

法令適用における検索拡張生成や検索プロンプト最適化の有効性の検証

丹羽 一絵^{1,a)} 内山 光彩^{1,b)} 秋本 一樹^{1,c)} 堀之内 響^{1,d)} 小野 智司^{1,e)}

概要: 大規模言語モデルは法務領域への適用が進む一方で、法令根拠の欠落に起因する誤答やハルシネーションが課題となる。検索拡張生成 (Retrieval-Augmented Generation: RAG) は外部知識の検索結果を根拠として回答を生成することでこれらの課題を緩和しうが、最終回答の性能は検索で参照される情報の品質に依存する。本研究では、検索クエリ設計および検索戦略の違いが回答品質に与える影響を実験的に検討する。具体的には、デジタル庁が公開する法令4択問題データセットを対象とし、プロンプト最適化によるクエリ生成を適用した RAG の有効性を検証する。

An Empirical Study on Retrieval-Augmented Generation and Search-Query Prompt Optimization for Legal Statute Application

Abstract: While large language models are increasingly being applied to the legal domain, incorrect answers and hallucinations due to insufficient grounding in legal authorities remain major challenges. Retrieval-Augmented Generation (RAG) can alleviate these issues by generating answers based on search results from external knowledge; however, answer quality depends on the quality of the information retrieved and referenced during generation. This study experimentally examines how differences in search query design and search strategy affect answer quality. Specifically, this study employs a dataset of multiple-choice legal questions released by the Japan's Digital Agency to verify the effectiveness of combining RAG and prompt optimization to generate retrieval queries.

1. はじめに

近年、大規模言語モデル (Large Language Models: LLM) の発展により、一般的な言語タスクのみならず、司法試験問題の解答や契約書レビューといった専門的な法務領域への応用が模索されている。しかし、LLM を単体で用いる場合、学習データに含まれない最新の法令への対応が困難である点や、ハルシネーションを生じうる点から、正確性に課題が残る。これに対し、外部知識を検索して統合する検索拡張生成 (Retrieval-Augmented Generation: RAG) は、生成時に根拠となる条文等を参照できるため、法令知識を要するタスクへの有効性が期待されている。

本研究の目的は、法令4択問題において、RAG の検索クエリ設計および検索戦略の違いが回答品質に与える影響を明らかにすることである。具体的には、デジタル庁が公開する法令4択問題データセット [1] を対象に、プロンプト最適化によるクエリ生成を適用した RAG の有効性を比較検証した。

本稿の構成は以下の通りである。第2節では、RAG の概要と課題および本稿で取り上げる検索クエリ設計の関連研究を整理する。第3節では、本研究で実装する推論方式とプロンプト設計手法を説明する。第4節では各手法の実験結果と考察を示し、第5節で結論を述べる。

2. 関連研究

2.1 RAG の基本枠組みと課題

Lewis らは、学習済みのモデルに外部知識を付与し、検索結果に基づいて回答を生成する RAG を提案した [2]。RAG は入力に基づき上位 k 文書を取得し、回答を生成する。外部知識を参照することで、LLM 内部の知識不足を

¹ 鹿児島大学
Korimoto, Kagoshima, Kagoshima 8900065, Japan
a) k6917121@kadai.jp
b) k7719654@kadai.jp
c) k1747415@kadai.jp
d) k0416336@kadai.jp
e) ono@ibe.kagoshima-u.ac.jp

補い、判断根拠の提示を行える点が特徴となる。一方で、検索失敗や根拠不足が最終回答の品質に直接的に波及する点、および検索クエリが固定的に与えられるため、文脈に応じた適応的な検索の実現が困難な点が課題として挙げられる。

2.2 GEPA によるプロンプト最適化

LLM の性能向上を目的として、強化学習や進化計算を用いてプロンプトを自動的に改善する手法が検討されている。Agrawal らが提案した GEPA (Genetic Pareto) は、実行ログに基づく自己反省を通じてプロンプトを更新、最適化する手法である [3]。GEPA は GRPO (Group Relative Policy Optimization) [4] 等の強化学習ベース手法と比較して数値報酬以外の観点も考慮しながら、探索を進められるため、比較的少ない試行回数で、改善が得られる可能性がある [3]。

