

# Agentic Vision in Gemini: 自律型AIの新境地

「見て説明する」から「見て、考えて、コードで解決する」へ

# 2026年のマクロトレンド：知能の産業化とエージェントへの転換



## 産業化 (Industrialization)

2026年1月、AIは「西部開拓時代」を終え、信頼性と実用性が求められる「産業時代」へ突入。

## 主権と実利 (Sovereignty & Utility)

企業は単なる会話能力ではなく、物理的な制約（電力、コスト）の中で確実な成果を出す「仕事をするAI」を求めている。

## 市場の要請 (Market Demand)

Product Hunt等での反響が示す通り、開発者が求めているのは「推測」ではなく「解決」である。

AIの役割は「チャットボット（対話）」から「エージェント（実行）」へと移行した。

# 従来の限界：確率論的な「推測」の壁

VLM（視覚言語モデル）は、なぜ細部を見落とし、数を間違えるのか。



- 幻覚 (Hallucination): オブジェクトのカウントや計算で嘘をつく
- 解像度の欠落: 細部を物理的に認識できない
- 検証不可能: 回答の根拠がブラックボックス



必要なのは「観光客」のような一瞥ではなく、「捜査官」のような調査である。

# Agentic Vision : 能動的に「調査」するAIの登場



Gemini 3 Flashは、画像を静的な入力としてではなく、「探索すべき環境」として扱う。



## 能動的調査 (Active)

答えを急がず、必要に応じて画像を操作（ズーム、クロップ）して情報を取得する。



## コードによる解決 (Code-driven)

```
def analyze_image(img_path):  
    results = model.predict(img_path)
