생성형 AI 운용을 위한 경영진 안내서

실험부터 구현까지: 실제 시나리오에 생성형 AI를 사용하는 방법

elastic | The Search AI Company

목차

생성형 AI로 달성하려는 것이 무엇인지 알고 있다면 여기에서 시작하세요

생성형 AI 여정을 시작하세요

생성형 AI(Generative AI)는 2023년에 등장한 가장 혁신적인 기술이었습니다. 산업 전반에 걸쳐 생성형 AI가 앞으로 몇 년 동안 거의 모든 것을 형성할 것으로 예측됩니다. 하지만 지금 생성형 AI를 작업에 활용하기 위해 코드를 해독했다고 말할 수 있는 사람은 몇 명이나 될까요?

기업들이 생성형 AI의 새로운 물결에 대처하는 동안, 일부 기업은 이미 결과를 보기 시작했습니다. 예를 들어, Cisco의 지원 엔지니어는 유사한 지원 사례, 내부 토론 포럼, 고객 문제와 관련된 지식 문서에서 정확도가 높고 요약된 답변을 즉시 찾을 수 있습니다. Cisco는 이미 재구상된 검색 솔루션을 통해 지원 요청의 90%를 해결하고 지원 엔지니어의 월 5,000시간을 절약하는 등 생성형 AI의 이점을 누리고 있습니다.[1]

전자 상거래 영역에서 생성형 AI가 활용되는 것을 보신 적이 있을 것입니다. 생성형 AI는 고객의 과거 구매 내역, 검색 기록, 선호도를 분석하여 챗봇을 통해 맞춤형 제품 추천을 생성할 수 있습니다. 그리고 백엔드에서 생성형 AI를 사용하면 고객 참여 및 유지율을 높이고 사기 탐지 기능을 향상시키는 등의 효과를 얻을 수 있습니다.

생성형 AI의 기능을 이해하고 이를 활용하는 방법을 결정하려면 데이터 활성화에 대한 단계별 안내서가 필요합니다. 이 eBook에서는 **현실이 되기는 어려운 희망 사항을 꿈꾸던 사람에서 AI 전문가**가 되는 여정을 안내해 드립니다. 생성형 AI를 사용하여 비즈니스 성과를 혁신하는 데 도움이 되는 로드맵으로 이 안내서를 생각해 주세요.

99% 의 조직만이 생성형 AI가 내부든 외부든 조직 내에서 변화를 주도할 잠재력이 있다고 믿습니다.[2]

아직은

32% 의 리더만이 조직에서 AI를 구현하는 능력에 자신감을 갖고 있습니다[3]

[1] Elastic, Cisco creates AI-powered search experiences with Elastic on Google Cloud 2024(Google Cloud 2024에서 Elastic을 사용하여 AI 기반 검색 환경을 만드는 Cisco)
[2] Elastic, The Elastic Generative AI Report(Elastic 생성형 AI 보고서)(2024)
[3] Russell Reynolds, Embracing the Unknown: How Leaders Engage with Generative AI in the Face of Uncertainty(미지의 세계를 포용: 불확실성에 직면하여 리더들이 생성형 AI에 참여하는 방법, (2024).

이 안내서에서 기대하실 수 있는 내용은 다음과 같습니다.

0단계

생성형 AI가 어떻게 도움이 될 수 있는지 물어봅니다. 생성형 AI를 통해 무엇을 달성하려고 하시나요?

1단계

이상적인 결과를 식별합니다. 여러분의 사용 사례에 대한 성공적인 구현은 어떤 모습인가요?

2단계

영향을 파악하고 성공 여부를 측정합니다. 어떤 조직 프로세스가 어떻게 영향을 받는지 고려하세요.

3단계

구현 전략을 선택합니다. 여러분의 옵션을 탐색해 보세요.

4단계

테스트를 시작하고 반복적인 접근 방식을 취합니다.

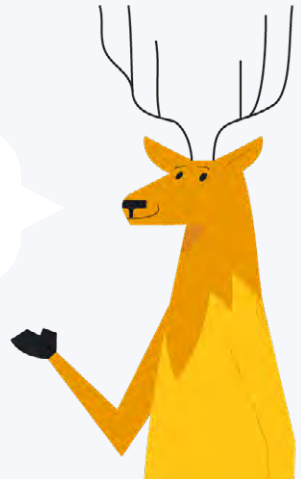
5단계

거버넌스 표준을 확립하고 데이터 안전 문제를 해결합니다.

6단계

팀의 여러 가지 조언을 반영해 타임라인을 설정합니다.

하지만 먼저 기본 사항부터 살펴보겠습니다.



제1부: 생성형 AI 환경 이해

이를 운영하기 위한 계획을 수립하기 위해 생성형 AI 전문가가 될 필요는 없습니다. 그러나 어떤 구성 요소가 작용하고 있는지 이해하면 프로세스 전반에 걸쳐 정보에 입각한 전략적인 결정을 내릴 수 있습니다. 기본 구성 요소를 배치해 보겠습니다.

생성형 AI란 무엇인가?

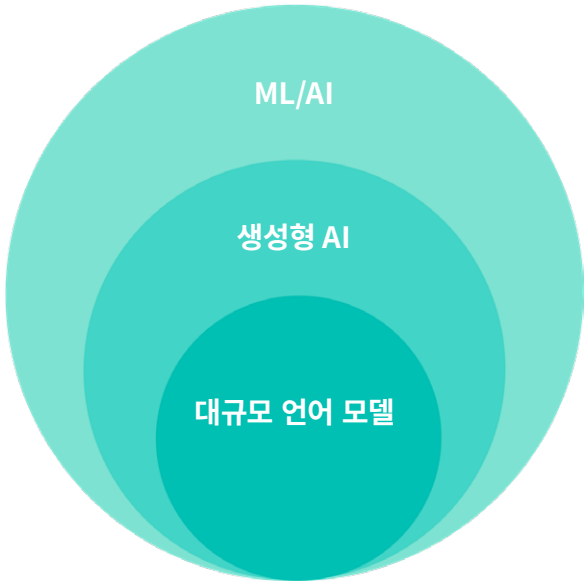
생성형 인공 지능(Generative Artificial Intelligence, Generative AI)은 메시지가 표시될 때 출력을 생성할 수 있는 딥 러닝 모델을 의미합니다. 이 기술의 생성 능력은 머신 러닝의 도움으로 달성되는 통계적으로 가능한 결과를 예측하는 능력에 달려 있다는 점을 이해하는 것이 중요합니다. **데이터는 생성형 AI 운영의 핵심이며 구현과 결과 모두에서 성공의 열쇠입니다.** 이에 대해서는 곧 자세히 설명하겠습니다.

머신 러닝이란 무엇인가?

머신 러닝(ML)은 알고리즘을 사용하여 데이터에서 지식을 도출하는 인공 지능의 하위 집합입니다. 이러한 알고리즘은 데이터를 분석하고 감독, 준지도 또는 비지도 컨텍스트에서 결정을 내릴 수 있는 패턴과 유사성을 "학습"합니다. 머신 러닝은 대규모 언어 모델과 같은 생성형 AI에 지속적으로 '학습'할 수 있는 기능을 제공하는 기반 기술입니다.

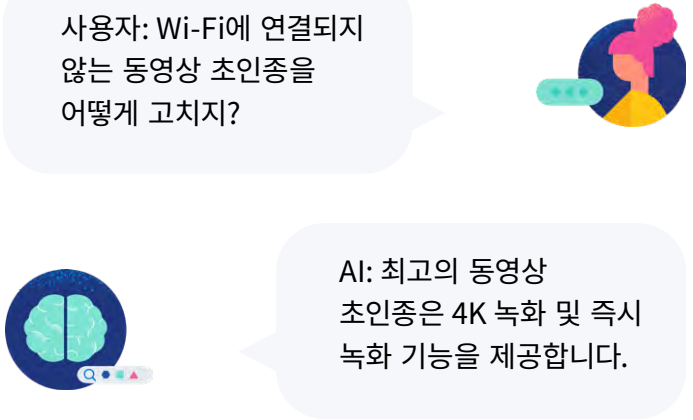
대규모 언어 모델(LLM)이란 무엇인가?

대규모 언어 모델(LLM)은 머신 러닝에 사용되는 계산 모델입니다. 인간의 언어를 구체적으로 다루는 생성형 AI의 일종입니다. 광범위한 공용 언어 데이터 세트에 대한 훈련을 받은 LLM은 텍스트 인식, 분석, 요약, 예측, 번역 또는 생성을 포함한 다양한 자연어 처리(NLP) 작업을 수행할 수 있습니다. 생성형 AI를 운용하는 맥락에서 LLM은 생성형 AI가 자연(또는 인간) 언어로 의사소통할 수 있도록 하는 것입니다.



환각에 대해 이야기해 보겠습니다.

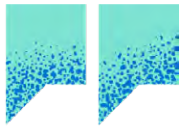
환각은 LLM이 생성한 부정확하거나 오해의 소지가 있는 결과입니다. ChatGPT가 가끔 의심스러운 답변을 하는 것이 현명할 것입니다. 출력은 합법적인 것처럼 보이지만… 실제로 그럴까요? LLM(ChatGPT는 LLM에 의해 구동됨)은 답변을 찾을 수 없더라도, 답변을 만드는 경향이 있습니다. 엔터프라이즈 애플리케이션에서 LLM 사용을 논의할 때 이 사각지대를 고려하는 것이 중요합니다. 생성된 출력이 관련성이 있고 정확하다는 것을 어떻게 보장할까요? 바로 이 부분에서 검색 증강 생성(RAG)이 등장합니다.



검색 증강 생성(RAG)이란 무엇인가?

검색 증강 생성(RAG)을 환각에 대한 방어선으로 생각하세요. LLM에서 생성된 출력은 벡터 데이터베이스에서 제공하는 관련성이 높은 검색을 사용하여 제공하는 특정 데이터 세트 또는 데이터 컨텍스트에서 정보를 검색하여 보강되거나 "확인"됩니다. 예를 들어, RAG를 통해 사용자 쿼리에 대한 응답으로 조직은 정책 문서를 검색하고 LLM에 관련 응답을 제공하므로 조직의 정책을 사용하여 질문에 응답할 수 있습니다. 환각에 대한 방어를 넘어 **RAG를 사용하면 독점 데이터 세트와 함께 생성형 AI를 사용할 수 있습니다**. 이것이 가장 큰 이점입니다.