本研究では、GEPA により最適化した検索用プロンプトを組み込む RAG と、クエリを固定設計した RAG を同一タスクで比較し、RAG 設計に GEPA を導入した場合の効果と課題を実証的に評価する。

2.3 LevelRAG による階層的アプローチ

検索品質向上のためのクエリ書き換えは、ベクトル検索のような意味類似度に基づく検索で有効である一方、単語レベルの検索を併用するハイブリッド検索では挙動が不安定になりうることが指摘されている。この課題に対して Zhang らは、特定の検索器に依存しない書き換えを可能にする階層的なアプローチとして LevelRAG を提案した [5]。ユーザ質問を分解し、小問ごとの検索結果の要約、統合を担う層と、検索器に応じたクエリ改良および検索実行を担う層の 2 階層に分けることで、検索器ごとの最適化と検索性能向上の両立を目指す。

本研究では、LevelRAG の着想に基づき、選択肢ごとに関連条文を取得して比較する階層的推論方式を実装し、他方式と正解率を比較する。

3. 研究方法

3.1 概要

本研究では、法令知識を要する 4 択問題を対象に、回答生成方法が異なる複数の推論方式を実装し、比較評価する。入力の問題文および選択肢集合 $L = \{a, b, c, d\}$ であり、出力は正しいと判断した選択肢ラベル $\hat{l} \in \{a, b, c, d\}$ である。比較対象は方式 1 から 5 の 5 方式であり、各方式は外部知識検索 (RAG) を用いるか、および RAG を用いる場合に検索クエリと検索戦略をどのように設計するかが異なる。方式 1 は RAG を用いないベースライン、方式 2 は正解文脈を付与した参照上限として位置づける。方式 3-5 は RAG を用いるが、検索クエリ生成と検索戦略が互

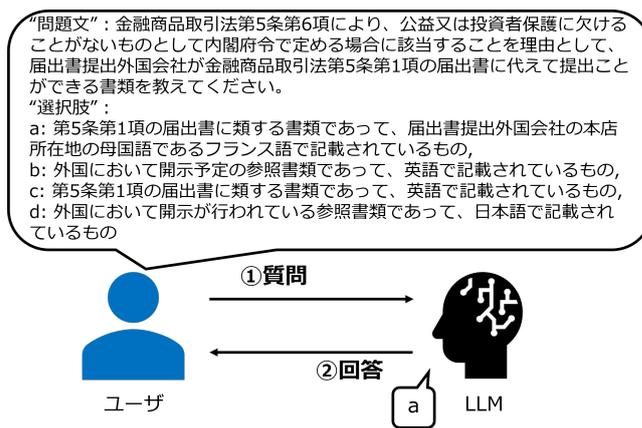


図 1 LLM 単体 (正解文脈なし) (方式 1) の全体像

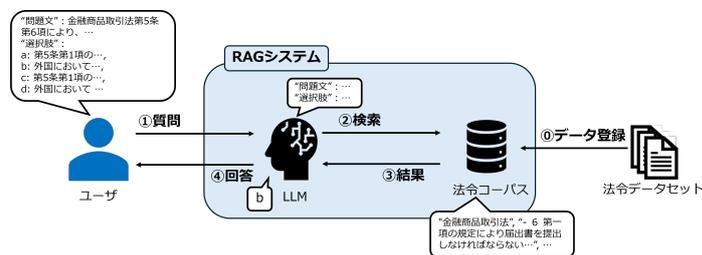


図 2 固定クエリ RAG (方式 3) の全体像

いに異なる比較対象である。

- (方式 1) LLM 単体 (正解文脈なし) : 検索は用いず、問題文と選択肢集合から LLM が直接回答を生成する。
- (方式 2) LLM 単体 (正解文脈付与) : 検索は用いず、問題文と選択肢集合に加えて、データセットに含まれる正解文脈を入力として付与し、LLM が回答を生成する。すなわち、RAG が理想的な検索を行った場合に相当する。
- (方式 3) 固定クエリ RAG : 固定の方針で検索クエリを構成し、法令コーパスから取得した条文を根拠として LLM が回答を生成する。
- (方式 4) GEPA によるクエリ生成プロンプト最適化 RAG : GEPA に基づきクエリ生成プロンプトを反復的に更新し、得られた最良プロンプトで検索を行う。取得条文を根拠として LLM が回答を生成する。
- (方式 5) 選択肢別評価 RAG : 選択肢ラベルごとに検索クエリを構成して条文を取得し、選択肢ラベルごとの評価結果を統合して最終回答を決定する。

