**УДК 004.93**
**DOI https://doi.org/10.32782/tnv-tech.2025.2.4**

# ВИКОРИСТАННЯ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ ДЛЯ ПЕРЕТВОРЕННЯ ПРИРОДНОЇ МОВИ В SQL ЗАПИТ

***Борисюк В. М.*** *– аспірант кафедри комп'ютерних наук*
*Вінницького національного технічного університету*
*ORCID ID: 0009-0002-1005-7492*

***Козловський А. В.*** *– доцент кафедри комп'ютерних наук*
*Вінницького національного технічного університету*
*ORCID ID: 0000-0001-9697-1511*

*Проблема автоматичного перетворення запитань, сформульованих природною мовою, у структуровані SQL-запити (технологія Text-to-SQL), залишається актуальною впродовж останніх десятиліть через постійне зростання обсягів даних і потребу в доступі до них з боку нефахових користувачів. Основними викликами цього завдання є складність інтерпретації запитів користувачів, необхідність врахування структур баз даних, неоднозначність природної мови, а також забезпечення високої точності синтезу SQL-запитів. Традиційні підходи, що поєднують ручну розробку правил із використанням глибоких нейронних мереж, продемонстрували значний прогрес, однак часто вимагають значних людських ресурсів для створення та підтримки правил, а також демонструють обмежену здатність до узагальнення на нові домени.*

*Подальший розвиток у цій сфері був пов'язаний із появою попередньо натренованих мовних моделей (PLM), які забезпечили суттєве покращення результатів у задачах Text-to-SQL за рахунок глибшого розуміння семантики природної мови. Проте зі зростанням складності схем баз даних і мовних формулювань виникає проблема: моделі, обмежені за розміром, часто генерують некоректні SQL-запити, що зумовлює потребу в складних оптимізаційних стратегіях та знижує масштабованість таких рішень. У цьому контексті великі мовні моделі (LLM) демонструють нові можливості завдяки своїй високій здатності до розуміння природної мови, багатозадачності, контекстної обізнаності та глибокого семантичного аналізу, що покращується зі зростанням розміру моделей.*

*У статті проведено ґрунтовний аналіз ключових технічних викликів, таких як лінгвістична неоднозначність, розуміння та репрезентація схеми бази даних, генерація рідкісних SQL-операцій і проблема узагальнення на різні домени. Описано основні етапи розвитку напряму Text-to-SQL, розглянуто актуальні набори даних, що охоплюють мультидоменні, багатомовні, контекстно-залежні та знання-доповнені задачі, а також наведено характеристики метрик оцінювання якості генерації SQL, серед яких Component Matching, Exact Matching, Execution Accuracy і Valid Efficiency Score. Також узагальнено останні наукові досягнення, зокрема інтеграцію великих мовних моделей, використання стратегій навчання з контекстом (in-context learning), донавчання (fine-tuning), аугментацію даних та багатозадачне налаштування.*

***Ключові слова:*** *Text-to-SQL, великі мовні моделі, генерація коду, машинне навчання, обробка природної мови, SQL.*

***Borysiuk V. M., Kozlovskyi A. V. Usage of large language models for natural language to SQL query conversion***

*The problem of automatically converting questions formulated in natural language into structured SQL queries (Text-to-SQL technology) has remained relevant for decades, driven by the continuous growth of data volumes and the need for access by non-expert users. The main challenges of this task include the complexity of interpreting user queries, the necessity of accounting for database structures, the ambiguity inherent in natural language, and the need to ensure high precision in SQL query synthesis. Traditional approaches combining manual rule engineering with deep neural networks have demonstrated significant progress but often require considerable human effort for rule development and maintenance, and they show limited generalization to new domains.*

*The subsequent development in this area has been marked by the advent of pre-trained language models (PLMs), which have considerably improved performance in Text-to-SQL tasks thanks to their deeper semantic understanding of natural language. However, as the complexity of database schemas and linguistic formulations increases, size-limited models often produce incorrect SQL queries, necessitating complex optimization strategies that reduce the scalability of such solutions. In this context, large language models (LLMs) open up new opportunities due to their remarkable natural language understanding capabilities, multitasking ability, contextual awareness, and deep semantic reasoning, which improve with increasing model size.*

