

## LLM付加価値基盤を支える技術的取り組み

ChatGPTの登場以降、大規模言語モデル(LLM)が世界的に注目されており、ビジネスシーンにおける活用も急速に進んでいます。NTTドコモにおいても、LLMを安全かつ便利に活用するための「LLM付加価値基盤」を開発し、ドコモグループ内へ提供しています。本稿ではLLM付加価値基盤について概観するとともに、LLM付加価値基盤の性能改善や利用状況分析に向けた取り組みとして、RAG (Retrieval-Augmented Generation) の精度向上、UI (User Interface) /UX (User Experience) の改善、ダッシュボードの構築について解説します。

### はじめに

生成AI(人工知能)が世界的に注目されています。特に、大規模言語モデル(LLM: Large Language Model)は、その性能の高さから、多くの企業で急速に導入が進んでいます。NTTドコモでも、業務効率化や新規事業創出を目的に「LLM付加価値基盤」を開発し、ドコモグループ内に向けて提供中です<sup>(1), (2)</sup>。

企業内でLLMを有効に活用するには、LLMの単なる提供だけでなく、アプリケーションやツールの開発、使いやすいUI (User Interface) の設計・提供、社内ユースケースの発掘と水平展開など、基盤そのものや基盤周辺機能、基盤運用プロセスの継続的な改善が肝要です。

NTTドコモでは、ビジネスニーズが高いRAG (Retrieval-Augmented Generation) に、検索精度向上と社内運用効率化の両側面から取り組んでいます。RAGは、LLMによるテキスト生成を情報検索システムの活用で検索・回答精度を高める技術であり、コンタクトセンタのオペレーション高度化やバックオフィスの業務効率化への貢献が大きく期待されています。LLM付加価値基盤では、社内ドキュメントの検索に対応したRAGを「オンライン文書検索」としてユーザへ提供しており、正確かつ有用な情報検索を可能としています。

また、LLM付加価値基盤の利便性向上につながる取り組みとして、独自UIの開発があります。LLM付加価値基盤の利用画面を、ユースケースに合わせた使い勝手の良いデザインとすることで、ユーザの定着や業務プロセスのさらなる効率化をねらっています。

さらに、ノウハウやナレッジの蓄積、LLM活用方法の新規創出、組織ごとの利用状況把握、ならびにそれらの社内展開によるLLM付加価値基盤の活用拡大をめざし、LLM付加価値基盤のユーザ利用ログ収集パイプライン\*<sup>1</sup>および可視化システム(ダッシュボード\*<sup>2</sup>)を構築して社内ユーザ向けに公開しています。

本稿では、LLM付加価値基盤を支える技術的取り組みを、RAG、UI/UX (User Experience)、ダッシュボードの観点から解説します。

### RAG：検索精度向上に向けての取り組み

#### ■ RAGアーキテクチャ

LLM付加価値基盤のRAGアーキテクチャは、ドキュメントを効率的に検索できるデータストアであるIndexの構築のためのアーキテクチャと、質問応答のためのアーキテクチャの2つに大別できます(図1)。

