Deloitte.

AI Tokenomics 서비스

사용량 가시화부터 예산·권한·비용책임 관리까지 연결하는 AI 비용관리 서비스

Deloitte Korea, Forensic & Resilience | September 2026

폭증하는 AI 사용량, 감당하기 어려워지는 기업 비용

AI 활용이 확대될수록 비용의 발생 주체를 명확히 하고, 예산 초과와 사용 권한을 일관된 기준으로 관리해야 합니다

경영진의 질문	현장의 공백
어느 부서가 무엇에 쓰고 있는가?	**비용 책임자 연결 부족** 공급업체·제품별 청구와 계정이 분산
다음 달에 얼마가 필요한가?	**예산 전망·증액 근거 부족** 사용자·업무량·실행량 변화가 늦게 반영
누가 어떤 모델과 도구를 쓰는가?	**허용 범위·실행 내역 확인 곤란** 사용자·에이전트의 권한과 기록이 분산

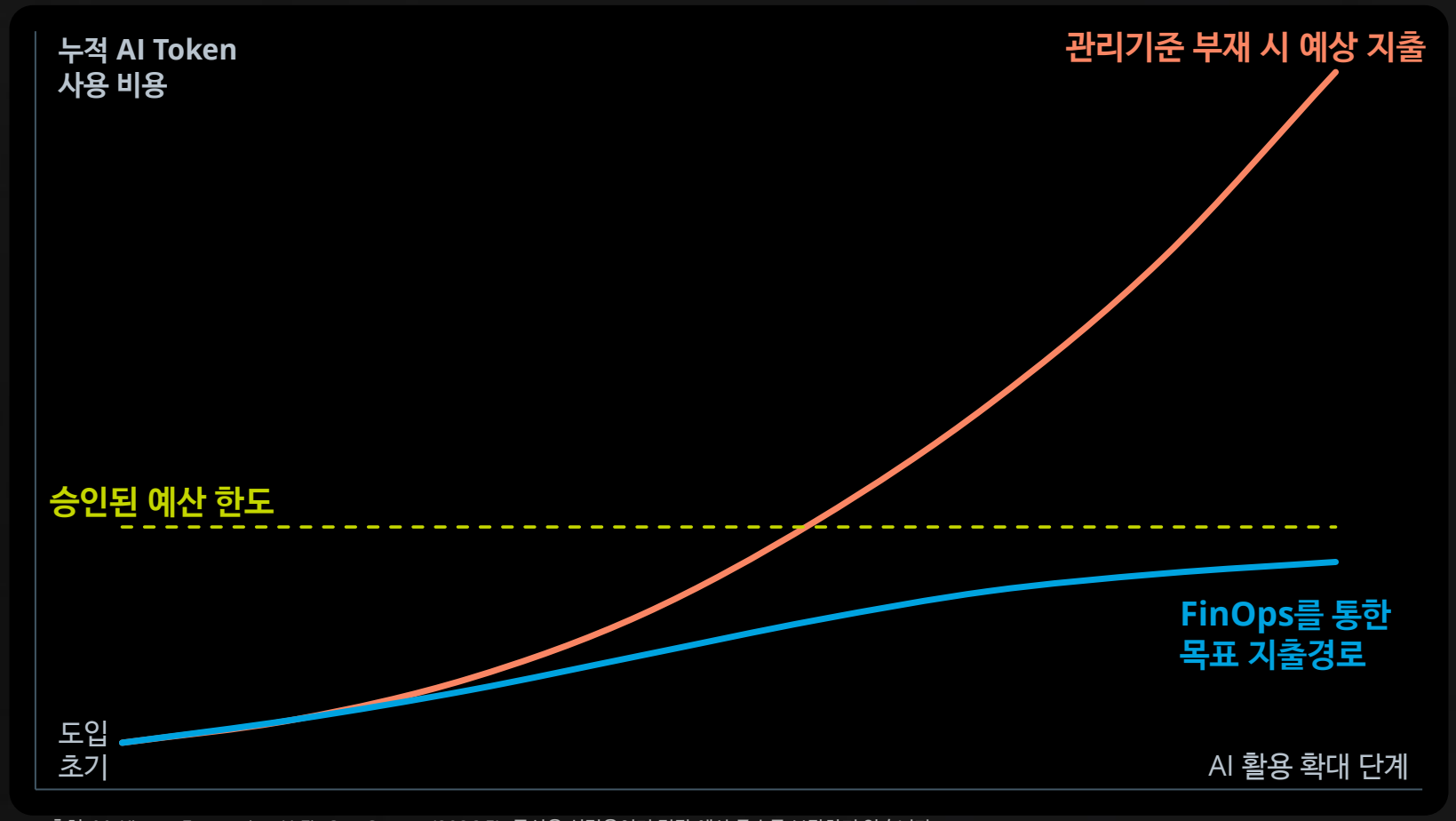
비용을 단순히 확인하는 것을 넘어,

예산과 권한에 따라 적시에 통제할 수 있는 관리체계가 필요합니다

© 2026. For information, contact Deloitte Anjin LLC

AI 활용 확대에 필요한 예산 통제

AI 업무와 에이전트 호출이 늘어날수록 지출이 빠르게 커질 수 있습니다. 사용량과 예산을 함께 관리해야 합니다



출처: McKinsey Enterprise AI FinOps Survey (2026.5). 곡선은 설명용이며 절감·예산 준수를 보장하지 않습니다

93%

설문 응답자가
AI 예산 대비 지출 초과를 경험

McKinsey, 2026년 5월 조사
5개 산업의 유효 응답자 75명

비용 발생주체 식별

어느 조직과 업무에서 비용이 발생하는가

예산 전망과 초과 대응

예산 소진전망 확인
조정안 승인 및 사용권한·한도 적용

© 2026. For information, contact Deloitte Anjin LLC