```

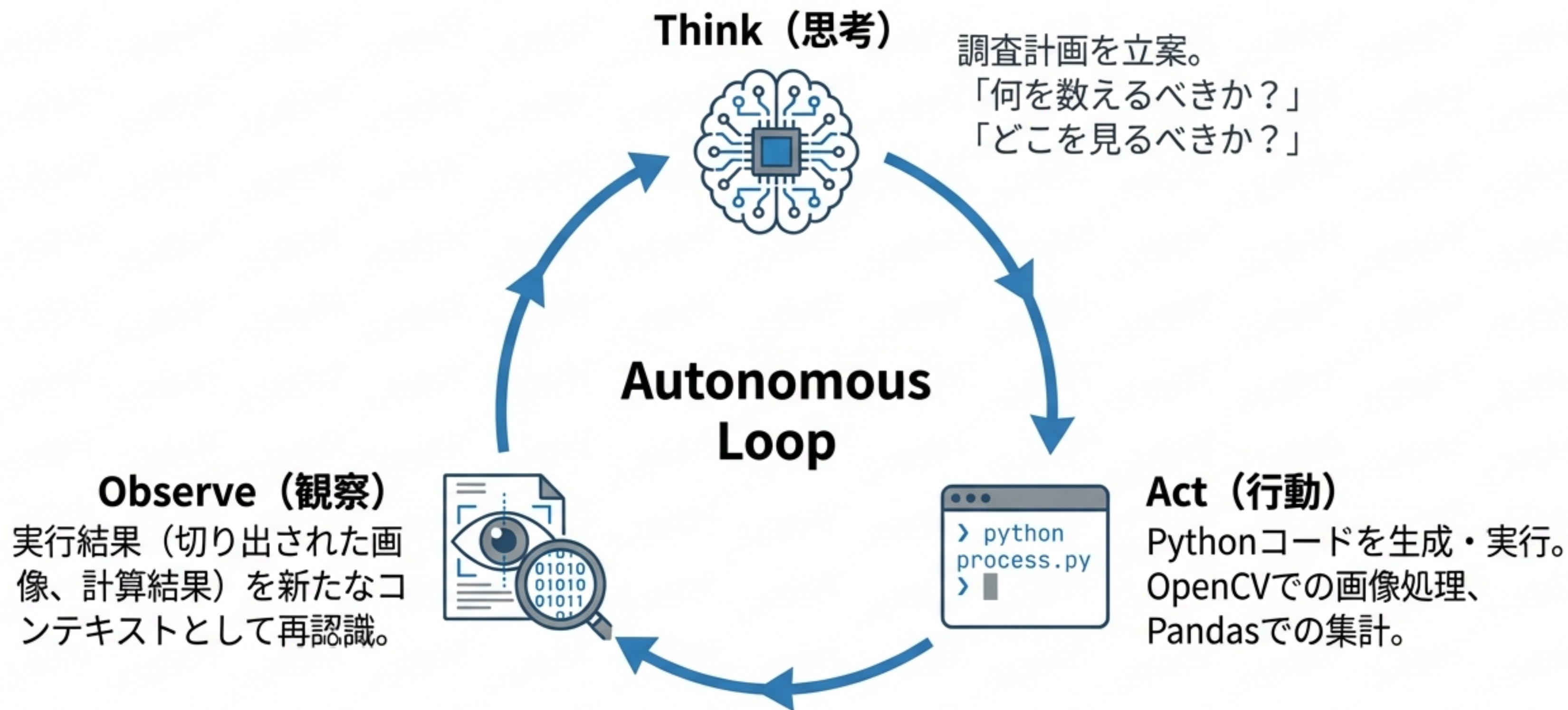
視覚情報をPythonコードで処理し、論理的な裏付けを持って回答する。



## パフォーマンス (Performance)

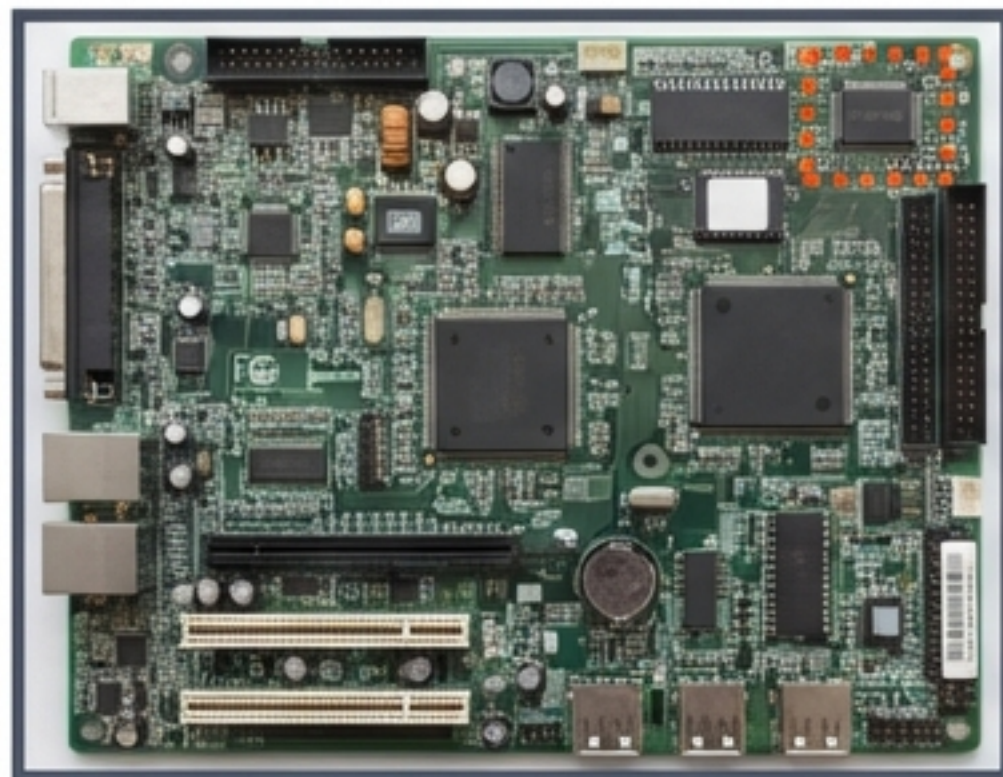
コード実行により、視覚ベンチマークにおいて5~10%の品質向上を一貫して記録。

# メカニズム：Think → Act → Observe の自律ループ



このループにより、AIは「直感 (Vision)」と「論理 (Code)」のギャップを埋める。

# 機能詳細①：能動的なズームと検査 (Active Inspection)



Step 1: 検知 (Detection)



```
img.crop((x, y, w, h))
```



Step 3: 再認識 (Re-observe)

Step 2: 行動 (Act)

シナリオ: マイクロチップのシリアル番号や、遠景の道路標識の読み取り。

シナリオ: マイクロチップのシリアル番号や、遠景の道路標識の読み取り。

AIは「解像度が足りず読めない」と判断すると、Pythonスクリプトで画像の特定領域を物理的にクロップし、高解像度で再読み込みを行う。

## 機能詳細②：視覚的数学と決定論的計算 (Visual Math)



ビジュアル・スクラッチパッド (Visual Scratchpad)

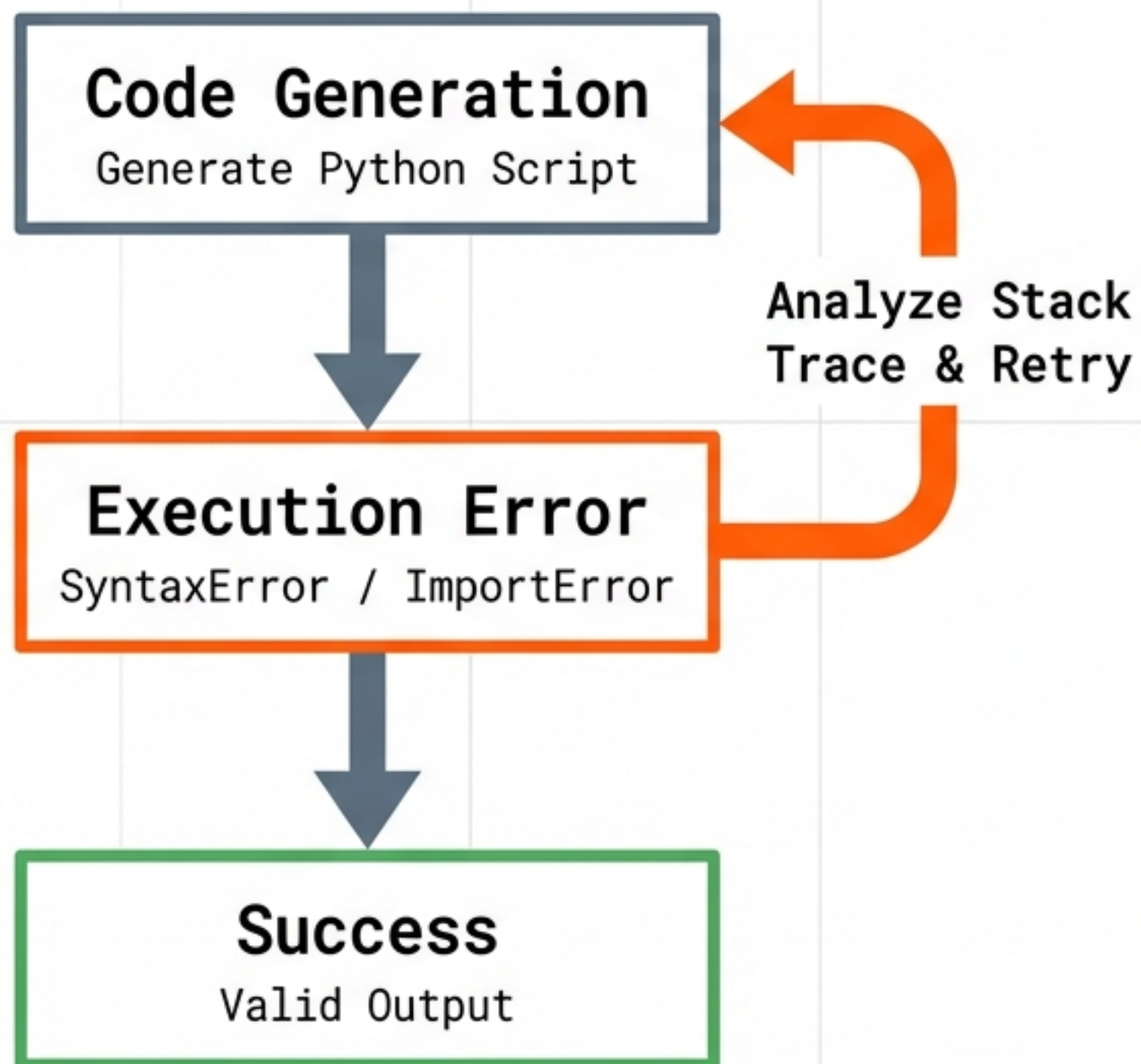
```
Terminal

>>> coins = detect_coins(image)
>>> count = len(coins)
>>> print(f"Total count: {count}")
Total count: 42
```

