

IBM 클라우드 기반의 신뢰할 수 있고 투명한 AI

귀사에 신뢰도와 투명성을 갖춘 AI를 구현, 실행 및 관리하십시오.



주요 특징

- AI의 정확성, 성능, 공정성의 추이 모니터링
- 모델에서 유해한 편향 탐지 및 대응
- 현업 지식 근로자의 AI 의사결정 이해, 감사, 설명 지원
- 신뢰할 수 있고 설명할 수 있는 결과를 통해 조직 내 AI 구현과 활용의 가속화 및 최적화

비즈니스 가치 제고를 위해 신뢰할 수 있고 투명한 AI의 구현, 실행 및 관리

IBM은 신뢰도와 투명성을 실현하는 새로운 기능을 통해 전사적 범위에서 AI 모델 및 애플리케이션을 구현, 실행, 관리하는 기업을 지원하고 있습니다.

AI 여정의 장애물

기업들은 AI가 당장 업계의 발전을 이끌 동력임을 점점 더 확신하고 있습니다. 콜센터 보안을 위해 AI 기반 챗봇을 구현하거나 문서 처리나 이미지 인식에 AI를 활용하여 백오피스 업무를 자동화하는 방식으로 이미 이 여정의 첫발을 내디딘 곳도 많습니다.

하지만 매우 성공적인 AI 프로젝트에서도 비즈니스 차원에서 기대했던 결과를 얻지 못하는 경우가 종종 있습니다. 최고의 전문성을 가진 데이터 과학 팀도 매년 소수의 모델만 프로덕션 환경에 배치할 수 있습니다. AI 구축 구조가 자칫 복잡해질 수 있고, 여전히 존재하는 기술력 및 툴 사용의 허점 때문에 데이터 과학, IT 운영, 비즈니스 팀의 협업이 어려워질 수 있기 때문입니다. 이러한 운영상의 문제뿐만 아니라 훨씬 더 중대한 신뢰도 및 투명성의 문제도 있습니다. AI 프로젝트에서 진정한 비즈니스 우위를 확보하려는 기업은 먼저 이 문제를 해결해야 합니다.

신뢰도 강화

지식 근로자가 비즈니스의 전 범위에서 의사결정 지원을 위해 자신 있게 AI를 사용하려면 먼저 AI를 신뢰하고 그 결과를 설명할 수 있어야 합니다. AI가 데이터를 받아 수치를 내놓는 블랙박스라면 기업이 그 수치가 의사결정에 유용한 길잡이인지 여부를 판단하기가 쉽지 않습니다. 게다가 고객, 감사자 또는 컴플라이언스 팀에 결과를 설명할 수도 없습니다.

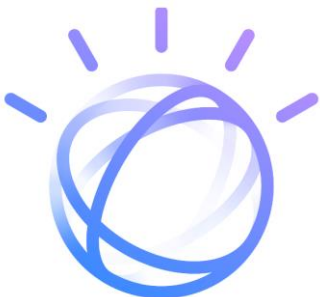




그림 A:
AI 애플리케이션 상태를 실시간으로 추적하여 정확성, 공정성, 성능 평가

현재 많은 유망한 모델이 프로덕션 단계까지 이르지 못하는 이유는 기업이 AI 결과를 확실히 이해하지 못해 신뢰할 수 없기 때문입니다. 기업의 기대 및 정책에 완전히 부합하지 않는 AI에 책임을 맡겼다가는 중대한 위험에 봉착하게 됩니다. 예를 들어 어떤 모델에서 특정 고객 집단에 대한 충분한 크기의 샘플이 학습 데이터에 포함되지 않아 이 집단이 부당하게 차별을 받는다면 막대한 경제적 손실 또는 이미지 실추로 이어질 수 있습니다.

공정성 보장

AI 모델이 아무리 세심하게 설계되었다고 일정 정도의 편향을 나타내기 마련입니다. 모델의 효용 가치는 그 모델이 학습한 데이터에 좌우됩니다. 학습 데이터 세트가 현실의 데이터를 100% 반영한다고 볼 수 없으므로 새롭게 훈련받은 모델이 프로덕션에서 제 기능을 하지 못할 위험성은 항상 있습니다. 게다가 데이터 도메인 대부분이 끊임없이 진화하므로 시간이 흐르면 모델의 정확성이 달라지게 됩니다.

