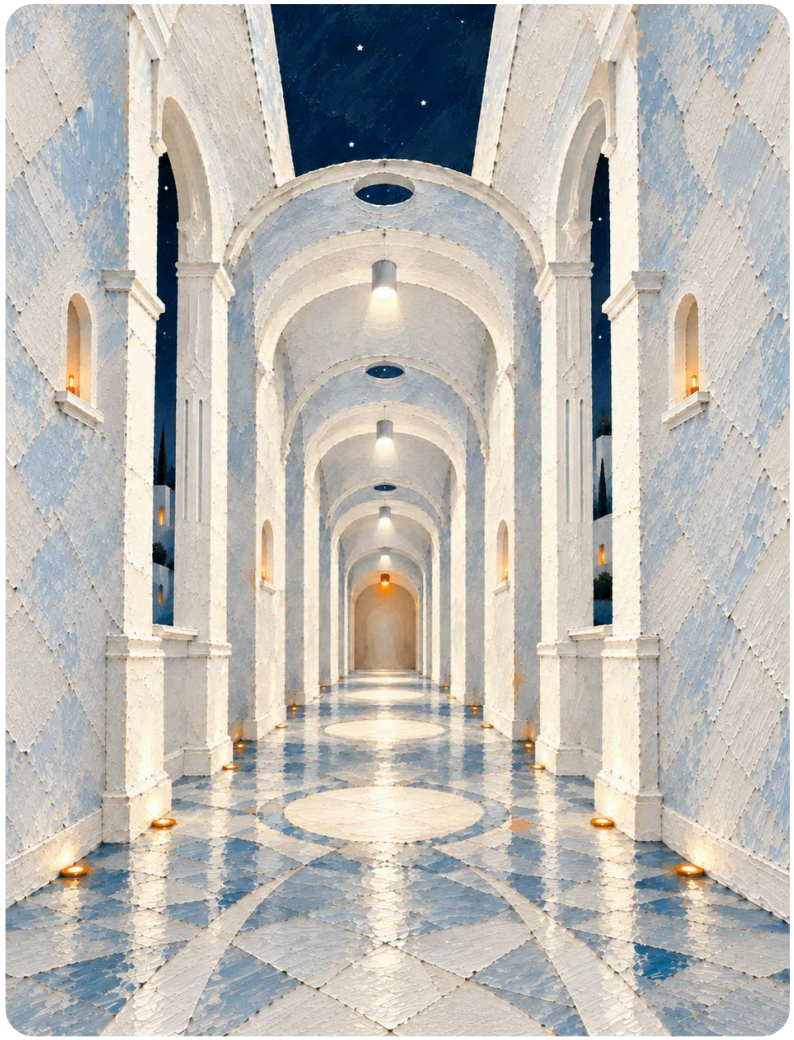
AI 게이트웨이

AIRouter

지능형 AI Gateway 솔루션

AI 비용은 청구서 전에 보고,
민감정보는 공급자로 가기 전에 걸러 냅니다.



이 브로슈어는 기능 목록이 아니라,
AI 모델을 운영하는 기업이 실제로
꺼내는 일곱 가지 고민에서 출발합니다.

고민 한 장, 해결 한 장씩
차례로 읽으실 수 있습니다.



들어가며

들어가며	AI 모델을 운영하는 기업의 일곱 가지 고민	p.03
솔루션	AIRouter는 무엇이고, 어디에서 고민을 푸는가	p.04

일곱 가지 고민과 해결

고민 01	AI 비용을 청구서를 받고서야 안다	p.05
고민 02	공급자마다 키 · 콘솔 · 계약이 따로다	p.07
고민 03	AI 비용이 어디서 새는지 모른다	p.08
고민 04	개인정보가 그대로 AI로 나갈까 걱정된다	p.10
고민 05	에이전트가 한 일을 사고 뒤에 설명할 수 없다	p.12
고민 06	모델 하나 바꾸려면 다시 배포해야 한다	p.14
고민 07	공급자 한 곳이 멈추면 서비스도 멈춘다	p.16

적용과 도입

적용 예시	AIRouter가 들어가면 달라지는 AI 운영	p.17
도입 순서	서비스 하나로 시작해 넓혀 가는 도입 순서	p.19
문의	데모 영상 · 도입 문의	p.20

자주 듣는 고민

AI 모델을 운영하는 기업이 먼저 꺼내는 일곱 가지 고민

여러 AI 모델을 서비스에 붙이는 순간,
관리 부서는 통제를, 개발 조직은
속도를 요구합니다.

두 쪽에서 반복해서 듣는 이야기를
모았습니다.



가장 마음에 걸리는 고민을 클릭하면,
그 고민을 AIRouter가 어떻게 푸는지로
바로 이동합니다.

