



생성형 AI에 대한 기업의 최근 적용 방향

김선수

AI/ML 사업개발 담당

Principal AI/ML GTM Specialist, World-wide Specialist Organization

Amazon Web Services

Generative AI (생성형 AI)

- 대화, 스토리, 이미지, 동영상, 음악 등 **콘텐츠**와 **아이디어**를 생성합니다.
- 방대한 데이터 뭉치들에 대해 사전 학습된 대규모 모델을 기반으로 하며, 일반적으로 **Foundation Models(FM)**이라고 합니다.

2023

The Year of POCs



생성형 AI가 무엇인가요?

안전한가요?

제가 프롬프트 엔지니어가 되어야 하나요?

모델을 어떻게 고릅니까?

어디서부터 시작하죠?



이게 내 사업에 어떤 의미인가요?

파운데이션 모델이 뭡니까?



어떤 모델을 시도해야 하나요?

FM이 뭐죠?

LLM이 무엇인가요?

Generative AI vs. Hype Cycle (2024년 기준)



2024

The Year of Production

(FOR SOME)



프로젝트 우선순위를 어떻게 하죠?

비용을 어떻게 줄일 수 있나요?

이것을 어떻게 실현합니까?

어떠한 최적화 방법을 사용합니까?



이걸 어떻게 확장하나요?

어떤 모델을 선택해야 합니까?

나만의 모델을 훈련해야 하나요?

위험을 관리하는 방법?



더 빨리 움직이는 방법?

생성형 AI의 활용 부문

대화 이력 검색

컨텐츠 로컬화

글, 이미지, 영상 생성

글 요약

코드 생성

문서 처리

컨텐츠 정리

합성 데이터 생성

관리 지원

이상 징후 감지

웹페이지용 이미지 생성

비디오 향상

음악 생성

영상 향상

애니메이션 생성

챗봇 & 가상 비서

상담원 지원

콜센터 분석

개인화

1

생산성 강화

2

운영 개선

3

창의/창조

4

고객 경험 강화

회사 내부 중심

대외 고객 중심

1 생산성 강화

다양한 배경 이미지 생성, 프로모션에 활용



- 사용자의 관점에 따라 다양하게 이미지 배경을 구사하여 개인 / 국가 / 문화 별로 적절한 서비스 제공이 가능

제품
프로모션



Inpaint + ControlNet
Inpaint + Canny model



C A N N Y



1 생산성 강화

다양한 배경 이미지 생성, 프로모션에 활용

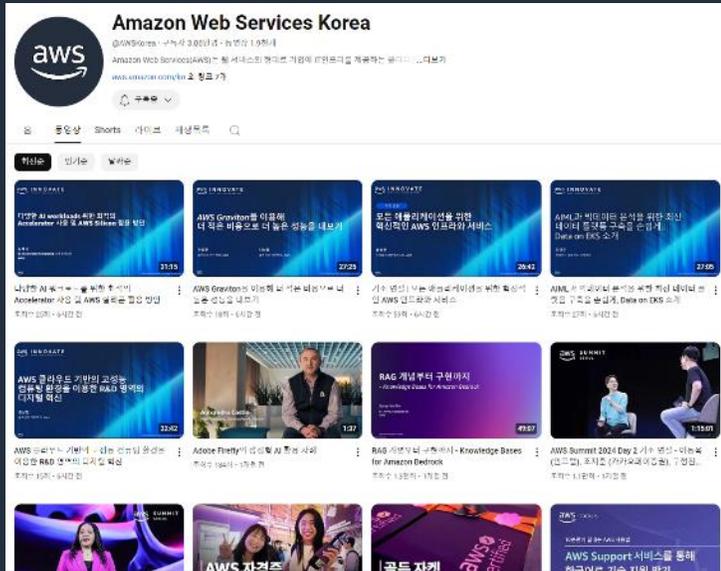


- Amazon Ads는 생성형 AI 기술을 사용하여 광고주가 고객에게 더 흥미롭고 관련성 높은 동영상 광고를 더 쉽게 만들 수 있는 새로운 기능을 발표
- 비디오 생성기는 추가 비용 없이 몇 분 만에 시각적으로 풍부한 동영상 콘텐츠를 만들 수 있음
- 비디오 생성기는 단일 제품 이미지를 사용하여 상품의 고유한 판매 제안 및 기능에 맞게 맞춤형 AI 생성 비디오를 큐레이팅하고, 아마존의 고유한 통찰력을 활용하여 제품 스토리를 생생하게 전달

