

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

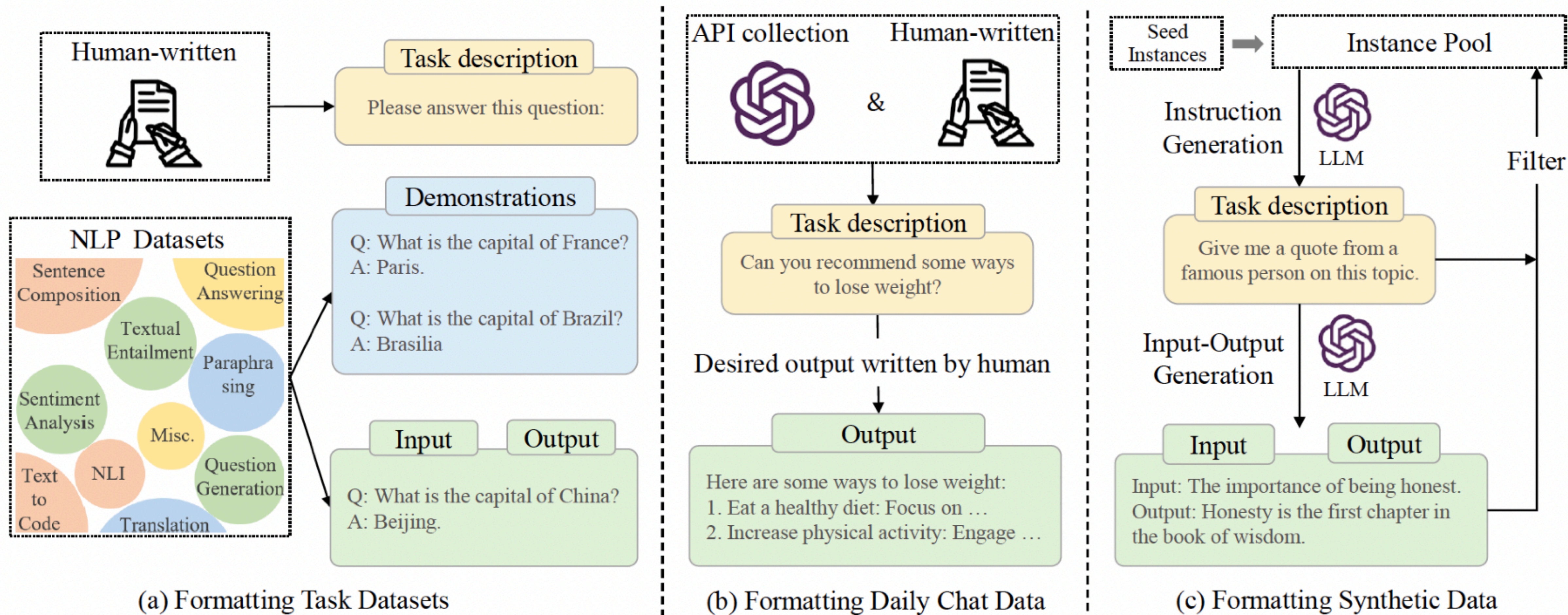


Fig. 8: An illustration of instance formatting and three different methods for constructing the instruction-formatted instances.