

# 7 AI 성공을 위해 꼭 알아야 할 데이터 인프라 핵심 질문

이 핵심 질문들을 통해 AI 도입을 간소화하고, ROI를 극대화하는 비즈니스  
혁신을 달성하세요





## 목차

|                                                                                 |    |
|---------------------------------------------------------------------------------|----|
| AI를 도입 성공을 위한 데이터 토대가 잘 마련되어 있나요?.....                                          | 3  |
| 1. 아직 아무도 생각해내지 못한 용도를 포함해 다양한 생성형 AI 활용 사례들이<br>우리에게 어떤 이점을 가져다줄 수 있을까요? ..... | 5  |
| 파인튜닝과 RAG: AI가 데이터를 '활용'할 수 있도록 지원하는 두 가지 방법 .....                              | 6  |
| 2. 비정형 데이터와 반정형 데이터를 모두 AI에서 활용할 수 있도록 하려면<br>어떻게 해야 할까요? .....                 | 7  |
| 3. 단일한 거버넌스 및 액세스 제어 전략을 통해 AI에 사용될 민감 데이터를<br>어떻게 보호할 수 있을까요? .....            | 8  |
| 생성형 AI 아키텍처: 컨버지드 데이터베이스 vs 다수의 단일 목적 데이터베이스<br>.....                           | 9  |
| 4. 다수의 유관 부서가 어떻게 표준 사항을 준수하면서도, 공통의 생성형<br>AI 전략을 위해 협업할 수 있을까요? .....         | 10 |
| 5. AI를 위한 데이터 가용성을 극대화하기 위해 클라우드 제공업체들의 강점을<br>어떻게 결합할 수 있을까요? .....            | 11 |
| 6. 파인튜닝 및 추론에 필요한 시스템을 어떻게 조달, 관리 및 비용 운영 해야할까요?<br>.....                       | 12 |
| 7. AI 계획에 대한 경영진의 지지를 확보했나요?<br>어떤 그룹과 협력할 것인가요? .....                          | 14 |
| Oracle이 도와드리겠습니다 .....                                                          | 17 |

# AI를 도입 성공을 위한 데이터 토대가 잘 마련되어 있나요?

Jeffrey Erickson  
Senior Writer

상상해 보세요. 이미 양자 물리학과 프랑스 소설을 이해하고 요약하는 데 능수능란한 대규모 언어 모델(LLM)이 귀사 고유의 운영 데이터 및 지식 기반에 대한 깊이 있는 전문성까지 얻게 된다면, 과연 어떨까요? 이제 데이터 검색, 결합 및 컨텍스트화는 더 이상 분석가의 업무가 아닙니다. 대신 AI 에이전트가 비즈니스 사용자와 자연어로 대화하며 인사이트를 발굴하고 즉시 실행에 옮겨주죠. 덕분에 IT 인력은 다른 중요한 업무에 활용될 수 있게 되며, 조직은 더욱 데이터 중심으로 변화하게 됩니다.

정말 멋지지 않나요? 그래서 세계에서 가장 야심차고, 많은 인재를 보유한, 게다가 자금력까지 풍부한 기술 기업들과 스타트업들이 고객에게 생성형 AI의 힘을 활용할 수 있도록 하기 위해 서로 경쟁하는 건 당연한 일입니다. 과장일 수도 있지만, 생성형 AI 모델을 통해 실현할 수 있는 데이터의 궁극적 가치 수준은 실로 엄청나다고 해도 과언이 아닙니다. 이는 기업 전반에 걸쳐 사람들이 데이터를 활용하고 액세스하는 방식을 근본적으로 바꿔놓을 것이기 때문이죠.

이 모든 일을 가능하게 하는 토대는 생성형 AI의 니즈를 충족하는 인프라입니다. 맞춤 도구와 기반 LLM, 기술 그리고 컴퓨트 시스템이 적절한 조합을 갖춘 인프라만이 빠른 응답을 제공하고, 이상 감지, 객체 인식과 같은 다양한 사용 사례를 지원할 수 있죠. 복잡한 AI 작업을 처리할 충분한 성능을 갖추었는지에 따라 성공 여부가 판가름 날 수 있습니다. 예를 들어, 초고성능 및 마이크로초 수준에 불과한 지연성을 가지는 클러스터 네트워크에 상호 연결된 GPU가 대표적인 사례입니다. 벡터 베이스 또한 필요합니다. 고객이 보유한 다양한 엔터프라이즈 지식 기반을 LLM과 연결하여 검색 증강 생성(RAG) 기능을 제공하는 AI 에이전트를 구축해야 하며, 이와 함께 사용자 친화적인 AI 에이전트 인터페이스도 구현해야 합니다

생성형 AI 모델이 데이터의 가치를 얼마나 크게 증폭시킬 수 있는지는 아무리 강조해도 지나치지 않습니다.

