에이전트는 미리 결정된 목표를 달성하기 위해 작동하는 코드 또는 메커니즘입니다. AI 에이전트의 예로는 채팅 봇, 스마트 홈, 재무에 사용되는 프로그래밍 방식 거래 소프트웨어 등의 코드에서 찾을 수 있습니다. (출처: 마이크로소프트)

#차별금지

특정 성별에게 강요되는 고정관념을 재현하지 않는다.

현재 각종 AI 스피커 등에서 여성의 목소리로 수행되는 과잉 친절함이 꼭 필요하다고 생각하지 않습니다. 여성에게 사회문화적으로 요구되어 온 친절한 목소리나 행동을 AI가 재현하지 않도록 하고, 성별이 특정되지 않는 목소리와 행동을 기본값으로 해야 합니다.

[관련 기사]
왜 음성비서는 여자 목소리일까. 기술은 사회를 반영하는 방식으로 설계된다. 오랫동안 사회에서 비서, 고객 응대 서비스, 도우미의 역할은 주로 여성이 맡아왔다. 음성 기계 처리, 로봇 등의 서비스를 설계하는 사람들은 대부분 남성이다.
— [ESC] 왜 음성비서는 여성 목소리일까, 한겨레, 2019년 4월 10일

#차별금지

AI 음성인식 기술은 목소리 높낮이를 기준으로 성별을 단정하지 않는다.

여성의 목소리 톤이 유달리 높은 것은 생물학적 차이로 볼 수 없습니다. 그보다는 여성의 높은 목소리 톤을 바람직하고 매력적인 것으로 간주하는 사회의 힘이 강력하기 때문입니다. 일례로 호주 연구팀이 1940년대와 1990년대 젊은 여성의 목소리를 비교 분석한 결과, 눈에 띌 만큼 여성의 목소리 톤이 낮아진 것으로 나타났는데, 이는 여성의 사회적 지위가 높아진 것과 관련 있습니다. 생물학적 특성이라고 간주되는 목소리조차도 사회적 맥락에 따라 구성되고 변할 수 있습니다. 그렇기 때문에 AI는 목소리로 화자의 성별을 단정해서는 안 됩니다.

#차별금지

AI 챗봇이 이용자의 차별·혐오발언을 감지하여 경고하는 시스템을 구축한다. 또한 시스템이 제대로 작동할 수 있도록 차별·혐오발언의 기준을 세운다.

네이버에는 악플을 감지하는 'AI클린봇'이 있습니다. 'AI클린봇'은 욕설뿐 아니라 모욕적인 표현이 담긴 댓글까지 AI 기술로 감지하여 자동으로 숨겨줍니다. 그런데 2020 도쿄올림픽에서 한국과 터키의 배구경기가 진행될 때 네이버 실시간 중계 댓글에서 터키 선수를 비하하는 댓글이 걸러지지 않았습니다. 네이버는 해당 표현을 비하로 보아야할지 사람마다 기준이 다르며, 표현의 자유가 지켜져야 한다는 해명을 하여 논란이 됐습니다.

이용자의 차별·혐오발언을 감지하고 블라인드 처리하는 기술은 그리 어렵지 않은 기술이라고 생각합니다. 이루다를 향한 이용자의 성희롱 발언을 걸러내는 기술도 불가능하지는 않았을 것입니다. 그러나 앞선 네이버 사례에서 알 수 있듯 차별·혐오발언의 기준을 세우는 것이 우선되어야 합니다. 기준에 대한 심도 있는 논의 속에서 시스템이 가동되어야 합니다.

#개인정보보호

정보 수집 대상자에게 어떤 목적으로 해당 데이터를 사용할지 고지하고, 원하지 않는 정보의 경우 제공하지 않을 수 있는 권리를 부여한다.

챗봇 이루다는 같은 회사에서 개발한 '연애의 과학'에서 쌓인 데이터를 기반으로 만들어졌습니다. '연애의 과학'은 연인 혹은 호감 있는 사람과의 대화를 제공하면 대화 패턴을 분석해 애정도를 분석해주는 앱이었습니다. 이루다는 이를 기반으로 무려 100억 건의 대화를 학습했다고 밝힌 바 있습니다. 하지만 '연애의 과학' 이용자는 자신의 대화가 챗봇 개발에 쓰인다는 점을 고지 받은 적이 없고(단순히 '신규 서비스 개발에 활용될 수 있다' 정도의 안내), 또 이용자(대화를 제공한 사람)와 대화한 상대방은 이러한 과정에 대한 안내를 전혀 받지 못했습니다. 기업은 이용자의 정보를 활용할 때 각 정보 주체에게 데이터의 활용 가능성을 최대한 상세히 안내하고, 제공을 원치 않는 정보의 경우 제외할 수 있도록 옵션을 제공해야 합니다.

#개인정보보호

초상권, 주소와 같은 개인정보가 노출되지 않도록 하고, 피해가 발생한 경우 AI 개발 주체는 이를 보상한다.

챗봇 이루다는 주소를 물어보았을 때, 실제 주소를 답변하여 논란이 되었습니다. 데이터로 활용된 사람의 개인정보가 노출된 것입니다. 또한, 한국 여성 가수의 얼굴을 포르노에 합성한 딥페이크 영상이 온라인에 유통되어 여성 가수들이 초상권 침해 피해를 입고 있다고 합니다. 네덜란드 사이버 보안업체인 '딥트레이스(Deeptrace)' 연구 결과에 따르면 발견된 딥페이크 영상의 96%가 '음란물'이었고 이 피해자 중 25%가 한국 여성 연예인이라고 합니다. 개인의 초상권을 비롯한 개인정보가 침해되지 않도록 법적인 장치를 마련하고, 피해가 발생할 경우 이에 대한 피해 보상을 해야 합니다. 또, 정보 제공자는 본인이 원하지 않는 정보의 제공을 거부할 권리가 있습니다. 정보 제공자의 정보삭제권은 반드시 법적으로 보장되어야 합니다.

#책무성

개발 주체는 언제나 AI에 대한 책임을 가진다.

개발 주체는 AI에 의해 발생한 피해 및 결과에 항상 사회적 책임을 지고 수정하거나 보완하여야 합니다. 특히 젠더 편향적인 AI에 대해서 책임감을 가지고 기술을 제어하고 대안을 제시하여 사회적 책임을 다해야 합니다.

AI 개발 및 이용 과정에서 발생하는 차별로 특정 집단에 불이익, 위협을 가하지 않으며 발생하는 결과에 대해서는 법적·사회적 책임을 진다.

자율 주행 운전의 경우 사고가 발생했을 때의 책임 주체에 관해 많은 논의가 진행되고 있습니다. 그러나 동시에 자율 주행으로 운전하는 경우에도 반드시 운전자가 탑승하여 운전에 바로 개입할 수 있도록 계속 경고를 주는 방식을 채택하고 있으며, 이는 책임 주체를 명확하게 하기 위함입니다. 다시 말해 어디까지나 AI가 운전자의 주행을 돕는 보조 역할임을 강조하는 것입니다. AI 개발 과정에서 발생하는 차별에 대해서도 법적·사회적 책임 주체를 명확히 하고 발생한 차별에 대해서는 구제, 배상 방안이 마련되어야 합니다.

#책무성

알고리즘의 설계부터 도출되는 결과에 이르기까지 차별, 불공정 등의 문제가 없는지 윤리심의기구가 관리 · 감독하고 그 결과를 공개한다.

유럽연합(EU)은 지난 4월 AI 규제법안(Artificial Intelligence Act)을 발표했습니다. EU는 '유럽인공지능위원회'를 신설하고 컨트롤타워 역할을 하게 되고, 회원국 국가감독기관장과 유럽 데이터 보호 감독기구가 위원으로 참석해 법안 규정 마련과 시행을 지원한다고 합니다. 한국도 정부와 기업으로부터 독립된 심의기구를 설치하여 AI 전반에 걸친 차별, 불공정 등을 관리 감독하고, 그 결과를 대중에 공개하여야 합니다.

AI 개발에 소요되는 환경적·사회적 자원을 고려하고 사회적 책임을 다한다.

