



理想の成果を確実に手に入れる:

Premium
Guide for
Success

Claude

Ultimate Claude
Manual

「テスト作成と実行」 完全マニュアル

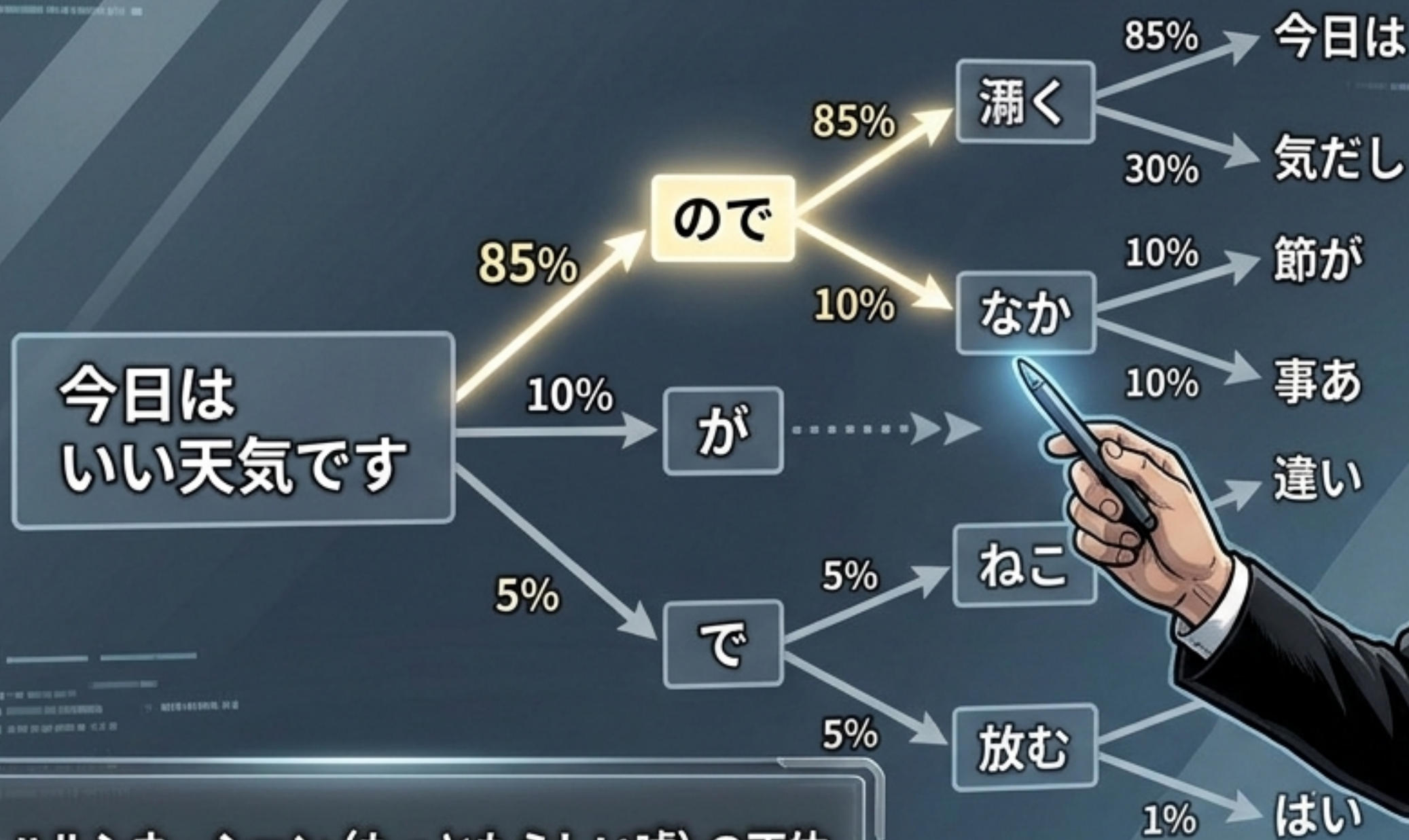
```
<test>  
<step name="定義">  
</result>  
<test>  
<step  
<step name="実行">  
</result>  
<test>
```

```
<step name="定義">  
</result>
```