## 에이전틱 AI란?

에이전틱 AI는 정보를 이해하고, 이해한 정보에 대응할 수 있을 뿐만 아니라, 적극적으로 목표 달성을 추구할 능력까지 지닌 인공지능을 의미합니다.

### 에이전틱 AI의 주요 특징



#### 사전 예방적 행동

에이전틱 AI는 단순히 외부의 프롬프트에 대응하는데 그치지 않고 스스로 행동을 시작할 수 있습니다.



#### 목표 지향적 행동

에이전틱 AI는 구체적인 목표를 가지며, 목표 달성을 위한 단계들을 스스로 구상할 수 있습니다.



#### 적응력

에이전틱 AI는 경험을 바탕으로 학습할 수 있으며, 피드백을 수용하고, 보다 효과적인 목표 달성을 위해 상황에 적응할 수 있습니다.



#### 자율성

에이전틱 AI는 주어진 매개변수 내에서 독립적으로 의사결정을 내리고 행동에 나설 수 있습니다.

생성형 AI는 무궁무진한 가능성을 약속합니다. 하지만 그 미래에는 인프라를 위한 투자가 선행되어야 하죠.

AI 지원에 필요한 데이터 센터의 규모에 대한 통계를 접해보셨을 겁니다. 대부분은 고객에게 훌륭한 AI 경험을 제공하기 위하여 하이퍼스케일 클라우드 제공업체에서 지출하며, 일부 리딩 기업들 또한 앞으로를 위한 대비 방안으로 적극적으로 데이터 인프라를 준비해 왔습니다.

그렇다면 지금 우리는 어떤 준비를 해야 할까요? 다음은 현재와 미래의 AI 가능성을 실현하는 과정에서 성공을 위해 많은 IT 아키텍트들이 자주 묻는 7가지 핵심 질문입니다.

# 1 아직 아무도 생각해내지 못한 용도를 포함해 다양한 생성형 AI 활용 사례들이 우리에게 어떤 이점을 가져다줄 수 있을까요?

고도로 개인화된 마케팅, 뛰어난 성능의 고객 서비스용 챗봇, 생산성을 높여주는 코딩 헬퍼 등 소수의 AI 프로젝트만으로도 관련 인프라 투자 비용을 충분히 정당화 할 수 있습니다. 잠재적 활용 사례는 매일 늘어나고 있습니다. 상상이 현실로 구현될 가능성이 높아졌기 때문입니다. 이 모든 것이 가능한 이유는 첨단 LLM 덕분입니다. 첨단 LLM은 이미 ERP, HCM, SCM 등 기업용 애플리케이션에 내장되어 있으며, 로보틱스, 헬스케어, 유전체학, 항공우주 등 연구 집약적 분야에서도 널리 활용되고 있습니다. 거의 모든 영역에서 사용되는 것이나 다름 없죠.

자, 이제는 이러한 LLM을 보완하여 AI를 귀사 비즈니스의 최고 전문가로 양성해야 합니다. 기업이 그동안 누적 관리해온 기록, 운영, 재무, 문서 데이터에 대한 접근 권한을 제공하면, 생성형 AI는 다양한 방식으로 기업 내 많은 사람들의 성과를 극대화하는 촉매제 역할을 할 것입니다.

그럼 어떤 모습이 될까요? 경영진들이 데이터와 '대화'하며 세부 내용을 파악하고, 이 과정에서 새로운 비즈니스 인사이트를 얻을 수 있게 됩니다. AI 에이전트가 소프트웨어 개발과 R&D 브레인스토밍 세션에서 파트너 역할을 하며 순식간에 프로토타입과 목업을 제작해줄 수도 있습니다.

AI 덕분에 영업 담당자들은 더 많은 영업 기회를 효율적으로 관리할 수 있게 됩니다. 잠재 고객 발굴, 커뮤니케이션, 각종 업무 프로세스가 한층 높아진 자동화 기능의 혜택을 받을 수 있기 때문입니다. 이러한 변화는 극히 일부분에 불과합니다. 제품, 재무, HR, 법무 팀의 브레인스토밍 세션에서도 생성형 AI가 데이터 인프라 계획을 위한 창의적인 활용 사례와 경쟁 우위의 출발점을 제공하게 될 가능성이 매우 높습니다.