AI 개발 과정에서는 보이지 않는 많은 자원이 소비되고 있습니다. 구글의 AI 엔지니어였던 팀닛 게브루는 대규모 언어처리 AI 모델이 엄청난 전력 소모를 유발해 지구온난화에 영향을 미친다고 이야기한 바 있습니다. 과도한 자원 낭비는 궁극적으로 사회적 약자에게 가장 큰 영향을 끼칩니다. 그렇기에 AI 개발 과정에서도 바로 눈앞에 보이지 않는 자원의 사용을 항상 고려하고 사회적 책임을 다할 필요가 있습니다.

#책무성

AI 개발 과정에 관여한 모든 노동자의 권리를 보장한다.

AI를 개발하는데 있어 꼭 필요한 것 중 하나는 데이터입니다. 가공되지 않은 원본 데이터 중 필요한 데이터를 선별하고 걸러내는 작업을 데이터 라벨링(data labeling)이라고 합니다. 예를 들어, 비행기와 새가 함께 보이는 이미지에서 새를 찾아내 박스로 표시하거나, 이미지 속 사물을 텍스트로 기록하는 등의 업무가 있습니다. 우리는 AI가 '자동'으로 여러 이미지 중 새를 선택하고, 자동차가 무엇인지를 정확하게 짚어낼 수 있다고 생각하지만 이러한 시스템을 구축해내기까지 수많은 인간의 노동이 존재합니다. 데이터 라벨링, 콘텐츠 검수 등은 굉장히 중요한 작업임에도 불구하고 노동자는 비가시화된 채로 노동을 수행합니다. AI를 개발한 기업은 엄청난 수익을 창출하지만, 노동자는 저임금 노동을 반복하게 됩니다. 기업은 이러한 노동 문제에 관심을 기울이고, 적극적으로 해결방안을 모색해야 합니다.

#다양성

다양성에 대한 공감을 AI 기술 개발의 기본 가치로 한다.

AI로 인한 차별이 일어나지 않기 위해서는 기술 윤리를 가져야 하며, 이 기술 윤리는 다양성과 성평등을 반드시 포함하여야 합니다. 0과 1의 이분법적 사고에서 완전히 탈피하여 기술 개발의 장애물을 유연하게 받아들이고 편향되지 않은 가치를 존중하여야 합니다.

AI에서 다양성의 가치를 고려한다는 것이 무엇을 의미하는지 구체화한다. 어떤 차별, 혐오를 포함하는 것인지 다양성의 세부적인 개념이 필요하다.

네이버 'AI 윤리 준칙'은 "네이버는 다양성의 가치를 고려하여 AI가 이용자를 포함한 모든 사람에게 부당한 차별을 하지 않도록 개발하고 이용하겠습니다.", 다음카카오 'AI 윤리'는 "카카오는 다양한 가치가 공존하는 사회를 지향합니다."라고 선언하고 있습니다. 그런데 이 선언에서 의미하는 다양성의 가치, 다양한 가치가 무엇인지 명확히 알기 어렵습니다. 기업에서 추구하는 다양성이 무엇인지 기업은 이용자에게 구체적으로 설명해야 합니다. 다양성은 고정된 가치가 아닙니다. 기업은 이용자와의 소통을 통해 지속적으로 다양성의 개념을 갱신하고 확장해야 합니다.

AI의 발화는 평등과 다양성을 포괄한다.

AI 시스템에 사용될 대화문을 작성하고 분류하는 일을 한 적이 있습니다. 대화문은 주제에 따라 여성, 남성 또는 '기타'로 반드시 분류되어야 했습니다. 고정관념에 기반하여 대화문을 분류하는 경우가 많았고, 연애/결혼 관련 대화는 여성으로, 직장 생활 관련 대화는 남성으로 분류되었습니다. 또 어린이 대화의 경우 어린이가 아닌 사람이 작성했기 때문에 대화의 주제가 학교나 가족으로 한정되고, 어린이를 미숙한 존재로 여기는 태도가 반영되었습니다.

이처럼 AI의 대화는 '일반적' 특징을 대변할 것이 요구되기 때문에 작성자와 검수자 모두 고정관념에 매몰될 수 있습니다. AI는 마치 탈-가치관, 객관성의 영역이라고 상상되지만 실제로는 성별 이분법적 분류와 고정 관념을 재생산하는데 끊임없이 동원됩니다. AI 시스템에서 사용된 대화문은 편향되지 않은 데이터로 보기 어렵고, 사회의 불평등을 그대로 반영한 것입니다. 따라서 AI 시스템의 대화문은 평등과 다양성의 가치를 계속 반영해나가야 합니다.

AI 스피커를 비롯해 음성 서비스를 제공하는 AI의 경우 여성/남성/성별을 구분할 수 없는 목소리를 모두 제공한다.

내비게이션 음성, AI 비서 등 서비스를 제공하거나 안내, 보조하는 음성은 대부분 여성 목소리가 기본값입니다. '이용자가 여성의 음성을 선호하기 때문'이라고 하지만, 컴퓨터 교육을 담당하는 AI 서비스는 모두가 남성 목소리를 더 좋아하는 것으로 나타났습니다. 이는 '선호도' 또한 성역할 고정관념에 기반한다는 점을 의미합니다. 굳이 여성/남성 음성을 나누어야 한다면 이용자에게 초기 설정 시 선택할 수 있도록 하고, 성별을 구분할 수 없는 목소리도 적극적으로 도입되어야 합니다.

[관련 기사]
인공지능이나 로봇 등 목소리를 통해 구현하는 서비스가 증가하면서 일부 사회단체들을 중심으로 목소리 선호도에 대한 논란이 벌어졌고 최근에 남성도 여성도 아닌 제3의 목소리가 개발됐다. 네덜란드의 버추 노르딕(Virtue Nordic)이라는 광고 회사는 사람들의 실제 목소리를 샘플링해 "Q"라는 새로운 중성의 목소리를 만들어냈다고 밝혔다.
— 너는 아니? 중성목소리…AI가 여성목소리인 이유, KBS, 2019년 5월1일

#다양성

▞ 다양한 정체성을 가진 AI를 만든다.

AI에 부여된 캐릭터는 대부분 '젊은' 남성 혹은 여성이고, '비서'의 역할을 하는 것으로 설정된 경우가 많은데, 이보다 더 다양한 정체성이 부여된 AI 에이전트가 있어야 합니다. 아이 혹은 노인일 수도 있고, 다양한 직종의 사람을 만들어낼 수도 있을 것 같습니다. 한 번에 하나의 캐릭터와 대화를 하는 것이 아니라, 대화 속에서 다양한 존재가 등장하는 방식이 될 수도 있겠죠? 이러한 시도는 일상에서 마주하는 수많은 사람과 그와 맺는 관계에 대해 생각해볼 수 있는 계기가 될 것이라고 생각합니다.

#설명책임 #투명성

▞ 투명성이 보장되는 AI를 만든다.

AI가 의사결정을 하는 방식, 통찰력 및 정보를 취득하는 과정을 투명하게 공개해야 합니다. 차별적인 결과가 도출되었을 때는 기술을 이용하는 조직이나 이용자가 요구할 때 그 결정을 내린 과정을 설명할 수 있어야 합니다.

▞ 사람이 의사결정을 할 때 요구되는 설명과 같이 AI 시스템이 무엇을, 어떤 이유로 하고 있는지 명확하고 완전하며, 이해하기 쉬운 설명을 제공한다.

AI 설계자의 편향이 알고리즘에 반영될 수 있고, 의도적/비의도적으로 차별적인 결과를 도출할 가능성이 있습니다. 그렇기 때문에 AI 시스템에 대한 구체적인 설명이 필요하고, 이는 기술을 잘 모르는 사람도 이해하기 쉬운 형태여야 합니다.

AI 이용자의 이해를 돕기 위해 기술을 설명하고 이를 주기적으로 업데이트한다. 이용자 피드백을 수집하여 기술을 이해하고 있는지 확인하고, 이해가 어려운 부분은 개선하여 다양한 이용자가 이해하기 쉬운 AI를 만든다.

이용자가 AI를 쉽게 이해할 수 있도록 설명하기 위해 어떠한 노력을 하고 있는지 명시가 필요합니다. 또한, 설명을 하고 있다면 이용자가 제대로 이해하고 있는지 확인하는 노력도 필요합니다. 이를 통해 이용자 의견을 수집하고 반영하여, 이용자 눈높이에 맞춰 AI 기술을 설명하고 투명하게 공개하는 것이 필요합니다.

