



의료 및 생명 과학에 Amazon Comprehend Medical 및 LLMs 사용

AWS 권장 가이드



AWS 권장 가이드: 의료 및 생명 과학에 Amazon Comprehend Medical 및 LLMs 사용

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon의 상표 및 트레이드 드레스는 Amazon 외 제품 또는 서비스와 함께, Amazon 브랜드 이미지를 떨어뜨리거나 고객에게 혼동을 일으킬 수 있는 방식으로 사용할 수 없습니다. Amazon이 소유하지 않은 기타 모든 상표는 Amazon과 제휴 관계이거나 관련이 있거나 후원 관계와 관계없이 해당 소유자의 자산입니다.

Table of Contents

소개	1
개요	1
대상 독자	2
목표	2
기술적 접근 방식	3
Amazon Comprehend Medical 사용	3
기능	4
사용 사례	5
Amazon Comprehend Medical과 LLMs 결합	6
아키텍처	6
사용 사례	8
모범 사례	8
프롬프트 - 엔지니어링	9
LLMs 사용	18
LLM 사용 사례	18
사용자 지정	19
LLM 선택	21
LLMs 미세 조정	24
비용 및 ROI 추정	25
전략 선택	25
데이터 세트 빌드	27
미세 조정	28
모니터링	29
접근 방식 선택	30
비즈니스 성숙도 고려 사항	31
LLMs 평가	33
훈련 및 테스트 데이터	33
Metrics	34
FAQ	36
Amazon Comprehend Medical과 LLM 중에서 선택하려면 어떻게 해야 하나요?	36
Amazon Comprehend Medical 결과를 LLM에 제공하려면 어떻게 해야 하나요?	36
Amazon Comprehend Medical LLMs과 함께 사용할 때의 모범 사례는 무엇입니까?	36
의료 사용 사례에 맞게 사전 훈련된 의료 LLM을 사용하거나 일반 LLM을 미세 조정해야 합니 까?	37

의료 NLP 작업에 대한 LLMs의 성능을 평가하려면 어떻게 해야 합니까?	37
복잡성이 높은 LLM 솔루션과 복잡성이 낮은 LLM 솔루션의 장단점은 무엇인가요?	37
다음 단계	38
AWS 리소스	38
기타 리소스	39
기여자	40
작성	40
검토	40
기술 작성	40
문서 기록	41
용어집	42
#	42
A	43
B	45
C	47
D	50
E	54
F	56
G	57
H	58
I	60
L	62
M	63
O	67
P	69
Q	72
R	72
S	75
T	78
U	80
V	80
W	81
Z	82
.....	lxxxiii

의료 및 생명 과학에 Amazon Comprehend Medical 및 LLMs 사용

Amazon Web Services(???기여자)

2025년 12월([문서 기록](#))

개요

의료 데이터의 양이 계속 증가하고 효율적이고 정확한 처리가 필요해지면서 인공 지능 및 기계 학습(AI/ML) 기술을 통해 자연어 처리(NLP)를 채택하게 되었습니다. 사전 훈련된 분류기 모델과 [대규모 언어 모델\(LLMs\)](#)은 임상 질문 답변, 보고서 요약, 인사이트 생성 등 다양한 의료 NLP 작업을 위한 강력한 도구로 등장했습니다. 그러나 의료 및 생명과학 도메인은 의료 용어, 도메인별 지식 및 규제 요구 사항의 복잡성으로 인해 고유한 문제를 안고 있습니다. 이 도메인에서 사전 훈련된 분류기 또는 LLMs 효과적으로 사용하려면 이러한 모델의 강점을 도메인별 리소스 및 기법과 결합하는 잘 설계된 접근 방식이 필요합니다.

의료 및 생명과학 분야의 업계 관행은 전통적으로 규칙 기반 시스템, 수동 코딩 및 전문가 검토 프로세스에 의존해 왔습니다. 이러한 시스템과 프로세스는 시간이 많이 걸리고 오류가 발생하기 쉽습니다. [Amazon Comprehend Medical](#) 및 [Amazon Bedrock](#)의 파운데이션 모델과 같은 AI 및 NLP 기술의 통합은 정확성과 일관성을 개선하는 동시에 의료 데이터를 처리하기 위한 효율적이고 확장 가능한 솔루션을 제공합니다.

이 가이드에서는 의료 산업의 지능형 자동화를 위한 Amazon Comprehend Medical 및 LLMs 사용을 살펴봅니다. 의료 코딩, 환자 정보 추출 및 레코드 요약 프로세스를 간소화하기 위한 모범 사례, 과제 및 실용적인 접근 방식을 간략하게 설명합니다. 의료 기관은 Amazon Comprehend Medical 및 LLMs의 기능을 사용하여 새로운 수준의 운영 효율성을 실현하고 비용을 절감하며 잠재적으로 환자 치료를 개선할 수 있습니다.

이 가이드에서는 의료 용어 이해, 도메인별 LLMs 사용, AI/ML 시스템의 제한 사항 해결과 같은 의료 도메인의 고유한 고려 사항을 자세히 설명합니다. 의료 IT 관리자, 아키텍트 및 기술 책임자가 조직의 준비 상태를 평가하고, 구현 옵션을 평가하고, 성공적인 자동화를 위해 적절한 AWS 서비스 및 도구를 사용할 수 있는 포괄적인 의사 결정 경로를 제공합니다.

이 가이드에 설명된 지침과 모범 사례를 따르면 의료 기관은 AI/ML 기술의 힘을 활용하는 동시에 의료 도메인의 복잡성을 탐색할 수 있습니다. 이 접근 방식은 윤리적 및 규제 지침 준수를 지원하고 의료 분

야에서 AI 시스템의 책임감 있는 사용을 장려합니다. 정확하고 비공개인 인사이트를 생성하도록 설계되었습니다.

대상 독자

이 가이드는 의료 데이터 분석 및 자동화를 위한 AI 기반 자연어 처리 솔루션을 구현하려는 기술 이해 관계자, 아키텍트, 기술 책임자 및 의사 결정자를 대상으로 합니다.

목표

의료 및 생명과학 조직은 Amazon Comprehend Medical 및 LLMs. 이러한 결과에는 일반적으로 운영 효율성 향상, 비용 절감, 환자 치료 개선이 포함됩니다. 이 섹션에서는 주요 비즈니스 목표와 이 가이드에 설명된 전략 및 모범 사례 구현의 관련 이점을 간략하게 설명합니다.

다음은 조직이 이 가이드의 지침과 모범 사례를 구현하여 달성할 수 있는 몇 가지 목표입니다.

- 개발 시간 단축 - 이 가이드의 궁극적인 목표는 비용과 함께 개발 시간을 줄이고, 기술 부채를 줄이고, POC의 잠재적 프로젝트 장애를 완화하는 것입니다. Amazon Comprehend Medical과 같은 주요 AI/ML 서비스와 의료 작업에 대한 LLM 사용의 장단점을 이해하면 기업은 시장 출시 시간을 단축하고 비즈니스 목표 달성 속도를 높일 수 있습니다.
- 의료 코딩 작업을 자동화하기 위한 정보 추출 - 환자 방문 후 코딩 전문가와 공급자는 주관적, 목표, 평가 및 계획(SOAP) 메모와 같은 의료 텍스트에서 인사이트를 추출할 수 있습니다. 이렇게 하면 수동 문서화 작업을 줄이고 공급자가 환자의 필요에 집중할 수 있습니다. Amazon Comprehend Medical의 개체 인식 기능을 LLMs와 결합하여 조직은 환자 기록, 임상 기록 및 기타 의료 데이터 소스에서 관련 의료 정보를 추출할 수 있습니다. 이렇게 하면 인적 오류를 최소화하고 일관된 관행을 촉진할 수 있습니다.
- 환자 기록 및 임상 문서 요약 - 환자 기록, 치료 계획 및 의료 결과를 자동으로 요약하면 의료 서비스 제공자에게 귀중한 시간을 절약할 수 있습니다. LLMs 포괄적이고 구조화된 임상 설명서를 생성하는데 도움이 될 수 있습니다. Amazon Comprehend Medical과 추가 컨텍스트를 가져오거나, 의료 도메인 LLM을 사용하거나, 의료 데이터로 LLM을 미세 조정할 수 있습니다. 이러한 접근 방식은 정확한 요약을 제공하고 설명서가 규정 준수 요구 사항 및 표준을 준수하는지 확인하는 데 도움이 될 수 있습니다.
- 임상 결정 및 환자 치료 지원 - Amazon Comprehend Medical에서 [온톨로지 연결](#)을 사용하고 LLMs 사용하여 공급자는 의학적 질문에 답변하거나 환자 치료 관련 권장 사항을 찾을 수 있습니다. 이를 통해 의료 전문가는 환자 결과를 개선하고 의료 오류 위험을 줄이는 정보에 입각한 결정을 내릴 수 있습니다.

의료 및 생명 과학을 위한 생성형 AI 및 NLP 접근 방식

자연어 처리(NLP)는 컴퓨터에 인간 언어를 해석, 조작 및 이해할 수 있는 기능을 제공하는 기계 학습 기술입니다. 의료 및 생명과학 조직은 환자 기록에서 대량의 데이터를 얻습니다. NLP 소프트웨어를 사용하여 데이터를 자동으로 처리할 수 있습니다. 예를 들어 NLP와 생성형 AI를 결합하여 의료 코딩을 간소화하고, 환자 정보를 추출하고, 레코드를 요약할 수 있습니다.

수행하려는 NLP 작업에 따라 사용 사례에 가장 적합한 아키텍처가 다를 수 있습니다. 이 가이드에서는 의료 및 생명과학 애플리케이션을 위한 AWS 다음과 같은 생성형 AI 및 NLP 옵션을 다룹니다.

- [Amazon Comprehend Medical 사용](#) - 대규모 언어 모델(LLM)과 통합하지 않고 Amazon Comprehend Medical을 독립적으로 사용하는 방법에 대해 알아봅니다.
- [Amazon Comprehend Medical과 대규모 언어 모델 결합](#) - 검색 증강 생성(RAG) 아키텍처에서 Amazon Comprehend Medical을 LLM과 결합하는 방법에 대해 알아봅니다.
- [의료 및 생명과학 사용 사례에 대규모 언어 모델 사용](#) - 미세 조정된 LLM 또는 RAG 아키텍처를 사용하여 의료 및 생명과학 애플리케이션에 LLM을 사용하는 방법에 대해 알아봅니다.

Amazon Comprehend Medical 사용

[Amazon Comprehend Medical](#)은 의사 기록, 퇴원 요약, 테스트 결과 및 사례 기록과 같은 비정형 임상 텍스트에서 유용한 정보를 감지하고 반환 AWS 서비스 하는입니다. 자연어 처리(NLP) 모델을 사용하여 개체를 감지합니다. 개체는 의학적 상태, 약물 또는 보호 대상 건강 정보(PHI)와 같은 의료 정보에 대한 텍스트 참조입니다.

Important

Amazon Comprehend Medical은 전문적인 의학적 조언, 진단 또는 치료를 대체할 수 없습니다. Amazon Comprehend Medical은 탐지된 엔터티의 정확성에 대한 신뢰 수준을 나타내는 신뢰도 점수를 제공합니다. 사용 사례에 적합한 신뢰 임계값을 식별하고, 높은 정확도가 필요한 상황에서는 높은 신뢰 임계값을 사용합니다. 특정 사용 사례의 경우 적절한 교육을 받은 인간 검토자가 결과를 검토하고 확인해야 합니다. 예를 들어 Amazon Comprehend Medical은 숙련된 의료 전문가로부터 정확성과 적절한 의학적 판단을 검토받은 후 환자 치료 시나리오에서만 사용해야 합니다.

AWS Management Console, AWS Command Line Interface (AWS CLI) 또는 SDK를 통해 Amazon Comprehend Medical에 액세스할 수 있습니다. AWS SDKs는 Java, Python, Ruby, .NET, iOS 및 Android와 같은 다양한 프로그래밍 언어 및 플랫폼에서 사용할 수 있습니다. SDKs를 사용하여 클라이언트 애플리케이션에서 Amazon Comprehend Medical에 프로그래밍 방식으로 액세스할 수 있습니다.

이 섹션에서는 Amazon Comprehend Medical의 주요 기능을 검토합니다. 또한 대규모 언어 모델(LLM)과 비교하여 이 서비스를 사용할 때의 이점에 대해서도 설명합니다.

Amazon Comprehend Medical 기능

Amazon Comprehend Medical은 거의 실시간 및 배치 추론을 위한 APIs를 제공합니다. 이러한 APIs의 의료 텍스트를 수집하고 의료 엔터티 인식을 사용하고 엔터티 관계를 식별하여 의료 NLP 작업에 대한 결과를 제공할 수 있습니다. 단일 파일에서 분석을 수행하거나 Amazon Simple Storage Service(Amazon S3) 버킷에 저장된 여러 파일에 대한 배치 분석으로 분석을 수행할 수 있습니다. Amazon Comprehend Medical은 동기 엔터티 감지를 위해 다음과 같은 텍스트 분석 API 작업을 제공합니다.

- [개체 감지](#) - 해부학, 의학적 상태, PHI 범주, 절차 및 시간 표현식과 같은 일반적인 의료 범주를 감지합니다.
- [PHI 감지](#) - 연령, 날짜, 이름 및 유사한 개인 정보와 같은 특정 엔터티를 감지합니다.

Amazon Comprehend Medical에는 임상 문서에 대한 배치 텍스트 분석을 수행하는 데 사용할 수 있는 여러 API 작업도 포함되어 있습니다. 이러한 API 작업을 사용하는 방법에 대한 자세한 내용은 [텍스트 분석 배치 APIs](#).

Amazon Comprehend Medical을 사용하면 임상 텍스트에서 엔터티를 탐지하고 해당 엔터티를 RxNorm, ICD-10-CM 및 SNOMED CT 지식 기반을 비롯한 표준화된 의료 온톨로지의 개념에 연결합니다. 단일 파일에서 분석을 수행하거나 Amazon S3 버킷에 저장된 대용량 문서 또는 여러 파일에 대한 배치 분석으로 분석을 수행할 수 있습니다. Amazon Comprehend Medical은 다음과 같은 온톨로지 연결 API 작업을 제공합니다.

- [InferICD10CM](#) - InferICD10CM 작업은 잠재적 의학적 상태를 감지하여 국제 질병 분류, 10차 개정, 임상 수정(ICD-10-CM) 2019 버전의 코드에 연결합니다. Amazon Comprehend Medical은 발견된 각 잠재적 의학적 상태에 대해 일치하는 ICD-10-CM 코드 및 설명을 나열합니다. 결과에 나열된 의학적 상태에는 결과에서 일치하는 개념에 대해 엔터티의 정확성에 대한 Amazon Comprehend Medical의 신뢰도를 나타내는 신뢰도 점수가 포함됩니다.
- [InferRxNorm](#) - InferRxNorm 작업은 환자 레코드에 엔터티로 나열된 약물을 식별합니다. 이것은 엔터티를 국립 의학 도서관의 RxNorm 데이터베이스에 있는 개념 식별자(RxCUI)에 연결합니다. 각

RxCUI는 강도와 투여 형태가 다르기 때문에 고유합니다. 결과에 나열된 의약품에는 RxNorm 지식 기반의 개념에 일치하는 엔터티의 정확성에 대한 Amazon Comprehend Medical의 신뢰도를 나타내는 신뢰도 점수가 포함됩니다. Amazon Comprehend Medical은 검출된 각 약물에 대해 잠재적으로 일치할 가능성이 가장 높은 RxCUI를 신뢰도 점수 내림차순으로 나열합니다.

- [InferSNOMEDCT](#) – InferSNOMEDCT 작업은 가능한 의료 개념을 엔터티로 식별하고 이를 2021년 3월 버전의 Systematized Nomenclature of Medicine, Clinical Terms(SNOMED CT)의 코드에 연결합니다. SNOMED CT는 의학적 상태 및 해부학, 의료 검사, 치료 및 절차를 포함한 의료 개념에 대한 어휘를 포괄적으로 제공합니다. Amazon Comprehend Medical은 일치하는 각 개념 ID에 대해 상위 5개 의료 개념을 반환하며, 각 개념에는 신뢰도 점수와 특성, 속성 등의 컨텍스트 정보가 포함되어 있습니다. SNOMED CT 개념 ID를 SNOMED CT 다중 계층 구조와 함께 사용하면 의료 코딩, 보고 또는 임상 분석을 위한 환자 임상 데이터를 구조화하는 데 사용할 수 있습니다.

자세한 내용은 Amazon Comprehend Medical 설명서의 [텍스트 분석 APIs](#) 및 [온톨로지 연결 APIs](#)를 참조하세요.

Amazon Comprehend Medical 사용 사례

독립 실행형 서비스인 Amazon Comprehend Medical은 조직의 사용 사례를 해결할 수 있습니다. Amazon Comprehend Medical은 다음과 같은 작업을 수행할 수 있습니다.

- 환자 레코드의 의료 코딩 지원
- 보호 대상 건강 정보(PHI) 데이터 감지
- 용량, 빈도 및 형태와 같은 속성을 포함한 약물 검증

Amazon Comprehend Medical 결과는 대부분의 의료 관행에 맞게 해석할 수 있습니다. 그러나 다음과 같은 제한이 있는 경우 대안을 고려해야 할 수 있습니다.

- 다른 개체 정의 - 예를 들어 약물 개체 FREQUENCY의 정의가 다를 수 있습니다. Amazon Comprehend Medical은 빈도에 따라 필요에 따라 예측하지만 조직에서 pro re nata(PRN)라는 용어를 사용할 수 있습니다.
- 결과 양이 너무 많음 - 예를 들어 환자 메모에는 여러 ICD-10-CM 코드에 매핑되는 여러 증상과 키워드가 자주 포함됩니다. 그러나 몇 가지 키워드는 진단에 적용할 수 없습니다. 이 경우 공급자는 수작업 처리 시간이 필요한 수많은 ICD-10-CM 엔터티와 신뢰도 점수를 평가해야 합니다.
- 사용자 지정 엔터티 또는 NLP 작업 - 예를 들어 공급자는 통증에 대해 필요에 따라 취하는 것과 같은 PRN 증거를 추출할 수 있습니다. Amazon Comprehend Medical에서는 사용할 수 없으므로 다른 AI/

ML 모델이 필요합니다. NLP 작업이 요약, 질문 응답, 감정 분석과 같은 엔터티 인식 범위를 벗어나는 경우 다른 AI/ML 솔루션이 필요합니다.

Amazon Comprehend Medical과 대규모 언어 모델 결합

[NEJM AI의 2024년 연구에](#) 따르면 의료 코딩 작업에 제로샷 프롬프트와 함께 LLM을 사용하면 일반적으로 성능이 저하되는 것으로 나타났습니다. Amazon Comprehend Medical을 LLM과 함께 사용하면 이러한 성능 문제를 완화하는 데 도움이 될 수 있습니다. Amazon Comprehend Medical 결과는 NLP 작업을 수행하는 LLM에 유용한 컨텍스트입니다. 예를 들어 Amazon Comprehend Medical의 컨텍스트를 대규모 언어 모델에 제공하면 다음과 같은 이점이 있습니다.

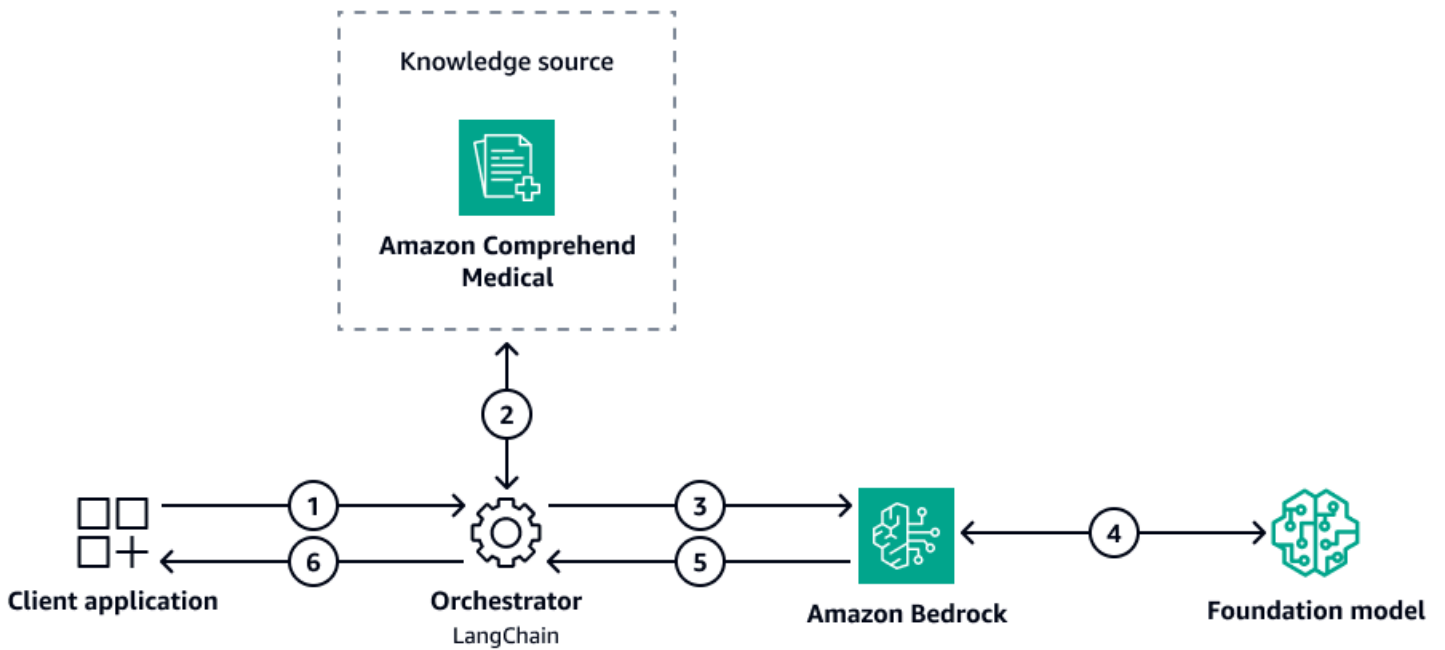
- Amazon Comprehend Medical의 초기 결과를 LLM의 컨텍스트로 사용하여 개체 선택의 정확도 향상
- 사용자 지정 엔터티 인식, 요약, 질문 응답 및 추가 사용 사례 구현

이 섹션에서는 검색 증강 생성(RAG) 접근 방식을 사용하여 Amazon Comprehend Medical을 LLM과 결합하는 방법을 설명합니다. Retrieval Augmented Generation(RAG)은 응답을 생성하기 전에 LLM이 훈련 데이터 소스 외부에 있는 신뢰할 수 있는 데이터 소스를 참조하는 생성형 AI 기술입니다. 자세한 내용은 [검색 증강 생성\(RAG\)이란 무엇인가요?](#)를 참조하세요.

이 접근 방식을 설명하기 위해 이 섹션에서는 ICD-10-CM과 관련된 의료(진료) 코딩의 예를 사용합니다. 여기에는 혁신을 가속화하는 데 도움이 되는 샘플 아키텍처와 프롬프트 엔지니어링 템플릿이 포함되어 있습니다. 또한 RAG 워크플로 내에서 Amazon Comprehend Medical을 사용하는 모범 사례도 포함되어 있습니다.

Amazon Comprehend Medical을 사용한 RAG 기반 아키텍처

다음 다이어그램은 환자 메모에서 ICD-10-CM 진단 코드를 식별하기 위한 RAG 접근 방식을 보여줍니다. Amazon Comprehend Medical을 지식 소스로 사용합니다. RAG 접근 방식에서 검색 방법은 일반적으로 해당 지식이 포함된 벡터 데이터베이스에서 정보를 검색합니다. 이 아키텍처는 벡터 데이터베이스 대신 검색 작업에 Amazon Comprehend Medical을 사용합니다. 오케스트레이터는 환자 메모 정보를 Amazon Comprehend Medical로 전송하고 ICD-10-CM 코드 정보를 검색합니다. 오케스트레이터는 Amazon Bedrock을 통해 컨텍스트를 다운스트림 파운데이션 모델(LLM)로 보냅니다. LLM은 ICD-10-CM 코드 정보를 사용하여 응답을 생성하고 해당 응답은 클라이언트 애플리케이션으로 다시 전송됩니다.



다이어그램은 다음 RAG 워크플로를 보여줍니다.