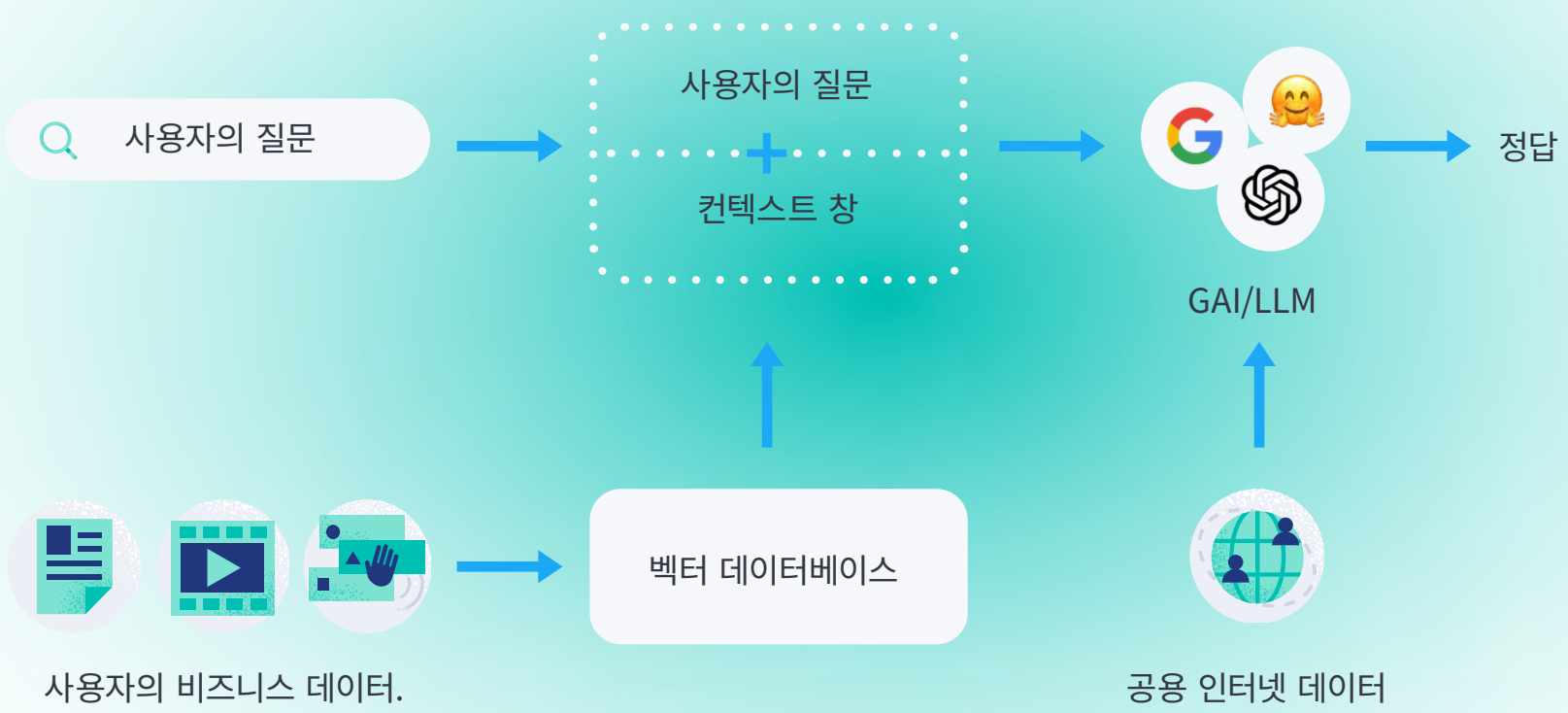
비즈니스 애플리케이션을 위한 생성형 AI를 운영하는 맥락에서 RAG는 여러 가지 이유로 중요합니다. RAG는 더 우수하고 관련성 높은 결과를 제공할 수 있으며 자체 독점 데이터를 부트스트랩하거나 활용할 수 있는 빠른 방법을 제공합니다. 또한 자체 LLM을 교육하거나 구축하는 것보다 비용 효율적입니다. 이러한 방식으로 RAG는 성공적인 생성형 AI 통합의 핵심입니다. RAG는 일반 LLM의 한계를 넘어 "차세대 검색 엔진"을 만듭니다.



> **"RAG는 차세대 검색 엔진을 만드는 데 도움이 됩니다."**
>
> — **Baha Azarmi**
> **Elastic의 글로벌 고객 엔지니어링 부사장**

검색 증강 생성(RAG)

RAG를 사용하면 독점 데이터 세트와 함께 생성형 AI를 사용할 수 있습니다.



RAG는 사용자 질문에 답변하는 새로운 방법입니다.

일반 검색

어떤 사람이 **용어**를 검색합니다

재택근무 정책

쿼리가 수행됩니다

```
query = {
"bool": {
"should": [
{
"text_expansion":{
"ml.inference.text_expanded_predicted_
value":{
"model_id":model_id.
"model_text":question
}
```

결과가 제시됩니다

	문서 제목	추가된 날짜
1	직원 윤리 규범	2010년 1월 1일
2	IT 사용 정책	2015년 1월 1일
3	홈페이지 작업 진행 중	2022년 1월 1일
	기타 등등...	

사용자는 문서를 선택하고 해당 내용을 읽습니다

직원 윤리 규범

Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum

RAG가 없는 생성형 AI

어떤 사람이 **질문**을 합니다

재택근무 정책은 무엇인가요?

쿼리가 수행됩니다

```
query = {
"bool": {
"should": [
{
"text_expansion":{
"ml.inference.text_expanded_predicted_
value":{
"model_id":model_id.
"model_text":question
}
```

여러분의 도메인과 관련이 없는 답변이 파생되었습니다

하이브리드 또는 하이브리드 근무를 하는 직원이 있는 경우 재택근무 정책이 필요합니다.

RAG 지원

어떤 사람이 **질문**을 합니다

재택근무 정책은 무엇인가요?

쿼리가 수행됩니다

```
query = {
"bool": {
"should": [
{
"text_expansion":{
"ml.inference.text_expanded_predicted_
value":{
"model_id":model_id.
"model_text":question
}
}
]
{
"match":{
"text": question
}
```

결과가 컨텍스트로 제공됩니다

	문서 제목	추가된 날짜
1	직원 윤리 규범	2010년 1월 1일
2	IT 사용 정책	2015년 1월 1일
3	홈페이지 작업 진행 중	2022년 1월 1일
	기타 등등...	

답변이 LLM의 검색 결과에서 파생됩니다

직원들이 효과적으로...기타 등등... 을 할 수 있는 경우, 재택근무를 권장합니다.

벡터 데이터베이스란 무엇인가?

벡터 데이터베이스는 벡터 임베딩 또는 단어, 이미지, 동영상의 숫자 표현을 저장합니다. 이러한 임베딩은 다차원적이며 쿼리의 의도와 문맥적 의미를 찾는 검색 유형인 의미론적 검색을 가능하게 합니다. 대조적으로, 텍스트 검색은 검색어의 키워드와 일치하는 결과만 찾습니다.

RAG의 맥락에서 벡터 데이터베이스는 생성형 AI에 제공되는 프롬프트를 기반으로 빠른 시맨틱 검색을 가능하게 합니다. 이것이 바로 RAG를 가능하게 하는 이유입니다.

생성형 AI는 NLP에 능숙하지만 기존 키워드 검색은 자연어를 수용하여 생성형AI에 제공할 최상의 결과를 제공할 수 없습니다. 따라서 생성형 AI에 원래 프롬프트와 의미론적으로 유사한 검색 결과를 제공하는 벡터 데이터베이스를 사용하면 생성형 AI가 더 관련성이 높은 답변을 생성할 수 있습니다. 벡터 데이터베이스를 생성형 AI가 정확한 정보로 질문에 답할 수 있게 해주는 지식 은행이라고 생각하세요.

그러나 생성형 AI는 벡터 데이터베이스에만 국한되지 않습니다. 생성형 AI는 RAG를 사용하여 관계형 데이터베이스, 그래프 데이터베이스, 문서 기반 데이터베이스 또는 키워드 검색 엔진을 활용할 수 있습니다. 여러분에게 가장 적합한 데이터베이스는 데이터의 성격, 사용되는 특정 알고리즘, 시스템의 성능 요구 사항에 따라 달라집니다. 예를 들어, 관계형 데이터베이스는 구조화된 데이터에 사용할 수 있는 반면, 그래프 데이터베이스는 복잡한 관계가 있는 데이터에 적합하며 전체 텍스트 검색에는 기존 검색 엔진에 적합합니다.

"모든 길은 하이브리드 검색으로 이어집니다."

Serena Chou
Elastic 제품 관리 이사

시맨틱 검색은 검색어의 의미와 일치하는 결과를 제공하는 반면, **키워드 검색**은 결과를 검색어의 정확한 키워드와 일치시키는 데 여전히 중요한 역할을 합니다. 하이브리드 검색은 시맨틱 검색에 자주 사용되는 벡터와 키워드 검색 기술을 모두 사용하여 생성형 AI에 가장 관련성이 높은 결과를 제공하는 방식입니다.

결론: 하이브리드 검색 솔루션은 조직의 생성형 AI 경험에 가장 관련성이 높은 결과를 제공할 가능성이 높습니다.

생성형 AI가 수행할 수 있는 작업

기본 기술과 기본 개에 대해 많이 이야기했지만, 생성형 AI가 정확히 무엇을 할 수 있을까요?

생성

생성형 AI는 훈련 데이터의 패턴을 학습함으로써 출력을 생성할 수 있습니다. 기존 데이터를 반복하여 새로운 아이디어, 이미지, 인사이트 등을 생성합니다.

요약

자연어 처리 기능 덕분에 생성형 AI는 텍스트를 분석하고 요약할 수 있습니다. 짧은 시간 내에 긴 문서를 검토해야 하시나요? 생성형 AI가 구조해 드립니다.