공급업체와 제품군에 따라 달라지는 AI 비용 구조

공급업체와 제품군마다 청구단위와 계약조건이 다르며, 동일한 API도 모델과 처리조건에 따라 비용구조가 달라집니다

01 공급업체

청구 단위의 대표 예시

OpenAI

좌석 · 크레딧 · 토큰

Anthropic

좌석 · 토큰

Google

좌석 · 쿼리 · 저장량

그 외 공급 경로

서로 다른 사용단위 간 직접 합산 불가

02 제품군

같은 공급업체도 과금 구조가 다름

구독 · ChatGPT Enterprise

계약별 좌석·사용량 조건
포함 범위와 추가 요금 확인

코딩 도구 · Codex

플랜·포함 사용량 확인
추가 사용·정산 조건 확인

API · OpenAI Platform

모델·사용량별 비용 산정
공개 요금표와 계약조건 확인

제품군별 계약범위와 청구기준의 개별 확인 필요

03 API 비용 구성과 계산 예시

항목별 사용량과 적용 단가를 확인하고 별도 도구 사용료를 합산

1 모델·입출력
모델과 입력·출력별 단가

2 처리 등급
처리 방식별 적용 조건

3 입력정보 길이
길이별 요율 적용 여부

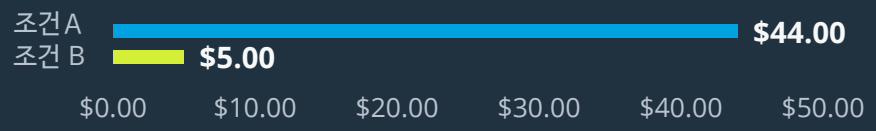
4 캐시 적용
캐시 사용량과 요율

5 지역 조건
지역별 적용 조건

6 도구 사용료
검색·저장·실행 등

비용 = 항목별 사용량 × 적용 단가의 합 + 별도 도구 사용료

동일 사용량의 조건별 비용 (USD)



동일 사용량 기준
비용 차이 예시
8.8배

조건 A: 우선 처리·긴 문맥·지역 처리 | 조건 B: 일괄 처리·짧은 문맥·지역 미지정

조직과 업무별로 비용을 배부하려면,

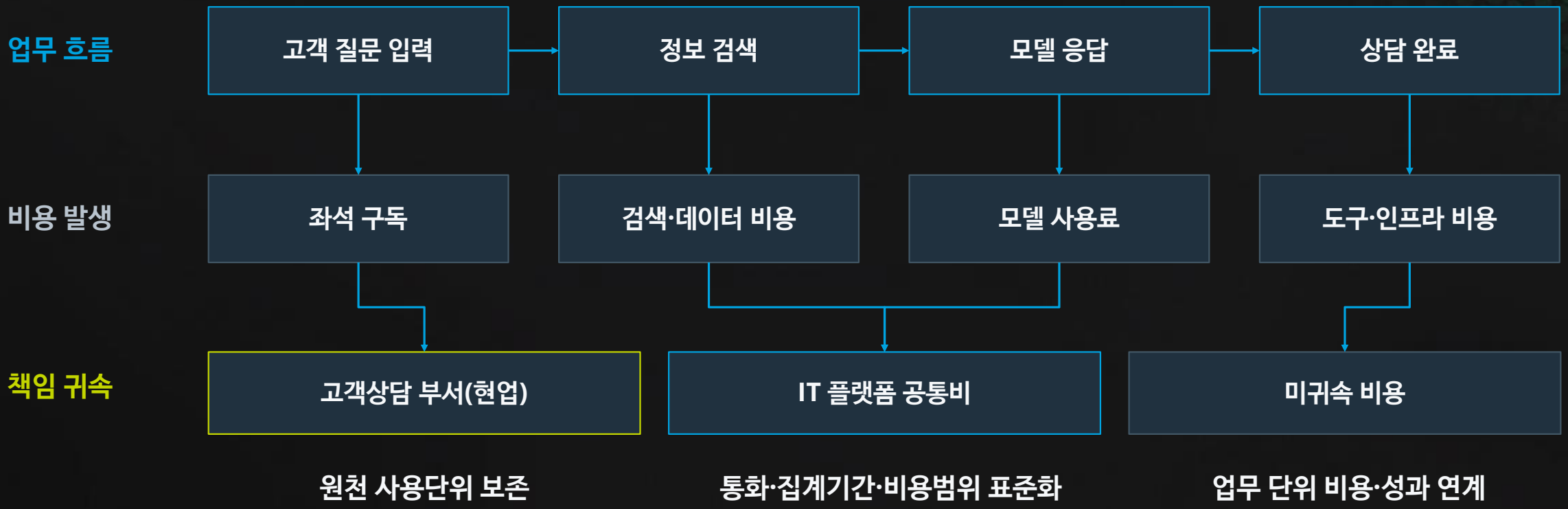
서로 다른 사용기록과 청구단위를 일관된 비용기준으로 표준화해야 합니다

청구 단위는 대표 예시이며 제품·계약별로 다릅니다. 비교 금액은 구조 설명용 가정으로, 실제 모델 단가나 할인율을 뜻하지 않습니다.