동종 업계의 사용 사례가 궁금하신가요?  
유전학적 분석 뿐 만 아니라 스포츠 중계에  
이르기까지 [다양한 케이스](#) 를 둘러보세요.

## 파인튜닝과 RAG: AI가 데이터를 '활용'할 수 있도록 지원하는 두 가지 방법

파인튜닝과 RAG는 생성형 AI 모델이 주어진 상황에 따라 사용자의 조직에 가장 적합한 응답을 제공할 수 있도록 합니다. 둘 사이에는 다음과 같은 차이점이 있습니다.



파인튜닝이란 Cohere의 Command, Meta의 Llama 3와 같은 범용 모델을 가져다가 비교적 작은 규모의, 도메인별 데이터 세트에 대한 추가 훈련을 여러 차례 진행하는 것을 의미합니다. 파인 튜닝은 코딩, 재무, 헬스케어 등 특정 도메인의 뉘앙스와 용어에 맞춰 훈련되므로 해당 작업에 대한 모델의 성능을 높여줍니다. 유일한 단점은, 비용과 GPU 기반 인프라가 필수적이라는 것입니다.



**RAG**는 특정 조직에 유의미한 출력을 제공하는 범용 AI 모델 지원용 아키텍처 프레임워크입니다. RAG는 쿼리에 대한 답변을 생성할 때 선택된, 유관한 데이터를 LLM에 제공합니다. 그 결과 LLM의 유창한 언어 능력과 조직 내부의 데이터로부터 확보한 인사이트를 결합해 보다 정밀하고 상황에 적합한 응답을 제공하는 AI 시스템이 탄생합니다. AI 모델 파인튜닝과는 달리 RAG는 기본 모델 수정 없이도 작동합니다. 게다가 LLM은 쿼리가 완료된 이후에는 제공된 데이터를 '잊어버리므로', 잠재적인 데이터 유출의 위험 또한 제거됩니다.

RAG와 파인튜닝 중 무엇이 귀사에 적합할까요?

[선택하는 방법 알아보기](#)



## 2 비정형 데이터와 반정형 데이터를 모두 AI에서 활용할 수 있도록 하려면 어떻게 해야 할까요?

AI가 비정형 및 반정형 데이터를 활용할 수 있게 하려면 데이터 수집 및 통합, 저장, 처리(클렌징, 특징 추출, 정상화, 세그먼트화 등)를 포함한 여러 단계의 프로세스를 거쳐야 합니다. 그래야만 데이터 레이크에 저장된 이미지, 문서, 음성, 영상 파일들을 파인튜닝, 벡터화, RAG에 사용할 수 있기 때문이죠.

**파인튜닝용 데이터: 특정 작업용 생성형 AI 모델의 파인튜닝을 위해서는 분야별로 데이터를 새롭게 제공해야 합니다.** 예를 들어, 의학 보고서 생성을 위해 생성형 AI 모델을 튜닝 중이라면, 관련한 올바른 용어와 맥락을 학습할 수 있도록 관련성이 높은 의학 보고서 데이터 세트가 필요합니다.

**지속적인 AI 출력을 위한 데이터:** AI를 지원하는 데이터 관리 시스템은 JSON 문서, 관계형 데이터, 반정형 데이터 등 광범위한 데이터 저장소에 사용자의 LLM이 접근할 수 있도록 해야합니다. 또한 벡터 임베딩을 데이터에 효율적으로 적용하며, 데이터베이스에 임베딩을 벡터로 저장할 뿐 만 아니라, 효율적인 [벡터 검색](#) 및 자연어 처리를 위해 데이터베이스를 쿼리할 수도 있어야 합니다. 이는 지속적인 데이터 수집을 필요로 하며, AI 모델 출력의 정확도 및 관련성을 향상시키는 데 도움이 되는 RAG 아키텍처를 포함한 통합 파이프라인이 필요할 수 있습니다.