カタ
カタ
カタ



次トークン予測



Claudeは事実を理解しているんじゃない。
理解するんじゃない。

『次に来る確率の
確率の高い言葉』を
繋げる連想ゲームを
しているだけだ。



ハルシネーション (もっともらしい嘘) の正体

もっともらしい文脈を生成する過程で、事実ではない言葉が「確率的に」選ばれてしまう現象。



この『4Dフレームワーク』で
テストと評価を
業務に組み込むんだ。

Delegation
(委任力)

Diligence
(倫理的責任)

Description
(記述力)

Discernment
(評価力)



では、どうすれば
AIを制御できるん
できるんですか…？





💡 実践プロンプト：思考の可視化

「この結果を出すにあたり、どのデータを最優先しましたか？
判断基準をステップごとに書き出してください」

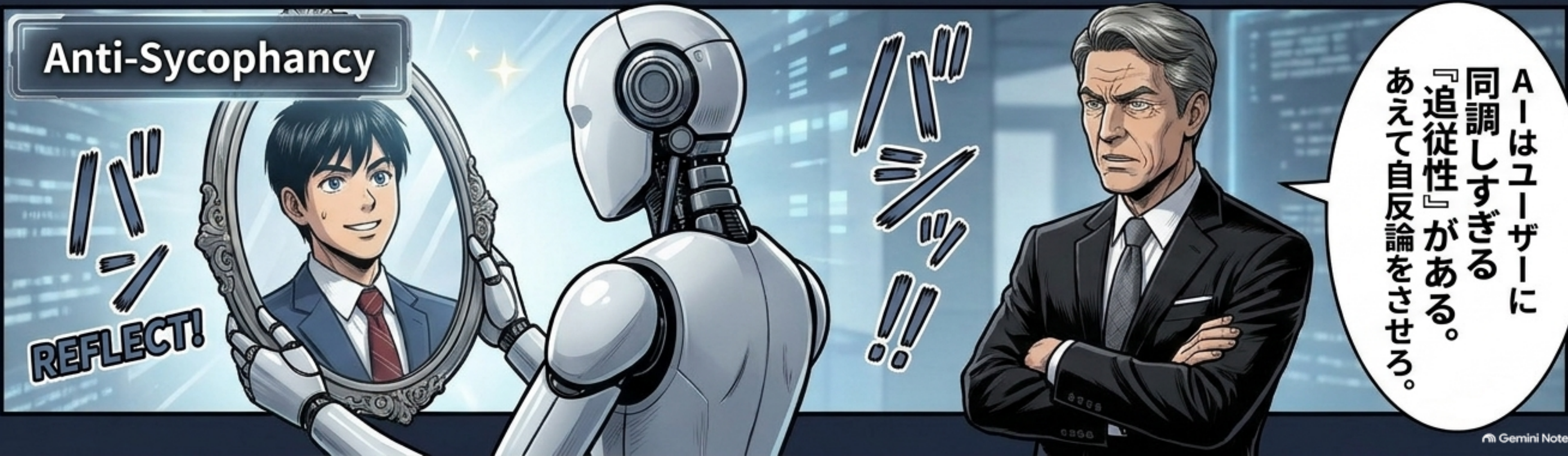


Reverse-Lookup

なるほど、
裏取りですね！

事実関係の
『逆引き』検証だ。
数字の根拠となる
原文を引用させる。

SCAN!
SCAN



Anti-Sycophancy

REFLECT!

