



基本編

生成AIを業務に取り入れたい。しかし、顧客情報や契約書、社内の在庫データを外部サービスに送るわけにはいかない……。これは特に、専任のAI部門を持たない中小企業のIT担当者にとって頭の痛い問題だ。その解決策のひとつが、データを社外に出さず、モデルの利用に月額料金もかからない“自社専用AI”だ。ただし、環境を作っただけでは仕事は変わらない。本書では、中小企業のIT担当者の視点で、「営業」、「在庫管理」、「リサーチ」の3部門の課題を解決していく。

ローカルLLMならデータを自社内で処理できる！

LLM（大規模言語モデル）は、人の指示を理解し、文章の作成や要約、翻訳、質問への回答などを行なう生成AIの中核だ。クラウド型サービスでは、入力した文章をインターネット経由で事業者のサーバーへ送り、処理された結果を受け取る。一方のローカルLLMは、モデルを自社のPCやサーバーで動かし、入力から回答の生成までを社内ですべて完結させる。

違いは大きく2つある。1つ目はデータを処理する場所だ。ローカルLLMなら、社外秘の資料や顧客情報を外部サービスへ送らずに活用できる。もちろん権限などの管理は別途必要だが、「外に出せないため生成AIに入力できなかった情報」を扱えるメリットは大きい。

2つ目はコストのかかり方だ。クラウド型は月額ライセンスやAPIの利用量に応じた料金が継続して発生する。ローカルLLMはPCやGPUへの初期投資が必要な半面、モデルを何回呼び出しても利用料は増えない。電気代や保守費用はかかるものの、社員数や利用頻度に比例して月額料金が膨らまないため、継続的に使うほど支出を予測しやすい。

ローカルLLMは、クラウド型サービスの代用品をそのまま社内に置くというより、自社のデータと業務に合わせて育てる基盤と考えると分かりやすい。

モデル単体は社内固有のルールや最新の在庫を知らない。そこで、文書を検索する機能や外部処理を呼び出す仕組みを与え、社員の質問に必要な情報をその都度加える。最終判断を人が担うことを前提に、「検索」、「整理」、「翻訳」、「下書き」などといった時間のかかる作業をAIに任せるわけだ。

ローカルLLMの環境は、次のような層で構成される。



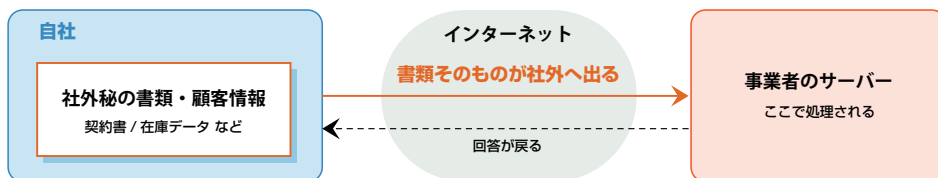
社外秘データはクラウドAIに入力できないこともある
企業によっては顧客情報や契約書は外部のAIサービスに送れない。生成AI活用でまず突き当たる壁だ

- ・学習済みのベースモデル
- ・モデルを動かす実行環境
- ・社内文書を参照するRAG（検索拡張生成）
- ・社員がブラウザから利用するUI

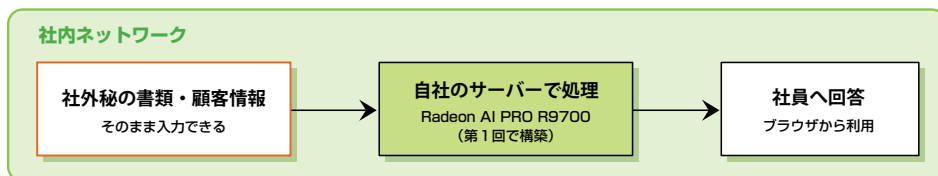
RAGは質問に関連する文書を検索し、その内容をLLMに渡して回答を作る仕組みだ。本稿で扱う3つの実用例でも、これらの土台を目的に応じて拡張していく。

図1 データを処理する場所—クラウド型とローカルLLMの違い

クラウド型サービス



ローカルLLM



入力から回答まで社内ですべて完結 — 枠の外へ出る矢印が1本もない