

なんでもかんでもRAGじゃない

よくある相談パターン

1. よく耳にするのは、「RAGを入れないとダメですよね？」という声。
2. けれど実際は、プロンプトだけで綺麗に回せるケースがたくさんあります。
3. 内容がパターン化できるなら、むしろRAGより精度が上がることも珍しくありません。

問い合わせ①：全部参照する？

- 最初に考えるのは、「AIに全部の情報を見てほしいのかどうか」。
- もし“全部読んだうえで判断してほしい”なら、部分検索であるRAGはあまり向いていません。
- 全文を読ませるアプローチのほうが自然です。

問い合わせ②：入力上限に収まる？

1. GPTやClaudeは10～20万字、Geminiなら100万字以上入る時代です。
2. 実は、全文を入れてもコストはごくわずか。
3. RAGを作り込むより、速くて安定していることが多いんですね。

問い合わせ③：パターンは何個？

- もしパターンが10～20種類くらいなら、テンプレート化が最適です。
- 業界別に作っておけば使い回しも簡単で、精度も運用も申し分ありません。

3つの問い合わせ

1. "全部見たい"ならそのままプロンプトへ。
2. "上限に収まる"なら全文を入れる。
3. "パターンが少ない"ならテンプレ化。
4. この条件がそろえれば、RAGは必要ありません。
5. もっとシンプルで強い方法があるからです。

RAG=正義ではない

1. RAGは万能ではありません。
2. 適材適所でこそ輝く技術です。
3. 「どの状況でもRAGが正解」というわけではないんです。

今日のポイント

1. RAGは“部分を取り出す技術”。
2. 全部読む用途には向きません。
3. 入力が収まるなら全文投入でOK。
4. パターンが少ないならテンプレ化が最善。
5. 要は、状況に合わせて賢く選ぶことが大切なんです。

RAGを使わない勇気を

1. 勢いでRAGを作つて失敗する例は少なくありません。
2. まずは「本当に必要か」を落ち着いて見極めること。
3. RAGを“使わない”判断こそ、自社データ活用の成功を近づける大事な一步です。