Business Software Alliance

AI 학습: AI 정책의 핵심 쟁점

지난 10년간의 기술 혁신은 인공지능(AI)을 단순한 '신흥 기술'의 단계에서 벗어나, 경제 전 분야의 기업 활동을 실질적으로 강화하는 상업적 현실로 발전시켰습니다. AI 도구를 활용해 항공사는 기체 청소를 더 효율적으로 수행하고, 농업은 방대한 기상 데이터를 분석해 수확량을 극대화합니다. AI를 통해 제조업체는 신제품을 실험하고, 사이버보안 기업은 신종 위협을 신속하게 탐지합니다. 건설사는 실제 도시를 기반으로 AI가 생성한 '디지털 트윈'을 구 축해, 설계도의 영향을 분석합니다.

모든 AI 도구가 올바르게 작동하기 위해서는 충분히 학습된 AI 모델이 필수적입니다. 따라서 경제 전반에서 활용되는 AI 시스템이 안정적으로 기능할 수 있도록, 이를 뒷받침하는 견고한 학습 과정을 지원하는 AI 정책의 마련이 필수적입니다.

AI 학습은 어떻게 이루어질까?

AI 모델을 '학생'에 비유해봅시다. 대규모 언어 모델(LLM)과 같은 AI 모델을 학습 시키는 방법 중 하나는, AI 모델에게 다양한 예시가 포함된 방대한 데이터를 제공하고 각 예시에 대한 정답을 알려주는 것입니다. 이후 AI 모델은 각 예시에 대한 질문을 받고, 그에 대한 답변을 개선하기 위한 피드백을 받습니다. 또 다른 방법은 AI 모델에게 명시적인 지침을 주지 않고 대량의 원시 데이터를 제공하여, AI 모델이 스스로 데이터 전반에서의 패턴과 구조를 학습하도록 하는 것입니다.

AI 모델 학습에서 무엇보다 중요한 것은 데이터의 품질입니다. AI 모델은 편향된 결과를 초래할 수 있는 제한적 데이터셋으로 학습할 때보다, 다양한 유형의 데이터가 포함된 폭넓은 데이터셋으로 학습될 때 훨씬 더 견고해지며, 목표한 결과를 보다 효과적으로 달성할 수 있습니다.

AI 학습은 AI 모델 개발 과정의 전체 생애주기 중 하나의 단계에 해당합니다

AI 프로젝트의 정의
AI 모델은 특정 작업 수행을 목표로 개발되기도 하고, 광범위한 AI 시스템을 구동하기 위한 범용 도구로 설계되기도 합니다. 개발 초기 단계에서 설계팀은 AI 프로젝트의 목적과 구조를 정의합니다.

→ **학습 데이터의 식별**
AI 모델은 일반적으로 방대한 양의 데이터를 기반으로 학습됩니다. 비지도 학습 방식으로 작동하는 경우에는 라벨이 지정되지 않은 데이터를 사용하기도 합니다. 반면, 특정한 작업 수행을 목표로 개발된 AI 모델의 경우에는 라벨이 붙은 구조화된 데이터가 활용됩니다.

→ **학습 데이터 준비**
원시 데이터는 학습에 활용되기 전 중복값, 이상값, 누락값, 일관되지 않은 서식 등을 처리하는 방식으로 정제·수정됩니다. 이후 데이터는 의미적·구조적으로 더 작은 단위인 '토큰'으로 변환되며, 최종적으로 학습 데이터는 토큰 시퀀스 형태로 구성됩니다.

→ **모델 정의**
학습 데이터가 처리된 후, 개발자는 AI 시스템의 기본 아키텍처를 설계합니다. 이 과정에는 알고리즘이 예측의 근거가 되는 패턴과 관계를 탐색하면서, 평가 대상이 될 학습 데이터의 주요 특징을 식별하는 단계가 포함됩니다. 또한, AI 시스템을 구동할 알고리즘 모델의 유형을 선택합니다.

→ **모델 검증, 테스트 및 수정**
AI 모델이 학습을 마친 후에는 의도한 대로 작동하는지 검증하고, 산출물이 예상 범위 내에 있는지 테스트해야 합니다. 검증과 테스트 결과에 따라 모델을 수정하거나 보완하는 과정이 뒤따를 수 있습니다.

