A DAILY PERSPECTIVE ON AI

AP *News*

AI의 편의, 조건까지 읽기

여행 계획부터 업무 응답까지 AI가 돕는 범위가 넓어지고 있습니다. 오늘은 제공 지역, 차단 규칙의 경계, 고객별 자료 접근, 문맥 재사용과 중단 복구를 살펴봅니다. 편리하다는 설명 옆에 실제 사용 조건과 확인 근거를 함께 둡니다.



AI 제작 · 편집 일러스트 · 실제 제품 화면 아님

THE DAILY BRIEF · 5 STORIES

전날인 2026년 11월 16일 공식 발표 4건을 선정했습니다. 이전 호의 원문·같은 이슈를 제외한 다섯 번째 기사로 11월 13일 고객 사례를 골랐으며, 선정 범위는 최근 72시간에서 7일로 넓혔습니다. Google 기능 발표 1건과 AWS 기술 지침·고객 사례 4건입니다. AI 작성 및 원문 확인.



AI 제작 · 편집 일러스트 · 실제 제품 화면 아님

01 · AI NEWS

Ask Maps, 여행 계획과 예약 가능 시간을 함께 찾기

Google Blog · 발표 2026-11-16

Google이 일정 작성·저장 목록 검색을 넓히고, 일부 지역에서 예약 가능 시간 찾기를 추가했습니다.

Google은 11월 16일 Ask Maps에 날짜별 여행 일정과 저장 목록 검색 등의 업데이트를 발표했습니다. 일정의 각 장소를 지도에서 확인하고 공유하거나 데스크톱에서 Google Docs로 내보낼 수 있습니다. Android·iOS에서 순차 제공하며, 유럽 30개국으로 확대를 시작해 전체 제공 범위를 200개 이상의 국가·지역으로 넓힌다고 설명했습니다. 여행 일정과 목록 검색은 Ask Maps 제공 지역이 대상입니다. 경로의 사전 안내와 예약 가능 시간 찾기는 미국에서 먼저 출시합니다. 모든 기능이 한국·싱가포르에서 즉시 제공된다는 뜻은 아닙니다.

AP의 관점 · AP의 AI 활용 교육이나 Lia Select 싱가포르 고객 동선 안내에 참고한다면, 가상 일정의 추천 장소와 실제 영업·예약 조건을 대조하는 실험을 제안합니다. 일정 초안 작성과 실제 예약 확정을 별도 단계로 확인하고, 계정·지역별 제공 여부도 함께 기록할 수 있습니다.

공식 원문 읽기 ↗



AI 제작 · 편집 일러스트 · 실제 제품 화면 아님

02 · AI NEWS

AI 차단 규칙, 넓은 분야보다 거절할 행동을 명확히

AWS Blog · 발표 2026-11-16

AWS는 Denied Topics의 과도한 차단과 누락을 줄이기 위한 작성·평가 방법을 설명했습니다.

AWS는 11월 16일 Bedrock Guardrails의 Denied Topics 작성 지침을 공개했습니다. 이 기능은 키워드가 아닌 의미로 범주를 판단하므로, 금융처럼 넓은 분야 이름이 정상 문의까지 막을 수 있다고 설명합니다. 이름·정의·예시를 함께 읽는 구조를 고려해 거절할 행동과 경계 사례를 구체화하고, 통과해야 할 질문도 평가 목록에 넣도록 안내합니다. 같은 요청의 반복이나 주제 목록 변경에 따라 결과가 달라질 수 있어 한 번의 성공을 정확성 증명으로 보지 않습니다. 중요한 판단은 다른 정책과 검토 경로도 함께 설계해야 합니다.

AP의 관점 · AP의 AI 교육용 상담 예제를 만든다면, 개인정보가 없는 가상 질문으로 허용·거절·경계 사례를 나눠 점검하는 실험을 제안합니다. 규칙의 표현을 바꿀 때 이전에 통과한 정상 문의가 새로 막히지 않는지도 같은 목록으로 확인해 볼 수 있습니다.

공식 원문 읽기 ↗



AI 제작 · 편집 일러스트 · 실제 제품 화면 아님

03 · AI NEWS

고객별 답변 말투와 자료 접근 권한을 분리하기

AWS Blog · 발표 2026-11-16

Managed Knowledge Base의 검색 지침은 고객별 응답 형식을 바꾸지만 자료 접근 자체를 제한하지는 않습니다.