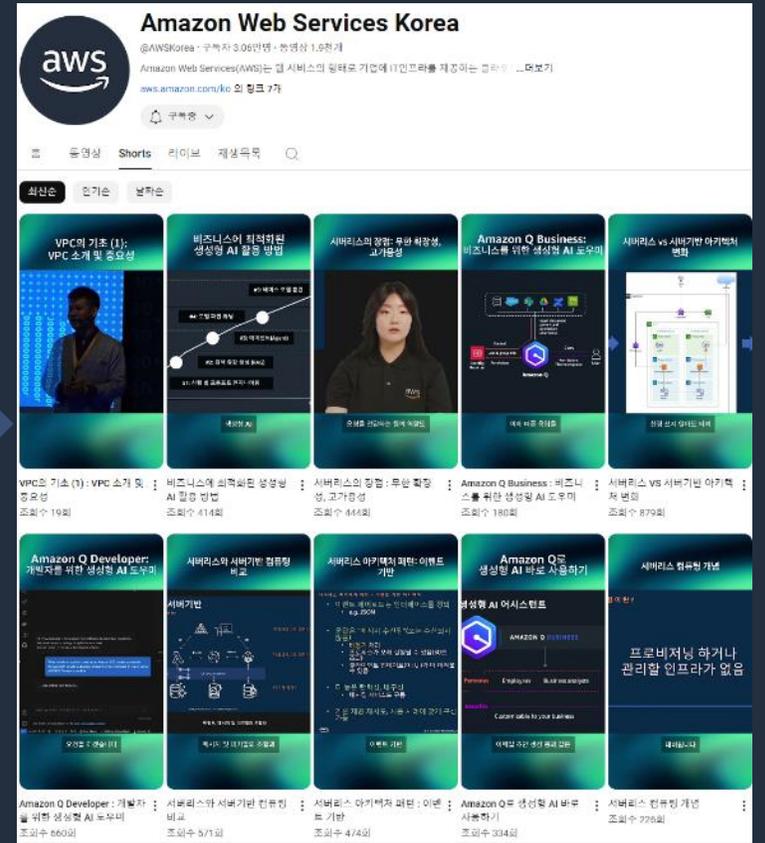


긴 영상이므로, 특정 기능에 대해서만 정보를 얻기 어려움

기존 AWS 채널의 영상을 짧은 기능 단위의 숏폼으로 자동화 편집

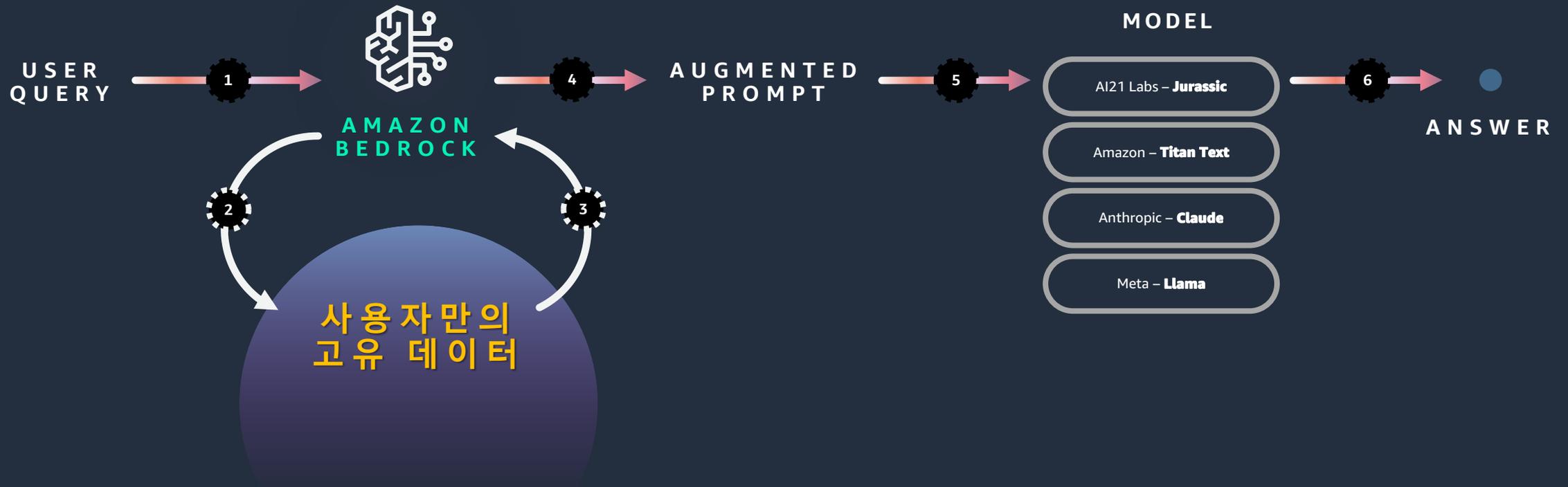


- AWS 서비스 활용**
Amazon Bedrock, Amazon Transcribe, 서버리스 아키텍처 활용하여 자동화 툴 제작
- 짧은 영상으로 집중**
1시간짜리 영상을 10개 내외의 짧은 핵심 콘텐츠들로 재구성
- 비용 최소화**
서버리스 아키텍처 구성, 비용 최소화
1시간 영상에 대해 \$1이내



Retrieval Augmented Generation (RAG - 검색증강생성)

Foundation Model을 **User의 데이터** 소스에 안전하게 연결하여 보다 **관련성 높은 응답 결과** 제공하는 방법



Example

Step 1: Upload document to platform



256 Page PDF guide

Step 2: Ask it questions

변압기 근처에서 불에 탄 종이를 발견했습니다. 이게 문제인가요?

DP 테스트는 어떻게 수행하나요?

Step 3: Get the answer

추가 정보 및 테스트 없이는 변압기 외부의 불에 탄 용지가 문제인지 확실하지 않습니다. 이 문서에는 DP 테스트 수행 방법에 대한 정보가 나와 있습니다...

See: pages 138, 162, 197, 216

DP 테스트를 수행하려면 중앙 페이즈 상단 근처의 편리한 위치에서 종이 단열재 샘플을 제거합니다.

See: page 243

2 운영 개선

생성형 AI 기반의 EDS (Equipment Diagnostics Service)

DOOSAN

EDS AI

안녕하세요. 설비 진단을 시작 할까요?

오늘 진단이 필요한 설비를 알려줘.

오늘의 설비 진단 현황입니다.
전체 설비 100대 중 정상 설비 96대, 이상 감지 2대, 점검이 필요한 설비 2대 입니다.

점검이 필요한 설비 목록을 보여줘.

점검이 필요한 설비 목록 입니다.
- F1공장 A-MTR-PUMP-01
- F3공장 A-TBN-03

Type your question...

직관적인 채팅 형태의 UX

실시간 설비 상태 조회

설비 과거 정비 이력 조회

제원 및 부품 정보 연동

점검 매뉴얼 검색 및 조회

진단 Report 생성



진단 Database



진단 Report



Sensor Data



Legacy



2 운영 개선

DDI EDS와 Amazon Bedrock 통합 시연 데모



The screenshot displays the DDI EDS dashboard with the following components:

- Header:** DOOSAN logo, navigation menu, and user profile (관리자님).
- Left Sidebar:** Monitoring (모니터링) menu with sub-items: 고장 진단 현황, 설비 모니터링, 디지털맵, 진단 Report, 진단 기준정보 관리, 시스템 관리, and 사용자 매뉴얼.
- Main Content Area:**
 - 고장 진단 현황 (Fault Diagnosis Status):**
 - 공장전체현황 (120): Overview of plant-wide status with warning (주의) and alert (경고) counts.
 - Factory Summary Table:
 - 설비별 현황 (Equipment Status):** Table listing equipment ID, date, and content.
 - 알람/레포트 현황 (Alarm/Report Status):** Summary of total alarms and reports.
 - 고장 발생 유형 분석 (Fault Occurrence Type Analysis):** Donut chart and bar chart showing fault distribution.

Factory	경고 (Warning)	주의 (Caution)	양호 (Good)
Factory1 (46)	0	19	27
Factory2 (68)	0	26	42
Factory3 (6)	0	0	6

설비 (Equipment)	날짜 (Date)	내용 (Content)
F1-EQ105A	2023-05-25 09:08:05	[Factory1] CHARGE PUMP
F1-EQ105B	2024-02-01 09:39:43	[Factory1] CHARGE PUMP
F1-EQ1501A	2022-12-31 00:00:00	[Factory1] CHARGE PUMP

Category	Count
알람 총 건수 (Total Alarms)	67 건
레포트 검토편 (Reports for Review)	10 건
레포트 미검토편 (Reports Not Reviewed)	57 건
누적 레포트 발생 (Cumulative Reports Occurred)	396건





2 운영 개선

반복적인 업무에 대한 개선 필요성

매장 파트너

POS가 고장 나면 어떻게 하나?
 매장 일 정산이 안되는데 어떻게 하지?
 각종 제증명 서류는 누구에게 청구하나?
 업무 택시는 어떻게 사용해요?
 법인카드 사용 지침은 어디서 확인하나요?
 관련 매뉴얼은 어디서 확인해요?