→ **사후 학습**
AI 모델은 초기 학습 이후에도 정제와 최적화 과정을 거치는 경우가 많습니다. 이 과정에서는 미세 조정(fine-tuning)이나 강화학습과 같은 기술이 활용될 수 있습니다. 이러한 사후 학습 기법은 모델을 특정 작업에 맞게 조정하거나, 전반적인 정확도·속도·효율성을 개선하는 데 사용됩니다.

AI 정책은 AI 학습에 어떤 영향을 미칠까?

입법자들은 기업이 AI 시스템을 개발하고 배포하는 과정에서 책임 있게 행동하도록 보장하기 위해 다양한 정책을 고려할 수 있습니다. 이러한 정책은 기업이 방대하고 다양한 유형의 데이터를 포함한 견고한 데이터셋으로 AI 모델을 학습시키도록 하고, 이를 장려하는 역할을 해야 합니다.

그러나 일부 정책은 의도와 달리, 기업이 광범위한 AI 모델을 충분히 학습시키는 과정을 제한할 수 있습니다. 이러한 제약은 다음과 같은 법률적 요인에서 비롯될 수 있습니다.

1. 저작권
2. 개인정보
3. 공개 요건
4. 국경 간 데이터 접근

1 저작권법

AI 모델 학습에 사용되는 데이터셋에는 일부 저작물이 포함될 수 있습니다. 그러나 대규모 AI 모델의 학습 과정은 개별 저작물의 창의적 표현을 모방하거나 재현하는 것과는 무관합니다. 대신, 방대한 데이터셋은 '토큰'이라고 불리는 작은 텍스트 단위의 조각으로 분해되어, AI 모델이 텍스트를 이해하고 생성하는 데 필요한 기초 단위를 형성합니다. 이후, AI모델은 데이터셋 내 수백만 또는 수십억 개의 토큰에 대한 확률, 상관관계, 추세 및 기타 패턴을 수학적으로 계산하는 연산 분석 과정을 거칩니다.

이러한 형태의 학습에서 핵심은 데이터가 저작물의 표현적 내용을 학습하기 위해 사용되는 것이 아니라 학습 과정 자체를 형성하는데 사용된다는 것입니다. 예를 들어, "사과"라는 단어가 사용된 충분한 문장을 토큰화하면, AI 모델은 주체가 잘못에 대해 미안함을 표현한다"의 의미로 단어를 사용한 경우와 '과일의 일종'을 의미하는 경우를 구별해 학습할 수 있습니다. 모델이 이 과정에서 저작권이 존재하는 작품의 문장의 일부를 분석했을 수 있지만, 그 목적은 보호의 대상인 표현을 학습하는 것이 아니라 언어적 상관관계를 학습하는 것입니다.

아울러, 유연한 혁신과 건전한 지식재산(IP) 생태계의 일환으로, 정부는 저작권자가 온라인에서 발견한 경우 침해 콘텐츠의 삭제를 요청할 수 있도록 해야 합니다. 이는 이미 미국, 일본, 싱가포르 등 국가에서는 "통지 및 삭제" 절차와 법적 절차를 시행하고 있습니다. 나아가, AI 개발자와 저작권자는 저작권자가 특정 웹사이트나 저작물의 일부를 AI 학습에 활용하지 않기를 원하는 경우, 그 의사를 명시적으로 표시할 수 있는 합의 기반의 기계판독형 제어체계를 마련하기 위해 지속적으로 협력해야 합니다.

권고사항:

- AI 모델의 학습 과정에서 저작권이 있는 작품의 일부에서 비롯된 자료가 포함되더라도, 모든 자료에 대한 비소모적 분석이 법적으로 허용되도록 하는 정책을 채택하거나 유지해야 합니다.
- 온라인에서 발견된 침해 콘텐츠를 삭제할 수 있도록 IP 보유자에게 법적 수단을 지속적으로 제공해야 합니다.
- 당사자가 AI 학습에 대한 '옵트아웃' 의사를 명시적으로 표시할 수 있도록 하는 자발적 통제 장치의 개발을 장려해야 합니다.

2 개인정보 보호법

개인정보 보호법은 기업이 개인 데이터를 수집하고 사용하는 방식을 제한함으로써, 개인의 권리를 보호하는 중요한 안전장치 역할을 합니다. 일부 경우 개인정보 및 데이터 보호 관련 법률은 기업이 개인 데이터를 처리하기 위해 합법적 목적 등 명확한 근거를 갖출 것을 요구하기도 합니다.

