

生成AIを活用した子育て分野のチャットボットサービス 実証運用第2弾のまとめ

令和6年12月23日

別府市 企画戦略部 情報政策課

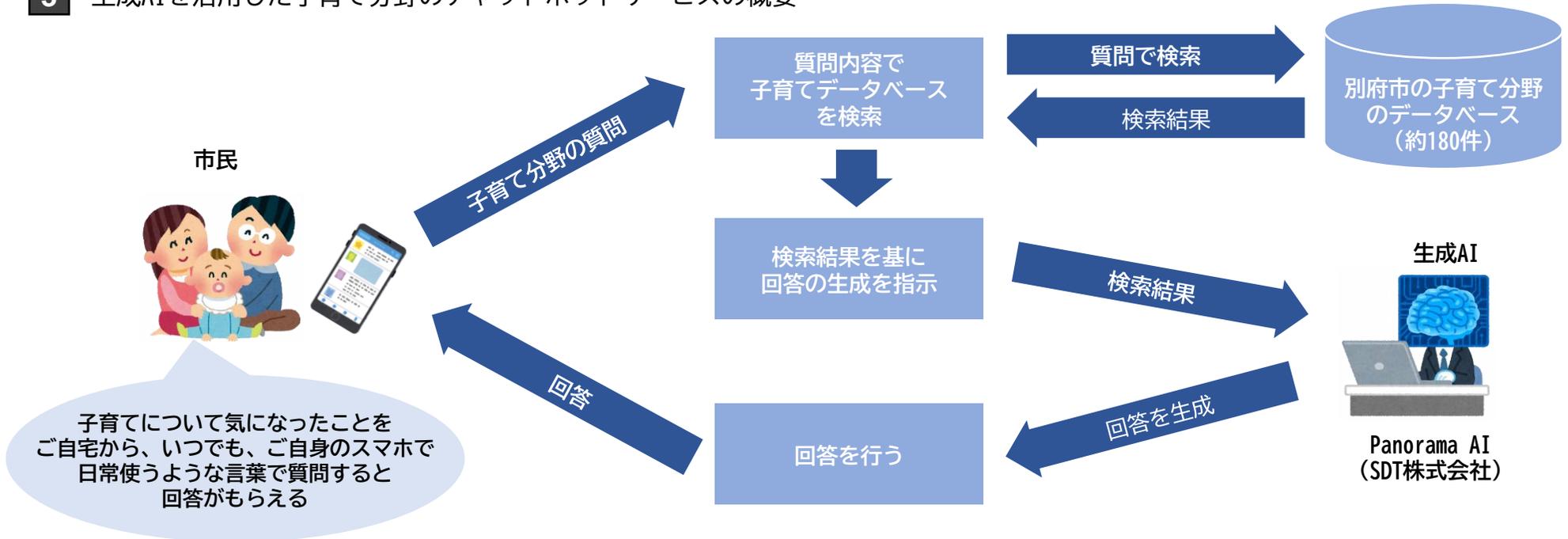
実証運用の概要

1 実証運用期間 : 令和6年8月20日 ~ 9月30日 (42日間)

2 実証運用実施の背景

- 「生成AIを活用した子育て分野のチャットボットサービス(*1)」の実証運用は、令和6年1月23日に“大分大学医学部”、“SDT株式会社”、別府市の三者で締結した「生成AIを活用した事業の実証運用に関する連携協定」に基づき実施しました。
- 本サービスは、市民がご自宅からいつでもご自身のスマートフォンやパソコン等を利用して子育てに分野に関する質問を受け付け回答することにより、デジタルファースト宣言の「いかにくていい市役所」の実現、市民の利便性の向上を目指した取り組みです。
- 令和6年3月～実証運用第1弾(*2)を実施し、その結果を踏まえ、実証運用第2弾を実施しました。

3 生成AIを活用した子育て分野のチャットボットサービスの概要



*1 チャットボットサービス: 「チャット」と「ボット」を組み合わせた言葉で、人工知能などを活用した「自動会話プログラム」のこと

*2 実証運用第1弾のまとめはこちらを参照 <https://www.city.beppu.oita.jp/doc/sisei/df/henkaku/ai2/matome.pdf>

