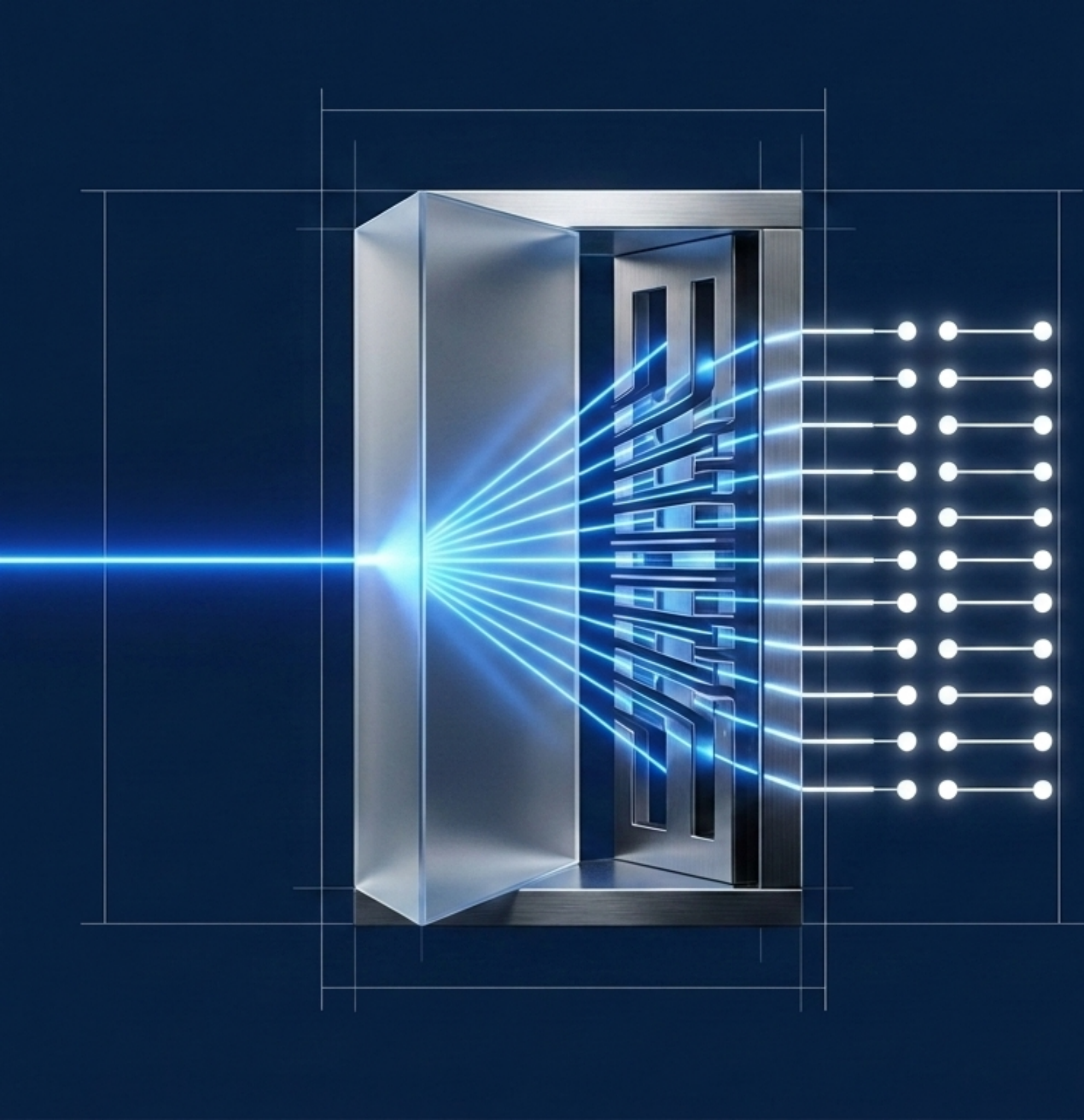




安全で低コストな チャットボット

「生成しないAI」が実現する
構造的なリスク排除と
圧倒的な業務効率化

エグゼクティブサマリー：約束される4つの価値



回答は「生成」しない

あらかじめ人が作成・確認した固定のFAQを一字一句その'返答。でっち上げ（ハルシネーション）の発生を原理的に排除。



「生成AI」は利用せず、 AIの利用はベクトル化のみ

AIの利用は「言葉の意味の数値化（ベクトル化）」のみ。