1. 클라이언트 애플리케이션은 환자 메모를 오케스트레이터에 쿼리로 전송합니다. 이러한 환자 참고 사항의 예로는 “환자는 X 박사의 71세 여성 환자입니다. 환자가 지난 저녁에 약 7일~8일의 복부 통증 이력이 있으며, 이는 지속적이었습니다. 이 여성에게는 명확한 열이나 오한이 없었고, 황달 이력도 없었습니다. 환자는 최근에 체중이 크게 줄었다는 사실을 부인합니다.”
2. 오케스트레이터는 Amazon Comprehend Medical을 사용하여 쿼리의 의료 정보와 관련된 ICD-10-CM 코드를 검색합니다. InferICD10CM API를 사용하여 환자 메모에서 ICD-10-CM 코드를 추출하고 유추합니다.
3. 오케스트레이터는 프롬프트 템플릿, 원본 쿼리 및 Amazon Comprehend Medical에서 검색된 ICD-10-CM 코드가 포함된 프롬프트를 구성합니다. 이 향상된 컨텍스트를 Amazon Bedrock으로 전송합니다.
4. Amazon Bedrock은 입력을 처리하고 파운데이션 모델을 사용하여 ICD-10-CM 코드와 쿼리의 해당 증거가 포함된 응답을 생성합니다. 생성된 응답에는 식별된 ICD-10-CM 코드와 각 코드를 지원하는 환자 메모의 증거가 포함됩니다. 다음은 샘플 응답입니다.

```
<response>
<icd10>
<code>R10.9</code>
<evidence>history of abdominal pain</evidence>
</icd10>
<icd10>
```

```
<code>R10.30</code>
<evidence>history of abdominal pain</evidence>
</icd10>
</response>
```

5. Amazon Bedrock은 생성된 응답을 오케스트레이터에 전송합니다.
6. 오케스트레이터는 응답을 클라이언트 애플리케이션으로 다시 전송하여 사용자가 응답을 검토할 수 있도록 합니다.

RAG 워크플로에서 Amazon Comprehend Medical을 사용하는 사용 사례

Amazon Comprehend Medical은 특정 NLP 작업을 수행할 수 있습니다. 자세한 내용은 [Amazon Comprehend Medical 사용 사례를 참조하세요](#).

다음과 같은 고급 사용 사례를 위해 Amazon Comprehend Medical을 RAG 워크플로에 통합할 수 있습니다.

- 추출된 의료 엔터티를 환자 레코드의 컨텍스트 정보와 결합하여 자세한 임상 요약 생성
- 코드 할당을 위해 온톨로지 연결 정보와 함께 추출된 개체를 사용하여 복잡한 사례에 대한 의료 코드 자동화
- 추출된 의료 엔터티를 사용하여 구조화되지 않은 텍스트에서 구조화된 임상 노트 생성을 자동화합니다.
- 추출된 약물 이름 및 속성을 기반으로 약물 부작용 분석
- 추출된 의료 정보를 up-to-date 연구 및 지침과 결합하는 지능형 임상 지원 시스템 개발

RAG 워크플로에서 Amazon Comprehend Medical을 사용하는 모범 사례

Amazon Comprehend Medical 결과를 LLM 프롬프트에 통합할 때는 모범 사례를 따르는 것이 중요합니다. 이렇게 하면 성능과 정확도가 향상될 수 있습니다. 다음은 주요 권장 사항입니다.

- Amazon Comprehend Medical 신뢰도 점수 이해 - Amazon Comprehend Medical은 탐지된 각 개체 및 온톨로지 연결에 대한 신뢰도 점수를 제공합니다. 이러한 점수의 의미를 이해하고 특정 사용 사례에 적합한 임계값을 설정하는 것이 중요합니다. 신뢰도 점수는 신뢰도가 낮은 엔터티를 필터링하여 노이즈를 줄이고 LLM 입력의 품질을 개선하는 데 도움이 됩니다.
- 프롬프트 엔지니어링에서 신뢰도 점수 사용 - LLM에 대한 프롬프트를 생성할 때 Amazon Comprehend Medical 신뢰도 점수를 추가 컨텍스트로 통합하는 것이 좋습니다. 이를 통해 LLM은 신

회도 수준에 따라 개체의 우선순위를 지정하거나 가중치를 부여하여 잠재적으로 출력 품질을 개선할 수 있습니다.

- 실측 데이터를 사용하여 Amazon Comprehend Medical 결과 평가 - 실측 데이터는 사실로 알려진 정보입니다. AI/ML 애플리케이션이 정확한 결과를 생성하고 있는지 검증하는 데 사용할 수 있습니다. Amazon Comprehend Medical 결과를 LLM 워크플로에 통합하기 전에 데이터의 대표 샘플에서 서비스의 성능을 평가합니다. 결과를 실측 주석과 비교하여 잠재적 불일치 또는 개선 영역을 식별합니다. 이 평가는 사용 사례에 대한 Amazon Comprehend Medical의 강점과 제한 사항을 이해하는 데 도움이 됩니다.
- 전략적으로 관련 정보 선택 - Amazon Comprehend Medical은 대량의 정보를 제공할 수 있지만 모든 정보가 작업과 관련이 있는 것은 아닙니다. 사용 사례와 가장 관련성이 높은 엔터티, 속성 및 메타데이터를 신중하게 선택합니다. LLM에 관련 없는 정보를 너무 많이 제공하면 노이즈가 발생하여 성능이 저하될 수 있습니다.
- 개체 정의 정렬 - Amazon Comprehend Medical에서 사용하는 개체 및 속성의 정의가 해석과 일치하는지 확인합니다. 불일치가 있는 경우 LLM에 추가 컨텍스트 또는 설명을 제공하여 Amazon Comprehend Medical 출력과 요구 사항 간의 격차를 해소하는 것이 좋습니다. Amazon Comprehend Medical 엔터티가 기대치를 충족하지 않는 경우 프롬프트에 추가 지침(및 가능한 예)을 포함하여 사용자 지정 엔터티 감지를 구현할 수 있습니다.
- 도메인별 지식 제공 - Amazon Comprehend Medical은 중요한 의료 정보를 제공하지만 특정 도메인의 모든 미묘한 차이를 캡처하지는 못할 수 있습니다. Amazon Comprehend Medical 결과를 온톨로지, 용어 또는 전문가가 선별한 데이터 세트와 같은 추가 도메인별 지식 소스로 보완하는 것이 좋습니다. 이를 통해 LLM에 대한 보다 포괄적인 컨텍스트가 제공됩니다.
- 윤리적 및 규제 지침 준수 - 의료 데이터를 처리할 때는 데이터 프라이버시, 보안 및 의료 분야에서 AI 시스템의 책임 있는 사용과 관련된 것과 같은 윤리적 원칙과 규제 지침을 준수하는 것이 중요합니다. 구현이 관련 법률과 업계 모범 사례를 준수하는지 확인합니다.

AI/ML 실무자는 이러한 모범 사례를 따르면 Amazon Comprehend Medical과 LLMs. 의료 NLP 작업의 경우 이러한 모범 사례는 잠재적 위험을 완화하고 성능을 개선할 수 있습니다.

Amazon Comprehend Medical 컨텍스트에 대한 프롬프트 엔지니어링

[프롬프트 엔지니어링](#)은 원하는 출력을 생성하도록 생성형 AI 솔루션을 안내하는 프롬프트를 설계하고 구체화하는 프로세스입니다. AI가 사용자와 더 의미 있게 상호 작용하도록 안내하는 가장 적절한 형식, 문구, 단어 및 기호를 선택합니다.

수행하는 API 작업에 따라 Amazon Comprehend Medical은 감지된 엔터티, 온톨로지 코드 및 설명, 신뢰도 점수를 반환합니다. 이러한 결과는 솔루션이 대상 LLM을 호출할 때 프롬프트 내에서 컨텍스트가 됩니다. 프롬프트 템플릿 내에 컨텍스트를 표시하도록 프롬프트를 엔지니어링해야 합니다.

Note

이 섹션의 예제 프롬프트는 [Anthropic 지침](#)을 따릅니다. 다른 LLM 공급자를 사용하는 경우 해당 공급자의 권장 사항을 따르십시오.

일반적으로 원본 의료 텍스트와 Amazon Comprehend Medical 결과를 모두 프롬프트에 삽입합니다. 다음은 일반적인 프롬프트 구조입니다.

```
<medical_text>
medical text
</medical_text>

<comprehend_medical_text_results>
comprehend medical text results
</comprehend_medical_text_results>

<prompt_instructions>
prompt instructions
</prompt_instructions>
```

이 섹션에서는 Amazon Comprehend Medical 결과를 다음과 같은 일반적인 의료 NLP 작업에 대한 프롬프트 컨텍스트로 포함하는 전략을 제공합니다.

- [Amazon Comprehend Medical 결과 필터링](#)
- [Amazon Comprehend Medical을 사용하여 의료 NLP 작업 확장](#)
- [Amazon Comprehend Medical을 사용하여 가드레일 적용](#)

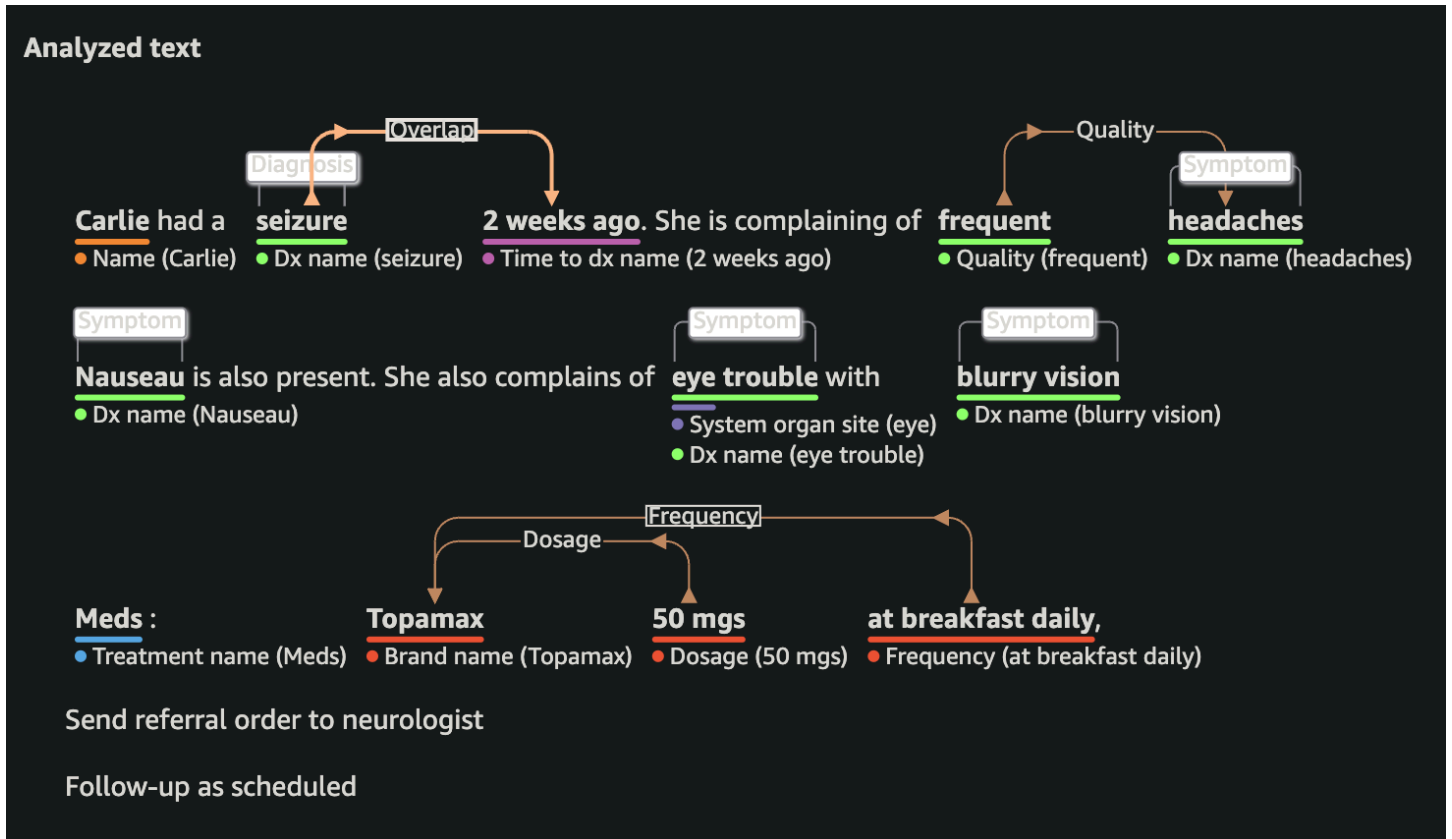
Amazon Comprehend Medical 결과 필터링

Amazon Comprehend Medical은 일반적으로 대량의 정보를 제공합니다. 의료 전문가가 검토해야 하는 결과 수를 줄일 수 있습니다. 이 경우 LLM을 사용하여 이러한 결과를 필터링할 수 있습니다. Amazon Comprehend Medical 엔터티에는 프롬프트를 설계할 때 필터링 메커니즘으로 사용할 수 있는 신뢰도 점수가 포함되어 있습니다.

다음은 환자 메모의 예입니다.

Carlie had a seizure 2 weeks ago. She is complaining of frequent headaches
 Nausea is also present. She also complains of eye trouble with blurry vision
 Meds : Topamax 50 mgs at breakfast daily,
 Send referral order to neurologist
 Follow-up as scheduled

이 환자 참고 사항에서 Amazon Comprehend Medical은 다음 개체를 감지합니다.



개체는 다음 ICD-10-CM 코드에 연결되어 있습니다.

카테고리	ICD-10-CM 코드	ICD-10-CM 설명	신뢰도 점수
압수	R56.9	지정되지 않은 총동	0.8348
압수	G40.909	뇌전증 지속증이 없는 상세 불명, 난치성 뇌전증	0.5424

압수	R56.00	단순성 심음 총동	0.4937
압수	G40.09	기타 발작	0.4397
압수	G40.409	뇌전증 지속 상태가 없 는 난치성이 아닌 기타 전신성 뇌전증 및 뇌전 증 증상	0.4138
머리카락	R51	머리카락	0.4067
머리카락	R51.9	상세 불명 머리카락	0.3844
머리카락	G44.52	새로운 일일 지속형두 통(NDPH)	0.3005
머리카락	G44	기타 골치 아픈 증상	0.2670
머리카락	G44.8	기타 명시된 골치 아픈 증상	0.2542

프롬프트에 ICD-10-CM 코드를 전달하여 LLM 정밀도를 높일 수 있습니다. 노이즈를 줄이기 위해 Amazon Comprehend Medical 결과에 포함된 신뢰도 점수를 사용하여 ICD-10-CM 코드를 필터링할 수 있습니다. 다음은 신뢰도 점수가 0.4보다 높은 ICD-10-CM 코드만 포함하는 예제 프롬프트입니다.

```
<patient_note>
Carlie had a seizure 2 weeks ago. She is complaining of frequent headaches
Nausea is also present. She also complains of eye trouble with blurry vision
Meds : Topamax 50 mgs at breakfast daily,
Send referral order to neurologist
Follow-up as scheduled
</patient_note>

<comprehend_medical_results>
<icd-10>
  <entity>
    <text>seizure</text>
    <code>
      <description>Unspecified convulsions</description>
      <code_value>R56.9</code_value>
      <score>0.8347607851028442</score>
```

```

</code>
<code>
  <description>Epilepsy, unspecified, not intractable, without status epilepticus</
description>
  <code_value>G40.909</code_value>
  <score>0.542376697063446</score>
</code>
<code>
  <description>Other seizures</description>
  <code_value>G40.89</code_value>
  <score>0.43966275453567505</score>
</code>
<code>
  <description>Other generalized epilepsy and epileptic syndromes, not intractable,
without status epilepticus</description>
  <code_value>G40.409</code_value>
  <score>0.41382506489753723</score>
</code>
</entity>
<entity>
  <text>headaches</text>
  <code>
    <description>Headache</description>
    <code_value>R51</code_value>
    <score>0.4066613018512726</score>
  </code>
</entity>
<entity>
  <text>Nausea</text>
  <code>
    <description>Nausea</description>
    <code_value>R11.0</code_value>
    <score>0.6460834741592407</score>
  </code>
</entity>
<entity>
  <text>eye trouble</text>
  <code>
    <description>Unspecified disorder of eye and adnexa</description>
    <code_value>H57.9</code_value>
    <score>0.6780954599380493</score>
  </code>
  <code>
    <description>Unspecified visual disturbance</description>

```

```

    <code_value>H53.9</code_value>
    <score>0.5871203541755676</score>
  </code>
  <code>
    <description>Unspecified disorder of binocular vision</description>
    <code_value>H53.30</code_value>
    <score>0.5539672374725342</score>
  </code>
</entity>
<entity>
  <text>blurry vision</text>
  <code>
    <description>Other visual disturbances</description>
    <code_value>H53.8</code_value>
    <score>0.9001834392547607</score>
  </code>
</entity>
</icd-10>
</comprehend_medical_results>

<prompt>
Given the patient note and Amazon Comprehend Medical ICD-10-CM code results above,
please select the most relevant ICD-10-CM diagnosis codes for the patient.
For each selected code, provide a brief explanation of why it is relevant based on the
information in the patient note.
</prompt>

```

Amazon Comprehend Medical을 사용하여 의료 NLP 작업 확장

의료 텍스트를 처리할 때 Amazon Comprehend Medical의 컨텍스트는 LLM이 더 나은 토큰을 선택하는 데 도움이 될 수 있습니다. 이 예제에서는 진단 증상을 의약품과 일치시키려고 합니다. 또한 피패널 검사와 관련된 용어와 같이 의료 검사와 관련된 텍스트를 찾고자 합니다. Amazon Comprehend Medical을 사용하여 엔터티와 약물 이름을 감지할 수 있습니다. 이 경우 Amazon Comprehend Medical용 [DetectEntitiesV2](#) 및 [InferRxNorm](#) APIs 사용합니다.

다음은 환자 메모의 예입니다.

```

Carlie had a seizure 2 weeks ago. She is complaining of increased frequent headaches
Given lyme disease symptoms such as muscle ache and stiff neck will order prescription.
Meds : Topamax 50 mgs at breakfast daily. Amoxicillan 25 mg by mouth twice a day
Place MRI radiology order at RadNet

```

진단 코드에 초점을 맞추기 위해 프롬프트에는 유형의 MEDICAL_CONDITION와 관련된 엔터티만 DX_NAME 사용됩니다. 관련성 때문에 다른 메타데이터는 제외됩니다. 약물 개체의 경우 추출된 속성과 함께 약물 이름이 포함됩니다. Amazon Comprehend Medical의 다른 약물 개체 메타데이터는 관련성으로 인해 제외됩니다. 다음은 필터링된 Amazon Comprehend Medical 결과를 사용하는 예제 프롬프트입니다. 프롬프트는 DX_NAME 유형이 있는 MEDICAL_CONDITION 엔터티에 중점을 둡니다. 이 프롬프트는 진단 코드를 의약품과 더 정확하게 연결하고 의료 주문 테스트를 더 정확하게 추출하도록 설계되었습니다.

```
<patient_note>
Charlie had a seizure 2 weeks ago. She is complaining of increased frequent headaches
Given lyme disease symptoms such as muscle ache and stiff neck will order
prescription.
Meds : Topamax 50 mgs at breakfast daily. Amoxicillan 25 mg by mouth twice a day
Place MRI radiology order at RadNet
</patient_note>

<detect_entity_results>
<entity>
  <text>seizure</text>
  <category>MEDICAL_CONDITION</category>
  <type>DX_NAME</type>
</entity>
<entity>
  <text>headaches</text>
  <category>MEDICAL_CONDITION</category>
  <type>DX_NAME</type>
</entity>
<entity>
  <text>lyme disease</text>
  <category>MEDICAL_CONDITION</category>
  <type>DX_NAME</type>
</entity>
<entity>
  <text>muscle ache</text>
  <category>MEDICAL_CONDITION</category>
  <type>DX_NAME</type>
</entity>
<entity>
  <text>stiff neck</text>
  <category>MEDICAL_CONDITION</category>
  <type>DX_NAME</type>
</entity>
```

```
</detect_entity_results>

<rx_results>
<entity>
  <text>Topamax</text>
  <category>MEDICATION</category>
  <type>BRAND_NAME</type>
  <attributes>
    <attribute>
      <type>FREQUENCY</type>
      <text>at breakfast daily</text>
    </attribute>
    <attribute>
      <type>DOSAGE</type>
      <text>50 mgs</text>
    </attribute>
    <attribute>
      <type>ROUTE_OR_MODE</type>
      <text>by mouth</text>
    </attribute>
  </attributes>
</entity>
<entity>
  <text>Amoxicillan</text>
  <category>MEDICATION</category>
  <type>GENERIC_NAME</type>
  <attributes>
    <attribute>
      <type>ROUTE_OR_MODE</type>
      <text>by mouth</text>
    </attribute>
    <attribute>
      <type>DOSAGE</type>
      <text>25 mg</text>
    </attribute>
    <attribute>
      <type>FREQUENCY</type>
      <text>twice a day</text>
    </attribute>
  </attributes>
</entity>
</rx_results>

<prompt>
```

```
Based on the patient note and the detected entities, can you please:
1. Link the diagnosis symptoms with the medications prescribed.
Provide your reasoning for the linkages.
2. Extract any entities related to medical order tests mentioned in the note.
</prompt>
```

Amazon Comprehend Medical을 사용하여 가드레일 적용

생성된 응답을 사용하기 전에 LLM 및 Amazon Comprehend Medical을 사용하여 가드레일을 생성할 수 있습니다. 수정되지 않은 의료 텍스트 또는 사후 처리된 의료 텍스트에서이 워크플로를 실행할 수 있습니다. 사용 사례에는 보호 대상 건강 정보(PHI) 해결, 할루시네이션 감지 또는 결과 게시를 위한 사용자 지정 정책 구현이 포함됩니다. 예를 들어 Amazon Comprehend Medical의 컨텍스트를 사용하여 PHI 데이터를 식별한 다음 LLM을 사용하여 해당 PHI 데이터를 제거할 수 있습니다.

다음은 PHI를 포함하는 환자 레코드의 정보 예제입니다.

```
Patient name: John Doe
Patient SSN: 123-34-5678
Patient DOB: 01/01/2024
Patient address: 123 Main St, Anytown USA
Exam details: good health. Pulse is 60 bpm. needs to work on diet with BMI of 190
```

다음은 Amazon Comprehend Medical 결과를 컨텍스트로 포함하는 예제 프롬프트입니다.

```
<original_text>
Patient name: John Doe
Patient SSN: 123-34-5678 Patient DOB: 01/01/2024
Patient address: 123 Main St, Anytown USA
Exam details: good health. Pulse is 60 bpm. needs to work on diet with BMI of 190
</original_text>

<comprehend_medical_phi_entities>
<entity>
  <text>John Doe</text>
  <category>PROTECTED_HEALTH_INFORMATION</category>
  <score>0.9967944025993347</score>
  <type>NAME</type>
</entity>
<entity>
  <text>123-34-5678</text>
  <category>PROTECTED_HEALTH_INFORMATION</category>
  <score>0.9998034834861755</score>
```

```

<type>ID</type>
</entity>
<entity>
  <text>01/01/2000</text>
  <category>PROTECTED_HEALTH_INFORMATION</category>
  <score>0.9964448809623718</score>
  <type>DATE</type>
</entity>
</comprehend_medical_phi_entities>

<instructions>
Using the provided original text and the Amazon Comprehend Medical PHI entities
detected, please analyze the text to determine if it contains any additional protected
health information (PHI) beyond the entities already identified. If additional PHI is
found, please list and categorize it. If no additional PHI is found, please state that
explicitly.
In addition if PHI is found, generate updated text with the PHI removed.
</instructions>

```

의료 및 생명과학 사용 사례에 대규모 언어 모델 사용

의료 및 생명과학 애플리케이션에 대규모 언어 모델(LLMs)을 사용하는 방법을 설명합니다. 일부 사용 사례에서는 생성형 AI 기능을 위해 대규모 언어 모델을 사용해야 합니다. state-of-the-art LLMs에도 장점과 제한이 있으며, 이 섹션의 권장 사항은 목표 결과를 달성하는 데 도움이 되도록 설계되었습니다.

결정 경로를 사용하여 도메인 지식 및 사용 가능한 훈련 데이터와 같은 요소를 고려하여 사용 사례에 적합한 LLM 솔루션을 결정할 수 있습니다. 또한 이 섹션에서는 널리 사용되는 사전 훈련된 의료 LLMs 과 선택 및 사용에 대한 모범 사례를 설명합니다. 또한 복잡한 고성능 솔루션과 더 간단하고 저렴한 접근 방식 간의 장단점을 설명합니다.

LLM 사용 사례

Amazon Comprehend Medical은 특정 NLP 작업을 수행할 수 있습니다. 자세한 내용은 [Amazon Comprehend Medical 사용 사례](#) 단원을 참조하십시오.

LLM의 논리적 및 생성형 AI 기능은 다음과 같은 고급 의료 및 생명과학 사용 사례에 필요할 수 있습니다.