실시간 가시성이 관건입니다. AI 모델의 운영 라이프사이클 전반에서 그 정확성, 성능, 공정성을 모니터링하고 그 결과가 나오기까지의 추론에 대한 현업 사용자의 이해를 돕는 데 분석을 활용한다면 AI 여정의 가장 큰 걸림돌을 해결할 수 있습니다.

설명 가능한 결정

많은 업종에서 규제 감시가 AI 도입의 어려움을 가중시킵니다. 기업이 모델의 공정성에 만족하고 그 결과를 신뢰할 수 있더라도, 규제 기관이 더 엄격한 접근법을 요구하는 경우가 많습니다.

이런 까닭으로 AI 기반 애플리케이션의 모든 모델, 데이터, 입출력의 계보를 철저히 추적하면서 모든 결정에 대한 AI 입력을 완전히 설명 가능한 상태로 만들어야 합니다. 초기 설계, 학습, 구축부터 운영 및 폐기까지 모든 AI 자산의 라이프사이클 전 범위에 대한 감사가 가능해야 합니다. 어떤 모델에 대해 그 모델 및 학습에 쓰인 데이터 세트를 구현한 팀, 프로덕션에서 받은 입력, 생성한 결과까지 파악할 수 있어야 합니다.

IBM 클라우드 기반 AI를 위한 IBM의 새로운 신뢰도 및 투명성 기능 소개

IBM이 새롭게 선보이는 편향 탐지/완화 및 설명 지원 기능은 조직 전반에서 AI가 어떻게 쓰이고 있는지 모니터링하여 AI를 구현, 실행, 관리하는 데이터 과학자 및 애플리케이션 개발자의 작업을 보강함으로써 AI의 신뢰도 및 투명성을 높입니다.

이 기능은 IBM 클라우드에서 실행 중인 모델에 최적화되었을 뿐만 아니라 다양한 머신 러닝 프레임워크 및 AI 구현 환경, 즉 TensorFlow, SparkML, AWS SageMaker, Microsoft Azure Machine Learning 등에도 사용할 수 있습니다.

기능의 작동 방식

AI 설명: 모든 비즈니스 사용자가 AI에서 어떤 결정에 도달하는 방식을 이해할 수 있도록 지원하여 이 기술을 신뢰할 수 있게 합니다. 이 기능은 모델에서 처리하는 데이터를 자동으로 로깅하므로 모든 결정 및 예측, 전체 데이터 및 모델 계보에 대한 철저한 추적이 가능해집니다. 이러한 데이터 로깅 덕분에 훨씬 더 효과적으로 감사 및 컴플라이언스 리포팅을 수행할 뿐만 아니라 강력한 분석도 해낼 수 있습니다. 사용자는 모든 비즈니스 트랜잭션에 대해 쿼리를 실행하고, 모델이 어떤 과정을 거쳐 권고안을 도출했는지 알 수 있습니다. 이 설명은 현업 사용자가 쉽게 이해하는 언어로 이루어집니다.

런타임 편향 탐지 및 완화: 어떤 기업의 AI가 불공정한 결과를 초래할 수 있는 어떤 편향도 받아들이거나 심화시키지 않게 합니다. 기업에서는 정교한 진단 서비스를 실행하여

모델의 정확성과 성능을 평가할 수 있습니다. IBM 연구소의 혁신적인 기술을 기반으로 이상 및 편향을 탐지하는 최첨단 기능을 통해 데이터와 모델 모두에서 유해한 편향을 찾아냅니다. 빌드타임 및 런타임에 편향 검사를 수행하여 최대한 일찍 문제점을 발견할 수 있습니다. 그런 다음 모델 재학습에 사용할 새 데이터 세트를 추천하는 방식으로 유해한 편향을 완화합니다.

유스 케이스

IBM이 새롭게 제공하는 AI를 위한 기능은 기업이 AI를 신뢰하고 설명하면서 비즈니스 문제를 해결하고 가치를 창출함과 동시에 위험을 획기적으로 줄일 수 있게 합니다. 다양한 업종에서 이 솔루션으로 누릴 수 있는 효과의 예는 다음과 같습니다.