文章を書くAI（生成AI／LLM）は一切不使用。



常時稼働コスト「ゼロ」

常時起動する高額な検索サーバーを廃止。費用は1問い合わせあたり数百万分の1ドル規模の「使った分だけ」。



24時間・自己解決の推進

追加の運用費をかけずに利用者の自己解決率を高め、
担当者の業務負担を劇的に軽減。

解決すべきジレンマ：効率化の要求と生成AIの罠



膨大な業務負担

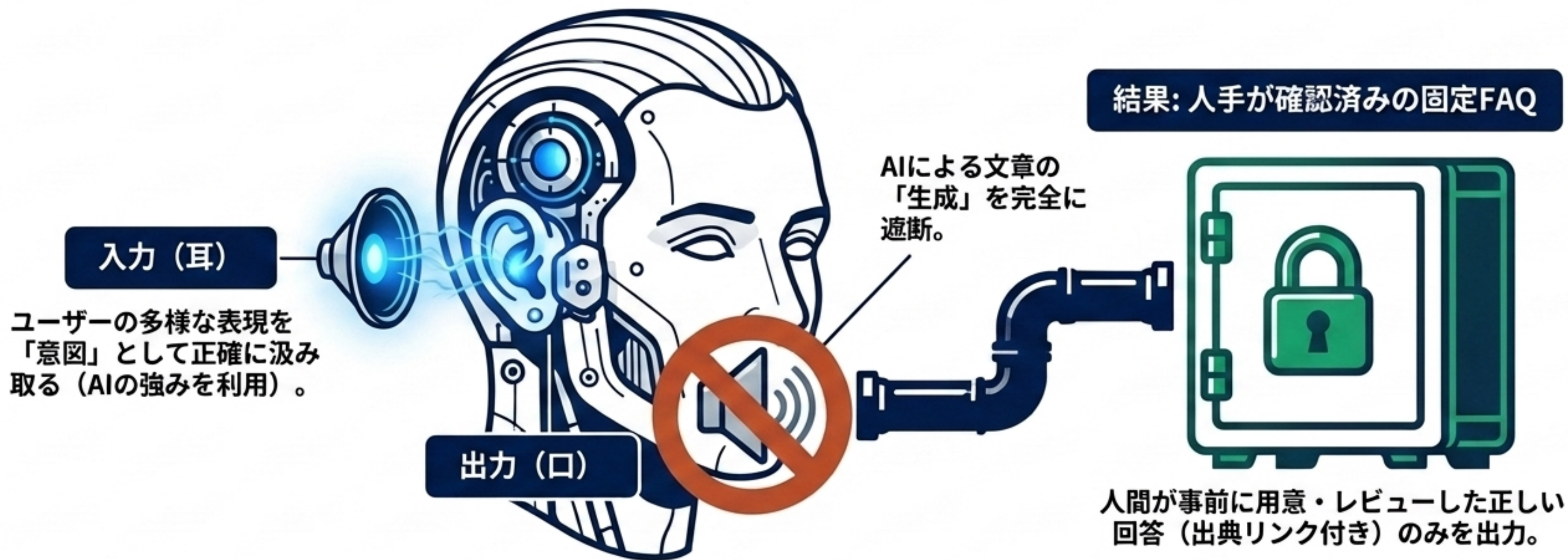
- よくある質問（件数多・定型的）が担当者の時間を奪う。
- 24時間即時対応への高まるニーズ。