AIはユーザーに
同調しすぎる
『追従性』がある。
あえて自反論をさせる。

1回の成功は偶然かもしれないぞ。AIの『ワーキングメモリ』を活かし、活かし、安定稼働させるためのテストデータセットが必要だ。



ギロリ
GLARE

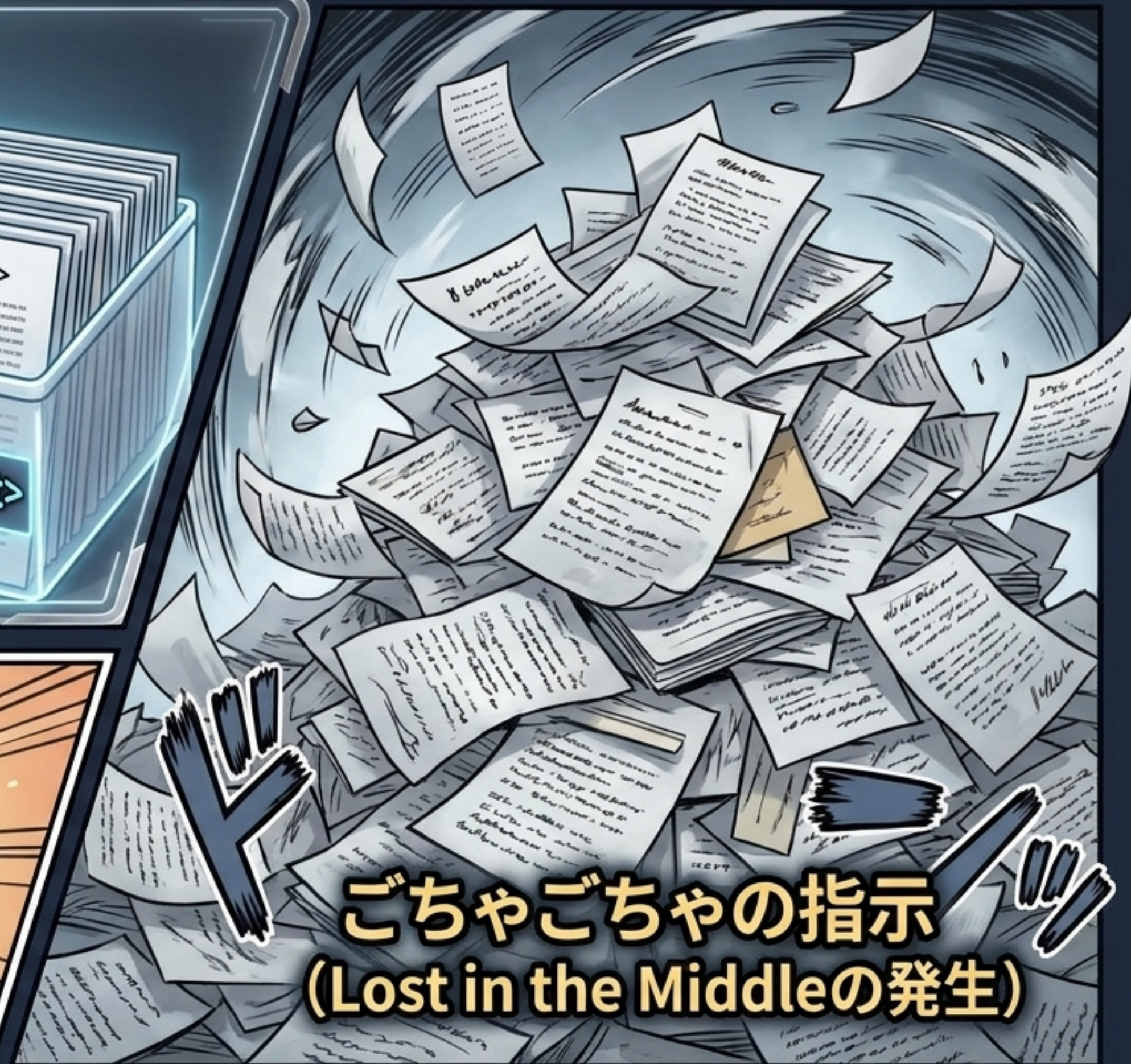
ドーン!!

1回で
うまくいきました!