....

지원센터

사내 규정 최신 버전은 어디에 있을까?
 지원센터 장비 사용 방법이 궁금해요
 복리후생은 어떻게 신청하나요?
 단순 질문/답변의 반복으로 인한 업무 증가
 신규/경력 입사 시, 알아야 할 것이 많은데
 어디서 뭘 확인해야 할지..

24,500+

전국 매장에 산재된
많은 파트너 수

200+

다양한 규정, 지침 문서.
AP, SCI 의 용어들

x10

정보검색 툴의 부재
반복적인 대응

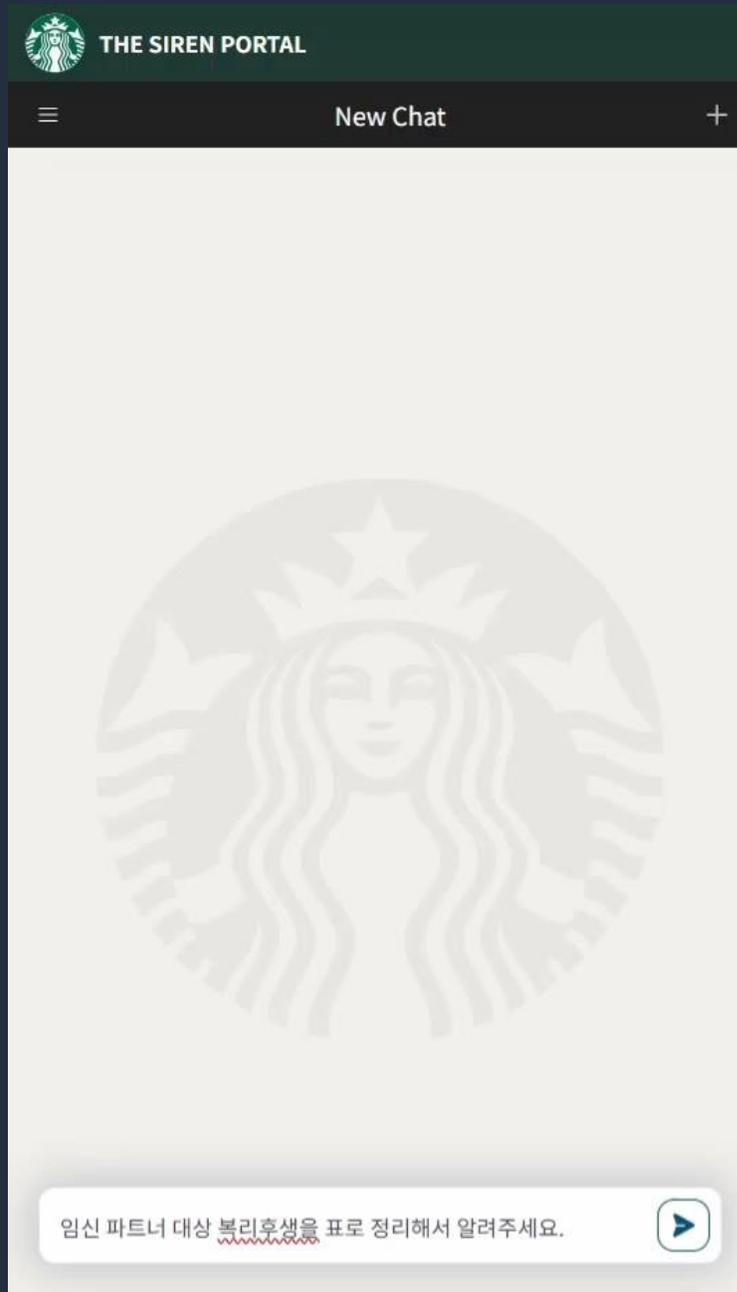
산재된 정보의 통합과 자동화 된 응답 체계가 필요

2

운영 개선

매장 파트너의
접근성을 고려

반응형 WEB



질문

답변

요약

참고 자료 (Source)

피드백 (좋아요/싫어요)



2 운영 개선

스타벅스 코리아의 사이렌포탈

30,000+

첫 일주일 간 질문 건수
현재 주 1만건 내외

95%+

Right Question
매장 운영 관련 24.8%
서비스 기준 매뉴얼 17%
메뉴 및 레시피 14.3%

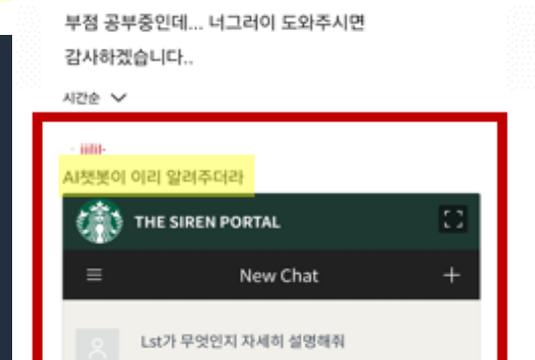
95%+

Right Answer
첫주 전수 검사

물음표 살인마인데 챗봇 좋다ㅋㅋㅋㅋ
궁금한 거 있으면 메모한 후 별백에서 못 찾으면 매장
선임한테 다 물어보거든? 근데 챗봇 있으니까 한
70퍼는 답변 다 해주네 ai가 짱이야

오 사이렌포탈 좋다..?
은근... 면접질문같은거 물어봤는데 대답해주네??
ㄴ 신기해...