- 사용자 지정 의료 엔터티 또는 텍스트 범주 분류
- 임상 질문에 대한 답변

- 의료 보고서 요약
- 의료 정보에서 인사이트 생성 및 탐지

사용자 지정 접근 방식

LLMs 구현되는 방식을 이해하는 것이 중요합니다. LLMs은 일반적으로 여러 도메인의 훈련 데이터를 포함하여 수십억 개의 파라미터로 훈련됩니다. 이 훈련을 통해 LLM은 대부분의 일반화된 작업을 처리할 수 있습니다. 그러나 도메인별 지식이 필요할 때 문제가 발생하는 경우가 많습니다. 의료 및 생명과학에 대한 도메인 지식의 예로는 정확한 답변을 생성하는 데 필요한 임상 코드, 의학 용어 및 건강 정보가 있습니다. 따라서 이러한 사용 사례에 대해 LLM을 있는 그대로 사용하면(도메인 지식을 보완하지 않고 제로샷 프롬프트) 부정확한 결과가 나올 수 있습니다. 이 문제를 해결하는 데 사용할 수 있는 몇 가지 일반적인 접근 방식은 프롬프트 엔지니어링, 검색 증강 생성(RAG) 및 미세 조정입니다.

프롬프트 엔지니어링

프롬프트 엔지니어링은 입력을 LLM에 맞게 조정하여 원하는 출력을 생성하도록 생성형 AI 솔루션을 안내하는 프로세스입니다. 관련 컨텍스트로 정확한 프롬프트를 생성하면 추론이 필요한 전문 의료 작업을 완료하기 위해 모델을 안내할 수 있습니다. 효과적인 프롬프트 엔지니어링은 모델 수정 없이도 의료 사용 사례의 모델 성능을 크게 개선할 수 있습니다. 프롬프트 엔지니어링에 대한 자세한 내용은 [Amazon Bedrock을 사용한 고급 프롬프트 엔지니어링 구현](#)(AWS 블로그 게시물)을 참조하세요. 퓨샷 프롬프트 및 chain-of-thought 프롬프트는 프롬프트 엔지니어링에 사용할 수 있는 기법입니다.

퓨샷 프롬프팅

퓨샷 프롬프트는 LLM에 원하는 입력 출력의 몇 가지 예를 제공한 후 유사한 작업을 수행하도록 요청하는 기법입니다. 의료 환경에서이 접근 방식은 의료 기관 인식 또는 임상 노트 요약과 같은 특수 작업에 특히 유용합니다. 프롬프트에 3~5개의 고품질 예제를 포함하면 의료 용어 및 도메인별 패턴에 대한 모델의 이해를 크게 개선할 수 있습니다. 몇 번의 프롬프트 예제는 [Amazon Bedrock의 LLMs을 참조하세요](#).AWS

예를 들어 임상 노트에서 약물 용량을 추출할 때 의료 전문가가 처방을 문서화하는 방법의 차이를 모델이 인식하는 데 도움이 되는 다양한 표기법 스타일의 예를 제공할 수 있습니다. 이 접근 방식은 표준화된 설명서 형식으로 작업하거나 데이터에 일관된 패턴이 존재할 때 특히 효과적입니다.

Chain-of-thought 프롬프트

Chain-of-thought(CoT) 프롬프트는 단계별 step-by-step 추론 프로세스를 통해 LLM을 안내합니다. 따라서 복잡한 의료 결정 지원 및 진단 추론 작업에 유용합니다. 임상 시나리오를 분석할 때 모델에 "단계별 사고"를 명시적으로 지시하면 의학적 추론 프로토콜을 따르고 진단 오류를 줄일 수 있습니다.

이 기법은 임상 추론에 차등 진단 또는 치료 계획과 같은 여러 논리적 단계가 필요한 경우에 적합합니다. 그러나 이 접근 방식에는 모델의 훈련 데이터 외부에서 고도로 전문화된 의료 지식을 처리하거나 중요한 치료 결정에 절대 정밀도가 필요한 경우 제한이 있습니다.

이러한 경우 CoT를 다른 접근 방식과 결합하면 더 나은 결과를 얻을 수 있습니다. 한 가지 옵션은 CoT를 자체 일관성 프롬프트와 결합하는 것입니다. 자세한 내용은 [Amazon Bedrock에서 자체 일관성 프롬프트를 사용하여 생성형 언어 모델의 성능 향상](#)(AWS 블로그 게시물)을 참조하세요. 또 다른 옵션은 ReAct 프롬프트와 같은 추론 프레임워크를 RAG와 결합하는 것입니다. 자세한 내용은 [RAG 및 ReAct 프롬프트를 사용하여 고급 생성형 AI 채팅 기반 어시스턴트 개발](#)(AWS 권고 가이드)을 참조하세요.

검색 증강 생성

Retrieval Augmented Generation(RAG)은 응답을 생성하기 전에 LLM이 훈련 데이터 소스 외부에 있는 신뢰할 수 있는 데이터 소스를 참조하는 생성형 AI 기술입니다. RAG 시스템은 지식 소스에서 의료 온톨로지 정보(예: 질병 국제 분류, 국가 약물 파일, 의료 주제 제목)를 검색할 수 있습니다. 이를 통해 LLM에 대한 추가 컨텍스트를 제공하여 의료 NLP 작업을 지원합니다.

[Amazon Comprehend Medical과 대규모 언어 모델 결합](#) 단원에서 설명한 대로 RAG 접근 방식을 사용하여 Amazon Comprehend Medical에서 컨텍스트를 검색할 수 있습니다. 다른 일반적인 지식 소스에는 Amazon OpenSearch Service, Amazon Kendra 또는 Amazon Aurora와 같은 데이터베이스 서비스에 저장된 의료 도메인 데이터가 포함됩니다. 이러한 지식 소스에서 정보를 추출하면 검색 성능, 특히 벡터 데이터베이스를 사용하는 의미 체계 쿼리에 영향을 미칠 수 있습니다.

도메인별 지식을 저장하고 검색하는 또 다른 옵션은 RAG 워크플로에서 [Amazon Q Business](#)를 사용하는 것입니다. Amazon Q Business는 내부 문서 리포지토리 또는 공개 웹 사이트(예: ICD-10 데이터의 경우 [CMS.gov](#))를 인덱싱할 수 있습니다. 그런 다음 Amazon Q Business는 쿼리를 LLM에 전달하기 전에 이러한 소스에서 관련 정보를 추출할 수 있습니다.

사용자 지정 RAG 워크플로를 구축하는 방법에는 여러 가지가 있습니다. 예를 들어 지식 소스에서 데이터를 검색하는 방법에는 여러 가지가 있습니다. 간소화를 위해 Amazon OpenSearch Service와 같은 벡터 데이터베이스를 사용하여 지식을 임베딩으로 저장하는 일반적인 검색 접근 방식을 사용하는 것이 좋습니다. 이렇게 하려면 문장 변환기와 같은 임베딩 모델을 사용하여 쿼리 및 벡터 데이터베이스에 저장된 지식에 대한 임베딩을 생성해야 합니다.

완전 관리형 및 사용자 지정 RAG 접근 방식에 대한 자세한 내용은 [의 증강 생성 옵션 및 아키텍처 검색을 참조하세요 AWS](#).

미세 조정

기존 모델을 미세 조정하려면 Amazon Titan, Mistral 또는 Llama 모델과 같은 LLM을 가져온 다음 모델을 사용자 지정 데이터에 맞게 조정해야 합니다. 미세 조정에는 다양한 기법이 있으며, 대부분 모델의 모든 파라미터를 수정하는 대신 몇 개의 파라미터만 수정하는 것이 포함됩니다. 이를 파라미터 효율적인 미세 조정(PEFT)이라고 합니다. 자세한 내용은 GitHub의 [Hugging Face PEFT](#)를 참조하세요.

다음은 의료 NLP 작업에 대해 LLM을 미세 조정하도록 선택할 수 있는 두 가지 일반적인 사용 사례입니다.

- **생성 작업** - 디코더 기반 모델은 생성형 AI 작업을 수행합니다. AI/ML 실무자는 실측 데이터를 사용하여 기존 LLM을 미세 조정합니다. 예를 들어 퍼블릭 의료 질문 응답 데이터 세트인 [MedQuAD](#)를 사용하여 LLM을 훈련할 수 있습니다. 미세 조정된 LLM에 쿼리를 호출할 때 LLM에 추가 컨텍스트를 제공하는 RAG 접근 방식이 필요하지 않습니다.
- **임베딩** - 인코더 기반 모델은 텍스트를 숫자 벡터로 변환하여 임베딩을 생성합니다. 이러한 인코더 기반 모델을 일반적으로 임베딩 모델이라고 합니다. 문장 변환 모델은 문장에 최적화된 특정 유형의 임베딩 모델입니다. 목표는 입력 텍스트에서 임베딩을 생성하는 것입니다. 그런 다음 임베딩은 의미 분석 또는 검색 작업에 사용됩니다. 임베딩 모델을 미세 조정하려면 훈련 데이터로 사용할 수 있는 문서와 같은 의료 지식이 있어야 합니다. 이는 문장 변환기 모델을 미세 조정하기 위해 유사성 또는 감정을 기반으로 한 텍스트 쌍으로 수행됩니다. 자세한 내용은 Hugging Face의 [문장 변환기 v3를 사용하여 임베딩 모델 훈련 및 미세 조정](#)을 참조하세요.

[Amazon SageMaker Ground Truth](#)를 사용하여 레이블이 지정된 고품질 훈련 데이터 세트를 구축할 수 있습니다. Ground Truth의 레이블이 지정된 데이터세트 결과를 사용하여 자체 모델을 훈련할 수 있습니다. 또한, 이러한 결과를 Amazon SageMaker AI 모델을 위한 훈련 데이터세트에 사용할 수도 있습니다. 명명된 엔터티 인식, 단일 레이블 텍스트 분류 및 다중 레이블 텍스트 분류에 대한 자세한 내용은 Amazon SageMaker AI 설명서의 [Ground Truth를 사용한 텍스트 레이블 지정](#)을 참조하세요.

미세 조정에 대한 자세한 내용은 이 가이드의 [의료 분야의 대규모 언어 모델 미세 조정](#)의 섹션을 참조하세요.

LLM 선택

[Amazon Bedrock](#)은 고성능 LLMs. 자세한 내용은 [Amazon Bedrock에서 지원되는 파운데이션 모델을 참조하세요](#). Amazon Bedrock에서 모델 평가 작업을 사용하여 여러 출력의 출력을 비교한 다음 사용 사례에 가장 적합한 모델을 선택할 수 있습니다. 자세한 내용은 [Amazon Bedrock 설명서의 Amazon Bedrock 평가를 사용하여 최상의 성능 모델 선택](#)을 참조하세요.

일부 LLMs이 제한적입니다. 사용 사례에서 Amazon Bedrock이 지원하지 않는 LLM 또는 LLM을 미세 조정해야 하는 경우 [Amazon SageMaker AI](#) 사용을 고려하세요. SageMaker AI에서는 미세 조정된 LLM을 사용하거나 의료 도메인 데이터에 대해 훈련된 사용자 지정 LLM을 선택할 수 있습니다.

다음 표에는 의료 도메인 데이터에 대해 훈련된 널리 사용되는 LLMs이 나열되어 있습니다.

LLM	작업	지식	아키텍처
BioBERT	정보 검색, 텍스트 분류 및 명명된 엔터티 인식	PubMed의 요약, PubMedCentral의 전체 텍스트 문서 및 일반 도메인 지식	인코더
ClinicalBERT	정보 검색, 텍스트 분류 및 명명된 엔터티 인식	전자 건강 기록(EHR) 시스템의 3,000,000개 이상의 환자 기록과 함께 대규모, 다중 센터 데이터 세트	인코더
ClinicalGPT	요약, 질문-응답 및 텍스트 생성	의료 기록, 도메인별 지식, 다원 대화 상담을 포함한 광범위하고 다양한 의료 데이터 세트	디코더
GatorTron-OG	요약, 질문-응답, 텍스트 생성 및 정보 검색	임상 기록 및 생체의학 문헌	인코더
메드-버트	정보 검색, 텍스트 분류 및 명명된 엔터티 인식	의료 텍스트, 임상 기록, 연구 논문 및 의료 관련 문서의 대규모 데이터 세트	인코더
메드-PaLM	의료 목적의 질문-응답	의료 및 생체의학 텍스트의 데이터 세트	디코더
medAlpaca	질문-응답 및 의료 대화 작업	의료 플래시카드, 위키 및 대화 데이터 세트와	디코더

같은 리소스를 포함하
는 다양한 의료 텍스트

BiomedBERT

정보 검색, 텍스트 분
류 및 명명된 엔터티
인식

PubMedCentral의
PubMedCentral 및 전
체 텍스트 문서의 독점
추상화

인코더

BioMedLM

요약, 질문-응답 및 텍
스트 생성

PubMed 지식 소스의
생체의학 문헌

디코더

다음은 사전 훈련된 의료 LLMs

- 훈련 데이터와 의료 NLP 작업과의 관련성을 이해합니다.
- LLM 아키텍처와 그 목적을 식별합니다. 인코더는 임베딩 및 NLP 작업에 적합합니다. 디코더는 생성 작업을 위한 것입니다.
- 사전 훈련된 의료 LLM을 호스팅하기 위한 인프라, 성능 및 비용 요구 사항을 평가합니다.
- 미세 조정이 필요한 경우 훈련 데이터에 대한 정확한 실측 정보 또는 지식을 확인하십시오. 개인 식별 정보(PII) 또는 보호 대상 건강 정보(PHI)를 마스킹하거나 수정해야 합니다.

실제 의료 NLP 작업은 지식 또는 의도한 사용 사례 측면에서 사전 훈련된 LLMs 다를 수 있습니다. 도메인별 LLM이 평가 벤치마크를 충족하지 않는 경우 자체 데이터 세트로 LLM을 미세 조정하거나 새 파운데이션 모델을 훈련할 수 있습니다. 새로운 파운데이션 모델을 훈련하는 것은 야심 차며 비용이 많이 드는 작업입니다. 대부분의 사용 사례에서는 기존 모델을 미세 조정하는 것이 좋습니다.

사전 훈련된 의료 LLM을 사용하거나 미세 조정할 때는 인프라, 보안 및 가드레일을 해결하는 것이 중요합니다.

인프라

Amazon Bedrock을 온디맨드 또는 배치 추론에 사용하는 것과 비교하여 사전 훈련된 의료 LLMs(일반적으로 Hugging Face에서)을 호스팅하려면 상당한 리소스가 필요합니다. 사전 훈련된 의료 LLMs 호스팅하려면 가속 컴퓨팅을 위한 ml.g5 인스턴스 또는 용 ml.inf2 인스턴스와 같은 하나 이상의 GPU가 있는 Amazon Elastic Compute Cloud(Amazon EC2) 인스턴스에서 실행되는 Amazon SageMaker AI 이미지를 사용하는 것이 일반적입니다 AWS Inferentia. GPUs 이는 LLMs 많은 양의 메모리와 디스크 공간을 소비하기 때문입니다.

보안 및 가드레일

비즈니스 규정 준수 요구 사항에 따라 Amazon Comprehend 및 Amazon Comprehend Medical을 사용하여 훈련 데이터에서 개인 식별 정보(PII) 및 보호 대상 건강 정보(PHI)를 마스킹하거나 수정하는 것이 좋습니다. 이렇게 하면 응답을 생성할 때 LLM이 기밀 데이터를 사용하는 것을 방지할 수 있습니다.

생성형 AI 애플리케이션에서 편향, 공정성 및 할루시네이션을 고려하고 평가하는 것이 좋습니다. 기존 LLM을 사용하든 미세 조정을 사용하든, 가드레일을 구현하여 유해한 응답을 방지합니다. 가드레일은 생성형 AI 애플리케이션 요구 사항 및 책임 있는 AI 정책에 맞게 사용자 지정하는 보호 장치입니다. 예를 들어 [Amazon Bedrock Guardrails](#)를 사용할 수 있습니다.

의료 분야의 대규모 언어 모델 미세 조정

이 섹션에 설명된 미세 조정 접근 방식은 윤리적 및 규제 지침 준수를 지원하고 의료 분야에서 AI 시스템의 책임감 있는 사용을 장려합니다. 정확하고 비공개인 인사이트를 생성하도록 설계되었습니다. 생성형 AI는 의료 서비스 제공을 혁신하고 있지만 정확도가 중요하고 규정 준수가 협상할 수 없는 임상 환경에서 off-the-shelf 모델이 부족한 경우가 많습니다. 도메인별 데이터로 파운데이션 모델을 미세 조정하면 이러한 격차가 커집니다. 이를 통해 엄격한 규제 표준을 준수하면서 의학 언어를 구사하는 AI 시스템을 만들 수 있습니다. 그러나 성공적인 미세 조정을 위해서는 민감한 데이터를 보호하고, 측정 가능한 결과로 AI 투자를 정당화하고, 빠르게 변화하는 의료 환경에서 임상 관련성을 유지하는 등 의료 분야의 고유한 문제를 신중하게 탐색해야 합니다.

경량 접근 방식이 한계에 도달하면 미세 조정이 전략적 투자가 됩니다. 정확도, 지연 시간 또는 운영 효율성의 이점이 필요한 상당한 컴퓨팅 및 엔지니어링 비용을 상쇄할 것으로 기대됩니다. 파운데이션 모델의 진행 속도는 빠르므로 미세 조정된 모델의 이점은 다음 메이저 모델 릴리스까지만 지속될 수 있습니다.

이 섹션에서는 의료 고객의 다음 두 가지 영향력이 큰 사용 사례에 AWS 대해 설명합니다.

- **임상 결정 지원 시스템** - 복잡한 환자 기록과 진화하는 지침을 이해하는 모델을 통해 진단 정확도를 높입니다. 미세 조정은 모델이 복잡한 환자 기록을 깊이 이해하고 특수 지침을 통합하는 데 도움이 될 수 있습니다. 이는 잠재적으로 모델 예측 오류를 줄일 수 있습니다. 그러나 이러한 이점을 민감한 대규모 데이터 세트에 대한 훈련 비용 및 고위험 임상 애플리케이션에 필요한 인프라와 비교해야 합니다. 특히 새 모델이 자주 출시될 때 향상된 정확도와 컨텍스트 인식이 투자를 정당화하나요?
- **의료 문서 분석** - 건강 보험 양도 및 책임에 관한 법률(HIPAA) 준수를 유지하면서 임상 기록, 영상 보고서 및 보험 문서 처리를 자동화합니다. 여기서 미세 조정을 통해 모델은 고유한 형식, 특수 약어 및 규제 요구 사항을 보다 효과적으로 처리할 수 있습니다. 수동 검토 시간이 단축되고 규정 준수가 개선되는 경우가 많습니다. 하지만 이러한 개선 사항이 미세 조정 리소스를 보장할 만큼 상당한지 평가

하는 것이 중요합니다. 프롬프트 엔지니어링 및 워크플로 오케스트레이션이 요구 사항을 충족할 수 있는지 확인합니다.

이러한 실제 시나리오는 초기 실험부터 모델 배포에 이르기까지 미세 조정 여정을 보여주는 동시에 모든 단계에서 의료의 고유한 요구 사항을 해결합니다.

비용 및 투자 수익 추정

다음은 LLM을 미세 조정할 때 고려해야 할 비용 요소입니다.

- 모델 크기 - 모델이 클수록 미세 조정 비용이 더 많이 듭니다.
- 데이터 세트 크기 - 미세 조정을 위한 데이터 세트 크기에 따라 컴퓨팅 비용 및 시간이 증가합니다.
- 미세 조정 전략 - 파라미터 효율적인 메서드는 전체 파라미터 업데이트에 비해 비용을 절감할 수 있습니다.

투자 수익률(ROI)을 계산할 때는 선택한 지표(예: 정확도)의 개선에 요청량(모델 사용 빈도)과 모델이 최신 버전으로 초과되기 전의 예상 기간을 곱한 값을 고려하세요.

또한 기본 LLM의 수명을 고려하세요. 6~12 개월마다 새로운 기본 모델이 나타납니다. 희귀 질병 감지기를 미세 조정하고 검증하는 데 8 개월이 걸리는 경우 최신 모델이 격차를 해소하기 전에 4 개월 동안만 우수한 성능을 얻을 수 있습니다.

사용 사례의 비용, ROI 및 잠재적 수명을 계산하여 데이터 기반 결정을 내릴 수 있습니다. 예를 들어, 임상 결정 지원 모델을 미세 조정하면 연간 수천 건의 사례에서 진단 오류가 눈에 띄게 감소하는 경우 투자가 빠르게 성과를 낼 수 있습니다. 반대로 프롬프트 엔지니어링만으로 문서 분석 워크플로를 목표 정확도에 가깝게 만드는 경우 차세대 모델이 도착할 때까지 미세 조정을 보류하는 것이 좋습니다.

미세 조정은 one-size-fits-all 아닙니다. 미세 조정을 결정하는 경우 올바른 접근 방식은 사용 사례, 데이터 및 리소스에 따라 달라집니다.

미세 조정 전략 선택

미세 조정이 의료 사용 사례에 적합한 접근 방식이라고 판단한 후 다음 단계는 가장 적합한 미세 조정 전략을 선택하는 것입니다. 몇 가지 접근 방식을 사용할 수 있습니다. 각 에는 의료 애플리케이션에 대한 고유한 장점과 장단점이 있습니다. 이러한 방법 중 선택하는 방법은 특정 목표, 사용 가능한 데이터 및 리소스 제약 조건에 따라 달라집니다.

훈련 목표

[도메인 적응형 사전 훈련\(DAPT\)](#)은 레이블이 지정되지 않은 대규모 도메인별 텍스트(예: 수백만 개의 의료 문서)에 대해 모델을 사전 훈련하는 비지도 방법입니다. 이 접근 방식은 의료 전문 약어와 방사선과 의사, 신경과 의사 및 기타 전문 공급자가 사용하는 용어를 이해하는 모델의 능력을 개선하는 데 매우 적합합니다. 그러나 DAPT에는 방대한 양의 데이터가 필요하며 특정 작업 출력을 처리하지 않습니다.

[감독 미세 조정\(SFT\)](#)은 구조화된 입력-출력 예제를 사용하여 명시적 지침을 따르도록 모델을 교육합니다. 이 접근 방식은 문서 요약 또는 임상 코딩과 같은 의료 문서 분석 워크플로에 적합합니다. 명령 튜닝은 원하는 출력과 페어링된 명시적 지침이 포함된 예제에 대해 모델이 훈련되는 일반적인 형태의 SFT입니다. 이렇게 하면 다양한 사용자 프롬프트를 이해하고 따르는 모델의 능력이 향상됩니다. 이 기법은 특정 임상 예제로 모델을 훈련시키기 때문에 의료 환경에서 특히 유용합니다. 주요 단점은 신중하게 레이블이 지정된 예제가 필요하다는 것입니다. 또한 미세 조정된 모델은 예제가 없는 옛지 케이스에서 어려움을 겪을 수 있습니다. Amazon SageMaker Jumpstart를 사용한 미세 조정에 대한 지침은 [Amazon SageMaker Jumpstart를 사용한 FLAN T5 XL에 대한 지침 미세 조정\(AWS 블로그 게시물\)](#)을 참조하세요.

[인적 피드백\(RLHF\)](#)을 통한 강화 학습은 전문가 피드백과 선호도를 기반으로 모델 동작을 최적화합니다. [근위 정책 최적화\(PPO\)](#) 또는 [직접 선호도 최적화\(DPO\)](#)와 같은 인적 선호도 및 방법을 기반으로 훈련된 보상 모델을 사용하여 모델을 최적화하는 동시에 파괴적인 업데이트를 방지합니다. RLHF는 출력을 임상 지침에 맞추고 권장 사항이 승인된 프로토콜 내에 있도록 하는 데 적합합니다. 이 접근 방식에는 피드백에 상당한 임상 시간이 필요하며 복잡한 훈련 파이프라인이 필요합니다. 그러나 RLHF는 의료 전문가가 AI 시스템이 통신하고 추천하는 방식을 구체화하는 데 도움이 되므로 의료 분야에서 특히 유용합니다. 예를 들어, 임상 의사는 피드백을 제공하여 모델이 적절한 병상 방식을 유지하고, 불확실성을 표현할 시기를 알고, 임상 지침을 준수하도록 할 수 있습니다. PPO와 같은 기법은 핵심 의료 지식을 보존하기 위해 파라미터 업데이트를 제한하면서 전문가 피드백을 기반으로 모델 동작을 반복적으로 최적화합니다. 이를 통해 모델은 환자에게 친숙한 언어로 복잡한 진단을 전달하는 동시에 즉각적인 의학적 치료를 위해 심각한 상태에 플래그를 지정할 수 있습니다. 이는 정확도와 커뮤니케이션 스타일이 모두 중요한 의료에 매우 중요합니다. RLHF에 대한 자세한 내용은 [인간 또는 AI 피드백에서 강화 학습을 통해 대규모 언어 모델 미세 조정\(AWS 블로그 게시물\)](#)을 참조하세요.

