

RAG

手軽にLLMの精度UP!

Retrieval-Augmented Generation

[検索]

[拡張]

[生成]

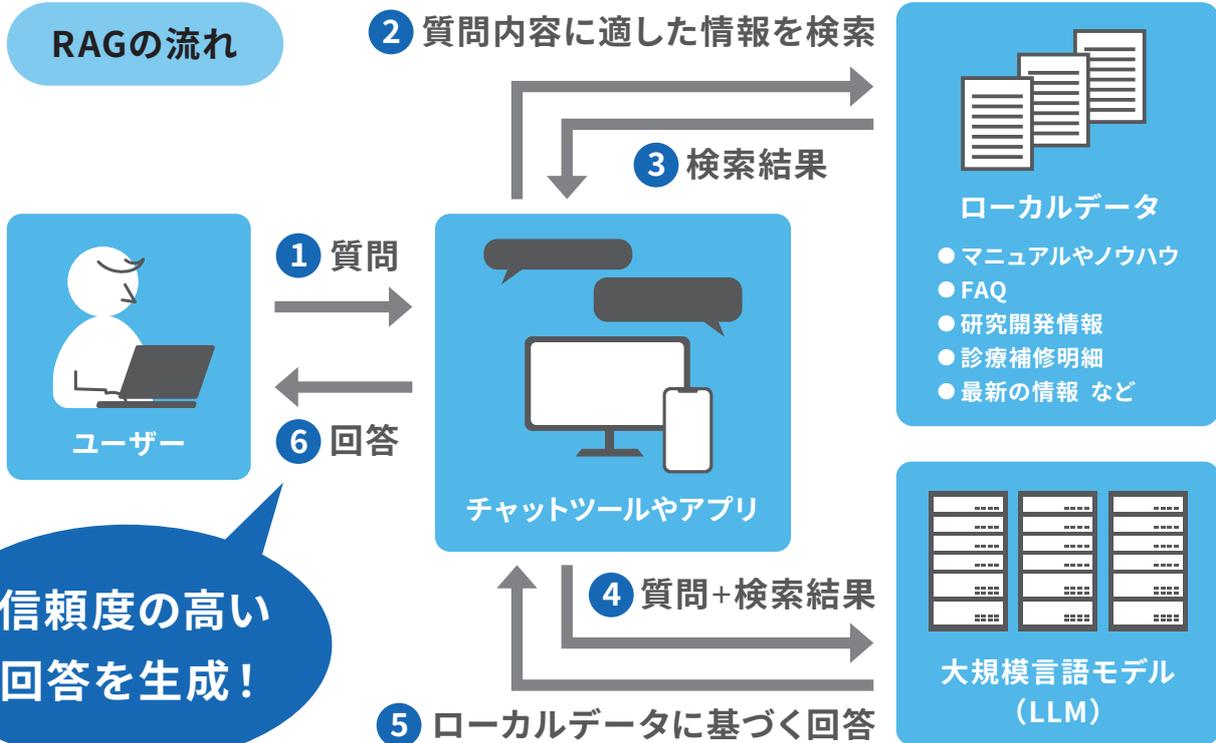
LLM出力結果を簡単に最新情報に更新! 根拠明確化でハルシネーション抑制!

[RAGとは]

もちろんPhi-3などのSLMでの活用も!

大規模言語モデル(LLM)に、ローカルデータへの検索を組み合わせることで、正確な情報に基づいた回答を生成させる技術です。

RAGの流れ



[RAGのメリット]

- 1) LLMが出力する結果を、手軽に最新情報に更新することができる
- 2) 信頼度の高いデータから情報を得るため、AIが学習データと相容れない誤情報を生成する“ハルシネーション(幻覚)”の抑制につながる

➔ **LLMの回答精度を向上させます!**

RAGを使わずローカルデータを反映させる方法にはファインチューニングがありますが...

- ・ ローカルデータによる追加学習(再学習): Deep Learningで行われるようなファインチューニングでは、回答にローカルデータが確実に選択される保証がない!
- ・ 大規模モデルでは処理に多大なりソースが必要。故に、気軽に再学習できないので学習データは古くなりがち。

最先端のRAGをオンプレミスで! あなたのデータを最大限に活用

最新AI技術のRAGをオンプレミス環境で導入しセキュリティとパフォーマンスを両立

[RAG×オンプレ 主なメリット]

- セキュリティ** クエリとデータが社内に留まるため、最高レベルのセキュリティを実現
- パフォーマンス** オンプレミス環境での高いパフォーマンス
- カスタマイズ** 企業のニーズに合わせた柔軟なカスタマイズが可能
- コスト** クラウドサービス費用の削減

[RAG×オンプレ 技術的な特徴]

- RAG** 高精度な情報検索と生成機能を統合し、より正確で有用な情報提供を実現
- オンプレミス**
 - ・自社のインフラを活用し、安定した運用と高速なアクセスを提供
 - ・ローカルデータの任意のタイミングでの更新に対応

blueqat RAG

オンプレ環境に
構築ご相談ください

- 最新のさまざまなカスタマイズが可能
- クラウド環境を活用する、様々なハイブリッド構成も可能

[導入プロセス]

お問い合わせから導入までの流れ

1.コンサルティング > 2.カスタマイズ > 3.導入 > 4.サポート

詳細な情報やデモのご依頼は、下記の連絡先までお気軽にお問い合わせください。

お問い合わせ 弊社サイトのお問い合わせフォームへご入力いただくか、下記までご連絡ください

メール vt-sales@v-t.co.jp **TEL** 03-6823-6789 (受付時間 平日10:00~17:00) **お問い合わせ**
フォーム ▶