특정 부분 모르겠으면 **봇 아용해봐여! 예상외로 답변 괜찮음**
ㅎㅎ
9분 · 👍 좋아요 · 💬 대댓글



나도 오늘 보고 뭐지? 했음 ㅋㅋ
3일 · 👍 좋아요 · 💬 대댓글

이거 대답 생각보다 잘해줌
3일 · 👍 좋아요 · 💬 대댓글

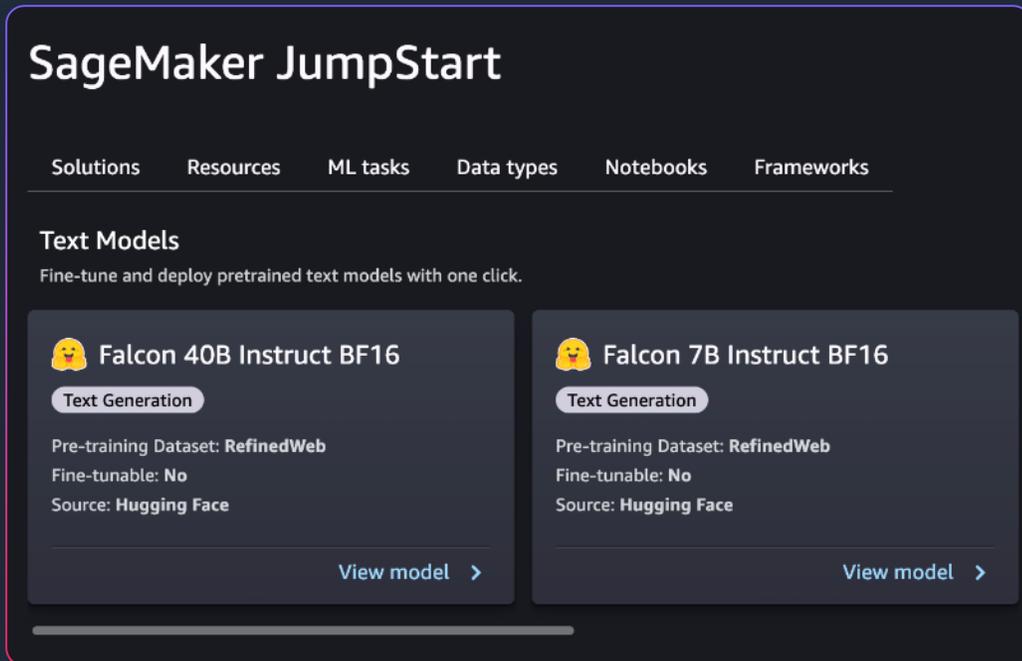


3

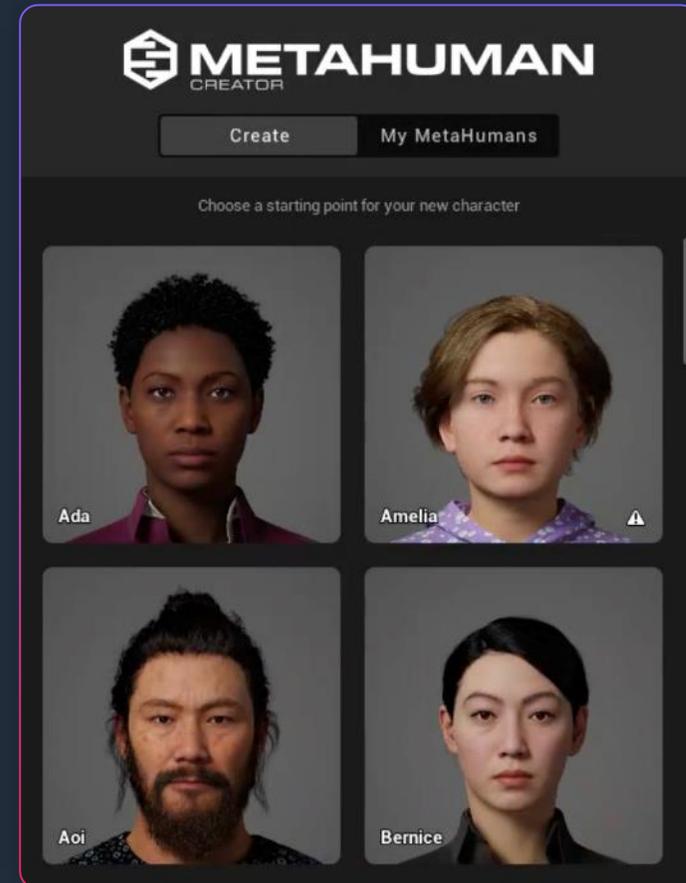
창의/창조

AI 기반 디지털 휴먼 구성

사용 중인 LLM은 쉽게 상호 교환할 수 있습니다.



디지털 휴먼 캐릭터도 전환할 수 있습니다.