고민 01 “AI 비용은 청구서를 받고서야 얼마인지 압니다.” 예산 통제 →

고민 02 “공급자마다 키도, 콘솔도, 계약도 따로입니다.” 공급자 통합 →

고민 03 “AI 비용이 어디서 새는지 모릅니다.” 비용 절감 →

고민 04 “고객 개인정보가 그대로 AI로 나갈까 걱정입니다.” 개인정보 →

고민 05 “에이전트가 한 일을 사고 뒤에 설명할 수 없습니다.” 에이전트 통제 →

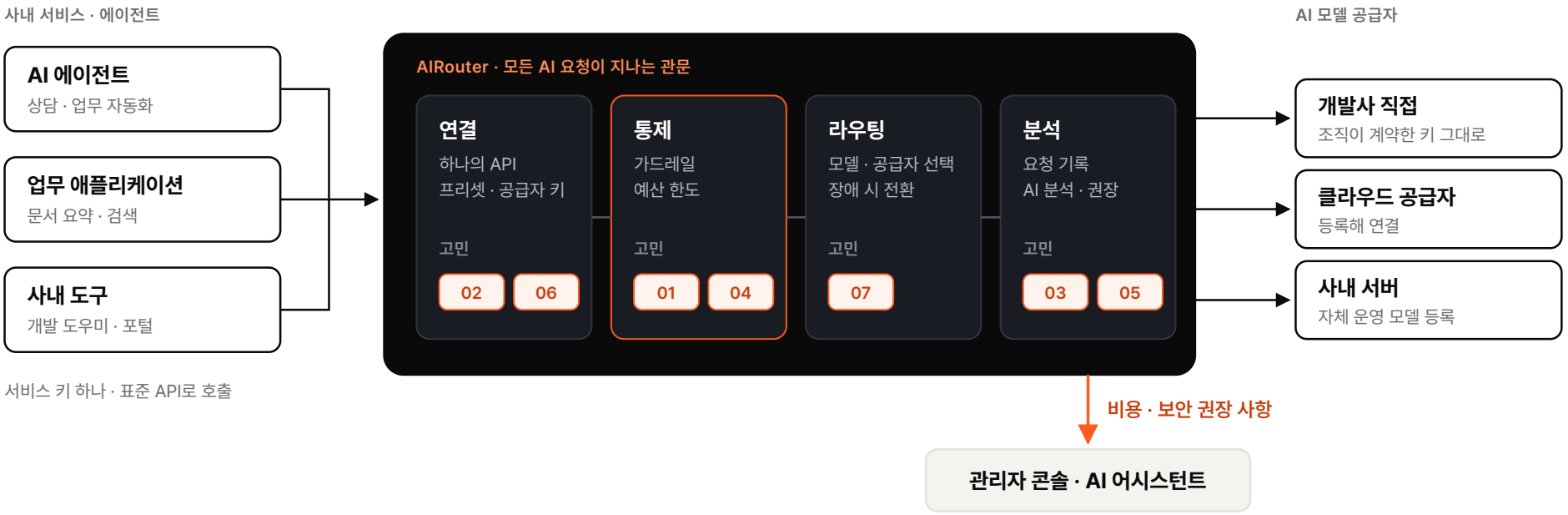
고민 06 “모델 하나 바꾸려면 코드를 고쳐 다시 배포합니다.” 모델 교체 →

고민 07 “공급자 한 곳이 멈추면 우리 서비스도 멈춥니다.” 장애 대응 →

하나의 관문

AIRouter는 이 고민들을 하나의 관문에서 풉니다.

지능형 AI Gateway 솔루션 공급자 콘솔과 서비스별 자체 개발로 흩어진 예산 · 보안 · 장애 대응을 한 관문에 모으고, 그 기록을 AI가 분석해 조치를 제시합니다.



'고민' 아래 숫자 버튼을 클릭하면 해당 고민으로 이동합니다.

고민 01 · 예산 통제

AI 비용은 청구서를 받고서야 얼마인지 압니다.

재무 · 구매 담당자가 가장 먼저 꺼내는 이야기입니다.



현장에서는 이런 일이 생깁니다

사용한 뒤에야 금액을 압니다

AI 비용은 요청이 쌓인 만큼 나가는데,
금액은 다음 달 청구서에서야 확인합니다.

누가 얼마나 사용했는지 나눌 수 없습니다

공급자 청구서는 계정 단위라,
부서 · 프로젝트별 사용을 따로 계산해야 합니다.

멈출 수단이 없습니다

한 서비스가 갑자기 많이 호출해도,
예산을 넘기 전에 막을 방법이 없습니다.

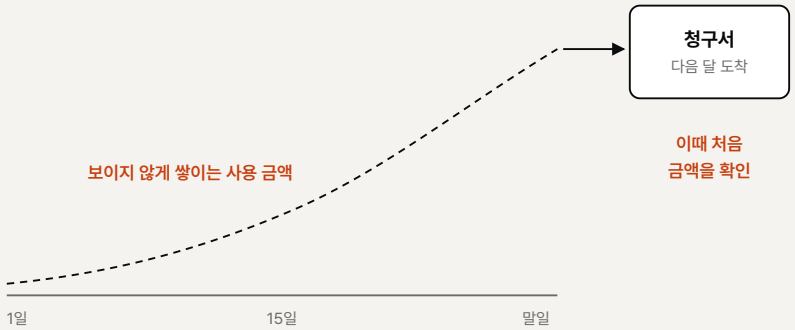
숨은 걱정 **"AI 예산의 통제권을 잃는 것"**

AIRouter의 해결 보기 →

해결 01 · 사후 확인이 아니라 사전 통제

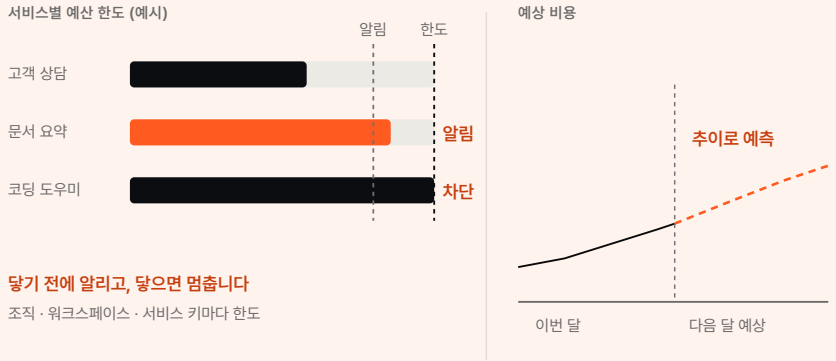
정해 둔 한도에 닿기 전에 알리고, 닿으면 호출을 멈춥니다.