탐색

생성형 AI의 핵심은 기본 검색 기술입니다. 이를 통해 생성형 AI 도구가 쿼리를 수신하고 광범위한 비공개 또는 공개 데이터를 검색하고 응답을 생성할 수 있습니다.

자동화

조직에서 서로 다른 서비스에 대해 서로 다른 두 가지 클라우드 플랫폼을 사용한다고 가정해 보겠습니다. 각 클라우드 플랫폼은 서로 다른 형식으로 로그를 생성합니다. 이 데이터를 동일한 형식으로 자동 변환하고 AI로 매핑함으로써 팀은 생성형 AI로 데이터를 요약하고 질문할 수 있습니다. IT 팀은 노동 집약적인 작업을 수행하는 대신 시스템 모니터링 및 관리에 집중할 수 있습니다.

0단계: 이유를 찾고 무엇이 가능한지 알아보기

가치를 추가할 수 있는 방법이 너무 많기 때문에 시작점에 초점을 맞추는 것이 중요합니다. 생성형 AI의 힘을 어떻게 활용하여 팀의 역량을 강화하고, 빠르게 변화하는 고객의 기대에 부응하며, 회사를 새로운 차원으로 끌어올리려면 어떻게 해야 할까요? 생성형 AI가 조직에 가장 큰 가치를 추가할 수 있는 한 영역에 초점을 맞추면 됩니다.

다음 질문을 고려해보세요.

1 어떤 문제를 해결하려고 하는가?

여러분의 비즈니스에서 특히 비효율적인 영역이 있나요? 직원들이 반복적인 작업에 많은 시간을 보내는 곳은 어디인가요? 내부 데이터베이스나 외부 엔진에서 기존 정보를 지속적으로 검색하고 있나요?

예를 들면, 다음과 같습니다.

직원들이 프로젝트 업데이트든 HR 관련 정보든 정보를 찾는 데 어려움을 겪고 계신가요? 보안 팀에 자동화할 수 있는 작업이 있어 보다 적극적인 자세를 취할 시간을 확보하실 수 있나요? 고객 서비스 엔지니어가 알려진 문제 및 최근 수정 사항에 대해 고객 서비스 상담원과 실시간으로 소통하지 않기 때문에 고객 서비스 프로세스에 엔트로피가 있나요?

알려드려요:

영향을 미칠 워크플로우를 담당하는 팀과 협력해야 합니다. 예를 들어, 비효율적인 HR 프로세스를 업데이트하기로 결정한 경우 이해관계자의 승인을 얻고 지원을 강화하려면 처음부터 HR을 도입하는 것이 중요합니다.

2 이것이 지식 기반 시스템으로 해결할 수 있는 문제인가?

지식 기반 시스템은 지원 기사든 내부 프로세스든 정보를 제공하는 콘텐츠의 모음입니다.

어떤 콘텐츠를 가져와야 하는지, 그리고 문제를 보다 효율적으로 해결하기 위해 해당 콘텐츠를 채굴할 수 있는지 여부를 고려하세요. 응답을 자동화하고 개인에게 맞춤화하는 것으로 충분한가요?

예를 들면, 다음과 같습니다. 직원들이 HR 관련 정보를 찾는 데 상당한 시간을 소비한다는 사실을 확인하셨습니다. 여러분의 팀에는 HR 전문가가 없기 때문에 직원들은 회사 인트라넷으로 리디렉션되며, 여기서 직원들은 일년에 남은 휴가 일수를 확인하기 위해 다양한 정책 문서를 읽어야 합니다. 이 문제를 해결하려면 HR 정책 문서와 같은 지식 기반 시스템이 필요하고 직원 개인 데이터에 액세스하여 응답을 개인에게 맞춤화해야 합니다.

"실 하나를 당기면 천 개의 실이 계속 나옵니다."

Baha Azarmi
Elastic 글로벌 고객 엔지니어링 담당 부사장

0단계는 어떤 프로세스가 생산성을 제한하는지 이해하는 것입니다. 단순한 작업을 줄임으로써 직원들이 창의력을 발휘할 수 있는 공간을 확보할 수 있습니다. 잘 실행된 구현은 책임 있는 구현입니다. **직원에게 투자하고 그에 따라 기술을 향상시키며 조정된 작업 흐름과 프로세스를 계획하는 것은 모두 생성형 AI와 성공적인 결합의 중요한 부분입니다.**

그 이유를 검색할 때, **간단하고 구체적으로 유지하세요.** 먼저 처리하고 싶은 생성형 AI 사용 사례를 식별하는 것은 생성형 AI를 운용하는 데 있어 훌륭한 초기 단계입니다. 그런 다음, 소규모 프로젝트를 통해 효과적인 구현이 가능해집니다.

예를 들어, 이전 HR 시나리오에서 생성형 AI는 다양한 사용 사례를 제공할 수 있습니다.

1

탐색

직원이 인터페이스에 쿼리합니다. 1년에 휴가가 며칠 남았지? 이에 대응하기 위해 AI는 검색을 수행하고 쿼리와 관련된 문서를 제시하고 HR 정책 문서와 직원 기록을 가져와야 합니다.

2

요약

한 단계 더 나아가 생성형 AI는 문서를 분석하고 직원을 위해 대화식 응답으로 요약할 수 있습니다. "올해 휴가는 10일 남았고, 4일의 추가 유급휴무일을 사용하실 수 있습니다. 유급 휴가 정책에 대한 인트라넷 페이지를 확인하세요."

3

생성 및 자동화

챗봇은 휴가 시간 사용 요청을 승인하거나 거부하는 응답을 작성하고 이유를 제시함으로써 관리자의 시간을 절약할 수 있습니다. 또한 캘린더 초대를 생성하고 시스템에 PTO 요청을 기록할 수도 있습니다.

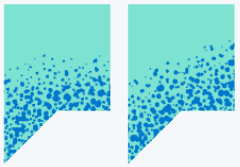
해당 업계에 맞게 채택하기

코파일럿, 어시스턴트, 봇 — 생성형 AI는 다양한 분야와 산업에서 귀중한 생산성 향상 서비스를 제공하는 다양한 "형태"를 취합니다. 보안 및 Observability 코파일럿으로서 또는 내부 및 외부 앱을 사용하는 생성형 AI는 기업이 효율성을 높이고 보안 노력을 업그레이드하며 고객 경험을 개선하고 경쟁 차별화를 가속화하는 데 도움이 됩니다.

데이터를 사용하여 최고의 AI 응답을 활용하고 기본 검색 기술의 힘을 활용함으로써 **회사는 반복 작업에 소요되는 시간을 줄이고 응답 시간을 단축하며 전반적으로 생산성을 향상시킬 수 있습니다.** 방정식에 RAG를 추가하면 문서 및 사용자 수준 권한을 존중하면서 안전한 생성형 AI 응답을 위해 독점 데이터를 활용할 수 있습니다.

갑자기 기술에 정통한 고객이 기대하는 것과 동일한 속도와 관련성을 통해 다음 단계의 속도와 관련성을 보게 됩니다. 여러분의 서비스가 이러한 기대에 부합하는지 확인하는 것이 중요합니다. 사용자 경험과 가장 관련성이 높은 방식으로 생성형 AI를 사용하는 것도 마찬가지입니다. 아무도 사용하지 않는 복잡한 건물에 투자하는 것보다 더 나쁜 것은 없습니다.

생성형 AI는 AI 어시스턴트 또는 보안 및/또는 Observability 코파일럿으로서 회사의 IT 인프라에 가장 일반적으로 통합됩니다.



생성형 AI는 근본적으로 정보 시스템에서 정보를 얻는 훨씬 더 인간적이고 직관적인 방법이기 때문에 모든 비즈니스에는 생성형 AI에 대한 기회가 있습니다.

Ash Kulkarni
Elastic의 CEO

AI 어시스턴트

내부 및 외부 앱을 통해 직원과 고객을 위한 생성형 AI 대화 기술을 최대한 활용하세요. AI 어시스턴트는 전문가, 퍼스널 쇼퍼, 일정 관리 담당자로서 모든 사용자에게 유연하고 적응력이 뛰어나며 개인적인 도움을 제공합니다.

보안 및 Observability 코파일럿

생성형 AI 코파일럿을 통해 Observability과 보안 기능을 강화하세요. IT 팀과 협력하도록 설계된 생성형 AI 코파일럿은 전문적인 문제 해결 파트너 역할을 합니다. 예를 들어, 보안 경보가 트리거된 이유에 대한 자세한 설명을 코파일럿에게 요청하고 공격을 분류하고 해결하기 위한 권장 단계를 얻을 수 있습니다 (조직에서 이전에 발생한 유사한 공격을 기반으로 함). 이러한 유형의 프롬프트는 조직에 대한 동적 런북을 생성할 수 있습니다.

이러한 통합을 통해 업계 전반의 기업은 개인에게 맞춤화, 자동화 및 생산성 잠재력을 높일 수 있으며, 이는 **세 가지 주요 생성형 AI 사용 사례**로 이어집니다.

운영 복원력 향상

운영 복원력은 시스템을 원활하게 실행하는 데 필수적입니다. 생성형 AI의 강화로 IT 팀은 근본 원인 분석을 가속화하고, 모든 환경에서 더 많은 데이터를 연관시켜 문제를 더 빠르게 찾아내고, 전용 검색 도구를 사용하여 대응 속도를 높일 수 있습니다. 이 모든 것이 비즈니스 연속성에 유리합니다.

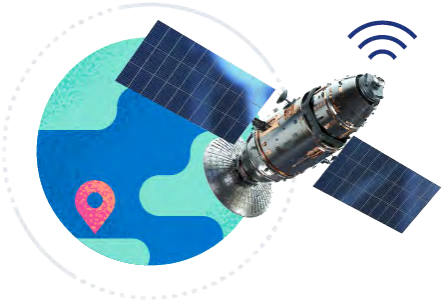
고객 경험 개선

고객 만족은 모든 비즈니스의 핵심입니다. 생성형 AI는 문제를 더 빠르게 해결하고 필요한 정보를 얻을 수 있는 도구를 팀에 제공하는 동시에 고객에게 개인에게 맞춤화된 관심과 관련 정보에 대한 빠른 액세스를 제공합니다. 그 결과는? 고객 경험이 향상되고 비즈니스 성과가 향상됩니다.

