

VOICE AI AGENT / FIELD REPORT

ごみ分別案内 Voice AI 実証実験結果報告書

実証期間 令和 8 年 7 月 1 日から 7 月 31 日まで

本資料について

本資料は、実証期間中の全 394 件の利用実績と、抽出した 191 件の通話内容に対する品質評価をまとめたものです。全件集計と抽出評価では対象範囲が異なるため、各図表に対象件数を明記しています。

令和 8 年 8 月

三次市

1 結果の概要

実証期間中には 394 件の着信があり、平日の日中以外の利用が半数を超えました。抽出した通話の品質評価では、一定の自動対応効果を確認した一方、音声認識や回答情報の充実など、継続的な改善が必要であることも明らかになりました。

394 件	53.8%	41.4%	71.5%
実証期間中の着信	平日日中以外の利用	AI 単独対応完了率	回答正答率

確認できたこと

- 休日や夜間を含め、時間を問わず問い合わせを受け付ける利用形態に一定の需要があった。
- 主な問い合わせは、品目の分別・出し方と、収集日・地区の確認だった。
- 抽出した 191 件のうち 79 件（41.4%）は、職員へ転送せず AI のみで対応を完了した。
- 実際の問い合わせから、改善を優先すべき品目、地区名、会話パターンを把握できた。

確認された主な課題

- 品目名や地区名などの音声認識
- 登録されていない品目や質問への対応
- 相づちなどによる発話の中断
- 誤案内を防ぐ回答ルールと職員転送の判断

結果の読み方

品質評価は全 394 件ではなく、評価期間を 4 区分して抽出した 191 件が対象です。また、正誤を判定できない通話を除くため、指標ごとに分母が異なります。

2 実証実験と評価方法

2.1 実証実験の概要

項目	内容
実証期間	令和 8 年 7 月 1 日から 7 月 31 日まで（31 日間）
対象業務	市民からのごみ分別、出し方、収集日、地区確認等の電話問い合わせ
対応方法	音声 AI が一次対応し、必要に応じて職員への転送を案内
利用実績の対象	実証期間中の全 394 件
品質評価の対象	全 394 件から期間を分けて抽出した 191 件

2.2 品質評価の抽出方法

評価期間を 4 つに区分し、各区分から 50 件、50 件、50 件、41 件の計 191 件を抽出して、通話ごとに判定しました。抽出評価は全通話の利用実績とは別に扱い、各指標では判定対象外（N/A）を分母から除いています。

2.3 評価指標

指標	定義	分母
事実に基づかない回答の非発生率	事実に基づかない内容を AI が案内しなかった割合	判定可能 175 件
回答正答率	分別区分や収集日等を正しく回答できた割合	正誤判定可能 130 件
AI 対応完了率	AI 単独完了、転送成功または対応自体に問題なしと判定した割合	判定可能 172 件
AI 単独対応完了率	職員へ転送せず AI のみで対応を完了した割合	抽出 191 件

重要

「AI 対応完了率」は、AI だけで問い合わせを解決した割合ではありません。AI のみで対応完了した通話に加え、転送成功や、回答できない場合に AI の対応自体には問題がなかった通話を含みます。