지금



사용과 청구 사이의 한 달 동안 비용이 보이지 않습니다.

AIRouter 도입 후



사용하는 순간 단가로 환산한 비용을 보고, 한도를 넘는 호출은 보내기 전에 거절합니다.

단계마다 예산 한도

조직 · 워크스페이스 · 서비스 키마다 사용할 수 있는 금액을 정해 둡니다.

닿기 전에 알림, 닿으면 차단

한도에 가까워지면 담당자에게 알리고, 한도에 닿으면 호출을 멈춥니다.

팀 · 프로젝트 · 모델별로

여러 공급자의 사용량과 단가로 환산한 비용을 한 곳에서 원하는 단위로 나눠 봅니다.

다음 달 비용 예측

지금까지의 추이로 다음 주 · 다음 달 예상 비용을 미리 보여 줍니다.

고객이 얻는 것

- ✓ 청구서보다 먼저 예상 비용을 압니다
- ✓ 정해 둔 예산 범위 안에서 운영합니다
- ✓ 부서별 사용을 근거로 예산을 나눕니다

고민 02 · 공급자 통합

공급자마다 키도, 콘솔도, 계약도 따로입니다.



서비스마다 따로 연동합니다
애플리케이션마다 공급자와 하나씩 연결해, 관리할 곳이 계속 늘어납니다.

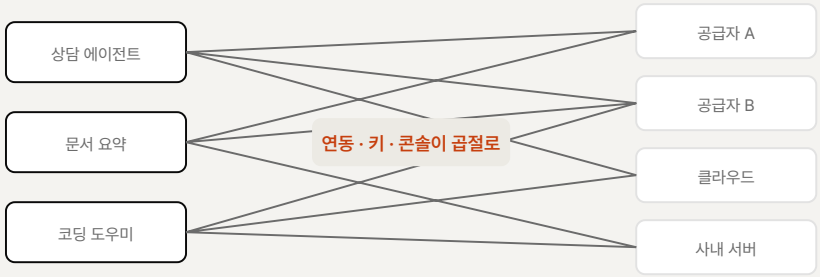
확인할 콘솔도 공급자 수만큼
사용량 · 비용 · 보안 설정을 공급자 콘솔마다 따로 봅니다.

숨은 걱정
"공급자가 늘수록 관리가 감당 안 되는 것"

해결 02 · 하나의 관문

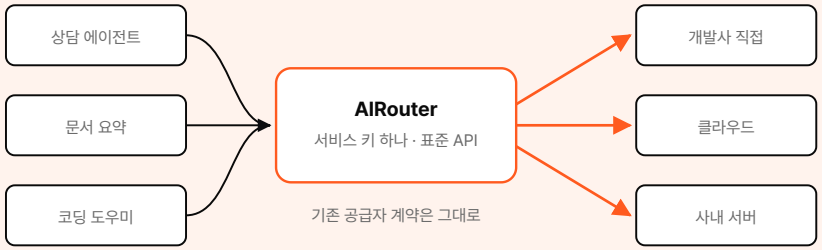
서비스 키 하나로, 여러 공급자의 모델을 사용합니다.

지금



서비스와 공급자가 늘 때마다 연결이 곱절로 늘어납니다.

AIRouter



서비스는 한 곳에만 연결하고, 공급자는 관문 뒤에서 관리합니다.

고객이 얻는 것

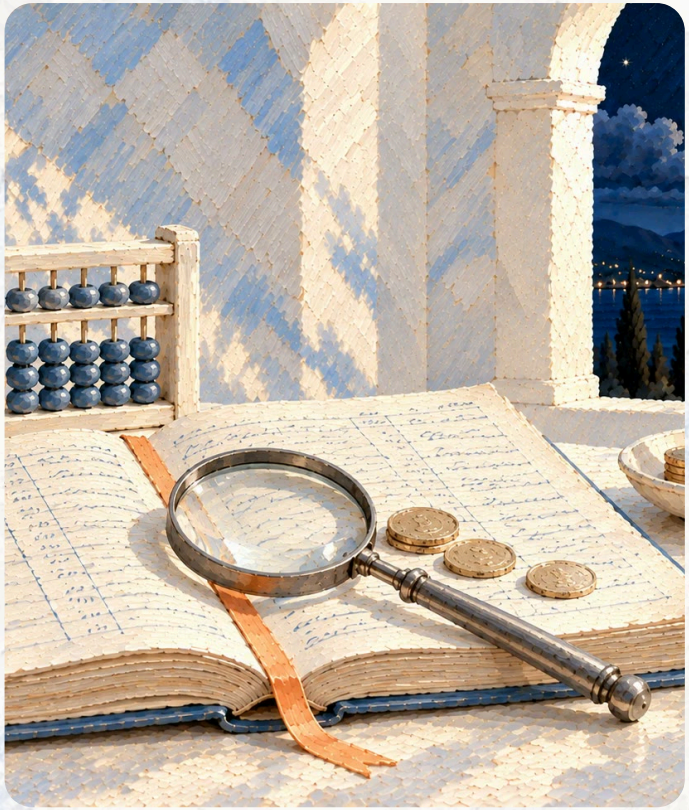
- ✓ 기존 공급자 계약을 그대로 사용합니다
- ✓ 공급자별 비용을 한눈에 봅니다

목차로 돌아가기

고민 03 · 비용 절감

AI 비용이 어디서 새는지 모릅니다.

