

事例6選! …生成AIを調査・分析に本格利用する

西川 仁

筆者は長年、知的財産関係の仕事をしてきました。知的財産の仕事は、膨大な特許データを扱います。特許データはIT化が進み広範囲で複雑な検索が可能です(特許情報が得られるウェブ・ページとしてはj-PlatPatやGoogle Patentなどがある)。しかし、対象特許群の絞り込みは簡単にできるようになったものの、隠れた技術や応用のヒントは特許を読まない限り見つかりません。特許特有の文体と文章構成と、膨大な文章量が素人を遠ざけてきました。企業の特許担当者や弁理士の専門領域です。

難解な特許のひとつをAIでできないか模索してきましたが、2022年のChatGPTの出現で一気に加速しました。本稿では、業務での生成AI活用について、現時点までの業務適用経験を元に紹介します。興味のある人は、自分の仕事で発生するデータを重ね合わせてみてください。共通点は見いだせるはずですが、

今回、利用した生成AIのタイプは、リスト中に[GPT-4]と書かれているものはOpenAI API経由による最新版GPT-4(2023年11月版)です。[Copilot]はマイクロソフト Copilot(無償版)です。

事例1 大量のドキュメントを要約させる

ここでの例は、特許調査において作成した特許集合から、詳細調査対象の特許を絞り込むための簡易な要約を生成AIに行わせるものです。全固体電池関連特許129件の課題と解決手段のポイントを抽出させています。特許データは前述の特許情報関係のウェブ・ページで検索(無償)できます。また、読者の会社で特許検索システムがあれば、そこでも特許データを検索できるはずですが、

● 方法…特許の対象項目をChatGPTのプロンプトに貼り付けるだけ

生成AIの扱えるトークン量には制限があるので、この例では、特許の対象項目(課題と解決手段が書かれている要約項目)だけをExcelからコピーしてプロンプトにペーストしています。ちなみに要約、課題と解決手段の3つのデータ項目を取り込むと、特許を50件に絞っても52,621文字になります。これを要約項目データだけにすると、129件全部でも24,731文字となり、最新版のGPT-4であれば1プロンプト範囲でも扱いやすくなります。

▶ 一気に要約させず10件単位で区切る

とはいえ、129件全部を一気に要約させようとする

と、生成AIの方で出力を自動的に抑えてしまいます。生成AIは、複雑で負荷が大きい処理になってしまう場合、出力を自動的に抑制しているようです。ここでの事例では少々手間はかかりますが、プロンプトで処理範囲を設定できるようにして、10件単位で区切って生成させるようにしています。そのため特許には事前に通し番号を振っています。

プロンプトでは単純にポイントの抽出としていますが、読者の観点(例:○○の観点でポイントを抽出)をプロンプトに追加できれば、自分のこだわりに近い回答が得やすくなります。

● 要約させた結果

リスト1-1(a)に示す[入力]はプロンプトです。見慣れない表記の仕方もかもしれませんが、気にしないでください。プロンプト表記のヒントについては後ほど説明しています。リスト1-1(b)がGPT-4の出力です。

▶ 解決手段の方法で分けた場合

さらに、製造手段以外の特許にフォーカスしたいので、リスト1-2では、リスト1-1の要約を基に解決手段が「製造手段」か「製造手段以外」かを仕分けさせています。リスト1-2(b)では誌面の都合上、途中を省