実証運用の概要

4 実証運用第1弾の結果および課題

■ 実証運用第1弾の結果

- ・ 利用件数 139 件、回答できた件数 75 件 (54%)
- ・ 回答のうち、正答数 70 件 (正答率93.3%)、回答内容に一部情報の不足があるもの (不十分な回答) が 7 件 (6.7%)
- ・ 生成AIが勝手に作り出したような完全に誤った回答は無かった
- ・ 回答ができなかった質問の中で、子育てに関するものが 42 件 (未回答全体の65.6%)

■ 実証運用第1弾での課題と評価

課題と対応の方向性

課題① 正しい回答を行う

対応の方向性：データベースの拡充を行う、利用上を常に監視しながら回答の調整を行う運用ルールを確立する

課題② 利用数を増やす

対応の方向性：利用者の入口を公式ホームページから利用可能にする、サービス内容と利用方法を積極的に周知する

課題③ 費用対効果に優れた最新のLLM(*1)を利用する

対応の方向性：生成AI関連の技術の動向を常に確認する、新たな技術の利用について検討する

評価

課題対応を進めることにより、市役所への問い合わせ対応に生成AIの活用は可能と評価

*1 LLM：「大規模言語モデル」 生成AIはこのLLMを応用して自然な会話ができるように特化した対話型のAI

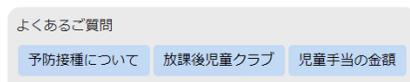
実証運用の概要

5 実証運用第1弾の課題への対応

課題①「正しい回答を行う」への対応

■ よくある質問を初期表示

市民から問合せのある質問については初期表示するように改善を図った。
よくある質問は時間帯（日中、夜間、週末）で表示内容を変更。



よくある質問（日中）



よくある質問（夜間）



よくある質問（週末）

課題②「利用数を増やす」への対応

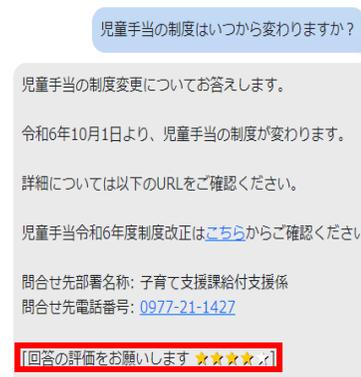
■ 別府市LINE公式アカウントのほか、公式ホームページにWEB小窓を設置、チラシの配布や別府市公式notelにて周知を図った。



別府市LINE公式アカウントメニューからの利用イメージ

■ 回答内容について評価アンケートを設置

回答内容について、利用者からの評価をもらえるように改善を図った。回答内容について5段階評価を行う。



評価アンケート



別府市ホームページトップページに設置したWEB小窓からの利用イメージ

実証運用の概要

5 実証運用第1弾の課題への対応（続き）

課題③「費用対効果に優れた最新のLLMを利用する」への対応

■ LLMの評価と変更

実証運用第1弾ではGemini 1.0 Proを利用したが、比較検討した結果、正確性を犠牲にすることなく、より柔軟に対応できるClaude3.5 Sonnetを採用した(*)。

* LLMの比較についてはこちらを参照 <https://beppu-city.note.jp/n/n8310f1a4427b>

質問

児童手当は、高校生はもらえますか？

回答

Claude3.5 Sonnet

児童手当についてお答えします。

児童手当は中学校修了までの児童（15歳到達後、最初の3月末までにある児童）を対象としています。したがって、高校生は児童手当の対象外となります。

問合せ先部署名称: 子育て支援課給付支援係
問合せ先電話番号: [0977-21-1427](tel:0977-21-1427)
詳細は[こちら](#)からご確認ください