AI 서비스를 운영하는 개발 · 재무 담당자가 함께 하는 이야기입니다.



현장에서는 이런 일이 생깁니다

같은 내용을 매번 새로 계산합니다

반복해서 보내는 긴 지시문에 캐시를 사용하지 않아,
할인받을 수 있는 비용을 그대로 냅니다.

실패한 요청에도 비용이 나갑니다

재시도 · 실패 · 잘린 응답에 들어간 비용이
청구서 안에 묻혀 보이지 않습니다.

쉬운 일에도 가장 비싼 모델

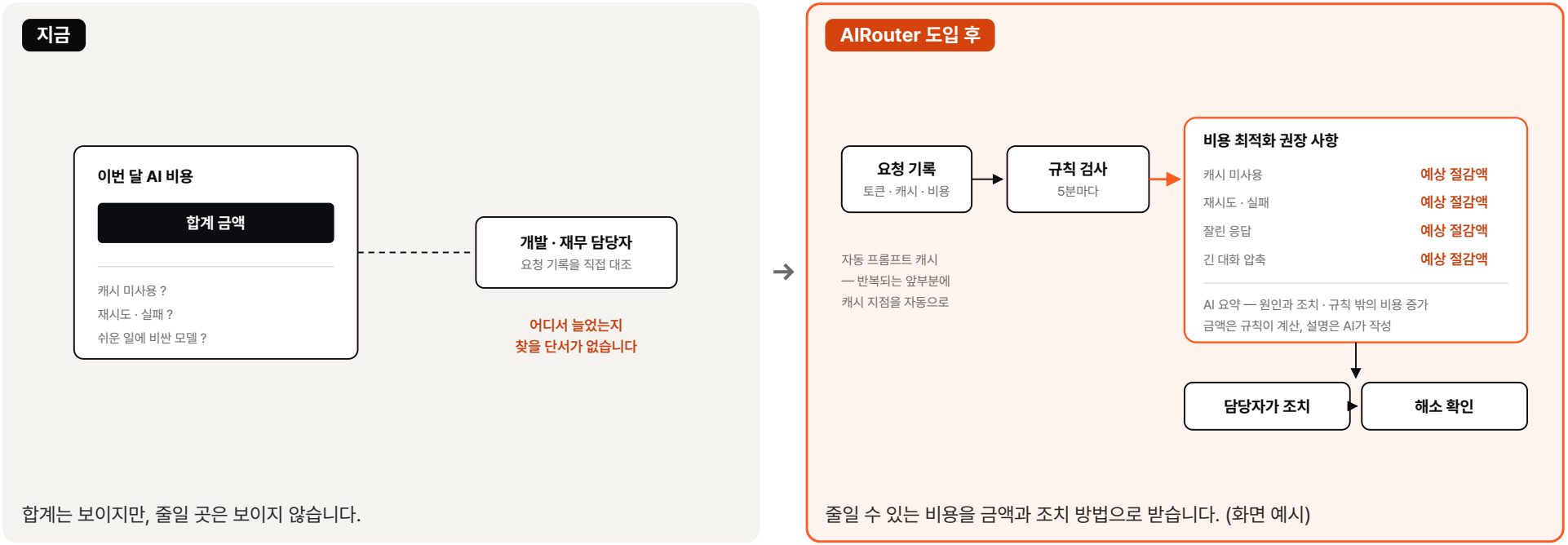
간단한 요약과 분류에도
최상위 모델이 사용됩니다.

숨은 걱정 **"줄일 수 있는 비용이 있는데도 찾지 못하는 것"**

AIRouter의 해결 보기 →

해결 03 · 줄일 수 있는 비용부터

요청 기록을 5분마다 검사해, 줄일 수 있는 비용과 방법을 알려 줍니다.



아홉 가지 비용 규칙

캐시 미사용 · 재시도 · 실패 ·
잘린 응답 · 긴 대화를 찾아 예상 절감액을
계산합니다.

자동 프롬프트 캐시

반복되는 앞부분에 캐시 지점을 붙이고,
같은 대화는 같은 공급자로 보냅니다.

조건에 맞는 모델로

입력 길이 · 도구 사용 · 이미지 포함 여부처럼
정해 둔 조건에 맞춰 모델을 나눠 보냅니다.

숫자는 규칙이, 설명은 AI가

금액과 건수는 요청 기록으로 계산하고,
AI는 원인과 조치를 정리합니다.

고객이 얻는 것

- ✓ 줄일 수 있는 비용을 예상 금액으로 봅니다
- ✓ 고칠 곳과 고치는 방법을 함께 받습니다
- ✓ 조치한 뒤 해소 여부까지 확인합니다

고민 04 · 개인정보 · 비밀 정보



고객 개인정보가 그대로 AI로 나갈까 걱정입니다.

보안 담당자가 AI 서비스 결재 전에 가장 먼저 묻는 질문입니다.

현장에서는 이런 일이 생깁니다

중간에서 확인할 곳이 없습니다

서비스마다 공급자로 바로 호출해,
무엇이 밖으로 나가는지 볼 지점이 없습니다.

읽은 문서 속 지시를 따릅니다

에이전트가 읽은 문서 · 검색 결과에 숨은 지시가
에이전트의 행동을 바꾸기도 합니다.