© 2026. For information, contact Deloitte Anjin LLC

하나의 AI 업무에서 발생하는 비용과 책임

고객상담 부서가 AI로 외부 고객의 질문에 답변하는 업무 예시이며 사용 비용을 현업 부담과 IT 공통비로 구분합니다



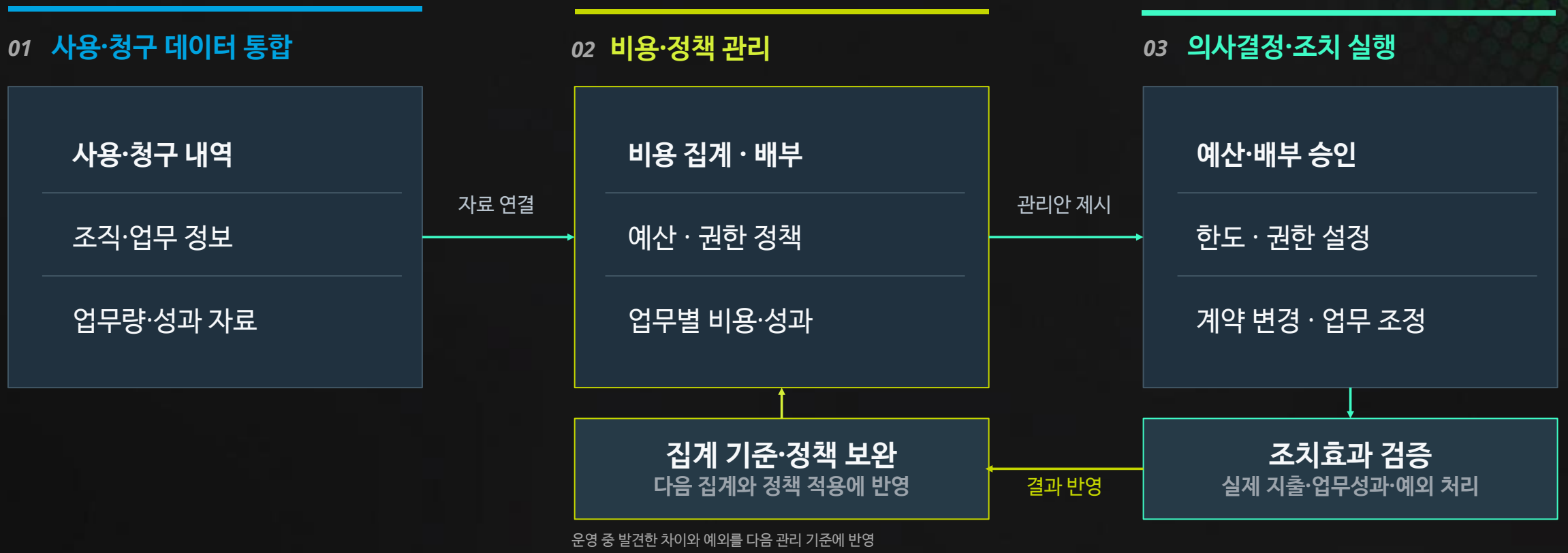
분산된 AI 비용을 기존 예산·회계체계와 연결하여 조직과 업무별 책임비용으로 전환합니다

고객 문의 응대 업무의 설명 예시. 별도로 발생한 비용만 포함하며 중복 계상을 제외합니다.

© 2026. For information, contact Deloitte Anjin LLC

AI Tokenomics 서비스 개요

사용·청구 데이터를 통합하여 비용과 정책을 관리하고, 승인된 조치의 실행결과를 다음 예산·권한·관리기준에 반영합니다



비용 확인에서 끝내지 않고, 승인·적용과 결과 확인까지 연결합니다

© 2026. For information, contact Deloitte Anjin LLC

서비스 수행 영역과 주요 산출물

고객의 관리 우선순위에 따라 필요한 서비스 영역을 선택하고, 실제 비용관리와 의사결정에 활용할 수 있는 운영기준과 결과물을 제공합니다

서비스 영역	주요 수행 내용	핵심 결과물	지원 의사결정
M1 비용 가시화·책임 귀속	사용·청구 연결 비용 집계·배부	**조직·업무별 비용 현황** **집계·배부 기준**	비용 책임 확인
M2 예산·사용 통제	예산·권한 정책 설계 선택 기능 설정·검증	**예산·권한 설정 내역** **시범 적용 결과**	예산·한도 조정
M3 계약·사용 효율 분석	미사용·중복 계약 검토 업무별 비용 비교	**계약 조정안** **업무별 비용 비교표**	계약 갱신·예산 검토
M4 운영체계 정립·내재화	운영 역할·절차 정리 교육·초기 운영 지원	**운영 매뉴얼·정기 보고** **인수 기준·확대 계획**	운영 인수·적용 확대

© 2026. For information, contact Deloitte Anjin LLC

AI Tokenomics 플랫폼 화면 예시 (1/2)

비용 현황과 예산·사용 통제, 계약·업무별 비용 분석을 한곳에서 연결하는 통합 관리 플랫폼을 제공합니다

한도 설정·예산 배분 화면 (상단)