Index構築のためのアーキテクチャは、取り込みを行うドキュ

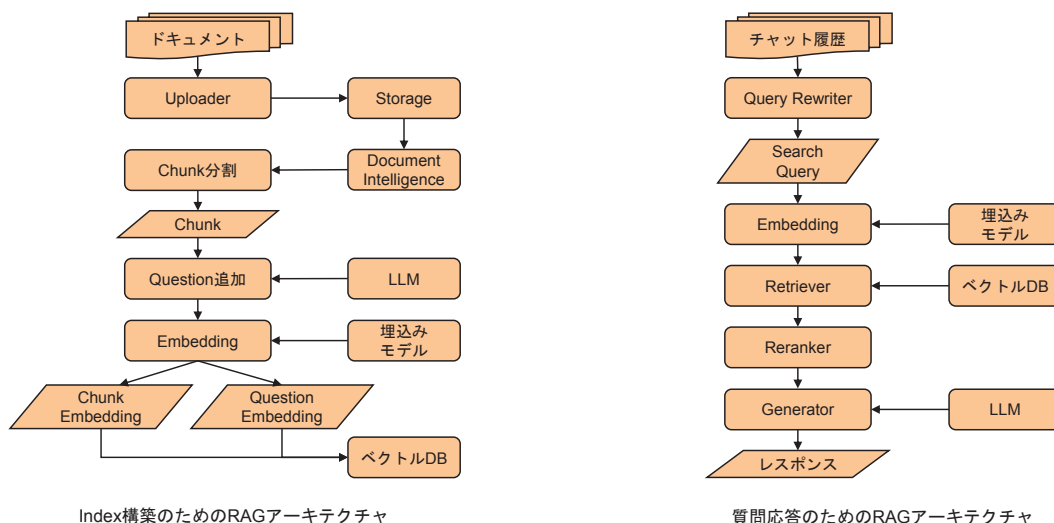


図1 RAGアーキテクチャ

表1 RetrieverとGeneratorの精度評価結果

			Generator		合計	全体「35」に占める割合
			正解	不正解		
評価セット1	Retriever	正解	24	1	25	71
		不正解	2	8	10	29
		合計	26	9	35	100
		全体「35」に占める割合	74	26	100	—
			Generator		合計	全体「16」に占める割合
			正解	不正解		
評価セット2	Retriever	正解	11	0	11	69
		不正解	0	5	5	31
		合計	11	5	16	100
		全体「16」に占める割合	69	31	100	—
			Generator		合計	全体「25」に占める割合
			正解	不正解		
評価セット3	Retriever	正解	25	0	25	100
		不正解	0	0	0	0
		合計	25	0	25	100
		全体「25」に占める割合	100	0	100	—

メントをアップロードするUploader機能、ドキュメントから想定され得る質問を生成し、ドキュメントと紐付けてIndexに格納しておくことでRetriever（後述）の検索精度を向上させるQuestion追加機能、そして、ベクトル検索のために、チャンク<sup>\*3</sup>や追加した質問をベクトル化してベクトルDB<sup>\*4</sup>に保存するChunk/Question Embedding機能の3つで構成されています。

他方、質問応答のためのアーキテクチャは、検索精度を上げるためクエリを拡張するQuery Rewriter機能、回答生成時に参照するドキュメントを検索するRetriever機能、Retrieverの出力結果を調整するReranker機能、そして、質問者に提示する回答を生成するGenerator機能の4つで構成されています。

#### ■精度評価

##### (1) 概要

LLM付加価値基盤においてRAGは、社内ドキュメントの検索やユーザ質問への回答といったビジネスユースを想定しているため、精度の高さが重要です。そこで、RAGの精度向上の取り組み

- \*1 パイプライン：一連の手順に従ってデータを順次処理する仕組みのこと。
- \*2 ダッシュボード：グラフや統計データなどを使って、特定の指標やデータを一目で把握できるように視覚的に表示するインターフェースのこと。
- \*3 チャンク：名詞句や動詞句などの文節のように、テキストを意味のある小さな単位やかたまりに分割したもの。
- \*4 ベクトルDB：テキスト、画像、音声などのデータをベクトル化して扱う際に利用されるデータベースのこと。類似性検索や異常検知などのタスクにおいて、大量のデータに対しても高速かつ安定した動作を行うことが特長です。

みとして、質問応答にかかわる要素技術の精度評価を行いました。

RAGは大きく分けて、関連ドキュメントを検索するRetrieverと、Retrieverの結果を受けて回答を生成するGeneratorの2つの要素技術で構成されています。これらの技術は直列処理になっており、先行するRetrieverの結果が不正解の場合、論理的には、後続するGeneratorの結果も不正解になります。そのため、最終出力であるGeneratorの結果を評価し、加えて精度低下の原因となり得る中間処理のRetrieverの結果も評価しました。評価については、「おおむね正解」か「明らかに不正解」かの2段階で行い、マトリクスにまとめました。

##### (2) 評価データセット

性能評価に向けて、社内業務マニュアルを基に3種類の評価データセットを作成しました。評価データセット1および2は、複数のPDFドキュメントで、評価データセット3は、Microsoft Word、Microsoft PowerPoint、Microsoft Excel形式のドキュメントで構成されています。いずれの評価データセットも、参照ドキュメントと評価用Q&Aから構成されており、RAGの精度評価作業に際しては参照ドキュメントをRAGのIndexに登録することで評価用Q&Aとして評価を行いました。