응답에도 비밀이 섞여 나옵니다

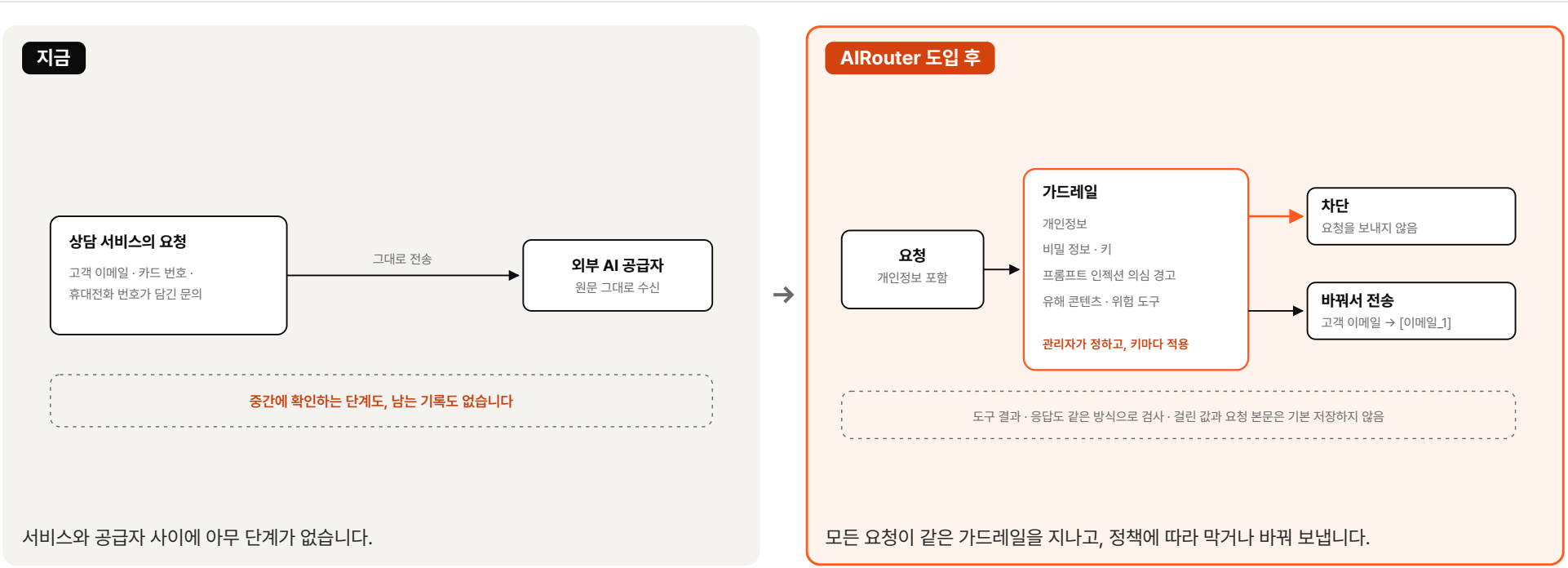
모델의 응답에 키 · 비밀번호 같은
비밀 정보가 섞여도 걸러 내지 못합니다.

숨은 걱정 **"유출 사고의 책임이 AI 도입을 결재한 사람에게 돌아오는 것"**

AIRouter의 해결 보기 →

해결 04 · 공급자로 가기 전에

주민등록번호 · 카드 번호 · 비밀 키는 공급자로 가기 전에 걸러 냅니다.



요청 · 도구 결과 · 응답 검사

나가는 요청뿐 아니라 에이전트가 받은 도구 결과와 모델의 응답까지 검사합니다.

막거나, 바꿔서 보냅니다

주민등록 · 외국인등록 · 여권 · 운전면허 · 계좌 번호 등은 기본으로 유형 표시로 바꿔 전송합니다.

개발자가 끌 수 없는 정책

가드레일은 관리자가 정해 서비스 키마다 적용하고, 호출하는 쪽에서 해제할 수 없습니다.

먼저 '기록만', 확인 뒤 차단

규칙을 기록만 하도록 걸어 두면, 차단했다면 막혔을 요청 수가 먼저 보입니다.

고객이 얻는 것

- ✓ 민감정보가 밖으로 나가기 전에 멈춥니다
- ✓ 업무를 멈추지 않고 정책을 시험합니다
- ✓ 모든 서비스에 같은 보안 기준

고민 05 · 에이전트 통제 · 감사



에이전트가 무엇을 했는지,
사고가 나면 설명할 수 없습니다.

에이전트를 업무에 넣으려는 IT · 보안 조직이 묻는 질문입니다.

현장에서는 이런 일이 생깁니다

작업 하나가 호출 수십 건이 됩니다

에이전트의 작업 하나가 모델 호출과
도구 호출 수십 건으로 이어져,
위험이 한 건에서는 보이지 않습니다.

위험한 도구도 그대로 실행됩니다

메일 발송 · 결제 · 삭제 · 명령 실행 같은
도구 호출을 중간에서 확인할 곳이 없습니다.

키가 어디서 사용되는지 모릅니다

처음 보는 IP에서의 사용이나
오래 쉬던 키의 재사용을 알아차리지 못합니다.