이 시스템들이 모두 제자리에 배치되면 어떤 모습일까요? 예를 들면, 영업 팀이 고객과의 음성 통화를 모두 저장한 뒤, AI에게 통화 속 고객의 기분을 분석하는 동시에 통화 요약을 생성하도록 지시할 수 있습니다. 한 손으로는 전화기를 붙들고 고객과 통화하며 관계를 형성하기 위해 애를 쓰면서, 다른 한 손으로는 알아볼 수도 없을 메모를 끄적일 필요가 없다는 얘기죠. 고객의 다음 번 주문 건적 요청에서 세부 정보를 빠뜨릴 일도 없습니다. 세부 내용 하나하나에 깃든 세심한 관심은 전과 같지만, 훨씬 더 정확하고 효율적으로 업무를 처리할 수 있게 됩니다.

**성공의 문을 열 유일한 열쇠:** 하나로 통합된 멀티모달 데이터베이스가 특수 시스템 전반에서의 데이터 이전, 재동기화, 재보안 없이도 모든 프로세스를 진행할 수 있게 해 줍니다. 이와 같은 통합 데이터 모델은 중요도 순서로 작업 속도를 비약적으로 높여주어, IT 팀이 이질적인 사일로로 통합하지 않아도 AI 벡터를 포함한 데이터 유형 전반에 대한 검색을 가능하게 합니다. AI 모델이 겉보기에는 관련 없어 보이는 수많은 핵심 정보들 사이의 미묘한 관계를 파악함에 따라서, 최종 검색 결과는 더욱 풍성해질 수 있습니다.

### 3 단일한 거버넌스 및 액세스 제어 전략을 통해 AI에 사용될 민감 데이터를 어떻게 보호할 수 있을까요?

생성형 AI 도구에 제공되는 데이터의 보안과 프라이버시 보호는 파인튜닝과 RAG뿐만 아니라 일상적인 사용 시에도 반드시 지켜져야 할 핵심 인프라 설계 원칙입니다.

생성형 AI 플랫폼에서 사용되는 데이터는 사용자 요구에 따라 익명화, 암호화, 마스킹 처리될 수 있습니다. AI 출력 생성에 사용되는 데이터 역시 요청자의 권한을 고려하여, 다양한 인증 절차를 통해 엄격하게 관리해야 합니다. 일부 기업은 전용 인프라나 온프레미스 데이터센터에 구축한 자체 AI 모델로 추론을 수행하여 민감한 데이터의 보안을 한층 강화하기도 합니다.

도움이 필요한 경우 담당 기술 제공업체 또는 표준 기관에 문의할 수 있습니다.

예를 들어, Oracle은 인공지능 안전 연구소 컨소시엄(Artificial Intelligence Safety Institute Consortium, AISIC)<sup>1</sup>을 통해 미국 국립 표준 기술 연구소(NIST)와 협업하는 전 세계 280여 기업 중 하나입니다. AISIC의 목표는 'AI 측정 및 정책을 위한, 과학 기반의 실증적 지침과 표준을 개발해 전 세계에 AI 안전을 위한 토대를 마련하는 것'입니다. 회원사들은 지침, 도구, 방법론, 프로토콜, 모범 사례를 함께 개발해 글로벌 공동체가 AI를 안전한 방식으로 개발 및 배포할 수 있도록 지원할 예정입니다.

현재 [AI를 중심으로 한 이와 같은 노력들이](#) 진행 중입니다.

<sup>1</sup> [NIST 인공지능 안전 연구소 컨소시엄\(Artificial Intelligence Safety Institute Consortium\)](#)



## 생성형 AI 아키텍처: 컨버지드 데이터베이스 vs 다수의 단일 목적 데이터베이스

일반적으로는 멀티모달 데이터베이스를 사용하면 생성형 AI 활용을 위한 광범위한 ETL 및 데이터 이전은 필요 없습니다. 오히려 하나의 데이터 플랫폼을 사용해 다양한 AI 프로젝트 및 워크로드를 지원할 수 있습니다. 덕분에 복잡성과 관리 오버헤드를 크게 줄이고, 동시에 데이터 이전이 야기할 수 있는 위험을 최소화할 수 있죠. 하나의 멀티모달 데이터베이스를 사용한다는 건 데이터의 형식이나 모델이 하나여야 한다는 뜻이 아닙니다. 다양한 유형의 애플리케이션에 사용할 데이터 아키텍처를 직접 선택할 수 있다는 뜻이죠.

## 4 다수의 유관 부서가 어떻게 표준 사항을 준수하면서도, 공통의 생성형 AI 전략을 위해 협업할 수 있을까요?

충분한 역량을 갖춘 기업은 범용 AI 모델을 자사 목적에 특화된 AI 모델로 개선할 수 있습니다. 사람들은 AI를 활용하고 싶어 합니다. 따라서 AI 활용에 필요한 도구를 제공하면, 생성형 AI 워크플로가 일회성 비공식 IT 프로젝트로 끝나지 않고 기업 전략의 핵심 요소로 자리잡을 가능성이 높아집니다.