화면 구성과 서비스 모듈 연결

1 비용 현황 요약

M1 비용 가시화·책임 귀속

- 전사 예산·배분 비율·이번 달 사용액·점검 대상 요약
- 토큰·비용 단위 전환

2 조직별 예산 배분

M2 예산·사용 통제

- 전사 › 본부 › 팀 › 개인 순으로 한도 배분
- 소진율 80% 기준선 표시, 초과 대상은 색상으로 구분

3 한도 조정

M2 예산·사용 통제

- 이번 달 한도·상위 미배분·사용 대비 소진율 확인 후 조정
- 조정 내용은 변경 요청으로 제출

4 한도 초과 시 동작

M2 예산·사용 통제

- 1단 경고·2단 한도 예외 승인·3단 소진 차단 중 선택
- 소진 차단으로 설정한 한도만 실행 차단

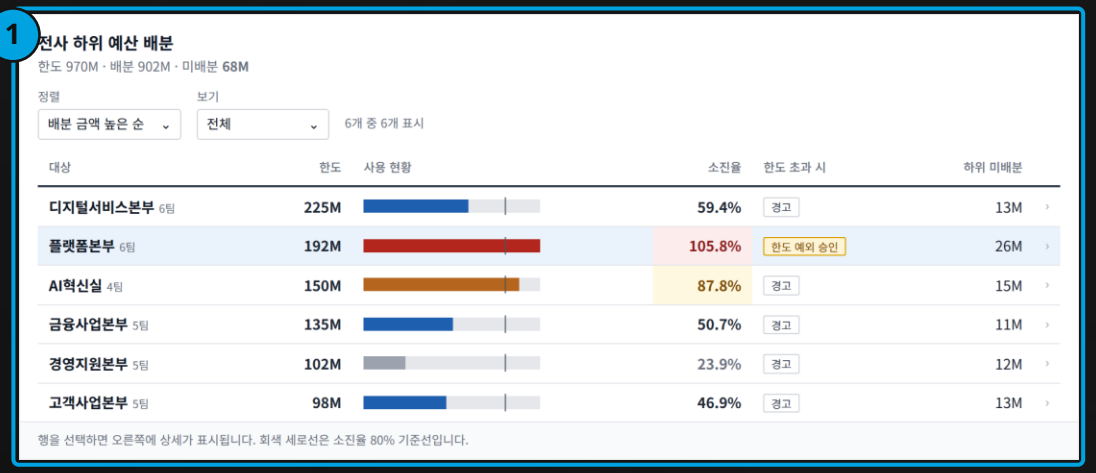
조직별 비용 현황을 확인하고 같은 화면에서 한도와 초과 시 동작을 설정합니다

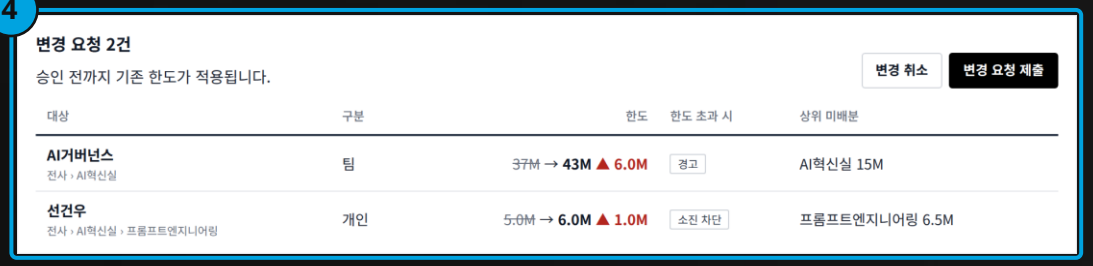
© 2026. For information, contact Deloitte Anjin LLC

AI Tokenomics 플랫폼 화면 예시 (2/2)

소진율 점검과 한도 예외 승인, 변경 요청 처리를 같은 화면에서 이어서 진행합니다

한도 설정·예산 배분 화면 (하단 구간)







1 하위 예산 배분 목록 — M1·M2

- 한도·사용 현황·소진율·미배분 확인
- 한도 초과 시 정책 표시, 80% 기준선

2 승인 대기 — M2

- 요청 한도·사유를 보고 승인/반려
- 일시적 증가는 기간을 한정해 예외 승인

3 남는 예산 회수 — M3

- 3개월 연속 소진율 21% → 한도 회수
- 회수한 예산은 우선 업무에 재배분

4 변경 요청 — M2

- 조정한 한도는 변경 요청으로 제출
- 승인 전까지 기존 한도 적용

M1 비용 가시화·책임 귀속 / M2 예산·사용 통제 / M3 계약·사용 효율 분석

한도 초과와 예외 요청을 승인·반려로 처리하고 남는 예산은 회수해 재배분합니다

© 2026. For information, contact Deloitte Anjin LLC

비용 가시화 및 예산 통제를 구현하는 게이트웨이 기능

M1은 사용량·비용을 집계하고, M2는 사용 통제와 예산 관리를 적용합니다. 고객 요구에 따라 기능을 조합합니다.