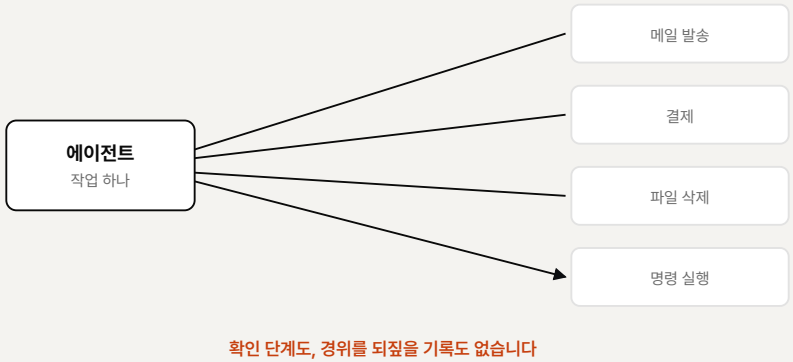
숨은 걱정 **"사고 뒤에 경위를 되짚을 기록이 없는 것"**

AIRouter의 해결 보기 →

해결 05 · 에이전트의 행동까지

위험한 도구 호출은 통제하고, 위험 징후는 기록에서 먼저 찾아 알립니다.

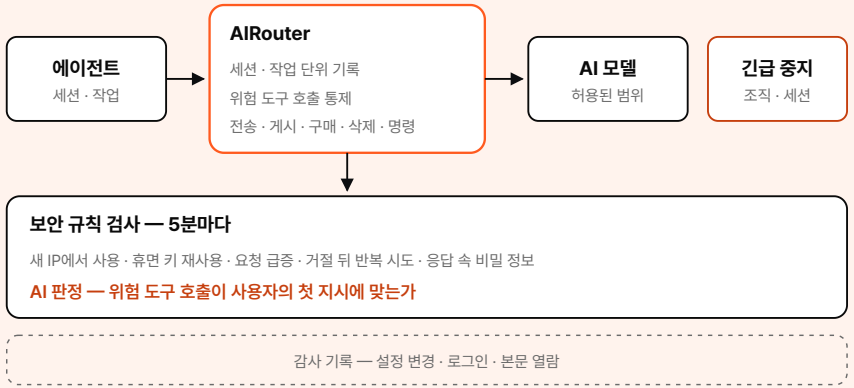
지금



확인 단계도, 경위를 되짚을 기록도 없습니다

에이전트와 도구 사이에 확인도 기록도 없습니다.

AIRouter 도입 후



모든 호출이 기록되고, 위험은 규칙과 AI가 함께 찾아 알립니다. (화면 예시)

위험 도구 호출 통제

전송 · 게시 · 구매 · 삭제 · 명령 실행 같은 도구 호출을 정책으로 통제합니다.

열다섯 가지 보안 규칙

새 IP · 휴면 키 · 요청 급증 · 거절 뒤 반복 시도 같은 징후를 5분마다 찾습니다.

지시에 맞는 행동인가

위험 도구를 부른 세션은 AI가 사용자의 첫 지시와 비교해 판정합니다.

즉시 중지와 감사 기록

조직 · 워크스페이스 · 세션 단위로 바로 멈추고, 설정 변경 · 로그인 · 본문 열람까지 기록합니다.

고객이 얻는 것

- ✓ 위험 징후를 사고 전에 알아차립니다
- ✓ 경위를 기록으로 설명합니다
- ✓ 문제가 생기면 바로 멈춥니다

고민 06 · 모델 교체



모델 하나 바꾸려면 코드를 고쳐 다시 배포합니다.

AI 서비스를 개발하는 팀이 가장 자주 하는 이야기입니다.

현장에서는 이런 일이 생깁니다

모델 이름이 코드에 박혀 있습니다

새 모델이 나올 때마다
수정 · 검증 · 배포를 반복합니다.

되돌리는 데도 배포가 필요합니다

바꾼 모델에 문제가 생기면,
원래대로 돌리는 데도 다시 배포해야 합니다.

공급자마다 호출 방식이 다릅니다

공급자를 옮기면 요청 형식과 응답 처리부터
다시 맞춰야 합니다.

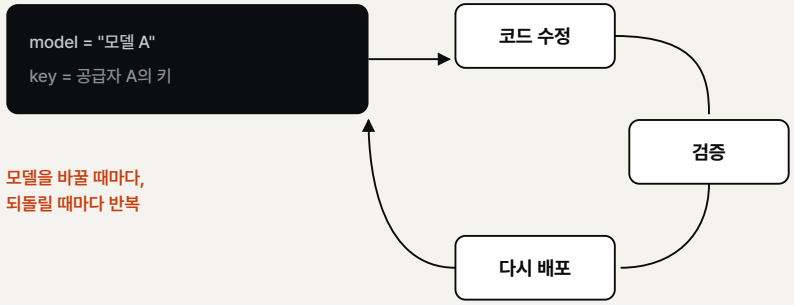
숨은 걱정 **"교체가 두려워 더 나은 모델을 사용하지 못하는 것"**

AIRouter의 해결 보기 →

해결 06 · 코드는 그대로, 운영은 콘솔에서

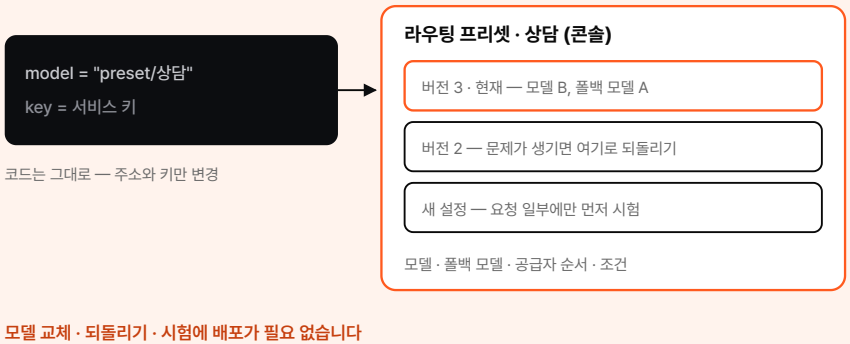
코드에는 프리셋 이름만 두고, 모델 교체와 되돌리기는 콘솔에서 합니다.

