

# 例文構築の主要な方法

- タスクデータセットのフォーマット
  - 既存の自然言語タスクのデータセットに、人手で書いたタスクの指示を追加する
  - 質問と回答のセットが既にある場合、「回答をベースに質問を考えて」と指示してデータを増やせる

# 例文構築の主要な方法

- 日々のチャットデータのフォーマット
  - 一般の自然言語処理データセットは多様性に欠け、実際の人間のニーズとマッチしない
  - ユーザが実際に入力した問い合わせ文や、人間のラベラーが生成した回答などを使用する
- 合成データのフォーマット
  - 手作業の負担を減らすために、既存の例文をLLMに入力してタスク指示と入出力分を自動で生成する