##### (3) 評価結果

###### ① RetrieverとGeneratorの精度評価結果

RetrieverとGeneratorの精度評価結果を表1に示します。

評価データセットすべてを通して、Retrieverが正解していて、

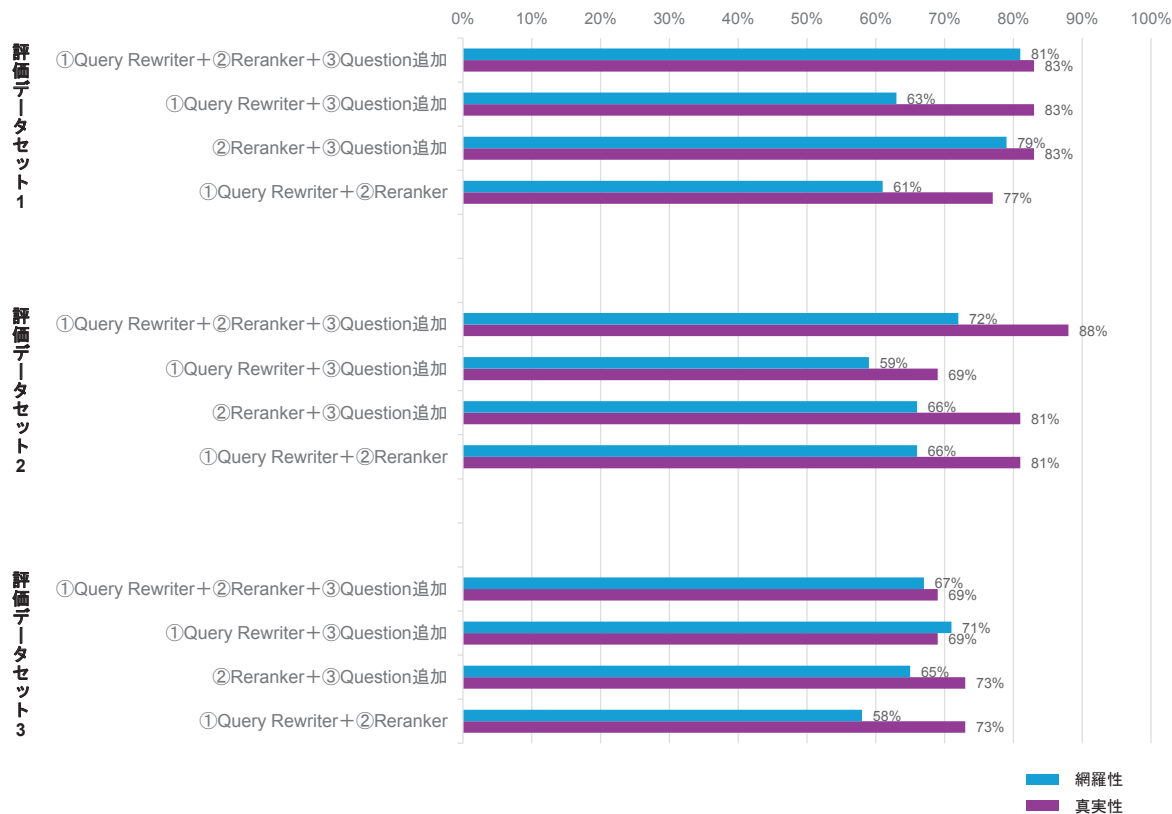


図2 RAGの要素技術別評価

Generatorが不正解になっている事例は1件だけでした。この結果から、Retrieverが正解していれば、Generatorはおおむね正しい回答を生成でき、RAGのボトルネックはRetrieverにあることが分かりました。これらの結果を踏まえ、NTTドコモではRetrieverの精度向上に注力しています。

② Query Rewriter, Reranker, Question追加の要素技術別評価

次に、Retrieverの精度向上のために実装した、Query Rewriter機能、Reranker機能、Question追加機能がどの程度精度向上に寄与しているのかを評価しました。具体的には、3つすべての機能があるケース、いずれかが欠落しているケースの合計4ケースの精度を網羅性（必要な情報が網羅されている回答の割合）と真実性（ハルシネーション\*5がない回答の割合）の2つの評価を行い、結果を比較しました（図2）。