구현 방법

전체 파라미터 업데이트에는 훈련 중 모든 모델 파라미터 업데이트가 포함됩니다. 이 접근 방식은 환자 기록, 실험실 결과 및 진화하는 지침을 심층적으로 통합해야 하는 임상 결정 지원 시스템에 가장 적합합니다. 단점에는 데이터 세트가 크고 다양하지 않은 경우 높은 컴퓨팅 비용과 과적합 위험이 포함됩니다.

[파라미터 효율성 미세 조정\(PEFT\)](#) 메서드는 파라미터의 하위 집합만 업데이트하여 과적합 또는 치명적인 언어 기능 손실을 방지합니다. 유형에는 [저순위 적응\(LoRA\)](#), 어댑터 및 접두사 튜닝이 포함됩니다. PEFT 메서드는 더 낮은 계산 비용, 더 빠른 훈련을 제공하며, 임상 결정 지원 모델을 새 병원의 프로토콜 또는 용어에 적용하는 등의 실험에 적합합니다. 주요 제한 사항은 전체 파라미터 업데이트에 비해 성능이 저하될 수 있다는 것입니다.

미세 조정 방법에 대한 자세한 내용은 [Amazon SageMaker AI의 고급 미세 조정 방법](#)(AWS 블로그 게시물)을 참조하세요.

미세 조정 데이터 세트 구축

미세 조정 데이터 세트의 품질과 다양성은 모델 성능, 안전 및 편향 방지에 매우 중요합니다. 다음은 이 데이터 세트를 구축할 때 고려해야 할 세 가지 중요한 영역입니다.

- 미세 조정 접근 방식을 기반으로 한 볼륨
- 도메인 전문가의 데이터 주석
- 데이터 세트의 다양성

다음 표와 같이 미세 조정을 위한 데이터 세트 크기 요구 사항은 수행 중인 미세 조정 유형에 따라 달라집니다.

미세 조정 전략	데이터 세트 크기
도메인 조정 사전 훈련	100,000개 이상의 도메인 텍스트
감독 미세 조정	레이블이 지정된 페어 10,000개 이상
인적 피드백을 통한 강화 학습	1,000개 이상의 전문가 기본 설정 페어

[AWS Glue](#), [Amazon EMR](#) 및 [Amazon SageMaker Data Wrangler](#)를 사용하여 데이터 추출 및 변환 프로세스를 자동화하여 소유한 데이터 세트를 큐레이션할 수 있습니다. 충분히 큰 데이터 세트를 큐레이션할 수 없는 경우를 AWS 계정 통해에 직접 데이터 세트를 검색하고 다운로드할 수 있습니다 [AWS Data Exchange](#). 타사 데이터 세트를 활용하기 전에 법률 고문에게 문의하세요.

의료 의사, 생물학자, 화학자와 같은 도메인 지식이 있는 전문가 주석자는 의료 및 생물학적 데이터의 누앙스를 모델 출력에 통합하기 위한 데이터 큐레이션 프로세스의 일부여야 합니다. [Amazon SageMaker Ground Truth](#)는 전문가가 데이터 세트에 주석을 달 수 있는 로우 코드 사용자 인터페이스를 제공합니다.

인구를 나타내는 데이터 세트는 편향을 방지하고 실제 결과를 반영하기 위한 의료 및 생명 과학 미세 조정 사용 사례에 필수적입니다. [AWS Glue 대화형 세션](#) 또는 [Amazon SageMaker 노트북 인스턴스](#)는 Jupyter 호환 노트북을 사용하여 데이터 세트를 반복적으로 탐색하고 변환을 미세 조정할 수 있는 강력한 방법을 제공합니다. 대화형 세션을 사용하면 로컬 환경에서 널리 사용되는 IDEs(통합 개발 환경)를 선택할 수 있습니다. 또는 [AWS Glue](#) 또는 [Amazon SageMaker Studio](#) 노트북으로 작업할 수 있습니다 AWS Management Console.

모델 미세 조정

AWS 는 성공적인 미세 조정에 중요한 [Amazon SageMaker AI](#) 및 [Amazon Bedrock](#)과 같은 서비스를 제공합니다.

SageMaker AI는 개발자와 데이터 과학자가 ML 모델을 빠르게 구축, 훈련 및 배포할 수 있도록 지원하는 완전관리형 기계 학습 서비스입니다. 미세 조정을 위한 SageMaker AI의 세 가지 유용한 기능은 다음과 같습니다.

- [SageMaker 훈련](#) - 광범위한 모델을 대규모로 효율적으로 훈련하는 데 도움이 되는 완전 관리형 ML 기능
- [SageMaker JumpStart](#) - ML 작업을 위한 사전 훈련된 모델, 내장 알고리즘 및 솔루션 템플릿을 제공하기 위해 SageMaker 훈련 작업을 기반으로 구축된 기능
- [SageMaker HyperPod](#) - 파운데이션 모델 및 LLMs의 분산 훈련을 위해 특별히 구축된 인프라 솔루션

Amazon Bedrock은 기본 보안, 개인 정보 보호 및 확장성 기능을 갖춘 API를 통해 고성능 파운데이션 모델에 대한 액세스를 제공하는 완전 관리형 서비스입니다. 이 서비스는 사용 가능한 여러 기본 모델을 미세 조정할 수 있는 기능을 제공합니다. 자세한 내용은 Amazon Bedrock 설명서의 [미세 조정 및 지속적인 사전 훈련을 위해 지원되는 모델 및 리전](#)을 참조하세요.

두 서비스 중 하나를 사용하여 미세 조정 프로세스에 접근하는 경우 기본 모델, 미세 조정 전략 및 인프라를 고려하세요.

기본 모델 선택

Anthropic Claude, Meta Llama, Amazon Nova와 같은 폐쇄 소스 모델은 관리형 규정 준수로 강력한 out-of-the-box 성능을 제공하지만 Amazon Bedrock과 같은 관리형 APIs와 같은 공급자 지원 옵션으로 미세 조정 유연성을 제한합니다. 이는 특히 규제 대상 의료 사용 사례의 경우 사용자 지정 가능성을 제한합니다. 반대로 Meta Llama와 같은 오픈 소스 모델은 Amazon SageMaker AI 서비스 전반에서 완전한 제어와 유연성을 제공하므로 특정 데이터 또는 워크플로 요구 사항에 맞게 모델을 사용자 지정, 감사 또는 심층 조정해야 하는 경우에 이상적입니다.

미세 조정 전략

간단한 명령 튜닝은 Amazon Bedrock [모델 사용자 지정](#) 또는 Amazon SageMaker JumpStart에서 처리할 수 있습니다. LoRA 또는 어댑터와 같은 복잡한 PEFT 접근 방식에는 Amazon Bedrock의 SageMaker 훈련 작업 또는 사용자 지정 미세 조정 기능이 필요합니다. SageMaker HyperPod는 매우 큰 모델에 대한 분산 훈련을 지원합니다.

인프라 규모 및 제어

Amazon Bedrock과 같은 완전관리형 서비스는 인프라 관리를 최소화하며 사용 편의성과 규정 준수를 우선시하는 조직에 적합합니다. SageMaker JumpStart와 같은 반관리형 옵션은 복잡성을 줄이면서 약간의 유연성을 제공합니다. 이러한 옵션은 신속한 프로토타입 생성 또는 사전 구축된 워크플로 사용에 적합합니다. 전체 제어 및 사용자 지정은 SageMaker 훈련 작업 및 HyperPod와 함께 제공되지만, 이러한 작업에는 더 많은 전문 지식이 필요하며 대규모 데이터세트에 맞게 확장해야 하거나 사용자 지정 파이프라인이 필요한 경우에 가장 적합합니다.

미세 조정된 모델 모니터링

의료 및 생명 과학에서 LLM 미세 조정을 모니터링하려면 여러 주요 성능 지표를 추적해야 합니다. 정확도는 기준 측정값을 제공하지만, 특히 분류 오류로 인해 상당한 결과가 발생하는 애플리케이션에서는 정밀도 및 재현율과 균형을 맞춰야 합니다. F1-score 의료 데이터 세트에서 흔히 발생할 수 있는 클래스 불균형 문제를 해결하는 데 도움이 됩니다. 자세한 내용은 이 안내서의 [의료 및 생명과학 애플리케이션을 위한 LLMs 평가](#) 섹션을 참조하세요.

보정 지표는 모델의 신뢰도 수준이 실제 확률과 일치하는지 확인하는 데 도움이 됩니다. [공정성 지표](#)는 다양한 환자 인구통계에서 잠재적 편향을 감지하는 데 도움이 될 수 있습니다.

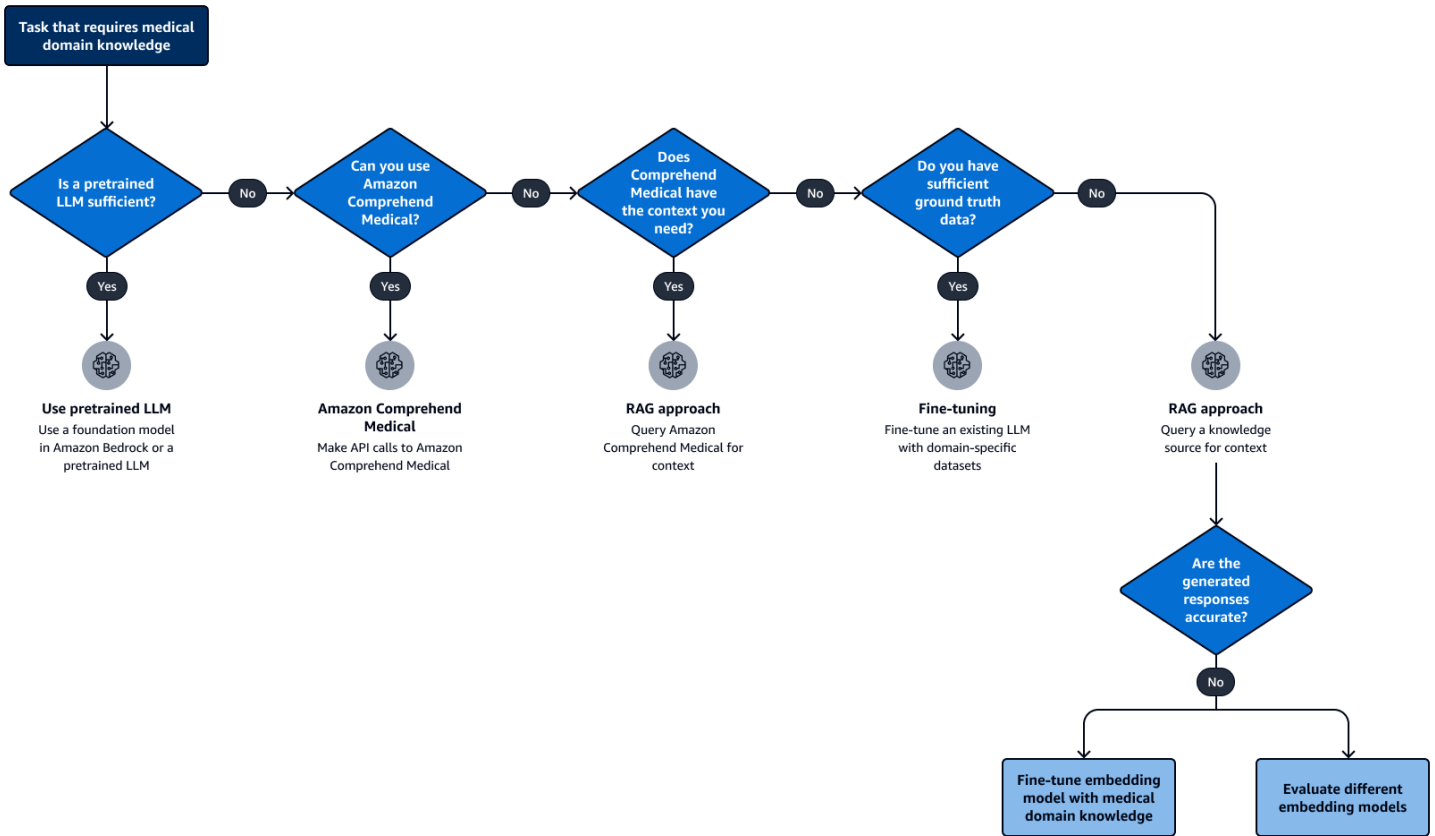
[MLflow](#)는 미세 조정 실험을 추적하는 데 도움이 되는 오픈 소스 솔루션입니다. MLflow는 Amazon SageMaker AI 내에서 기본적으로 지원되므로 훈련 실행의 지표를 시각적으로 비교할 수 있습니다. Amazon Bedrock의 미세 조정 작업의 경우 지표가 Amazon CloudWatch로 스트리밍되므로 CloudWatch 콘솔에서 지표를 시각화할 수 있습니다.

의료 및 생명 과학을 위한 NLP 접근 방식 선택

이 [의료 및 생명 과학을 위한 생성형 AI 및 NLP 접근 방식](#) 섹션에서는 의료 및 생명과학 애플리케이션의 자연어 처리(NLP) 작업을 해결하기 위한 다음 접근 방식을 설명합니다.

- Amazon Comprehend Medical 사용
- 검색 증강 생성(RAG) 워크플로에서 Amazon Comprehend Medical과 LLM 결합
- 미세 조정된 LLM 사용
- RAG 워크플로 사용

의료 도메인 작업에 대한 LLMs의 알려진 제한 사항과 사용 사례를 평가하여 작업에 가장 적합한 접근 방식을 선택할 수 있습니다. 다음 의사 결정 트리는 의료 NLP 작업에 대한 LLM 접근 방식을 선택하는데 도움이 될 수 있습니다.



이 다이어그램은 다음 워크플로를 보여줍니다.

1. 의료 및 생명과학 사용 사례의 경우 NLP 작업에 특정 도메인 지식이 필요한지 여부를 식별합니다. 필요에 따라 주제 전문가(SMEs).

2. 일반 LLM 또는 의료 데이터 세트에 대해 훈련된 모델을 사용할 수 있는 경우 Amazon Bedrock 또는 사전 훈련된 LLM에서 사용 가능한 파운데이션 모델을 사용합니다. 자세한 내용은 이 안내서의 [LLM 선택](#) 섹션을 참조하세요.
3. Amazon Comprehend Medical의 개체 감지 및 온톨로지 연결 기능이 사용 사례를 해결하는 경우 Amazon Comprehend Medical APIs 사용합니다. 자세한 내용은 이 안내서의 [Amazon Comprehend Medical 사용](#) 섹션을 참조하세요.
4. Amazon Comprehend Medical에는 필요한 컨텍스트가 있지만 사용 사례를 지원하지 않는 경우가 있습니다. 예를 들어 다른 엔터티 정의가 필요하거나, 많은 수의 결과를 받거나, 사용자 지정 엔터티가 필요하거나, 사용자 지정 NLP 작업이 필요할 수 있습니다. 이 경우 RAG 접근 방식을 사용하여 컨텍스트에 대해 Amazon Comprehend Medical을 쿼리합니다. 자세한 내용은 이 안내서의 [Amazon Comprehend Medical과 대규모 언어 모델 결합](#) 섹션을 참조하세요.
5. 충분한 양의 실측 데이터가 있는 경우 기존 LLM을 미세 조정합니다. 자세한 내용은 이 안내서의 [사용자 지정 접근 방식](#) 섹션을 참조하세요.
6. 다른 접근 방식이 NLP 작업 목표를 충족하지 못하는 경우 RAG 솔루션을 구현합니다. 자세한 내용은 이 안내서의 [사용자 지정 접근 방식](#) 섹션을 참조하세요.
7. RAG 솔루션을 구현한 후 생성된 응답이 정확한지 평가합니다. 자세한 내용은 이 안내서의 [의료 및 생명과학 애플리케이션을 위한 LLMs 평가](#) 섹션을 참조하세요. Amazon Titan Text Embeddings 모델 또는 [all-MiniLM-L6-v2](#)와 같은 일반 문장 변환기 모델로 시작하는 것이 일반적입니다. 그러나 도메인 컨텍스트 부족으로 인해 이러한 모델은 텍스트의 의학 용어를 캡처하지 못할 수 있습니다. 필요한 경우 다음 조정 사항을 고려하세요.
 - a. 다른 임베딩 모델 평가
 - b. 도메인별 데이터 세트를 사용하여 임베딩 모델 미세 조정

비즈니스 성숙도 고려 사항

의료 및 생명과학 애플리케이션에 LLM 솔루션을 적용할 때는 비즈니스 성숙도가 중요합니다. 이러한 조직은 수락 기준에 따라 LLMs 구현할 때 다양한 수준의 복잡성에 직면합니다. AI/ML 리소스가 부족한 조직은 LLM 솔루션을 구축하기 위해 계약업체 지원에 투자하는 경우가 많습니다. 이러한 상황에서는 다음과 같은 장단점을 이해하는 것이 중요합니다.

- 높은 비용 및 유지 관리를 위한 고성능 - 엄격한 성능 표준을 충족하기 위해 미세 조정되거나 사용자 지정 LLMs 포함된 복잡한 솔루션이 필요할 수 있습니다. 그러나 이로 인해 비용 및 유지 관리 요구 사항이 증가합니다. 이러한 정교한 솔루션을 유지하려면 전문 리소스를 고용하거나 계약업체와 협력해야 할 수 있습니다. 이렇게 하면 개발 속도가 느려질 수 있습니다.

- 저렴한 비용 및 유지 관리를 위한 우수한 성능 - 또는 Amazon Bedrock 또는 Amazon Comprehend Medical과 같은 서비스가 허용 가능한 성능을 제공한다는 것을 알 수 있습니다. 이러한 LLMs 또는 접근 방식은 완벽한 결과를 제공할 수 있지만 이러한 솔루션은 종종 일관된 고품질 결과를 제공할 수 있습니다. 이러한 솔루션은 비용이 절감되고 유지 관리 부담이 줄어듭니다. 이렇게 하면 개발이 가속화될 수 있습니다.

더 간단하고 저렴한 접근 방식이 수락 기준을 충족하는 고품질 결과를 지속적으로 제공하는 경우 성능 향상이 비용, 유지 관리 및 시간 절충의 가치가 있는지 고려하세요. 그러나 더 간단한 솔루션이 목표 성능에 크게 미치지 못하고 조직이 복잡한 솔루션 및 유지 관리 요구 사항에 대한 투자 용량이 부족한 경우 더 많은 리소스 또는 대체 솔루션을 사용할 수 있을 때까지 AI/ML 개발을 연기하는 것이 좋습니다.

또한 LLM을 사용하는 모든 의료 NLP 솔루션의 경우 지속적인 모니터링 및 평가를 수행하는 것이 좋습니다. 시간이 지남에 따라 사용자의 피드백을 평가하고 정기적인 평가를 구현하여 솔루션이 비즈니스 목표를 계속 충족하는지 확인합니다.

의료 및 생명과학 애플리케이션을 위한 LLMs 평가

이 섹션에서는 의료 및 생명과학 사용 사례에서 대규모 언어 모델(LLMs)을 평가하기 위한 요구 사항 및 고려 사항에 대한 포괄적인 개요를 제공합니다.

편향을 완화하고 LLM 생성 응답의 정확성을 검증하려면 실측 데이터와 SME 피드백을 사용하는 것이 중요합니다. 이 섹션에서는 훈련 및 테스트 데이터를 수집하고 큐레이션하는 모범 사례를 설명합니다. 또한 가드레일을 구현하고 데이터 편향과 공정성을 측정하는 데 도움이 됩니다. 또한 텍스트 분류, 명명된 엔터티 인식, 텍스트 생성과 같은 일반적인 의료 자연어 처리(NLP) 작업과 관련 평가 지표에 대해서도 설명합니다.

또한 훈련 실험 단계 및 프로덕션 후 단계에서 LLM 평가를 수행하기 위한 워크플로를 제공합니다. 모델 모니터링 및 LLM 작업은 이 평가 프로세스의 중요한 요소입니다.

의료 NLP 작업을 위한 훈련 및 테스트 데이터

Medical NLP 태스크는 일반적으로 의료 코포라(예: PubMed) 또는 환자 정보(예: 임상 환자 방문 노트)를 사용하여 인사이트를 분류, 요약 및 생성합니다. 의사, 의료 관리자 또는 기술자와 같은 의료 담당자는 전문 지식과 관점이 다양합니다. 이러한 의료진 간의 주관성으로 인해 훈련 및 테스트 데이터 세트가 작을수록 편향의 위험이 있습니다. 이러한 위험을 완화하려면 다음 모범 사례를 따르는 것이 좋습니다.

- 사전 훈련된 LLM 솔루션을 사용할 때는 적절한 양의 테스트 데이터가 있는지 확인합니다. 테스트 데이터는 실제 의료 데이터와 매우 유사해야 합니다. 작업에 따라 레코드의 범위는 20~100개입니다.
- LLM을 미세 조정할 때 대상 의료 도메인의 다양한 SMEs에서 충분한 수의 레이블이 지정된(실측) 레코드를 수집합니다. 일반적인 시작점은 100개 이상의 고품질 레코드입니다. 그러나 작업의 복잡성과 정확도 수락 기준을 고려할 때 더 많은 레코드가 필요할 수 있습니다.
- 의료 사용 사례에 필요한 경우 가드레일을 구현하고 데이터 편향과 공정성을 측정합니다. 예를 들어 LLM이 환자의 인종 프로파일로 인한 오진을 방지하는지 확인합니다. 자세한 내용은 이 설명서의 [보안 및 가드레일](#) 섹션을 참조하세요.

Anthropic과 같은 많은 AI 연구 및 개발 회사는 이미 유해성을 방지하기 위해 파운데이션 모델에 가드레일을 구현했습니다. 유해성 감지를 사용하여 입력 프롬프트와 LLMs. 자세한 내용은 Amazon Comprehend 설명서의 [유해성 감지](#) 및 Amazon Bedrock 설명서의 [가드레일을 참조하세요](#).

생성형 AI 작업에는 할루시네이션 위험이 있습니다. 분류와 같은 NLP 작업을 수행하여 이 위험을 완화할 수 있습니다. 텍스트 유사성 지표와 같은 고급 기술을 사용할 수도 있습니다. [BertScore](#)는 일반적

로 채택되는 텍스트 유사성 지표입니다. 할루시네이션을 완화하는 데 사용할 수 있는 기법에 대한 자세한 내용은 [대규모 언어 모델의 할루시네이션 완화 기법에 대한 포괄적인 설문 조사를 참조하세요](#).

의료 NLP 작업에 대한 지표

훈련 및 테스트를 위해 실측 데이터 및 SME 제공 레이블을 설정한 후 정량화 가능한 지표를 생성할 수 있습니다. 스트레스 테스트 및 LLM 결과 검토와 같은 정성적 프로세스를 통해 품질을 확인하는 것은 빠른 개발에 유용합니다. 그러나 지표는 향후 LLM 운영을 지원하는 정량적 벤치마크 역할을 하며 각 프로덕션 릴리스의 성능 벤치마크 역할을 합니다.

의료 작업을 이해하는 것이 중요합니다. 지표는 일반적으로 다음 일반 NLP 작업 중 하나에 매핑됩니다.

- 텍스트 분류 - LLM은 입력 프롬프트 및 제공된 컨텍스트에 따라 텍스트를 하나 이상의 사전 정의된 범주로 분류합니다. 예를 들어, 통증 척도를 사용하여 통증 범주를 분류합니다. 텍스트 분류 지표의 예는 다음과 같습니다.
 - [정확도](#)
 - [정밀도](#), 매크로 정밀도라고도 함
 - [매크로 재현율](#)이라고도 하는 재현율
 - 매크로 [F1 점수](#)라고도 하는 F1 점수
 - [해밍 손실](#)
- NER(명명된 엔터티 인식) - 텍스트 추출이라고도 하는 명명된 엔터티 인식은 비정형 텍스트에 언급된 명명된 엔터티를 미리 정의된 범주로 찾아 분류하는 프로세스입니다. 예를 들어 환자 레코드에서 약물 이름을 추출합니다. NER 지표의 예는 다음과 같습니다.
 - [정확도](#)
 - [정밀도](#)
 - [리콜](#)
 - [F1 점수](#)
 - [해밍 손실](#)
- 생성 - LLM은 프롬프트와 제공된 컨텍스트를 처리하여 새 텍스트를 생성합니다. 생성에는 요약 작업 또는 질문 응답 작업이 포함됩니다. 생성 지표의 예는 다음과 같습니다.
 - [Gisting Evaluation을 위한 Recall-Oriented Understud\(ROUGE\)](#)
 - [명시적 ORdering을 사용한 번역 평가 지표\(METEOR\)](#)
 - [연구 대상 이중 언어 평가\(BLEU\)\(번역용\)](#)

- 코사인 유사성이라고도 하는 [문자열 거리](#)

의료 및 생명과학 사용 사례에 대한 FAQ

다음은 의료 NLP 작업에 Amazon Comprehend Medical 또는 LLMs를 사용하는 것과 관련하여 자주 묻는 질문입니다.