ブランドを毀損するリスク

- 一般的な生成AIチャットが引き起こす「それらしい嘘（ハルシネーション）」。
- 不適切な誘導や機密漏洩の懸念。

結論：一般的な生成AIの仕組みを採用しては、安全性と省コストを両立できない

パラダイムシフト：AIを「耳」としてのみ使い、 「口」としては使わない



Key Message

AIを「なんでもできる魔法」としてではなく、「厳密な照合フィルター」として
極限まで利用範囲を絞り込むことが、リスクとコストを同時にゼロに近づける工夫です。

3つの検索モード比較：なぜAI（ベクトル化）が必要なのか

	意味を汲み取れるか	AIを利用するか	コスト
最適 ① キーワード検索 （ベクトル）	○ 意図を理解する	あり（意味のベクトル化のみ）	使った分だけ極小
補助 ② カテゴリから探す	✕ 階層を辿るのみ	なし	ゼロ
対比用 ③ 部分一致検索	✕ 入力文字そのまま	なし	ゼロ

自由な言葉での問い合わせに対し、②や③では「表記ゆれ」に対応できず回答を取りこぼします。

検証：従来型検索（部分一致）の限界

③ 部分一致



入力：「更新の周期」
結果：0件（該当なし）

入力：「資格を辞めたい」
結果：0件（該当なし）