Gemini 1.0 Pro

申し訳ありません。回答がわかりません。

Command-R+

児童手当についてお答えします。

別府市では中学校修了までの児童（15歳到達後、最初の3月末までにある児童）を対象に児童手当が支給されます。

6 実証運用第2弾の目的

生成AIを活用したチャットボットサービスの本格運用を目指し、第1弾における課題対応の検証および本格運用に向けた運用面を確立する

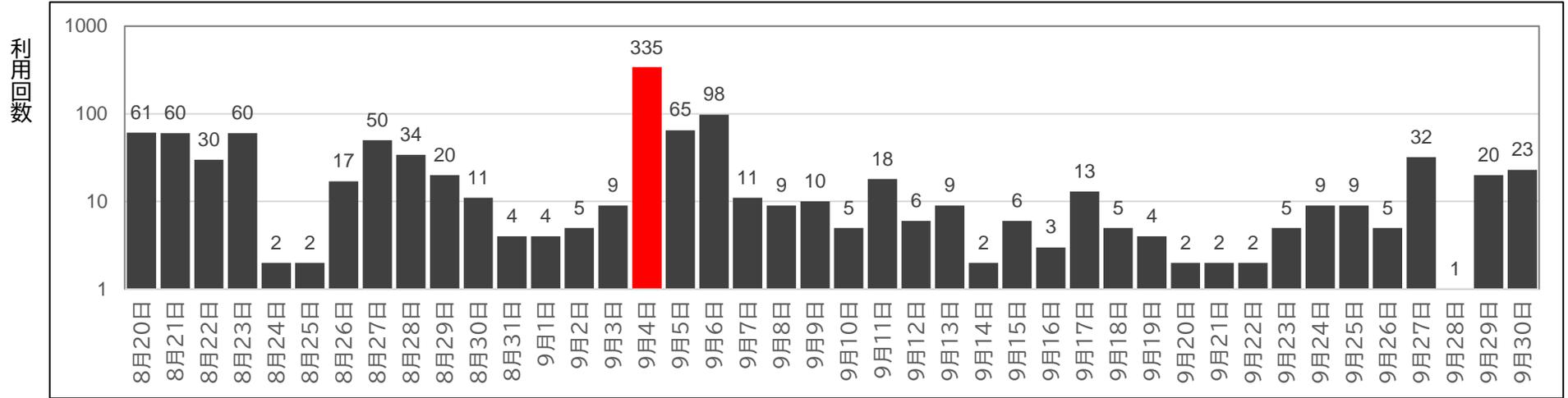
7 実証運用第2弾で確認したいこと

- ① 実証運用第1弾の課題に対応することで、自然言語での質問に対してより正しい情報を回答する
- ② 利用数が増加したか
- ③ 本格運用に向けた運用手順を確立できるか