Amazon Comprehend Medical과 LLM 중에서 선택하려면 어떻게 해야 하나요?

의료 텍스트 내에서 의료 엔터티를 감지하는 경우 [Amazon Comprehend Medical 설명서](#)를 검토하여 추출할 수 있는 의료 엔터티와 사용 사례를 해결하는 [온톨로지](#)가 있는지 확인합니다. 그렇지 않은 경우 LLM 사용을 고려하세요. 자세한 내용은 이 설명서의 [Amazon Comprehend Medical 사용 사례](#) 및 [LLM 사용 사례](#) 단원을 참조하세요.

Amazon Comprehend Medical 결과를 LLM에 제공하려면 어떻게 해야 하나요?

Amazon Comprehend Medical 결과를 LLM 프롬프트 내에 컨텍스트로 통합할 수 있습니다. 이를 통해 LLM에 대한 추가 의학적 지식과 용어가 제공됩니다. 제공된 컨텍스트는 개체 인식, 요약 또는 질문 응답과 같은 작업에 대한 LLM의 성능을 개선할 수 있습니다. 이 가이드에서는 Amazon Comprehend Medical 결과로 프롬프트를 구성하는 방법에 대한 몇 가지 예를 제공합니다. 자세한 내용은 이 안내서의 [Amazon Comprehend Medical과 대규모 언어 모델 결합](#) 섹션을 참조하세요.

Amazon Comprehend Medical LLMs과 함께 사용할 때의 모범 사례는 무엇입니까?

Amazon Comprehend Medical 신뢰도 점수를 사용하여 프롬프트 내의 개체를 필터링하거나 우선순위를 지정하는 것이 좋습니다. 또한 특정 데이터에 대한 성능을 평가하고 개체 정의가 요구 사항에 맞는지 확인하는 것이 중요합니다. Amazon Comprehend Medical을 도메인별 지식 소스와 결합하면 LLM의 성능이 더욱 향상될 수 있습니다. 자세한 내용은 이 안내서의 [RAG 워크플로에서 Amazon Comprehend Medical을 사용하는 모범 사례](#) 섹션을 참조하세요.

의료 사용 사례에 맞게 사전 훈련된 의료 LLM을 사용하거나 일반 LLM을 미세 조정해야 합니까?

결정은 특정 요구 사항과 고품질 훈련 데이터의 가용성에 따라 달라집니다. 사전 훈련된 의료 LLMs 좋은 출발점을 제공할 수 있습니다. 그러나 도메인별 데이터로 미세 조정해야 할 수도 있습니다. 레이블이 지정된 데이터가 충분한 경우 일반 LLM을 미세 조정하는 것이 실행 가능한 옵션일 수 있습니다. 자세한 내용은 이 설명서의 [LLM 선택](#) 및 [의료 및 생명 과학을 위한 NLP 접근 방식 선택](#) 단원을 참조하세요.

의료 NLP 작업에 대한 LLMs의 성능을 평가하려면 어떻게 해야 합니까?

텍스트 분류 및 명명된 개체 인식 작업에 정확도, 정밀도, 재현율 및 F1 점수와 같은 정량적 지표를 사용하는 것이 좋습니다. 텍스트 생성 작업에 ROUGE 및 METEOR를 사용할 수 있습니다. 주제 전문가가 신뢰할 수 있는 실측 데이터에 레이블을 지정하고 시간 경과에 따른 모델 성능 모니터링을 위한 프로세스를 구현하는 것이 중요합니다. 자세한 내용은 이 안내서의 [의료 및 생명과학 애플리케이션을 위한 LLMs 평가](#) 섹션을 참조하세요.

복잡성이 높은 LLM 솔루션과 복잡성이 낮은 LLM 솔루션의 장단점은 무엇인가요?

LLM을 미세 조정하거나 사용자 지정 LLM을 구축하는 것은 매우 복잡한 솔루션입니다. 이러한 접근 방식은 성능을 개선할 수 있지만 비용 및 유지 관리 요구 사항이 더 높습니다. 사전 훈련된 LLMs 또는 Amazon Comprehend Medical을 사용하는 것과 같은 더 간단한 솔루션은 더 낮은 비용과 더 빠른 개발 주기로 허용 가능한 성능을 제공할 수 있습니다. 그러나 이러한 접근 방식은 일부 사용 사례에 대한 엄격한 정확도 요구 사항을 충족하지 못할 수 있습니다. 자세한 내용은 이 안내서의 [비즈니스 성숙도 고려 사항](#) 섹션을 참조하세요.

다음 단계 및 리소스

이 가이드는 AWS 서비스를 사용하여 프로덕션 환경에서 실제 애플리케이션을 위한 의료 NLP 및 생성형 AI 작업을 자동화하는 데 도움이 됩니다. Amazon Comprehend Medical, Amazon Bedrock에서 지원되는 LLMs, 사전 훈련된 의료 LLMs 또는 미세 조정된 LLMs을 사용하여 의료 및 생명과학 비즈니스 목표를 달성하는 방법을 설명합니다. 이 가이드에서는 다음 접근 방식의 장점과 제한 사항에 대해 설명합니다.

- Amazon Comprehend Medical을 독립적으로 사용
- LLM에 Amazon Comprehend Medical 결과 제공
- 검색 증강 생성(RAG) 접근 방식에서 사전 훈련된 일반 LLM 또는 의료 LLM 사용
- 일반 LLM 또는 의료 LLM 미세 조정

이 가이드의 [의사 결정 트리](#)와 [비즈니스 성숙도 고려](#) 사항을 사용하여 조직의 AI/ML 성숙도 수준에 따라 이러한 접근 방식 중에서 선택합니다. Amazon Comprehend Medical 및 Amazon Bedrock LLMs 강력한 기능을 제공하지만 제대로 구현하고 평가하는 경우에만 성공합니다. 이 가이드에 설명된 [평가 정보](#) 및 [지표](#)를 사용하여 솔루션의 성능을 검증합니다.

다음 단계에서는 의료 IT 관리자, 아키텍트 및 기술 책임자가 AI/ML 실무자와 협력하여 NLP 의료 작업을 식별하는 것이 좋습니다. 이 가이드를 사용하여 개발 경로를 선택한 다음 적절한 AWS 서비스 및 기능을 사용하여 자동화된 솔루션을 성공적으로 구현합니다 AWS.

AWS 리소스

- Amazon Comprehend Medical 설명서:
 - [개발자 안내서](#)
 - [API Reference](#)
- [Amazon Bedrock 설명서](#)
 - [Amazon Bedrock 모델 평가](#)
 - [Amazon Bedrock의 미세 조정](#)
- [Amazon SageMaker AI에서 모델 미세 조정](#)
- [Amazon SageMaker Ground Truth](#)
- [Amazon Comprehend 유해성 감지](#)
- [AWS 의료 역량 파트너](#)

기타 리소스

- [Medical-LLM 리더보드 열기](#)
- [의료용 대규모 언어 모델 설문 조사: 데이터, 기술, 애플리케이션부터 책임 및 관리에 이르기까지](#)
- [대규모 언어 모델은 불량한 의료 코드 제공자 - 의료 코드 쿼리 벤치마킹](#)
- [초보자에서 전문가로: 의료 지식을 일반 LLMs으로 모델링](#)

기여자

작성

- Joe King, AWS 선임 데이터 과학자
- Ankith Ede, AWS 솔루션 아키텍트
- Clement Perrot, AWS 선임 생성형 AI 전략가
- Jillian Forde, AWS 선임 솔루션 아키텍트
- Rajesh Sitaraman, 수 AWS 석 배달 컨설턴트
- Ross Claytor, AWS 책임 응용 과학자
- Shivesh Ummat, AWS 솔루션 아키텍트

검토

- Dilshad Raihan Akkam Veettil, AWS 선임 데이터 과학자
- Joseph Cottingham, AWS 딥 러닝 아키텍트

기술 작성

- GxP AbouHarb, AWS 선임 기술 작성자

문서 기록

아래 표에 이 가이드의 주요 변경 사항이 설명되어 있습니다. 향후 업데이트에 대한 알림을 받으려면 [RSS 피드](#)를 구독하십시오.

변경 사항	설명	날짜
새 섹션	의료 분야의 대규모 언어 모델 미세 조정 섹션 과 프롬프트 엔지니어링 섹션을 추가했습니다.	2025년 12월 5일
최초 게시	—	2024년 12월 16일

AWS 권장 가이드 용어집

다음은 AWS 권장 가이드에서 제공하는 전략, 가이드 및 패턴에서 일반적으로 사용되는 용어입니다. 용어집 항목을 제안하려면 용어집 끝에 있는 피드백 제공 링크를 사용하십시오.

숫자

7가지 전략

애플리케이션을 클라우드로 이전하기 위한 7가지 일반적인 마이그레이션 전략 이러한 전략은 Gartner가 2011년에 파악한 5가지 전략을 기반으로 하며 다음으로 구성됩니다.

- 리팩터링/리아키텍트 - 클라우드 네이티브 기능을 최대한 활용하여 애플리케이션을 이동하고 해당 아키텍처를 수정함으로써 민첩성, 성능 및 확장성을 개선합니다. 여기에는 일반적으로 운영 체제와 데이터베이스 이식이 포함됩니다. 예: 온프레미스 Oracle 데이터베이스를 Amazon Aurora PostgreSQL 호환 에디션으로 마이그레이션합니다.
- 리플랫폼(리프트 앤드 리세이프) - 애플리케이션을 클라우드로 이동하고 일정 수준의 최적화를 도입하여 클라우드 기능을 활용합니다. 예: 온프레미스 Oracle 데이터베이스를 AWS 클라우드의 Amazon Relational Database Service(Amazon RDS) for Oracle로 마이그레이션합니다.
- 재구매(드롭 앤드 쇼프) - 일반적으로 기존 라이선스에서 SaaS 모델로 전환하여 다른 제품으로 전환합니다. 예: 고객 관계 관리(CRM) 시스템을 Salesforce.com으로 마이그레이션합니다.
- 리호스팅(리프트 앤드 시프트) - 애플리케이션을 변경하지 않고 클라우드로 이동하여 클라우드 기능을 활용합니다. 예: 온프레미스 Oracle 데이터베이스를 AWS 클라우드클라우드의 EC2 인스턴스에 있는 Oracle로 마이그레이션합니다.
- 재배포(하이퍼바이저 수준의 리프트 앤 시프트) - 새 하드웨어를 구매하거나, 애플리케이션을 다시 작성하거나, 기존 운영을 수정하지 않고도 인프라를 클라우드로 이동합니다. 온프레미스 플랫폼에서 동일한 플랫폼의 클라우드 서비스로 서버를 마이그레이션합니다. 예: Microsoft Hyper-V 애플리케이션을 로 마이그레이션합니다 AWS.
- 유지(보관) - 소스 환경에 애플리케이션을 유지합니다. 대규모 리팩터링이 필요하고 해당 작업을 나중에 연기하려는 애플리케이션과 비즈니스 차원에서 마이그레이션할 이유가 없어 유지하려는 레거시 애플리케이션이 여기에 포함될 수 있습니다.
- 사용 중지 - 소스 환경에서 더 이상 필요하지 않은 애플리케이션을 폐기하거나 제거합니다.

A

ABAC

[속성 기반 액세스 제어](#)를 참조하세요.

추상화된 서비스

[관리형 서비스](#)를 참조하세요.

ACID

[원자성, 일관성, 격리성, 내구성](#)을 참조하세요.

능동-능동 마이그레이션

양방향 복제 도구 또는 이중 쓰기 작업을 사용하여 소스 데이터베이스와 대상 데이터베이스가 동기화된 상태로 유지되고, 두 데이터베이스 모두 마이그레이션 중 연결 애플리케이션의 트랜잭션을 처리하는 데이터베이스 마이그레이션 방법입니다. 이 방법은 일회성 전환이 필요한 대신 소규모의 제어된 배치로 마이그레이션을 지원합니다. 더 유연하지만 [액티브 패시브 마이그레이션](#)보다 더 많은 작업이 필요합니다.

능동-수동 마이그레이션

소스 데이터베이스와 대상 데이터베이스가 동기화된 상태로 유지되지만 소스 데이터베이스만 연결 애플리케이션의 트랜잭션을 처리하고 데이터는 대상 데이터베이스로 복제되는 데이터베이스 마이그레이션 방법입니다. 대상 데이터베이스는 마이그레이션 중 어떤 트랜잭션도 허용하지 않습니다.

집계 함수

행 그룹에서 작동하고 그룹에 대한 단일 반환 값을 계산하는 SQL 함수입니다. 집계 함수의 예로 SUM 및 MAX가 있습니다.

AI

[인공 지능](#)을 참조하세요.

AI Ops

[인공 지능 운영](#)을 참조하세요.

익명화

데이터세트에서 개인 정보를 영구적으로 삭제하는 프로세스입니다. 익명화는 개인 정보 보호에 도움이 될 수 있습니다. 익명화된 데이터는 더 이상 개인 데이터로 간주되지 않습니다.

안티 패턴

솔루션이 다른 솔루션보다 비생산적이거나 비효율적이거나 덜 효과적이어서 반복되는 문제에 자주 사용되는 솔루션입니다.

애플리케이션 제어

맬웨어로부터 시스템을 보호하기 위해 승인된 애플리케이션만 사용하도록 허용하는 보안 접근 방식입니다.

애플리케이션 포트폴리오

애플리케이션 구축 및 유지 관리 비용과 애플리케이션의 비즈니스 가치를 비롯하여 조직에서 사용하는 각 애플리케이션에 대한 세부 정보 모음입니다. 이 정보는 [포트폴리오 탐색 및 분석 프로세스](#)의 핵심이며 마이그레이션, 현대화 및 최적화할 애플리케이션을 식별하고 우선순위를 정하는 데 도움이 됩니다.

인공 지능

컴퓨터 기술을 사용하여 학습, 문제 해결, 패턴 인식 등 일반적으로 인간과 관련된 인지 기능을 수행하는 것을 전문으로 하는 컴퓨터 과학 분야입니다. 자세한 내용은 [What is Artificial Intelligence?](#)를 참조하십시오.

인공 지능 운영(AIOps)

기계 학습 기법을 사용하여 운영 문제를 해결하고, 운영 인시던트 및 사용자 개입을 줄이고, 서비스 품질을 높이는 프로세스입니다. AWS 마이그레이션 전략에서 AIOps가 사용되는 방법에 대한 자세한 내용은 [운영 통합 가이드](#)를 참조하십시오.

비대칭 암호화

한 쌍의 키, 즉 암호화를 위한 퍼블릭 키와 복호화를 위한 프라이빗 키를 사용하는 암호화 알고리즘입니다. 퍼블릭 키는 복호화에 사용되지 않으므로 공유할 수 있지만 프라이빗 키에 대한 액세스는 엄격히 제한되어야 합니다.

원자성, 일관성, 격리성, 내구성(ACID)

오류, 정전 또는 기타 문제가 발생한 경우에도 데이터베이스의 데이터 유효성과 운영 신뢰성을 보장하는 소프트웨어 속성 세트입니다.

ABAC(속성 기반 액세스 제어)

부서, 직무, 팀 이름 등의 사용자 속성을 기반으로 세분화된 권한을 생성하는 방식입니다. 자세한 내용은 AWS Identity and Access Management (IAM) 설명서의 [용 ABAC AWS](#)를 참조하세요.

신뢰할 수 있는 데이터 소스

가장 신뢰할 수 있는 정보 소스로 간주되는 기본 버전의 데이터를 저장하는 위치입니다. 익명화, 편집 또는 가명화와 같은 데이터 처리 또는 수정의 목적으로 신뢰할 수 있는 데이터 소스의 데이터를 다른 위치로 복사할 수 있습니다.

가용 영역

다른 가용 영역의 장애로부터 격리 AWS 리전 되고 동일한 리전의 다른 가용 영역에 저림하고 지연 시간이 짧은 네트워크 연결을 제공하는 내의 고유한 위치입니다.

AWS 클라우드 채택 프레임워크(AWS CAF)

조직이 클라우드로 성공적으로 전환 AWS 하기 위한 효율적이고 효과적인 계획을 개발하는 데 도움이 되는 지침 및 모범 사례 프레임워크입니다. AWS CAF는 지침을 비즈니스, 사람, 거버넌스, 플랫폼, 보안 및 운영이라는 6가지 중점 영역으로 구성합니다. 비즈니스, 사람 및 거버넌스 관점은 비즈니스 기술과 프로세스에 초점을 맞추고, 플랫폼, 보안 및 운영 관점은 전문 기술과 프로세스에 중점을 둡니다. 예를 들어, 사람 관점은 인사(HR), 직원 배치 기능 및 인력 관리를 담당하는 이해관계자를 대상으로 합니다. 이러한 관점에서 AWS CAF는 성공적인 클라우드 채택을 위해 조직을 준비하는 데 도움이 되는 인력 개발, 교육 및 커뮤니케이션에 대한 지침을 제공합니다. 자세한 내용은 [AWS CAF 웹사이트](#)와 [AWS CAF 백서](#)를 참조하세요.

AWS 워크로드 검증 프레임워크(AWS WQF)

데이터베이스 마이그레이션 워크로드를 평가하고, 마이그레이션 전략을 권장하고, 작업 견적을 제공하는 도구입니다. AWS WQF는 AWS Schema Conversion Tool (AWS SCT)에 포함되어 있습니다. 데이터베이스 스키마 및 코드 객체, 애플리케이션 코드, 종속성 및 성능 특성을 분석하고 평가 보고서를 제공합니다.

B

악성 봇

개인 또는 조직을 방해하거나 해를 입히기 위한 [봇](#)입니다.

BCP

[비즈니스 연속성 계획](#)을 참조하세요.

동작 그래프

리소스 동작과 시간 경과에 따른 상호 작용에 대한 통합된 대화형 뷰입니다. Amazon Detective에서 동작 그래프를 사용하여 실패한 로그인 시도, 의심스러운 API 직접 호출 및 유사한 작업을 검사할 수 있습니다. 자세한 내용은 Detective 설명서의 [Data in a behavior graph](#)를 참조하십시오.

빅 엔디안 시스템

가장 중요한 바이트를 먼저 저장하는 시스템입니다. [엔디안](#)도 참조하세요.

바이너리 분류

바이너리 결과(가능한 두 클래스 중 하나)를 예측하는 프로세스입니다. 예를 들어, ML 모델이 “이 이메일이 스팸인가요, 스팸이 아닌가요?”, ‘이 제품은 책임가요, 자동차인가요?’ 등의 문제를 예측해야 할 수 있습니다.

블룸 필터

요소가 세트의 멤버인지 여부를 테스트하는 데 사용되는 메모리 효율성이 높은 확률론적 데이터 구조입니다.

블루/그린(Blue/Green) 배포

동일하지만 별개의 두 환경을 생성하는 배포 전략입니다. 하나의 환경(파란색)에서 현재 애플리케이션 버전을 실행하고 새 애플리케이션 버전은 다른 환경(녹색)에서 실행합니다. 이 전략을 사용하면 영향을 최소화하면서 신속하게 롤백할 수 있습니다.

bot

인터넷을 통해 자동화된 태스크를 실행하고 인적 활동이나 상호 작용을 시뮬레이션하는 소프트웨어 애플리케이션입니다. 인터넷에서 정보를 인덱싱하는 웹 크롤러와 같이 유용하거나 이로운 봇도 있습니다. 악성 봇이라고 하는 다른 일부 봇은 개인 또는 조직을 방해하거나 해를 입히기 위한 봇입니다.

봇넷

[맬웨어](#)에 감염되고 봇 허더 또는 봇 운영자와 같은 단일 당사자가 제어하는 [봇](#) 네트워크입니다. 봇넷은 봇의 규모와 봇의 영향 범위를 확대하는 가장 잘 알려진 메커니즘입니다.

브랜치

코드 리포지토리의 포함된 영역입니다. 리포지토리에 생성되는 첫 번째 브랜치가 기본 브랜치입니다. 기존 브랜치에서 새 브랜치를 생성한 다음 새 브랜치에서 기능을 개발하거나 버그를 수정할 수 있습니다. 기능을 구축하기 위해 생성하는 브랜치를 일반적으로 기능 브랜치라고 합니다. 기능을 출시할 준비가 되면 기능 브랜치를 기본 브랜치에 다시 병합합니다. 자세한 내용은 [About branches](#)(GitHub 설명서)를 참조하십시오.

긴급 액세스 권한

예외적인 상황에서 승인된 프로세스를 통해 사용자가 일반적으로 액세스할 권한이 없는데 액세스할 수 있는 빠른 방법입니다. 자세한 내용은 AWS Well-Architected 지침의 [Implement break-glass procedures](#) 지표를 참조하세요.

브라운필드 전략

사용자 환경의 기존 인프라 시스템 아키텍처에 브라운필드 전략을 채택할 때는 현재 시스템 및 인프라의 제약 조건을 중심으로 아키텍처를 설계합니다. 기존 인프라를 확장하는 경우 브라운필드 전략과 [그린필드](#) 전략을 혼합할 수 있습니다.

버퍼 캐시

가장 자주 액세스하는 데이터가 저장되는 메모리 영역입니다.

사업 역량

기업이 가치를 창출하기 위해 하는 일(예: 영업, 고객 서비스 또는 마케팅)입니다. 마이크로서비스 아키텍처 및 개발 결정은 비즈니스 역량에 따라 이루어질 수 있습니다. 자세한 내용은 백서의 [AWS에서 컨테이너화된 마이크로서비스 실행의 비즈니스 역량 중심의 구성화](#) 섹션을 참조하십시오.

비즈니스 연속성 계획(BCP)

대규모 마이그레이션과 같은 중단 이벤트가 운영에 미치는 잠재적 영향을 해결하고 비즈니스가 신속하게 운영을 재개할 수 있도록 지원하는 계획입니다.

C

CAF

[AWS Cloud Adoption Framework](#)를 참조하세요.

카나리 배포

최종 사용자에게 제공하는 느린 증분 릴리스 버전입니다. 확신이 들면 새 버전을 배포하고 현재 버전을 완전히 교체합니다.

CCoE

[클라우드 혁신 센터](#)를 참조하세요.

CDC

[데이터 캡처 변경](#)을 참조하세요.

변경 데이터 캡처(CDC)

데이터베이스 테이블과 같은 데이터 소스의 변경 내용을 추적하고 변경 사항에 대한 메타데이터를 기록하는 프로세스입니다. 대상 시스템의 변경 내용을 감사하거나 복제하여 동기화를 유지하는 등의 다양한 용도로 CDC를 사용할 수 있습니다.

카오스 엔지니어링

시스템의 복원력을 테스트하기 위해 의도적으로 장애나 중단 이벤트를 도입합니다. [AWS Fault Injection Service \(AWS FIS\)](#)를 사용하여 AWS 워크로드에 스트레스를 주고 응답을 평가하는 실험을 수행할 수 있습니다.

CI/CD

[지속적 통합 및 지속적 전송](#)을 참조하세요.

분류

예측을 생성하는 데 도움이 되는 분류 프로세스입니다. 분류 문제에 대한 ML 모델은 이산 값을 예측합니다. 이산 값은 항상 서로 다릅니다. 예를 들어, 모델이 이미지에 자동차가 있는지 여부를 평가해야 할 수 있습니다.

클라이언트측 암호화

대상이 데이터를 AWS 서비스 수신하기 전에 로컬에서 데이터를 암호화합니다.

클라우드 혁신 센터(CCoE)

클라우드 모범 사례 개발, 리소스 동원, 마이그레이션 타임라인 설정, 대규모 혁신을 통한 조직 선도 등 조직 전체에서 클라우드 채택 노력을 추진하는 다분야 팀입니다. 자세한 내용은 AWS 클라우드 엔터프라이즈 전략 블로그의 [CCoE 게시물](#)을 참조하세요.

클라우드 컴퓨팅

원격 데이터 스토리지와 IoT 디바이스 관리에 일반적으로 사용되는 클라우드 기술 클라우드 컴퓨팅은 일반적으로 [엣지 컴퓨팅](#) 기술에 연결되어 있습니다.

클라우드 운영 모델

IT 조직에서 하나 이상의 클라우드 환경을 구축, 성숙화 및 최적화하는 데 사용되는 운영 모델입니다. 자세한 내용은 [클라우드 운영 모델 구축](#)을 참조하십시오.

클라우드 채택 단계

조직이 AWS 클라우드로 마이그레이션할 때 일반적으로 거치는 4단계는 다음과 같습니다.