보안 위험 완화

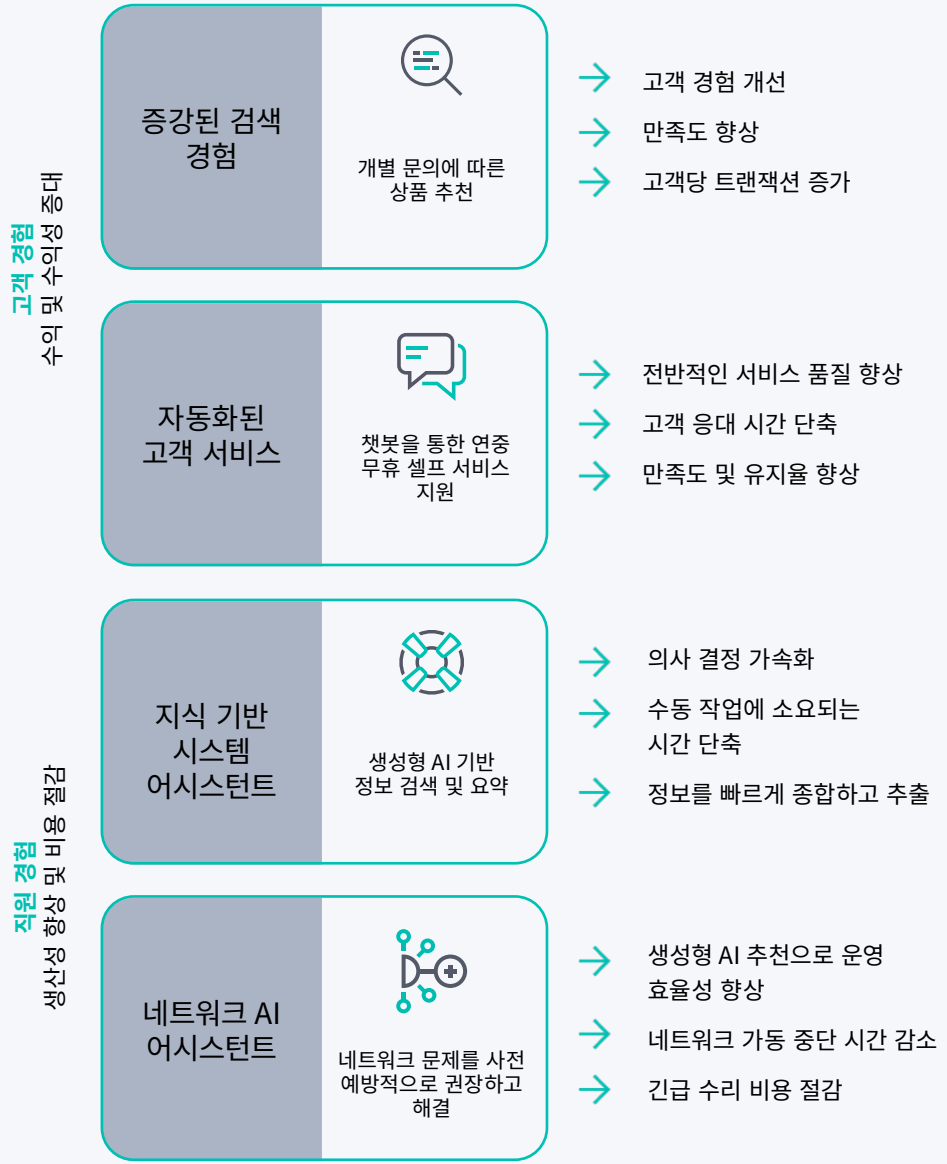
디지털 세계가 엄청난 속도로 발전함에 따라 새롭고 정교한 보안 위협이 등장하고 있습니다. 이에 대처하려면 위협에 대응하고 관리하기 위한 전문성은 물론 역동적이고 선제적인 조치가 필요합니다. 생성형 AI는 보안 팀과 운영을 강화할 뿐만 아니라 경보를 자동화하고 사전 대응적인 자세를 유지할 수 있습니다.

산업 전반에 걸쳐 생성형 AI는 쿼리에 대해 개인에게 맞춤화되고 관련성이 높으며 규범적인 응답을 제공함으로써 기존 직원 및 고객 경험을 강화할 수 있습니다. 어떤 분야에 있든 생성형 AI를 운용하여 검색을 강화하고 데이터의 새로운 기능을 활용할 수 있는 방법이 있습니다.



통신

통신 회사의 경우 생성형 AI는 600억 달러를 초과하는 경제적 가치를 창출할 것으로 예상됩니다.[4] 생성형 AI를 통해 통신 회사는 직원과 고객이 웹 사이트나 내부 작업장 지식 기반 시스템에 쿼리하여 개인에게 맞춤화되고 관련 있는 응답을 빠르게 얻을 수 있도록 할 수 있습니다. 그 결과는? 더 나은 고객 서비스와 향상된 생산성.



[4] McKinsey, 과대광고를 넘어서: 기술, 미디어, 통신 분야에서 AI 및 생성형 AI의 잠재력 포착, (2024).

금융 서비스

생성형 AI를 통해 금융 서비스 회사는 고객 및 직원 경험을 더욱 개인에게 맞춤화할 수 있습니다. 고객 경험, 사기 방지 및 자동화의 향상은 금융 서비스 업계에서 2,500억 달러가 넘는 경제적 가치를 창출할 것으로 예상됩니다.[5]

[5]McKinsey, 생성형 AI의 경제적 잠재력: 차세대 생산성 개척지(2023년).

고객 경험 수익 및 수익성 증대

소매 금융 어시스턴트	생성형 AI 기반 정보 검색 및 요약	→ 개인 금융 가시성 확대 → 더 높은 전환율을 위한 맞춤형 제안 제공 → 만족도 및 유지율 향상
향상된 고객 서비스	네트워크 문제를 사전 예방적으로 권장하고 해결	→ 전반적인 서비스 품질 향상 → 고객 응대 시간 단축 → 유지율 증가

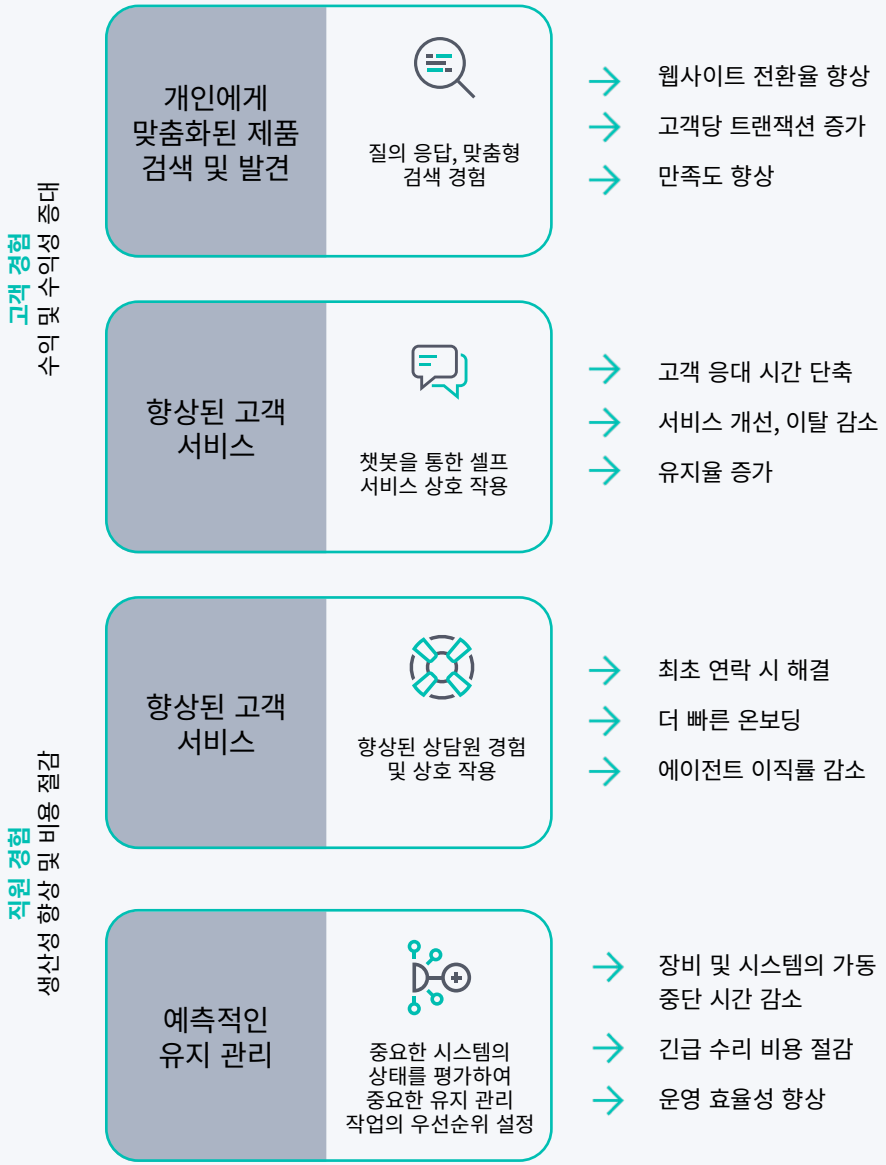
직원 경험 생산성 향상, 비용 및 위험 감소

사기 탐지 요약	이상 징후 탐지/ 트랜잭션 요약 및 차선책	→ 사기 탐지의 정확성과 속도 향상 → 작업 자동화로 비용 절감 → 재정적 손실 감소
가상 어시스턴트	NLP 기반 정보 검색 및 요약	→ 의사 결정 가속화 → 수동 작업에 소요되는 시간 단축 → 정보를 빠르게 종합하고 추출

소매

소매업에 가장 매력적인 생성형 AI는 검색 관련성을 높이고 추가 제품을 추천하며 채널 전반에 걸쳐 개인에게 맞춤화된 후속 조치를 보내 고객 유지율을 높일 것을 약속합니다. "장바구니에 넣어두신 물건을 잊으셨군요!"라는 이메일 메시지를 받아보신 적이 있으신가요? AI는 더 나은 추천과 보다 개인에게 맞춤화된 제품 검색을 위해 이를 자동화하고 개선할 수 있습니다.