推論 (Probabilistic) から  
計算 (Deterministic) へ。

AIは対象にバウンディングボックスを描画し、`len(boxes)`  
のようにPythonで数値を計算するため、数え間違い (ハル  
シネーション) が起きない。

## 機能詳細③：エラー自己修復ループ (Self-Correction)

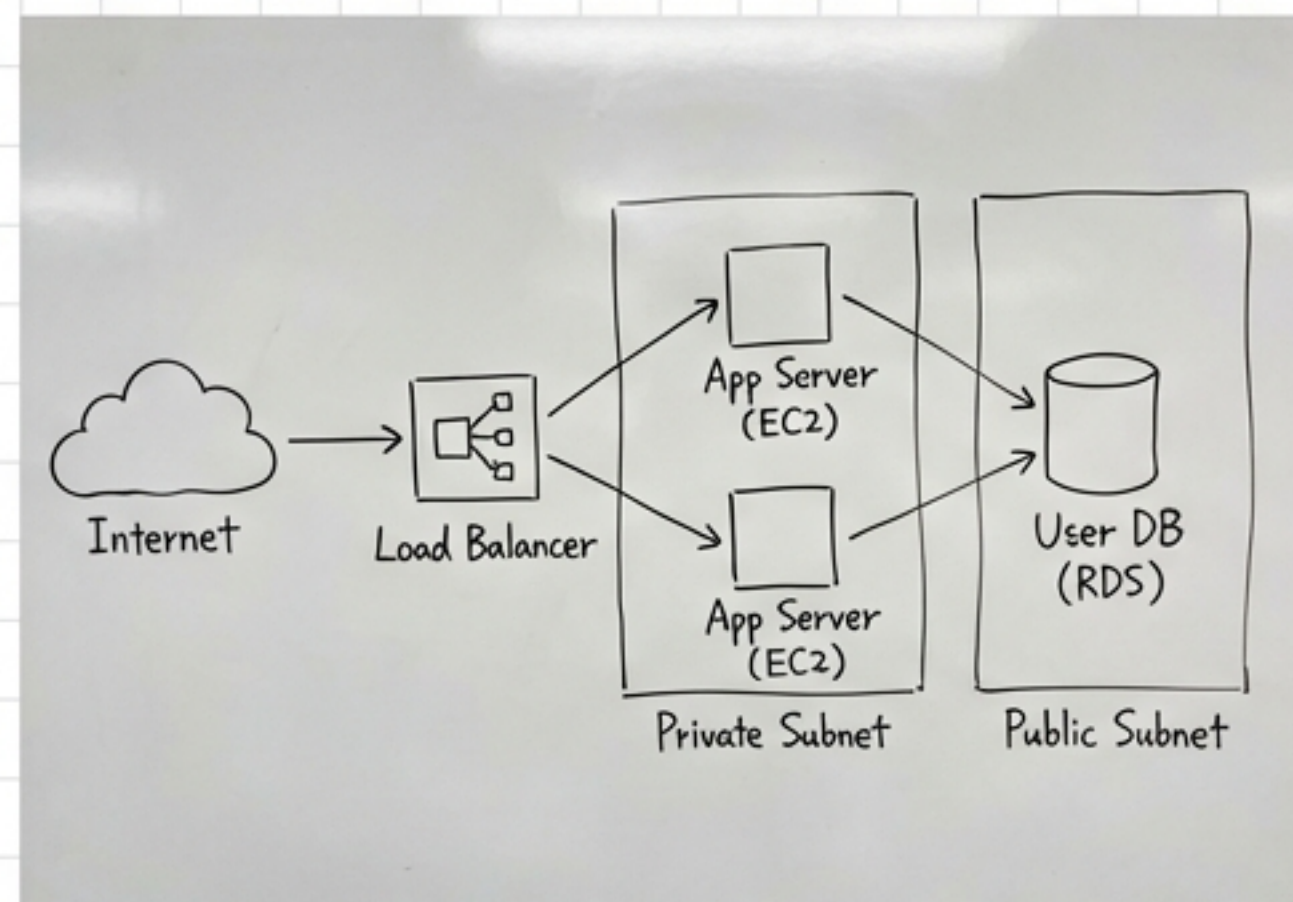


**問題:** 生成されたコードが構文エラーやライブラリエラーを起こすことは避けられない。

**解決:** Geminiはエラーのスタックトレースを読み取り、自律的にコードを修正して再実行する（最大3回まで）。

# ユースケース A：設計レビューの自動化と構造化

## Role: The Auditor (監査役)



Component	Type	Connection	Security Risk
User DB	Database	Public Subnet	HIGH (No Encryption)
App Server	EC2	Private Subnet	Low