- 프로젝트 - 개념 증명 및 학습 목적으로 몇 가지 클라우드 관련 프로젝트 실행
- 기반 - 클라우드 채택 확장을 위한 기초 투자(예: 랜딩 존 생성, CCoE 정의, 운영 모델 구축)
- 마이그레이션 - 개별 애플리케이션 마이그레이션
- Re-invention - 제품 및 서비스 최적화와 클라우드 혁신

이러한 단계는 Stephen Orban이 블로그 게시물 [The Journey Toward Cloud-First and the Stages of Adoption](#) on the AWS 클라우드 Enterprise Strategy 블로그에서 정의했습니다. AWS 마이그레이션 전략과 어떤 관련이 있는지에 대한 자세한 내용은 [마이그레이션 준비 가이드](#)를 참조하세요.

CMDB

[구성 관리 데이터베이스](#)를 참조하세요.

코드 리포지토리

소스 코드와 설명서, 샘플, 스크립트 등의 기타 자산이 버전 관리 프로세스를 통해 저장되고 업데이트되는 위치입니다. 일반적인 클라우드 리포지토리로 GitHub 또는 Bitbucket Cloud가 포함됩니다. 코드의 각 버전을 브랜치라고 합니다. 마이크로서비스 구조에서 각 리포지토리는 단일 기능 전용입니다. 단일 CI/CD 파이프라인은 여러 리포지토리를 사용할 수 있습니다.

콜드 캐시

비어 있거나, 제대로 채워지지 않았거나, 오래되었거나 관련 없는 데이터를 포함하는 버퍼 캐시입니다. 주 메모리나 디스크에서 데이터베이스 인스턴스를 읽어야 하기 때문에 성능에 영향을 미치며, 이는 버퍼 캐시에서 읽는 것보다 느립니다.

콜드 데이터

거의 액세스되지 않고 일반적으로 과거 데이터인 데이터. 이런 종류의 데이터를 쿼리할 때는 일반적으로 느린 쿼리가 허용됩니다. 이 데이터를 성능이 낮고 비용이 저렴한 스토리지 계층 또는 클래스로 옮기면 비용을 절감할 수 있습니다.

컴퓨터 비전(CV)

기계 학습을 사용하여 디지털 이미지 및 비디오와 같은 시각적 형식에서 정보를 분석하고 추출하는 [AI](#) 필드입니다. 예를 들어 Amazon SageMaker AI는 CV에 대한 이미지 처리 알고리즘을 제공합니다.

구성 드리프트

워크로드의 경우 구성이 예상되는 상태에서 변경됩니다. 이로 인해 워크로드가 규정을 준수하지 않을 수 있으며, 이는 일반적으로 점진적이고 의도되지 않은 작업입니다.

구성 관리 데이터베이스(CMDB)

하드웨어 및 소프트웨어 구성 요소와 해당 구성을 포함하여 데이터베이스와 해당 IT 환경에 대한 정보를 저장하고 관리하는 리포지토리입니다. 일반적으로 마이그레이션의 포트폴리오 탐색 및 분석 단계에서 CMDB의 데이터를 사용합니다.

규정 준수 팩

규정 준수 및 보안 검사를 사용자 지정하기 위해 조합할 수 있는 AWS Config 규칙 및 수정 작업 모음입니다. YAML 템플릿을 사용하여 적합성 팩을 AWS 계정 및 리전 또는 조직 전체에 단일 엔터티로 배포할 수 있습니다. 자세한 내용은 AWS Config 설명서의 [적합성 팩](#)을 참조하세요.

지속적 통합 및 지속적 전달(CI/CD)

소프트웨어 릴리스 프로세스의 소스, 빌드, 테스트, 스테이징 및 프로덕션 단계를 자동화하는 프로세스입니다. CI/CD는 일반적으로 파이프라인으로 설명됩니다. CI/CD를 통해 프로세스를 자동화하고, 생산성을 높이고, 코드 품질을 개선하고, 더 빠르게 제공할 수 있습니다. 자세한 내용은 [지속적 전달의 이점](#)을 참조하십시오. CD는 지속적 배포를 의미하기도 합니다. 자세한 내용은 [지속적 전달\(Continuous Delivery\)](#)과 [지속적인 개발](#)을 참조하십시오.

CV

[컴퓨터 비전](#)을 참조하세요.

D

저장 데이터

스토리지에 있는 데이터와 같이 네트워크에 고정되어 있는 데이터입니다.

데이터 분류

중요도와 민감도를 기준으로 네트워크의 데이터를 식별하고 분류하는 프로세스입니다. 이 프로세스는 데이터에 대한 적절한 보호 및 보존 제어를 결정하는 데 도움이 되므로 사이버 보안 위험 관리 전략의 중요한 구성 요소입니다. 데이터 분류는 AWS Well-Architected Framework의 보안 원칙 구성 요소입니다. 자세한 내용은 [데이터 분류](#)를 참조하십시오.

데이터 드리프트

프로덕션 데이터와 ML 모델 학습에 사용된 데이터 간의 상당한 차이 또는 시간 경과에 따른 입력 데이터의 의미 있는 변화. 데이터 드리프트는 ML 모델 예측의 전반적인 품질, 정확성 및 공정성을 저하시킬 수 있습니다.

전송 중 데이터

네트워크를 통과하고 있는 데이터입니다. 네트워크 리소스 사이를 이동 중인 데이터를 예로 들 수 있습니다.

데이터 메시

중앙 집중식 관리 및 거버넌스를 통해 분산되고 탈중앙화된 데이터 소유권을 제공하는 아키텍처 프레임워크입니다.

데이터 최소화

꼭 필요한 데이터만 수집하고 처리하는 원칙입니다. 에서 데이터를 최소화하면 개인 정보 보호 위험, 비용 및 분석 탄소 발자국을 줄일 AWS 클라우드 수 있습니다.

데이터 경계

신뢰할 수 있는 자격 증명만 예상 네트워크에서 신뢰할 수 있는 리소스에 액세스하도록 하는 데 도움이 되는 AWS 환경의 예방 가드레일 세트입니다. 자세한 내용은 [데이터 경계 구축을 참조하세요 AWS](#).

데이터 사전 처리

원시 데이터를 ML 모델이 쉽게 구문 분석할 수 있는 형식으로 변환하는 것입니다. 데이터를 사전 처리한다는 것은 특정 열이나 행을 제거하고 누락된 값, 일관성이 없는 값 또는 중복 값을 처리함을 의미할 수 있습니다.

데이터 출처

라이프사이클 전반에 걸쳐 데이터의 출처와 기록을 추적하는 프로세스(예: 데이터 생성, 전송, 저장 방법).

데이터 주체

데이터를 수집 및 처리하는 개인입니다.

데이터 웨어하우스

분석과 같은 비즈니스 인텔리전스를 지원하는 데이터 관리 시스템입니다. 데이터 웨어하우스에는 보통 많은 양의 기록 데이터가 포함되며 일반적으로 쿼리 및 분석에 사용됩니다.

데이터 정의 언어(DDL)

데이터베이스에서 테이블 및 객체의 구조를 만들거나 수정하기 위한 명령문 또는 명령입니다.

데이터베이스 조작 언어(DML)

데이터베이스에서 정보를 수정(삽입, 업데이트 및 삭제)하기 위한 명령문 또는 명령입니다.

DDL

[데이터 정의 언어](#)를 참조하세요.

딥 앙상블

예측을 위해 여러 딥 러닝 모델을 결합하는 것입니다. 딥 앙상블을 사용하여 더 정확한 예측을 얻거나 예측의 불확실성을 추정할 수 있습니다.

딥 러닝

여러 계층의 인공 신경망을 사용하여 입력 데이터와 관심 대상 변수 간의 매핑을 식별하는 ML 하위 분야입니다.

심층 방어

네트워크와 그 안의 데이터 기밀성, 무결성 및 가용성을 보호하기 위해 컴퓨터 네트워크 전체에 일련의 보안 메커니즘과 제어를 신중하게 계층화하는 정보 보안 접근 방식입니다. 이 전략을 채택하면 AWS Organizations 구조의 여러 계층에 여러 제어를 AWS 추가하여 리소스를 보호할 수 있습니다. 예를 들어, 심층 방어 접근 방식은 다단계 인증, 네트워크 세분화 및 암호화를 결합할 수 있습니다.

위임된 관리자

에서 AWS Organizations 호환되는 서비스는 AWS 멤버 계정을 등록하여 조직의 계정을 관리하고 해당 서비스에 대한 권한을 관리할 수 있습니다. 이러한 계정을 해당 서비스의 위임된 관리자라고 합니다. 자세한 내용과 호환되는 서비스 목록은 AWS Organizations 설명서의 [AWS Organizations 와 함께 사용할 수 있는 AWS 서비스](#)를 참조하십시오.

배포

대상 환경에서 애플리케이션, 새 기능 또는 코드 수정 사항을 사용할 수 있도록 하는 프로세스입니다. 배포에는 코드 베이스의 변경 사항을 구현한 다음 애플리케이션 환경에서 해당 코드베이스를 구축하고 실행하는 작업이 포함됩니다.

개발 환경

[환경](#)을 참조하세요.

탐지 제어

이벤트 발생 후 탐지, 기록 및 알림을 수행하도록 설계된 보안 제어입니다. 이러한 제어는 기존의 예방적 제어를 우회한 보안 이벤트를 알리는 2차 방어선입니다. 자세한 내용은 AWS에서 보안 제어 구현의 [탐지 제어](#)를 참조하세요.

개발 가치 흐름 매핑 (DVSM)

소프트웨어 개발 라이프사이클에서 속도와 품질에 부정적인 영향을 미치는 제약 조건을 식별하고 우선 순위를 지정하는 데 사용되는 프로세스입니다. DVSM은 원래 린 제조 방식을 위해 설계된 가치 흐름 매핑 프로세스를 확장합니다. 소프트웨어 개발 프로세스를 통해 가치를 창출하고 이동하는 데 필요한 단계와 팀에 중점을 둡니다.

디지털 트윈

건물, 공장, 산업 장비 또는 생산 라인과 같은 실제 시스템을 가상으로 표현한 것입니다. 디지털 트윈은 예측 유지 보수, 원격 모니터링, 생산 최적화를 지원합니다.

차원 테이블

[스타 스키마](#)에서 팩트 테이블의 정량적 데이터에 대한 데이터 속성을 포함하는 더 작은 테이블을 말합니다. 차원 테이블 속성은 일반적으로 텍스트 필드나 텍스트처럼 동작하는 개별 숫자입니다. 이러한 속성은 보통 쿼리 제약, 필터링 및 결과 세트 레이블 지정에 사용됩니다.

재해

워크로드 또는 시스템이 기본 배포 위치에서 비즈니스 목표를 달성하지 못하게 방해하는 이벤트입니다. 이러한 이벤트는 자연재해, 기술적 오류, 의도하지 않은 구성 오류 또는 멀웨어 공격과 같은 사람의 행동으로 인한 결과일 수 있습니다.

재해 복구(DR)

[재해](#)로 인한 가동 중지 시간 및 데이터 손실을 최소화하기 위해 사용하는 전략 및 프로세스입니다. 자세한 내용은 AWS Well-Architected Framework의 [Disaster Recovery of Workloads on AWS: Recovery in the Cloud](#)를 참조하세요.

DML

[데이터베이스 조작 언어](#)를 참조하세요.

도메인 기반 설계

구성 요소를 각 구성 요소가 제공하는 진화하는 도메인 또는 핵심 비즈니스 목표에 연결하여 복잡한 소프트웨어 시스템을 개발하는 접근 방식입니다. 이 개념은 에릭 에반스에 의해 그의 저서인 도메인 기반 디자인: 소프트웨어 중심의 복잡성 해결(Boston: Addison-Wesley Professional, 2003)에서 소개되었습니다. Strangler Fig 패턴과 함께 도메인 기반 설계를 사용하는 방법에 대한 자세한 내용은 [컨테이너 및 Amazon API Gateway를 사용하여 기존의 Microsoft ASP.NET\(ASMX\) 웹 서비스를 점진적으로 현대화하는 방법](#)을 참조하십시오.

DR

[재해 복구](#)를 참조하세요.

드리프트 감지

기준이 되는 구성과의 편차 추적을 말합니다. 예를 들어 AWS CloudFormation 를 사용하여 [시스템 리소스의 드리프트를 감지](#)하거나 사용하여 AWS Control Tower 거버넌스 요구 사항 준수에 영향을 미칠 수 있는 [랜딩 존의 변경 사항을 감지](#)할 수 있습니다.

DVSM

[개발 가치 흐름 매핑](#)을 참조하세요.

E

EDA

[탐색 데이터 분석](#)을 참조하세요.

EDI

[전자 데이터 교환](#)을 참조하세요.

엣지 컴퓨팅

IoT 네트워크의 엣지에서 스마트 디바이스의 컴퓨팅 성능을 개선하는 기술 엣지 컴퓨팅은 [클라우드 컴퓨팅](#)에 비해 보다 통신 지연 시간을 줄이고 응답 시간을 개선할 수 있습니다.

전자 데이터 교환(EDI)

조직 간 비즈니스 문서의 자동화된 교환을 나타냅니다. 자세한 내용은 [전자 데이터 교환\(EDI\)이란 무엇인가요?](#)를 참조하세요.

암호화

사람이 읽을 수 있는 일반 텍스트 데이터를 사이버텍스트로 변환하는 컴퓨팅 프로세스입니다.

암호화 키

암호화 알고리즘에 의해 생성되는 무작위 비트의 암호화 문자열입니다. 키의 길이는 다양할 수 있으며 각 키는 예측할 수 없고 고유하게 설계되었습니다.

엔디안

컴퓨터 메모리에 바이트가 저장되는 순서입니다. 빅 엔디안 시스템은 가장 중요한 바이트를 먼저 저장합니다. 리틀 엔디안 시스템은 가장 덜 중요한 바이트를 먼저 저장합니다.

엔드포인트

[서비스 엔드포인트](#)를 참조하세요.

엔드포인트 서비스

Virtual Private Cloud(VPC)에서 호스팅하여 다른 사용자와 공유할 수 있는 서비스입니다. 를 사용하여 엔드포인트 서비스를 생성하고 다른 AWS 계정 또는 AWS Identity and Access Management (IAM) 보안 주체에 권한을 AWS PrivateLink 부여할 수 있습니다. 이러한 계정 또는 보안 주체는 인터페이스 VPC 엔드포인트를 생성하여 엔드포인트 서비스에 비공개로 연결할 수 있습니다. 자세한 내용은 Amazon Virtual Private Cloud(VPC) 설명서의 [엔드포인트 서비스 생성](#)을 참조하십시오.

엔터프라이즈 리소스 계획(ERP)

엔터프라이즈의 주요 비즈니스 프로세스(예: 회계, [MES](#), 프로젝트 관리)를 자동화하고 관리하는 시스템입니다.

봉투 암호화

암호화 키를 다른 암호화 키로 암호화하는 프로세스입니다. 자세한 내용은 AWS Key Management Service (AWS KMS) 설명서의 [봉투 암호화](#)를 참조하세요.

환경

실행 중인 애플리케이션의 인스턴스입니다. 다음은 클라우드 컴퓨팅의 일반적인 환경 유형입니다.

- 개발 환경 - 애플리케이션 유지 관리를 담당하는 핵심 팀만 사용할 수 있는 실행 중인 애플리케이션의 인스턴스입니다. 개발 환경은 변경 사항을 상위 환경으로 승격하기 전에 테스트하는 데 사용됩니다. 이러한 유형의 환경을 테스트 환경이라고도 합니다.
- 하위 환경 - 초기 빌드 및 테스트에 사용되는 환경을 비롯한 애플리케이션의 모든 개발 환경입니다.
- 프로덕션 환경 - 최종 사용자가 액세스할 수 있는 실행 중인 애플리케이션의 인스턴스입니다. CI/CD 파이프라인에서 프로덕션 환경이 마지막 배포 환경입니다.
- 상위 환경 - 핵심 개발 팀 이외의 사용자가 액세스할 수 있는 모든 환경입니다. 프로덕션 환경, 프로덕션 이전 환경 및 사용자 수용 테스트를 위한 환경이 여기에 포함될 수 있습니다.

에픽

애자일 방법론에서 작업을 구성하고 우선순위를 정하는 데 도움이 되는 기능적 범주입니다. 에픽은 요구 사항 및 구현 작업에 대한 개괄적인 설명을 제공합니다. 예를 들어, AWS CAF 보안 에픽에는 ID 및 액세스 관리, 탐지 제어, 인프라 보안, 데이터 보호 및 인시던트 대응이 포함됩니다. AWS 마 이그레이션 전략의 에픽에 대한 자세한 내용은 [프로그램 구현 가이드](#)를 참조하십시오.

ERP

[엔터프라이즈 리소스 계획](#)을 참조하세요.

탐색 데이터 분석(EDA)

데이터 세트를 분석하여 주요 특성을 파악하는 프로세스입니다. 데이터를 수집 또는 집계한 다음 초기 조사를 수행하여 패턴을 찾고, 이상을 탐지하고, 가정을 확인합니다. EDA는 요약 통계를 계산하고 데이터 시각화를 생성하여 수행됩니다.

F

팩트 테이블

[스타 스키마](#)의 중앙 테이블입니다. 비즈니스 운영에 대한 정량적 데이터를 저장합니다. 일반적으로 팩트 테이블은 측정값이 있는 열 및 차원 테이블에 대한 외래 키가 있는 열과 같이 두 가지 열 유형을 포함합니다.

빠른 실패

개발 수명 주기를 줄이기 위해 빈번한 증분 테스트를 사용하는 철학입니다. 애자일 접근 방식의 핵심입니다.

장애 격리 경계

에서 장애의 영향을 제한하고 워크로드의 복원력을 개선하는 데 도움이 되는 가용 영역, AWS 리전 컨트롤 플레인 또는 데이터 플레인과 같은 AWS 클라우드경계입니다. 자세한 내용은 [AWS 장애 격리 경계](#)를 참조하세요.

기능 브랜치

[브랜치](#)를 참조하세요.

기능

예측에 사용하는 입력 데이터입니다. 예를 들어, 제조 환경에서 기능은 제조 라인에서 주기적으로 캡처되는 이미지일 수 있습니다.

기능 중요도

모델의 예측에 특성이 얼마나 중요한지를 나타냅니다. 이는 일반적으로 SHAP(Shapley Additive Descriptions) 및 통합 그래디언트와 같은 다양한 기법을 통해 계산할 수 있는 수치 점수로 표현됩니다. 자세한 내용은 [기계 학습 모델 해석 가능성을 참조하세요 AWS](#).

기능 변환

추가 소스로 데이터를 보강하거나, 값을 조정하거나, 단일 데이터 필드에서 여러 정보 세트를 추출하는 등 ML 프로세스를 위해 데이터를 최적화하는 것입니다. 이를 통해 ML 모델이 데이터를 활용

할 수 있습니다. 예를 들어, 날짜 '2021-05-27 00:15:37'을 '2021년', '5월', '목', '15일'로 분류하면 학습 알고리즘이 다양한 데이터 구성 요소와 관련된 미묘한 패턴을 학습하는 데 도움이 됩니다.

퓨샷 프롬프팅

유사한 태스크를 수행하도록 요청하기 전에 [LLM](#)에 태스크와 원하는 출력을 보여주는 몇 가지 예제를 제공합니다. 이 기법은 모델이 프롬프트에 포함된 예제(샷)에서 학습하는 컨텍스트 내 학습을 적용합니다. 퓨샷 프롬프팅은 특정 형식 지정, 추론 또는 분야별 지식이 필요한 태스크에 효과적일 수 있습니다. [제로샷 프롬프팅](#)도 참조하세요.

FGAC

[세분화된 액세스 제어](#)를 참조하세요.

세분화된 액세스 제어(FGAC)

여러 조건을 사용하여 액세스 요청을 허용하거나 거부합니다.

플래시컷 마이그레이션

단계적 접근 방식을 사용하는 대신 [변경 데이터 캡처](#)를 통해 지속적 데이터 복제를 사용하여 최단 시간에 데이터를 마이그레이션하는 데이터베이스 마이그레이션 방법입니다. 목표는 가동 중지 시간을 최소화하는 것입니다.

FM

[파운데이션 모델](#)을 참조하세요.

파운데이션 모델(FM)

일반화되고 레이블이 지정되지 않은 데이터의 대규모 데이터세트에서 훈련된 대규모 딥 러닝 신경망입니다. FM은 언어 이해, 텍스트 및 이미지 생성, 자연어 대화와 같은 다양한 일반 태스크를 수행할 수 있습니다. 자세한 내용은 [파운데이션 모델이란?](#)을 참조하세요.

G

생성형 AI

대량의 데이터에서 훈련되었으며 간단한 텍스트 프롬프트를 사용하여 이미지, 비디오, 텍스트, 오디오와 같은 새 콘텐츠와 아티팩트를 생성할 수 있는 [AI](#) 모델의 하위 세트입니다. 자세한 내용은 [생성형 AI란 무엇인가요?](#)를 참조하세요.

지리적 차단

[지리적 제한](#)을 참조하세요.

지리적 제한(지리적 차단)

Amazon CloudFront에서 특정 국가의 사용자가 콘텐츠 배포에 액세스하지 못하도록 하는 옵션입니다. 허용 목록 또는 차단 목록을 사용하여 승인된 국가와 차단된 국가를 지정할 수 있습니다. 자세한 내용은 CloudFront 설명서의 [콘텐츠의 지리적 배포 제한](#)을 참조하십시오.

Gitflow 워크플로

하위 환경과 상위 환경이 소스 코드 리포지토리의 서로 다른 브랜치를 사용하는 방식입니다. Gitflow 워크플로는 레거시로 간주되며 [트렁크 기반 워크플로](#)는 선호되는 현대적 접근 방식입니다.

골든 이미지

시스템 또는 소프트웨어의 새 인스턴스를 배포하기 위한 템플릿으로 사용되는 해당 시스템 또는 소프트웨어의 스냅샷입니다. 예를 들어 제조 분야에서는 골든 이미지를 사용하여 여러 디바이스에서 소프트웨어를 프로비저닝할 수 있으며 이를 통해 딥이스 제조 작업의 속도, 확장성 및 생산성을 개선할 수 있습니다.

브라운필드 전략

새로운 환경에서 기존 인프라의 부재 시스템 아키텍처에 대한 그린필드 전략을 채택할 때 [브라운필드](#)라고도 하는 기존 인프라와의 호환성 제한 없이 모든 새로운 기술을 선택할 수 있습니다. 기존 인프라를 확장하는 경우 브라운필드 전략과 그린필드 전략을 혼합할 수 있습니다.

가드레일

조직 단위(OU) 전체에서 리소스, 정책 및 규정 준수를 관리하는 데 도움이 되는 중요 규칙입니다. 예방 가드레일은 규정 준수 표준에 부합하도록 정책을 시행하며, 서비스 제어 정책과 IAM 권한 경계를 사용하여 구현됩니다. 탐지 가드레일은 정책 위반 및 규정 준수 문제를 감지하고 해결을 위한 알림을 생성하며, 이는 AWS Config Amazon GuardDuty AWS Security Hub CSPM, , AWS Trusted Advisor Amazon Inspector 및 사용자 지정 AWS Lambda 검사를 사용하여 구현됩니다.

H

HA

[고가용성](#)을 참조하세요.

이기종 데이터베이스 마이그레이션

다른 데이터베이스 엔진을 사용하는 대상 데이터베이스로 소스 데이터베이스 마이그레이션(예: Oracle에서 Amazon Aurora로) 이기종 마이그레이션은 일반적으로 리아키텍트 작업의 일부이며 스

키마를 변환하는 것은 복잡한 작업일 수 있습니다. AWS 는 스키마 변환에 도움이 되는 [AWS SCT를](#) 제공합니다.

높은 가용성(HA)

문제나 재해 발생 시 개입 없이 지속적으로 운영할 수 있는 워크로드의 능력. HA 시스템은 자동으로 장애 조치되고, 지속적으로 고품질 성능을 제공하고, 성능에 미치는 영향을 최소화하면서 다양한 부하와 장애를 처리하도록 설계되었습니다.

히스토리언 현대화

제조 산업의 요구 사항을 더 잘 충족하도록 운영 기술(OT) 시스템을 현대화하고 업그레이드하는 데 사용되는 접근 방식입니다. 히스토리언은 공장의 다양한 출처에서 데이터를 수집하고 저장하는 데 사용되는 일종의 데이터베이스입니다.

홀드아웃 데이터

[기계 학습](#) 모델을 훈련하는 데 사용되는 데이터세트에서 보류되는 레이블이 지정된 기록 데이터의 일부입니다. 홀드아웃 데이터를 사용하여 모델 예측을 홀드아웃 데이터와 비교해 모델 성능을 평가할 수 있습니다.

동종 데이터베이스 마이그레이션

동일한 데이터베이스 엔진을 공유하는 대상 데이터베이스로 소스 데이터베이스 마이그레이션(예: Microsoft SQL Server에서 Amazon RDS for SQL Server로) 동종 마이그레이션은 일반적으로 리호스팅 또는 리플랫폼 작업의 일부입니다. 네이티브 데이터베이스 유틸리티를 사용하여 스키마를 마이그레이션할 수 있습니다.