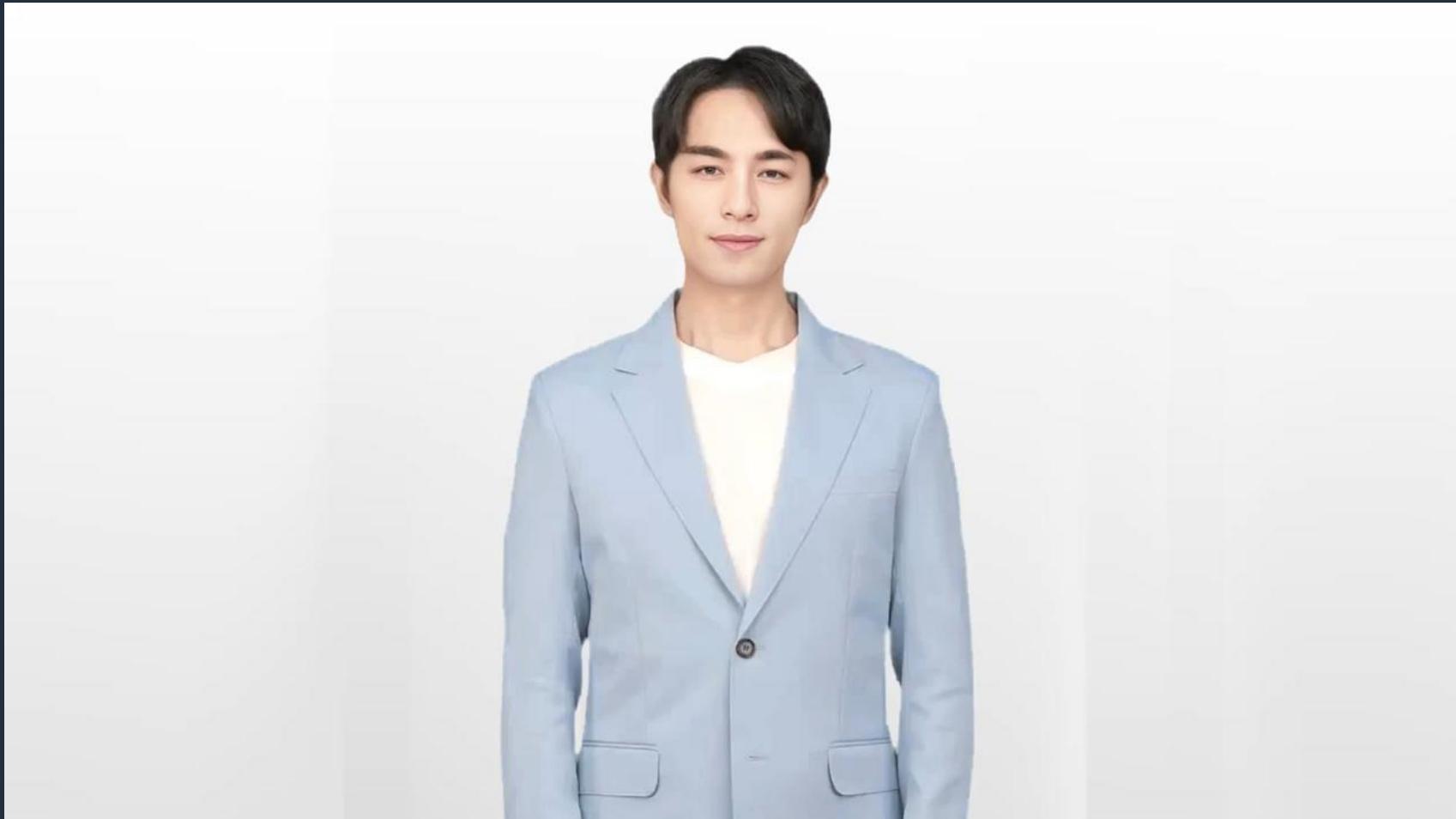


3 창의/창조

디지털 휴먼 생성 모델 추론 비용 절감 사례



- 입 모양 생성 모델을 AWS Inferentia (AWS 실리콘 칩 기반)를 이용해서 모델 추론 비용을 50% 이상 줄임



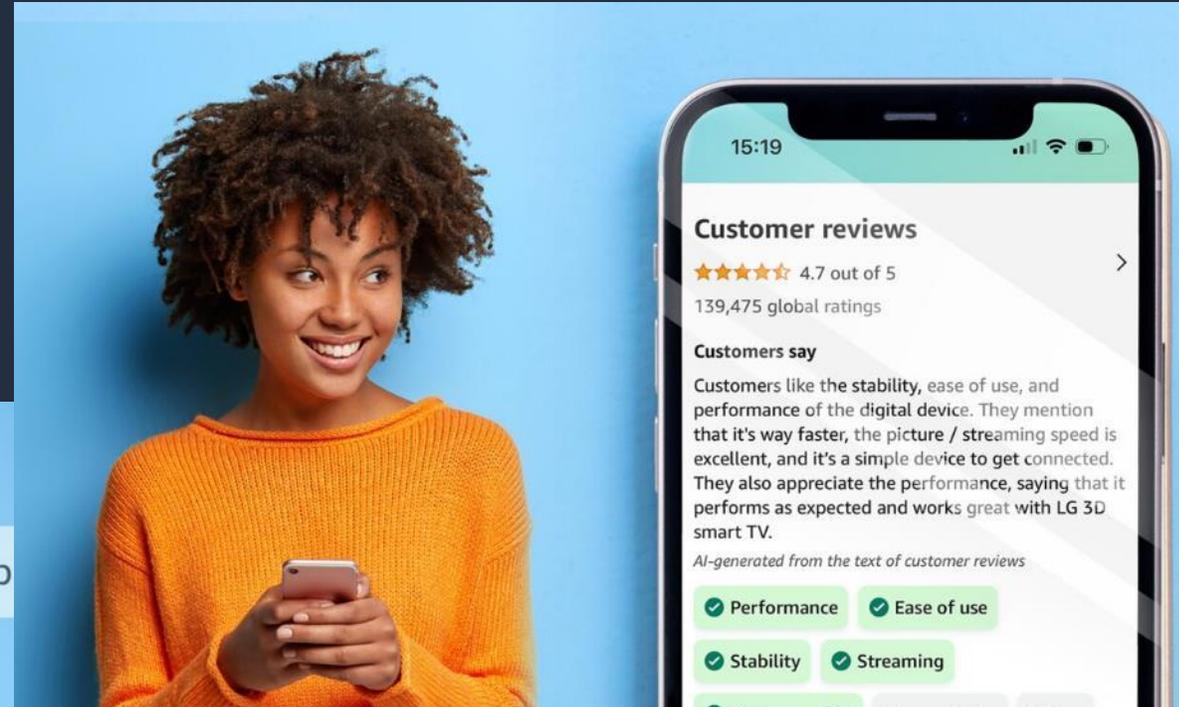
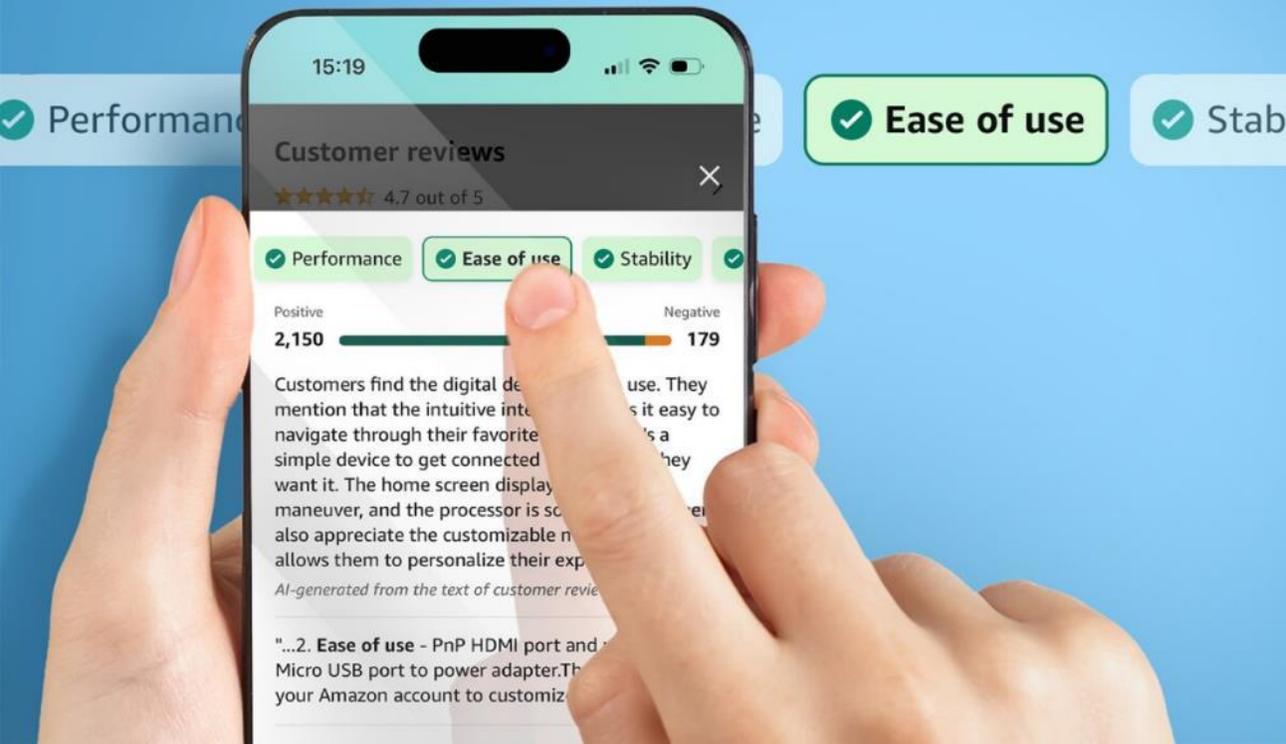
4 고객 경험 강화