3 利用実績

3.1 基本集計

394 件	7 時間 32 分 1 秒	68.8 秒	61.5 秒
総着信件数	総通話時間	平均通話時間	中央値

通話の 48.2%は 60 秒以内、91.4%は 120 秒以内でした。0 秒の通話が 3 件あり、いずれも接続切断として記録されています。

3.2 日別の利用状況

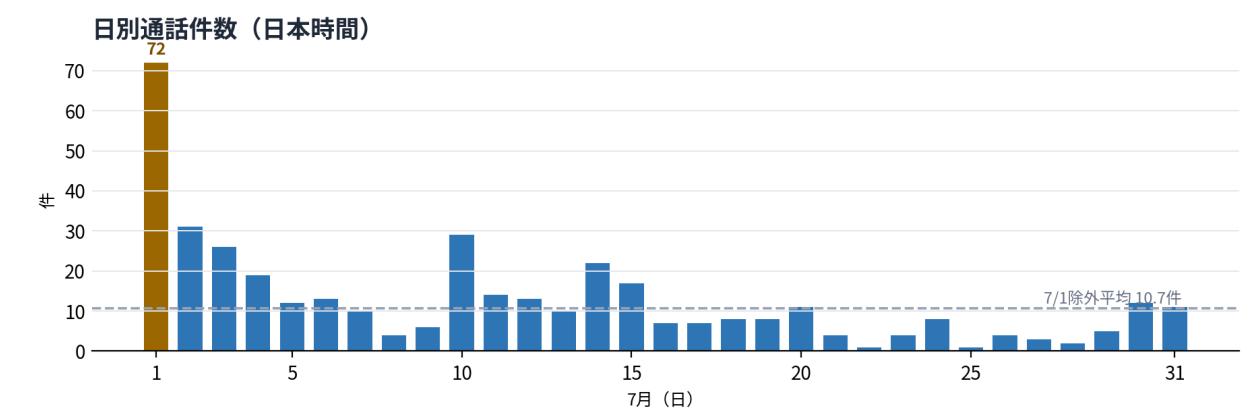


図 1 日別通話件数（日本時間、全 394 件）

期間	通話件数	総通話時間	職員への転送
7 月 1 日から 7 日	183 件	217.6 分	5 件
7 月 8 日から 14 日	98 件	110.2 分	5 件
7 月 15 日から 21 日	62 件	72.3 分	7 件
7 月 22 日から 28 日	23 件	22.7 分	1 件
7 月 29 日から 31 日	28 件	29.2 分	2 件

7 月 1 日は 72 件で、全体の 18.3%を占めました。初日を除く 30 日間の平均は 10.7 件／日です。実証開始時の周知や試行的利用が影響した可能性があるため、日別の増減だけから継続的な需要を判断することはできません。

4 利用時間帯と 24 時間対応の可能性

182 件	212 件	53.8%	79 件
平日 8:30 から 17:15	上記以外	平日日中以外の割合	土日の利用

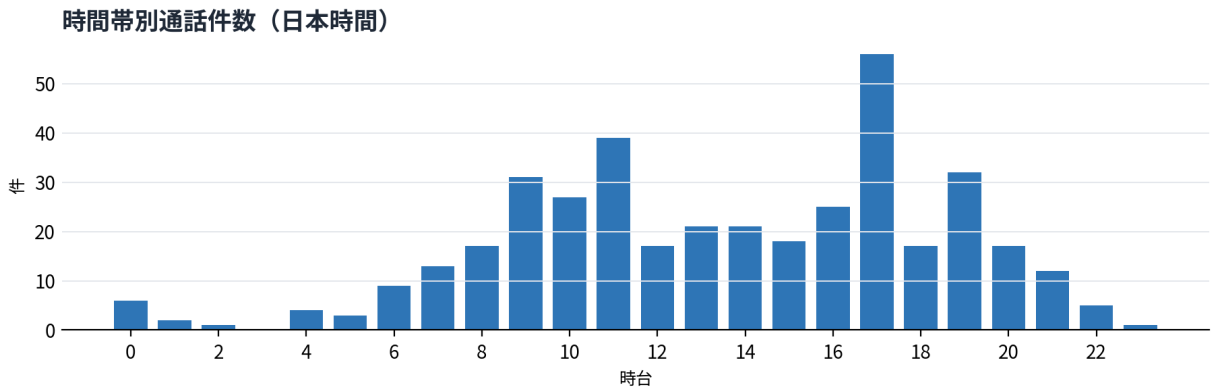


図 2 時間帯別通話件数（日本時間、全 394 件）

平日の午前 8 時 30 分から午後 5 時 15 分まで（7 月 20 日の祝日を除く）に該当する通話は 182 件、それ以外は 212 件でした。休日・夜間・早朝を含む時間帯の利用が 53.8%を占めており、24 時間問い合わせを受け付けられることは音声 AI の主な利点の一つと考えられます。

集計上の注意

営業時間区分は効果を確認するための便宜的な区分です。休日、ごみ収集日、周知施策などの影響を切り分けたものではありません。

4.1 通話終了理由

終了理由	件数	構成比	解釈上の注意
利用者が終了	187 件	47.5%	正常終了と途中離脱を区別できない
AI が終了	184 件	46.7%	正常完了の可能性はあるが、解決結果を示すラベルではない
職員へ転送	20 件	5.1%	転送後の解決状況は含まない
接続切断	3 件	0.8%	全件 0 秒の接続事象