전자 상거래 판매를 촉진하기 위한 차세대 고객 경험을 구축하든, 직원에게 최신 기술을 제공하여 생산성을 향상시키든, 생성형 AI는 소매업체에 2,400억 달러가 넘는 경제적 가치를 창출할 것으로 예상됩니다.[5]



[5]McKinsey, 생성형 AI의 경제적 잠재력: 차세대 생산성 개척지(2023년).

사례 연구: HSE

HSE는 유럽 라이브 커머스 분야의 선도적인 브랜드 중 하나입니다.[6]

"홈쇼핑 유럽[HSE]의 상업적 성공은 웹사이트 개인별 맞춤화 및 관련성에서 시작됩니다."

Peter Strasser
HSE 소프트웨어 개발자

기회

모든 전자 상거래 비즈니스와 마찬가지로 검색 기능은 고객 경험과 판매에 있어 기본입니다. HSE는 다양한 채널에서 발생하는 고객 여정에 맞춰 고객이 제품을 접하게 된 위치를 반영하는 다양한 검색어를 제공해야 합니다.

HSE는 생성형 AI와 LLM을 사용하여 고객 쿼리의 의미론적 의미를 추출하고 기존 키워드 매칭을 보완하는 결과를 생성했습니다.

결과

HSE는 더욱 정확하고 관련성이 높은 검색 결과 덕분에 **클릭률이 4% 증가**하고 **고객 만족도가 8% 증가**했습니다.

인사이트

고객 검색 경험과 같이 이미 개선하고 싶은 영역에 집중하세요. 생성형 AI를 통합하여 개인에게 맞춤화 및 관련성을 통해 경험을 한 단계 더 발전시킬 수 있는 방법을 알아보세요.

[6] Elastic, HSE는 AWS에서 Elasticsearch를 사용하여 유지 관리 시간을 42% 줄이면서 고객 만족도를 높였습니다(2024).

자동차 및 제조

자동차 및 제조 산업 프로세스의 모든 단계는 AI를 통해 간소화될 수 있으며 예상 경제적 가치는 1,700억 달러를 초과할 수 있습니다.[5] 생성형 AI는 산업을 제품 연구 및 개발 혁신에서 맞춤형 고객 유지 전략으로 변화시킬 수 있는 잠재력을 가지고 있습니다. 하늘을 하는 자동차? 아마도 가능할 수도 있습니다!

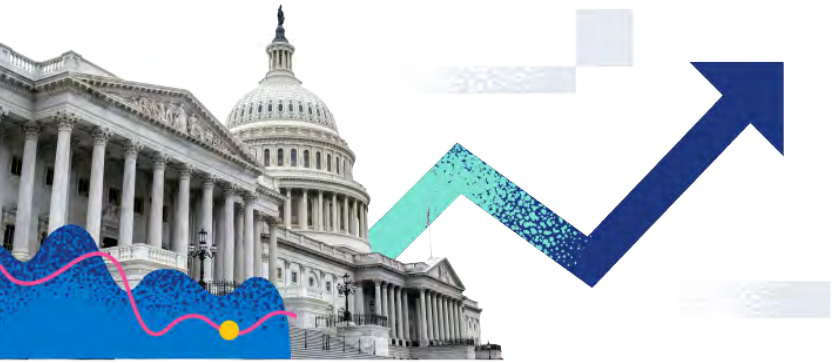
[5]McKinsey, 생성형 AI의 경제적 잠재력: 차세대 생산성 개척지(2023년).

고객 경험 수익 및 수익성 증대

대화형 디지털 매뉴얼	가상 제품 어시스턴트	→ 제품 기능, 유지보수, 문제 해결에 대한 실시간 답변 → 지원 문의 감소 → 만족도 향상
향상된 고객 서비스	챗봇을 통한 셀프 서비스 상호 작용	→ 고객 응대 시간 단축 → 서비스 개선, 이탈 감소 → 유지율 증가

직원 경험 생산성 향상 및 비용 절감

운영 기술 최적화	예측적인 유지 관리: 문제 및 해결 방법 요약	→ 문제를 빠르게 식별하고 해결 → 운영 개선 효율성과 의사 결정 → 제조 비용 절감
제품 정서 분석	제품 비율 요약 및 개선 권고	→ 고객 PoV를 통해 제품 제안 개선 → 신제품 제공의 가치 실현 시간 단축



공공 부문

생성형 AI는 생성형 AI를 기관 데이터와 안전하게 연결함으로써 임무 결과를 크게 가속화하고, 시민 서비스를 개선하며, 정부 분석가와 보안 전문가를 적시에 올바른 데이터에 더 잘 연결할 수 있습니다.

시민 대상 애플리케이션에는 다음이 포함됩니다.

- 공공 서비스에 대한 개인 맞춤형 액세스
- 간소화된 온라인 시민 경험
- 향상된 접근성 및 셀프 서비스 옵션

직원용 애플리케이션에는 다음이 포함됩니다.

- 더욱 정확한 조사와 정보 제공
- 수동 프로세스 및 워크플로를 자동화하여 생산성 향상
- 보다 효율적인 조달 프로세스

작업 부하 감소
수동 프로세스 및 워크플로우 자동화

규정 준수
역할 기반 데이터 액세스 활성화

실시간 상황 인식
보다 정확한 결정 내리기

직원 생산성
적시에 적절한 정보 찾기

시민 경험
맞춤형 디지털 상호작용을 통해 신뢰 구축

공공 서비스
접근성 및 셀프서비스 옵션 향상

동적 인텔리전스
임무 검색 및 인사이트 가속화

사이버 보안
실시간 위험 평가 및 분석 수행

사례 연구: Relativity

Relativity는 기업, 법률 회사, 대행사가 전자 증거개시 및 법률 검색을 위해 데이터를 저장하고 활용하는 데 도움이 됩니다.[7]

"현재 Relativity 고객이 직면하고 있는 가장 큰 과제는 이기종 데이터 소스에서 데이터가 폭발적으로 증가한다는 것입니다. 다양한 통신 모드에서 생성된 데이터의 차이로 인해 문제가 더욱 복잡해졌습니다."

Brittany Roush
수석 제품 관리자

기회

보안 우선 입장을 유지하면서 데이터를 통합하려면 Relativity가 필요했습니다. 데이터, 소스, 복잡성이 폭발적으로 증가함에 따라 기존의 키워드 검색 접근 방식은 효과적이지 않았습니다. 바로 이 지점에서 RAG가 필요합니다.

결과

Relativity는 RAG와 벡터 데이터베이스를 함께 사용하여 독점 데이터를 기반으로 구축된 검색 환경을 구현하고 사용자에게 빠르고 관련성이 높으며 정확한 검색 환경을 제공했습니다. 생성형 AI 솔루션은 PCI, DSS, SOC2, HIPAA 등의 규정 준수 표준을 충족합니다.

인사이트

규모를 염두에 두고 구축하세요. 작게 시작하면 생성형 AI의 기능을 식별하고 가장 관련성이 높은 애플리케이션을 연마하는 데 도움이 될 수 있습니다. 최적의 지점을 찾으면 한계가 없습니다.

[7]Elastic, Relativity는 Elasticsearch와 Azure OpenAI를 사용하여 오늘(2024년) 미래형 검색 환경을 구축합니다.

따라서 여러분은 산업 전반에 걸쳐 생성형 AI의 엄청난 경제적 잠재력을 이해하고 계십니다. 잠재적인 사용 사례를 염두에 둘 수도 있습니다. 바라건대, 여러분도 "이유"를 알고 계실 것입니다.

그러나 생성형 AI를 구현하는 것은 힘들고 파괴적인 프로세스처럼 보일 수 있습니다. 개인 정보 보호 문제, 규정 준수 영역에서 수행해야 할 일부 발품 작업, 사람들이 업무를 수행하는 방식에 대한 변경 사항이 있습니다. 책임 있는 운영을 위해서는 교육, 기술 향상, 인력의 부분적인 재구성이 필요합니다.

이러한 과제에도 불구하고 생성형 AI가 기업에 가져올 수 있는 가치는 부인할 수 없습니다. 경쟁력을 유지하려면 구현이 불가피합니다. 좋은 소식은? 완전히 생산 준비가 되지 않은 테스트를 통해 가치 실현 시간을 단축할 수 있습니다. 즉, 이제 시작할 시간입니다.

