

なんでもかんでも

RAGじゃない!

よくある相談パターン

- 自社データ=RAGという思い込み
- 実はプロンプト設計で十分
- パターン化で解決できる
- RAG導入前の3つの問い

問い①：全部参照する？

問い②：入力上限に収まる？

- GPT/Claude：10～20万文字
- Gemini：100万文字以上
- 全部入れても低コスト
- 10万文字で約10円
- RAG構築より安く確実

問い③：パターンは何個？

- 10～20個程度なら
- テンプレート化が最適
- 業界別プロンプト作成
- 1日で完成、使い回し可能
- RAGより精度高く運用楽

3つの問いまとめ

- 全部見たい？→プロンプト
- AIで読める量？→全部入れる
- パターン少ない？→テンプレ化
- 当てはまるならRAG不要
- プロンプト設計が早く安い

RAG=正義ではない

今日のポイント

- RAGは一部を引っ張る仕組み
- 全部見る用途には不向き
- 上限内なら全部入れる
- パターン少ないならテンプレ化
- 適材適所で使い分ける

RAGを使わない勇気を

無駄にRAGを作って成果が出ない失敗を防ぐ。自社データ活用の第一歩として適切な選択を。