① ベクトル検索

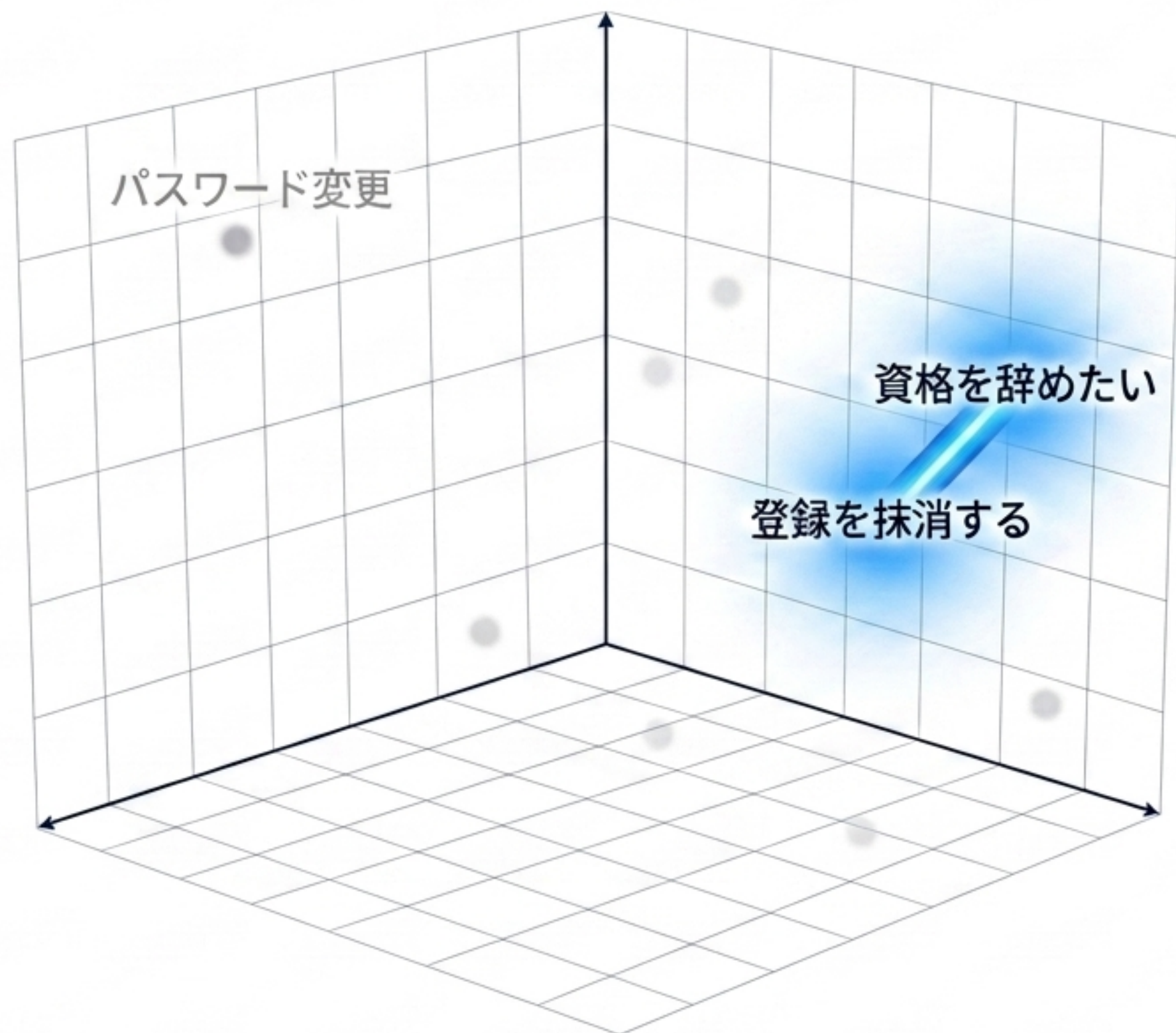


入力：「更新の周期」
結果：「更新期限」のFAQを意味で捉えて回答

入力：「資格を辞めたい」
結果：「登録の抹消手続き」のFAQを的確に提示

言い換えや同義語を吸収する「意味の検索」がなければ、自己解決率は上がりません。

技術的証明：AIによる「意味のベクトル化（空間配置）」



1. 数値化

Amazon Bedrockの埋め込みモデル（Titan Embeddings）が、入力された文章を「1024個の数値の並び」に変換します。

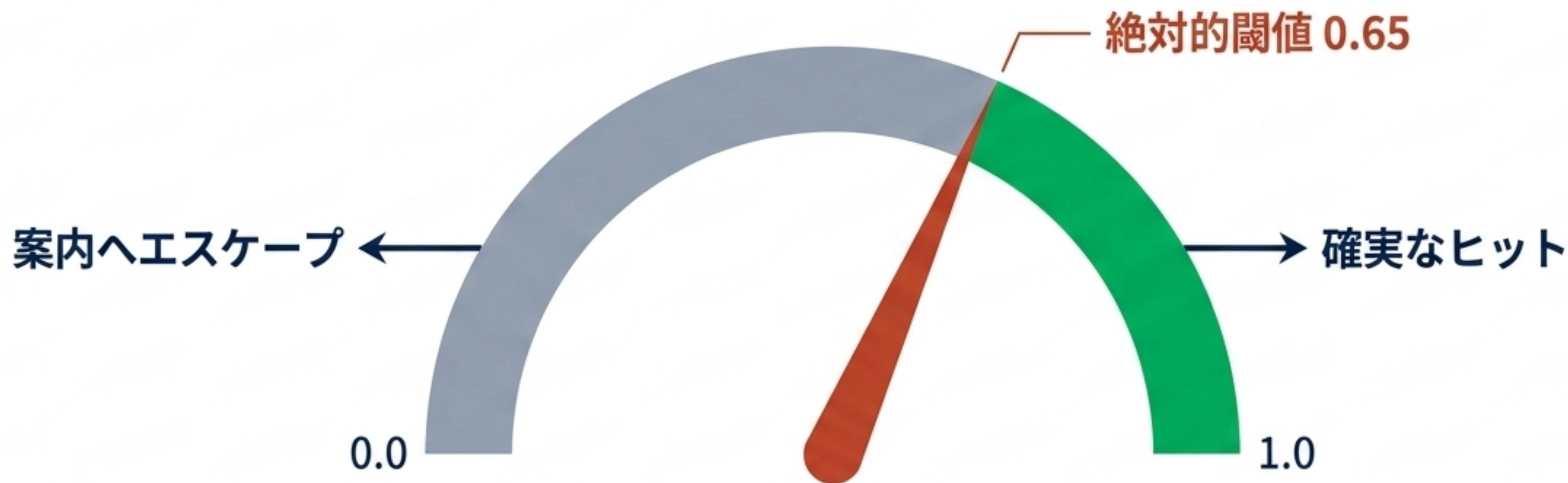
2. 空間上の配置

似た意味の文章は、数値空間（マップ）上で極めて近い距離に配置されます。

3. 類似度の計算（コサイン類似度）

ベクトル化された質問とFAQの「近さ」を、「コサイン類似度」というスコア（0～1）で計算します。スコアが1に近いほど意味が似ています。

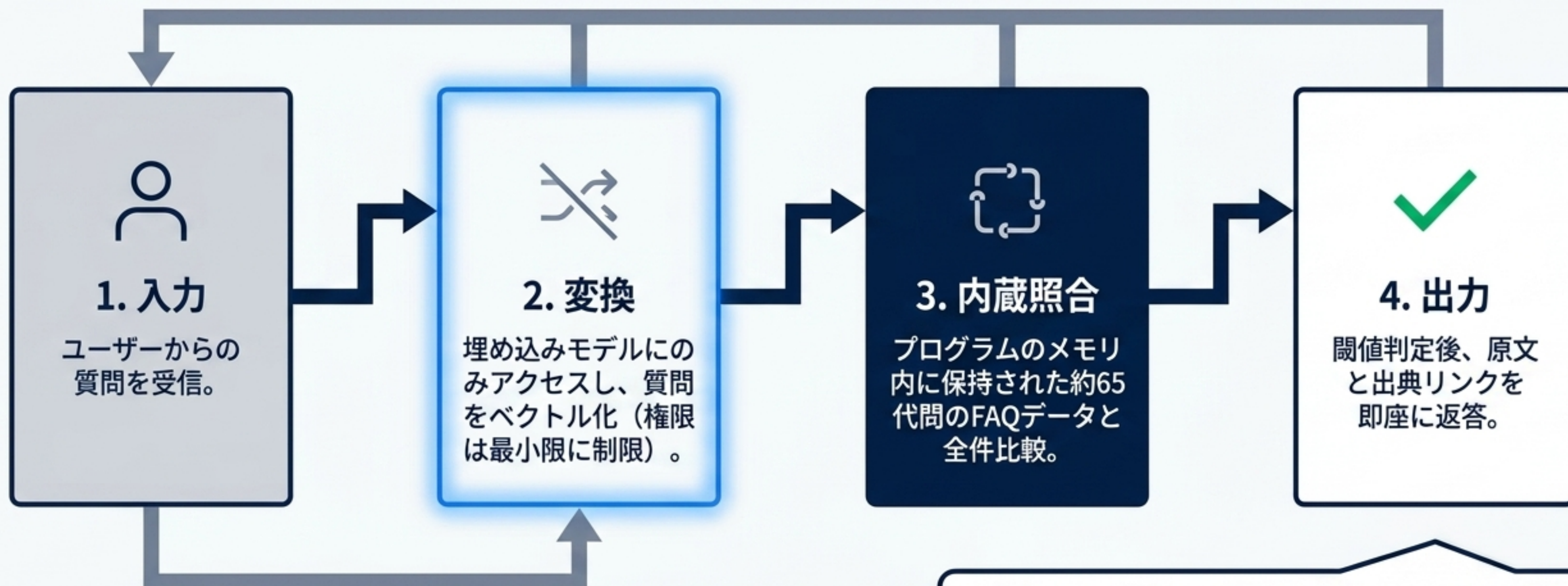
精密なフィルター：絶対的閾値「0.65」による安全装置