지금



모델 하나가 바뀌어도 개발 일정이 움직입니다.

AIRouter 도입 후



모델 운영은 콘솔로 옮기고, 개발은 서비스에 집중합니다. (화면 예시)

주소와 키만 바꾸면 연결

널리 사용되는 두 가지 AI API 형식과 호환되어, 기존 코드는 주소와 키만 바꾸면 동작합니다.

라우팅 프리셋

모델 · 폴백 모델 · 공급자 순서 · 조건을 콘솔에 저장하고, 코드에서는 이름만 부릅니다.

버전과 되돌리기

프리셋을 버전으로 관리해, 문제가 생기면 바로 이전 버전으로 돌립니다.

일부 요청으로 먼저 시험

새 설정을 요청 일부에만 적용해 비용과 응답 속도를 확인한 뒤 넓힙니다.

고객이 얻는 것

- ✓ 모델 교체에 배포가 필요 없습니다
- ✓ 문제가 생기면 바로 되돌립니다
- ✓ 더 나은 모델을 부담 없이 시험합니다

고민 07 · 장애 대응

공급자 한 곳이 멈추면 서비스도 함께 멈춥니다.



호출 한도 초과가 사용자에게 그대로

공급자의 호출 한도를 넘으면,
그 오류가 사용자 화면까지 갑니다.

대체 로직을 서비스마다 만듭니다

재시도와 다른 공급자로의 전환을
개발팀이 서비스마다 직접 만듭니다.

숨은 걱정
"AI 공급자의 장애가 우리 서비스의 장애가 되는 것"

해결 07 · 허용된 범위 안에서 전환

공급자에 문제가 생기면, 허용된 다른 공급자로 넘어갑니다.



장애 시 자동 전환

응답 전에 실패하면 다음 공급자,
그다음 폴백 모델로 넘어갑니다.

실패가 잦은 곳은 잠시 제외

짧은 시간에 실패가 반복된
공급자로는 잠시 보내지 않습니다.

정책 밖으로는 보내지 않습니다

허용 목록 · 데이터 미저장 ·
학습 미사용 조건 밖으로는
보내지 않고 거절합니다.

고객이 얻는 것

- ✓ 장애 대응을 서비스마다 만들지 않습니다
- ✓ 전환도 보안 정책 안에서만 일어납니다

목차로 돌아가기

업무 적용 예시

AIRouter가 들어가면, AI 운영이 이렇게 달라집니다.

에이전트 개발 · 출시 전 점검

비용과 보안 문제를 출시 뒤에 알지 않습니다

고객이 얻는 것 — **출시 전에** 비용과 보안 문제를 고칩니다.

지금은
상담 에이전트에 드는 비용과 부르는 도구는, 실제 고객이 사용한 뒤에야 드러납니다.

→

AIRouter가 맡는 일
개발 환경에서 실행한 세션을 추적해 캐시 낭비 · 위험 도구 호출 · 가드레일 판정을 찾고, 원인과 조치를 정리합니다.

→

담당자가 하는 일
지적된 곳을 고치고 다시 실행해, 문제가 사라졌는지 확인한 뒤 출시합니다.

이 장면을 만드는 AIRouter: 해결 03 비용 분석 · 해결 05 도구 통제 · 해결 04 가드레일

연결하는 것: 개발 워크스페이스 · 서비스 키 · 세션 · 작업 기록

재무 · 부서별 예산

부서별 AI 비용을 손으로 나누지 않습니다

고객이 얻는 것 — **부서별 사용을 근거로** 예산을 나눕니다.

지금은
공급자 콘솔마다 사용량을 내려받아 부서별로 나누고, 스프레드시트로 월간 보고서를 만듭니다.

→

AIRouter가 맡는 일
워크스페이스별 한도를 지키게 하고, 대화로 요청하면 표 · 그래프가 들어간 월간 보고서를 만들어 줍니다.

→

담당자가 하는 일
보고서를 확인하고, AI 어시스턴트가 제안한 예산 조정을 확인해 실행합니다.

이 장면을 만드는 AIRouter: 해결 01 예산 통제 · 해결 02 한 곳에서 집계 · 해결 03 비용 분석

연결하는 것: 워크스페이스 · 예산 한도 · AI 어시스턴트

업무 적용 예시입니다. 귀사 서비스에 맞춘 범위와 효과는 도입 문의(20쪽)에서 함께 설계합니다.

업무 적용 예시 · 이어서

같은 방식으로, 다른 업무에서도 일합니다.

보안 · 규제 산업

정책을 바로 차단으로 걸지 않고, 먼저 시험합니다

고객이 얻는 것 — **업무를 멈추지 않고** 보안 정책을 강화합니다.

지금은
차단 규칙을 갑자기 켜면 업무가 막힐까 걱정되어, 결국 어떤 규칙도 켜지 못합니다.

→

AIRouter가 맡는 일
먼저 규칙을 기록만 하도록 걸고, 차단했다면 막혔을 요청 수를 서비스 키별로 보여 줍니다.

→

담당자가 하는 일
영향을 확인한 뒤 차단으로 바꾸고, 감사 기록을 보안 심사 자료로 사용합니다.