또한 많은 경우 AI 도구는 개인과 무관한 정보를 바탕으로 학습된다는 점을 고려하면, AI 학습은 개인정보 보호법 적용 대상이 아닙니다. 예를 들어, 기상 패턴을 분석하도록 학습된 AI 시스템은 방대한 양의 기상 데이터를 처리하지만 이는 개인에 관한 데이터가 아니므로 일반적으로 개인정보 보호법의 보호 범위에 포함되지 않습니다. 다만, 특정 작업의 성능을 개선하기 위해 더 구체적인 데이터로 AI 모델을 학습시키는 경우가 있으며, 이 과정에서 개인 데이터를 사용하여 모델을 미세 조정할 수 있습니다.

개인정보 보호법이 적용되는 경우라 하더라도 기업이 적절한 보호 조치를 마련한다면 개인과 관련된 데이터 (이하 '개인 데이터') 를 바탕으로 AI 도구를 학습시키는 행위 자체를 금지해서는 안 됩니다. 예를 들어, 사람과 관련된 정보를 분석하기 위해 AI 도구를 사용하는 기업은 개인정보 보호법의 적용을 받는 개인 데이터를 활용하여 AI 도구를 미세 조정하고자 할 수 있습니다. 예컨대, 고객 서비스 불만 사항을 여러 부서로 전달하도록 설계된 AI 도구는 고객의 구매 이력·위치·

과거 불만 사항 등과 같은 개인 데이터를 식별하도록 미세 조정이 가능 할 때 더 효과적으로 작동할 것입니다. 이처럼 개인 데이터를 활용한 AI 시스템의 미세 조정은 시스템이 의도한 대로 작동하게 합니다. 따라서 개인정보 보호법은 개인 데이터를 활용한 책임 있는 AI 학습 활동을 지원하는 방향으로 운영되어야 합니다.

권고사항:

- 적절한 보호 조치를 전제로, 개인 데이터를 AI 도구 학습에 활용할 수 있도록 보장해야 합니다.
- 균형 잡힌 개인정보 보호 체계를 마련하고, 모든 데이터 처리에 대해 일률적인 동의를 요구하지 않아야 합니다. 개인정보 보호법은 정당한 이익 등 다양한 목적으로 개인 데이터를 처리할 수 있음을 명시적으로 허용해야 하며, 이를 통해 사회적으로 유익한 연구가 위축되거나 AI 도구의 학습용 데이터가 왜곡되는 상황을 방지해야 합니다.
- AI 학습을 데이터 처리의 정당한 목적으로 인정해야 합니다.
- AI 관련 활동에서, 개인 데이터를 처리하는 기업이 프라이버시 설계와 책임성 기반의 개인정보 보호 체계를 구축하도록 장려해야 합니다.

3 공개 요건

법률은 기업이 AI 도구의 개발 및 활용 방식에 대한 정보 공개를 요구할 수 있습니다. 이러한 조치는 경우에 따라 AI 기술에 대한 신뢰를 제고하는데 기여할 수 있습니다. 그러나 기업이 AI 모델 학습에 사용된 데이터셋의 수집 및 큐레이션 방법에 관한 구체적 정보를 공개해야 한다면, 이는 핵심 영업비밀 보호에 중대한 위협이 될 수 있습니다. 예를 들어, AI 모델이 학습한 구체적인 데이터셋과 해당 데이터의 큐레이션 방식을 공개하는 것은 결국 기업의 AI 모델 학습 방식 자체의 공개로 이어질 수 있습니다. 이 문제는 특히 모델이 개발 또는 배포된 후, 특정하고 제한된 데이터셋을 사용하여 "미세 조정"을 거쳐 모델의 작동 방식을 개선하는 경우 두드러집니다. 이는 영업비밀 유출 뿐만 아니라, 악의적 침해 주체에게 잠재적인 사이버 경로를 제공할 수도 있습니다.