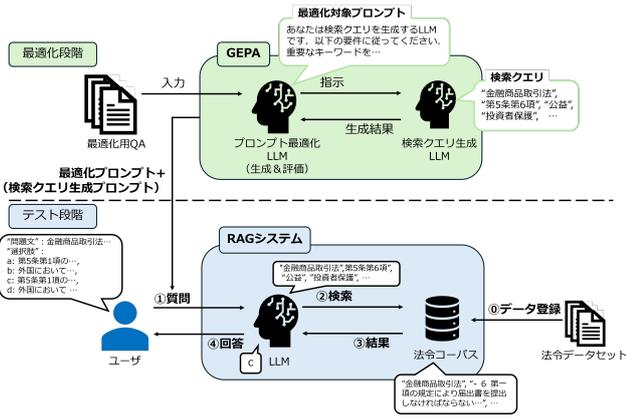


図 3 GEPA によるクエリ生成プロンプト最適化 RAG (方式 4) の全体像

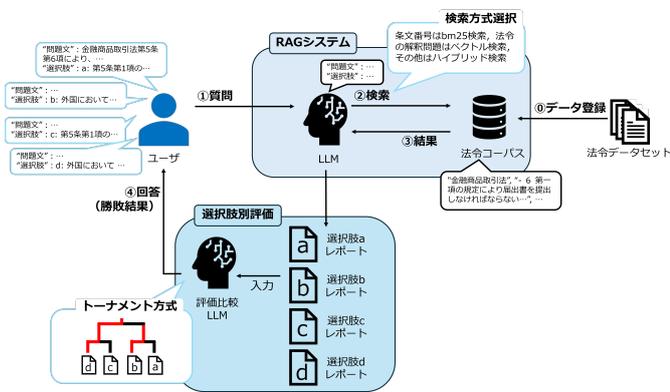


図 4 選択肢別評価 RAG (方式 5) の全体像

3.2 方式 1: LLM 単体 (正解文脈なし)

3.2.1 概要

方式 1 の全体像を図 1 に示す。方式 1 は、外部知識検索を用いず、問題文と選択肢集合のみを入力として LLM が直接回答を生成する方式である。本方式は、検索や改良手法を導入した方式 2-5 と比較するためのベースラインとして位置づける。方式 2 は方式 1 の入力にデータセットに含まれる正解文脈を加えた手法であるため、具体的な構成と処理手順についての記載は省略する。

3.2.2 構成と処理手順

データセットから取得した問題文と選択肢集合を LLM に与える。LLM は正しいと判断した選択肢ラベルを 1 つ出力する。検索処理や根拠文書の取得は行わない。

3.2.3 入力内容と出力形式

入力内容はデータセットから取得する。出力形式や推論時の設定 (例: temperature, Top- k) は他方式と可能な限り揃え、方式間の差が検索の有無や検索戦略の違いに起因するよう統制する。

3.3 方式 3: 固定クエリ RAG

3.3.1 概要

方式 3 の全体像を図 2 に示す。方式 3 は、RAG を導入

し、法令コーパスから取得した関連条文を根拠として LLM が最終回答を生成する方式である。方式 1 との比較により、RAG 導入の効果を評価可能にする。

3.3.2 構成と処理手順

方式 3 は、(1) 質問、(2) 検索、(3) 関連条文に基づく最終回答生成、の 3 段階で処理を行う。まず、問題文と選択肢集合から、あらかじめ定められた方針に従って検索クエリを構成し、法令コーパスから関連条文を検索する。次に、ベクトル検索により関連条文を上位 $k = 5$ 件取得する。最後に、取得した条文を問題文と選択肢集合とともに LLM に与え、正しいと判断した選択肢ラベルを出力させる。