*This study provides a thorough analysis of key technical challenges, such as linguistic ambiguity, database schema understanding and representation, generation of rare SQL operations, and the issue of cross-domain generalization. It outlines the main stages of the Text-to-SQL field development, reviews current datasets covering multi-domain, multilingual, context-dependent, and knowledge-augmented tasks, and describes the characteristics of evaluation metrics, including Component Matching, Exact Matching, Execution Accuracy, and Valid Efficiency Score. Furthermore, the paper summarizes recent scientific achievements, such as the integration of LLMs, use of in-context learning strategies, fine-tuning, data augmentation, and multitask training. It identifies several open problems, including improving generation accuracy, robustness to errors, and adaptation to new domains, and outlines promising future research directions, including the development of hybrid systems, architecture improvements, and enhanced training methods.*

***Key words:*** *Text-to-SQL, LLM, code generation, machine learning, ML, NLP, SQL.*

**Вступ.** Завдання Text-to-SQL передбачає перетворення запитів, сформульованих природною мовою, у SQL-запити, що можуть бути виконані в базах даних. Це одна з фундаментальних проблем обробки природної мови, яка має високу практичну цінність у створенні інтерфейсів природної мови до баз даних (NLIDB). Такі системи надають змогу навіть нефаховим користувачам отримувати доступ до структурованих даних без знання синтаксису SQL.

Завдяки широкому розповсюдженню SQL у професійному середовищі, автоматизація його генерації на основі запитів природною мовою стала предметом активних досліджень. Розвиток великих мовних моделей (LLM) відкрив нові можливості для реалізації Text-to-SQL через парадигми навчання з контекстом (in-context learning) та донавчання (fine-tuning), що дозволило суттєво підвищити точність генерації запитів навіть у складних сценаріях.

Попередні підходи до реалізації Text-to-SQL включали використання шаблонів і правил, що добре працювали у простих випадках, але були обмеженими в умовах зростаючої складності структур баз даних. Подальший розвиток глибоких нейронних мереж і попередньо натренованих мовних моделей (PLM) дозволив системам автоматично навчатися трансформації природньої мови у SQL. Однак саме великі мовні моделі, завдяки своїй масштабованості та здатності до глибокого семантичного аналізу, стали рушієм нової хвилі досліджень у цій галузі.

Запропоновано систематизовану таксономію відповідних методів. Така класифікація забезпечує технічний аналіз реалізацій і дозволяє виявити внесок окремих підходів у підвищення якості генерації SQL-запитів. У межах огляду також здійснено узагальнення наявних підходів з порівняльним аналізом LLM-орієнтованих і традиційних рішень, а також парадигм навчання з контекстом (in-context learning) і донавчання (fine-tuning) з урахуванням перспектив їх подальшого розвитку. Структура та зміст огляду представлені у вигляді дерева таксономії на рисунку 1.

**Огляд процесу перетворення природньої мови в SQL запит**

Процес перетворення приданої мови в SQL запит полягає у перетворенні запитань, сформульованих природною мовою, у SQL-запити, які можуть бути виконані в реляційній базі даних. Такий підхід має потенціал зробити дані більш доступними, надаючи користувачам можливість взаємодіяти з базами даних без