- 🧠 ・ 質問とFAQの「近さ」をコサイン類似度スコア（0～1）で計算。
- ⚙️ ・ 0.65 以上: 「該当あり」として、FAQの原文をそのまま返答。
- ✅ ・ 0.65 未満: 「該当なし」と厳格に判断し、無理に答えず「窓口へご連絡ください」と丁寧に案内。

この閾値は実測に基づきチューニングされています。最も近い候補を無理に返す「当てずっぽう」の設計を排除し、「分からない時は答えない」安全性をシステムレベルで担保しています。

全体アーキテクチャ：外部依存をなくした閉域照合プロセス



検索サーバーを常時稼働させない「内蔵型」設計が、圧倒的なコスト削減の鍵です。

セキュリティとガバナンス：リスク比較マトリクス

	一般的な生成AIチャット	本システム
でっち上げ (ハルシネーション)	あり得る	原理的に発生しない（生成しない） 
不適切な誘導・ プロンプト攻撃	リスクあり	新たな文面は出ない（固定集合から選ぶのみ）
機密の生成・漏洩	注意が必要	学習も生成もしないため該当せず
説明責任・監査性	出力の再現が困難	人手作成の固定回答＋出典提示で容易
知的財産権侵害リスク	学習データ由来の懸念	自社コンテンツのみ利用で完全に排除