評価データセットによって、各要素技術の効果に差がみられました。このことから、ドキュメント形式ごとに適性の高い要素技術と低い要素技術があると考えられます。今後は、各要素技術が得意とするドキュメント形式の傾向を分析していくとともに、ド

\*5 ハルシネーション：AIが、事実と異なる情報、あるいは実際には存在しない情報を生成する現象のこと。

キュメント形式の種類に応じて、要素技術を柔軟に組み合わせて利用できるようにアーキテクチャの設計を検討していきます。

## RAG：社内運用の利便性向上

### ■適性度チェックシートの展開

RAGの導入を検討している組織が自律的にRAGの導入是非の判断や導入対象業務の選定を行えるよう、「適性度チェックシート」を独自に作成してNTTドコモ社内で配布しています。

適性度チェックシートは表2に示すように、6つのチェック項目と易・中・難の3段階の導入難易度からなる適性度のマトリクスで構成されており、利用者は、自身の想定するユースケースをマトリクスの該当セルに当てはめていくことで、当該業務に対するRAGの適性度を確認できます。

### ■Q&A検索エンジンとの連携

十分な量のQ&Aデータを準備できるのであれば、基本的な情報や問合せ回数の多い質問に対しては、RAGを活用するよりもQ&Aデータを直接参照して回答するほうが正確かつ低コストな場合が多いです。そこで、RAGアーキテクチャとQ&A検索エンジンとの連携を進めています。

Q&A検索エンジンとの連携の一環として、RAGの検索結果と

表2 適正度チェックシート

	易	中	難	理由
1. 利用者の有する知識	専門家	ある程度知識のある社員	知識のない社員、個人ユーザなど	利用者がRAGの出力内容の正誤を判定可能
2. 問合せ内容	Q&Aやマニュアルにある問合せ	過去履歴にある問合せ	個別ケース、属人的な問合せ	正確な回答には体系的にまとめられた参照データが必要
3. 出力内容に求められる精度	たたき台レベルでよい	おおむね近い回答、または参考ドキュメントをセットで提示したい	誤答不可	RAGはハルシネーションや網羅的な回答生成が苦手
4. リアルタイム性	夜間バッチ処理でも可	1分程度のレスポンスタイムは許容できる	数秒単位での即時性が求められる	RAGはレスポンスに時間がかかる
5. 検索対象となる社内ドキュメントの形式	テキスト中心のWord, pdfなど	テキストと図表が混在するWord, pdfなど	図表中心のPowerPoint, Excelなど	RAGは図表の読み取りが苦手
6. 検索対象となる社内ドキュメントの量	少	中	多	ドキュメントの量が増加するにつれて、精度が低下する傾向あり

同時にQ&A検索エンジンの結果を提示するGUI (Graphical User Interface) を作成しました。Web検索のように、Q&A検索エンジンでヒットした類似質問とそれに紐づく回答を、関連する項目として提示します。

## UI/UX

### ■カスタマージャーニー分析

NTTドコモでは、LLM付加価値基盤のユーザ定着や業務効率化、利便性向上を実現するためのUI/UXの改善の一環として、社内の多様な利用シーンを想定したGUIデザインを独自に作成し、提供中です。

今回、これらの汎用的なGUIの提供に加え、特定のユースケースにマッチした専用GUIの新規作成と作成に向けたカスタマージャーニー分析<sup>\*6</sup>を実施しました。専用GUIを作成するユースケースとして、LLMによる大きな業務効率化が期待される「オンライン文書検索を用いた社内の事務処理に関する問合せ回答」を選定し、ユーザ（LLMを活用して、問合せに対する回答を検索・作成する作業）の協力を得ながら次のような調査を行いました。

まず、ユーザに対してアンケートを実施し、業務内容や作業頻度、LLMには任せられない工程などを確認しました。次に、ユーザが実際に行っている業務の観察を行い、どのような流れでどのような動きをしているか、どのような工程に時間がかかっているかを確認しました。最後に、ユーザインタビューを行い、ユーザが不満を感じて効率化したいと感じるポイントなどを確認しました。