성과를 극대화하는 한 가지 방법은 투명성을 제공하고 부서 간 일관성을 높이는 CoE(전문 센터; center of excellence)를 운영하는 것입니다. CoE는 생성형 AI 관련 모범 사례, 도구, 기술을 통합하여 비즈니스 문제 해결에 활용할 수 있는 방향을 제시합니다.



[기업 내에 AI CoE를 구축하는 방법 알아보기](#)

이러한 협업을 가능하게 할 핵심 기술은 파인튜닝과 모델 배포의 기술적 단계에서 협업을 촉진하는 데이터 사이언스 플랫폼입니다. 하이퍼스케일 클라우드 제공업체들은 생성형 AI 플랫폼을 보유하고, 여기서 데이터 사이언티스트들은 협업하며 Cohere, Meta 등 공급업체의 강력한 기반 모델을 개선할 수 있습니다. 또한 이런 플랫폼에서는 데이터 사이언티스트와 IT 팀이 서비스 전반에 걸쳐 모델, 데이터셋, 데이터 라벨을 저장하고 공유하며 재사용할 수 있습니다.

하이퍼스케일러들은 또한 모델 카탈로그를 제공하고, 파인튜닝된 LLM에 데이터와 이를 실행할 수 있는 컴퓨트 인프라를 제공합니다. 이제 각 사업부는 기업 조직 내 타 부서가 거둔 과거의 AI 성공 사례에 함께할 수 있게 됩니다.

CoE와 선도적인 데이터 과학 플랫폼을 모두 갖춘 조직은 지속적인 학습 및 개선의 문화를 조성할 수 있습니다. Oracle을 포함한 제공업체들은 광범위한 사용 사례에 LLM을 유기적으로 통합할 수 있게 해 주는, [Oracle Cloud Infrastructure \(OCI\) Generative AI](#)와 같은 완전 관리형 서비스를 제공합니다. 자체 데이터 세트를 활용해 기본 모델을 파인튜닝하여 커스텀 모델을 생성할 수도 있습니다.

## 5 AI를 위한 데이터 가용성을 극대화하기 위해 클라우드 제공업체들의 강점을 어떻게 결합할 수 있을까요?

현재 Google Gemini나 Microsoft의 Copilot을 고려 중이신가요? 아니면 AWS에서 오픈 소스 모델을 또는 OCI에서 Cohere의 기반 모델을 호스팅하기로 결정했을 수도 있겠습니다. 기업의 온프레미스 또는 클라우드 데이터 거버넌스 전략이 Oracle Database 기술을 중심으로 구축된 기업들에게는 이제 Oracle과 타 하이퍼스케일러 간의 새로운 협약으로 인해 폭넓은 가능성의 세계가 펼쳐집니다. OCI에서 뿐만 아니라 각 제공업체의 데이터센터 내에서 Oracle Database 서비스를 사용할 수 있게 해 주는 Oracle과 Azure, Google Cloud, AWS 간의 혁신적인 멀티클라우드 관계 덕분에 고객은 원하는 모델을 자유롭게 선택할 수 있고, 낮은 지연성 및 관리 오버헤드의 이점도 누릴 수 있습니다.

귀사가 구상한 바로 그 AI 모델을, 원하는 곳에서 실행할 수 있게 된다면, 그리고 데이터에 쉽고 안전하게 액세스할 수 있다면 지금은 상상조차 하지 못할 생성형 AI 사용 사례들을 바탕으로 성공을 거둘 가능성이 더욱 높아집니다. 각 팀은 각자의 작업에 적합한 최고의 서비스를 활용할 수 있고, 보안과 회복 탄력성은 유지하면서, 동시에 비용은 절감할 수 있게 됩니다.





## 6 파인튜닝 및 추론에 필요한 시스템을 어떻게 조달, 관리하고 비용까지 감당할 수 있을까요?

생성형 AI의 파인튜닝 및 배포는 컴퓨트 집약적인 작업입니다. 각각의 상호 작용에는 엄청난 양의 데이터 세트에 대한 복잡한 계산이 필요할 수 있습니다. 이때 수십억 개의 매개변수를 포함하는 LLM이 정보 조작 및 분석 역량을 강화하기 위해 특수 컴퓨팅 시스템을 활용합니다. 모델의 복잡성이 커질수록 데이터를 처리하고, 유의미한 출력을 제공하는 데 필요한 연산의 양도 많아집니다. 따라서 모델 파인튜닝 및 추론에 드는 컴퓨트 비용을 낮추면서도, 고품질의 결과를 제공하기 위한 계획이 필요합니다.