コスト構造の変革：常時課金ゼロ・使った分だけの極小コスト

常時稼働する
検索サーバー
(月額約\$42)が
サイレントに
固定費として
計上されていた。

New Reality
(従量課金)
1,000問処理しても
合計約 **\$0.001**

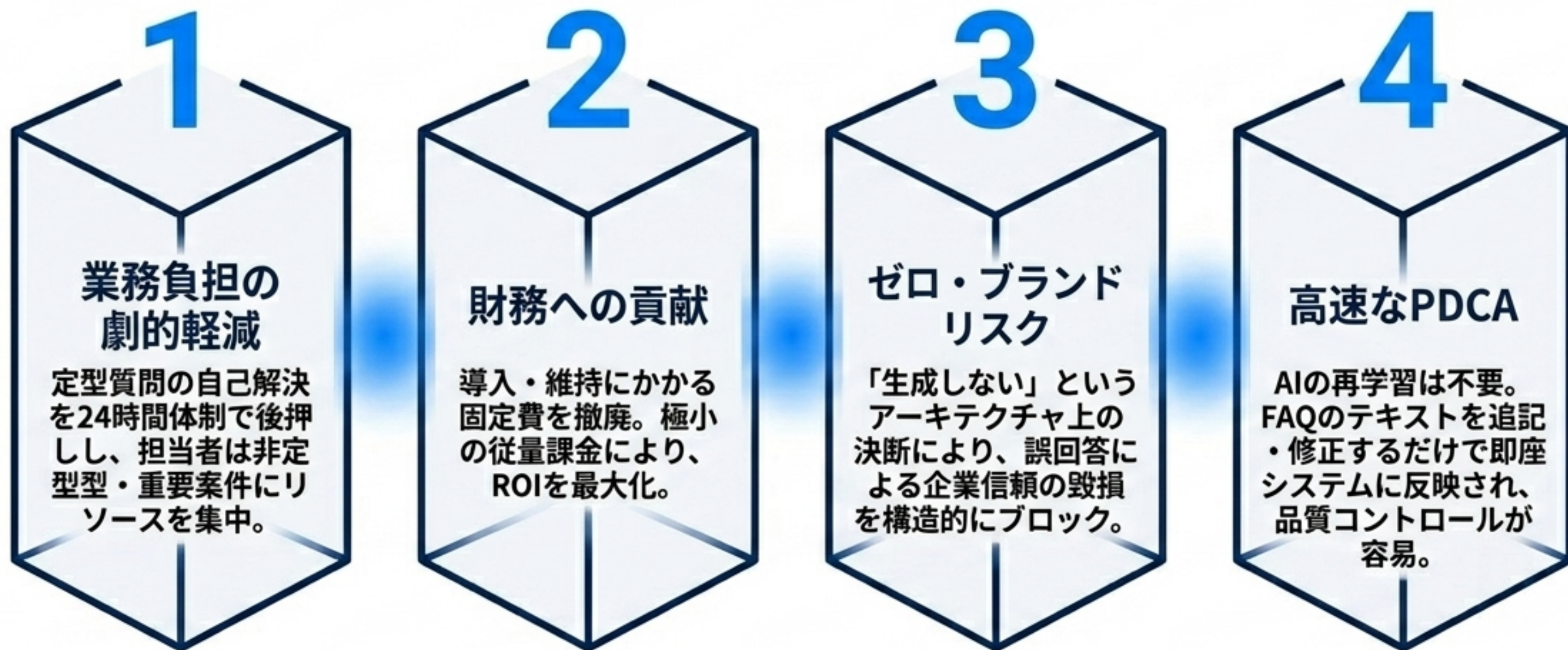
1万問でも約 **\$0.02**

(1問あたりの処理コスト：
約数百万分の1ドル～
10万分の1ドル)

After
サーバー廃止により
固定費 \$0

予算を圧迫するサイレントな固定費を完全に排除。
「コストをかけずに業務負担を減らす」を現実のアーキテクチャで証明しました。

期待されるビジネスインパクト と今後の展望



AIの真価は「何でもできること」ではなく、
「適切な制約のもとで、最も確実な結果を出すこと」にあります。