Interactive Text-to-SQL Generation via Editable Step-by-Step Explanations

Yuan Tian
Purdue University

Zheng Zhang
University of Notre Dame

Zheng Ning
University of Notre Dame

Toby Jia-Jun Li
University of Notre Dame

Jonathan K. Kummerfeld
University of Sydney

Tianyi Zhang
Purdue University

Introduction

Generating a database query that accurately answers a question (i.e., text-to-SQL) is challenging. Even the latest LLM-based models often make mistakes. Prior work has explored user feedback, but either in constrained ways that are hard to use, or using free text which is nearly as difficult to interpret as the original question.

We introduce **editable explanations** of SQL. Each part of our explanation has a direct mapping to tokens in the query. This means when a user edits the explanation, we know which part of the query needs updating. This results in a system that has great flexibility and accuracy, as shown in both a simulated user experiment and a study with real users.

Code generation systems make mistakes

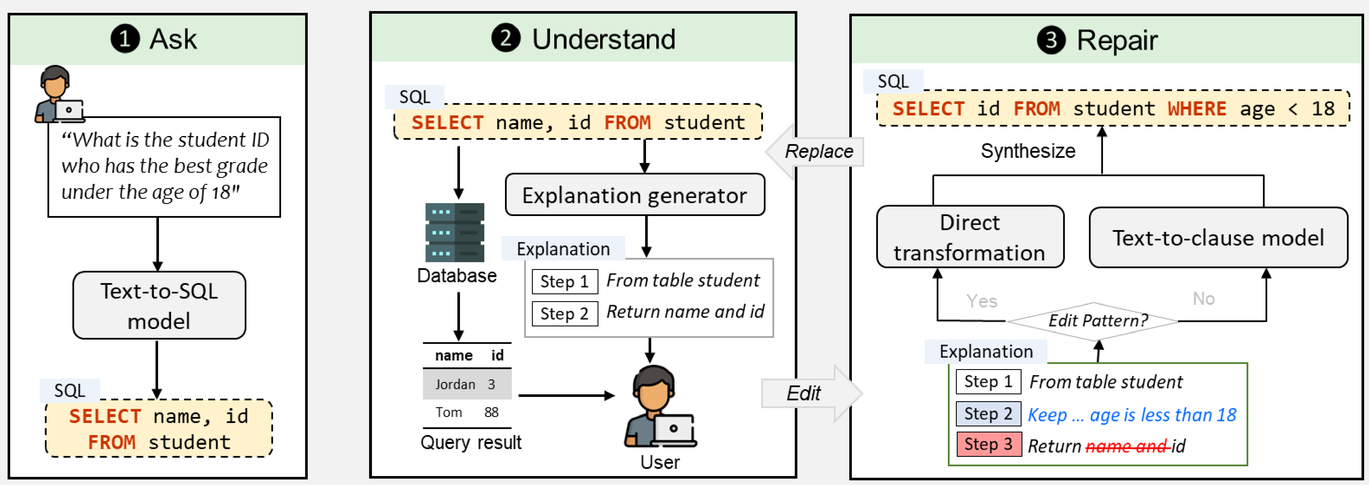
Editable explanations

A flexible and natural way for users to recognize and fix errors

Code, Paper, Data

Contact: tian211@purdue.edu

Method

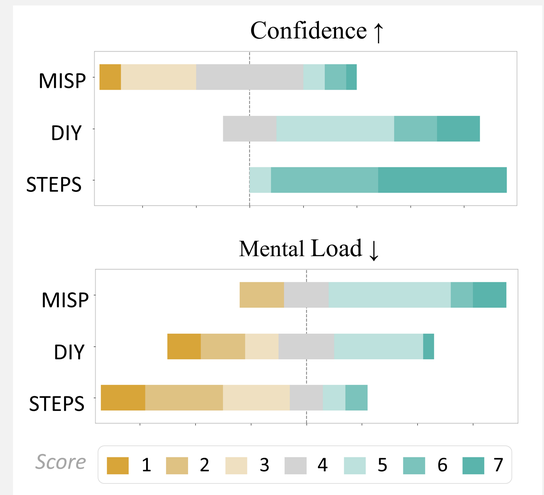


Results

Automated User Simulation

	Exact match accuracy
AI-only Methods	
SmBoP [Rubin and Berant, 2021]	0.745
Graphix-3B + PICARD [Li et al., 2023]	0.740
SHiP + PICARD [Zhao et al., 2022]	0.772
DIN-SQL + GPT-4 [Pourreza et al.,2023]	0.601
EditSQL [Zhang et al., 2019]	0.576
Human-in-the-Loop Methods	
EditSQL + MISP [Yao et al., 2019]	0.644
EditSQL + DIY [Narechania et al., 2021]	0.647
EditSQL + NL-EDIT [Elgohary et al., 2021]	0.666
EditSQL + STEPS (Ours)	**0.979**

User Study



	Complete	Correct	Accuracy	Skipped
MISP	3.0	1.7	0.57	1.4
DIY	5.4	3.5	0.68	0.8
STEPS	**6.7**	**5.7**	**0.86**	**0.3**