많은 기업들은 시스템을 자체 개발하는 데 드는 비용과 전문성을 감당하는 대신, 데이터셋을 정제하고 특정 작업에 적합하게 만들기 위해 주요 클라우드 제공업체의 제품 등 잘 알려진 데이터 사이언스 플랫폼을 선택합니다.

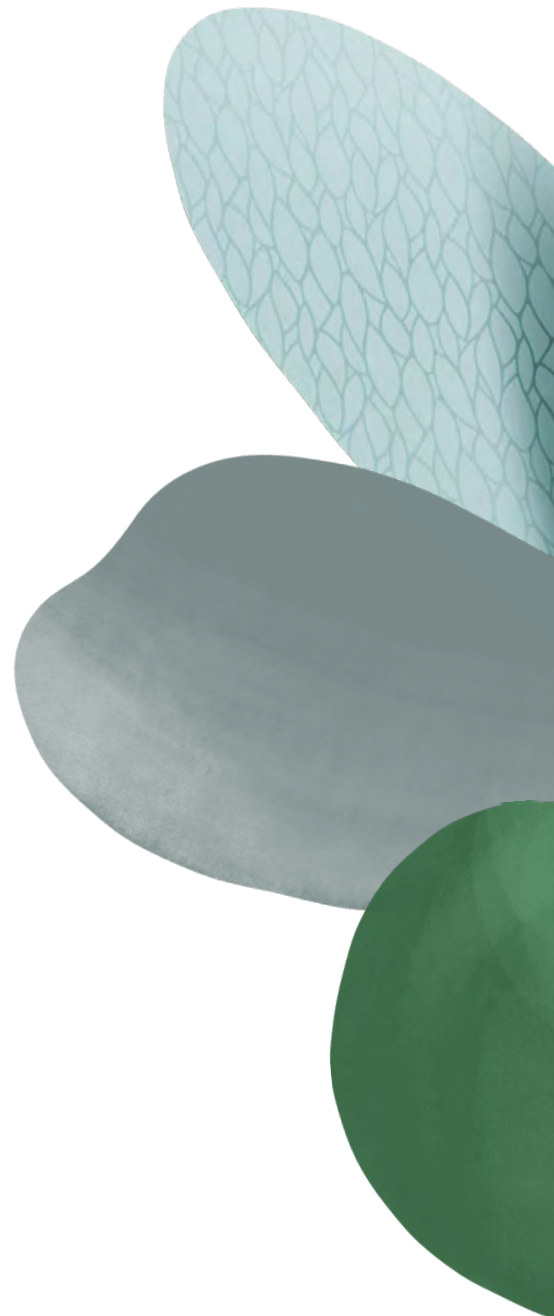
그다음, 선택한 클라우드 제공업체의 AI 인프라 서비스를 이해해야 합니다. 널리 사용되는 오픈 소스 도구, 생성형 AI 기반 모델 및 기타 필요한 기술들은 대체로 손쉽게 접근하여 활용하거나 통합할 수 있습니다.

오픈 소스 커뮤니티와 하이퍼스케일러들은 생성형 AI의 비용을 낮추기 위해 많은 노력을 기울이고 있습니다. 클라우드에서는 AI 워크로드 가속화에 필요한 강력한 GPU 및 클러스터 네트워킹을 활용할 수 있고, 자동 확장, 스팟 인스턴스, 예약된 인스턴스, 최적화된 작업 일정 관리, 효율적인 워크로드 배포, 멀티테넌시, 모델 최적화와 같은 기능들도 제공됩니다. 이와 같은 기술들은 사용자가 원하는 시점에 언제든지 워크로드 수요에 컴퓨트 파워를 정확히 일치시킬 수 있어 리소스 활용도를 극대화할 수 있습니다. 기업은 사용한 만큼의 리소스에 대해서만 비용을 지불하면 되죠.