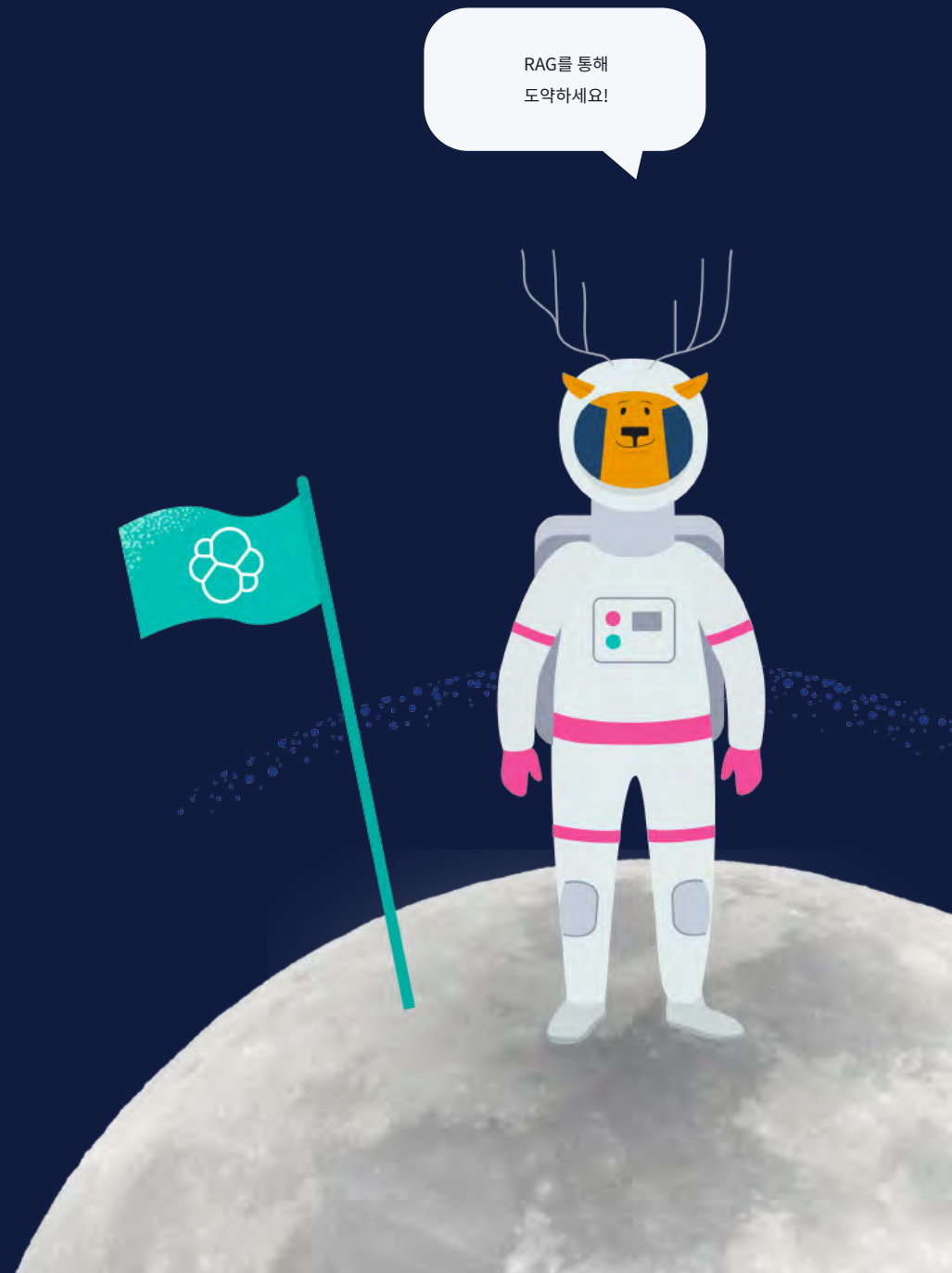
제2부:

생성형 AI 운용

머신을 위한 작은 한 걸음 — 조직을 위한 거대한 도약

생성형 AI의 운용은 한꺼번에 이루어지지 않습니다. 이는 계획과 명확한 결과가 필요한 반복적인 접근 방식입니다. 단일 생성형 AI 프로젝트(작은 단계)로 시작하면 피할 수 없는 학습 곡선을 통해 팀에 권한을 부여하고 프로세스를 워크숍으로 진행하고 기술을 미세 조정하며 우려 사항을 해결하여 팀과 회사가 성공할 수 있도록, 거대하게 도약할 수 있도록 준비할 수 있습니다.

이제 이를 실현하는 방법은 다음과 같습니다.



단계 1

이상적인 결과 식별

여러분은 문제를 해결했습니다. 여러분은 비효율적인 프로세스를 최적화하려고 한다는 것을 알고 있습니다. 이제 사용자가 솔루션과 상호 작용하는 방식에 대해 생각해야 합니다. 검색 애플리케이션이나 챗봇을 강화하고 계신가요? 팀이나 고객과 상호 작용할 수 있는 새로운 방법을 찾고 계신가요?

여러분의 사고 과정은 다음과 같을 수 있습니다.

- 더 많은 고객 유지를 원하고 있습니다.
- 개인에게 맞춤화된 제품 검색 및 발견 애플리케이션을 구현하기로 결정했습니다.
- 성공을 위한 지표를 만듭니다. 이것을 "하위 이유"라고 생각하세요.

  생성형 AI를 활용하고 싶습니다. 이유가 뭘까요? 개인에게 맞춤화된 제품을 검색하고 발견하기 위해서입니다. 이유가 뭘까요? **이상적인 결과는 다음과 같습니다.** 이러한 새로운 방식으로 데이터와 상호 작용함으로써 고객은 검색 기록과 위치를 기반으로 필요한 제품을 쉽게 찾고, 원할 수도 있는 제품을 발견할 수 있습니다. 결과적으로 고객 유지율이 향상됩니다.
- 이제 첫 번째 생성형 AI 프로젝트를 운영화하는 크고 큰 작업을 시작합니다.

스스로에게 이렇게 물어보세요.

데이터와 상호 작용하는 이 새로운 방법을 통해 어떤 조치와 결과를 만들 수 있을까?

이 질문에 대한 답은 목표를 설정하는 데 도움이 될 것입니다. 이상적인 결과를 식별하면 프로젝트의 "좋은" 모습이 결정되고 더 큰 규모에서는 회사의 모습이 결정됩니다.

단계 2

영향 파악. 성공 측정.

생성형 AI 운용에 성공하려면 "좋은" 것이 자신에게 무엇을 의미하는지 측정하는 데 도움이 되는 일련의 KPI를 설정해야 합니다. 생성형 AI가 조직의 생산성 바늘을 어떻게 움직이는지 이해하는 것은 많은 성과 지표 중 하나일 뿐입니다.

기타에는 고객 지원 맥락에서의 리뷰로 측정되는 고객 만족도 증가, 지원 티켓 감소, 해결 시간 단축 등이 포함될 수 있습니다. 테스트 중인 사용 사례에 따라 해당 성능 지표를 설정해야 합니다. 테스트 과정의 모든 단계에 이러한 내용을 포함시키는 것은 여러분과 팀의 진행 상황을 이해하는 데 매우 중요합니다.

기본 성과 지표

1

생산성 영향

사용 사례로 인한 생산성 변화를 측정합니다. 생성형 AI를 사용하지 않고 필요한 시간과 특정 작업을 완료하는 데 필요한 시간을 비교하세요.

2

확장성

사용량 및 수요 증가에 따라 모델이 얼마나 잘 확장되는지 평가합니다. 여전히 안정적이고 정확하게 작동하고 있나요?

3

요점

생성형 AI 구현이 비즈니스 비용 측면에서 어떤 영향을 미쳤는지 평가합니다. 기록된 고객 불만 사항 수, 판매 변경 사항 등 특정 비즈니스 지표를 이 검토에 포함할 수 있습니다.

4

규정 준수

생성형 AI의 데이터 개인정보 보호 규정 준수 여부를 지속적으로 모니터링하세요.

5

고객 만족도

고객 이탈, 매출 증가, 브랜드 충성도 유지 등의 비즈니스 지표를 검토하고 고객 피드백을 검토합니다.

이러한 지표를 사용하여 프로젝트가 실현 가능하고, 실행 가능하며, 확장 가능하고, 합리적인 가격인지 판단하세요. 이러한 지표는 ROI를 결정하는 데 도움이 되며 향후 사용 사례를 확장함에 따라 확장될 수 있습니다.

단계 3 모델 선택(앞으로 나아갈 길)

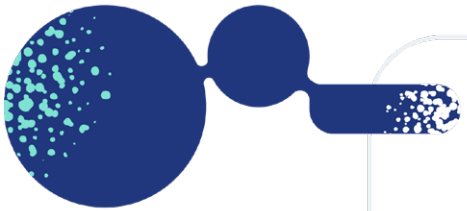
비즈니스 요구 사항을 충족하는 생성형 AI 아키텍처를 어떻게 구축할까요? 비용, 언어, IT 생태계, 배포 기능 및 타임라인, 데이터 개인 정보 보호 규정, 거버넌스 등 많은 요소가 선택에 영향을 미칩니다. 이러한 이유로 단순하고 구체적인 사용 사례부터 시작하여 좁은 입장을 취하는 것이 중요합니다.

생성형 AI를 운용하려면, 다음 구성 요소가 필요합니다.

- **완전 관리형 클라우드 인프라는** 민첩성을 높이고 비용 효율성을 높이며 낭비되는 리소스를 줄여줍니다. 칩과 하드웨어는 엄청난 속도로 발전하고 있습니다. 자체 AI 데이터 센터 구축에 투자하면 몇 달 안에 쓸모 없게 될 수 있습니다.
- **LLM**은 생성형 AI가 자연어로 의사소통하고 이해할 수 있도록 하는 기반이 될 것입니다.
- 독점 데이터의 올바른 컨텍스트로 LLM을 강화하는 데 사용할 수 있는 벡터, 하이브리드 및 기존 키워드 검색을 포함하는 **데이터 플랫폼**.
- 데이터를 강화하고 LLM과 검색 엔진에 전달할 수 있는 **광범위한 API**.