5 問い合わせ内容

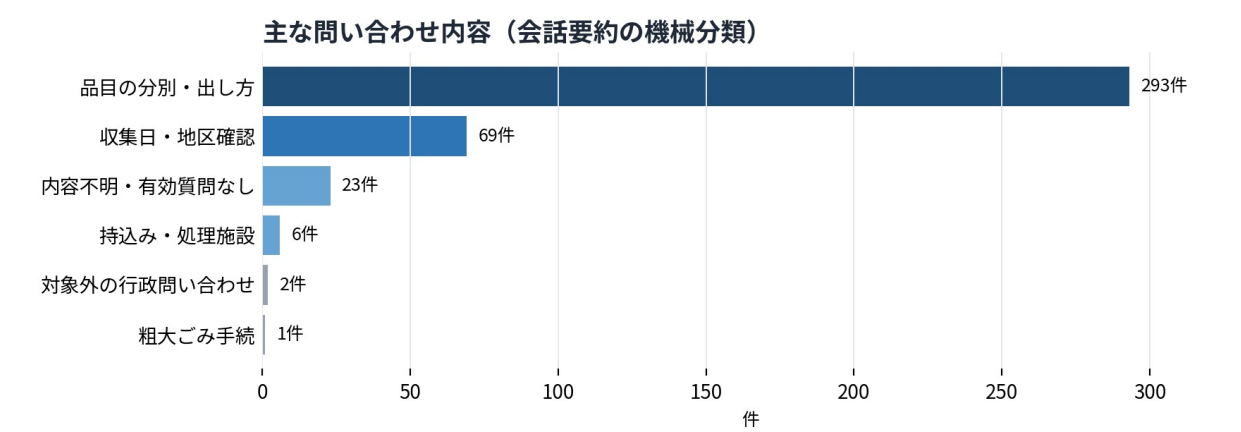


図3 主な問い合わせ内容（会話要約の機械分類、全 394 件）

分類	件数	構成比
品目の分別・出し方	293 件	74.4%
収集日・地区確認	69 件	17.5%
持込み・粗大ごみ手続	7 件	1.8%
対象外の行政問い合わせ	2 件	0.5%
内容不明・有効質問なし	23 件	5.8%

品目の分別や出し方に関する問い合わせが約4分の3を占めました。複数計上した頻出テーマでは、収集日・地区、粗大ごみ、プラスチック・容器包装、電池・充電電池などが多く、今後の回答情報の整備では、これらを優先することが有効と考えられます。

分類方法

自動生成された会話要約を、語句と優先順位ルールにより1通話1分類とした参考集計です。複数の質問が含まれる場合は主な内容のみを示し、全文音声を手分類した結果ではありません。