설명 가능한 AI를 만들기 위해 다양한 기술을 개발하고 도입한다.

AI 알고리즘을 인간이 설명할 수 있는, 해석 가능한 방법으로 접근할 수 있게 하는 대리모형(surrogate model) 등을 제안합니다. 대리모형이란 AI 모델과 결과를 유사하게 내면서도 인간이 이해할 수 있게 원리가 밝혀져 있는 모델을 말합니다. 대리모형은 설명가능한 인공지능(XAI) 방법론 중 하나로, 이를 통해 대략적으로나마 본래 학습에 이용한 AI의 판단을 유추할 수 있습니다. 금융권에서는 대출심사 등에서 설명가능한 인공지능(XAI) 기법을 시범적으로 이용하고 있습니다. 이를 통해 기존의 금융 서비스에 적용되었던 편향을 찾아내고, 보완하는 것이 가능해질 것입니다.

인간의 일상에 관여하는 AI는 학습한 데이터의 출처와 데이터 유형(사진, 목소리, 언어 등), 데이터에 담긴 인구학적 요소(성별, 나이 등) 등을 공개한다.

미국에서 안면인식 AI의 오류로 무고한 흑인이 범인으로 체포되는 일이 반복되고 있습니다. 이는 AI가 흑인의 안면을 제대로 인식하지 못하여 발생한 일로, AI로 인해 인종차별이 심화되었다는 것을 보여줍니다. 안면인식, 음성인식 등 인간의 일상에 관여하는 AI 기술은 어떠한 데이터를 기반으로 하고 있는지를 상세하게 공개해야 합니다.

AI를 활용한 교육을 실시할 경우 이를 명확하게 고지한다. 또, 교육 대상자에게 AI 교육이 작동되는 방식을 안내한다.

인간 교사와 AI 교사의 다름을 충분히 이해할 수 없는 이용자에게는 반드시 AI에 대한 교육을 실시해야 합니다.

[관련 기사]
교원그룹이 유·초등 대상 인공지능(AI) 선생님을 선보인다. 16일 교원그룹은 실사형 'AI튜터'를 도입한다고 밝혔다. AI튜터는 실사형 인공지능 기술을 적용해 영상 합성으로 제작한 가상 교사다. 학습 몰입감을 위해 캐릭터가 아닌 실제 인물을 AI튜터로 구현했다. 이 기술은 교원그룹이 4분기 출시를 앞둔 에듀테크 학습 프로그램에 적용될 예정이다.
— "실제와 99% 유사"…인공지능 선생님 나온다, 서울경제, 2021년 9월 16일

▞ 인간의 의사결정(예. 채용, 판결, 투표 등)에 AI 기술을 보조적으로 활용하는 것은 가능하지만, 인간의 의사결정을 AI로 완전히 대체해서는 안 된다.

인간의 의사결정은 검토 여지가 있지만, AI의 의사결정은 왜 그러한 결과가 나왔는지 불투명합니다. 그렇기에 AI 기술에 의사결정을 맡긴 후에는 이를 번복할 가능성이 낮아집니다. AI 의사결정을 검토할 수 있는 단계를 두는 방식도 있겠지만, 인간이 잘못된 의사결정을 하는 경우가 있더라도 인간의 판단에 맡겨야 하는 영역은 지켜져야 합니다.

▞ AI를 사용한 결과를 노출할 경우 이를 밝히고, 사용하지 않은 결과값을 노출하는 옵션도 이용자에게 제공한다.

검색, 쇼핑 등 큐레이션이 필요한 서비스나 SNS 타임라인을 노출하는 방식 등 AI는 수많은 정보를 선별하기 위해 사용되고 있습니다. AI의 선별 기준은 ▲서비스 제공자가 임의로 정한 기준(예. 포털의 스폰서 광고 우선 노출), ▲데이터 학습에 따른 결과(예. 클릭수가 높은 기사) 등을 따르기 쉬운데, 이용자는 이 결과를 비판적으로 수용하기 어려울 수 있습니다. 여성, 어머니, 부인, 여학생 등 특정 키워드를 검색했을 때 첫 페이지에 나오는 이미지는 관련도가 높은 순이라 표시가 되는데, 관련도란 무엇일까요? 길거리를 검색했을 때 왜 여성의 뒷모습 사진이 첫 페이지에 표시 될까요? 이용자는 AI 알고리즘을 알권리가 있고, 그 기준에 따른 결과가 옳지 않다면 정정을 요청할 수 있어야 합니다.

▞ 동의 없이 실존하는 인물을 그대로 모방하는 서비스를 제공해서는 안 된다.

딥페이크는 AI 딥러닝 기술을 이용해 가짜 영상이나 목소리, 사진을 만들어내는 기술입니다. 딥페이크를 활용한 정치인, 연예인 등의 음성을 듣거나, 영상을 본 적이 있으실 것입니다. 최근에는 이러한 기술을 저렴한 가격으로 손쉽게 이용할 수 있는 서비스를 제공하는 스타트업도 생겨났습니다. 이러한 서비스는 반드시 당사자의 동의를 기반으로 합법적인 영역에서 이용되어야 합니다.

그러나 딥페이크는 여성 연예인의 얼굴, 일반인 여성의 얼굴을 포르노 영상에 합성하는 범죄에 이용되는 등의 문제를 야기하고 있습니다. 유명인의 영상으로 '가짜 뉴스'를 생산하여 유포하는 것 또한 심각한 문제로 지적됩니다. 그렇기 때문에 딥페이크 기술을 활용할 때에는 당사자의 동의가 반드시 필요하며, 동의 없이 활용하여 발생한 피해를 막을 수 있는 강력한 규제가 필요합니다.

#기술통제권 #고위험AI

AI 면접의 결과만으로 채용을 결정해서는 안 된다.

2020년 말 국회입법조사처의 보고서에 따르면 민간 기관 339곳, 공공기관 93곳이 채용절차에서 AI 면접을 활용하고 있다고 합니다. 잡코리아의 2020년 조사에 따르면 구직자의 48.1%가 'AI 면접이 공정하다고 생각한다'고 응답했습니다.(대면면접 28.3%) 하지만 AI의 공정성은 공정하지 않은 사회에서는 도달하기 어려운 목표입니다. 채용 성차별이 있는 사회의 AI 채용 시스템은 여성을 뽑지 않는다는 것이 아마존의 사례에서 확인된 바 있습니다. 또한 최근 한 정부 기관에서 AI 면접 진행 중 프로그램 오류가 발생하여 응시자가 면접에서 불합격한 사례가 있었습니다. 이 기관은 공정성이 훼손된다며 재응시 기회를 주지 않았고, 이후 감사원의 감사가 진행되기도 하였으나 추가 대응은 없었다고 합니다. 이처럼 문제제기조차 하기 어려운 AI 면접의 결과가 채용을 결정하는 중요한 요소가 되어서는 안됩니다.

참고 기사

'AI 면접'의 확산…사회적 신뢰 얻기 위해 해야할 것, SBS, 2021년 2월 25일

'AI면접' vs '대면면접' 공정성 비교, 전북일보, 2020년 9월 7일

#교육

AI를 개발, 운영하는 과정에 평등과 다양성의 가치가 반영되도록 개발 주체에게 충분한 교육을 제공한다.

AI 시스템 자체는 하나의 컴퓨터 프로그램처럼 운영되더라도 말 그대로 '인공' 지능이기 때문에 이를 개발하고 관리하는 인간의 가치관이 반영될 수밖에 없습니다. 충분한 교육을 제공하여 평등한 가치관이 반영될 수 있도록 하고, 더불어 다양한 배경의 인력이 AI 개발과 운영에 참여할 수 있어야 합니다. 여기에서 나아가 STEM* 분야에 더 다양한 이들이 진출할 수 있어야 하고, 특히 주요 결정을 내리는 과정에 이들의 의견이 반영될 수 있는 환경이 마련되어야 합니다.

---

*STEM: Science, Technology, Engineering, Mathematics

#교육

AI 설계자는 성별, 인종, 장애에 따른 차별 방지 교육을 받는다.