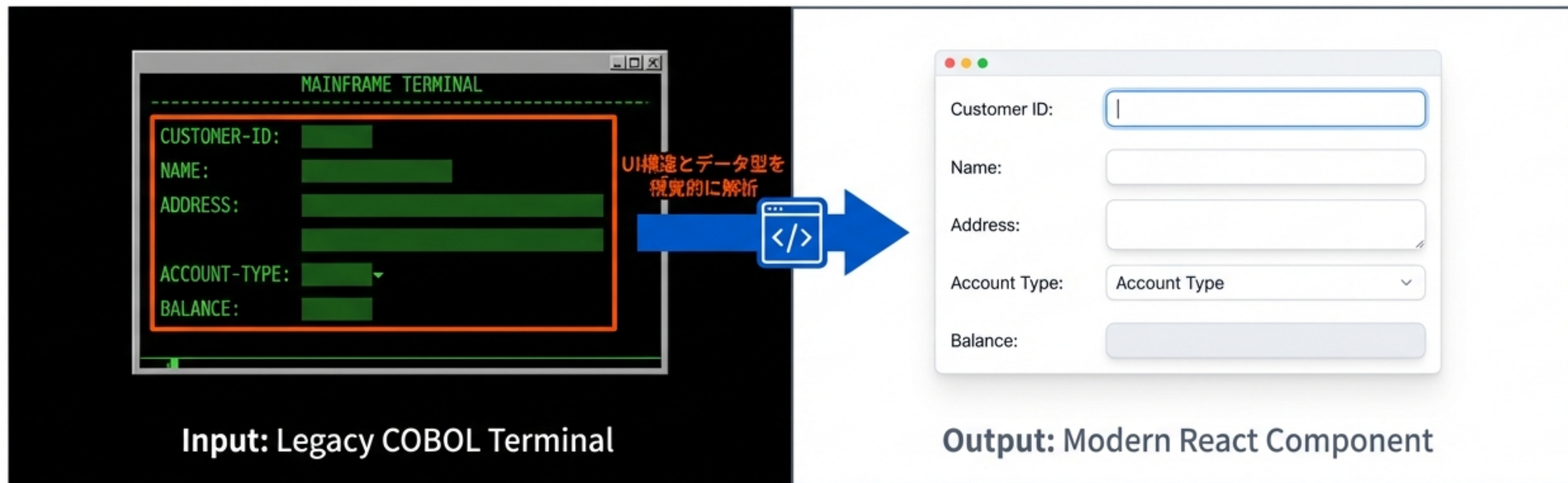
Workflow: アーキテクチャ図を入力 → ノードと接続を抽出して構造化データ (CSV) に変換 →