実証運用の利用状況

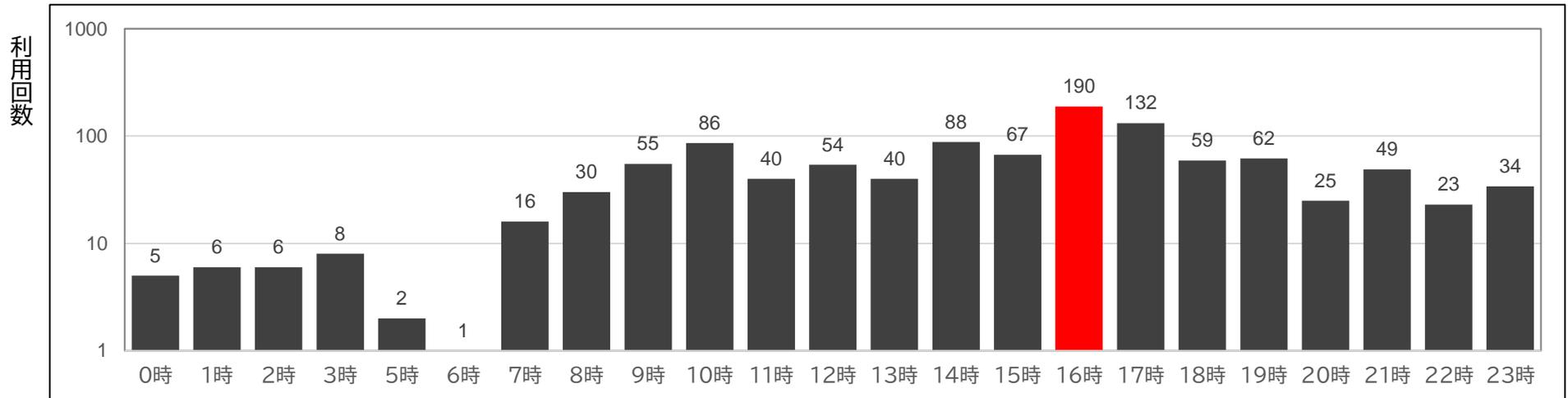
■ 利用回数：1,078 回

■ 日ごとの利用回数：9月4日が最多 335 回



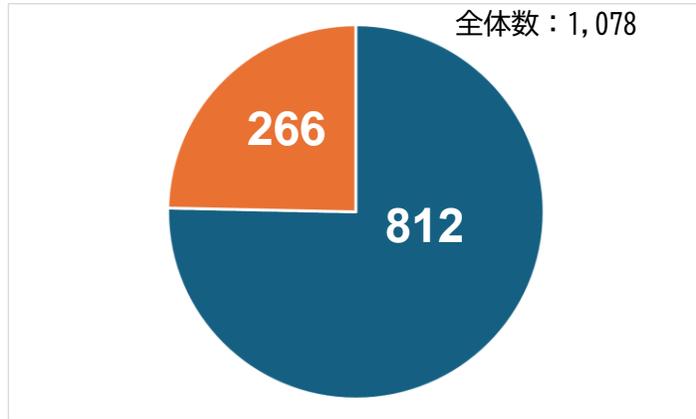
9月4日 別府市LINE公式アカウントの「友だち」全員に対して情報発信を行った

■ 時間帯ごとの利用回数：16時台が最多 190 回



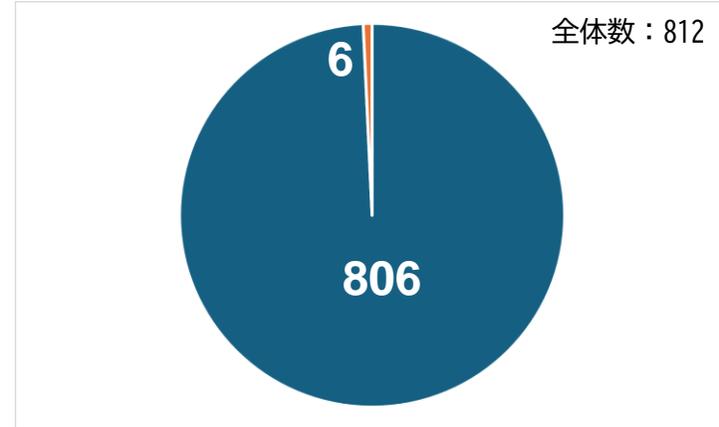
実証運用の利用状況

■ 質問に対して回答した件数



■ 回答ができた ■ 回答ができなかった

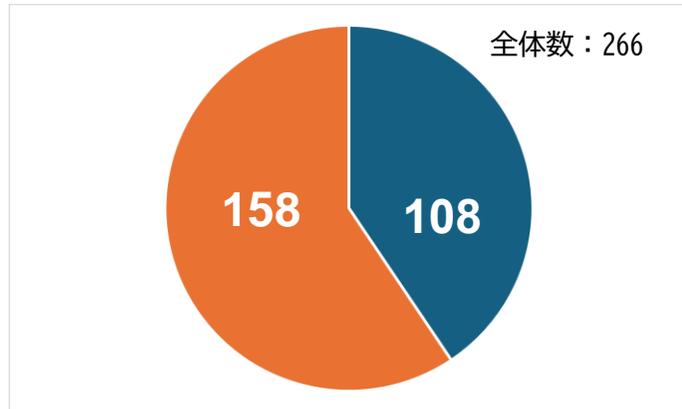
■ 回答した件数の内、正しい回答の件数



■ 正しい回答 ■ 不十分な回答

※不十分な回答：
回答内容に一部情報の
不足があるもの