AI는 뛰어난 기술이지만 인간이 만들어 내는 기술이라는 것을 인지해야 합니다. 차별적이지 않고, 편향되지 않은 학습 데이터가 제공되기 위해서는 담당자 교육이 무엇보다도 중요합니다. '중립'이라는 모호한 기준 대신 인권 중심, 성평등 가치가 반영된 가이드라인을 작성하고 AI 설계자 및 관리자가 숙지할 수 있도록 해야 합니다.

[관련 기사]
인공지능에 의한 차별과 편견을 경계해야 하는 것은 인공지능이 알고리즘과 데이터를 기반으로 작동하는 기술 구조 때문이다. 알고리즘은 개발자의 사고방식과 편견을 반영하고, 학습 대상 데이터의 속성과 한계를 드러내기 때문이다. 인공지능 개발자들의 인구통계적 속성상 백인, 남자, 고소득층, 영어 사용자 위주의 편향이 만들어진다는 지적이 있다.
— 인공지능, 몰랐던 '영화 속 배역 성차별'도 콕 찍어낸다, 한겨레, 2017년 3월 20일

#영향평가

AI의 성평등 정도를 평가하는 지표를 정립한다.

성평등을 정량적으로 측정할 수 있는 지표를 만들어 적용해야 합니다. 헐리웃 영화 내 성별 간 대사 지분, 출연 시간, 대화의 질 등의 차이를 지적했다는 '지나 데이비스(GD) 포용 지수'*와 같이, 성평등에 저해되는 표현이 얼마나 나왔는지 빈도기반으로 체크하는 지표를 생각해봤습니다. 텍스트와 음성 데이터(Speech-to-Text)의 경우, 빈도기반(frequency based)이나 의미기반(semantic based) 지표를 정립할 수 있습니다.

[관련 기사]
*지나 데이비스 포용 지수: 지나 데이비스는 2007년 '지나 데이비스 미디어젠더연구소'를 설립하고, 영화와 드라마에서 남녀 배우의 출연 시간과 대사 분량, 역할과 대화의 중요도 등을 분석해 이를 지수화 하는 작업을 해왔다.
— 인공지능, 몰랐던 '영화 속 배역 성차별'도 콕 찍어낸다, 한겨레, 2017년 3월 20일

AI 서비스를 대중에게 공개하기 전에 외부 윤리심의기구를 통해 성인지감수성, 인권감수성을 기반으로 한 심의를 받는다.

AI 기술은 궁극적으로 인간의 번영과 공익에 기여해야 합니다. 만들어진 AI 기술이 누군가를 차별하는 등 인간에게 피해를 입힐 가능성이 있다면 수정·보완되어야 합니다.

#영향평가

시민참여형 AI 기술 검수시스템을 마련한다. 다양성이 확보되는 알고리즘을 설계하기 위해 다양한 주체가 참여하는 'AI 알고리즘 편향 검증 기구'를 설치한다.

많은 기업이 사회 전반을 포용할 수 있는 AI를 개발하겠다고 말합니다. 단순히 선언에 그치지 않고 이를 위해 어떠한 노력을 하고 있는지가 구체적으로 명시될 때 이용자가 기술을 신뢰할 수 있습니다. 해당 AI가 차별적 정보를 재생산할 우려가 있는지 전문가뿐만 아니라 이용자 또한 검수에 함께할 수 있는 시스템을 만들어야 합니다. 이 검증 기구에는 성별·연령·지역 등의 특성을 고려해 다양한 이용자가 참여해야 하며, 이들이 이해하고 검증하기 쉽도록 설명이 제공되어야 합니다. 기업은 이를 통해 AI가 사회적 소수자에게 미칠 영향을 파악하고, 다양성이 확보되는 알고리즘을 설계해야 합니다. 또 이러한 개발 과정을 모든 이용자가 확인할 수 있도록 공개할 것을 요구합니다.

#정부와의회

AI 기술로 인한 차별을 금지할 수 있도록 포괄적 차별금지법을 제정한다.

유럽의 경우 차별의 범주가 법적으로 명시되어 있기 때문에 법에 기초해서 데이터 관리나 알고리즘으로 발생한 차별을 문제제기할 수 있습니다. 반면 한국은 차별에 대한 기준조차 확보하고 있지 못한 사회입니다. AI를 매개로 해서 조금 더 강력하게 차별금지법 제정의 필요성을 목소리 높일 수 있는 공간이 펼쳐지길 바랍니다.

개인정보 침해나 데이터 독점을 막을 수 있는 데이터다양성위원회와 같은 사회적 규제 기구를 설치한다.

데이터는 전적으로 사적인 소유가 될 수 없고, 독점적으로 관리되어서는 안 된다고 생각합니다. 기업을 규제하는 가이드라인을 통해 데이터 독점을 막아내기에는 강제력이 부족합니다. 이에 거대 테크 기업과 온라인 플랫폼의 개인정보 침해, 데이터 독점을 막을 수 있는 정부 차원의 데이터다양성위원회와 같은 사회적 규제 기구가 필요합니다. 데이터다양성위원회는 데이터의 다양성을 판단할 수 있는 객관적 기준을 세우고, AI 훈련 데이터셋이나 알고리즘이 차별을 재생산하지 않도록 하고, 이용자가 AI를 이해할 수 있도록 기업에게 설명책임을 부여하는 등의 역할을 수행해야 합니다.

#정부와의회

▞ 정부 기금으로 AI 서비스를 개발할 경우, 심사 과정에서 AI의 사회적 영향력을 예측한 보고서를 반드시 제출한다.

정부가 발표한 디지털 뉴딜 계획에 따르면 데이터·네트워크·인공지능 고도화에 9조 9천 억원이 투입됩니다. 정부 기금으로 AI 기술을 개발할 경우, 해당 기술의 사회적 영향력을 예측·평가한 보고서를 반드시 제출하도록 해야 합니다. 정부 예산으로 차별을 재생산하는 AI 기술이 개발되는 것을 막아내기 위해서 이러한 조치는 꼭 필요합니다. 또한 해당 기술이 어떤 윤리적 기준을 가지고 개발이 되었는지도 투명하게 공개하여야 합니다. 이러한 절차가 점차 민간 기업에까지 확대될 수 있어야 합니다.

▞ 정부기관에서 지원하는 AI 사업의 인적 구성은 특정 성별이 60%를 넘지 않는다.

현재 AI는 많은 국가 예산이 투입되는 분야 중 하나입니다. 이러한 분야는 성별영향평가를 통해 성평등 수준을 측정하는데, 이 중 구성원의 성비는 중요한 지표입니다. IT 업계는 종사자 대부분이 남성인 직군입니다. 성비 불균형과 성인지감수성 부족은 AI로 인한 성차별이 발생하는 이유로 지적되기도 합니다. 따라서 AI 분야의 정부 사업비 지원 조건 중 하나로 구성원의 성비가 균형적이어야 한다는 항목을 포함하면 AI의 성차별성을 낮출 수 있는 기반이 될 것입니다.

#이용자의무 #시민참여

▞ AI에 관한 문제제기와 해결에 있어 다양한 주체가 참여한다.

AI 개발과 이용에 관여된 여러 사람과 집단이 함께 문제 제기하고 해결에 참여할 수 있어야 합니다. 기술 생산 주체는 민주적인 의사소통을 위한 채널을 열고, 시민사회는 기술 리터러시에 입각해 기술을 비판적으로 바라보고 논의할 수 있어야 합니다.

▞ AI 기술 이용자가 지켜야할 가이드라인을 마련한다.

AI 스피커의 보급률이 급격하게 증가하고 있고, 어린이도 손쉽게 이 기기를 사용합니다. 성별과 연령을 막론하고, 대부분의 사람이 AI 스피커를 이용할 때 명령조로 이야기를 합니다. AI 스피커에게 전달하려는 내용을 정확하게 인식시키기 위해 사용되는 말투로 생각되기는 하지만, 기술을 대하는 방식이 인간을 대하는 방식에 끼치게 될 영향이 우려됐습니다. 예를 들어, 명령조로 AI 스피커를 이용하는 방식이 서비스직 노동자를 대하는 태도에 반영될 수도 있겠다는 생각을 하게 되었습니다. 대부분의 AI 가이드라인이 개발하는 사람에 초점이 맞춰져 있다면 AI 이용자가 지켜야할 가이드라인도 함께 고민되길 바랍니다.