호출 주체

- 사용자
- 에이전트
- 코딩 도구
- 자동화 작업

게이트웨이 기반 3대 관리 기능

01 사용량·비용 집계

사용자와 업무별 비용 현황

공급업체·조직·업무별 집계
적용 단가로 비용 산정
청구 내역과 집계 결과 비교

02 사용 통제

허용 범위와 사용 한도

사용자·조직·업무별 한도
경고·제한·승인 연계
허용 모델·도구 설정

03 예산 관리

예산 대비 사용 현황

조직·프로젝트·업무별 배분
소진율과 예상 비용 확인
예산 초과 대응 관리

공급업체

- OpenAI
- Anthropic
- Google
- 그 외 공급 경로

관리목적별 적용모델 예시

기본 사용 계측을 바탕으로 필요한 기능 선택

고객 유형	집계	통제	예산	주요 요구
A 사용 현황 파악(비용 가시화형)	●	–	–	공급업체·조직별 사용량과 비용 확인
B 과다 사용 방지(사용 통제형)	●	●	–	사용자·조직별 사용 한도 적용
C 예산 관리(예산 관리형)	●	–	●	예산 배분·소진 전망·초과 알림
D 통합 관리(통합 운영형)	●	●	●	비용 현황과 예산·사용 정책 통합 운영

필요한 기능부터 도입하고, 관리 범위에 맞춰 같은 게이트웨이에서 확장합니다

직접 통제는 게이트웨이 연계 경로에 적용합니다. 구독형 서비스 등은 제공되는 사용·청구 데이터로 집계합니다.

© 2026. For information, contact Deloitte Anjin LLC

청구비용 검증과 조직·업무별 책임비용 배부

사용기록과 청구내역을 대사하여 비용의 정확성을 검증하고, 합의된 배부기준에 따라 조직과 업무별 책임비용을 산정합니다



직접 귀속		공통비 배부		미귀속		관리대상 비용
700만원	+	200만원	+	100만원	=	1,000만원
추적성		**일관성**				**책임성**
원자료·계약·기간 연결		통화·단가·비용기준 통일				배부 근거·미배부 사유 기록

설명용 가정. 동일한 관리비용 기준이며 청구기준과 선납 배분 후 비용은 구분합니다.

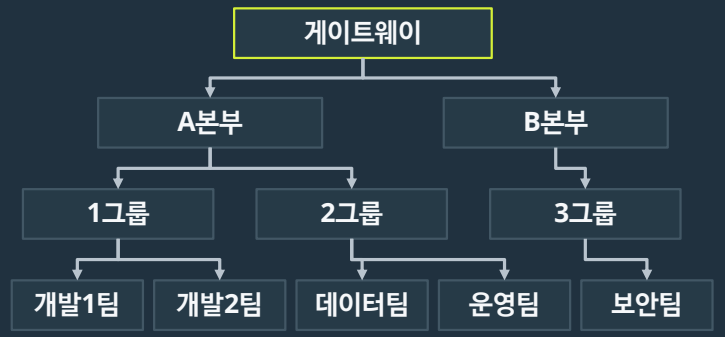
© 2026. For information, contact Deloitte Anjin LLC

조직과 업무에 맞춘 세 가지 예산 관리 방식

예산 책임구조와 업무 수행방식에 따라 조직 중심, 업무 중심 또는 혼합형 관리방식을 선택하고, 경고·승인·제한정책을 함께 적용합니다

01 조직 책임형

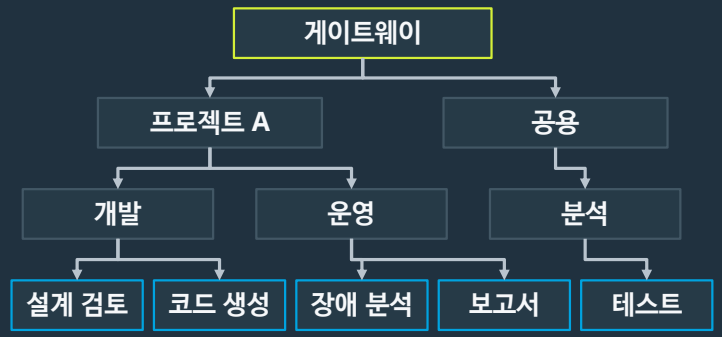
조직별 예산 한도 설정



도입 목적	부서 예산·회계 책임 관리
예산 기준	본부·그룹·팀별 한도
관리 지표	부서별 지출과 잔여 예산
운영 조건	업무별 분석은 업무정보 필요

02 업무·프로젝트 책임형

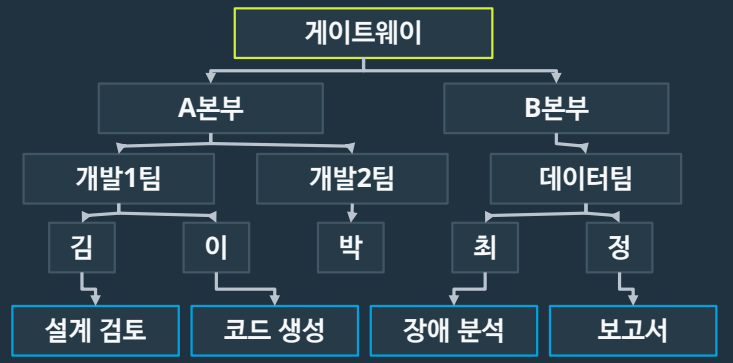
프로젝트·업무별 예산 배분



도입 목적	프로젝트·업무별 정산
예산 기준	프로젝트·업무별 배분
관리 지표	업무별 비용과 예산 소진
운영 조건	부서 승인권자와 연결 필요

03 조직·업무 연계형

조직 한도 안에서 업무별 배분



도입 목적	부서 예산과 업무 비용 관리
예산 기준	조직 한도와 업무별 배분
관리 지표	조직·업무 비용 함께 확인
운영 조건	연결정보 갱신·중복 합산 방지

정책 예시 : 80% 소진 시 알림 | 초과 전망 시 조정 승인 | 한도 도달 시 제한 또는 승인된 예외 적용

80%는 설명용 설정값입니다. 금액 예산과 요청 한도의 변환·갱신 기준, 복수 한도의 적용 우선순위는 별도로 정합니다.