「DBが公開されていないか？」などのルールベース検証を自動実行。

Benefit: 図面という「非構造化データ」を、機械が検証可能な「構造化データ」へ変換。

# ユースケース B : レガシーシステムの現代化 (Legacy Modernization)

## Role: The Translator (翻訳家)



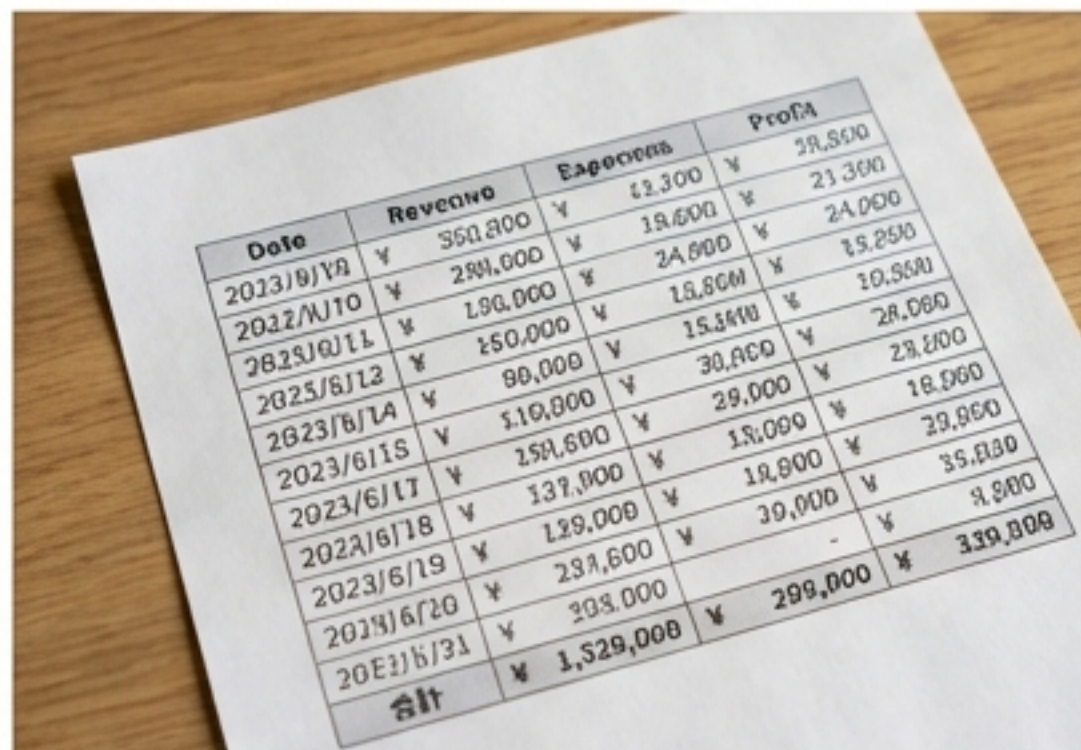
**Scenario:** 仕様書が存在しない1980年代の画面。

**Workflow:** スクリーンショットからUI構造とデータ型を視覚的に解析し、現代的なWebフレームワークのコードを生成。

**Benefit:** ドキュメント不足によるDXの停滞を打破。

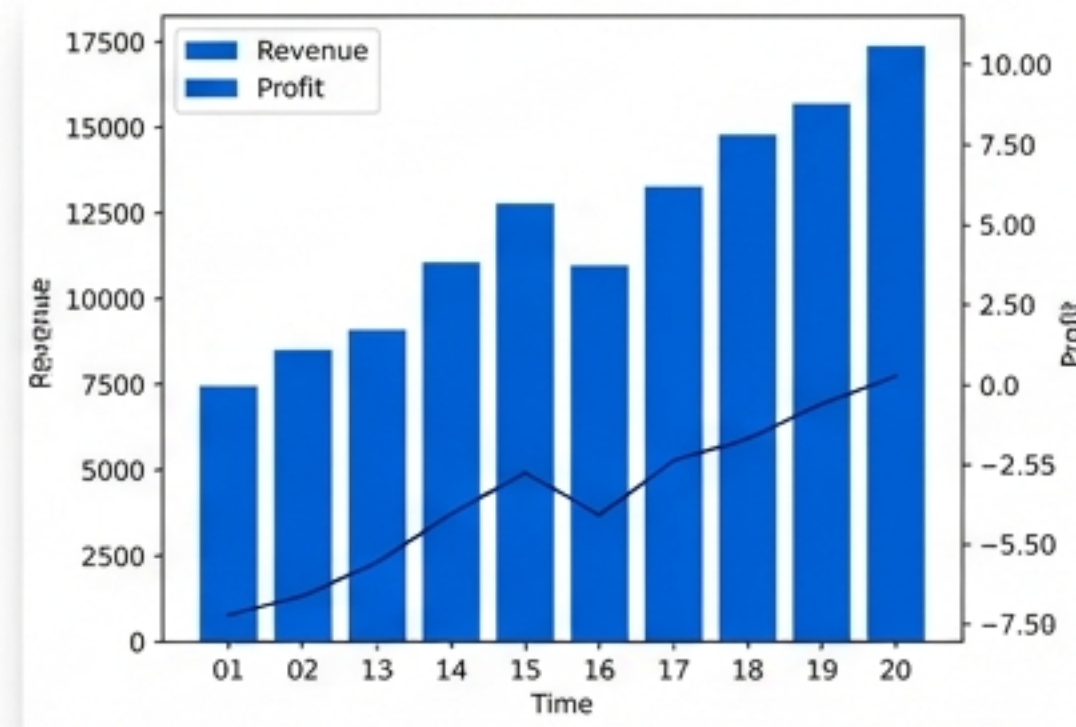
# ユースケース C：物理データのデジタル資産化

Role: The Analyst (分析官)



Date	Revenue	Expenses	Profit
2023/6/09	¥ 350,000	¥ 11,300	¥ 28,500
2022/6/10	¥ 280,000	¥ 19,600	¥ 21,300
2023/6/11	¥ 290,000	¥ 24,000	¥ 24,000
2023/6/12	¥ 150,000	¥ 15,500	¥ 15,250
2023/6/13	¥ 80,000	¥ 30,000	¥ 10,500
2023/6/14	¥ 110,000	¥ 29,000	¥ 28,000
2023/6/15	¥ 250,000	¥ 18,000	¥ 28,200
2023/6/17	¥ 137,000	¥ 19,900	¥ 18,000
2023/6/18	¥ 129,000	¥ 30,000	¥ 35,000
2023/6/19	¥ 231,600	-	¥ 9,000
2023/6/20	¥ 303,000	¥ 299,000	¥ 339,000
合計	¥ 1,529,000	¥ 299,000	¥ 339,000

