

# CS 846 Software Engineering for Big Data and AI

## Talking to Databases: A Literature Review on How Large Language Models are Transforming Text-to-SQL translation

Jun Lim

20870249

The project presents a comprehensive reviews of the integration of Large Language Models (LLMs) with database querying, particularly focusing on the transformation of natural language to SQL queries. The analogy of motivation of LLMs acting as a translator in an international market conveys the potential of this technology to make text-to-SQL interaction more accessible.

### Strengths

- **Clear Problem Statement:** The motivation for the research is well-articulated, identifying key areas where LLMs contribute to database querying. The literature gap is clearly identified, providing a strong justification for the research.
- **Structured Presentation/Proposal:** The presentation is visually engaging, with relevant illustrations that complement the text.
- **Strong Research Value:** The project could help lay a solid foundation for future work as well as providing good understanding in text-to-sql of LLM, and a good starting point for other researchers to expand upon.

### Weaknesses

- **Specifics of Methodology:** The project lacks specific plans on how it is to be carried out (but only on a general/brief level overview), could put more thoughts onto how to conduct a critical + rigorous comprehensive review.
- **Diversity of Sources:** The research might benefit from a more diverse range of sources, including non-academic publications and industry reports, to provide a more comprehensive view of the field.
- **Follow-Up Research:** There's a lack of clear directives for follow-up research. Outlining next steps or potential research projects could provide pathways for continued exploration in this field.

### Suggestions for Potential Improvements

- Include threats to validity, discussing about the limitations of the literature reviews and acknowledging the scope of the research findings.
- Include quantitative analysis, discuss about some quantitative data or analysis (metrics or statistical results), this will help bolster the credibility of the research.
- Add future research on how future research could be expanded from this study of LLM database query.
- Adopt a rigorous methodological framework of conducting literature reviews such as more inclusion/exclusion criteria(s) (instead of just past 8 years), adding broader search string keywords (such as using wildcards for search), this will help make the research to be stronger.
- Include which database/publication/venues the papers queried from.