예산 관리 구조를 선택하고, 경고·승인·제한 기준까지 함께 설정합니다

© 2026. For information, contact Deloitte Anjin LLC

사용자·AI 에이전트의 권한과 사용 한도

사용자와 에이전트의 신원을 업무에 연결하고, 허용된 모델·도구와 사용 한도를 적용하며 처리 이력을 남깁니다

사용자와 업무를 기준으로 권한과 예산을 관리하는 구조

에이전트 통제 계층에서 정책을 관리하고, 게이트웨이 연계 경로에 적용합니다.

01 사용주체 식별

기존 SSO·IdP 인증체계 연계

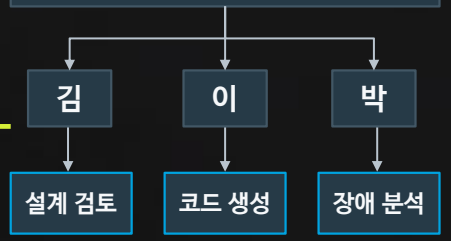
사용자·에이전트 식별

실행 주체와 책임자 연결
신뢰할 수 있는 식별정보 확인

식별정보가 있는 요청은
사용자별 비용 귀속

02 조직·업무 귀속

조직·예산 책임



업무별 권한과 예산 설정
조직 변경 시 연결정보 갱신

03 권한·예산정책 적용

허용 모델
업무별 모델과 등급

허용 도구·작업
에이전트 작업과 위임 범위

예산·사용 한도
사용량·호출 한도와 정책 조치

04 정책 집행·감사추적

게이트웨이 요청 확인
허용·차단·경고·승인 연계
관리 경로에서 정책 적용
외부 경로는 직접 통제 범위 구분

감사 기록
요청 주체·적용 정책·처리 결과
관리 경로의 사용량과 비용 기록

비용 관리와 연결 사용자·업무별 집계와 청구 비교에 활용하며, 에이전트는 고유 신원과 위임 범위에 따라 권한을 구분합니다.