3.3.3 法令コーパスと検索単位

本研究では、日本の現行法令テキストを網羅的に収録した e-Gov 法令検索の法令データをコーパスとして用いる [6]。同コーパスは法令体系や条文番号、本文などの構造情報を XML として保持しているため、条、項単位への分割や参照関係を意識した前処理を行いやすい。本研究では、法令データを条、項単位に分割して文書化し、埋め込みベクトルを付与してベクトルデータベースに格納する。検索では近傍検索により関連条文を取得する。

3.4 方式 4: GEPA によるクエリ生成プロンプト最適化 RAG

3.4.1 概要

方式 4 の全体像を図 3 に示す。方式 4 は、GEPA に基づく自動最適化により、RAG における検索クエリ生成プロンプトを更新し、得られた最良プロンプトを用いて検索および回答を行う方式である。訓練データに対する実行結果からフィードバックを生成し、プロンプトを反復的に改良する。なお、方式 4 における最適化の対象は検索クエリ生成に用いるプロンプトに限定する。方式 3 との比較により、検索クエリ改良の効果を評価可能にする。

3.4.2 構成と処理手順

方式 4 は、(i) 訓練用データに対する推論結果を収集、(ii) 誤答または参照法令の未取得例に対して失敗要因を診断するフィードバックを生成、(iii) フィードバックに基づきプロンプト候補を生成、(iv) 検証データで性能の高い候補を保持、というループを反復する。反復後に得られた最良のプロンプトを用いて、(1) 問題文、選択肢集合から検索クエリを生成し、(2) 法令コーパスを検索して関連条文を取得し、(3) 取得条文を根拠として LLM に最終回答の選択肢ラベルを生成させる。

3.4.3 探索設定

目的関数は、検索で得た取得文書本文と、データセットに含まれる正解文脈との一致スコアとする。一致スコアは、取得文書ごとに正解文脈と取得文書本文の類似度 s を計算し、その最大値を当該問題の best_score とした。類似度 s は次式で算出する。

$$s = \min\left(1.0, \sum_{i \in \{legal, keyword, text\}} w^{(i)} s^{(i)}\right).$$

ここで、 $s^{(legal)}$ は正解文脈と取得文書本文の間で、法令名および「第〇条」、「第〇項」、「第〇号」といった条文番号表現の一致度を示す。 $s^{(keyword)}$ は、正解文脈および取得文書から抽出した2文字以上の日本語文字列をキーワード集合とみなし、その共通部分の割合として算出する。 $s^{(text)}$ は、正解文脈を一定長の断片に分割し、各断片が取得文書に含まれるかを調べ、含まれる断片の割合として算出する。これらに重み(0, 4, 0.4, 0.2)を乗じ、1を上限としてクリップすることで最終的な類似度を得る。

3.5 選択肢別評価 RAG (方式5)

3.5.1 概要

方式5の全体像を図4に示す。方式5は、選択肢ラベルごとに検索を行って法令コーパスから条文を取得し、選択肢ラベル別の評価結果を統合して最終回答を決定する方式である。LevelRAGの階層的設計の着想に基づき[5]、選択肢ラベル別に複数の情報を収集し、比較することで、検索品質の向上を期待する。方式3および4との比較により、検索戦略の改良効果を評価可能にする。

3.5.2 構成と処理手順

本方式では、問題文と選択肢集合から検索クエリを生成し、問題の種類およびクエリの性質に応じて3種類の検索方法を適用して法令コーパスから関連条文を取得する。具体的には、法令の解釈を問う問題にはベクトル検索を用い、クエリに条文番号が含まれる場合にはbm25検索を用いる。一方、クエリに条文番号が含まれず、条文の内容や適用事例を問う問題にはハイブリッド検索を用いる。本方式のハイブリッド検索では、同一クエリに対してbm25検索とベクトル検索を並列に実行し、得られたランキングをReciprocal Rank Fusion (RRF)により統合して上位文書を取得する[7-9]。次に、取得文書を根拠として、当該選択肢ラベルが妥当かどうかの判断と理由をレポートとして出力する。1問当たり計4件の評価結果を入力として、最終的に採用する選択肢ラベルを決定する。最終回答の決定にはトーナメント方式を用いる。具体的には、(a,b)と(c,d)をそれぞれ比較して勝者を選出し、最後に勝者同士を比較して最終回答を生成する。