## 파인튜닝 역시 점점 정교화되고 있습니다.

T-Few라고 알려진 이 방법은 기존 파인튜닝 방식 대비 짧은 훈련 기간 및 적은 컴퓨팅 리소스를 필요로 합니다. 하지만 여전히 높은 정확도를 자랑하죠. 마지막으로, 선택한 모델과 인프라가 가장 효율적인 방식으로 작동하는 지 입증할 수 있어야 합니다.

클라우드 제공업체의 생성형 AI 서비스 또는 TensorFlow Serving, TorchServe와 같은 **효율적인 모델 서비스 프레임워크를 통해 추론 프로세스를 모니터링 및 최적화할 수 있습니다.**





## 7 우리 AI 계획에 대한 경영진의 지지를 확보했나요? 어떤 그룹과 협력할 것인가요?

귀사의 AI 전략은 현재 사용 중인 엔터프라이즈 시스템을 바탕으로 시작될 가능성이 매우 높습니다. CRM, HCM, ERP 및 [기타 다양한 핵심 애플리케이션](#)들에 생성형 AI 기능이 내장되고, AI 에이전트가 공통의 워크플로에 적용되는 방식이죠. 지능형 문서 인식 기능 덕분에 공급업체 송장을 더욱 빠르게 처리하고, 수작업의 필요성을 크게 줄일 수 있게 된다면 어떨까요? 생성형 내러티브는 보고 및 분석 내용을 더욱 깊이 이해할 수 있게 지원할 것이고, 위험 및 성과 분석과 같은 영역에서는 광범위한 지능형 자동화가 제공될 것입니다. 이와 같은 기능들이 앞으로의 발전 방향을 제시할 테지만, 저마다의 고유한 워크플로에 AI를 적용하기 위해서는 보다 광범위한 전략이 필요합니다. 앞서 논의한 CoE가 여기에 포함될 수 있죠.

생성형 AI 서비스와 AI 에이전트를 비즈니스 운영에 적용하기 위해서는 경영진, 부서장, IT, 법무 및 규제 준수 팀을 포함한 조직 전반의 승인과 리소스가 필요합니다. 생산성 향상 및 경쟁 우위에 중점을 두고 각 업무 부서의 이해관계자들과의 협업을 통해 비즈니스 사례를 개발하면 조율된 접근 방식을 마련할 수 있습니다.

AI 워크플로의 기초가 마련되었다면, 이제 IT와 데이터 사이언스 팀과 협업하여 필요한 아키텍처의 유형을 파악할 시간입니다. 이때 다음과 같은 질문들을 고려해볼 수 있습니다: LLM은 어디에 배치될 것인가? 비용은 누가 부담할 것인가? 데이터는 어떻게 흘러야 하는가? 새로운 벡터 데이터베이스나 RAG 아키텍처가 필요한가? LLM을 특정 부서나 업무용으로 파인튜닝해야 하는가? 데이터 스토리지와 컴퓨팅 파워는 어디서 확보할 것인가? 우리의 네트워크 설계와 기타 아키텍처 결정으로 지속적인 AI 추론에 필요한 저지연성과 높은 처리량을 달성할 수 있는가?

일정, 마일스톤, 필요한 리소스 목록과 데이터 프라이버시, 보안 및 규제 준수, 비용 고려 사항 등의 위험 요소들을 포함한 상세 제안서를 준비해야 합니다. 또한 신뢰하는 클라우드 서비스 제공업체로부터 어떤 지원을 받을 수 있는지도 검토해야 합니다.

잘 정제된 데이터의 지속적인 공급이 수반되지 않는다면, 효과적인 AI 모델 구축 또한 불가능하다는 점을 명심해야 합니다

따라서 AI 구현을 귀사 전반에 걸친 데이터 전략의 확장판이라고 생각해야 합니다. 경영진의 공감을 얻기 위해서는 각 팀이 세운 AI 계획이 조직 전반의 비즈니스 전략 및 데이터 거버넌스 계획과 긴밀히 연계되어 있음을 입증해야 할 것입니다. 이때 데이터 이전이나 관리 리소스 추가 등의 불필요한 복잡성이 더해지는 일은 피해야 합니다.



# Oracle 솔루션

기존 데이터 인프라로 최첨단 생성형 AI 모델을 빠르고 경제적으로 파인튜닝하고 운영하고 싶다면, Oracle과 함께하세요.