AWS는 11월 16일 Bedrock Managed Knowledge Base에서 고객별 검색 지침을 적용하는 방법을 소개했습니다. 지침은 페르소나·답변 형식·검색 방향·답변 정책으로 구성되며, 검토를 거쳐 활성화하고 고정된 버전을 요청에 지정할 수 있습니다. 그러나 지침만으로 검색 가능한 문서가 제한되지는 않습니다. 인증된 고객을 기준으로 서버가 지침과 메타데이터 필터를 선택하거나 지식베이스를 분리해야 합니다. 형식을 지정해도 결과의 형식 검증은 별도로 필요하며, 정확한 날짜·금액 계산은 앱에서 처리하도록 안내합니다.

AP의 관점 · AP의 고객사별 교육 자료나 Lia Select 안내 검색을 설계한다면, 같은 공개 질문을 서로 다른 가상 고객으로 조회하는 실험을 제안합니다. 말투가 맞는지와 다른 고객의 자료가 섞이지 않는지를 별도 기준으로 확인하고, 사용한 지침 버전도 결과와 함께 남길 수 있습니다.

공식 원문 읽기 ↗



AI 제작 · 편집 일러스트 · 실제 제품 화면 아님

04 · AI NEWS

반복 문맥은 재사용하되, 캐시 적중을 직접 확인하기

AWS Blog · 발표 2026-11-16

AWS는 공개 가중치 모델의 프롬프트 캐시를 자동·명시 방식으로 사용하는 예제를 공개했습니다.

AWS는 11월 16일 Bedrock의 공개 가중치 모델에 명시적 프롬프트 캐싱을 사용하는 방법을 소개했습니다. 캐시는 이전 요청과 일치하는 앞부분의 계산 상태를 재사용하며, 같은 답변을 복사하는 기능이 아닙니다. 자동 캐싱은 가능한 범위에서 동작하고, 명시 방식은 재사용할 구간의 끝을 지정합니다. 지원 모델·리전·최소 길이·보존 시간은 각각 확인해야 합니다. 캐시 읽기는 낮은 요율이 적용되지만 모델에 따라 쓰기가 일반 입력보다 비쌀 수 있어, 실제 읽기·쓰기 사용량으로 효과를 확인하도록 안내합니다.

AP의 관점 · AP의 교육 자료 검색이나 웹 서비스 비용 구조를 검토한다면, 고정 자료와 매번 달라지는 질문을 나눠 같은 과제의 사용량을 비교하는 설계를 제안합니다. 응답 품질·지연·전체 비용을 함께 기록하면 재사용이 도움이 되는 구간과 그렇지 않은 구간을 구분할 수 있습니다.

공식 원문 읽기 ↗



AI 제작 · 편집 일러스트 · 실제 제품 화면 아님

05 · AI NEWS

저렴한 음성 처리 자원, 중단 복구까지 함께 설계하기

AWS Blog · 발표 2026-11-13

AssemblyAI 사례는 Spot 자원의 비용 이점과 작업 큐를 통한 복구를 함께 설명합니다.

AWS는 11월 13일 AssemblyAI의 음성 인식 추론 운영 사례를 공개했습니다. ECS에서 Spot과 On-Demand GPU 자원을 함께 쓰고 여러 인스턴스 계열로 용량을 분산합니다. 중단 예고를 받으면 끝낼 수 없는 작업을 SQS 큐로 돌려 다른 작업자가 이어받으며, 재시도 한도를 넘은 작업은 별도 실패 큐에서 조사하도록 구성합니다. 소개된 70% 절감은 이 고객의 On-Demand 대비 추론 컴퓨팅 비용 결과이며 전체 서비스 비용의 보장 수치가 아닙니다. 음성 자료의 지역별 보관 조건도 따로 유지한 사례입니다.

AP의 관점 · AP의 콘텐츠 전사나 교육 영상 자막 처리를 설계한다면, 공개된 짧은 음성으로 중단 뒤 결과가 한 번만 완성되는지 확인하는 실험을 제안합니다. 저렴한 실행 자원뿐 아니라 재시도 한도·중복 결과·실패 조사 비용까지 포함해 운영 방식을 비교할 수 있습니다.

공식 원문 읽기 ↗

ONE THING TO TAKE AWAY

AP의 시선

새 기능의 사용 가능 지역, 답변 규칙과 자료 권한, 캐시와 실제 비용은 각각 확인할 조건입니다. 자동 처리가 중간에 멈췄을 때 결과를 어떻게 보존하고 이어갈지도 운영 품질에 포함됩니다. AP의 교육·콘텐츠·웹 서비스와 Lia Select 안내에 참고한다면, 작은 공개 과제 하나에 사용 조건·성공 근거·실패 후 처리 방법을 함께 적는 실험을 제안합니다. 편의 기능을 더하는 일과 그 결과를 믿을 수 있도록 확인하는 일을 같은 설계에 담아 볼 수 있습니다.