3.5.3 情報収集と検索結果の評価

各選択肢ラベルに対して、初回検索で取得した上位 $k=5$ 件のうち、検索クエリに含まれる施行令/規則、法令名で取りこぼしがあれば追加で検索し、初回検索結果と情報を統合する。次に、重複文書を削除し、検索結果を再度ランク付けした上で、上位 $k=10$ 件を取得する。各選択肢ラベルについて、検索結果に基づく判定を「正しい」「正解候補」「正解ではない」「誤り」の4段階で出力させ、トーナ

ファイル名

金商法 第2章 選択式 関連法令 問題番号 57

問題文

金融商品取引法第5条第6項により、公益又は投資者保護に欠けることがないものとして内閣府令で定める場合に該当することを理由として、届出書提出外国会社が金融商品取引法第5条第1項の届出書に代えて提出することができる書類を教えてください。

選択肢集合

- 第5条第1項の届出書に類する書類であって、届出書提出外国会社の本店所在地の母国語であるフランス語で記載されているもの
- 外国において開示予定の参照書類であって、英語で記載されているもの
- 第5条第1項の届出書に類する書類であって、英語で記載されているもの
- 外国において開示が行われている参照書類であって、日本語で記載されているもの

正答ラベル

c)

正解文脈

金融商品取引法
第5条
第6項

第一項の規定により届出書を提出しなければならない外国会社(以下「届出書提出外国会社」という。)は、公益又は投資者保護に欠けることがないものとして内閣府令で定める場合には、同項の届出書に代えて、内閣府令で定めるところにより、次に掲げる書類を提出することができる。第2号外国において開示が行われている参照書類又は第一項の届出書に類する書類であつて英語で記載されているもの

図5 法令4択問題データセットに含まれる問題の例

メント方式で最終回答を決定する。

4. 実験

4.1 問題データセット

実験には、デジタル庁公開の法令4択問題データセット[1]を用いた。データセットには問題文、選択肢集合、正答ラベル、正解文脈が含まれる。データは学習用に100問と評価用に40問に分割し、実験を行った。問題例を図5に示す。

4.2 実験設定

表1に、比較対象である5方式、すなわち、LLM単体(正解文脈なし)、LLM単体(正解文脈付与)、固定クエリRAG、GEPAによるクエリ生成プロンプト最適化RAG、

表 1 実験設定

手法	RAG	GEPA	選択肢別評価
方式 1: LLM 単体	なし	なし	なし
方式 2: LLM+正解文脈	なし	なし	なし
方式 3: LLM+RAG	あり	なし	なし
方式 4: LLM+RAG+GEPA	あり	あり	なし
方式 5: LLM+RAG+選択肢別評価	あり	なし	あり

表 2 正解率

手法	Accuracy	(正解数)
方式 1	52.5%	(21/40)
方式 2	77.5%	(31/40)
方式 3	60.0%	(24/40)
方式 4	67.5%	(27/40)
方式 5	60.0%	(24/40)

選択肢別評価 RAG の概要を示す。対象タスク、データ分割、推論設定は全方式で共通とし、検索の有無、および GEPA による最適化の有無を主な相違点とする。LLM は Qwen3:8b を用い、推論はローカル環境で実行した。法令コーパスには、e-Gov 法令検索の法令データを用い、条、項、号ごとに分割して構造化した上で登録した [6]。RAG の検索アルゴリズムには選択肢別評価以外はベクトル検索を採用し、上位 $k = 5$ 件を取得した。選択肢別評価では、条文番号は bm25 検索、法令の解釈問題はベクトル検索、その他はハイブリッド検索を採用し、上位 $k = 10$ 件を取得した。全ての方式において、qwen3:8b を LLM として用いた。

