

生成 AI を用いた対話サービスの評価

1232106 塚本啓太郎

1232031 大塚玲奈

1232112 富永彩渚

1232090 相馬海人

近年、AI に関する要素技術の発展により、機械学習を用いたさまざまなアプリケーションが社会の多様な場面で利用されるようになってきている。中でも、自然言語でユーザーと対話を行う対話サービスは、企業の問い合わせ対応を行うチャットボットをはじめ、生成 AI の発展に伴い、柔軟な質疑応答を可能とする形態へと進化している。このような対話サービスは、ユーザーとのインタフェースとして重要な役割を担う一方で、実際に開発・導入する前段階において、その有効性や適用可能性をどのように評価するかという点については、十分に整理されていないという課題がある。特に、対話サービスの評価は、開発経験や専門的知識を持つ評価者に依存しやすく、誰もが容易に行えるものとは言い難い。

本研究は、専門的な知識や開発経験を持たない人でも、対話サービスの有効性を評価できる方法として、対話シナリオを入力とし、生成 AI に対話サービスの評価を行わせる手法を提案する。具体的には、想定ユーザーと対話アプリのやりとりを記載した対話シナリオを生成 AI に提示し、対話サービスの価値・コスト・リスクの観点から評価を行わせる。その際、評価の立場、評価軸および評価項目の定義、出力形式をプロンプト内で明確に指定することで、生成 AI の判断基準を統一し、評価結果のばらつきを抑えることを目指した。提案手法の有効性を検証するために、スポーツ練習メニュー、観光案内、教員指導案、ライフプラン設計、健康・栄養サポートといった複数の対話シナリオを対象とし、ChatGPT および Google Gemini を用いて評価実験を行った。その結果、評価の立場や評価軸を明確に指定しない初期段階のプロンプトと比較して、提案手法を用いた場合には、評価結果の形式や内容が安定し、異なる対話シナリオ間および生成 AI 間での比較可能性が向上することが確認された。

以上より、本研究で提案した手法は、生成 AI を用いた対話サービス評価において、プロンプト設計が評価結果の安定性と再現性に大きく影響することを示すとともに、対話サービスの初期設計段階におけるアイデア検討や優先度判断を支援する基礎的な枠組みを提供するものである。