#기본원칙

☑ 어떤 AI 기술을 개발할지 페미니즘 관점에서의 논의가 필요하다.

☑ AI 기술은 개인의 프라이버시와 친밀감을 침해하지 않는다.

☑ AI 기술로 인해 발생한/할 차별과 편향의 사회적 맥락을 고려한다.

☑ AI로 인한 수익을 사회적으로 환원하는 것이 필요하다.

☑ AI 기술은 인간의 자율성을 존중하고 공동체의 미래와 더 나은 삶을 위해 기여한다.

#데이터편향 #알고리즘편향

☑ 데이터 편향을 줄이고 차별적이지 않은 데이터를 수집한다. 전문가 집단의 검수 등 데이터 편향을 줄이기 위한 시스템이 마련한다.

☑ AI 추천 시스템은 이용자에게 차별적인 결과를 제시하지 않는다.

☑ 성별, 인종, 성정체성, 성적지향, 장애에 관한 자료 수집이 필요한 경우 최대한 다양한 현실을 반영할 수 있는 선택지를 마련한다.

☑ 번역모델에서 젠더 편향을 반영하지 않도록 각별히 주의한다.

☑ AI 의료 기술은 인종, 성별, 국적 등 다양한 데이터를 수집해 편향되지 않은 결과를 도출한다.

☑ 성차별 언어나 이미지를 데이터 학습 단계에서 걸러낸다.

#차별금지

☑ AI에 영향을 미치는 권력 구조를 인식하고, 부당한 권력 구조에 맞선다.

☑ AI 알고리즘에 의하여 의도적 차별뿐만 아니라 비의도적 차별 또한 일어나지 않도록 한다.

☑ AI 대화 기술은 성차별에 기반한 응답을 하지 않는다.

☑ AI 정체성은 성역할 고정관념, 성별 이분법에 기반하지 않으며, 소수자와 약자의 정체성을 도구화·상품화하지 않는다.

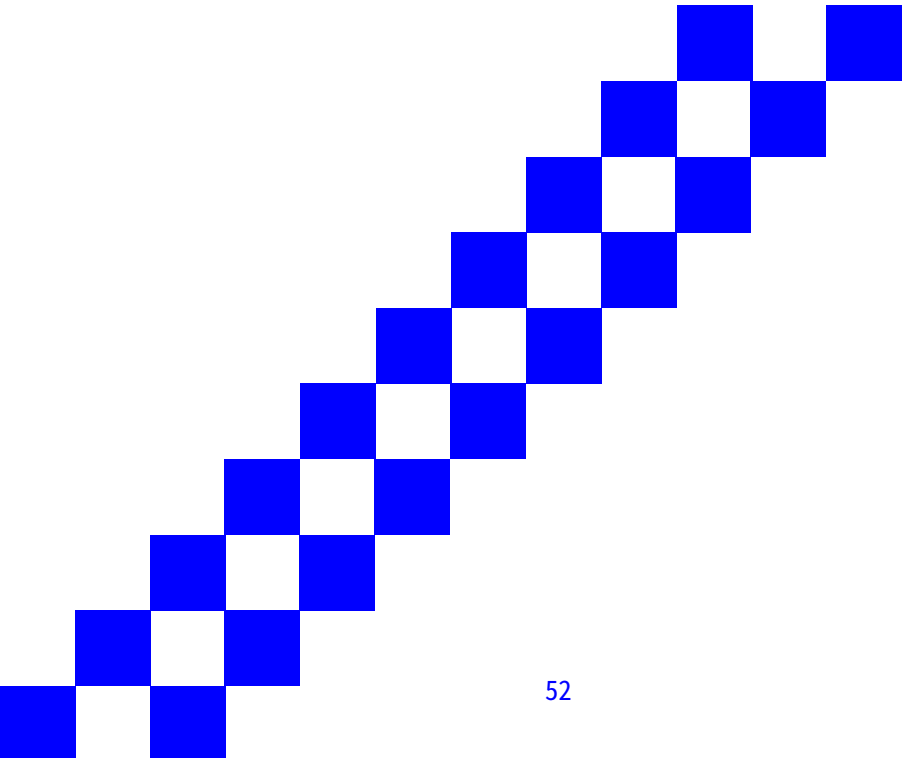
☑ AI를 인간화하여 표현하는 방식(단어 사용, 음성, 외형 구성)에 있어 인종, 종교, 장애, 성정체성, 성적지향, 사상, 정치 성향에 대한 사회적 편견을 반영하지 않는다.

☑ 특정 성별에게 강요되는 고정관념을 재현하지 않는다.

☑ AI 음성인식 기술은 목소리 높낮이를 기준으로 성별을 단정하지 않는다.

#차별금지

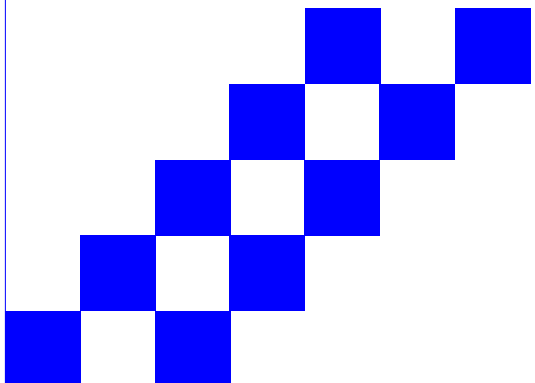
☑ AI 챗봇이 이용자의 차별·혐오발언을 감지하여 경고하는 시스템을 구축한다. 또한 시스템이 제대로 작동할 수 있도록 차별·혐오발언의 기준을 세운다.



#개인정보보호

☑ 정보 수집 대상자에게 어떤 목적으로 해당 데이터를 사용할지 고지하고, 원하지 않는 정보의 경우 제공하지 않을 수 있는 권리를 부여한다.

☑ 초상권, 주소와 같은 개인정보가 노출되지 않도록 하고, 피해가 발생한 경우 AI 개발 주체는 이를 보상한다.



#책무성

☑ 개발 주체는 언제나 AI에 대한 책임을 가진다.

☑ AI 개발 및 이용 과정에서 발생하는 차별로 특정 집단에 불이익, 위협을 가하지 않으며 발생하는 결과에 대해서는 법적·사회적 책임을 진다.

☑ 알고리즘의 설계부터 도출되는 결과에 이르기까지 차별, 불공정 등의 문제가 없는지 윤리심의기구가 관리 · 감독하고 그 결과를 공개한다.

☑ AI 개발에 소요되는 환경적·사회적 자원을 고려하고 사회적 책임을 다한다.

☑ AI 개발 과정에 관여한 모든 노동자의 권리를 보장한다.

#다양성

☑ 다양성에 대한 공감을 AI 기술 개발의 기본 가치로 한다.

☑ AI에서 다양성의 가치를 고려한다는 것이 무엇을 의미하는지 구체화 한다. 어떤 차별, 혐오를 포함하는 것인지 다양성의 세부적인 개념이 필요하다.

☑ AI의 발화는 평등과 다양성을 포괄한다.

☑ AI 스피커를 비롯해 음성 서비스를 제공하는 AI의 경우 여성/남성/성별을 구분할 수 없는 목소리를 모두 제공한다.

☑ 다양한 정체성을 가진 AI를 만든다.

#설명책임 #투명성

☑ 투명성이 보장되는 AI를 만든다.

☑ 사람이 의사결정을 할 때 요구되는 설명과 같이 AI 시스템이 무엇을, 어떤 이유로 하고 있는지 명확하고 완전하며, 이해하기 쉬운 설명을 제공한다.

☑ AI 이용자의 이해를 돕기 위해 기술을 설명하고 이를 주기적으로 업데이트한다. 이용자 피드백을 수집하여 기술을 이해하고 있는지 확인하고, 이해가 어려운 부분은 개선하여 다양한 이용자가 이해하기 쉬운 AI를 만든다.

☑ 설명 가능한 AI를 만들기 위해 다양한 기술을 개발하고 도입한다.