評価指標は 4 択の正解率 (Accuracy) とした。すなわち、LLM が最終回答として出力する選択肢ラベルと、データセットに含まれる正答ラベルの一致数を 40 で除して算出した。

4.3 実験結果

表 2 に実験結果を示す。方式 1 と方式 3 の結果に着目すると、RAG の導入により、LLM 単体 (52.5%) と比べて正解率が 7.5 ポイント向上したことがわかる。

続いて、方式 3 と 4 を比較すると、GEPA を追加することで正解率は 67.5% となり、方式 3 の LLM+RAG から 7.5 ポイント改善したことがわかる。

一方、選択肢別評価 RAG は 60.0% に留まり、固定クエリ RAG と同等の正解率であった。

4.4 考察

誤答を以下の要因に分類し、各方式がどの要因に影響したかを考察する。

- (1) 検索失敗：関連条文が十分に取得できなかったケース。
- (2) 根拠不一致：取得した上位条文が問題に無関係であったケース。

ファイル名

金商法 第 2 章 選択式 関連法令 問題番号 57

問題文

金融商品取引法第 5 条第 6 項により、公益又は投資者保護に欠けることがないものとして内閣府令で定める場合に該当することを理由として、届出書提出外国会社が金融商品取引法第 5 条第 1 項の届出書に代えて提出することができる書類を教えてください。

取得条文 (抜粋)

金融商品取引法 昭和二十三年法律第二十五号
第二章 企業内容等の開示

図 6 固定クエリ RAG で改善に成功した問題例

ファイル名

金商法 第 1 章 選択式 関連法令 問題番号 44

最適化前クエリ

金融商品取引法第 2 条第 1 項 有価証券の定義
例外

最適化後クエリ

金融商品取引法 第 2 条第 1 項 有価証券 定義
異なるもの 相互会社社債券 株券 外国の者の
証券 新株予約権証券 信託受益権

図 7 GEPA で改善した検索クエリ例

- (3) 回答生成失敗：最終回答として選択肢ラベルを所定形式で出力できなかったケース。

LLM 単体から固定クエリ RAG により結果が改善した 1 問を図 6 に提示する。図 6 のファイル名はテストデータの識別子である法令名、章、問題番号を示す。図 6 の問題文は本実験における入力にあたり、取得条文は RAG を通して法令コーパスから取得した関連条文トップ 1 を抜粋して記載した。図 6 のファイル名の法令名と取得条文の法令名が一致していることから、関連条文を参照できていることが示唆される。固定クエリ RAG の導入により、根拠に基づいた回答を生成しやすくなったと考えられる。

固定クエリから GEPA に変更したことで、結果が改善したケースのプロンプト最適化前後の検索クエリを図 7 に示す。図 7 において、最適化後クエリの太字は、最適化によって新たに生成されたクエリである。最適化前後のクエリを比較すると、「金融商品取引法第 2 条第 1 項」が「金融商品取引法」と「第 2 条第 1 項」に、「有価証券の定義」が「有価証券」と「定義」に分割されていることが確認できる。検索クエリ生成プロンプトの最適化により、クエリの種類が増加し、単語の粒度が細くなることで、関連条文の網羅性が向上した可能性がある。

ファイル名 借地借家法 第 3 章 選択式 根拠条文 問題番号 16
最適化前クエリ 借地借家法 第 34 条 第 1 項 建物賃貸借契約 終了後の転借人に対する通知義務と過量の対 抗権
最適化後クエリ 借地借家法 第 34 条 第 1 項 通知 過量 借地権 存続期間 借地借家法第 34 条第 1 項 借地借 家法第 35 条第 1 項 借地借家法第 34 条第 2 項 借地借家法第 35 条第 2 項