또한, 일부 공개 요건은 현실적으로 이행이 불가능할 수 있습니다. AI 개발사업자에게 AI 학습 데이터셋과 간접적으로 관련된 모든 저작물 및 저작권자 정보를 공개하도록 요구하는 경우입니다. . 저작권은 창작과 동시에 발생하기때문에 음식점 리뷰, 블로그, 디지털 기록 등 사실상 모든 콘텐츠가 잠재적으로 저작권 보호 대상이 될 수 있습니다. 그러나 개별 저작물의 출처, 보호 기간, 소유권 및/또는 라이선스 세부 사항 등은 명백하게 드러나 있지 않고 현실적으로 파악하기 어렵습니다. 예를 들어, 하나의 작품에 여러 명의 중복된 권리자가 존재할 수 있으며, 명확한 권리자가 없는 "권리자 불명 저작물 (Orphan works)"이 있을 수 있습니다. 따라서 대규모 AI를 개발하거나 사용할 때, 토큰화 된 데이터셋의 일부가 저작권 보호대상의 일부분일지 여부를 세부적으로 판단하는 것은 사실상 불가능합니다. 더 나아가, 이러한 저작권 기반 공개 의무는 일반적으로 AI 학습에 적용되는 공정 이용 등 저작권 보호의 예외 조항과 충돌할 가능성도 있습니다.

권고사항:

- 공개 수준을 규제 기관, 사용자, 대중 등 대상에 따라 달리해야 합니다. 공개 의무는 저위험 AI 도구 보다는 고위험 AI 도구의 활용에 초점을 맞춰야 합니다.
- 투명성 의무가 기업의 영업비밀 공개 요구를 강제하는 것이 아님을 명확히 하여 영업비밀을 보호해야 합니다.
- 공개 요건은 소비자가 AI 도구를 이해하는 데 도움이 되는 고차원적 정보에 집중하되, 학습에 사용된 데이터셋의 세부 내용을 과도하게 요구하지 않아야 합니다.

4 AI 학습을 위한 국경 간 데이터 접근

AI 모델은 여러 국가에서 개발·배포되며, 학습에 사용되는 데이터 역시 전 세계 다양한 출처에서 수집됩니다. 이러한 맥락에서 다양성은 AI 모델의 정확도와 신뢰성을 향상시킬 수 있기에, 기업은 다양한 데이터 출처를 활용할 수 있도록 장려되어야 합니다. 따라서, 국경 간 데이터 이전 제한은 학습 데이터 접근을 방해하거나, 기업이 제한된 범위의 데이터셋으로 AI 모델을 학습시키도록 유도할 수 있으며, AI 도구의 산출물이 왜곡되는 부정적인 결과를 초래할 수 있습니다.

또한, 국경을 초월한 다양한 데이터를 AI 훈련에 활용하는 것은 매우 중요합니다. 예를 들어 의료 분야에서는 여러 국가의 의료 데이터를 기반으로 학습된 AI 도구가 환자의 진료 품질을 개선할 수 있고, 신종 질병 발생에 보다 효과적으로 대응할 수 있습니다. 사이버 보안 분야에서도 전 세계의 사건 데이터를 학습한 AI 도구가 새로운 지역에서 활동하는 악의적인 행위자를 더 빠르게 식별할 수 있습니다. 따라서, 정책은 전 세계적으로 활용되는 AI 도구의 성능을 향상시키기 위해, 여러 국가의 학습 데이터를 수집할 수 있도록 장려해야 합니다.

권고사항:

- 책임있고 신뢰할 수 있는 국경 간 데이터 이전을 장려해야 합니다.
- 데이터 현지화 요건 부과는 지양해야 합니다.
- 국가 간 조화를 통해 확장 가능한 AI 개발을 지원해야 합니다.

데이터의 중요성 - AI 모델 학습에서 데이터의 역할

AI 모델의 올바른 학습을 위해서는 데이터가 필수적입니다. 데이터가 없다면 모델의 학습은 불가능합니다. 이와 같이 데이터는 AI 학습에서 가장 중요한 요소이기에 개발자는 AI 모델 학습에 사용될 데이터셋을 신중하게 선정하고 큐레이션합니다. 학습 데이터를 식별하고 큐레이션하는 데 필요한 요소는 다음과 같습니다.

» 데이터의 품질. AI 모델에 검증되지 않은 저품질 데이터가 대량으로 입력되면, 산출물도 품질이 떨어질 가능성이 높습니다. AI 도구를 학습시키는 데 필요한 "양질의 데이터"를 식별하는 기준은 해당 AI 도구의 목적과 용도에 따라 달라집니다.

» 데이터의 양. AI는 일반적으로 방대한 양의 데이터를 학습함으로써 광범위한 데이터 포인트 기반을 구축합니다. 이러한 대규모 학습 데이터의 사용은, 모델이 다양한 시나리오로를 학습하고 이상값을 식별하는 데 도움이 됩니다.

