

テキスト生成 AI の適用検討プロセスの研究

アブストラクト：

近年、テキスト生成 AI の利用領域と利用者層は拡大し、多くの企業が業務への適用を検討している。学術論文や企業のレポートなどでは様々な適用事例が紹介されており、適用検討にあたってのヒントとはなるものの、明確な適用可否の判断に利用するのは難しいというのが実態である。

本研究では「テキスト生成 AI の具体的適用可否判断プロセスが明確となっていない」という課題に対し、「業務要求を評価指標ごとに数値化しテキスト生成 AI の性能と比較する仕組みを確立すれば、テキスト生成 AI の適用可否を判断するプロセスを定義することができる」との仮説を立て、この判断を行うためのプロセスを定義した。

本プロセスは、Nejumi LLM リーダーボードの評価指標を利用してテキスト生成 AI を適用したい業務の業務要求を数値化し、リーダーボード上の LLM の性能と比較するものである。

検証にあたっては、メール文作成や翻訳等テキスト生成 AI の適用実績がある業務を例として、数値化した業務要求と LLM の性能の差分値を取得した。この値と、検証対象業務を LLM で実機検証した結果に対するアンケートの値を比較し、両者の組み合わせの妥当性を検証した。

その結果、いずれの業務でも妥当との評価が得られたため、本プロセスが有効であると結論づけた。

また、この手法を実際の適用検討時に活用できるよう、検証ツール「テキスト生成 AI 適用検討ガイドライン」として整備した。

キーワード：

テキスト生成 AI、Nejumi LLM リーダーボード、適用検討、業務要求、評価指標