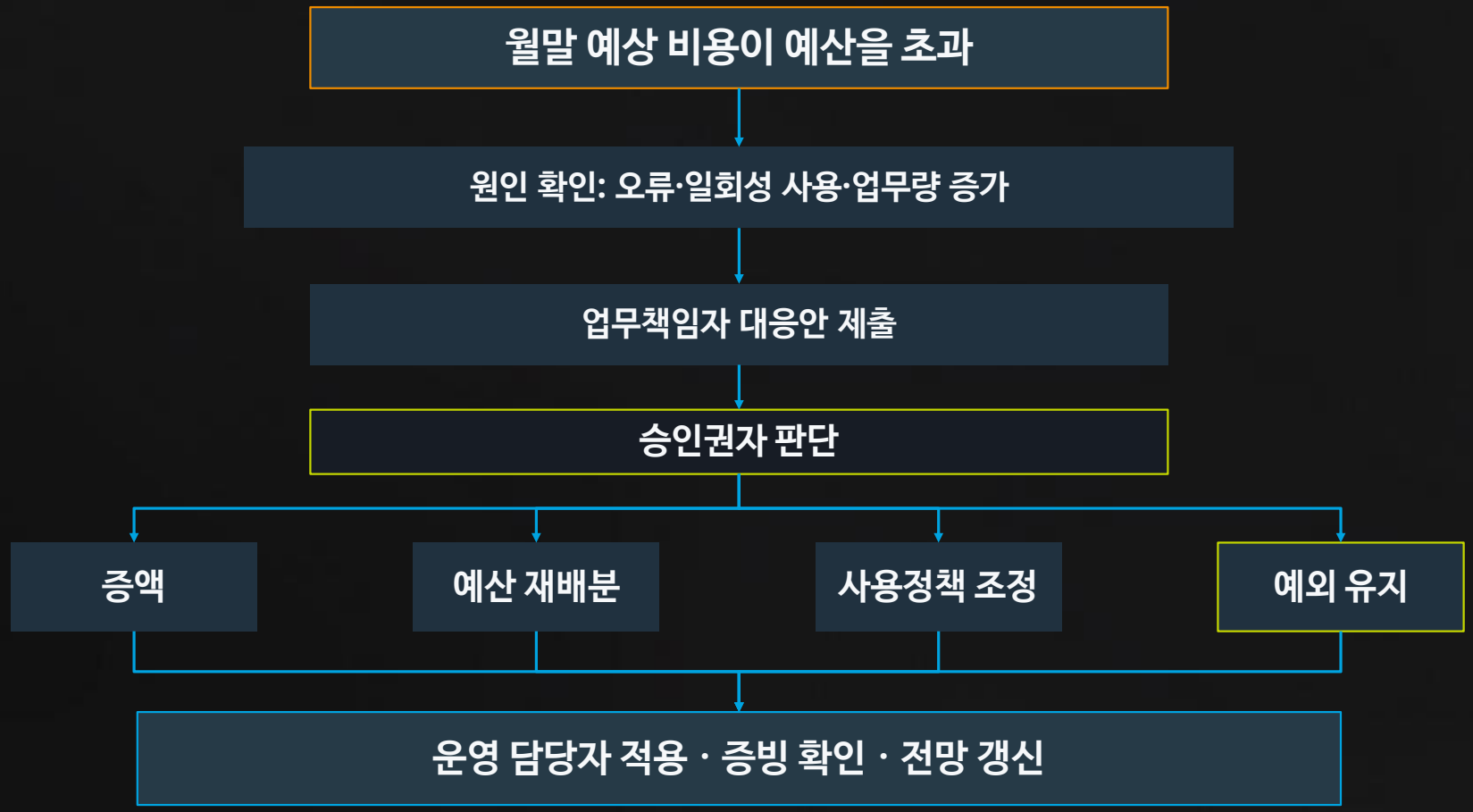
사용자와 에이전트의 사용 범위를 업무별 권한과 예산 정책에 연결합니다

권한 정책과 지원 기능은 적용 환경에서 검증합니다. 식별정보가 없는 요청은 미식별로 관리하고 직접 통제 범위를 명시합니다.

© 2026. For information, contact Deloitte Anjin LLC

예산 초과와 예외 요청의 처리 절차

예산 소진전망과 업무 중요도를 기준으로 대응방안을 검토하고, 승인·실행·사후확인까지 일관된 절차로 관리합니다



의사결정별

필수 관리기록

- 사유와 영향
- 승인자와 실행자
- 적용기간과 만료일
- 조치 결과와 증빙

대시보드 경고와 기술적 사용 제한은 별도 기능입니다. 데이터 갱신 시각과 실제 집행 수단을 확인합니다.

© 2026. For information, contact Deloitte Anjin LLC

AI 비용 효율화를 위한 4대 관리조치

사용·계약 현황에서 비용 누수요인을 식별하고, 계정·계약·사용정책을 조정하여 불필요한 지출을 체계적으로 개선합니다

식별된 비용 이슈	개선 조치	기대 효과
미사용·저활용 유료계정	실사용자 중심으로 계정 재배치 갱신 시 라이선스 구매수량 재산정	신규 구매수요 절감 갱신 후 구독비용 최적화
기능이 중복되는 AI 서비스	업무별 필수기능과 활용도를 비교하여 계약 통합 서비스 전환비용과 해지조건을 함께 검토	중복 구독비용 절감
실제 수요와 불일치하는 사용약정	최근 사용추이와 향후 수요전망에 기반하여 차기 약정량 재산정	미활용 약정비용 최소화
비정상적·일시적 사용량 급증	오류 및 실제 업무증가 여부를 구분하여 원인 분석 불필요 사용은 제한하고 필요 사용은 예외 승인	비용 누수를 방지하면서 핵심업무의 연속성 유지

비용 누수는 최소화하고, 경영적 우선순위가 높은 업무에 AI 예산을 재배분합니다

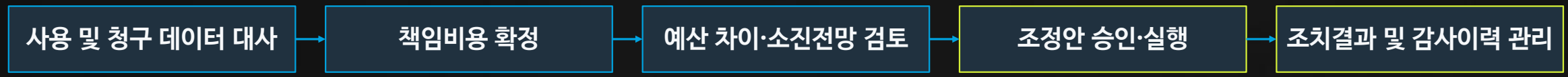
사용량이 줄어도 청구액이 바로 줄지 않을 수 있습니다. 계약 변경 시점과 실제 청구액을 함께 확인합니다.

© 2026. For information, contact Deloitte Anjin LLC

AI 비용관리를 위한 역할·책임 및 운영체계

고객 조직의 역할과 의사결정 권한을 명확히 하고, 표준화된 운영절차를 통해 비용·예산·정책조치를 지속적으로 관리합니다

주요 의사결정 사안	검토·조정 책임	의사결정 권한	실행 책임
비용 마감·배부	비용관리 담당	재무 책임자	비용관리·회계
프로젝트 증액	업무책임자·비용관리	예산 승인권자	예산·플랫폼
계정·사용정책 변경	업무책임자·플랫폼	정책 승인권자	플랫폼 담당
계약 갱신	구매·업무책임자	구매 전결권자	구매 담당



수시 이상 징후·예외 만료 확인

분기 업무 경제성과 계약·예산 검토

Deloitte 수행범위 : 관리기준 및 운영모델 설계 | 적용기능 구체화 | 초기 운영 정착 | 고객 자립운영 전환 검증

최종 예산·계약 승인권은 고객에게 있습니다.

© 2026. For information, contact Deloitte Anjin LLC

진단부터 시범 적용과 운영 정착까지

우선순위가 높은 공급경로와 업무를 대상으로 비용집계·예산·권한정책을 검증한 후, 검증결과에 따라 적용범위를 단계적으로 확대합니다

01 진단

주요 활동

계약·데이터·책임
현황 확인

고객 참여

자료 제공
담당자 인터뷰

관리대상·우선과제 합의

02 설계

주요 활동

원장·배부·통제
운영 기준 정의

고객 참여

배부 · 예산 · 성과
기준 협의

기준서·적용범위 승인

03 시범 적용

주요 활동

선택 기능 설정
집계·통제 검증

고객 참여

정책 실행
결과 검증

집계·정책 적용 결과 검증

04 운영 정착·확장

주요 활동

운영 인수
이행과제 정리

고객 참여

운영 책임자
지정 · 인수

담당자·일정·확대순서 확정

우선 관리대상과 검증기준을 합의한 후 단계적으로 적용범위를 확대합니다

수행기간은 데이터 접근성과 연계·적용 범위에 따라 산정합니다.