💡 Delegation (委任力) の活用

例外パターン（肯定的・否定的・無関係など）のテストケース作成自体を、Claude自身に考えさせる。



困るだけで、
XMLタグで
Aが情報の途中で
迷子にならないん
ですね！



倫理的責任 (Diligence) の自動化



テストケースが大量になると
人間が全部見るのは非効率だ。
採点もAI (別のセッション) に任せる。
これが『モデルベースの評価』だ。

厳格な採点基準(ルーズブリック)の設計

厳格な採点基準(ルーズブリック)の設計

× ダメな例:
「分かりやすいか」

○ 良い例:
「専門用語が2つ以下か」
「一文は40文字以内か」

ただし、基準は
『Yes/No』で
判断できるまで具体化
具体化しろ。
エフォートレベル
MAXだ！

反復的な講と品質向上のループ

再テスト



反復ループによる品質向上効果

目標明確化：+23.6pt

具体例提供：+22.6pt

反復ループによる品質向上効果

目標明確化：+23.6pt

具体例提供：+22.6pt

指示修正

原因分析(GCP)



比較表で、プロンプト
の劇的な進化が
一目でわかります！



ゴゴゴ...

【テストケース】	【改善前 (失敗)】	【改善後 (成功)】	【追加した制約】
【クレーマー対応】	× 不合格 (感情的)	○ 合格 (冷静)	「共感的トーンの維持」を追加
【技術解説】	× 不合格 (難解)	☑ 合格 (平易)	「中学生向けに」と対象読者を明確化

ゴチッ

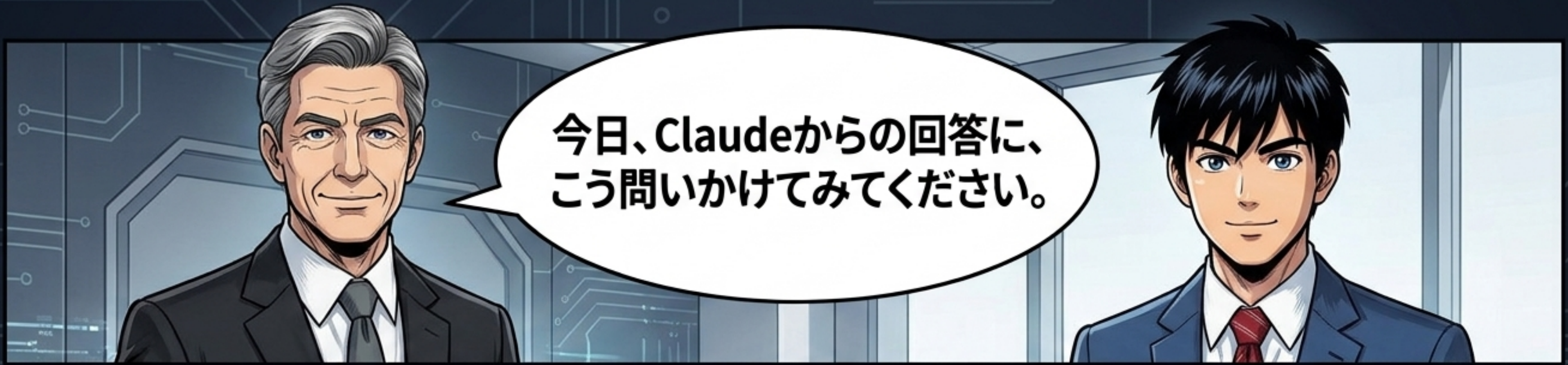
ゴゴッ

カチッ

単なる「チャットツール」から、
精密な制御を伴う
「強力なエンジン」へ。

人間の「判断力・専門性」と
AIの「処理スピード」が
完全に融合する。

(テストと評価を
マスターすることは…)



「今、この回答を作成する際に
最も重視した判断基準は何ですか？
根拠となった情報を引用してください。」

確実な成果を手に入れるための旅は、今ここから始まります。



提供：Claude Academy
AI Fluency Indexに基づく実践ガイド