## [☞ Oracle Cloud Infrastructure](#)



### Oracle AI Infrastructure

OCI는 최고 수준의 AI 인프라로 포괄적인 AI 서비스와 최첨단 생성형 AI 혁신을 제공합니다. 사전 구축된 모델을 선택하거나, 오픈소스 또는 상용 LLM을 활용한 생성형 AI 서비스로 모델을 파인튜닝하고, 자체 기업 데이터로 성능을 향상시킬 수 있습니다. 여기에 업계 최고의 데이터베이스 관리 시스템까지 더해집니다.

[☞ 더 알아보기](#)



### Oracle Database 23ai

Oracle Database 23ai는 트랜잭션과 분석을 위한 멀티모달 데이터베이스로, 내장 벡터 데이터베이스는 물론 JSON, 관계형, 그래프, 지리공간 데이터, 인데이터베이스 머신러닝을 모두 지원합니다. OCI 전용 서비스이자 Oracle 멀티클라우드 파트너 데이터센터에서 OCI 서비스로 제공되는 고유한 최적화 플랫폼을 통해 벡터를 처리하고 Oracle Database를 실행할 수 있습니다.

[☞ 더 알아보기](#)



### OCI Data Science 및 GenAI 서비스

OCI Data Science는 데이터 사이언티스트들이 선호하는 오픈소스 프레임워크에서 협업하며 머신러닝 모델을 구축, 훈련, 배포, 관리하거나 인데이터베이스 ML을 활용할 수 있게 하는 클라우드 서비스입니다. OCI 생성형 AI 서비스는 고객이 선택한 언어 모델을 다양한 활용 사례에 통합하는 완전 관리형 서비스입니다.

[☞ 더 알아보기](#)

OCI에서는 더욱 빠르고 비용 효율적인 파인튜닝, 추론, 배치 처리를 실행할 수 있습니다. OCI는 대규모 AI 워크로드를 더욱 빠르게, 과도한 컴퓨트 비용 없이 실행할 수 있는 확장성과 성능을 제공합니다.

# Oracle은 최적의 파트너입니다

독보적인 멀티클라우드 파트너십을 통해 원활한 데이터 호환을 지원합니다.

전 세계의 많은 기업들이 Oracle Exadata Database Service에서 애플리케이션을 신속히 구축 및 실행하고, Oracle Autonomous Database 기능들의 이점을 누리기 위해 OCI에서 Oracle Database 서비스를 활용하고 있습니다. 우리의 멀티클라우드 파트너십 덕분에 기업들은 타 하이퍼스케일 클라우드에서 실행되는 애플리케이션에서도 Oracle Database가 제공하는 단일 운영 환경의 단순성, 보안성, 저지연성 이점을 누릴 수 있습니다.

AWS에서는 Amazon Bedrock과 같은 서비스를 통해 Exadata Database Service 및 Autonomous Database에서 애플리케이션을 신속히 구축 및 실행할 수 있고, Azure에서는 Azure OpenAI와 같은 서비스를 통해 Exadata Database Service와 Autonomous Database를 실행할 수 있습니다. Google Cloud에서는 Google Cloud의 Vertex AI 및 Gemini 기반 모델과 같은 서비스를 통해 Exadata Database Service 및 Autonomous Database를 실행할 수 있습니다.

## 문의처

한국오라클 대표번호 02-2194-8000, 또는 [oracle.com/kr](https://oracle.com/kr) 웹사이트를 통해 Oracle 담당자에게 연락하실 수 있습니다.

북미 지역 외 국가인 경우 [oracle.com/kr/contact](https://oracle.com/kr/contact)에서 현지 지사를 찾을 수 있습니다.

Copyright © 2025 Oracle, Java, MySQL, NetSuite는 Oracle 및/또는 그 계열사의 등록 상표입니다. 기타 명칭들은 각 명칭을 소유한 기업의 상표일 수 있습니다. 본 문서는 정보 제공 목적으로만 제공되며 본 문서의 내용은 예고 없이 변경될 수 있습니다. Oracle은 본 문서의 무오류성을 보증하지 않습니다. 또한 본 문서에는 상업성 또는 특정 용도 수행을 위한 적합성과 관련된 암시적 보증 및 조건을 비롯한 구두상의 표현 또는 법 규정에 의한 어떠한 보증 또는 조건도 포함되어 있지 않습니다. Oracle은 본 문서로 인한 법적 책임을 일체 지지 않으며, 본 문서로 인한 직접 또는 간접적 계약 구속력 역시 일체 발생하지 않습니다. 본 문서는 Oracle의 사전 서면 승인 없이 전자적, 기계적 및 기타 모든 형태 또는 수단을 통해 복제 또는 전송될 수 없습니다.