이 장면을 만드는 AIRouter
해결 04 가드레일
해결 05 보안 규칙
해결 01 예산 통제

연결하는 것
가드레일
감사 기록
알림 웹훅

서비스 운영 · 모델 교체

새 모델을 전체에 바로 적용하지 않습니다

고객이 얻는 것 — **배포 없이** 더 나은 모델로 옮겨 갑니다.

지금은
더 나은 모델이 나와도 전체 교체가 부담되어, 오래된 모델을 계속 사용합니다.

→

AIRouter가 맡는 일
라우팅 프리셋의 새 버전을 요청 일부에만 적용하고, 공급자 장애 때는 허용된 다른 공급자로 넘깁니다.

→

담당자가 하는 일
요청 기록으로 비용과 응답 속도를 비교한 뒤, 전체로 넓히거나 되돌립니다.

이 장면을 만드는 AIRouter
해결 06 라우팅 프리셋
해결 07 자동 전환
해결 03 비용 분석

연결하는 것
라우팅 프리셋
요청 기록
폴백 모델

업무 적용 예시입니다. 귀사 서비스에 맞춘 범위와 효과는 도입 문의(20쪽)에서 함께 설계합니다.

도입 순서

모든 서비스를 한 번에 옮기지 않습니다.
서비스 하나를 연결해 확인하고, 넓혀 갑니다.

1

현황을 함께 점검

어떤 서비스가 어떤 공급자를 사용하는지,
키와 비용은 어디서 관리되는지 살펴봅니다.

고객이 얻는 것
흩어진 AI 호출과 키의 현황

2

서비스 하나 연결

기존 코드에서 주소와 키만 바꿔
서비스 하나를 연결합니다. 가드레일은
'기록만'으로 먼저 겁니다.

고객이 얻는 것
주소와 키만 되돌리면 원래대로 돌아가는
부담 없는 확인

3

정책 켜기

예산 한도와 라우팅 프리셋을 정하고,
영향을 확인한 가드레일 규칙부터
차단으로 바꿉니다.

고객이 얻는 것
정해 둔 예산 · 보안 기준 안의 운영

4

다른 서비스로 확산

부서 · 프로젝트별로 워크스페이스와
서비스 키를 나눠, 같은 정책 아래
넓혀 갑니다.

고객이 얻는 것
AI 호출 전체를 한 곳에서 보는 운영

첫 상담은 아래 자료 일부만 있어도 충분합니다

사용 중인 AI 공급자
계약 중인 AI 공급자와
클라우드 목록

AI를 호출하는 서비스
에이전트 · 애플리케이션 ·
사내 도구 목록

월 AI 비용 규모
공급자별 대략의
사용 금액

보안 기준
외부 전송이 막혀야 하는 정보와
사내 보안 정책

SaaS로 가볍게 시작
서비스 키를 받아 바로 연결하고,
기존 공급자 계약은 그대로 사용합니다.

전용 환경 협의
보안 요건이 높은 조직은
고객 전용 클라우드 구성을 협의합니다.

요청 본문은 기본 미저장
요청 본문은 조직이 켤 때만,
정한 기간 동안 저장합니다.

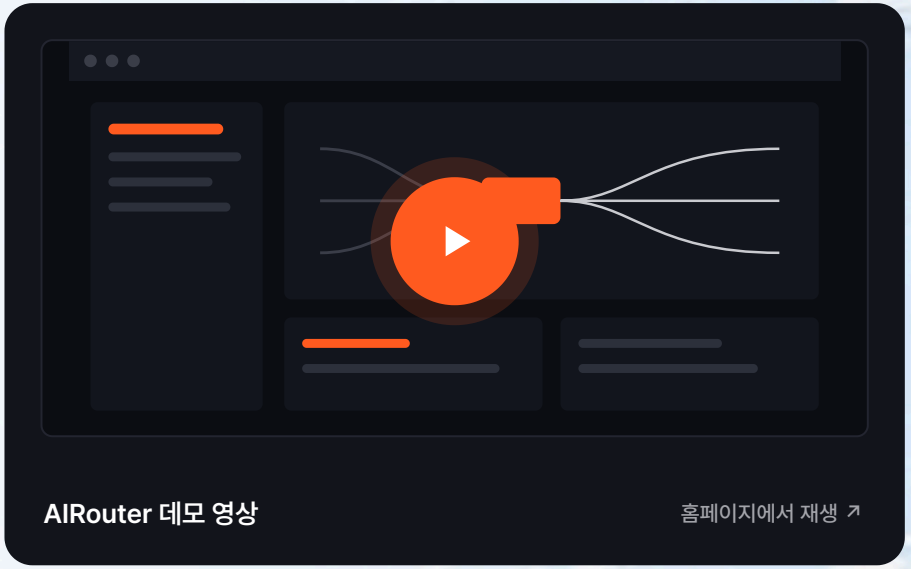
상담 안내

귀사의 고민부터 들려주세요.

AI를 호출하는 서비스 하나를 알려 주시면, 그 서비스를 기준으로
비용이 새는 곳 · 밖으로 나가는 민감정보 · 장애 시 전환 경로를
데모로 짚어 드립니다. 자료는 19쪽 목록 일부면 충분하고,
도입 비용 산정 기준도 함께 안내합니다.
먼저 살펴보고 싶다면 데모 영상을 확인해 보세요.

도입 문의하기 →

▶ 데모 영상 보기



문의 때 가장 급한 고민 하나를 알려 주세요

- 예산 통제
- 공급자 통합
- 비용 절감
- 개인정보
- 에이전트 통제
- 모델 교체
- 장애 대응

이메일
sales@vicsnc.com

웹사이트
vicsnc.com