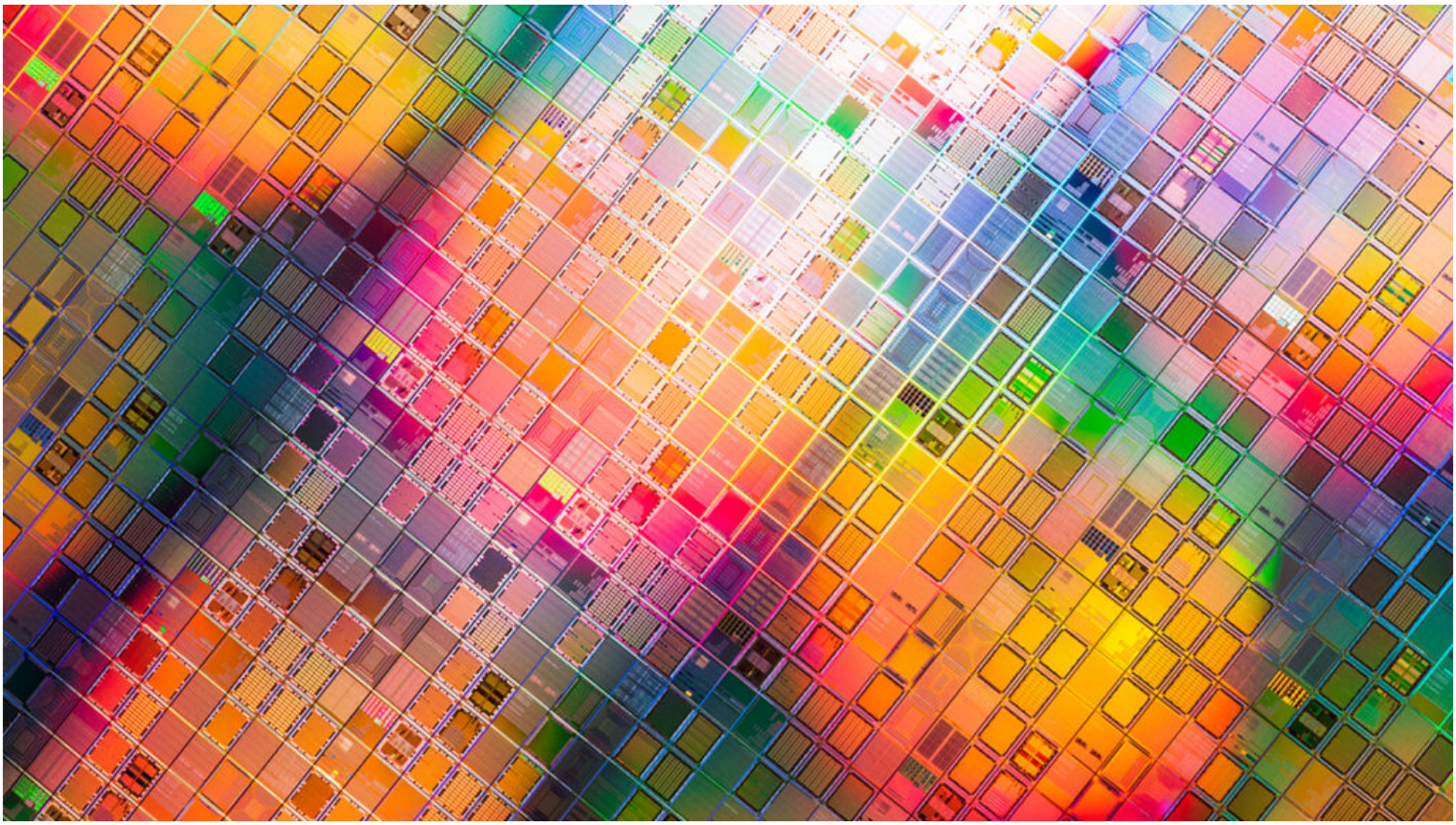
데이터 사이언스

구글이 AI 윤리에 접근하는 방식

도널드 마틴 주니어 (Donald Martin Jr.),앤드루 무어 (Andrew Moore)

디지털 | 2020. 11. 23.



AI는 다양한 산업에서 완전히 새로운 기회를 창출해내며 기술 진보의 주요 원동력으로 자리 잡았습니다. 공학적 관점에서 봤을 때, AI는 그저 조금 더 진보한 데이터 엔지니어링 기법에 불과합니다. 좋은 AI 프로젝트는 자동차에 비유하자면 대부분 매끈한 레이싱카보다는 지저분한 픽업트럭에 가깝습니다. 생산 라인의 안전성을 5% 정도 개선하기도 하고, 영화를 조금 더 정확하게 추천할 수 있도록 하는, 묵묵하고 듬직하게 작동하는 그런 기술이죠.

하지만 다른 어떤 기술보다도 AI를 사용하는 기술자는 의도치 않은 악영향을 끼칠 위험이 아주, 아주 높습니다. AI에는 불공정한 편향성을 확대하고, 내재된 편향성을 훨씬 파괴적으로 만들 힘이 있기 때문이죠.