생성형 AI 기반 리뷰 요약 서비스



New AI-generated customer review highlights are the latest in a string of review innovations.

고객수 1.3억명 / 리뷰수 1.5조개 / 초당 리뷰 45개



- 상품 상세 페이지 바로 위에 짧은 단락을 제공하여, 작성된 리뷰에서 자주 언급되는 제품 기능과 고객 심리를 강조하여 한눈에 판단할 수 있도록 도와줌
- 생성형 AI 리뷰 하이라이트는 검증된 구매에서 신뢰할 수 있는 리뷰 코퍼스만 (Corpus) 사용

생성형 AI 접근에 있어서 공통적으로 기대하는 부분



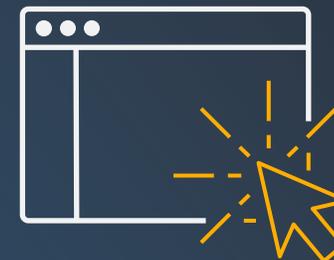
1 사용 편의 + 보안

Foundation Model을
선택하고, 보안성을
확보하면서
사용 최적화



2 비용 효율성

최적의 성능과 비용을
보장하기 위한
인프라 관리

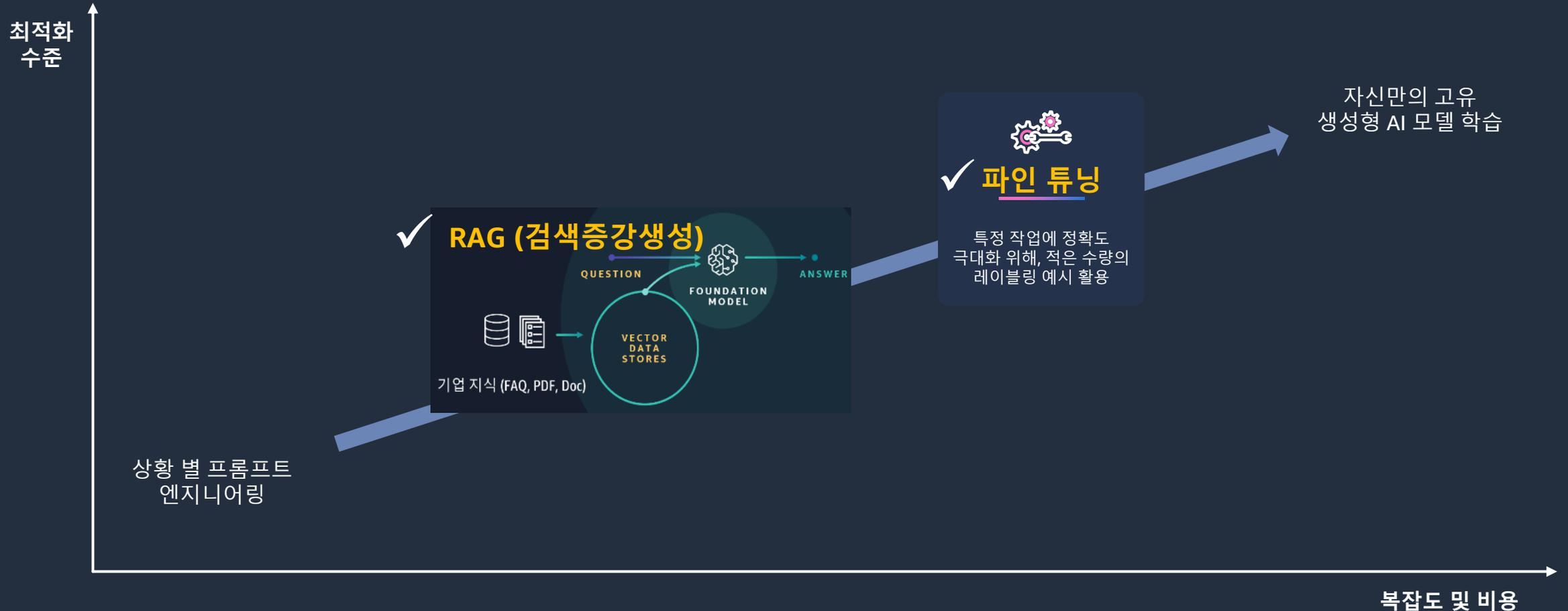


3 더욱 빠른 가치 제공

공통적인 비즈니스
작업들을 위한
생성형 AI 서비스 시작

맞춤형 최적화에 대한 요구

'우리에게 적합한', '우리 용어 이해하는' 생성형 AI 적용을 위해 (1) **RAG**, 또는 (2) **파인튜닝** 등을 필수 활용





생성형 AI 애플리케이션



생성형 AI 애플리케이션

Data Foundation