\*6 カスタマージャーニー分析：より良いユーザ体験の提供やユーザ満足度の向上を目的とし、ユーザが商品やサービスを知ってから使用するまでの一連のプロセスを詳細に解析する手法のこと。

これらの調査を通じて、「問合せ回答業務の習熟度により作業プロセスが異なっており、特に習熟度の低いユーザにとってはLLMをうまく活用することで大きな業務効率化が見込めること」が明らかになりました。また、ハルシネーションはLLMをビジネスシーンで利用するうえでの課題の1つとしてよく知られていますが、問合せに回答する業務での活用においては、ハルシネーションに起因する誤案内の防止が特に重要であるという示唆が得られました。

これらの結果から、ユーザのニーズにこたえ、業務効率化の実現やユーザの定着を図るためには、LLMを利用して回答するだけのシンプルな設計ではなく高度なUI設計が必要であることが分かりました。

### ■分析に基づくデザイン作成

カスタマージャーニー分析から得られた知見を基に独自に開発したものがオンライン文書検索ユースケース用GUIです。ユーザは、本専用GUIを通じて検索を実施し、マニュアルなどの文書を基にしたLLMの回答およびヒットした文書自体を確認できます。

ユーザがLLMの回答の正しさを確認できるように、オンライン文書検索ユースケース用GUIでは、LLMの生成した回答に加え、オンライン文書検索でヒットした文書が上位3件まで表示される、文書のダウンロードができる、ブックマークや検索履歴機能により別の検索条件の結果との比較を実施しやすくするなどの特長を備えています。

このように、特定のユースケースを想定した専用GUIを、通常のLLM利用の汎用GUIとは別に提供することで、ユーザのさらなる定着や業務プロセスの効率化をめざします。

