

例文構築のための重要な要素

- 指示の質と多様性が重要

- ❖ 命令のスケーリング

- タスク数を増やすことで、LLMの汎化能力が(一定のレベルに達するまでは)大幅に向上することがわかっている

- ❖ フォーマットの設計

- 入出力のデータに加えてタスクの説明や例を追加すると、大幅な改善につながる
- しかし、それ以外の要素(避けるべきことや理由など)を組み込んでも改善しないかむしろ悪影響

インストラクション・チューニングの戦略

- データ分布のバランスをとる
 - 種類の異なるタスクの比率のバランスをとる
- 事前学習との組み合わせ
 - チューニングを安定させるため、インストラクションチューニングの際に事前学習データを組み込む方法もある