6 抽出 191 件の品質評価

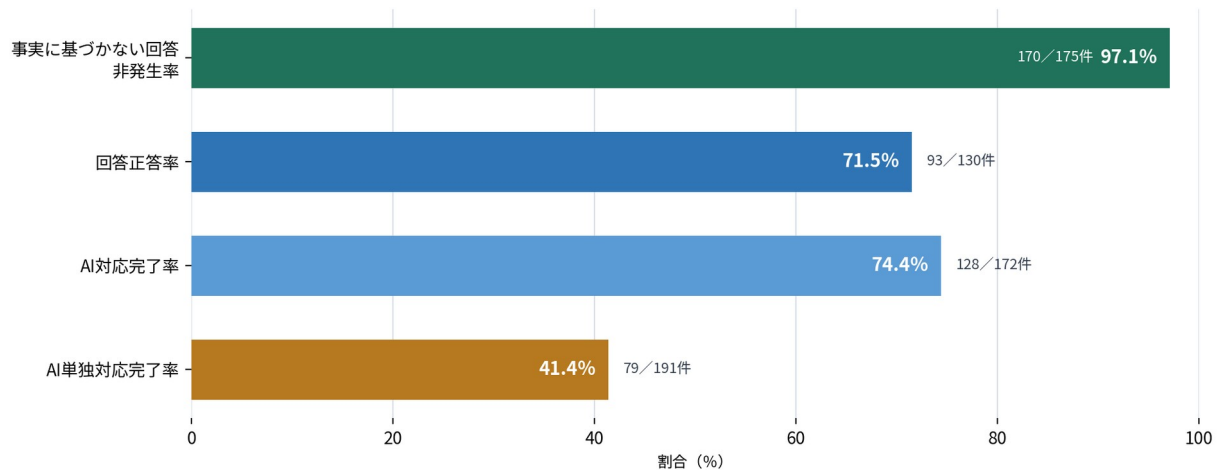


図 4 品質評価結果（指標ごとに分母が異なります）

指標	結果	判定内訳	判定対象外
事実に基づかない回答の非発生率	97.1%	発生なし 170 件 / 発生 5 件	16 件
回答正答率	71.5%	正答 93 件 / 誤答・回答失敗 37 件	61 件
AI 対応完了率	74.4%	対応完了 128 件 / 対応上の問題 44 件	19 件
AI 単独対応完了率	41.4%	AI のみで完了 79 件 / 抽出 191 件	—

6.1 回答正答率の対象外

回答正答率では、回答に必要な情報が未整備で正誤を判定できなかった 48 件と、会話が成立しなかったなど判定できなかった 13 件の計 61 件を分母から除いています。71.5%は、全 191 件に対する正答の割合ではありません。

6.2 事実に基づかない回答

判定可能な 175 件のうち 5 件で、事実に基づかない回答が確認されました。非発生率だけでなく、発生した事例を再発防止につなげることが重要です。

7 AI 対応完了率の内訳

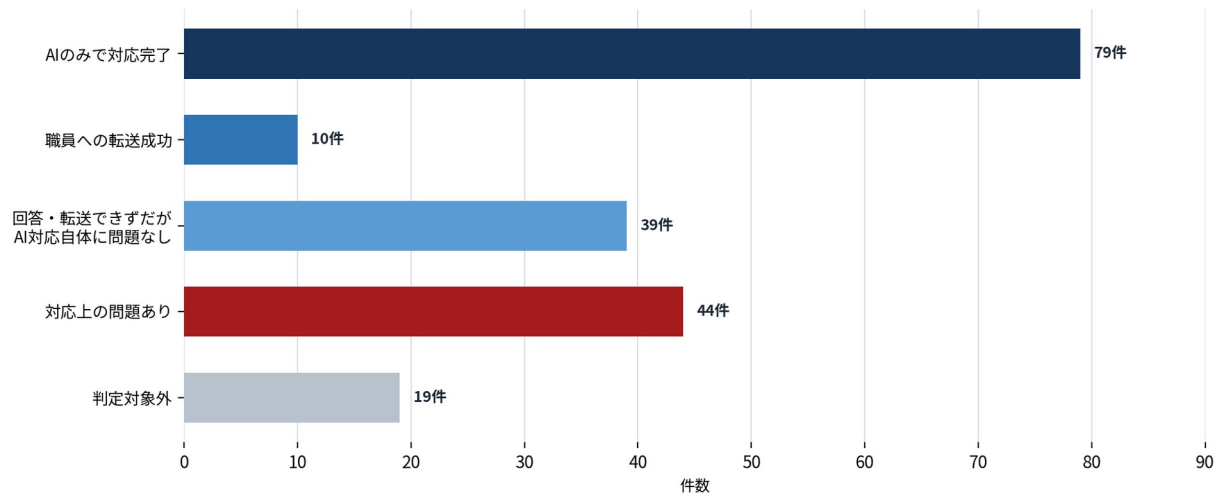


図5 抽出 191 件の対応結果

AI 対応完了と判定した 128 件は、次の 3 種類で構成されます。

区分	件数	意味
AI のみで対応完了	79 件	職員へ転送せず、AI のみで対応を完了
職員への転送成功	10 件	必要に応じて職員へ転送
AI 対応自体に問題なし	39 件	回答・転送できなかったが、回答不能時の案内など AI の対応自体に問題なし
対応上の問題あり	44 件	誤答、回答失敗、誤転送などを確認
判定対象外	19 件	会話が成立しないなど、対応結果を判定できない

指標の解釈

AI 対応完了率 74.4%を「AI が 74.4%の問い合わせを解決した」と読み替えることはできません。AI だけで対応を完了した割合は 41.4%です。

7.1 職員への転送

全 394 件のうち、システム上、職員への転送で終了した通話は 20 件（5.1%）でした。ただし、AI 導入前の電話対応時間や転送後の職員対応時間がないため、この数値だけから職員負担の削減率を算定することはできません。