☑ 인간의 일상에 관여하는 AI는 학습한 데이터의 출처와 데이터 유형(사진, 목소리, 언어 등), 데이터에 담긴 인구학적 요소(성별, 나이 등) 등을 공개한다.

☑ AI를 활용한 교육을 실시할 경우 이를 명확하게 고지한다. 또, 교육 대상자에게 AI 교육이 작동되는 방식을 안내한다.

#기술통제권 #고위험AI

☑ 인간의 의사결정(예. 채용, 판결, 투표 등)에 AI 기술을 보조적으로 활용하는 것은 가능하지만, 인간의 의사결정을 AI로 완전히 대체해서는 안 된다.

☑ AI를 사용한 결과를 노출할 경우 이를 밝히고, 사용하지 않은 결과 값을 노출하는 옵션도 이용자에게 제공한다.

☑ 동의 없이 실존하는 인물을 그대로 모방하는 서비스를 제공해서는 안 된다.

☑ AI 면접의 결과만으로 채용을 결정해서는 안 된다.

#교육

☑ AI를 개발, 운영하는 과정에 평등과 다양성의 가치가 반영되도록 개발 주체에게 충분한 교육을 제공한다.

☑ AI 설계자는 성별, 인종, 장애에 따른 차별 방지 교육을 받는다.

#영향평가

☑ AI의 성평등 정도를 평가하는 지표를 정립한다.

☑ AI 서비스를 대중에게 공개하기 전에 외부 윤리심의기구를 통해 성인지감수성, 인권감수성을 기반으로 한 심의를 받는다.

☑ 시민참여형 AI 기술 검수시스템을 마련한다. 다양성이 확보되는 알고리즘을 설계하기 위해 다양한 주체가 참여하는 ‘AI 알고리즘 편향 검증 기구’를 설치한다.

#정부와의회

☑ AI 기술로 인한 차별을 금지할 수 있도록 포괄적 차별금지법을 제정한다.

☑ 개인정보 침해나 데이터 독점을 막을 수 있는 데이터다양성위원회와 같은 사회적 규제 기구를 설치한다.

☑ 정부 기금으로 AI 서비스를 개발할 경우, 심사 과정에서 AI의 사회적 영향력을 예측한 보고서를 반드시 제출한다.

☑ 정부기관에서 지원하는 AI 사업의 인적 구성은 특정 성별이 60%를 넘지 않는다.

#이용자의무 #시민참여

☑ AI에 관한 문제제기와 해결에 있어 다양한 주체가 참여한다.

☑ AI 기술 이용자가 지켜야할 가이드라인을 마련한다.

가이드라인은 변화의 '시작'이다.

조경숙
테크-페미 활동가, IT개발자

처음 IT 회사에 입사했을 때만 해도, 나는 페미니스트 기술자가 되겠다고 굳게 마음먹고 있었다. 기술 연수를 받을 때, 나는 개발 언어의 개념을 하나하나 노트에 꼼꼼하게 정리하고 프로그램을 짜며 나중에 현업에서 어떤 서비스를 맡게 될지 설레했다. 그러나 정작 현업에 투입되고서는 오류를 잡는 데에 급급해 서비스 개선에 크게 관심을 두지 못했다. 당시 내가 담당하던 건 사내시스템으로, 인사를 관리할 수 있는 인사 시스템이나 회계/재무 정보를 연동해 보여주는 회계 시스템 등 다양한 서비스로 구성되어 있었다. 그때 내 관심사는 오로지 버그를 잡는 것뿐이었다. 복잡하게 오가는 요청(Request)과 응답(Response) 사이, 어디에 문제가 있는지 정확히 파악하는 것만이 내 주된 관심사였다.

이런 태도가 바뀌게 된 건 사내시스템을 벗어나 대외서비스를 담당하게 되면서부터다. 사용자 평균 연령대가 30대였던 사내시스템에 비해 대외서비스는 연령대가 10대부터 70대까지 다양했다. 웹사이트가 제대로 작동하는 것뿐만 아니라, 웹사이트에 게시된 콘텐츠가 문제되지 않을지 검수하는 일도 중요했다. 일단 대외 서비스는 웹사이트의 폰트 사이즈부터 다르다. 사내시스템의 기본 폰트가 12~14pt라면 대외 서비스의 기본 폰트는 16~18pt다. 사용하는 연령대가 다르므로, 누구든 편히 볼 수 있도록 글씨 크기를 키운 것이다. 누구나 편리하게 사용할 수 있는 서비스를 개발하는 건, 서비스의 사소한 영역에서부터 찬찬히 살피는 일이다. 말 그대로 SM(System Management)인 셈이다.

다양한 사용자가 평등하게 이용할 수 있는 서비스를 만드는 건 정말 어려운 일이다. 많은 사용자의 테스트를 거치며 어떤 방식의 서비스를 구현해야 누구나 편리하게 사용할 수 있을지 연구와 사례조사, 개발과 테스트 등을 거쳐야 한다. 그렇게 만들어진 것 중 하나가 바로 '웹 접근성'이다. 웹 접근성 가이드라인은 웹사이트가 색약 사용자를

위해 명도 대비 기준을 맞추고, 시각장애인을 위해 이미지 대체 텍스트를 제공하며, 키보드만으로 버튼 등에 접근할 수 있도록 안내하고 있다. 선구적으로 가이드라인을 만든 이들이 있기 때문에, 현재는 많은 공공기관과 기업에서 웹 접근성을 준수해 웹사이트를 개발한다. 이 가이드라인이 존재함으로써, 보다 많은 사람이 풍성하게 웹에 접근할 수 있게 되었다.

웹 이후로도 새로운 서비스는 계속해서 출시된다. 인스타그램, 페이스북, 틱톡과 같은 SNS에 이어 유튜브, AI에 이르기까지 신규 서비스는 굉장히 다양하고 빠르게 등장한다. 그리고 이러한 새로운 서비스에는 새로운 위험성이 뒤따른다. 인스타그램이 장소 태그를 붙이면서 과연 사이버 스토킹의 위험성을 생각했을까? 페이스북이 '함께 아는 친구'를 보여줄 때 프라이버시 침해에 대해 떠올렸을까? 나아가 AI의 기계학습에 우리의 개인정보가 몽땅 빨려 들어갈 줄 대체 누가 알았을까? 모든 서비스가 이런 사건들에 처음부터 사전 대응할 수는 없었겠지만, 사례가 보고되었다면 이후로는 비슷한 사건이 재발하지 않도록 '어떻게든' 조치해야 한다.

십대여성인권센터 IT지원단 'Women Do IT팀'은 오랜 스터디 기간을 거쳐 '디지털 성폭력 방지를 위한 서비스 개발자 가이드 깨톡(teen-it.kr)'을 제작했다. 개별 기능의 특성과 위험성을 가능한 구체적으로 분석하고, 기능별로 대안을 적시한 가이드다. 1차 가이드를 만든 뒤에는 IT업계 종사자를 대상으로 설문을 돌려 회신을 받았다. 이후 설문 결과를 수집·분석하여 가이드에 반영하고 다시 2차 완성본을 내는 과정을 거쳤다. 이렇게 만들어진 가이드라인은 다시 IT업계 안에서 회자되었고, 일부 기업에서는 실제로 적용을 검토하는 등 실무에서 논의가 되었다고 한다. 가이드라인이 매뉴얼로서 작동하면 좋겠지만, 현장마다 상황이

다르니 가이드라인이 적합하지 않을 수 있다. 그러나 가이드라인이 서비스와 관련해 개발진이 나눠야 할 논의의 물꼬를 터주었다면 그것 자체로도 유의미한 결과라고 볼 수 있다. 우리가 개발한 가이드는 웹/앱 서비스를 대상으로 한다. 그러나 최근 이슈가 되고 있는 AI는 기존의 웹/앱 서비스와는 또 다른 위험성을 갖고 있다. AI는 사용자의 데이터를 대량으로 학습하는 만큼, 개인정보의 수집과 활용·서비스 개발·BM(Business Model)의 영역까지 꼼꼼하게 논의될 필요가 있다.