스토리지

거버넌스
& 컴플라이언스

데이터베이스,
분석,
& 데이터 레이크

데이터 통합

[예시] 더욱 개인화된 가상 에이전트

목적

맞춤형 여행 일정을 만들고자 하는 온라인 여행사

The data

고객 프로필

 지난 여행

 Web 기록

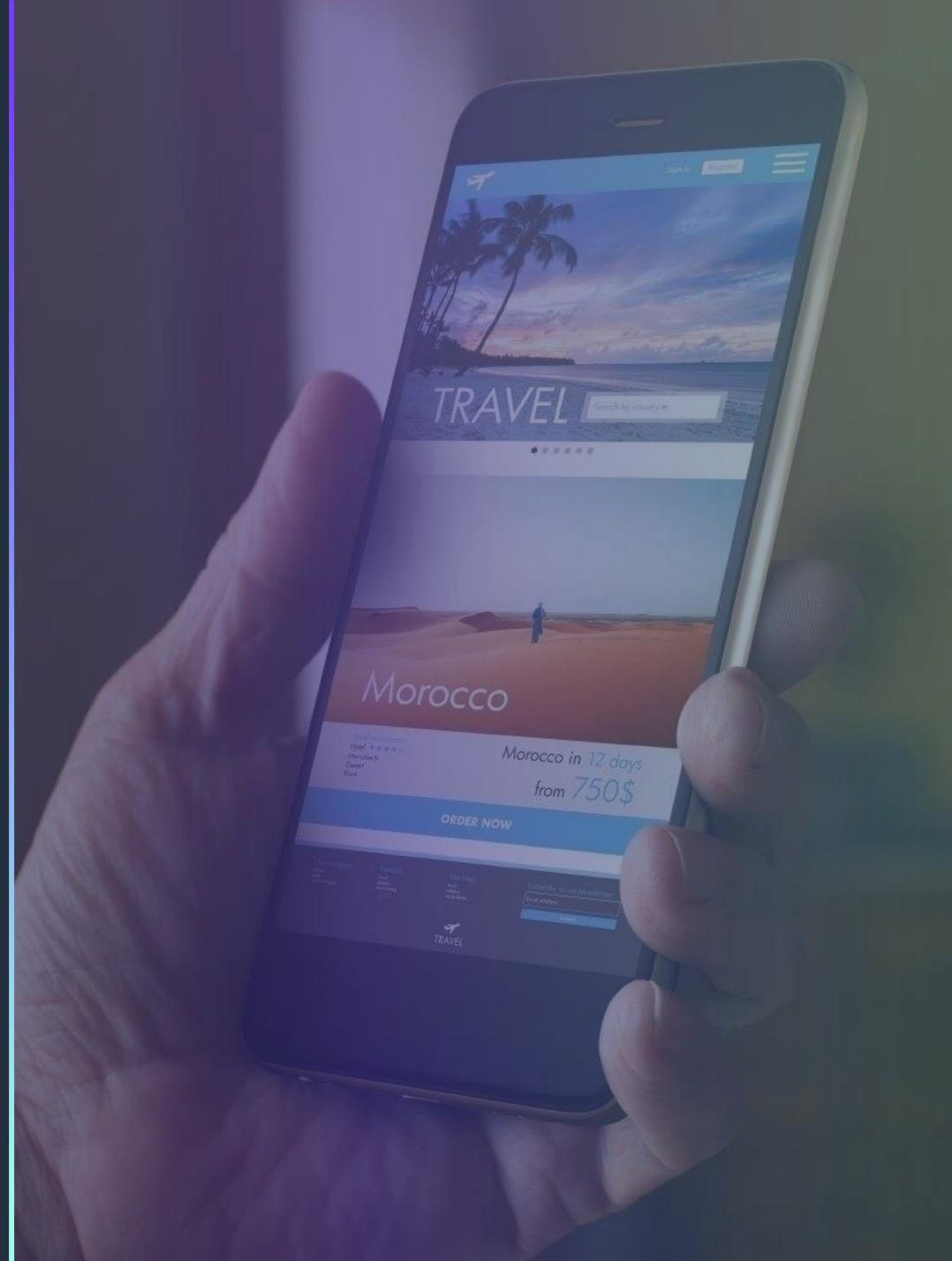
 여행 선호사항

기업 데이터

 항공 및 호텔 인벤토리

 프로모션

 유사한 여행자의 세부사항



생성형 AI의 미래는 어떻게 될까요?

에이전트

멀티모달

다양한 모델

AI 정책과 표준

AWS Summit New York 2024 신규 발표 내용 (2024.07)

기업이 실제 프로덕션 환경에 도입하는 데에 적용할 수 있는 Amazon Bedrock 관련한 새로운 기능을 소개하였습니다.

1. 생성형 AI
모델 최적화

2. 데이터 연결
성능 강화

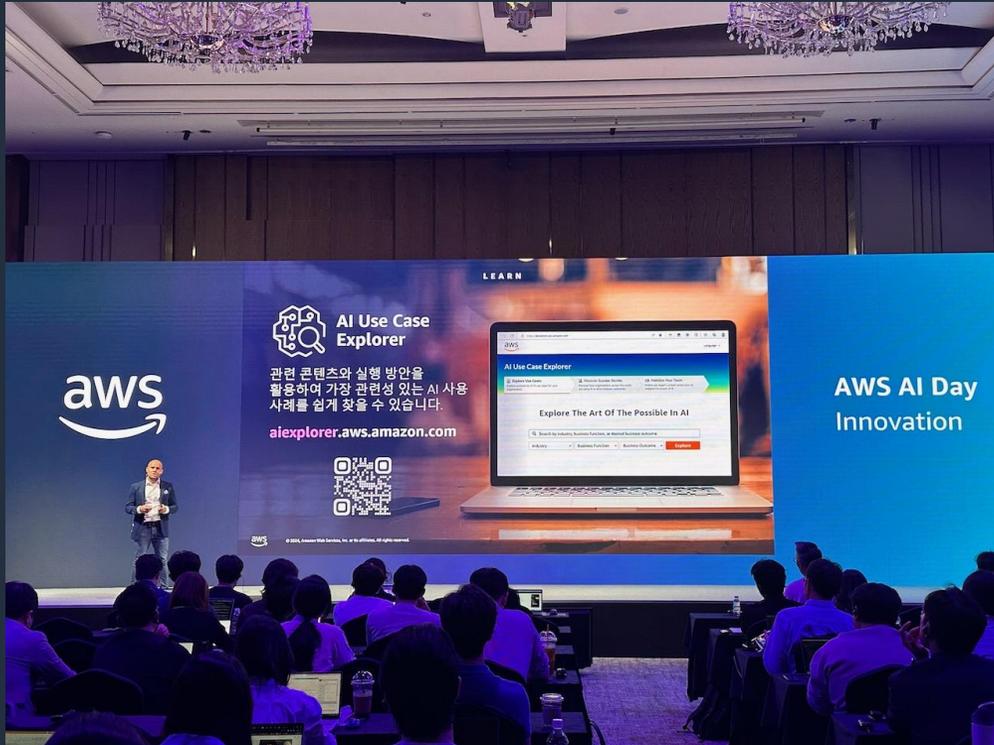
3. 책임감 있는
AI 기능 강화

4. 실행 능력
강화

Multi-modal Foundation Model



Bedrock 내의 Claude의 Multi-modal 기능



AI Claude 3.5 Sonnet v1 | On-demand
Change

해당 이미지에 대해 설명해줘



이 이미지는 AWS AI Day Innovation이라는 행사의 한 장면을 보여줍니다.