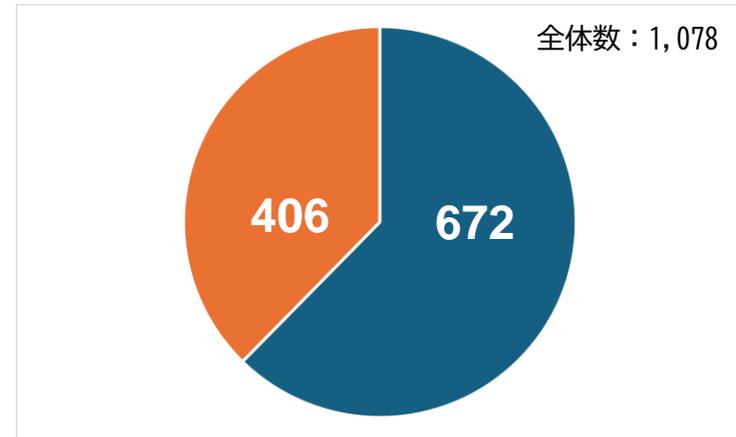
■ 未回答の質問分野の件数



■ 子育て ■ 対象外

※対象外：
質問内容が子育て分野ではないもの

■ 利用した経由



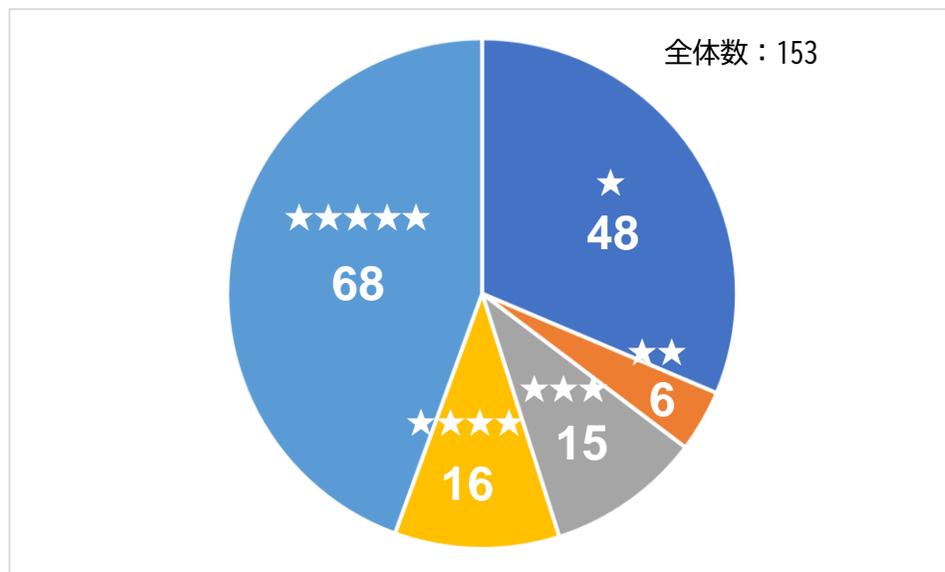
■ LINE ■ WEB

実証運用の利用状況

■ 回答が不十分と評価した質問とその理由

N0	質問	回答が不十分とした理由
1	1さいはんけんしん 場所	本来は正しい回答だったが、今年度のみ改修のため健診場所が変更となっていた
2	障害福祉課（*同一の質問が2件あり）	業務内容について限定した回答になっていた
3	給付金	一部回答にリンク切れを案内していた
4	中学生の子どもが一人いますが、どういった支援がありますか？	質問者の意図と違う回答と判断
5	こどもがなきやみません。どうすればよいか？	質問者の意図と違う回答と判断

