

内部実証結果の概要

Copilot Studio を活用した AI チャットボット作成等業務委託

本資料は、令和 8 年度に実施した生成 AI チャットボットの内部実証について、回答の正確性評価及び分野別の主な課題を整理したものである。受注者は、本業務における課題の原因分析及び改善の基礎資料として活用すること。

1 内部実証の概要

内部実証期間: 令和 8 年 6 月下旬から約 1 か月間

対象分野: 福祉の総合相談窓口相談業務、会計事務の 2 分野

ボット作成者: サービスデザイン課の職員

利用対象: 制度所管課の職員、サービスデザイン課の職員 外

ボットの構成: RAG(検索拡張生成)により、各種事務マニュアル・手引・例規等をナレッジとして参照する構成とし、用途は検索・回答に限定した。

指示文: 制度の該当可否や最終判断を断定せず、確認事項・確認先を示す等の最低限の指示文のみを設定した。

外部情報の参照: 「根拠のない応答を許可する」及び「Web の情報を使用する」機能はオフとした。ただし、福祉の総合相談窓口相談業務については、特定の URL(外部サイト)のみをナレッジソースとして追加した。

ナレッジの参照方法: ナレッジは、SharePoint 上の所定フォルダをナレッジソースとして接続し、当該フォルダ内に格納した資料を参照する構成とした。

ログ取得: 利用状況及び回答内容を把握するため、利用日時、会話 ID、利用者、質問内容等を記録した。

2 回答の正確性評価

内部実証期間中の利用ログ及びフィードバックに基づく、分野別の回答の正確性評価は次のとおりである。

分野	評価件数	○ 概ね正答	△ 情報不足等	× 誤り
福祉	43	24(約 56%)	16(約 37%)	3(約 7%)
会計	34	17(約 50%)	6(約 18%)	11(約 32%)

※ 正確性評価: 制度所管課の職員が、チャットボットの回答に対し、業務上の知見や根拠資料との整合を確認し、「○ 概ね正答」「△ 情報不足等」「× 誤り」の 3 区分で評価した。なお、利用された全ての質問について正確性評価を行ったものではなく、制度所管課から評価入力があったものを集計している。

3 分野別の主な課題

内部実証を通じて把握された、分野別の主な課題は次のとおりである。

なお、記載した課題は内部実証時に確認された事象を整理したものであり、その原因については、ナレッジの内容・構成、検索の挙動、指示文、対話設計等の観点から、本業務において分析するものとする。

(1) 福祉の総合相談窓口相談業務

- 単一の制度・窓口に関する案内は、比較的正確に回答できる傾向がある。
- 複数の制度が関係する複合的な相談では、一般的な支援策の列举にとどまり、相談者の実情に即した回答になりにくい。
- 本来と異なる窓口・関係機関を案内する誤りや、論点がずれた回答が生じる場合がある。
- 施策の分岐点となる情報(障がいの種別、年齢、住所、所得等)について、追加で確認する仕組みが求められる。
- 制度判断を確定させるより、確認すべき事項や関係機関の候補を提示する使い方が適している。

(2) 会計事務

- 参照すべき資料が明確な単一論点の質問は、比較的安定して回答できる。
- 参照すべきナレッジを明示的に指定しても、無関係なナレッジを参照して回答する誤りが発生する。
- 類似用語や複数の手引き・様式が関係する場面で、資料の取り違いによる誤回答が目立つ。
- 前の質問の文脈が維持されず、追加の聞き直しでも正しく話を引き継げないことがある。
- 回答の参照箇所を尋ねても、無関係な文書を参照する場合がある。