가이드라인은 함께 만드는 것이다. 웹 접근성 가이드도 처음엔 W3C*가 내놓았지만, 국내에도 웹 접근성을 연구하는 기관이 생기는 등 다양한 곳에서 지식을 보태며 내용이 점차 풍성해졌다. 서비스를 사용하는 이와 만드는 이가 가이드라인을 통해 함께 대화를 나누어 간다면, IT 업계 안에 유의미한 변화가 일어날 것이다. 가이드라인은 우리가 마주할 변화의 시작이다.

---

* W3C(World Wide Web Consortium): 웹의 장기적 성장을 보장하기 위해 개방형 표준을 개발하는 국제 커뮤니티이다. (출처: W3C 홈페이지)

AI는 페미니스트의 친구가 될 수 있을까?

모리
민우회 회원

나는 일상에서 많은 부분을 AI에 기대고 있다. 하루에 적어도 30분은 쳐다보는 유튜브에선 내가 좋아할 만한 콘텐츠를 알아서 추천해준다. 회의록을 작성할 땐 구글문서나 클로바노트의 음성 기록 서비스(*음성을 텍스트로 자동 변환시켜준다)를 활용해 시간을 줄인다. 구글포토에 사진을 백업하고 검색어나 얼굴인식을 통해 원하는 사진을 찾는다. 그 외에도 교통 상황에 따라 빠른 길을 알려주는 내비게이션이나 내가 살 만한 물건을 상단에 띄워주는 쇼핑몰도 모두 AI 기반의 시스템이다. 내가 훨씬 많은 시간을 들여야 할 일을, 그들은 몇 초도 걸리지 않고 해결해준다. 아마 이 편리함 이전으로 돌아갈 순 없을 것이다.

물론 그만큼 섬뜩한 순간도 있다. “영어 공부나 할까?”란 말을 내뱉은 날부터 페이스북이나 인스타그램에 전화 영어 광고가 노출되었고, ‘애플워치’를 검색한 이후 며칠 동안은 애플워치 악세사리 광고를 봐야만 했다. 그럴 땐 내가 AI를 이용하는 게 아니라 AI가 나를 자신의 활동을 위한 실험대상으로 삼는 것만 같았다. 내가 넣은 검색어가 곧 나를 말하는 시대다. 그 데이터가 나를 구성하고, 내가 어떤 사람인지 정의한다. 자동차 관련 용품을 검색하면 AI는 나를 남성으로 인식하고 남성 용품을 함께 추천하곤 했다. 성별 고정관념은 이미 ‘후진 것’이 되었는데, 여전히 다수에게 통하기에 ‘빅데이터’로 살아남은 것이다.

‘섬뜩함’이 좀 더 가까워졌던 순간이 있다. 회사 건물을 출입할 때면 QR코드 인증을 하는데 “인증 되었습니다”라는 건조한 말투가 어느 날 ‘솔톤’의 간지러운 여성 목소리로 변해있었다. 전자출입명부가 확인되었다는 메시지에서조차 ‘대접’ 받고 싶어하는 사람이 있기에 누군가 이 목소리를 바꾸었을 것이다. 과장된 친절함, 웃음을 머금은 목소리. 기계의 재현을 결국 인간 여성인 나에게도 기대할 거란 생각이 들었다. 이미 일상 깊이 들어온 AI가 내 삶을 해치지 않으려면 무엇을 어떻게

해야 하는지 고민하고 싶었다. 그래서 민우회에서 준비한 '페미니스트가 함께 만드는 AI 가이드라인' 라운드 테이블에 참여했다. 권력을 가진 이들이 데이터를 독점하는 것, 소수자를 배제하는 기술이 등장하는 것을 어떻게 막을 수 있을지 걱정했는데 함께 이야기를 나누면서 조금은 안심되었다. 이렇게 고민하고 실천하려는 사람이 있다면, 모든 인간에게 도움되는 기술이 만들어질 수 있을 것 같았다.

모임을 마치고 할머니 생각이 났다. 아흔 살이 다 된 할머니가 요즘 가장 아끼고 좋아하는 대상은 '아리(SK인공지능 스피커)'다. 눈도 잘 보이지 않고, 귀도 잘 들리지 않는 할머니에게 아리는 TV 드라마 시청 시간을 알려주고, 볼륨을 적절히 조정해주며 짜증도 내지 않는 최고의 비서다. 할머니 혼자 누운 잠자리에서 책을 읽어주고, 아침 컨디션을 물어봐 주기도 한다. 아리가 페미니스트라면 어떨까? 여성 작가의 좋은 책을 골라 읽어주고, 드라마 속 성차별적 표현에 대해 의견을 남기며 할머니와 대화할 수 있다면? 나이 든 여성을 함부로 대하는 무례한 사람과 대신 싸워줄 수 있다면? 상상만으로도 즐거웠다.

모임에서 나누었던 이야기를 떠올려 보면 '페미니스트 AI'가 영 허무맹랑한 이야긴 아닐 것이다. 시간은 좀 걸리더라도 페미니스트 AI를 필요로 하는 사람이 생겨날 거고, 수요가 생기면 그에 맞는 상품이 개발될 수 있지 않을까? 물론 그 전에 페미니스트 시각에서 가이드라인을 만들고, 실무자에 대한 교육을 강화하는 과정이 필요하겠지만 말이다. AI를 만드는 것도 인간, AI의 딥러닝을 돕는 데이터를 만드는 것도 인간이다. AI는 멀리서 뚝 떨어지는 게 아니라 우리가 사는 사회를 반영할 수밖에 없다. 나는 AI의 배움목록에 페미니스트의 언어를 하나라도 더 추가하기 위해 오늘도 열심히 온라인에 글을 남기려 한다.

언젠가 "요즘 세상에 그런 말 하면 큰일나요."라고 나 대신 말해 줄 인공지능 비서를 기대하며.

활동을 마치며

AI 가이드라인을 진행하면서 기술에 대한 확신보다 의심이 더 많이 생겼어요. AI가 편향될 수 있고, 불공정할 수 있고, AI를 설계하는 업체는 무분별하게 많은 개인정보를 가지고 있으면서 개인정보 처리에는 소홀한 점 등을 확인할 수 있었어요. 기술이 발전하면서 기술에 더욱 많은 정보를 저장하고 의지해 왔었는데요. 기술에 경각심을 가지게 된 계기였어요. 기술이 좀 더 윤리적으로 발전되기를, 누구도 불편하지 않게 서비스를 이용할 수 있기를 바라는 마음으로 AI 가이드라인 작업을 참여했습니다. 덕분에 개인정보동의 약관을 꼼꼼하게 읽는 습관이 생기기도 했고, 개인정보보호에 대한 중요성을 느끼게 되었습니다. 많은 분들을 만나볼 수 있는 기회였고, 각계의 개발자와 시민 분들을 만나 같이 문제의식을 공유할 수 있어 좋았습니다. 이 글을 보는 시민분들도 저희와 함께 고민해보셨으면 좋겠습니다. — 단호박

한국여성민우회 성평등미디어팀이 [AI는 성차별이 뭔지 알까?] 활동을 시작한 지 벌써 일 년이 다 되어가네요. AI 가이드라인을 만들기로 했는데, 과연 누가 볼까? 강제력이 없는데 어떤 의미일까? 의문을 갖기도 했었습니다. 하지만 페미니스트 공학도, 엔지니어를 만나면서 인공지능과 차별의 문제를 말하는 판을 만들고 자료를 생산해내는 의미를 알게 되었습니다. 기술이 우리 삶에 미칠 영향력을 고민하는 페미니스트 시민들이 각자의 자리에서 자신이 무엇을 할 수 있을지 고민하고 있다는 것도 알게 되었습니다. 세상을 바꾸는 것의 시작은 나를 바꾸는 것이라던데, 저도 많이 배우며 바뀐 시간이었습니다. AI는 성차별이 뭔지 모르더라도 사람은 성차별이 뭔지 알아야 합니다! — 보라