■ 評価アンケート（5段階評価）の結果



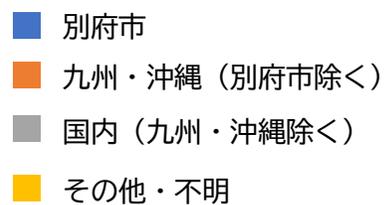
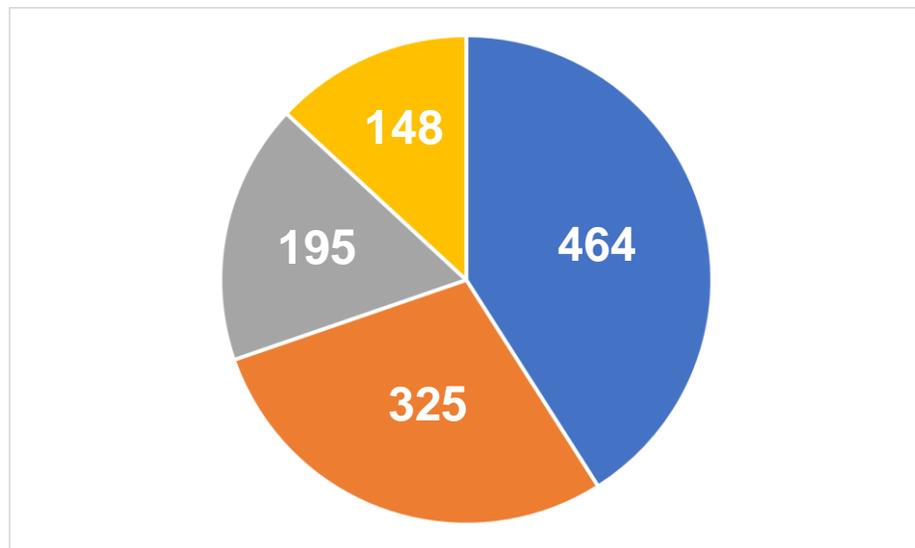
実証運用の利用状況

■ 実証運用第1弾との比較

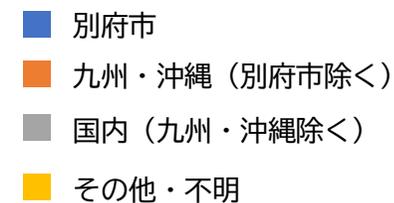
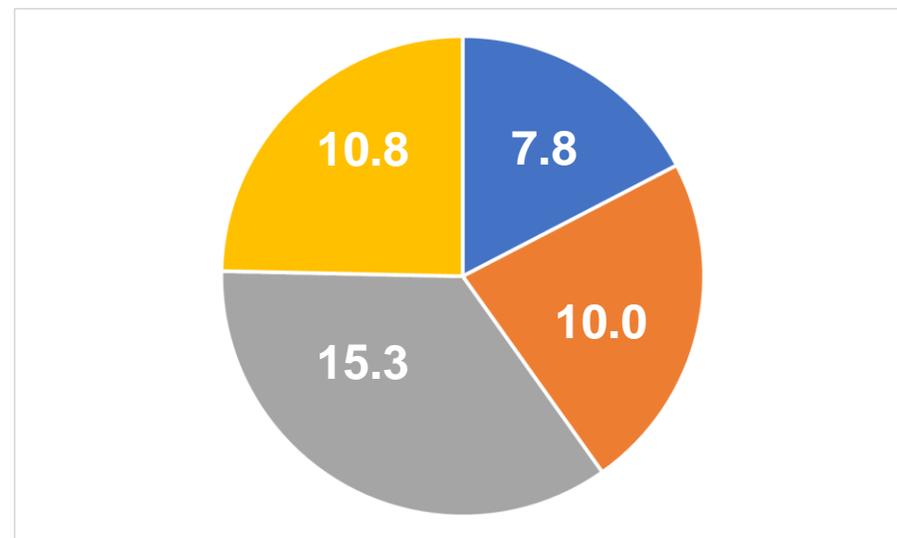
	実証運用第1弾	実証運用第2弾	差分
質問数	139	1,078	+ 939
回答できた数 (回答率)	75 (54.0%)	812 (75.3%)	+ 21.3 ポイント
正しい回答 (正答率)	70 (93.3%)	806 (99.3%)	+ 6.0 ポイント
誤った回答	5 (6.7%)	6 (0.7%)	△ 6.0 ポイント
一部不正確	5 (6.7%)	6 (0.7%)	△ 6.0 ポイント
完全間違い	0 (0%)	0 (0%)	増減なし
未回答数 (未回答率)	64 (46.0%)	266 (24.7%)	△ 21.3 ポイント
子育て分野の質問	42 (65.6%)	108 (40.6%)	△ 25.0 ポイント
子育て分野以外の質問	22 (34.4%)	158 (59.4%)	+ 25.0 ポイント