구글의 AI 기술자로서, 우리는 AI 기술이 개발되고 사용되는 방식이 앞으로 우리 사회에 얼마나 큰 영향을 미칠지 잘 알고 있습니다. 그렇기에 좋은 사용 지침을 만드는 것이 필수적입니다. 책임감 있게 기술을 개발하는 것과 혹시 모를 잠재적 편향성을 방지하는 것이 첫걸음입니다. 이 두 가지 모두 기술자들이 몇 걸음 더 멀리 내다봐야만 가능한 일입니다. “이 자동화 방식이 비용을 15% 절감할 수 있을까?”가 아니라 “이 변화가 우리가 일하는 도시와 시민들, 특히 취약 계층의 사람들에게 어떤 영향을 끼칠까?” 하는 질문을 해야 한다는 겁니다.

여기에는 오래된 방식의 접근이 필요합니다. 기계가 아닌 인간 데이터 과학자들이 데이터세트와 모델에 들어

가는 변수가 어떻게 생성되는지 이해하는 것 말입니다. 여기에는 데이터에 반영된 집단뿐 아니라 데이터에 영향을 받는 집단과의 협업이 필수적입니다. 공동체 구성원이나 AI가 사용되는 복잡한 시스템을 이해하는 전문가들 말입니다.

잘못된 인과관계 분석은 불공정한 편향성을 낳는다

첨단 AI 기술에서 공정성을 확보하겠다는 목표를 어떻게 하면 구체화할 수 있을까요? AI는 종종 우리가 예상하지 못한 방식으로 작동하기도 하는데 말입니다. 가장 먼저, 컴퓨터 과학자들이 AI가 개발되고 사용되는 맥락을 보다 잘 이해해기 위해 노력해야 합니다.

불공정한 편향성을 식별하고 측정하는 기법이 많이 발전하기는 했지만 아직도 인과관계에 대한 잘못된 분석은 소수자들에게 악영향을 끼칠 수 있습니다. 인과관계에 대한 잘못된 분석이란 무엇일까요? 예를 들어, 중세시대에는 병에 걸린 사람들에게는 머릿니가 적다는 것을 보고 머릿니가 사람에게 이롭다고 생각했습니다. 실제로는 머릿니가 열이 나는 사람들을 피하는 것이었는데 말입니다.

이와 같이 상관관계를 인과관계로 착각하는 실수는 의료 제도나 형사 제도와 같은 중요한 영역에서 엄청난 피해를 초래할 수 있습니다. AI 시스템 개발자들은 대부분 사회과학을 공부한 경험이 적고, 자신들의 기술이 해결하고자 하는 문제의 바탕에 깔린 사회구조를 이해하지 못하는 경우가 많습니다. 지나치게 단순하거나 부정확하게 인과관계를 분석한 채로 중요한 사회적 요소가 배제된 AI 시스템을 개발하고, 이로 인해 의도치 않게 피해를 초래한다는 겁니다.

예를 들어 보겠습니다. 한 연구진은 미국 의료보험 업계에서 널리 쓰이는 알고리즘이 흑인 환자들에게 불리하게 편중돼 있다는 것을 발견했는데요. 이는 알고리즘을 설계한 기술자들이 의료 혜택이 많이 필요한 사람일수록 보험에 지출한 액수가 높을 거라는 잘못된 가정을 내렸기 때문이라고 합니다. 이는 흑인들의 의료보험 지출액이 적으니 그들은 의료 혜택이 크게 필요하지 않다는 결론으로 이어질 수 있습니다.

하지만 사실 흑인 환자들이 의료보험 지출을 적게 하는 데는 다른 여러 가지 이유가 있을 수 있습니다. 이분들이 의료보험 체계를 믿지 못하거나, 혹은 경제적으로 부담하기 어렵기 때문일 수 있지요.

상관관계와 인과관계를 혼동하는 것은 흔한 일입니다. 하지만 딥러닝 컴퓨터의 경우는 문제가 더 심각합니다. 딥러닝 컴퓨터는 데이터를 예측하는 가장 정확한 방법을 찾기 위해 수십억 개의 가능한 상관관계를 탐색하는데요. 이 과정에서 잘못된 인과관계를 가정할 경우의 수도 수십억 개가 되기 때문입니다. 설상가상으로, 섀플리 분석과 같은 최신 기법을 사용하더라도 왜 잘못된 인과관계가 도출됐는지 파악하는 것은 아주 어렵습니다. 데이터 자체만 가지고는 슈퍼컴퓨터를 사용하더라도 어디서, 어떤 가정이 틀렸는지 알아낼 수 없다는 겁니다.

이런 이유로 과학계에서는 수동적으로 데이터만 보고 자연적인 인과관계를 찾아냈다고 하는 것은 절대 인정받지 못합니다. 가설을 세우고 인과관계를 엄밀하게 도출하기 위해 실험을 진행하는 과정을 거쳐야만 하지요.