대출 승인 절차 간소화

보통 주택 대출을 승인받는 데 30일가량 소요되는데, 대출 기관의 업무가 많은 기간에는 더 오래 걸리기도 합니다. 이 프로세스에서는 각 신청 건에 대해 수십 차례의 조회를 수행하여 연체 또는 채무 불이행 위험을 평가합니다.

AI 모델은 다양한 데이터를 활용하여 대출 기관이 위험도가 낮은 신청자를 빨리 승인하고 고위험 신청자를 식별함으로써 고객 서비스를 개선하고 수익을 늘리며 손실 위험을 낮출 수 있게 합니다. 그러나 이러한 모델이 개발 단계에서는 뛰어난 결과를 내놓을 때가 많지만, 모델을 확장하여 전체 대출 승인 프로세스를 안정적으로 지원하려면 상당한 어려움이 따를 수도 있습니다.

IBM의 새로운 편향 탐지 및 설명 기능으로 이러한 문제를 해결할 수 있습니다. 즉 현업 사용자가 모델에서 생성된 위험 점수를 확인할 뿐만 아니라 그 점수 산정의 로직을 이해하도록 인사이트를 제공하는 것입니다. 모델이 정확하며 그 결과가 공정함을 알고 있으면 더 자신 있게 대출 승인/거부 결정을 내릴 수 있습니다. 게다가 고객이나 규제 기관이 어떤 결론의 근거를 물을 경우 대출 기관은 그 모델에서 결론에 이르게 된 과정을 손쉽게 설명할 수 있습니다.

신뢰할 수 있는 AI

AI 를 위한 신뢰도 및 투명성 기능으로 모델에 대한 완전한 가시성을 확보하고, AI에 대한 완전한 설명을 제공하여 공정한 결과를 보장할 뿐만 아니라 비즈니스 프로세스 소유자에게 AI의 의사결정 지원 기능에 대한 확신을 심어줄 수 있습니다. 기업은 이러한 기능을 통해 AI가 내리는 결정을 심도 있게 이해하고, 자신감을 갖고 더 빨리, 광범위하게 AI를 활용할 수 있습니다.

추가 정보

IBM 클라우드에서 제공하는 새로운 AI 신뢰도 및 투명성 기능에 관한 자세한 내용은 IBM 영업 담당자 또는 IBM 비즈니스 파트너에게 문의하거나 다음 웹 사이트에서 확인하십시오. ibm.com/watson/trust-transparency



© Copyright IBM Corporation 2018

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
2018년 9월

IBM, IBM 로고, ibm.com은 전세계 여러 국가에 등록된 International Business Machines Corp.의 상표 또는 등록상표입니다. 기타 제품 및 서비스 이름은 IBM 또는 타사의 상표입니다. 현재 IBM 상표 목록은 웹 "저작권 및 상표 정보"(ibm.com/legal/copytrade.shtml)에 있습니다.

이 문서는 최초 발행일을 기준으로 하며, 통지 없이 언제든지 변경될 수 있습니다. IBM이 영업하는 모든 국가에서 모든 오퍼링이 제공되는 것은 아닙니다.

이 문서의 정보는 상품성, 특정 목적에의 적합성에 대한 보증 및 타인의 권리 침해에 대한 보증이나 조건을 포함하여(단, 이에 한하지 않음) 명시적이든 묵시적이든 일체의 보증 없이 "현상대로" 제공됩니다. IBM 제품은 제품이 제공되는 계약의 조건에 따라 보증됩니다.

IBM이 제시하는 계획, 방향 및 의도에 대한 IBM의 모든 언급은 특별한 통지 없이 변경 또는 철회될 수 있습니다. 향후 제품에 대한 정보는 IBM 제품의 일반적인 방향성을 설명하고자 하는 것이며, 이러한 정보는 제품 구입을 결정하는 근거가 될 수 없습니다. 향후 제품에 대한 정보는 임의의 자료, 코드 또는 기능을 제공하기 위한 약속이나 법적 책임사항을 의미하지 않습니다. 향후 제품에 대한 정보는 계약서의 일부로 통합될 수 없습니다. IBM 제품에 대해 기술된 향후 기능의 개발, 릴리스 및 출시 시기는 전적으로 IBM의 결정에 따릅니다.



Please Recycle

