

# 데이터 운용 관리 체크리스트

본격적인 인공지능(AI)의 시대가 열렸습니다. 머신러닝, 딥러닝과 같은 데이터 중심의 기술이 비즈니스 수행 방식에서 중요한 역할을 하고 있으며, 새로운 AI 툴을 제대로 활용하기 위해 귀사의 데이터 “하우스”가 제 기능을 하고 있는지 점검해볼 필요가 있습니다.

완벽한 데이터 하우스 관리를 위해 먼저 검토해야 할 사항들이 있습니다. 이 체크리스트는 데이터 운용 관리(data housekeeping)의 2가지 핵심 단계인 훈련과 추론으로 나뉩니다.

이 단계를 수행하여 AI 마스터가 될 수 있습니다. AI의 개념 검증 단계에서 본격적인 프로덕션 모드로 전환하고 더 큰 규모로 확장할 방법에 대해서는 IDC 보고서, [AI에 최적화된 인프라에서 더 빨리 AI를 구축](#)

## 훈련

AI 준비를 위한 훈련 단계에서는 데이터 세트를 이해하기 위한 알고리즘을 개발합니다. 여기서는 기존 데이터를 수집하고 AI를 활용하여 새로운 기능을 학습하는 것이 관건입니다.

- AI를 사용하여 해결할 구체적인 비즈니스 문제를 파악합니다. 효과적인 학습을 위해 작은 프로젝트부터 시작합니다.
- 문제 해결을 위한 데이터를 관련 소스에서 찾아냅니다. (모든 데이터가 한곳에 있는 경우는 거의 없습니다.)
- 관련 데이터를 찾는 데 걸리는 시간을 최대한 줄이기 위해 데이터에 메타데이터 태그를 추가하면서 준비합니다.
- 사용할 모든 데이터 세트에서 데이터가 제대로 동기화되었고 연결되었는지 확인합니다. (시간 동기화도 포함됩니다.)
- 고객에게 중요한 데이터 및 기타 개인 정보는 따로 플래그를 지정하여 반드시 안전하게 보호하고 모든 거버넌스 및 규정 요건을 이행할 수 있게 합니다. (메타데이터 태그 지정 프로세스가 도움이 될 수 있습니다.)
- 사용하려는 데이터의 유형 및 그 형식에 적합한 개발 환경을 선택합니다. (이를테면 이미지, 비디오, 자유 형식 텍스트, 오디오 등은 저마다 알맞은 환경이 있을 것입니다.)
- 리포지토리의 데이터 세트를 개발 환경으로 가져옵니다.
- 더 원활한 모델 개발 프로세스를 위해 데이터를 두 그룹으로 나눕니다. 각각 “train” 폴더와 “test” 폴더에 둡니다.
- 데이터의 원래 위치 및 소스를 파악하는 방식으로 데이터를 추적 가능한 상태로 유지합니다. 이 프로세스를 자동화하는 툴을 사용하는 것도 좋습니다.
- 기본적인 데이터 정제화(data hygiene) 작업을 수행하여 모델 개발을 위한 데이터를 준비합니다. 이를테면 누락된 데이터 항목을 채우고 null 상태의 항목을 제거합니다.
- 예측에 대한 답을 이미 알고 있는 데이터의 서브셋 샘플(“훈련 세트”)을 사용하여 예측을 위한 데이터 준비에서 거쳐야 할 사전 처리 단계를 모두 확인합니다.
- 이 훈련 세트의 지식을 활용하여 정확도 점수를 계산합니다. 그 결과에 따라, 지금까지 이 모델이 명시적으로 훈련받지 않았던 새로운 데이터에도 자신 있게 모델을 적용할 수 있습니다.

## 추론

비즈니스 문제 해결에 사용할 모델을 개발했다면 훈련에서 추론으로 진행할 차례입니다. 이 단계에서는 성공적인 모델을 얻고 그 모델을 새로운 데이터에 적용하게 되는데, 그러기 위해서는 지속적인 데이터 운용 관리도 필요합니다.

- AI 모델을 데이터에 가까운 곳에 두어 레이턴시를 단축하고 대역폭 요구사항을 줄이며 전반적인 모델 성능을 향상시킵니다.
- 효율적인 데이터 파이프라인 프로세스를 개발하고, 유입되는 데이터에 메타데이터 레이블을 지정합니다. 이러한 방법으로 새로운 데이터를 수집한 다음 지속적인 모델 업그레이드에 활용할 수 있습니다.
- 데이터를 연결하고 동기화할 수 있도록 태그를 지정합니다. 예컨대 데이터에 타임 시퀀스를 부여한 경우, 여러 데이터 세트를 포괄하여 동기화하거나, 유입되는 모든 데이터를 대상으로 ‘고객의 이름’과 같은 필드 하나를 선택하여 연결할 수도 있습니다.
- 장기적인 데이터 라이프사이클 보관 계획을 마련합니다. 여기서는 데이터가 유입되고 보관되는 양과 속도를 효과적으로 관리할 방법을 결정합니다.
- 최고 데이터 책임자(CDO)를 두고 미래의 AI, 딥러닝, 기타 데이터 기반 프로젝트를 위해 조직의 데이터를 지속적으로 관리할 책임을 부여할 수도 있습니다.