2019년 지금은 세상을 떠난 윤정주 활동가의 제보로 한국여성민우회 성평등미디어팀의 AI 대응 활동이 시작되었습니다. AI 스피커의 성차별성을 문제제기하는 활동이 이 소책자를 만드는 활동으로 이어져 온 것인데요. 방송의 문제점을 지적하는 활동을 주되게 해왔기에 AI라는 낯선 기술을 우리가 비판적으로 분석하는 활동을 할 수 있을지 걱정이 많이 됐습니다. 그런데 올 한해 AI를 파헤치면 파헤칠수록 AI 기술을 잘 모를 수 있어도, AI가 나아가야 할 방향을 제시하는 것은 페미니스트가 해야 한다는 확신을 가지게 되었습니다. 기술은 눈부신 속도로 발전하지만 거기에 제동을 거는 것은 페미니스트가 안 하면 누가 할까...는 자의식 과잉이겠지만 정말로 그런 생각이 들었습니다. AI 이야기하는 자리에 누가 올까라는 걱정과는 달리 많은 페미니스트의 참여로 이 활동을 마무리하게 되어 기쁩니다. 앞으로도 AI에 대해, 지금은 상상할 수 없는 신기술에 대해 페미니즘 관점으로 씹고 뜯고 맛보고 즐기는 활동을 함께 할 수 있길 바랍니다! — 윤소

[AI는 성차별이 뭔지 알까?]라는 제목을 보면서 아니 근데(?) 나는 AI가 뭔지 알고 있나? 잘 모르는데 AI한테(??) 저런 걸 물어봐도 되나? 생각했던 적이 있습니다. 활동을 하면서 저뿐만 아니라 많은 사람이 AI를 낯설고 어려운, 어떤 순간엔 좀 무서워하기도 한다고 느껴졌는데요. (당연히) AI 전에도 인류는 많은 기술을 만들어 내고 활용해왔더라고요. 기술이 우리의 삶을 더 나은 방향으로 나아가게 할 것이라는 낙관을 놓지 않으며, 더 많은 페미니스트가 기술을 개발하고, 감시하고, 변화시키는 데 목소리 내면 좋겠습니다. — 은사자

참고자료

AI와 차별이라는 키워드를 더욱 깊이 이해할 수 있는 다큐멘터리와 책을 소개합니다.

알고리즘의 편견: Coded Bias

감독: 샬리니 칸타야

MIT 미디어 랩의 연구원 조이 부올람위니가 안면 인식 소프트웨어에서 이상한 점을 발견한다. 왜 흑인 여성의 얼굴은 제대로 인식하지 못하는 것일까. 조이의 경험을 바탕으로 발전하고 있는 알고리즘 기술에 숨겨진 인종, 성차별에 대한 문제점들을 파헤치고 있는 여러 여성 과학자들, 활동가를 카메라에 담는다. 이들은 기술이 얼마나 은밀하게 차별을 공고히 하는지 밝혀낸다. 디지털 기술은 가시적이었던 차별을 은밀하게 코드에 숨겨 확산하고 있음을 확인할 수 있다.

대량살상 수학무기

작가: 캐시오닐 | 옮긴이: 김정혜 | 출판사: 흐름출판

민주주의를 위협하는, 아무도 알려주지 않았던 빅데이터 이야기. 불평등을 확산하고, 민주주의를 위협하는 WMD(Weapons of Math Destruction)의 특징을 상세한 사례와 분석을 통해 파헤친다.

보이지 않는 여성들

작가: 캐럴라인 크리아도 페레스 | 옮긴이: 황가한 | 출판사: 웅진지식하우스

스마트폰을 자꾸 떨어뜨리는가? 사무실 냉방 온도가 낮아 감기를 달고 사는가? 마스크나 안전벨트를 착용하면 너무 헐겁거나 꽉 끼고, 처방받은 약이 어쩐지 효과를 보이지 않는가? 그렇다면 당신은 여성일 가능성이 높다. 이 책은 남성을 위해, 남성에 의해 설계된 이 세계가 어떻게 인구의 반, 여성을 배제하는지 증명한다. 남자를 인간의 디폴트값으로 여기는 사고방식 때문에 여성과 관련된 지식과 정보는 제대로 수집되지 않는다. 그렇게 생겨난 데이터 공백은 여자들을 가난하게 만들고 아프게 만들고 때로는 죽이기까지 한다.

자동화된 불평등

작가: 버지니아 유뱅크스 | 옮긴이: 김영선 | 출판사: 북트리거

가난한 사람들을 표적으로 삼는 자동화 시스템의 실체를 폭로하는 책이다. 뉴욕주립대학교 정치학 부교수 버지니아 유뱅크스는 법 집행부터 의료보험, 사회복지사업까지 미국의 공공 정책에 도입된 자동화 기술이 시민권 및 인권, 경제 형평성에 어떤 영향을 미치는지 낱낱이 보여 준다.

빅나인

작가: 에이미 웹 | 옮긴이: 채인택 | 출판사: 토트

AI는 이미 재정·금융 시스템과 전력망 그리고 유통 공급 체인의 중추를 차지했다. 우리가 교통 체증 속을 빠져나가도록 알려 주고, 잘못 입력한 단어의 정확한 의미를 찾아 주며, 무엇을 사고, 보고 들어야 할지를 결정해 주는 보이지 않는 인프라다. 또한 우리의 미래를 만들고 있는 기술이기도 하다. 그렇다면 20년 뒤, 50년 뒤 AI와 더불어 살고 있는 인류는 어떤 모습일까. AI는 의료, 주택, 농업, 교통, 스포츠, 심지어 사랑에까지 우리 삶의 모든 측면에 개입하고 있다. AI에 대한 당신의 생각은 무엇인가? 낙관적인가, 실용적인가, 아니면 파국적인가? 미래학자 에이미 웹은 실재 데이터와 연구 자료를 바탕으로 모델링한 3개의 미래 시나리오를 펼쳐 보이며 AI의 미래를 어떻게 준비할지 이야기한다.

데이터 프라이버시

저자: 니혼게이자이신문 데이터경제취재반 | 옮긴이: 전선영 |
출판사: 머스트리드북

우리는 최신 데이터 기술로 편리한 생활을 누리고 생산성을 높이는 대가로 중요한 개인정보를 기업에 내준다. 자기도 모르는 사이 사생활이 침범당할 위험에 노출돼 있다. 아이디 제휴에서 스코어링, 프로파일링, 딥페이크, 표적형 사이버 공격까지 빠르게 진화하는 기술에 대한 불안도 제기되고 있다. 데이터의 사용 방법에 따라 새로운 격차 사회가 출현할 수 있다는 우려도 끊이지 않는다. 데이터가 가져오는 경제성장과 편리한 사회를 향한 기대는 여전히 크지만, 개인 생활과 사회를 갉아먹는 부작용도 더는 간과할 수 없게 되었다. 일터와 가정에서 하루하루를 살아가는 한 개인으로서 우리는 어떤 변화를 겪고, 밀려오는 변혁의 큰 파도에 맞서 어떻게 대처해야 할까. 디지털 기술의 진보 끝에 나타나는 사회는 어떤 모습이고, 데이터 경제는 정말로 풍요로운 미래를 약속할까.

*각 소개글은 넷플릭스와 출판사 소개글을 인용·발췌했습니다.

민우회는 당신의 목소리가, 삶이 곧 운동이 되는 곳
지금보다 좀 더 나은, 다른 세상을 꿈꾸는 당신과 함께 합니다.

한국여성민우회는 각자의 존엄성을 지키며 차별 없이 평등하게 공존하는 세상을 향해
성평등한 노동권, 일과 생활의 균형을 위한 활동
여성이 자신의 몸과 건강의 주체가 되는 활동
성인지적 관점의 미디어 감시 활동
성평등 관점으로 복지국가를 기획하는 활동
성폭력 없는 세상을 만드는 반성폭력 활동
더불어 사는 민주사회를 위한 사회개혁 활동
풀뿌리로부터의 변화를 만드는 신나는 지역여성운동을 만들어 갑니다.

발행처 한국여성민우회
발행인 강혜란 최진협
만든이 문미향 박지수 신혜정 이윤소
디자인 양민영
발행일 2021년 10월
문의 media@womenlink.or.kr 02-737-5763

본 사업은 한국여성재단이 지원하는
사업입니다.

한국여성민우회

[03969] 서울시 마포구 월드컵로26길 39
시민공간 나루 3층(성산동)
전화 02-737-5763
홈페이지 womenlink.or.kr
이메일 minwoo@womenlink.or.kr
트위터 @womenlink 인스타그램
@women_link 페이스북 womenlink1987
3천원 문자후원 #2540-3838