핫 데이터

자주 액세스하는 데이터(예: 실시간 데이터 또는 최근 번역 데이터). 일반적으로 이 데이터에는 빠른 쿼리 응답을 제공하기 위한 고성능 스토리지 계층 또는 클래스가 필요합니다.

핫픽스

프로덕션 환경의 중요한 문제를 해결하기 위한 긴급 수정입니다. 핫픽스는 긴급하기 때문에 일반적인 DevOps 릴리스 워크플로 외부에서 실행됩니다.

하이퍼케어 기간

전환 직후 마이그레이션 팀이 문제를 해결하기 위해 클라우드에서 마이그레이션된 애플리케이션을 관리하고 모니터링하는 기간입니다. 일반적으로 이 기간은 1~4일입니다. 하이퍼케어 기간이 끝나면 마이그레이션 팀은 일반적으로 애플리케이션에 대한 책임을 클라우드 운영 팀에 넘깁니다.

I

IaC

[코드형 인프라](#)를 참조하세요.

자격 증명 기반 정책

AWS 클라우드 환경 내에서 권한을 정의하는 하나 이상의 IAM 보안 주체에 연결된 정책입니다.

유휴 애플리케이션

90일 동안 평균 CPU 및 메모리 사용량이 5~20%인 애플리케이션입니다. 마이그레이션 프로젝트에서는 이러한 애플리케이션을 사용 중지하거나 온프레미스에 유지하는 것이 일반적입니다.

IIoT

[산업용 사물 인터넷](#)을 참조하세요.

변경 불가능한 인프라

기존 인프라를 업데이트, 패치 또는 수정하는 대신 프로덕션 워크로드에 대한 새 인프라를 배포하는 모델입니다. 변경 불가능한 인프라는 [변경 가능한 인프라](#)보다 본질적으로 더 일관되고 안정적이며 예측 가능합니다. 자세한 내용은 AWS Well-Architected Framework의 [변경 불가능한 인프라를 사용하여 배포](#) 모범 사례를 참조하세요.

인바운드(수신) VPC

AWS 다중 계정 아키텍처에서 애플리케이션 외부에서 네트워크 연결을 수락, 검사 및 라우팅하는 VPC입니다. [AWS Security Reference Architecture](#)에서는 애플리케이션과 더 넓은 인터넷 간의 양방향 인터페이스를 보호하기 위해 인바운드, 아웃바운드 및 검사 VPC로 네트워크 계정을 설정할 것을 권장합니다.

증분 마이그레이션

한 번에 전체 전환을 수행하는 대신 애플리케이션을 조금씩 마이그레이션하는 전환 전략입니다. 예를 들어, 처음에는 소수의 마이크로서비스나 사용자만 새 시스템으로 이동할 수 있습니다. 모든 것이 제대로 작동하는지 확인한 후에는 레거시 시스템을 폐기할 수 있을 때까지 추가 마이크로서비스 또는 사용자를 점진적으로 이동할 수 있습니다. 이 전략을 사용하면 대규모 마이그레이션과 관련된 위험을 줄일 수 있습니다.

Industry 4.0

연결성, 실시간 데이터, 자동화, 분석 및 AI/ML의 발전을 통해 제조 프로세스의 현대화를 나타내기 위해 2016년에 [Klaus Schwab](#)에서 도입한 용어입니다.

인프라

애플리케이션의 환경 내에 포함된 모든 리소스와 자산입니다.

코드형 인프라(IaC)

구성 파일 세트를 통해 애플리케이션의 인프라를 프로비저닝하고 관리하는 프로세스입니다. IaC는 새로운 환경의 반복 가능성, 신뢰성 및 일관성을 위해 인프라 관리를 중앙 집중화하고, 리소스를 표준화하고, 빠르게 확장할 수 있도록 설계되었습니다.

산업용 사물 인터넷(IIoT)

제조, 에너지, 자동차, 의료, 생명과학, 농업 등의 산업 부문에서 인터넷에 연결된 센서 및 디바이스의 사용 자세한 내용은 [산업용 사물 인터넷\(IoT\) 디지털 트랜스포메이션 전략 구축](#)을 참조하십시오.

검사 VPC

AWS 다중 계정 아키텍처에서는 VPC(동일하거나 다른 AWS 리전), 인터넷 및 온프레미스 네트워크 간의 네트워크 트래픽 검사를 관리하는 중앙 집중식 VPCs입니다. [AWS Security Reference Architecture](#)에서는 애플리케이션과 더 넓은 인터넷 간의 양방향 인터페이스를 보호하기 위해 인바운드, 아웃바운드 및 검사 VPC로 네트워크 계정을 설정할 것을 권장합니다.

사물 인터넷(IoT)

인터넷이나 로컬 통신 네트워크를 통해 다른 디바이스 및 시스템과 통신하는 센서 또는 프로세서가 내장된 연결된 물리적 객체의 네트워크 자세한 내용은 [IoT란?](#)을 참조하십시오.

해석력

모델의 예측이 입력에 따라 어떻게 달라지는지를 사람이 이해할 수 있는 정도를 설명하는 기계 학습 모델의 특성입니다. 자세한 내용은 [기계 학습 모델 해석 가능성을 참조하세요 AWS](#).

IoT

[사물 인터넷](#)을 참조하세요.

IT 정보 라이브러리(ITIL)

IT 서비스를 제공하고 이러한 서비스를 비즈니스 요구 사항에 맞게 조정하기 위한 일련의 모범 사례 ITIL은 ITSM의 기반을 제공합니다.

IT 서비스 관리(ITSM)

조직의 IT 서비스 설계, 구현, 관리 및 지원과 관련된 활동 클라우드 운영을 ITSM 도구와 통합하는 방법에 대한 자세한 내용은 [운영 통합 가이드](#)를 참조하십시오.

ITIL

[IT 정보 라이브러리](#)를 참조하세요.

ITSM

[IT 서비스 관리](#)를 참조하세요.

L

레이블 기반 액세스 제어(LBAC)

사용자 및 데이터 자체에 각각 보안 레이블 값을 명시적으로 할당하는 필수 액세스 제어(MAC)를 구현한 것입니다. 사용자 보안 레이블과 데이터 보안 레이블 간의 교차 부분에 따라 사용자가 볼 수 있는 행과 열이 결정됩니다.

랜딩 존

랜딩 존은 확장 가능하고 안전한 잘 설계된 다중 계정 AWS 환경입니다. 조직은 여기에서부터 보안 및 인프라 환경에 대한 확신을 가지고 워크로드와 애플리케이션을 신속하게 시작하고 배포할 수 있습니다. 랜딩 존에 대한 자세한 내용은 [안전하고 확장 가능한 다중 계정 AWS 환경 설정](#)을 참조하십시오.

대규모 언어 모델(LLM)

방대한 양의 데이터에서 사전 훈련된 딥 러닝 [AI](#) 모델입니다. LLM은 질문에 대한 답변, 문서 요약, 텍스트를 다른 언어로 번역, 문장 완성과 같은 여러 태스크를 수행할 수 있습니다. 자세한 내용은 [대규모 언어 모델\(LLM\)이란 무엇인가요?](#)를 참조하세요.

대규모 마이그레이션

300대 이상의 서버 마이그레이션입니다.

LBAC

[레이블 기반 액세스 제어](#)를 참조하세요.

최소 권한

작업을 수행하는 데 필요한 최소 권한을 부여하는 보안 모범 사례입니다. 자세한 내용은 IAM 설명서의 [최소 권한 적용](#)을 참조하십시오.

리프트 앤드 시프트

[7R](#)을 참조하세요.

리틀 엔디안 시스템

가장 덜 중요한 바이트를 먼저 저장하는 시스템입니다. [엔디안](#)도 참조하세요.

LLM

[대규모 언어 모델](#)을 참조하세요.

하위 환경

[환경](#)을 참조하세요.

M

기계 학습(ML)

패턴 인식 및 학습에 알고리즘과 기법을 사용하는 인공지능의 한 유형입니다. ML은 사물 인터넷 (IoT) 데이터와 같은 기록된 데이터를 분석하고 학습하여 패턴을 기반으로 통계 모델을 생성합니다. 자세한 내용은 [기계 학습](#)을 참조하십시오.

기본 브랜치

[브랜치](#)를 참조하세요.

맬웨어

컴퓨터 보안 또는 프라이버시를 위협하도록 설계된 소프트웨어입니다. 맬웨어는 컴퓨터 시스템을 방해하거나 민감한 정보를 유출하거나 무단 액세스 권한을 확보할 수 있습니다. 맬웨어의 예로 바이러스, 웜, 랜섬웨어, 트로이 목마, 스파이웨어, 키로거 등이 있습니다.

관리형 서비스

AWS 서비스는 인프라 계층, 운영 체제 및 플랫폼을 AWS 운영하고, 사용자는 엔드포인트에 액세스하여 데이터를 저장하고 검색합니다. 관리형 서비스의 예로 Amazon Simple Storage Service(Amazon S3) 및 Amazon DynamoDB가 있습니다. 이를 추상화된 서비스라고도 합니다.

제조 실행 시스템(MES)

원자재를 생산 현장에서 완제품으로 변환하는 생산 프로세스를 추적, 모니터링, 문서화 및 제어하기 위한 소프트웨어 시스템입니다.

MAP

[Migration Acceleration Program](#)을 참조하세요.

메커니즘

도구를 생성하고 도구 채택을 유도한 다음 조정을 위해 결과를 검사하는 전체 프로세스입니다. 메커니즘은 작동 시 자체적으로 강화하고 개선하는 주기입니다. 자세한 내용은 AWS Well-Architected Framework의 [메커니즘 구축](#)을 참조하세요.

멤버 계정

조직의 일부인 관리 계정을 AWS 계정 제외한 모든 계정. AWS Organizations 하나의 계정은 한 번에 하나의 조직 멤버만 될 수 있습니다.

MES

[제조 실행 시스템](#)을 참조하세요.

메시지 큐 원격 분석 전송(MQTT)

리소스 제약이 있는 [IoT](#) 디바이스에 대한 [게시 및 구독](#) 패턴을 기반으로 하는 경량 Machine-to-Machine(M2M) 통신 프로토콜입니다.

마이크로서비스

잘 정의된 API를 통해 통신하고 일반적으로 소규모 자체 팀이 소유하는 소규모 독립 서비스입니다. 예를 들어, 보험 시스템에는 영업, 마케팅 등의 비즈니스 역량이나 구매, 청구, 분석 등의 하위 영역에 매핑되는 마이크로 서비스가 포함될 수 있습니다. 마이크로서비스의 이점으로 민첩성, 유연한 확장, 손쉬운 배포, 재사용 가능한 코드, 복원력 등이 있습니다. 자세한 내용은 [AWS 서버리스 서비스를 사용하여 마이크로서비스 통합을 참조하세요](#).

마이크로서비스 아키텍처

각 애플리케이션 프로세스를 마이크로서비스로 실행하는 독립 구성 요소를 사용하여 애플리케이션을 구축하는 접근 방식입니다. 이러한 마이크로서비스는 경량 API를 사용하여 잘 정의된 인터페이스를 통해 통신합니다. 애플리케이션의 특정 기능에 대한 수요에 맞게 이 아키텍처의 각 마이크로 서비스를 업데이트, 배포 및 조정할 수 있습니다. 자세한 내용은 [에서 마이크로서비스 구현을 참조하세요 AWS](#).

Migration Acceleration Program(MAP)

조직이 클라우드로 전환하기 위한 강력한 운영 기반을 구축하고 초기 마이그레이션 비용을 상쇄하는 데 도움이 되는 컨설팅 지원, 교육 및 서비스를 제공하는 AWS 프로그램입니다. MAP에는 레거시 마이그레이션을 체계적인 방식으로 실행하기 위한 마이그레이션 방법론과 일반적인 마이그레이션 시나리오를 자동화하고 가속화하는 도구 세트가 포함되어 있습니다.

대규모 마이그레이션

애플리케이션 포트폴리오의 대다수를 웨이브를 통해 클라우드로 이동하는 프로세스로, 각 웨이브에서 더 많은 애플리케이션이 더 빠른 속도로 이동합니다. 이 단계에서는 이전 단계에서 배운 모범 사례와 교훈을 사용하여 팀, 도구 및 프로세스의 마이그레이션 팩토리를 구현하여 자동화 및 민첩한 제공을 통해 워크로드 마이그레이션을 간소화합니다. 이것은 [AWS 마이그레이션 전략](#)의 세 번째 단계입니다.

마이그레이션 팩토리

자동화되고 민첩한 접근 방식을 통해 워크로드 마이그레이션을 간소화하는 다기능 팀입니다. 마이그레이션 팩토리 팀에는 일반적으로 스프린트에서 일하는 운영, 비즈니스 분석가 및 소유자, 마이그레이션 엔지니어, 개발자, DevOps 전문가가 포함됩니다. 엔터프라이즈 애플리케이션 포트폴리오의 20~50%는 공장 접근 방식으로 최적화할 수 있는 반복되는 패턴으로 구성되어 있습니다. 자세한 내용은 이 콘텐츠 세트의 [클라우드 마이그레이션 팩토리 가이드](#)와 [마이그레이션 팩토리에 대한 설명](#)을 참조하십시오.

마이그레이션 메타데이터

마이그레이션을 완료하는 데 필요한 애플리케이션 및 서버에 대한 정보 각 마이그레이션 패턴에는 서로 다른 마이그레이션 메타데이터 세트가 필요합니다. 마이그레이션 메타데이터의 예로는 대상 서브넷, 보안 그룹 및 AWS 계정이 있습니다.

마이그레이션 패턴

사용되는 마이그레이션 전략, 마이그레이션 대상, 마이그레이션 애플리케이션 또는 서비스를 자세히 설명하는 반복 가능한 마이그레이션 작업입니다. 예: AWS Application Migration Service를 사용하여 Amazon EC2로 마이그레이션을 리호스팅합니다.

Migration Portfolio Assessment(MPA)

AWS 클라우드로 마이그레이션하는 비즈니스 사례를 검증하기 위한 정보를 제공하는 온라인 도구입니다. MPA는 상세한 포트폴리오 평가(서버 적정 규모 조정, 가격 책정, TCO 비교, 마이그레이션 비용 분석)와 마이그레이션 계획(애플리케이션 데이터 분석 및 데이터 수집, 애플리케이션 그룹화, 마이그레이션 우선순위 지정, 웨이브 계획)을 제공합니다. [MPA 도구](#)(로그인 필요)는 모든 AWS 컨설턴트와 APN 파트너 컨설턴트가 무료로 사용할 수 있습니다.

마이그레이션 준비 상태 평가(MRA)

AWS CAF를 사용하여 조직의 클라우드 준비 상태에 대한 인사이트를 얻고, 강점과 약점을 식별하고, 식별된 격차를 해소하기 위한 행동 계획을 수립하는 프로세스입니다. 자세한 내용은 [마이그레이션 준비 가이드](#)를 참조하십시오. MRA는 [AWS 마이그레이션 전략](#)의 첫 번째 단계입니다.

마이그레이션 전략

워크로드를 AWS 클라우드로 마이그레이션하는 데 사용되는 접근 방식입니다. 자세한 내용은 이 용어집의 [7R 항목](#)과 [조직을 동원하여 대규모 마이그레이션 가속화](#)를 참조하세요.

ML

[기계 학습](#)을 참조하세요.

현대화

비용을 절감하고 효율성을 높이고 혁신을 활용하기 위해 구식(레거시 또는 모놀리식) 애플리케이션과 해당 인프라를 클라우드의 민첩하고 탄력적이고 가용성이 높은 시스템으로 전환하는 것입니다. 자세한 내용은 [AWS 클라우드에서 애플리케이션을 현대화하기 위한 전략](#)을 참조하세요.

현대화 준비 상태 평가

조직 애플리케이션의 현대화 준비 상태를 파악하고, 이점, 위험 및 종속성을 식별하고, 조직이 해당 애플리케이션의 향후 상태를 얼마나 잘 지원할 수 있는지를 확인하는 데 도움이 되는 평가입니다. 평가 결과는 대상 아키텍처의 청사진, 현대화 프로세스의 개발 단계와 마일스톤을 자세히 설명하는 로드맵 및 파악된 격차를 해소하기 위한 실행 계획입니다. 자세한 내용은 [AWS 클라우드에서 애플리케이션의 현대화 준비 상태 평가](#)를 참조하세요.

모놀리식 애플리케이션(모놀리식 유형)

긴밀하게 연결된 프로세스를 사용하여 단일 서비스로 실행되는 애플리케이션입니다. 모놀리식 애플리케이션에는 몇 가지 단점이 있습니다. 한 애플리케이션 기능에 대한 수요가 급증하면 전체 아키텍처 규모를 조정해야 합니다. 코드 베이스가 커지면 모놀리식 애플리케이션의 기능을 추가하거나 개선하는 것도 더 복잡해집니다. 이러한 문제를 해결하기 위해 마이크로서비스 아키텍처를 사용할 수 있습니다. 자세한 내용은 [마이크로서비스로 모놀리식 유형 분해](#)를 참조하십시오.

MPA

[Migration Portfolio Assessment](#)를 참조하세요.

MQTT

[메시지 큐 원격 분석 전송](#)을 참조하세요.

멀티클래스 분류

여러 클래스에 대한 예측(2개 이상의 결과 중 하나 예측)을 생성하는 데 도움이 되는 프로세스입니다. 예를 들어, ML 모델이 '이 제품은 책인가요, 자동차인가요, 휴대폰인가요?' 또는 '이 고객이 가장 관심을 갖는 제품 범주는 무엇인가요?'라고 물을 수 있습니다.

변경 가능한 인프라

프로덕션 워크로드에 대한 기존 인프라를 업데이트하고 수정하는 모델입니다. 일관성, 신뢰성 및 예측 가능성을 높이기 위해 AWS Well-Architected Framework에서는 [변경 불가능한 인프라](#)를 모범 사례로 사용할 것을 권장합니다.

O

OAC

[오리진 액세스 제어](#)를 참조하세요.

OAI

[오리진 액세스 ID](#)를 참조하세요.

OCM

[조직 변경 관리](#)를 참조하세요.

오프라인 마이그레이션

마이그레이션 프로세스 중 소스 워크로드가 중단되는 마이그레이션 방법입니다. 이 방법은 가동 중지 증가를 수반하며 일반적으로 작고 중요하지 않은 워크로드에 사용됩니다.

O

[운영 통합](#)을 참조하세요.

OLA

[운영 수준 계약](#)을 참조하세요.

온라인 마이그레이션

소스 워크로드를 오프라인 상태로 전환하지 않고 대상 시스템에 복사하는 마이그레이션 방법입니다. 워크로드에 연결된 애플리케이션은 마이그레이션 중에도 계속 작동할 수 있습니다. 이 방법은 가동 중지 차단 또는 최소화를 수반하며 일반적으로 중요한 프로덕션 워크로드에 사용됩니다.

OPC-UA

[Open Process Communications - Unified Architecture\(OPC-UA\)](#)를 참조하세요.

Open Process Communications - Unified Architecture(OPC-UA)

산업 자동화를 위한 Machine-to-Machine(M2M) 통신 프로토콜입니다. OPC-UA는 데이터 암호화, 인증 및 권한 부여 체계에 관한 상호 운용성 표준을 제공합니다.

운영 수준 협약(OLA)

서비스 수준에 관한 계약(SLA)을 지원하기 위해 직무 IT 그룹이 서로에게 제공하기로 약속한 내용을 명확히 하는 계약입니다.

운영 준비 상태 검토(ORR)

인시던트 및 잠재적 장애의 범위를 이해, 평가 또는 예방하거나 줄이는 데 도움이 되는 질문 체크리스트 및 관련 모범 사례입니다. 자세한 내용은 AWS Well-Architected Framework의 [운영 준비 상태 검토\(ORR\)](#)를 참조하세요.

운영 기술(OT)

물리적 환경에서 작동하여 산업 운영, 장비 및 인프라를 제어하는 하드웨어 및 소프트웨어 시스템입니다. 제조 분야에서 OT 및 정보 기술(IT) 시스템의 통합은 [Industry 4.0](#) 트랜스포메이션의 주요 중점 사항입니다.

운영 통합(OI)

클라우드에서 운영을 현대화하는 프로세스로 준비 계획, 자동화 및 통합을 수반합니다. 자세한 내용은 [운영 통합 가이드](#)를 참조하십시오.

조직 트레일

조직 AWS 계정 내 모든에 대한 모든 이벤트를 로깅 AWS CloudTrail 하는에서 생성된 추적입니다 AWS Organizations. 이 트레일은 조직에 속한 각 AWS 계정에 생성되고 각 계정의 활동을 추적합니다. 자세한 내용은 CloudTrail 설명서의 [Creating a trail for an organization](#)을 참조하십시오.

조직 변경 관리(OCM)

사람, 문화 및 리더십 관점에서 중대하고 파괴적인 비즈니스 혁신을 관리하기 위한 프레임워크입니다. OCM은 변화 채택을 가속화하고, 과도기적 문제를 해결하고, 문화 및 조직적 변화를 주도함으로써 조직이 새로운 시스템 및 전략을 준비하고 전환할 수 있도록 지원합니다. AWS 마이그레이션 전략에서는 클라우드 채택 프로젝트에 필요한 변경 속도 때문에이 프레임워크를 인력 가속화라고 합니다. 자세한 내용은 [사용 가이드](#)를 참조하십시오.

오리진 액세스 제어(OAC)

CloudFront에서 Amazon Simple Storage Service(S3) 콘텐츠를 보호하기 위해 액세스를 제한하는 고급 옵션입니다. OAC는 AWS KMS (SSE-KMS)를 사용한 모든 서버 측 암호화 AWS 리전와 S3 버킷에 대한 동적 PUT 및 DELETE 요청에서 모든 S3 버킷을 지원합니다.

오리진 액세스 ID(OAI)

CloudFront에서 Amazon S3 콘텐츠를 보호하기 위해 액세스를 제한하는 옵션입니다. OAI를 사용하면 CloudFront는 Amazon S3가 인증할 수 있는 보안 주체를 생성합니다. 인증된 보안 주체는 특

정 CloudFront 배포를 통해서만 S3 버킷의 콘텐츠에 액세스할 수 있습니다. 더 세분화되고 향상된 액세스 제어를 제공하는 [OAC](#)도 참조하십시오.

ORR

[운영 준비 상태 검토](#)를 참조하세요.

OT

[운영 기술](#)을 참조하세요.

아웃바운드(송신) VPC

AWS 다중 계정 아키텍처에서 애플리케이션 내에서 시작된 네트워크 연결을 처리하는 VPC입니다. [AWS Security Reference Architecture](#)에서는 애플리케이션과 더 넓은 인터넷 간의 양방향 인터페이스를 보호하기 위해 인바운드, 아웃바운드 및 검사 VPC로 네트워크 계정을 설정할 것을 권장합니다.

P

권한 경계

사용자나 역할이 가질 수 있는 최대 권한을 설정하기 위해 IAM 보안 주체에 연결되는 IAM 관리 정책입니다. 자세한 내용은 IAM 설명서의 [권한 경계](#)를 참조하십시오.

개인 식별 정보(PII)

직접 보거나 다른 관련 데이터와 함께 짝을 지을 때 개인의 신원을 합리적으로 추론하는 데 사용할 수 있는 정보입니다. PII의 예로는 이름, 주소, 연락처 정보 등이 있습니다.

PII

[개인 식별 정보](#)를 참조하세요.

플레이북

클라우드에서 핵심 운영 기능을 제공하는 등 마이그레이션과 관련된 작업을 캡처하는 일련의 사전 정의된 단계입니다. 플레이북은 스크립트, 자동화된 런북 또는 현대화된 환경을 운영하는 데 필요한 프로세스나 단계 요약의 형태를 취할 수 있습니다.

PLC

[프로그래밍 가능 로직 컨트롤러](#)를 참조하세요.

PLM

[제품 수명 주기 관리](#)를 참조하세요.

정책

권한 정의([ID 기반 정책](#) 참조), 액세스 조건 지정([리소스 기반 정책](#) 참조), AWS Organizations 내 조직의 모든 계정에 대한 최대 권한 정의([서비스 제어 정책](#) 참조)와 같은 작업을 수행할 수 있는 객체입니다.

다국어 지속성

데이터 액세스 패턴 및 기타 요구 사항을 기반으로 독립적으로 마이크로서비스의 데이터 스토리지 기술 선택. 마이크로서비스가 동일한 데이터 스토리지 기술을 사용하는 경우 구현 문제가 발생하거나 성능이 저하될 수 있습니다. 요구 사항에 가장 적합한 데이터 저장소를 사용하면 마이크로서비스를 더 쉽게 구현하고 성능과 확장성을 높일 수 있습니다.

포트폴리오 평가

마이그레이션을 계획하기 위해 애플리케이션 포트폴리오를 검색 및 분석하고 우선순위를 정하는 프로세스입니다. 자세한 내용은 [마이그레이션 준비 상태 평가](#)를 참조하십시오.

