

5-3-4)。

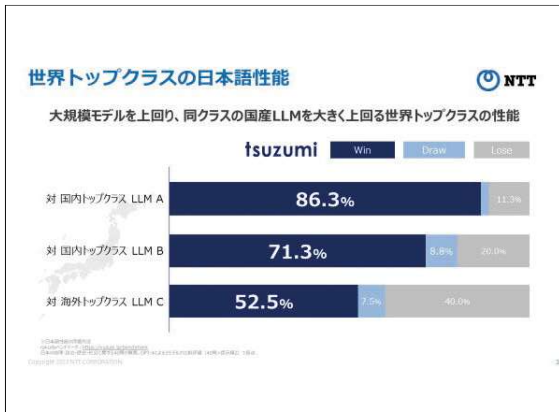
4つ目の特徴は「マルチモーダル」、すなわち、複数のデータ形式(モード)を処理できることである。

LLMというと、典型的には、テキストで質問を入力するとAIが回答をテキストで返してくるという利用シーンがイメージされるが、テキスト情報だけでなく、音声、画像、映像などの情報も取り扱うことができれば、応用範囲をさ

らに拡大することが可能になる。

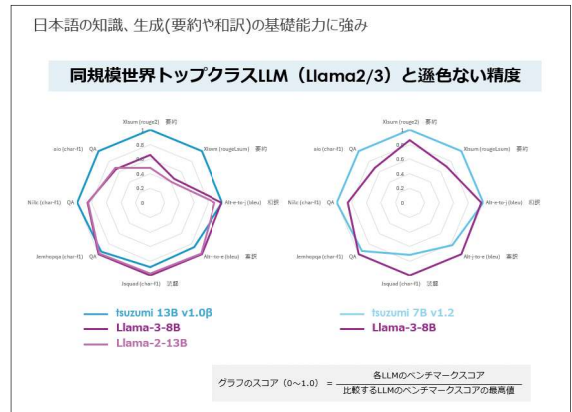
例えば、画像認識技術を活用して、様式の異なる見積書を読み取る機能を追加すれば、人力でデータを構造化する作業を省略し、そのままAIに情報をインプットすることが可能になる(図表5-3-5)。同様に、図表が含まれるスライド資料の中から、情報を読み取ることができれば、ウェブ上で多数公開されている、さまざまな企業のプレゼンター

図表5-3-2 ▶tsuzumiの日本語性能①



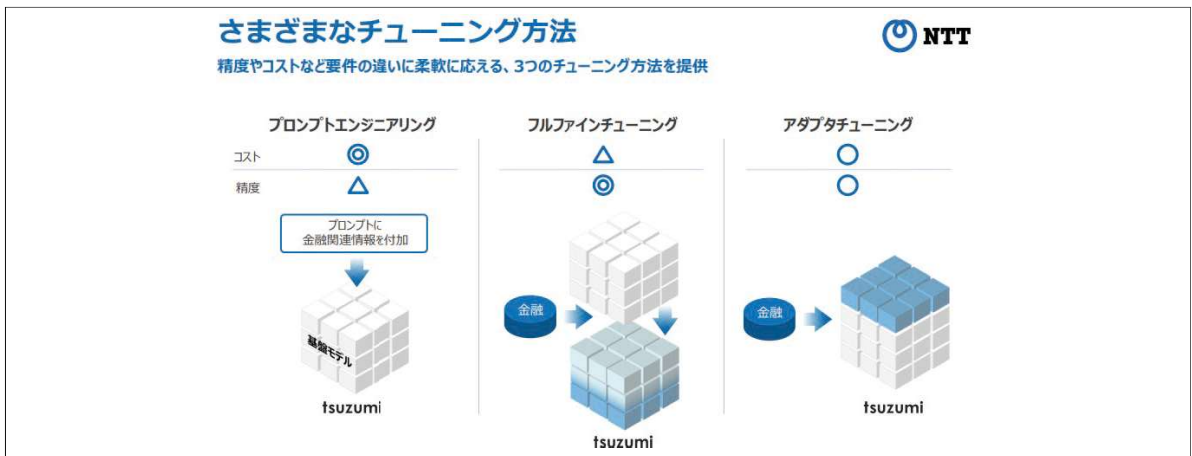
出所：NTT「NTT独自の大規模言語モデル『tsuzumi』を用いた商用サービスを2024年3月提供開始」(2023年11月1日)

図表5-3-3 ▶tsuzumiの日本語性能②



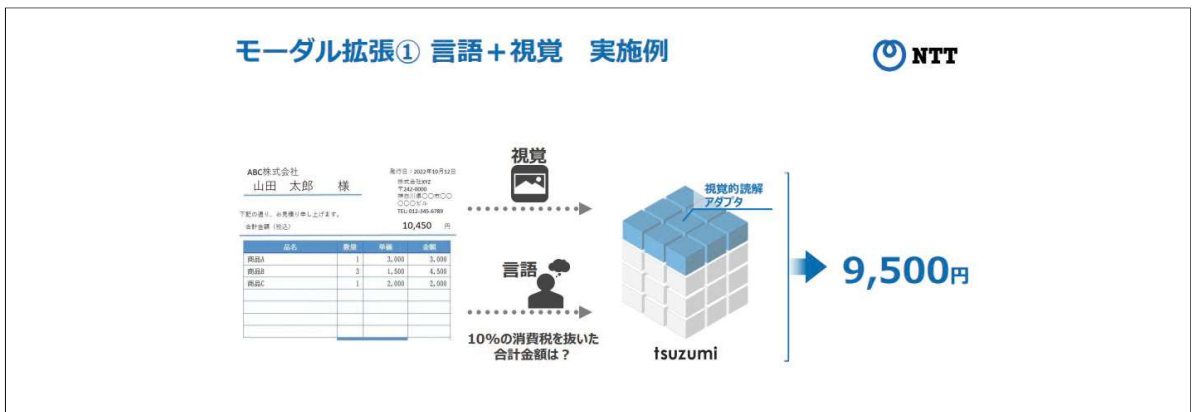
出所：NTT R&D フォーラム 2024 基調講演 木下部門長講演資料

図表5-3-4 ▶tsuzumiのチューニング方法



出所：NTT「NTT独自の大規模言語モデル『tsuzumi』を用いた商用サービスを2024年3月提供開始」(2023年11月1日) 記者説明会資料「NTT版LLM『tsuzumi』の概要」

図表5-3-5 ▶マルチモーダルの例(+視覚)



出所：NTT「NTT独自の大規模言語モデル『tsuzumi』を用いた商用サービスを2024年3月提供開始」(2023年11月1日) 記者説明会資料「NTT版LLM『tsuzumi』の概要」