OCR + Pandas + Matplotlib

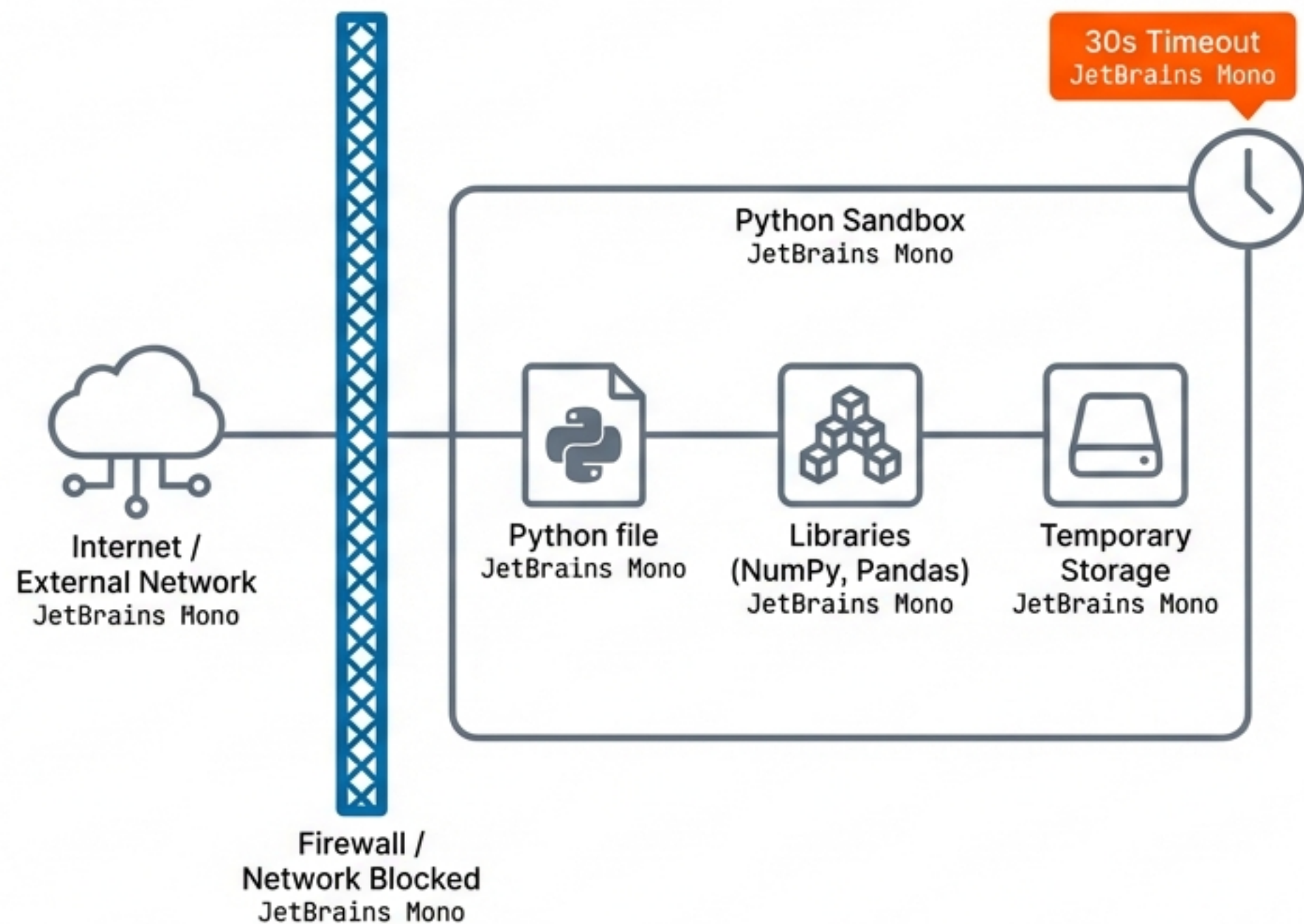


**Scenario:** 紙の財務レポート、ホワイトボードの手書き表。

**Workflow:** 画像から数値を読み取りPandas DataFrameへ格納。基本統計量を計算し、Matplotlibでその場でグラフ画像を描画。

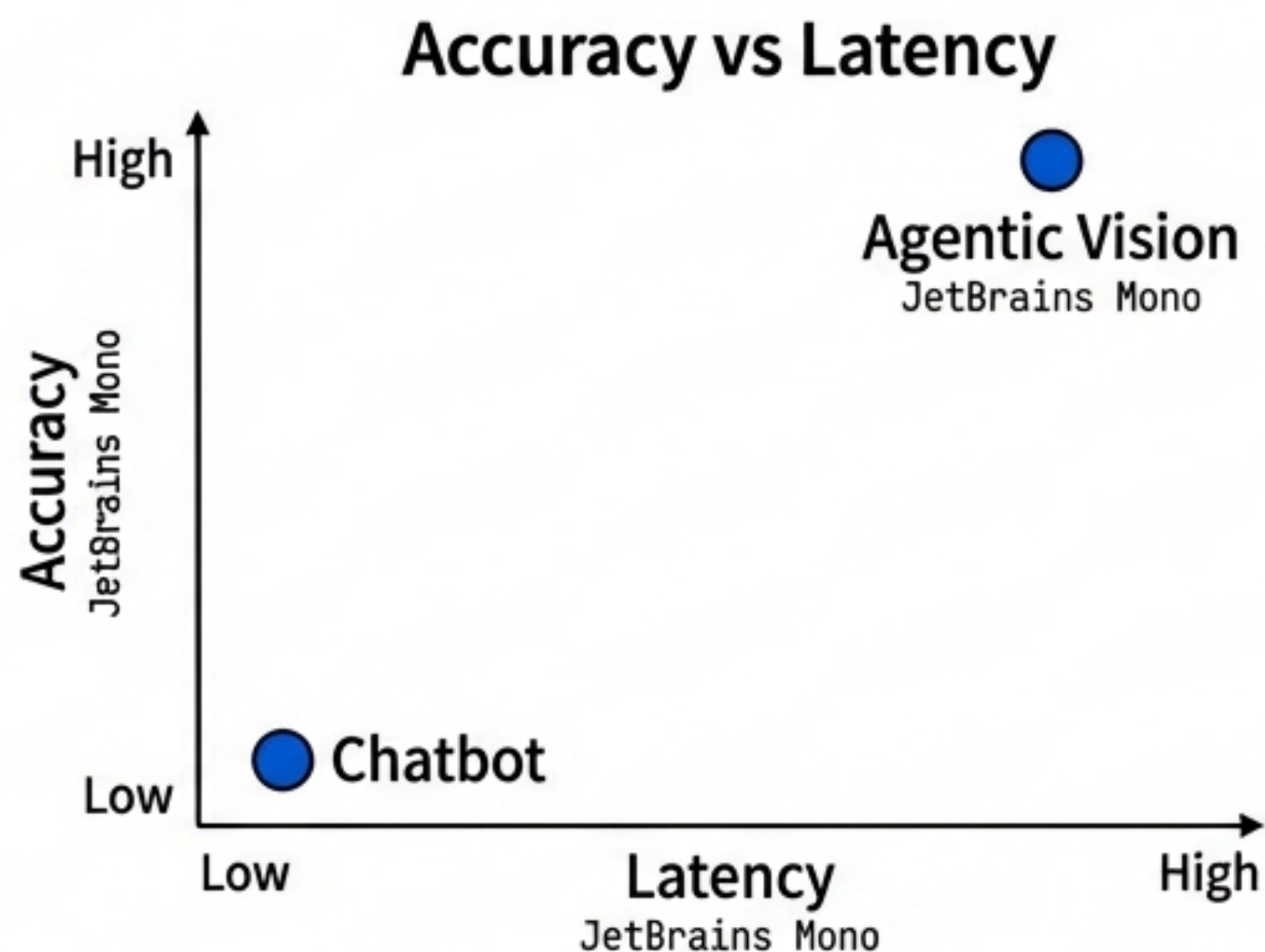
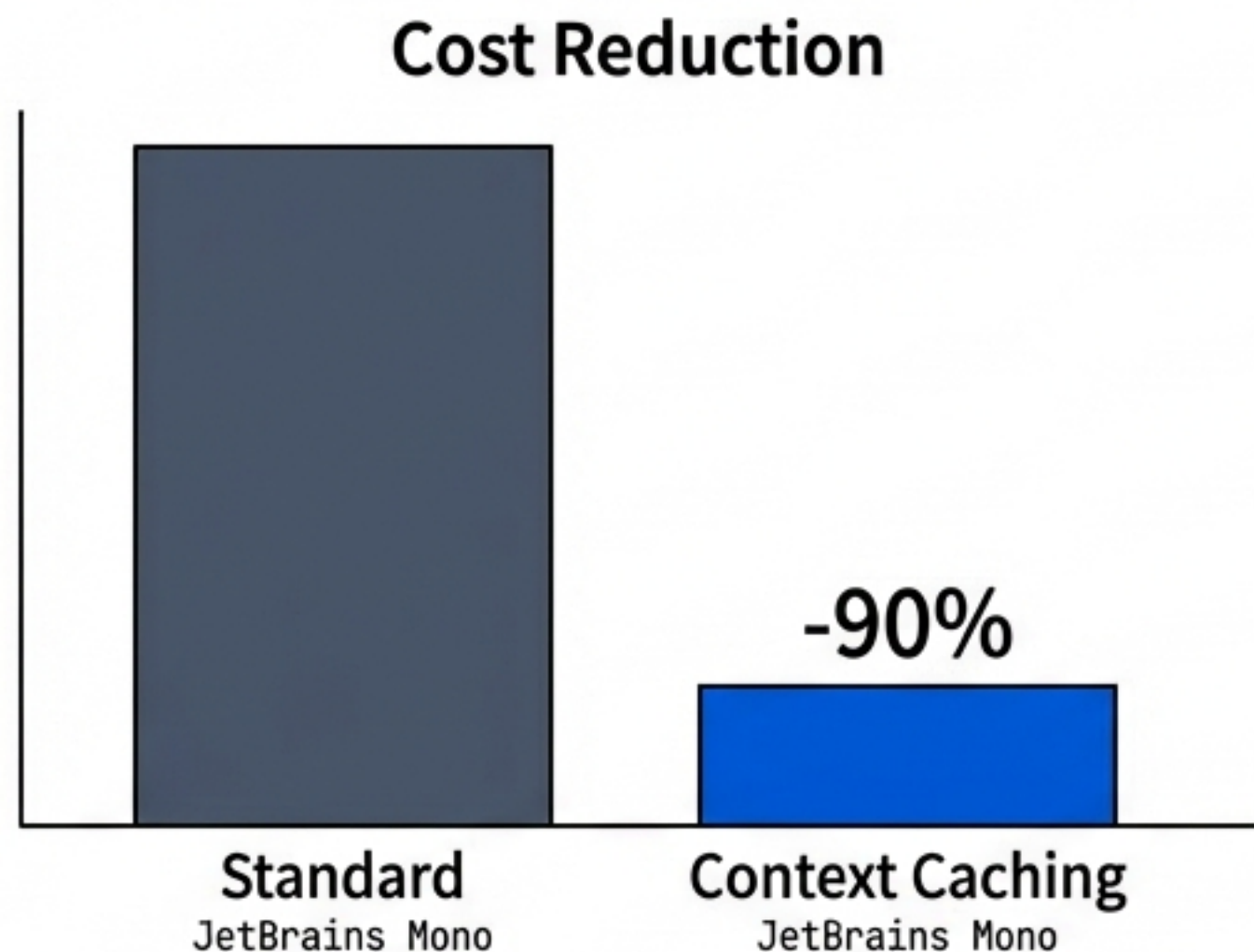
**Benefit:** 「ただの画像」から「意思決定のためのインサイト」を数秒で引き出す。

# アーキテクチャとセキュリティ：安全なサンドボックス実行



- Python Sandbox: ネットワークアクセスが遮断された隔離環境。データ流出リスクなし。
- Stateless: 各実行は独立的だが、セッション内でコンテキストを維持可能。
- Libraries: NumPy, Pandas, Matplotlib, SymPyなどの科学計算ライブラリをプリインストール。