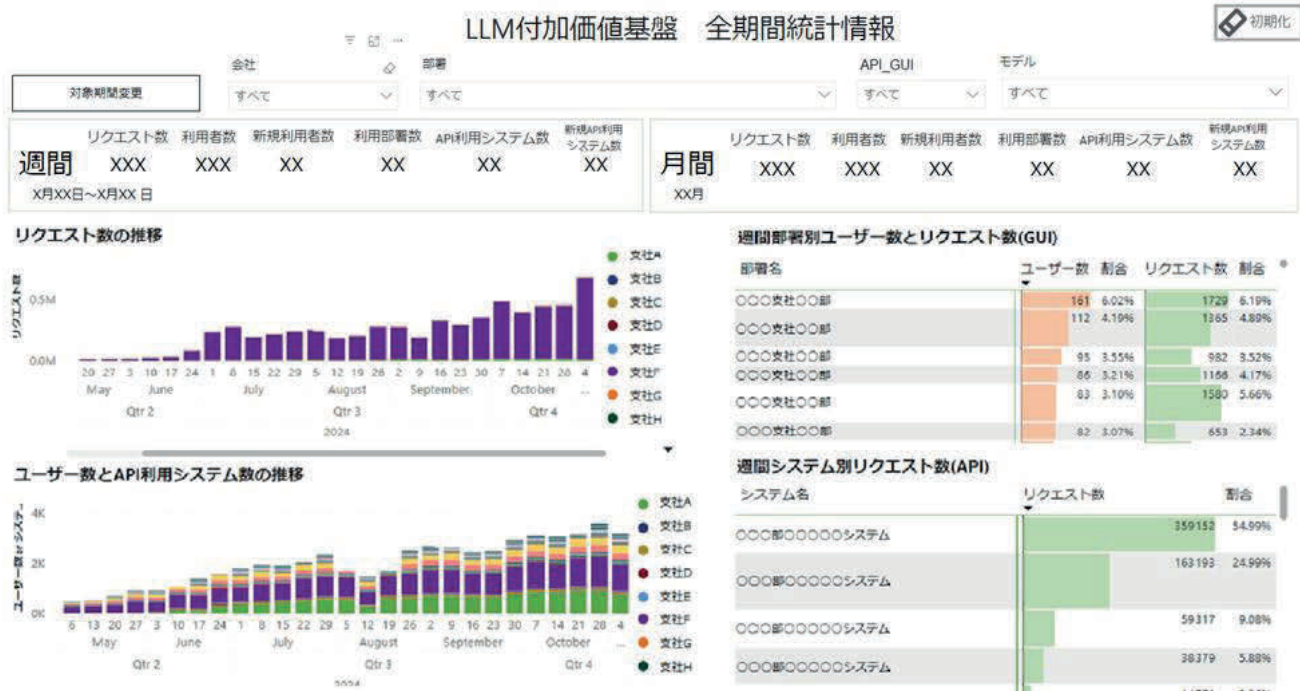


図3 全期間統計情報ページ

## ダッシュボード

### ■構築の背景

LLM付加価値基盤の利用状況をモニタリングすることは、基盤開発者によるサービスの継続的な改善や入力内容の監査、LLM活用推進担当者による施策の効果測定、LLMを活用した有用なユースケースのユーザ向け展開を行ううえで重要です。LLM付加価値基盤のリリース当初は、定期的に手動で組織ごとのユーザ数やリクエスト数などの統計情報やLLMの主な活用事例のレポートを作成していましたが、利用量の増加に伴いログが膨大になり、手作業での分析が非効率となったため、利用状況を自動的に可視化するダッシュボードを構築し、LLM付加価値基盤のユーザ向けに公開しています。

### ■可視化の概要

#### (1) ダッシュボードの構成

本稿執筆時点でのLLM付加価値基盤ダッシュボードは、「全期間統計情報」「週間利用傾向ランキング」「週間活用事例一覧」「週間ユーザプロンプト詳細」の4つのページから構成されています。なお、ダッシュボードはDatabricks<sup>(3)</sup>とMicrosoft社が提供するPower BI<sup>(4)</sup>を利用して構築しています。

#### (2) 全期間統計情報ページ

全期間統計情報ページは、利用会社部署ごとのユーザ数・リクエスト数などの統計情報をまとめたページです(図3)。本ページを活用することで、LLM活用推進施策の実施や新機能のリリー

スなどの効果測定、経営層などのステークホルダへの利用状況の迅速な報告などができます。

各指標の統計値からは、直近の1週間および1カ月間の利用状況を確認できます(図3上)。また、リクエスト数ないしユーザ数の推移グラフから、過去から現在までの状況を確認、また期間ごとに比較することなどができます(図3左下)。部署別・システム別の統計情報からは、部署やシステムごとの直近の1週間のユーザ数・リクエスト数をランキング形式で確認できます(図3右下)。

#### (3) 週間利用傾向ランキングページ

週間利用傾向ランキングページは、ユーザがLLM付加価値基盤をどのようなタスクや業務領域で活用しているかをまとめたページです(図4)。本ページを活用することで、部署ごともたはLLMのモデルごとに、活用されている主なタスクや業務領域を大局的に把握できます。

グラフからは、ユーザの活用タスクと活用業務領域をランキング形式で把握できます(図4左)。データは、各ユーザのプロンプトをLLMで分類した結果に基づいています。LLMに与えるプロンプトは、活用タスクであれば「コーディング」「コンテンツ生成」、業務領域であれば「コールセンター」「コンサルティング」といったように、それぞれあらかじめ定義されたカテゴリへ分類されるよう設計しています。

ワードクラウド<sup>\*7</sup>とキーワードランキングからは、ユーザの入力プロンプトから抽出した頻出キーワードを定性的・定量的に把

## LLM付加価値基盤 週間利用傾向ランキング

※本ページはプロンプト全件から10000件ランダム抽出して分析した結果を表示。

### 活用タスクランキング

タスク	リクエスト数	割合
コーディング	2877	29.49%
質問応答	1532	15.70%
コンテンツ生成	1095	11.22%
データ分析	823	8.43%
情報検索	662	6.78%
文章校正	588	6.03%
テキスト分類	550	5.64%
文章要約	459	4.70%
翻訳	440	4.51%
ソリューション提案	326	3.34%
意思決定支援	264	2.71%
数値計算	118	1.21%

### 活用業務領域ランキング

業務領域	リクエスト数	割合
エンジニアリング	2717	28.05%
データ分析	1471	15.19%
コールセンター	1355	13.99%
コンサルティング	1187	12.25%
その他	1015	10.48%
企画提案	347	3.58%
マーケティング	322	3.32%
システム開発	307	3.17%
営業	256	2.64%
総務	233	2.41%
人事	183	1.89%
プロジェクトマネジメント	154	1.59%
UI/UX設計	72	0.74%