» 데이터의 다양성. 학습 데이터는 다양한 데이터 포인트를 대표하는 광범위한 단면을 포함해야 하며, 이를 통해 AI 모델 학습의 정확도를 높일 수 있습니다. 다양하고 포괄적인 학습 데이터셋은 AI 모델이 더 깊이 있는 지식을 갖추도록 하고, 편향된 정보로 인해 출력이 왜곡되는 현상을 방지합니다. 예를 들어, 인구 통계적·언어적 다양성을 확보하면 AI 도구는 다양한 집단에서 보다 균형 잡힌 성과를 발휘할 수 있습니다.

AI 모델은 어떻게 학습할까?

AI 모델의 학습은 적절한 데이터에 달려있습니다. 예를 들어, AI 모델이 개와 고양이를 구별하도록 학습될 경우, 모델의 정확도는 학습에 사용된 데이터의 양과 유형에 따라 달라집니다.

» 데이터의 양. AI 모델에 개 사진 5,000장과 고양이 사진 5,000장을 제공하면, 사진을 각각 50장씩 제공했을 때보다 더 정확하게 작동합니다.

» 데이터의 품질. 개와 고양이 얼굴의 클로즈업 사진과 전신 사진을 모두 제공한다면 모델의 정확도가 더 높아집니다. 반면, 발, 귀 또는 신체의 일부만 찍힌 사진으로 학습할 경우 정확도가 떨어집니다.

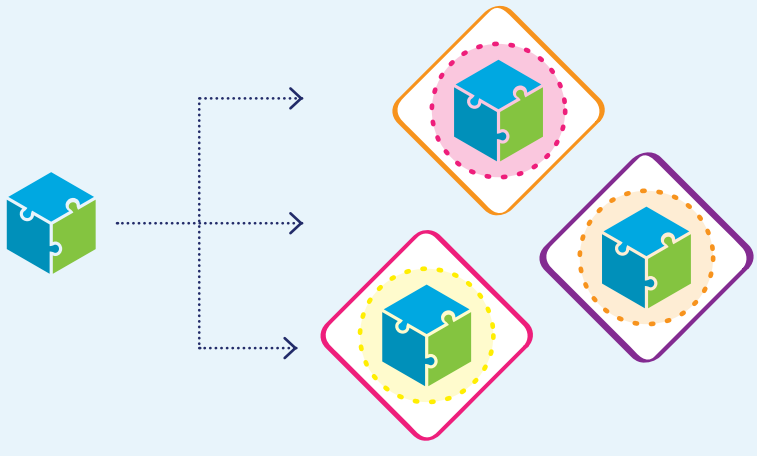
» 데이터의 다양성. AI 모델은 다양한 품종의 개와 고양이를 보여주는 데이터로 학습했을 때 가장 효과적으로 작동합니다. 반대로, 모델이 래브라도 리트리버 사진만 학습했다면, 모든 품종의 이미지가 포함된 데이터셋으로 학습한 모델에 비해 치와와와 같은 소형견과 고양이를 구별하기 어려울 것입니다.

» 특화 데이터. 모델이 학습을 마친 후에는 특정 목적에 맞게 미세 조정할 수도 있습니다. 예를 들어, 사용자가 희귀한 벵갈 고양이만 식별하도록 모델을 활용하고자 할 경우, 수백 장 이상의 벵갈 고양이 사진을 추가로 제공하여 모델을 미세 조정함으로써, 더 구체적이고 맞춤화된 작업에 대한 이해도를 높일 수 있습니다.

AI 모델 학습은 다양한 AI 시스템의 발전에 기여합니다

많은 AI 모델은 '범용 도구'로 개발됩니다. 이러한 범용 모델은 이후 특정 목적에 맞게 설계된 다양한 AI 시스템의 핵심 기반이 됩니다. 예를 들어, 하나의 대규모 언어 AI 모델은 번역 앱, AI 챗봇, 소프트웨어 코드를 개발하는 AI 시스템 등 수백 개의 다양한 AI 애플리케이션에 활용될 수 있습니다.

따라서, AI 모델의 학습을 촉진하는 것은 해당 AI 모델 자체의 발전뿐 아니라, 이를 기반으로 구축되는 다양한 특화형 AI 시스템에도 도움이 됩니다.



AI 모델 AI 시스템