기업이 AI 검색에 필요한 재료

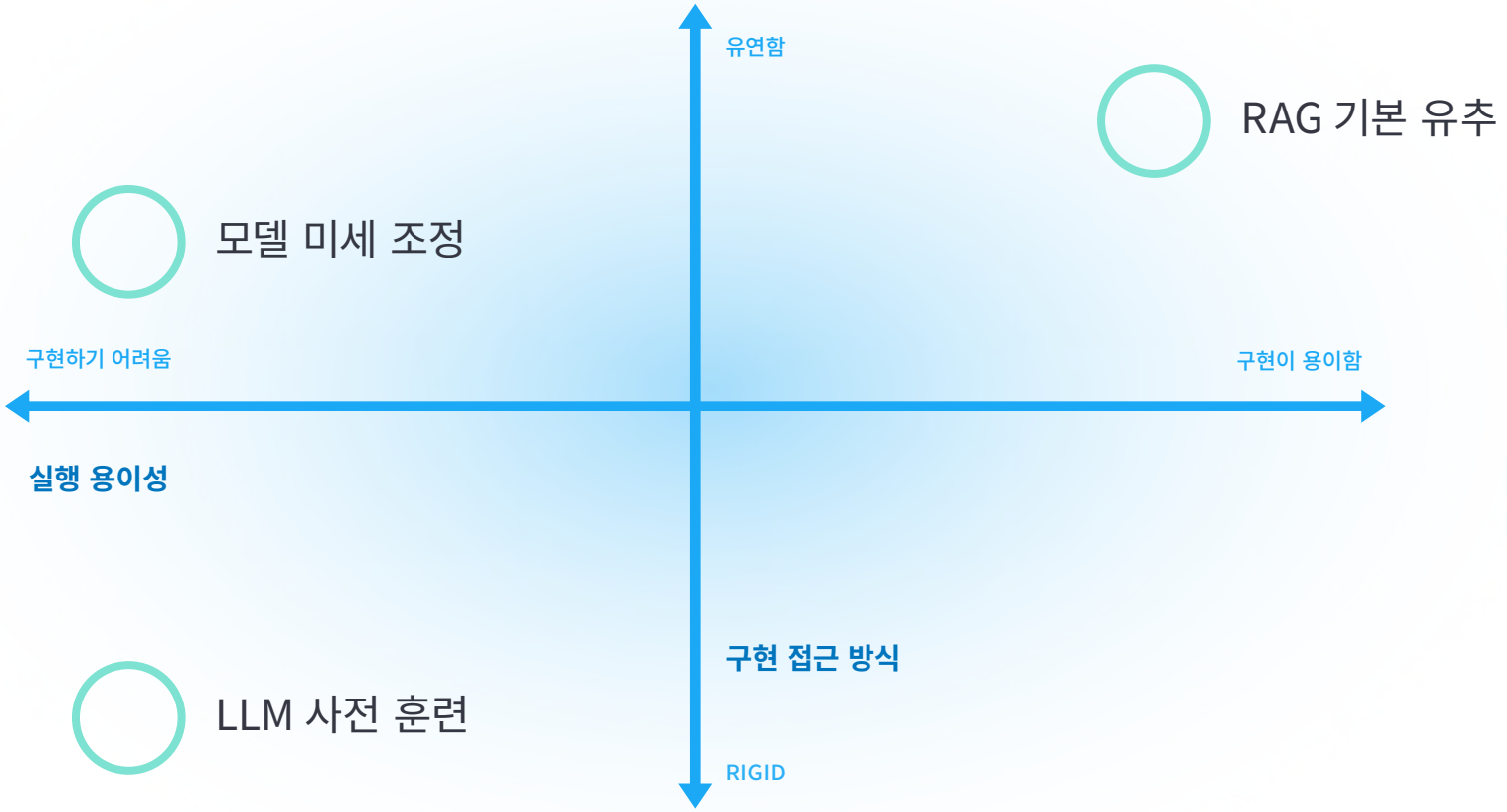
자체 모델을 미세 조정하든, 자체 벡터 데이터베이스를 가져오든, 자체 모델을 가져오든, 아니면 그 조합을 가져오든, 이러한 구성 요소를 결합하는 방법에 따라 구현 방식이 결정되고, 타임라인, 테스트 복잡성 및 실행 여부에 영향을 미칩니다. 팀을 보완해야 할 수도 있습니다.



	LLM 사전 훈련	모델 미세 조정	RAG
	이러한 리소스 집약적 접근 방식에는 대규모 데이터 세트에 대해 대규모 언어 모델을 교육하여 처음부터 시작하는 작업이 수반됩니다.	이 접근 방식은 검색 엔진 및 벡터 데이터베이스와 함께 기존 LLM을 사용하여 독점 데이터에 컨텍스트를 제공합니다.	이 프로세스에서는 기존의 사전 훈련된 LLM과 일련의 기술을 사용하여 요구 사항에 맞게 모델을 조정합니다.
비용	$$$$	$$$	$$
배포 시간	긴 시간, 몇 달 만에 배포	보통, 몇 주 안에 배포	신속, 며칠 만에 배포
데이터 개인정보 보호	LLM에 중요한 학습 자료를 제공할 만큼 충분한 데이터 세트가 있나요? 그렇지 않다면, 공개 데이터가 필요합니다. 공개 데이터와 비공개 데이터를 결합하시겠어요?	LLM에 중요한 학습 자료를 제공할 만큼 충분한 데이터 세트가 있나요? 그렇지 않다면, 공개 데이터가 필요합니다. 공개 데이터와 비공개 데이터를 결합하시겠어요?	이 접근 방식을 사용하면 개인 데이터를 비공개로 유지할 수 있습니다.
정확성과 관련성	일관되게 보장하기가 어려움	모델이 조정된 특정 작업에 대한 정확성과 관련성을 더 쉽게 보장할 수 있습니다.	RAG의 이점은 특히 출처를 인용하거나 사용자에게 답변이 없을 때 알려줌으로써 환각을 "필드"하는 능력에 있습니다.

다음 옵션을 고려하세요.

사전 학습,
미세 조정,
RAG



앞으로 나아갈 올바른 길 선택

모든 사람에게 올바른 방법은 없습니다. 따라서 설명된 목표와 비교하여 결정을 내렸는지 확인하세요. 이전 단계에서 이러한 목표가 관련 당사자와 일치하는지 확인합니다. 궁극적으로 비즈니스 이해관계자와 팀에게 명확하게 설명할 수 있는 경로가 필요합니다.

보다 광범위한 점검 및 대규모 프로젝트를 추구하지 않는 한 처음부터 LLM을 구축하고 미세 조정하는 것은 리소스 집약적입니다. 질문이 넘쳐날 것입니다. 검색 엔진을 보완하기 위해 벡터 데이터베이스가 필요한가? 검색 엔진을 업그레이드하고, 임베딩을 생성 및 저장하고, 검색 기능을 계속 강화할 수 있는 논리를 구축할 수 있는가? 이것을 추천 시스템으로 어떻게 확장하는가?

하이브리드 검색과 시맨틱 검색을 갖춘 클라우드 기반 솔루션의 경우, 매우 간단할 수 있습니다. 기존 LLM에 연결하면 RAG를 사용하여 고객에게 보다 관련성 높은 검색 환경을 만들 수 있습니다.

고려해야 할 사항은 다음과 같습니다.

- 여러분의 IT 환경에 이미 보유하고 있는 사항을 조사하세요. 인프라 재설계가 전혀 필요하지 않은 경우가 많습니다.

- OOTB LLM, OOTB 벡터 데이터베이스, OOTB 올인 패키지와 같은 기본(OOTB) 솔루션을 고려해보세요.

  장점: 블랙박스 기술을 사용하면 더 빨리 시작하고 실행할 수 있습니다.

  단점: 최소한의 사용자 정의가 가능하기 때문에 동일한 방식으로 확장할 수 없습니다.

- 유연성을 제공하고 중단을 최소화하는 보완 제품을 찾아보세요. 검색 관련성과 성능을 벤치마킹하고 모델을 교환하여 어떤 것이 가장 적합한지 확인하고 싶을 것입니다.

단계 4

빨리 시도하고, 빨리 실패하기

빠르게 진화하는 디지털 생태계는 생성형 AI 프로젝트에 움직이는 부분이 많다는 것을 의미합니다. 사전 훈련된 LLM에 대해 제어할 수 있는 것은 제한적일 뿐만 아니라 아키텍처를 조작할 수 있는 유연성도 제한되어 있습니다.

이제 반복적인 접근 방식을 취해야 할 때입니다. 사용 사례가 있고, 원하는 결과를 설정했으며, KPI를 설정하고, 생성형 AI 프로젝트 구현 방법을 고려했습니다.

기억해 두세요

기본적으로 생성형 AI를 운용하는 것은 여러분의 데이터에서 답변을 얻는 것입니다. 규정 준수를 최우선으로 생각하세요. 개인 정보 보호 정책을 손상시킬 수 있는 테스트를 설정하고 계신가요? 위험성이 낮은 테스트인가요?

이 단계에서 여러분이 원하는 것은…

- **피드백 루프 구축:** 누가 누구에게 무엇을 보고하는지 설정하고 프로젝트의 주요 이해관계자를 식별하세요.
- **LLM 보강:** LLM이 벡터 데이터베이스에 저장된 올바른 정보에 액세스할 수 있는지 확인하세요. 벡터 데이터베이스를 사용하면 가장 관련성이 높은 정보를 신속하게 제공하여 LLM을 강화할 수 있습니다.
- **사용자 경험 미세 조정:** 사용자 친화적인 인터페이스로 작업하고 계속 테스트하세요. 궁극적으로 생성형 AI는 직원과 고객에게 서비스를 제공하기 위해 존재합니다. 애플리케이션과 사용자에게 적합한 인터페이스를 구축하는 것은 성공적인 생성형 AI 프로젝트에 매우 중요하며 확장성을 보장합니다.
- **확장 가능한 참조 아키텍처 설정:** 생성형 AI 프로젝트를 테스트하는 동안 큰 그림을 주시하세요. 프로젝트를 확장할 때 아키텍처는 어떤 모습이며, 추가 사용 사례로 확장하면 어떤 모습일까요?

예를 들어, 처음부터 벡터 데이터베이스를 구축하는 것이 무거운 작업처럼 느껴진다면 다운로드 가능한 데이터베이스를 찾아볼 수 있습니다. 그렇습니다. 이런 데이터베이스도 존재합니다. 해당 벡터 데이터베이스를 사용하면 다음 단계인 하이브리드 검색의 잠금을 해제할 수 있습니다. 검색 애플리케이션에서 의미론적 검색을 사용하면 차세대 AI 프로젝트 프로토타입을 테스트할 수 있습니다. 이는 작게 시작하고 반복하는 것의 힘을 보여주는 예입니다.

단계 5

거버넌스 및 운영

생성형 AI 프로젝트는 데이터 개인 정보 보호 및 규정 준수부터 윤리적 고려 사항, 품질 관리, 위험 관리에 이르기까지 자체적인 과제를 안고 있습니다. 잠재적인 장애물을 예측하고 프로젝트가 비즈니스 목표에 부합하는지 확인해야 합니다.

거버넌스 및 운영 검토의 일환으로 다음과 같은 다양한 요소를 고려해야 합니다.