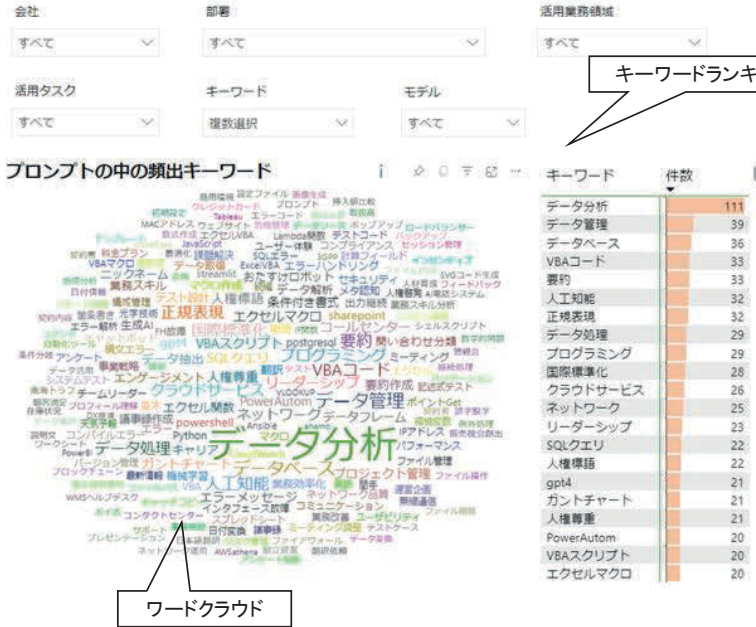


図4 週間利用傾向ランキングページ

握できます (図4右)。

#### (4) 週間活用事例一覧ページ

週間活用事例一覧ページは、LLMの主な活用事例やリクエスト数をユーザごとに可視化したページです。本ページを通じて、ユーザがほかの社員のLLMの活用状況を把握することで、自身の業務へのさらなる活用を推進することができます。

週間活用事例一覧ページは、週間利用傾向ランキングページと同様にLLMを活用し、ユーザ全体の過去のプロンプトから主なLLM活用事例を抽出・要約した結果を表示しています。例えば、「タイムスケジュールを作成する」「プログラムを修正する」「問合せへの回答文を作成する」などが実際に要約された結果です。

#### (5) 週間ユーザプロンプト詳細ページ

週間ユーザプロンプト詳細ページは、ユーザのプロンプト内容を開覧できるページです。本ページは、LLM付加価値基盤やダッシュボードの開発者だけに限定して公開しています。

本ページを活用することで、事業推進に資する有用な活用事例をプロンプト集として蓄積、またプロンプトに重大な秘匿情報や入力が禁止されている情報が含まれていないかを監査することができます。

## おわりに

本稿では、LLM付加価値基盤を支える技術的取り組みとして、RAGの検索精度向上と運用効率化の取り組み、UI/UXの実装・提供、ダッシュボード構築について解説しました。企業内でLLMを活用するためには、先進的技術の調査や実装のみならず、安全性や利便性など多岐に渡る観点からの検証と改善が重要です。

今後は、RAGの精度改善と運用効率化、UI/UXの機能改善と専用UIの作成・提供、高度な分析を可能とするダッシュボード機能の追加をめざします。

### ■参考文献

- 1) [https://www.docomo.ne.jp/binary/pdf/info/news\\_release/topics\\_230821\\_00.pdf](https://www.docomo.ne.jp/binary/pdf/info/news_release/topics_230821_00.pdf)
- 2) [https://www.docomo.ne.jp/corporate/technology/rd/technical\\_journal/bn/vol32\\_1/001.html](https://www.docomo.ne.jp/corporate/technology/rd/technical_journal/bn/vol32_1/001.html)
- 3) <https://www.databricks.com/jp>
- 4) <https://www.microsoft.com/ja-jp/power-platform/products/power-bi>

### ◆問い合わせ先

NTTドコモ  
R&D戦略部

\*7 ワードクラウド：単語の出現頻度を文字の大きさと表示する可視化手法で、文字が大きいほど出現回数が多いことを表します。