화면 중앙에는 큰 프레젠테이션 스크린이 있고, 그 위에 "AI Use Case Explorer"라는 AWS의 새로운 도구에 대한 정보가 표시되어 있습니다. 스크린에는 이 도구의 웹사이트 스크린샷과 QR 코드, 그리고 한국어로 된 설명이 포함되어 있습니다.

무대 왼쪽에는 발표자로 보이는 한 남성이 서 있습니다. 그는 정장 차림을 하고 있으며 청중을 향해 말하고 있는 것으로 보입니다.

청중석에는 많은 사람들이 앉아 있으며, 대부분이 노트북을 사용하고 있는 것으로 보입니다. 행사장의 조명은 보라색 계열로, 전체적으로 차분하면서도 미래지향적인 분위기를 자아내고 있습니다.

천장에는 화려한 샹들리에가 있어 고급스러운 행사장임을 알 수 있습니다.

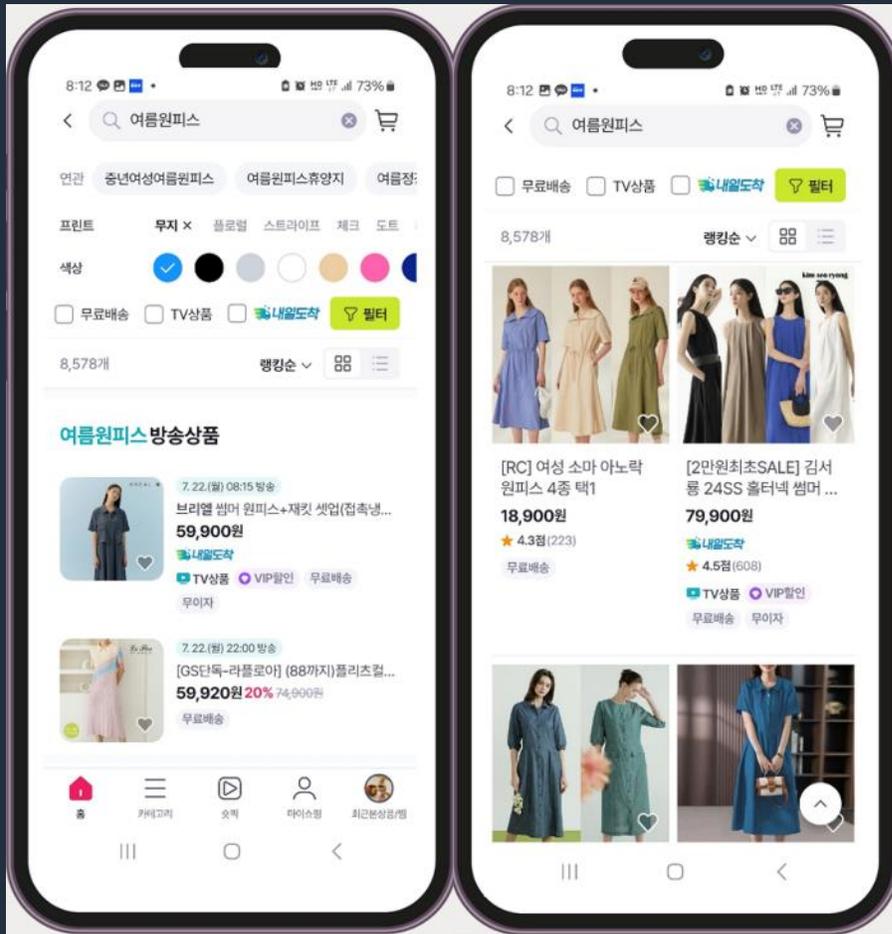
이 이미지는 AWS가 주최한 AI 관련 컨퍼런스나 세미나의 한 장면을 포착한 것으로 보이며, 최신 AI 기술과 그 활용 사례에 대해 소개하고 있는 것 같습니다.



멀티모달 기반 패션 검색 시스템 구현 사례



약 1억 개의 상품 DB에서, 패션 상품 데이터를 효과적으로 분류하고 검색 시스템에 활용하기 위해, Claude 3 Haiku 모델 도입



비용 효율성

Claude 3 Haiku 모델은, GS SHOP의 요구사항 (상품당 약 1200개의 토큰)에 부합하는 경제적인 솔루션



뛰어난 성능

이미지와 텍스트 정보를 활용하여 의상 속성을 정확하게 분류. 품질 검증 결과, 상위 모델과 비교해도 성능 차이가 없었으며, 빠른 응답 속도를 제공



다양한 태스크 적용 가능

Claude 3는 자연어 처리, 텍스트 생성, 요약, 번역뿐만 아니라 이미지 멀티모달까지 다양한 태스크에 활용 가능한 범용 모델



보안 강화

Amazon Bedrock은 AWS PrivateLink를 통해 VPC에 연결될 수 있어, 모든 데이터 전송이 AWS 내부 네트워크를 통해 이뤄짐. 외부 접근을 차단하여 시스템의 보안성을 크게 강화



손쉬운 통합

GS SHOP 업무 시스템이 AWS 환경에 있어, Bedrock 서비스와의 통합이 용이. 새로운 시스템 도입 시 통합 및 연동 작업의 복잡성을 크게 낮춤



AWS의 생성형 AI 관련 서비스와 인프라

[3] LLM 및 기타 FM을 활용하는 애플리케이션



Amazon Q Business



Amazon Q Developer



Amazon Q in QuickSight



Amazon Q in Connect



Amazon Q in AWS Supply Chain

[2] LLM 및 기타 FM을 사용하여 구축할 수 있는 도구



Amazon Bedrock

Guardrails | Agents | Studio | Customization Capabilities

Broadest selection of models

[1] Foundation model 학습과 추론을 위한 기반 인프라



GPUs



Trainium



Inferentia



SageMaker



UltraClusters



EFA



EC2 Capacity Blocks



Nitro



Neuron



FM Consumer



FM Tuner



FM Provider

생성형 AI 제언

WHY
왜 하는가?

생성형 AI 활용사례
정의 및 발굴

- AWS의 Digital Innovation Program Workshop (Amazon의 Working Backwards 방식의 발굴)

WHAT
무엇을 할 것인가?

생성형 AI 기술
검토 및 선택

- AWS의 다양한 Foundation Model 확인, 점검
- AWS 기반의 다양한 교육 프로그램 기반 학습

HOW
어떻게 할 것인가?

생성형 AI 프로젝트
진행 및 운영

- AWS Professional Service와 프로젝트 추진
- AWS Partner Network 협업 통한 프로젝트 진행



Thank you!

김선수

AI/ML 사업개발 담당

Amazon Web Services