8 確認された課題と改善の方向性

課題	確認された状況	改善の方向性
音声認識	品目名や地区名を正しく聞き取れず、回答できない事例	品目・地区名辞書の充実、認識モデルと聞き返し方法の改善
回答情報の未整備	登録されていない品目や想定外の質問に回答できない事例	頻度と安全性を踏まえて優先順位を付け、回答情報を追加
会話の中断	相づちなどを利用者の新たな発話と捉え、案内が中断する事例	相づちを考慮した会話制御の改善
誤案内	分別区分等を誤って案内する事例	回答ルール、確認手順、正解情報との照合を強化
職員への転送	転送の可否判断、転送提案、転送後の評価が不十分	難しい質問は正確性を優先して転送し、結果を記録

8.1 改善の優先順位

- 安全性や市民生活への影響が大きい誤案内の防止
- 問い合わせの多い品目、電池類、材質・大きさで区分が変わる品目の回答整備
- 地区名や品目名の聞き取り精度と、聞き返し方の改善
- AI で確実に回答できない場合に職員へつなぐ判断ルールの強化
- 改善前後を同じ評価方法で比較できる判定記録の継続

基本的な考え方

回答できる範囲を広げるだけでなく、回答の確実性が低い場合に無理に回答せず、職員確認へつなぐことも品質向上として評価します。

9 評価上の限界と今後の検証

評価領域	現時点	理由・今後必要な確認
利用実績・稼働状況	評価可能	件数、日時、通話時間、終了理由を全件集計
平日日中以外の利用	評価可能	日本時間により営業時間相当を区分
問い合わせ内容	参考評価	自動生成された会話要約を機械分類
回答品質	抽出評価	191 件を対象とし、指標ごとに判定対象外を除外
利用者満足度	評価不可	通話後アンケート等を実施していない
職員負担削減	評価不可	導入前の対応時間、転送後の対応時間、確認工数が不足
費用対効果	評価不可	運用費用と削減時間等の比較が必要

9.1 今後確認する事項

- 誤案内、転送、短時間終了など、重要事例の重点確認
- 音声・全文文字起こしと正解情報を照合した継続評価
- 改善前後で同じ基準を用いた回答正答率・単独対応完了率の比較
- 従来の電話対応件数と職員対応時間の把握
- 必要に応じた利用者の意見・満足度の確認

9.2 まとめ

今回の実証では、平日日中以外にも多く利用され、音声 AI による 24 時間の一次対応に一定の可能性が確認されました。一方、回答正答率や個別の誤案内事例からは、音声認識、回答情報、会話制御、職員転送の改善が必要です。今後は、正確性と安全性を優先しながら、利便性と職員負担軽減の両面を検証します。

個人情報の取扱い

本資料には、発信元電話番号、Call ID、個別の会話要約など、通話を特定できる情報は掲載していません。

10 集計条件・資料

10.1 集計条件

項目	条件
期間	令和 8 年 7 月 1 日から 7 月 31 日まで
利用実績	1 行を 1 通話として全 394 件を集計
時刻	日本時間を使用
平日日中	平日 8 時 30 分から 17 時 15 分まで。7 月 20 日の祝日を除外
問い合わせ分類	自動生成された会話要約を、優先順位付きルールで 1 通話 1 分類
品質評価	期間を 4 区分して計 191 件を抽出し、通話単位で判定
割合	原則として小数第 2 位を四捨五入し、小数第 1 位まで表示

10.2 参照資料

- ・ 三次市「ごみ分別案内 Voice AI 実証実験レポート 第 2 次集計案」
- ・ 株式会社 Verbex「PoC 実施報告書（市民向けごみ収集問い合わせ対応）」
- ・ 株式会社 Verbex「三次市様 課題管理シート 市民向け利用」
- ・ 実証期間中のコールログおよび通話履歴

10.3 用語

用語	説明
音声 AI（Voice AI）	電話で利用者の発話を認識し、音声で案内する仕組み
音声認識	利用者の発話を文字情報として認識する処理
回答情報	AI が回答を作成する際に参照する、ごみ分別等の登録情報
ハルシネーション	事実に基づかない内容を AI が回答すること
N/A（判定対象外）	会話が成立しない、必要な回答情報がないなど、指標を判定できないもの