実証運用の利用状況（Googleアナリティクスデータ）

■ 地域別アクティブユーザー



■ 平均エンゲージメント時間（単位：分）



- ・ アクティブユーザー サイト内で特定の行動を取ったユーザーの数
- ・ 平均エンゲージメント時間 アクティブユーザーがWebサイトやアプリを実際に見ていたり、読んでいたり、使用している平均時間

利用状況のまとめと実証運用の評価

■ 確認したいこと① 実証運用第1弾の課題に対応することで、自然言語での質問に対して、より正しい情報を回答する

- ・回答ができた割合が **75.3 %**、第1弾の実証運用と比較し **21.3** ポイント上昇した
- ・回答ができた内、正しい回答は **99.3 %**、第1弾の実証運用と比較し **6** ポイント上昇した
- ・一部不正解の回答は **6** 件、回答数の内 **0.7 %**、第1弾の実証運用と比較し **6** ポイント減少した
- ・生成AIが勝手に作り出したような完全に誤った回答はなかった
- ・回答ができなかった質問の中で子育てに関するものが **108** 件、未回答中 **40.6 %**、第1弾の実証運用と比較し **21.3** ポイント減少した
- ・回答できなかったものとして、子育てに関する悩みについての質問が一定数あった

■ 確認したいこと② 利用数が増加したか

- ・利用数は **1,078** 回、1日平均 **26.3** 回であり、実証運用1弾と比較し **939** 件増加した
- ・現状サービス提供中の別府市LINE公式アカウントのシナリオ形式のチャットボット（1日平均5.5回）と比較しても利用数は増加している
- ・別府市LINE公式アカウントで実証運用についてメッセージ配信した9月4日の利用数が多く、周知を行うと利用数が伸びる傾向が見受けられる
- ・WEB小窓経由の割合も **37.7 %**あり、利用者数が増えた一因と考えられる
- ・実証運用第2弾では市外からの利用も一定数見られ、note等での情報発信が利用数増加につながったものと推測する

■ 確認したいこと③ 本格運用に向けた運用手順を確立できるか

- ・実証運用期間中、質問と回答内容を確認し、実証運用期間中にRAG用のデータベースの調整を **8** 件実施した
- ・確認作業の期間は、最初は毎日実施していたが、調整作業対象が減少してきたため、2週間後には日々の確認を週1度の確認に変更した

■ 評価

- ・回答の精度は上がった評価する（実証運用を通して精度を上げていくための経験値が蓄積された）
- ・利用数は増加したと評価する（入口を増やす、適切なタイミングでの周知は重要）
- ・運用手順についてのポイントは整理できたと評価する（実施しながらの成長を目指し定期的な確認作業は重要）

まとめ

今回の実証運用においても、生成AIが勝手に作り出したような回答はなく、正しい回答は99.3%、不十分な回答（回答内容に一部情報の不足があったもの）が0.7%と、実証運用第1弾と比較しても、利用数、正答率ともに向上がみられた。

利用者に不利益を与えるような回答は見受けられなかったが、子育て関連の質問で回答ができなかったものが全体の約10%あった。

正答率は向上しているが、まだ、回答すべき質問に回答できていない事例も見受けられ、今後も継続した精度の向上、回答率の向上に向けた取り組みが必要であると感じている。

精度の向上、回答率の向上に向けては、LLMの選択、データの拡充が必要であるものの、同時に重要なこととして、サービスを開始してから生の質問を受けながら調整していく作業であると感じている。

そのためには、利用状況を監視しながらデータの調整を行なっていく運用体制の確立など、継続した取り組みが行える“仕組みづくり”が必要と感じている。

二度の実証運用を経験して、生成AIの能力は素晴らしいものがあり市役所の問合せ対応業務の入口としての実装が可能であると評価している。

今後は、本番稼働を目指し、回答できなかった質問を中心に子育て分野のデータを拡充しながら、さらに、子育て分野以外の幅広いデータを整備しながら、市民がいつでも、どこからでも行政サービスに対して質問し正しく回答し、必要な手続きを案内していく、デジタル総合窓口の構築に取り組んでいきたい。