背景：

現在、「サプライチェーン評価事業」の準備を進めています。

この事業の登場人物は、

「自社のｃを評価したい事業者（お客様）」と

「サプライチェーンに組み込まれているため評価される側の事業者（対象者/複数）」と

「評価するための作業一式を請け負う事業者（ｃ）」

の３者です。

大まかな事業フローは、以下の通り。

・お客様は、サプライチェーンを評価するための質問事項を日経リサーチ提供する。

・日経リサーチは、提供された項目を元に調査票を作成し、対象者に送付する。

・対象者は、調査票に回答すると同時に、記載内容の根拠になる情報を提供する。

・日経リサーチは、回答内容と、根拠になる情報の記載を確認する。

・お客様は日経リサーチから提供された調査結果によりサプライチェーン評価を完了する。

このフローにおいては、以下のシステムは開発が別途進行中。

①アンケート調査票（エクセル）の送付、回収、データ化のシステム

調査票のメール送付、エクセルファイルのアップロード、保管

②記載内容の根拠情報（エビデンス）の回収、管理システム

xls,doc,ppt,pdf等のファイルと、WEBサイトURLなど情報のアップロード＆保管

③調査票からデータ抽出をし、設問とエビデンスとのつながりを管理/出力する

「案件12のA社回答で」「設問１の根拠はxxxxx.xls」「設問2の根拠はhttps://www.…」

（その他、エクセルやVBAで当面は実行可能なものはビジネスの稼働後に開発。）

今回、貴社に検討開発を依頼したいと考えているのは、以下の仕組みです。

例えば、調査の内容に

設問：「サステナビリティに関するトップコミットメントメッセージが発信されている」

であった場合は

上記③のシステムから出力された情報をもとに

「当該ファイルorWEBサイトに、その情報が記載されているか」

を判定し、判定の根拠になった文言や情報を合致度順に複数見つけ

それらをファイルとして出力する仕組み。

★「追加学習」に回せるほどのデータはたまっていない。

★設問文の複雑さを考えると、プロンプトで対応できる内容ではないかという想定。

★プロンプトでFew(Many)-Shotができるくらいの知見はあるので、LLMを噛ませて検索用の言葉を作るなどの対応を検討。

この仕組みがなぜ必要かというと、

エビデンスとして渡された資料が膨大なページ数になる可能性が高いためです。

特にサステナビリティに関する内容は「統合報告書」と言われる資料に記載されることが多いのですが、

これは、百ページ超になることもあります。その内容をすべて目視して探すことが

人間の工数を必要としてしまい非効率なためです。

設問自体が、有無を問うものですので、出力内容は

「関連する記述がある（当該箇所を出力できる）」か「(みつから)ない」かで構いません。