- **비용 관리:** 1,000개 토큰 단위로 요금이 청구됩니다. 하나는 메시지를 보내는 데, 하나는 응답에 드는 비용입니다.
- **로깅:** 품질 관리를 위해 모델과 고객 간의 커뮤니케이션을 확인하려면 모든 응답을 기록해야 합니다.
- **응답 정서 설정:** LLM 응답의 정서를 파악하여 회사의 목소리 톤과 브랜드가 일치하도록 합니다(또 다른 중요한 품질 관리 단계).
- **환각 모니터링:** 환각에는 부정확하거나 오해의 소지가 있는 정보가 포함되지만 증오심 표현이나 챗봇의 반사회적 행동도 포함될 수 있습니다.
- **결정적이지 않은 답변 신고:** 응답의 품질과 관련성을 모니터링하는 것은 품질 관리에 매우 중요합니다. 이는 어떤 애플리케이션이 다른 애플리케이션보다 더 많은 사람의 개입을 필요로 하는지 이해하고 확장 시점에 따라 그에 따라 계획을 세울 수 있는 기회입니다.

AI의 편견

생성형 AI 모델은 훈련된 데이터에 의존합니다. 훈련 데이터에 편향과 제한 사항이 포함되어 있는 경우, 이는 출력에 반영됩니다.

조직은 모델이 훈련되는 데이터를 신중하게 고려 및 제한하거나 요구 사항에 맞게 맞춤화되고 특화된 모델을 사용하여 이러한 위험을 완화할 수 있습니다. 즉, 이 기술을 프로그래밍하거나 모델이 훈련된 데이터를 관리하는 사람에게도 편견이 있습니다.

어떤 상황에서든 편견은 근절하기 어렵습니다. 이는 조직이 이 문제를 해결하고 솔루션의 일부로 비판적 사고를 수행하도록 사용자를 교육해서는 안 된다는 의미는 아닙니다.

또한 법무팀의 참여를 기대하고 그들의 작업을 개념 증명에 포함시키세요. 이들의 참여가 테스트 단계의 속도를 늦추는 것처럼 보일 수도 있지만 책임감 있고 윤리적이며 규정을 준수하는 구현을 위해서는 철저하고 효율적인 검토 프로세스를 확립하는 것이 중요합니다.

데이터 안전에 관한 사항

매일 조직에 영향을 미치는 보안 위협으로 인해 데이터 안전이 무엇보다 중요합니다. 고객은 자신의 데이터에 대해 여러분을 신뢰합니다. 이것이 바로 많은 기업이 제로 트러스트 프레임워크에서 운영되는 이유입니다. 이는 조직의 네트워크 경계 내부든 외부든 사용자와 기기를 자동으로 또는 묵시적으로 신뢰해서는 안 된다는 원칙을 따릅니다.

보안을 최적화하려면, 다음을 수행할 수 있습니다.

1. **RAG 접근 방식을 취할 수 있습니다.** RAG 모델은 검색 메커니즘을 활용하여 입력 프롬프트의 맥락을 더 잘 이해합니다. 이를 통해 민감한 세부정보를 생략하는 보다 상황에 맞는 적절한 응답을 얻을 수 있습니다. 문서 수준 및 역할 기반 보안을 갖춘 데이터 플랫폼과 함께 RAG를 사용하면 권한이 존중됩니다.

2. **Observability 솔루션에 투자하거나 확장할 수 있습니다.** 신뢰 문제를 해결하세요. 모니터링 기능으로 데이터 추적을 추적하고 생성형 AI 프로젝트에서 생성된 응답을 모니터링하세요. 여러분의 데이터는 어디로 가고 있으며 생성형 AI는 고객에게 무엇을 말하고 있나요?

궁극적으로 생성형 AI를 생태계에 도입하려면 새로운 운영 프로토콜과 그에 따른 새로운 정책을 수립해야 합니다. 보다 효율적인 프로세스와 더 높은 수익을 통해 일상적인 작업을 수행하면서 절약된 시간을 이러한 노력에 다시 투입할 수 있습니다.

단계 6 타임라인 설정. 벤치마크 제공.

기간을 대략적으로 설명합니다. 1분기를 가정해 보겠습니다. 그 기간 내에서, 30일과 90일에 목표 지점 표시를 설정하세요. 해당 분기를 사용하여 생성형 AI 기반 사용 사례의 가치를 입증하세요.

30일까지는 첫 번째 테스트를 시작하고 싶으실 것입니다. 이것은 어떻게 보일까요?

- [ ] 사용 사례를 선택했습니다
- [ ] 작업에 소규모 팀을 할당했습니다
- [ ] 필요에 따라 교육 세션을 진행했습니다
- [ ] 원하는 결과를 얻었습니다
- [ ] 프로토타입 인터페이스를 구축했습니다

90일이 되면 첫 번째 사용 사례를 시작할 준비가 됩니다. 이것은 어떻게 보일까요?

- [ ] 몇몇 내부 사람들에게 테스트를 공개했습니다
- [ ] 생성된 출력을 테스트, 조정 및 측정했습니다
- [ ] 사용자가 인터페이스와 상호 작용하는 방식을 지속적으로 모니터링했습니다
- [ ] 품질 결과물을 구성하는 요소에 대한 일련의 지침을 설정했습니다
- [ ] 몇 가지 주요 성과 지표에 대한 데이터를 수집했습니다
- [ ] 이니셔티브의 가치를 측정했습니다

이러한 작업은 대략적인 벤치마크 역할을 해야 합니다. 회사의 특정 요구 사항(팀 구성, 작업 중이거나 스택에 추가하는 기술)은 첫 번째 사용 사례를 배포하고 인사이트를 수집할 수 있는 속도에 영향을 미칩니다.

이 단계에서는 다음을 고려하세요.

1. **오류율:** 오류율을 측정합니다. 생성형 AI가 정확하고 관련성이 높은 출력을 생성하고 있나요? 이는 생성형 AI를 미세 조정하는 데 중요합니다.
2. **훈련 시간 및 비용:** 모델 훈련에 필요한 시간과 리소스를 측정하세요. 이렇게 하면 효율적인 테스트 기간이 보장되므로 운영 시간이 더 빨라집니다.
3. **사람의 개입:** 생성형 AI는 인간 참여형(Human-In-The-Loop)을 통해서만 작동하나요? 신뢰성과 정확성을 유지하려면 얼마나 많은 감독이 필요한가요?
4. **응답 시간 및 출력 품질:** 생성형 AI가 출력을 얼마나 빨리 제공하는지 측정하고 설정된 규칙 또는 지침과 출력의 품질을 비교합니다.

그렇게 하면, 성공을 운영하고 확장할 준비가 된 것입니다.

새로운 시대의 시작

수많은 업계 리더들이 이미 생성형 AI의 이점을 확인하고 있으며, 더 많은 사람들이 이러한 성공을 재현하고 변화하는 고객 기대에 부응하기 위해 노력하고 있습니다. 생성형 AI의 혁신은 빠르게 진행되고 있습니다. 하지만 기본이 없으면 아무데도 갈 수 없습니다.

생성형 AI를 구현하는 데 가장 잘 맞는 방법을 전략화하면 흥미롭지만 관련성이 없는 혁신에 방해받지 않고 데이터의 힘을 활용하는 데 도움이 될 수 있습니다. 생성형 AI를 가장 효과적으로 운영하려면 이를 단계적으로 구현하는 데 시간과 리소스를 투자하세요. 새로운 기술과 목적을 통합하는 것은 투자 대비 최고의 수익을 위한 비결입니다. 또한 운영 요구 사항에 맞게 AI 도구를 사용자 정의하고 조정하면 관련성과 효율성이 보장됩니다. 이는 우선 생성형 AI의 "소명"입니다.

데이터 개인 정보 보호 및 보안부터 민감도 및 윤리에 이르기까지 생성형 AI를 책임감 있게 구현해야 할 필요성을 명심하세요. 생성형 AI는 세계 경제에 최대 수조 달러의 가치를 더하는 것 외에도 인력을 대중화하고 직원의 기술을 향상시킬 수 있는 기회를 제공합니다. 생성형 AI 선구자일 뿐만 아니라 회사를 위한 새로운 비즈니스 프로세스 개발의 선구자로서 자리매김하세요.

이제 시작해 보겠습니다.

첫 번째 생성형 AI 사용 사례를 식별하는 것은 팀 전체의 노력입니다. 처음부터 함께 작업하려면 보안 팀, IT 팀, 개발 팀, 사업 부문 팀이 필요합니다. **Elastic이 도움을 드릴 수 있는 방법은 다음과 같습니다.**

보안팀과 함께

의사의 생산성을 높이고 위험을 줄일 수 있습니다. 보안 사용 사례에 맞게 생성형 AI를 운영하는 것은 개방형 플랫폼에 대한 통합 접근 방식으로 시작됩니다. Elastic Security로 생성형 AI의 힘을 활용하여 보안 팀의 요구 사항에 맞는 환경을 만들 수 있습니다.

Elastic Security를 만나보세요

SRE 및 IT 운영팀과 함께

eqSRE와 엔지니어에게 가장 관련성이 높은 정보를 더 빠르게 찾아볼 수 있는 대화형 자연어 채팅 인터페이스를 활용할 수 있는 기능을 제공하세요. 독점 데이터와 런북을 기반으로 하는 컨텍스트 인식 대화형 채팅 경험을 위해 대화형 AI를 Elastic Observability 및 고급 머신 러닝과 결합하는 방법을 알아보세요.

Elastic Observability를 만나보세요

개발팀과 함께

개발팀의 툴킷을 강화하여 Elastic Search가 포함된 고도로 개인에게 맞춤화된 챗봇과 같은 셀프 서비스 옵션을 통해 고객 지원을 제공할 수 있습니다. 서로 다른 데이터 소스에서 답변을 찾는 데 도움이 되는 생성형 AI 경험을 포함하여 케이스를 신속하게 해결할 수 있는 동일한 훌륭한 검색 도구를 사용하여 고객 서비스 상담원의 역량을 강화하세요. 지식 기반 시스템에 대한 강력한 검색을 구현하는 방법을 알아보세요. .

Elastic Search를 만나보세요

elastic | The Search AI Company