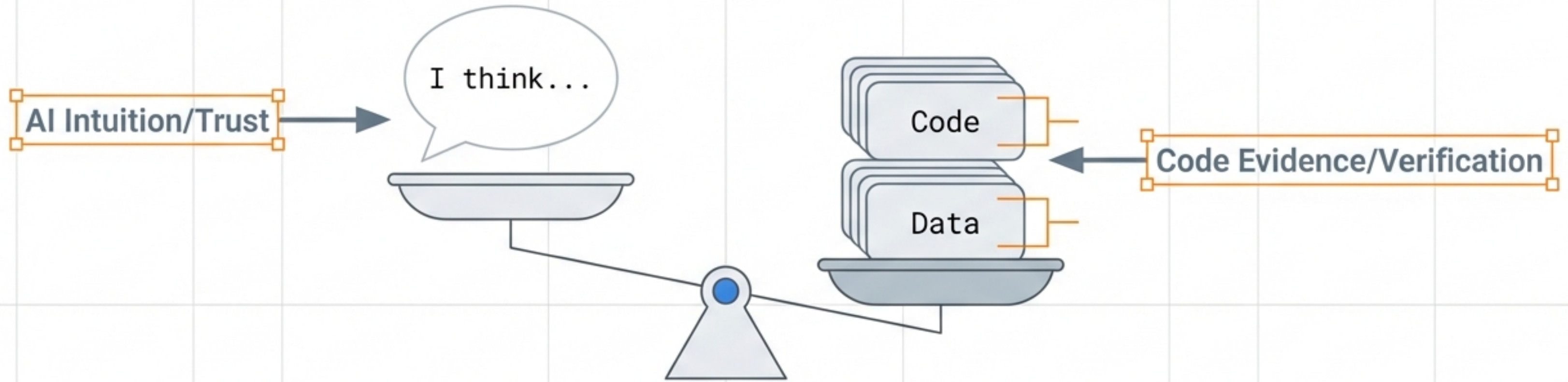
# コスト戦略とパフォーマンスの最適化



**Cost Efficiency:** 繰り返し参照する画像やルールセットをキャッシュすることで、コストを最大90%削減可能。

**Trade-off:** ループ処理により応答時間は増加するが、精度と検証可能性は飛躍的に向上する。即答性より正確性が求められるバックオフィス業務に最適。

# 戦略的示唆：「説明」から「証明」へのパラダイムシフト

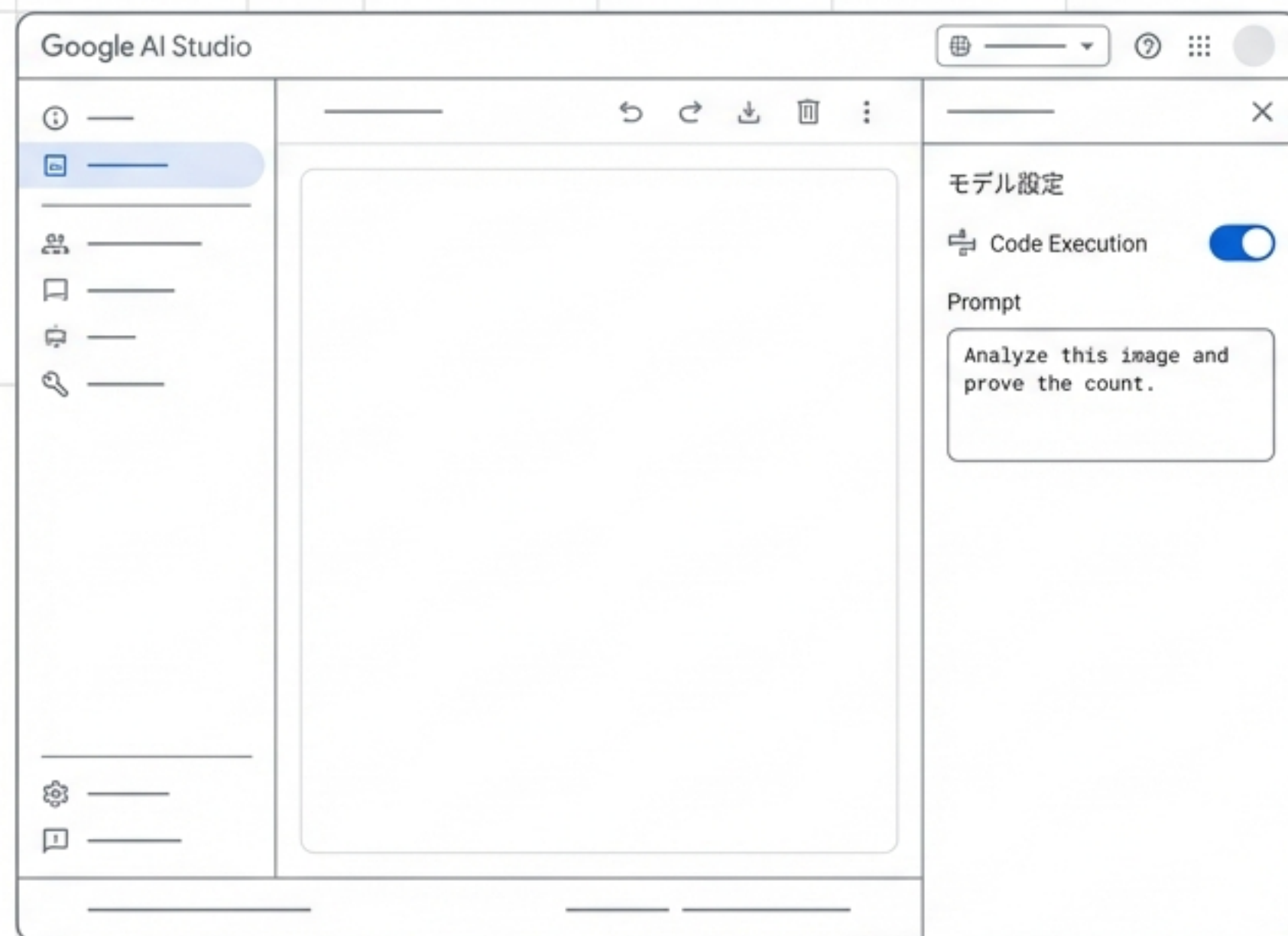


The Big Picture: Agentic Visionの真価は、機能ではなく「信頼の構造変化」にある。

- Before: 「AIがそう言っているから正しいはず」 (信頼ベース)
- After: 「計算過程とコードが示されているから正しい」 (検証ベース)

Prediction: 2027年までに、高責任領域（法務、医療、エンジニアリング）では、コードによる裏付け（Evidence Layer）を持つAIエージェントが標準となる。

# 次のアクション：Google AI Studioで「調査」を開始する



1. **Try:** Google AI Studioで `Code Execution` を有効化する。
2. **Challenge:** 従来は見えなかった「小さな文字」や「複雑な計算」を画像で投げしてみる。
3. **Deploy:** Vertex AI経由で業務ワークフロー（レビュー、データ入力）に組み込む。

The AI that investigates.