인과관계가 잘못 도출되는 문제를 해결하기 위해서는 한 걸음 뒤로 물러날 필요가 있습니다. 컴퓨터 과학자들

이 자신의 기술이 개발되고 사용되는 사회적 맥락을 이해하고 반영하기 위해 보다 많은 노력을 기울여야 한다는 뜻입니다.

구글에서 우리는 이런 노력의 기초를 다지기 시작했습니다. 최근 딥마인드, 구글 AI, 그리고 구글의 신용 및 안전팀이 함께 논문을 발표했는데요. 사회적 맥락을 이해하고자 할 때는 사회적 맥락이 눈에 잘 보이지 않는 피드백 메커니즘에 따라 움직이고, 역동성, 복합성, 비선형성, 적응성을 갖추고 있다는 사실을 인정하는 것에서부터 시작돼야 합니다. 우리는 모두 이런 사회 시스템의 일부이지만 그 어떤 개인이나 알고리즘도 이를 온전히 관측하거나 이해할 수는 없습니다. 따라서 불가피한 맹점을 해결하고 책임감 있는 혁신을 이끌기 위해 기술자들은 사회학, 행동과학, 인문학 분야의 전문가들 및 취약 계층의 대표 등 다양한 관계자 협업해야만 합니다. 이런 일은 제품 디자인이 시작되기 전 개발 첫 단계부터 진행돼야 하고 알고리즘의 편향성에 가장 취약한 계층과의 파트너십을 통해 이뤄져야 합니다.

이와 같이 복잡한 사회 체계를 이해하고자 하는 접근법을 '공동체 기반 시스템 역학 (community-based system dynamics, CBSD)'이라고 부릅니다. 이를 위해서는 필요한 관계자들을 참여시킬 수 있는 새로운 네트워크를 구축해야만 합니다.

CBSD는 복잡한 문제를 함께 이해하고 규정하기 위해 정성적, 정량적 방법론을 치밀하게 적용합니다. 참여하는 모든 이에게 이득이 될 수 있는 공정하고 윤리적인 방식으로 다양한 공동체와 협업할 수 있는 역량을 기르는 것이 최우선입니다.

쉬운 일은 아닐 겁니다. 하지만 사회의 가장 취약한 계층에게 가장 중요한 문제를 깊이 이해함으로써 얻은 통찰은 모든 이에게 보다 안전하고 이로운 기술 진보를 이끌어낼 수 있습니다.

'만들 수 있으니 만든다는' 마인드세트에서 '만들어야 하는 것을 만드는' 마인드세트로

제품 개발 및 디자인 과정에서 특정 집단을 간과한다면 그 집단은 결과물로 만들어진 제품의 혜택 또한 적게 받게 됩니다. 지금 우리는 AI의 미래가 어떤 모습일지 디자인하고 있습니다. 포괄적이고 평등한 모습일까요? 아니면 우리 사회의 가장 불공평하고 부정의한 요소를 반영하게 될까요? 보다 정의로운 AI의 미래는 거저 얻어지는 것이 아닙니다. 우리가 노력해야만 합니다. 우리가 그리는 기술의 비전은 모든 관점과 경험, 그리고 구조적 불공정성이 고려되는 것입니다.

우리는 다양한 계층의 관점을 반영하기 위해 여러 방면에서 노력하고 있습니다. 인권 실사 프로세스나 리서치 스프린트(research sprint)뿐 아니라 취약 계층에서 직접적으로 의견을 반영할 수 있는 제도를 운영하고 있습니다. 또한 여성 기술자들을 위한 WiML(Women in ML), 라틴계 기술자들을 위한 Latinx in AI 등 다양성과 평등을 제고하기 위한 단체들도 있습니다. 흑인 단체 Black in AI나 성소수자 단체 Queer in AI와 같이 구글의 연구진이 함께 창립하거나 이끌고 있는 단체들도 빼놓을 수 없죠.

AI 기술이 우리의 이상에 부합하도록 하기 위해서는 무엇을 만들지에 대한 AI 업계의 인식 전환이 필요합니다. '만들 수 있으니 만든다는' 마인드세트에서 '만들어야 하는 것을 만드는' 마인드세트로 바뀌어야 합니다.

문제를 깊이 있게 이해하는 것에 근본적으로 집중하고, 소외된 공동체와 윤리적으로 협업하기 위해 노력해야 합니다. 이를 통해 우리가 만든 알고리즘에 사용되는 데이터와 우리가 해결하고자 하는 문제에 대해 보다 신뢰도 있게 이해할 수 있습니다. 이런 이해를 바탕으로 모든 분야의 단체들이 포괄적이고, 평등하면서, 사회적으로 이로운 범위 내에서 새로운 기회를 창조해낼 수 있을 것입니다.

원문: *AI Engineers Need to Think Beyond Engineering*

도널드 마틴 주니어는 구글의 선임 기술 프로그램 매니저이자 사회적 기술 전략가입니다.
앤드루 무어는 구글 클라우드 AI 및 산업 솔루션 부서의 장입니다.