

目次

- RAG過信への警鐘と本稿の目的
- 問1:網羅的な参照は必要か
- 問2:元データは入力上限内か
- 問3:パターン数は何個か
- 実務への落とし込み

RAG過信への警鐘と本稿の目的

- 「自社データ=RAG」は早合点
- RAG不要の案件が多数存在
- プロンプト設計のみで解決可能
- RAG導入前の3つの問を確認
- 簡便で高精度な代替を検討

問1:網羅的な参照は必要か

- チェックリスト型はRAG不適
- 全観点をプロンプトに明示
- ルール型はプログラム化
- 解釈型は生成AIで判断
- 20項目以上は分割実行

問2:元データは入力上限内か

- 上限内なら全投入が最適
- 抽出漏れ・取り違えを回避
- 数百回程度なら総コスト有利
- 大量トランザクションで再検討
- RAGは巨大データ向き

問3:パターン数は何個か

- 10～20個ならテンプレ運用
- 業界別テンプレートを作成
- 人が選択する方が高精度
- RAGの初期コストを回避
- パターン管理で開発軽量化

実務への落とし込み

- 網羅チェック必要→RAG不要
- 上限内データ→全投入
- 管理可能パターン→テンプレ
- RAG導入は最後の選択肢
- 失敗回避は観点の明示化

本日のポイント

- RAGは必要部分だけ引く道具
- 網羅チェックには不向き
- 上限内は全投入が最適
- RAGは巨大・高頻度向き
- それ以外はプロンプト設計