© 2026. For information, contact Deloitte Anjin LLC

AI 비용을 투명하게 파악하고, 책임 있게 관리합니다

전사 AI 사용량과 비용을 투명하게 파악하고, 조직과 업무의 우선순위에 따라 예산·권한·사용범위를 체계적으로 관리합니다

비용 가시화

누가 어떤 업무에서
얼마의 비용을
발생시키는가

예산 배분 · 관리

어느 조직과 업무에
얼마의 예산을
배분할 것인가

사용권한 통제

누가 어떤 모델과 도구를
어느 범위까지
사용할 수 있는가

초기 적용을 위한 핵심 결정

우선 관리할 AI 서비스·업무·조직과 적용범위를 함께 정의합니다

초기 논의를 위한 준비 정보 : 현재 사용 중인 AI 서비스 및 계약 현황 · 최근 청구내역과 사용량 데이터 · 우선적으로 해결할 비용·예산·권한 관리과제

© 2026. For information, contact Deloitte Anjin LLC

AI 사용을 OP 혁신으로 잇는 3단계 추진 방향

AI 사용을 한 관문에 모아 업무별 성과를 측정하고, 우수 방식을 표준으로 확산해 위임 범위를 넓힙니다

	STEP 1	STEP 2	STEP 3 · 궁극 목표
경영 목표	업무 단위의 비용 성과 측정 (Unit Economics)	상위 Agent 기반 업무 상향 표준화	AI Autonomous 확산 · OP 혁신
경영 과제	AI 비용 사용처·성과 파악	팀 간 우수 활용 방식 확산	AI 업무 위임 범위 설정
수행 내용	• 에이전트·업무별 AI 비용 집계 • 처리 결과와 연결해 ROI 측정 • 미사용 계정·고비용 모델 조정	• 팀별 비용·시간·품질 비교 • 상위 에이전트 중 우수 방식 선별 • 전사 공통 에이전트로 확산	• 실행·승인 이력으로 위임 후보 선별 • 책임자 승인 후 자동 처리 확대 • 기한·중단·회수 조건으로 통제
산출물 화면 예시	**업무별 원가·성과표** 계약서 검토 — 처리 시간 장애 요약 — 처리 건수 회의록 정리 — 연결 전	**계약서 검토·팀별 비교** 법무팀 영업팀 재무팀 → 공통 에이전트 계약서 검토	**업무별 위임 단계** 장애 요약 게시 — 자동 처리 주간 품질 보고 — 사후 확인 전사 채널 공지 — 매번 승인
	↑ 비용·결과 기록	↑ 업무 라벨·사용량	↑ 승인·반려 이력

기반

임직원 에이전트 → **AI 게이트웨이 · 연결된 AI 호출이 모두 거쳐 가는 관문**: 한도 통제 | 사용 기록 | 업무 라벨 | 에이전트 카탈로그 | 승인·위임 | 감사 기록 → LLM·AI 앱 사내 시스템

비용·성과 측정부터 시작하고, 검증된 업무 순서로 표준화와 위임 범위를 넓힙니다

ROI는 결과 시스템이 연결된 업무에서 측정하고, 팀 비교는 개인 평가에 쓰지 않습니다. 화면은 예시이며 실측값이 아닙니다.

© 2026. For information, contact Deloitte Anjin LLC

Deloitte.

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited (DTTL), its global network of member firms, and their related entities (collectively, the "Deloitte organization"). DTTL (also referred to as "Deloitte Global") and each of its member firms and related entities are legally separate and independent entities, which cannot obligate or bind each other in respect of third parties. DTTL and each DTTL member firm and related entity is liable only for its own acts and omissions, and not those of each other. DTTL does not provide services to clients. Please see www.deloitte.com/about to learn more.

Deloitte Asia Pacific Limited is a company limited by guarantee and a member firm of DTTL. Members of Deloitte Asia Pacific Limited and their related entities, each of which is a separate and independent legal entity, provide services from more than 100 cities across the region, including Auckland, Bangkok, Beijing, Bengaluru, Hanoi, Hong Kong, Jakarta, Kuala Lumpur, Manila, Melbourne, Mumbai, New Delhi, Osaka, Seoul, Shanghai, Singapore, Sydney, Taipei and Tokyo.

This communication contains general information only, and none of Deloitte Touche Tohmatsu Limited (DTTL), its global network of member firms or their related entities (collectively, the "Deloitte organization") is, by means of this communication, rendering professional advice or services. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser.

No representations, warranties or undertakings (express or implied) are given as to the accuracy or completeness of the information in this communication, and none of DTTL, its member firms, related entities, employees or agents shall be liable or responsible for any loss or damage whatsoever arising directly or indirectly in connection with any person relying on this communication. DTTL and each of its member firms, and their related entities, are legally separate and independent entities.

© 2026. For information, contact Deloitte Anjin LLC

MAKING AN IMPACT THAT MATTERS since 1845