図 8 GEPA で不正解だった問題の検索クエリ例

一方で、LLM 単体（正解文脈なし）と固定クエリ RAG で正解したにもかかわらず、GEPA では誤答したケースも確認された。GEPA で誤答したケースのファイル名と最適化前後の検索クエリを図 8 に示す。図 8 から、最適化後に「〇〇法第〇条第〇項」というクエリが複数存在していることが確認できる。これらの結果から、クエリの細分化は (1) 検索失敗と (2) 根拠不一致の抑制に寄与したと考えられる。

方式 5 の選択肢別評価 RAG では、bm25 検索、ベクトル検索、ハイブリッド検索 (bm25+ベクトル) のうち、bm25 検索が頻繁に適用されていた。これは、問題文あるいは選択肢に条文番号が含まれるケースが多かったため、単語検索にあたる bm25 検索 [7] が適用され、(2) 根拠不一致の抑制につながったと考えられる。一方、選択肢別評価 RAG は検索、評価を多段に行うため、途中段階の誤りである (1) 検索失敗および (2) 根拠不一致の影響が累積し、最終判断に波及しやすい可能性がある。

正解率の上限として位置付けた方式 3 の LLM 単体（正解文脈付与）は、比較手法の中で最良 (77.5%) であった一方で、100%には至らなかった。その一因として、(3) の回答生成失敗が確認された。本実験では最終回答として出力冒頭に選択肢ラベルのいずれか 1 文字を出力するよう指示を行っていたが、この形式に従わない出力が一部で発生した。

今後は、正規表現によるラベル抽出と再試行により出力形式を厳格化するとともに、検索クエリの細分化と追加生成、およびクエリに含まれる条文番号に基づく語彙ベース検索の強制を組み合わせることで、検索失敗および根拠不一致の低減を図る。

5. 結論

本研究では、デジタル庁公開の法令 4 択問題データセットを対象に、軽量 LLM (Qwen3:8b) を用いたローカル実

行可能な検索拡張生成 (RAG) システムを構築し、検索クエリ生成およびプロンプト設計の工夫が回答に与える影響を検証した。具体的には、GEPA による検索クエリ生成プロンプトの自動最適化、ならびに選択肢ラベルを個別評価し、得られた根拠情報から回答を決定する選択肢別評価 RAG を実装し、比較評価を行った。

今後の課題として、GEPA と選択肢別評価 RAG の組み合わせが未検証な点が挙げられる。複数の RAG 検索アルゴリズムを用いる選択肢別評価 RAG における検索クエリ生成プロンプトを GEPA により最適化することで、取得条文の適合度を高め、最終的な正解率を向上させることができると考える。

参考文献

- [1] Government of Japan Digital Agency. 日本の法令に関する多肢選択式 qa データセット (lawqa-jp). GitHub repository. accessed: 2025-12-26.
- [2] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 33, pp. 9459–9474, 2020.
- [3] Lakshya A. Agrawal, Shangyin Tan, Dilara Soylu, Noah Ziems, Rishi Khare, Krista Opsahl-Ong, Arnav Singhvi, Herumb Shandilya, Michael J. Ryan, Meng Jiang, Christopher Potts, Koushik Sen, Alexandros G. Dimakis, Ion Stoica, Dan Klein, Matei Zaharia, and Omar Khat-tab. GEPA: Reflective prompt evolution can outperform reinforcement learning. *arXiv preprint arXiv:2507.19457*, 2025.
- [4] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models, 2024.
- [5] Zhuocheng Zhang, Yang Feng, and Min Zhang. Levelrag: Enhancing retrieval-augmented generation with multi-hop logic planning over rewriting augmented searchers, 2025.
- [6] デジタル庁. e-gov 法令検索: Xml 一括ダウンロード. Web ページ. 参照日: 2026-01-21.
- [7] Stephen Robertson and Hugo Zaragoza. The probabilistic relevance framework: Bm25 and beyond. *Found. Trends Inf. Retr.*, Vol. 3, No. 4, p. 333–389, April 2009.
- [8] Vladimir Karpukhin, Barlas Öğuz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. Dense passage retrieval for open-domain question answering, 2020.
- [9] Gordon V. Cormack, Charles L A Clarke, and Stefan Büttcher. Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, p. 758–759, New York, NY, USA, 2009. Association for Computing Machinery.