조건자

보통 WHERE 절에 있는 true 또는 false를 반환하는 쿼리 조건입니다.

푸시다운 조건자

전송 전에 쿼리의 데이터를 필터링하는 데이터베이스 쿼리 최적화 기법입니다. 이렇게 하면 관계형 데이터베이스에서 검색하고 처리해야 하는 데이터의 양이 줄고 쿼리 성능이 향상됩니다.

예방적 제어

이벤트 발생을 방지하도록 설계된 보안 제어입니다. 이 제어는 네트워크에 대한 무단 액세스나 원치 않는 변경을 방지하는 데 도움이 되는 1차 방어선입니다. 자세한 내용은 Implementing security controls on AWS의 [Preventative controls](#)를 참조하십시오.

보안 주체

작업을 수행하고 리소스에 액세스할 수 있는 AWS 있는의 엔터티입니다. 이 엔터티는 일반적으로 , AWS 계정 IAM 역할 또는 사용자의 루트 사용자입니다. 자세한 내용은 IAM 설명서의 [역할 용어 및 개념](#)의 보안 주체를 참조하십시오.

개인 정보 보호 중심 설계

전체 개발 프로세스에서 개인 정보를 고려하는 시스템 엔지니어링에서의 접근 방식입니다.

프라이빗 호스팅 영역

Amazon Route 53에서 하나 이상의 VPC 내 도메인과 하위 도메인에 대한 DNS 쿼리에 응답하는 방법에 대한 정보가 담긴 컨테이너입니다. 자세한 내용은 Route 53 설명서의 [프라이빗 호스팅 영역 작업](#)을 참조하십시오.

선제적 제어

규정 미준수 리소스의 배포를 방지하도록 설계된 [보안 제어](#)입니다. 이러한 제어는 리소스를 프로비저닝하기 전에 리소스를 스캔합니다. 리소스가 제어를 준수하지 않으면 프로비저닝되지 않습니다. 자세한 내용은 AWS Control Tower 설명서의 [제어 참조 가이드](#)를 참조하고 보안 [제어 구현의 사전 예방적 제어](#)를 참조하세요. AWS

제품 수명 주기 관리(PLM)

설계, 개발 및 출시부터 성장 및 성숙도를 거쳐 거부 및 제거에 이르기까지 전체 수명 주기 동안 제품의 데이터 및 프로세스 관리를 나타냅니다.

프로덕션 환경

[환경](#)을 참조하세요.

프로그래밍 가능 로직 컨트롤러(PLC)

제조 분야에서 기계를 모니터링하고 제조 프로세스를 자동화하는 매우 안정적이고 적응력이 뛰어난 컴퓨터입니다.

프롬프트 체이닝

한 [LLM](#) 프롬프트의 출력을 다음 프롬프트의 입력으로 사용하여 더 나은 응답을 생성합니다. 이 기법은 복잡한 작업을 하위 태스크로 나누거나 예비 응답을 반복적으로 세부 조정하거나 확장하는 데 사용됩니다. 이를 통해 모델 응답의 정확성과 관련성을 개선하고 보다 세분화되고 개인화된 결과를 얻을 수 있습니다.

가명화

데이터세트의 개인 식별자를 자리 표시자 값으로 바꾸는 프로세스입니다. 가명화는 개인 정보를 보호하는 데 도움이 될 수 있습니다. 가명화된 데이터는 여전히 개인 데이터로 간주됩니다.

게시/구독(pub/sub)

여러 마이크로서비스에서 비동기 통신을 지원하여 확장성과 응답성을 개선하는 패턴입니다. 예를 들어 마이크로서비스 기반 [MES](#)에서 마이크로서비스는 다른 마이크로서비스가 구독할 수 있는 채널에 이벤트 메시지를 게시할 수 있습니다. 시스템은 게시 서비스를 변경하지 않고도 새 마이크로서비스를 추가할 수 있습니다.

Q

쿼리 계획

SQL 관계형 데이터베이스 시스템의 데이터에 액세스하는 데 사용되는 명령어와 같은 일련의 단계입니다.

쿼리 계획 회귀

데이터베이스 서비스 최적화 프로그램이 데이터베이스 환경을 변경하기 전보다 덜 최적의 계획을 선택하는 경우입니다. 통계, 제한 사항, 환경 설정, 쿼리 파라미터 바인딩 및 데이터베이스 엔진 업데이트의 변경으로 인해 발생할 수 있습니다.

R

RACI 매트릭스

[Responsible, Accountable, Consulted, Informed\(RACI\)](#)를 참조하세요.

RAG

[검색 증강 생성](#)을 참조하세요.

랜섬웨어

결제가 완료될 때까지 컴퓨터 시스템이나 데이터에 대한 액세스를 차단하도록 설계된 악성 소프트웨어입니다.

RASCI 매트릭스

[Responsible, Accountable, Consulted, Informed\(RACI\)](#)를 참조하세요.

RCAC

[행 및 열 액세스 제어](#)를 참조하세요.

읽기 전용 복제본

읽기 전용 용도로 사용되는 데이터베이스의 사본입니다. 쿼리를 읽기 전용 복제본으로 라우팅하여 기본 데이터베이스의 로드를 줄일 수 있습니다.

리아키텍팅

[7R](#)을 참조하세요.

Recovery Point Objective(RPO)

마지막 데이터 복구 시점 이후 허용되는 최대 시간입니다. 이에 따라 마지막 복구 시점과 서비스 중단 사이에 허용되는 데이터 손실로 간주되는 범위가 결정됩니다.

Recovery Time Objective(RTO)

서비스 중단과 서비스 복원 사이의 허용 가능한 지연 시간입니다.

리팩터링

[7R](#)을 참조하세요.

리전

지리적 영역의 AWS 리소스 모음입니다. 각 AWS 리전은 내결함성, 안정성 및 복원력을 제공하기 위해 서로 격리되고 독립적입니다. 자세한 내용은 [계정에서 사용할 수 있는 AWS 리전 지정](#)을 참조하세요.

회귀

숫자 값을 예측하는 ML 기법입니다. 예를 들어, '이 집은 얼마에 팔릴까?'라는 문제를 풀기 위해 ML 모델은 선형 회귀 모델을 사용하여 주택에 대해 알려진 사실(예: 면적)을 기반으로 주택의 매매 가격을 예측할 수 있습니다.

리호스팅

[7R](#)을 참조하세요.

릴리스

배포 프로세스에서 변경 사항을 프로덕션 환경으로 승격시키는 행위입니다.

재배치

[7R](#)을 참조하세요.

리플랫폼

[7R](#)을 참조하세요.

재구매

[7R](#)을 참조하세요.

복원력

중단에 저항하거나 중단을 복구할 수 있는 애플리케이션의 기능입니다. [고가용성](#) 및 [재해 복구](#)는 AWS 클라우드에서 복원력을 계획할 때 일반적인 고려 사항입니다. 자세한 내용은 [AWS 클라우드 복원력](#)을 참조하세요.

리소스 기반 정책

Amazon S3 버킷, 엔드포인트, 암호화 키 등의 리소스에 연결된 정책입니다. 이 유형의 정책은 액세스가 허용된 보안 주체, 지원되는 작업 및 충족해야 하는 기타 조건을 지정합니다.

RACI(Responsible, Accountable, Consulted, Informed) 매트릭스

마이그레이션 활동 및 클라우드 운영에 참여하는 모든 당사자의 역할과 책임을 정의하는 매트릭스입니다. 매트릭스 이름은 매트릭스에 정의된 책임 유형에서 파생됩니다. 실무 담당자 (R), 의사 결정권자 (A), 업무 수행 조언자 (C), 결과 통보 대상자 (I). 지원자는 (S) 선택사항입니다. 지원자를 포함하면 매트릭스를 RASCI 매트릭스라고 하고, 지원자를 제외하면 RACI 매트릭스라고 합니다.

대응 제어

보안 기준에서 벗어나거나 부정적인 이벤트를 해결하도록 설계된 보안 제어입니다. 자세한 내용은 AWS에서 보안 제어 구현의 [대응 제어](#)를 참조하세요.

retain

[7R](#)을 참조하세요.

사용 중지

[7R](#)을 참조하세요.

검색 증강 세대(RAG)

응답을 생성하기 전에 [LLM](#)이 훈련 데이터 소스 외부에 있는 신뢰할 수 있는 데이터 소스를 참조하는 [생성형 AI](#) 기술입니다. 예를 들어 RAG 모델은 조직의 지식 기반 또는 사용자 지정 데이터에 대한 시맨틱 검색을 수행할 수 있습니다. 자세한 내용은 [검색 증강 생성\(RAG\)이란 무엇인가요?](#)를 참조하세요.

교체

공격자가 자격 증명에 액세스하는 것을 더욱 어렵게 만들기 위해 [보안 암호](#)를 주기적으로 업데이트 하는 프로세스입니다.

행 및 열 액세스 제어(RCAC)

액세스 규칙이 정의된 기본적이고 유연한 SQL 표현식을 사용합니다. RCAC는 행 권한과 열 마스크로 구성됩니다.

RPO

[목표 복구 시점\(RPO\)](#)을 참조하세요.

RTO

[목표 복구 시간\(RTO\)](#)을 참조하세요.

런북

특정 작업을 수행하는 데 필요한 일련의 수동 또는 자동 절차입니다. 일반적으로 오류율이 높은 반복 작업이나 절차를 간소화하기 위해 런북을 만듭니다.

S

SAML 2.0

많은 ID 제공업체(idP)에서 사용하는 개방형 표준입니다. 이 기능을 사용하면 연동 SSO(Single Sign-On)를 AWS Management Console 사용할 수 있으므로 사용자는 조직의 모든 사용자에게 대해 IAM에서 사용자를 생성하지 않고도 로그인하거나 AWS API 작업을 호출할 수 있습니다. SAML 2.0 기반 페더레이션에 대한 자세한 내용은 IAM 설명서의 [SAML 2.0 기반 페더레이션 정보](#)를 참조하십시오.

SCADA

[감독 제어 및 데이터 획득](#)을 참조하세요.

SCP

[서비스 제어 정책](#)을 참조하세요.

보안 암호

에는 암호화된 형식으로 저장하는 암호 또는 사용자 자격 증명과 같은 AWS Secrets Manager 기밀 또는 제한된 정보가 있습니다. 보안 암호 값과 메타데이터로 구성됩니다. 보안 암호 값은 바이너리, 단일 문자열 또는 여러 문자열일 수 있습니다. 자세한 내용은 AWS Secrets Manager 설명서의 [Secrets Manager 보안 암호란 무엇인가요?](#)를 참조하세요.

보안 중심 설계

전체 개발 프로세스에서 보안을 고려하는 시스템 엔지니어링에서의 접근 방식입니다.

보안 제어

위험 행위자가 보안 취약성을 악용하는 능력을 방지, 탐지 또는 감소시키는 기술적 또는 관리적 가드레일입니다. 보안 제어는 [예방](#), [감지](#), [대응](#), [선제적](#)과 같은 기본적인 네 가지 보안 제어 유형으로 구분됩니다.

보안 강화

공격 표면을 줄여 공격에 대한 저항력을 높이는 프로세스입니다. 더 이상 필요하지 않은 리소스 제거, 최소 권한 부여의 보안 모범 사례 구현, 구성 파일의 불필요한 기능 비활성화 등의 작업이 여기에 포함될 수 있습니다.

보안 정보 및 이벤트 관리(SIEM) 시스템

보안 정보 관리(SIM)와 보안 이벤트 관리(SEM) 시스템을 결합하는 도구 및 서비스입니다. SIEM 시스템은 서버, 네트워크, 디바이스 및 기타 소스에서 데이터를 수집, 모니터링 및 분석하여 위협과 보안 침해를 탐지하고 알림을 생성합니다.

보안 응답 자동화

보안 이벤트에 자동으로 응답하거나 이를 해결하도록 설계된 사전 정의되고 프로그래밍된 작업입니다. 이러한 자동화는 보안 모범 사례를 구현하는 데 도움이 되는 [탐지 또는 대응](#) AWS 보안 제어 역할을 합니다. 자동화된 응답 작업의 예로 VPC 보안 그룹 수정, Amazon EC2 인스턴스 패치 적용 또는 자격 증명 교체 등이 있습니다.

서버 측 암호화

대상에서 데이터를 수신하는 AWS 서비스에 의한 데이터 암호화.

서비스 제어 정책(SCP)

AWS Organizations에 속한 조직의 모든 계정에 대한 권한을 중앙 집중식으로 제어하는 정책입니다. SCP는 관리자가 사용자 또는 역할에 위임할 수 있는 작업에 대해 제한을 설정하거나 가드레일을 정의합니다. SCP를 허용 목록 또는 거부 목록으로 사용하여 허용하거나 금지할 서비스 또는 작업을 지정할 수 있습니다. 자세한 내용은 AWS Organizations 설명서의 [서비스 제어 정책을](#) 참조하세요.

서비스 엔드포인트

에 대한 진입점의 URL입니다 AWS 서비스. 엔드포인트를 사용하여 대상 서비스에 프로그래밍 방식으로 연결할 수 있습니다. 자세한 내용은 AWS 일반 참조의 [AWS 서비스 엔드포인트](#)를 참조하십시오.

서비스 수준에 관한 계약(SLA)

IT 팀이 고객에게 제공하기로 약속한 내용(예: 서비스 가동 시간 및 성능)을 명시한 계약입니다.

서비스 수준 지표(SLI)

오류 발생률, 가용성 또는 처리량과 같은 서비스의 성능 측면에 대한 측정값입니다.

서비스 수준 목표(SLO)

[서비스 수준 지표](#)로 측정되는 서비스의 상태를 나타내는 목표 지표입니다.

공동 책임 모델

클라우드 보안 및 규정 준수를 AWS 위해와 공유하는 책임을 설명하는 모델입니다. AWS 는 클라우드의 보안을 담당하는 반면, 사용자는 클라우드의 보안을 담당합니다. 자세한 내용은 [공동 책임 모델](#)을 참조하십시오.

SIEM

[보안 정보 및 이벤트 관리 시스템](#)을 참조하세요.

단일 장애점(SPOF)

애플리케이션을 중단시킬 수 있는 애플리케이션의 중요한 단일 구성 요소에서 발생하는 장애입니다.

SLA

[서비스 수준 계약](#)을 참조하세요.

SLI

[서비스 수준 지표](#)를 참조하세요.

SLO

[서비스 수준 목표](#)를 참조하세요.

분할 앤 시드 모델

현대화 프로젝트를 확장하고 가속화하기 위한 패턴입니다. 새로운 기능과 제품 릴리스가 정의되면 핵심 팀이 분할되어 새로운 제품 팀이 만들어집니다. 이를 통해 조직의 역량과 서비스 규모를 조정하고, 개발자 생산성을 개선하고, 신속한 혁신을 지원할 수 있습니다. 자세한 내용은 [AWS 클라우드에서 애플리케이션을 현대화하기 위한 단계별 접근 방식](#)을 참조하세요.

SPOF

[단일 장애점](#)을 참조하세요.

스타 스키마

하나의 큰 팩트 테이블을 사용하여 트랜잭션 또는 측정된 데이터를 저장하고 하나 이상의 더 작은 차원 테이블을 사용하여 데이터 속성을 저장하는 데이터베이스 조직 구조입니다. 이 구조는 [데이터 웨어하우스](#)에서 또는 비즈니스 인텔리전스 목적으로 사용하도록 설계되었습니다.

Strangler Fig 패턴

레거시 시스템을 폐기할 수 있을 때까지 시스템 기능을 점진적으로 다시 작성하고 교체하여 모놀리식 시스템을 현대화하기 위한 접근 방식. 이 패턴은 무화과 덩굴이 나무로 자라 결국 숙주를 압도하고 대체하는 것과 비슷합니다. [Martin Fowler](#)가 모놀리식 시스템을 다시 작성할 때 위험을 관리하는 방법으로 이 패턴을 도입했습니다. 이 패턴을 적용하는 방법의 예는 [컨테이너 및 Amazon API Gateway를 사용하여 기존의 Microsoft ASP.NET\(ASMX\) 웹 서비스를 점진적으로 현대화하는 방법](#)을 참조하십시오.

서브넷

VPC의 IP 주소 범위입니다. 서브넷은 단일 가용 영역에 상주해야 합니다.

감독 제어 및 데이터 획득(SCADA)

제조 분야에서 하드웨어와 소프트웨어를 사용하여 물리적 자산과 프로덕션 작업을 모니터링하는 시스템입니다.

대칭 암호화

동일한 키를 사용하여 데이터를 암호화하고 복호화하는 암호화 알고리즘입니다.

합성 테스트

사용자 상호 작용을 시뮬레이션하여 잠재적 문제를 감지하거나 성능을 모니터링하는 방식으로 진행되는 시스템 테스트입니다. [Amazon CloudWatch Synthetics](#)를 사용하여 이러한 테스트를 생성할 수 있습니다.

시스템 프롬프트

[LLM](#)에 컨텍스트, 명령 또는 지침을 제공하여 동작을 지시하는 기법입니다. 시스템 프롬프트는 컨텍스트를 설정하고 사용자와의 상호 작용을 위한 규칙을 설정하는 데 도움이 됩니다.

T

tags

AWS 리소스를 구성하기 위한 메타데이터 역할을 하는 키-값 페어입니다. 태그를 사용하면 리소스를 손쉽게 관리, 식별, 정리, 검색, 필터링할 수 있습니다. 자세한 내용은 [AWS 리소스에 태그 지정](#)을 참조하십시오.

대상 변수

지도 ML에서 예측하려는 값으로, 결과 변수라고도 합니다. 예를 들어, 제조 설정에서 대상 변수는 제품 결함일 수 있습니다.

작업 목록

런북을 통해 진행 상황을 추적하는 데 사용되는 도구입니다. 작업 목록에는 런북의 개요와 완료해야 할 일반 작업 목록이 포함되어 있습니다. 각 일반 작업에 대한 예상 소요 시간, 소유자 및 진행 상황이 작업 목록에 포함됩니다.

테스트 환경

[환경](#)을 참조하세요.

훈련

ML 모델이 학습할 수 있는 데이터를 제공하는 것입니다. 훈련 데이터에는 정답이 포함되어야 합니다. 학습 알고리즘은 훈련 데이터에서 대상(예측하려는 답)에 입력 데이터 속성을 매핑하는 패턴을 찾고, 이러한 패턴을 캡처하는 ML 모델을 출력합니다. 그런 다음 ML 모델을 사용하여 대상을 모르는 새 데이터에 대한 예측을 할 수 있습니다.

Transit Gateway

VPC와 온프레미스 네트워크를 상호 연결하는 데 사용할 수 있는 네트워크 전송 허브입니다. 자세한 내용은 AWS Transit Gateway 설명서의 [전송 게이트웨이란 무엇입니까?](#)를 참조하세요.

트렁크 기반 워크플로

개발자가 기능 브랜치에서 로컬로 기능을 구축하고 테스트한 다음 해당 변경 사항을 기본 브랜치에 병합하는 접근 방식입니다. 이후 기본 브랜치는 개발, 프로덕션 이전 및 프로덕션 환경에 순차적으로 구축됩니다.

신뢰할 수 있는 액세스

사용자를 대신하여 AWS Organizations 및 해당 계정에서 조직에서 작업을 수행하도록 지정하는 서비스에 대한 권한 부여. 신뢰할 수 있는 서비스는 필요할 때 각 계정에 서비스 연결 역할을 생성하여 관리 작업을 수행합니다. 자세한 내용은 설명서의 [다른 AWS 서비스와 AWS Organizations 함께 사용](#)을 참조하세요 AWS Organizations .

튜닝

ML 모델의 정확도를 높이기 위해 훈련 프로세스의 측면을 여러 변경하는 것입니다. 예를 들어, 레이블링 세트를 생성하고 레이블을 추가한 다음 다양한 설정에서 이러한 단계를 여러 번 반복하여 모델을 최적화하는 방식으로 ML 모델을 훈련할 수 있습니다.

피자 두 판 팀

피자 두 판이면 충분한 소규모 DevOps 팀. 피자 두 판 팀 규모는 소프트웨어 개발에 있어 가능한 최상의 공동 작업 기회를 보장합니다.

U

불확실성

예측 ML 모델의 신뢰성을 저해할 수 있는 부정확하거나 불완전하거나 알려지지 않은 정보를 나타내는 개념입니다. 불확실성에는 두 가지 유형이 있습니다. 인식론적 불확실성은 제한적이고 불완전한 데이터에 의해 발생하는 반면, 우연한 불확실성은 데이터에 내재된 노이즈와 무작위성에 의해 발생합니다.

차별화되지 않은 작업

애플리케이션을 만들고 운영하는 데 필요하지만 최종 사용자에게 직접적인 가치를 제공하거나 경쟁 우위를 제공하지 못하는 작업을 헤비 리프팅이라고도 합니다. 차별화되지 않은 작업의 예로는 조달, 유지보수, 용량 계획 등이 있습니다.

상위 환경

[환경](#)을 참조하세요.

V

정리

스토리지를 회수하고 성능을 향상시키기 위해 증분 업데이트 후 정리 작업을 수반하는 데이터베이스 유지 관리 작업입니다.

버전 제어

리포지토리의 소스 코드 변경과 같은 변경 사항을 추적하는 프로세스 및 도구입니다.

VPC 피어링

프라이빗 IP 주소를 사용하여 트래픽을 라우팅할 수 있게 하는 두 VPC 간의 연결입니다. 자세한 내용은 Amazon VPC 설명서의 [VPC 피어링이란?](#)을 참조하십시오.

취약성

시스템 보안을 손상시키는 소프트웨어 또는 하드웨어 결함입니다.

W

웜 캐시

자주 액세스하는 최신 관련 데이터를 포함하는 버퍼 캐시입니다. 버퍼 캐시에서 데이터베이스 인스턴스를 읽을 수 있기 때문에 주 메모리나 디스크에서 읽는 것보다 빠릅니다.

웜 데이터

자주 액세스하지 않는 데이터입니다. 이런 종류의 데이터를 쿼리할 때는 일반적으로 적절히 느린 쿼리가 허용됩니다.

창 함수

현재 레코드와 어떤 식으로든 관련된 행 그룹에서 계산을 수행하는 SQL 함수입니다. 창 함수는 이동 평균을 계산하거나 현재 행의 상대적 위치를 기반으로 행 값에 액세스하는 등의 태스크를 처리하는 데 유용합니다.

워크로드

고객 대면 애플리케이션이나 백엔드 프로세스 같이 비즈니스 가치를 창출하는 리소스 및 코드 모음입니다.

워크스트림

마이그레이션 프로젝트에서 특정 작업 세트를 담당하는 직무 그룹입니다. 각 워크스트림은 독립적이지만 프로젝트의 다른 워크스트림을 지원합니다. 예를 들어, 포트폴리오 워크스트림은 애플리케이션 우선순위 지정, 웨이브 계획, 마이그레이션 메타데이터 수집을 담당합니다. 포트폴리오 워크스트림은 이러한 자산을 마이그레이션 워크스트림에 전달하고, 마이그레이션 워크스트림은 서버와 애플리케이션을 마이그레이션합니다.

WORM

[Write Once, Read Many\(WORM\)](#)를 참조하세요.

WQF

[AWS Workload Qualification Framework](#)를 참조하세요.

Write Once Read Many(WORM)

데이터를 한 번 쓰고 데이터가 삭제되거나 수정되지 않도록 하는 스토리지 모델입니다. 권한 있는 사용자는 필요한 만큼 여러 번 데이터를 읽을 수 있지만 데이터를 변경할 수는 없습니다. 이 데이터 스토리지 인프라는 [변경 불가능](#)한 항목으로 간주됩니다.

Z

제로데이 익스플로잇

[제로데이 취약성](#)을 악용하는 공격(일반적으로 맬웨어)입니다.

제로데이 취약성

프로덕션 시스템의 명백한 결함 또는 취약성입니다. 위협 행위자는 이러한 유형의 취약성을 사용하여 시스템을 공격할 수 있습니다. 개발자는 공격의 결과로 취약성을 인지하는 경우가 많습니다.

제로샷 프롬프팅

태스크를 수행하기 위해 [LLM](#)에 명령을 제공하지만 안내에 도움이 되는 예제(샷)는 제공하지 않습니다. LLM은 사전 훈련된 지식을 사용하여 태스크를 처리해야 합니다. 제로샷 프롬프팅의 효과는 태스크의 복잡성과 프롬프트의 품질에 따라 달라집니다. [퓨샷 프롬프팅](#)도 참조하세요.

좀비 애플리케이션

평균 CPU 및 메모리 사용량이 5% 미만인 애플리케이션입니다. 마이그레이션 프로젝트에서는 이러한 애플리케이션을 사용 중지하는 것이 일반적입니다.

기계 번역으로 제공되는 번역입니다. 제공된 번역과 원본 영어의 내용이 상충하는 경우에는 영어 버전이 우선합니다.