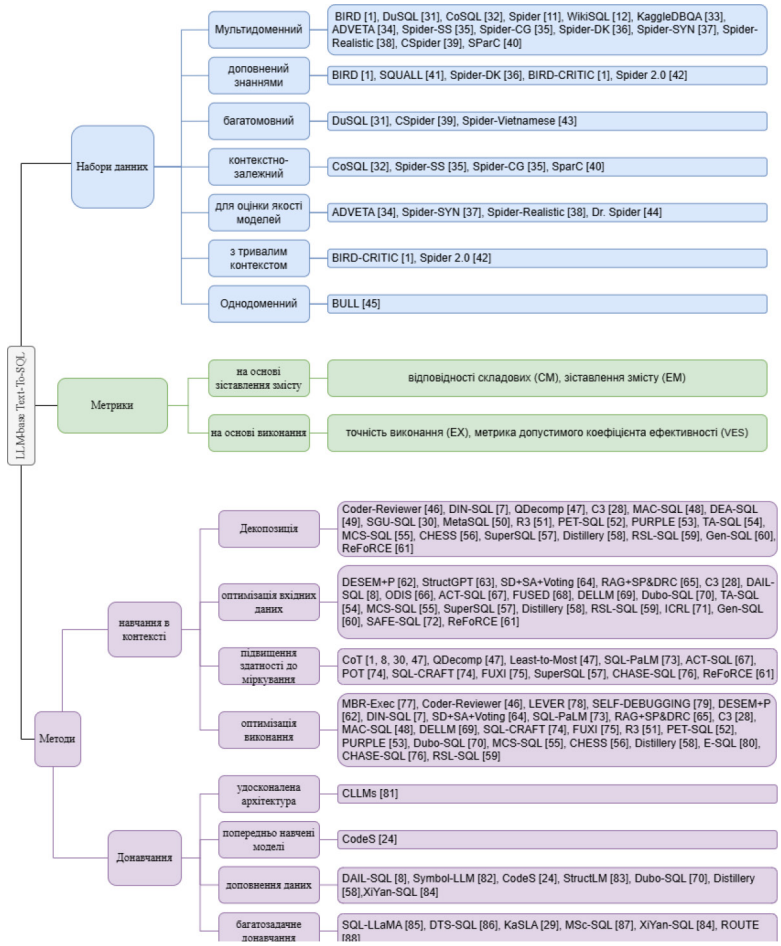


*Рис. 1. Дерево таксономії дослідження у сфері перетворення природної мови в SQL на основі великих мовних моделей (LLM)*

потреби у володінні спеціалізованими знаннями програмування на мові SQL [65]. Це особливо актуально для таких сфер, як бізнес-аналітика, клієнтська підтримка та наукові дослідження, де можливість швидкого доступу до потрібної інформації користувачами без технічної підготовки суттєво підвищує ефективність аналізу даних та прийняття рішень.

**Виклики в реалізації задачі Text-to-SQL.** Основні технічні труднощі можна узагальнити так:

1. Лінгвістична складність і неоднозначність.
2. Розуміння й репрезентація схеми бази даних.
3. Рідкісні та складні SQL-операції.
4. Узагальнення на різні домени.

**Лінгвістична складність і неоднозначність.** Запити сформовані природною мовою часто містять складні мовні конструкції – вкладені підрядні речення, еліпсис тощо, що ускладнює точне відображення їхніх компонентів у відповідних частинах SQL-запитів [4]. Крім того, природна мова є за своєю природою

неоднозначною: одне і те ж запитання може мати кілька коректних інтерпретацій [66]. Розв'язання таких неоднозначностей і точне виявлення інтенції користувача вимагає глибокого розуміння мови, контексту та предметної області [1].

**Розуміння й репрезентація схеми бази даних.** Для генерації коректних SQL-запитів системи Text-to-SQL повинні мати глибоке уявлення про схему бази даних: назви таблиць, стовпців, а також взаємозв'язки між таблицями. Формалізація та кодування такої інформації є складним завданням, оскільки структури БД часто бувають комплексними та суттєво відрізняються в різних доменах [4].

**Рідкісні та складні SQL-операції.** Деякі SQL-запити містять складні конструкції, як-от вкладені підзапити, зовнішні об'єднання (outer join), віконні функції тощо. Такі конструкції є рідкісними у навчальних даних і створюють додаткові труднощі для генерації коректного синтаксису. Розробка моделей, здатних до узагальнення на широкий спектр SQL-операцій, включно з рідкісними та складними, є ключовим викликом.

**Узагальнення на різні домени.** Системи Text-to-SQL часто втрачають ефективність при застосуванні до нових баз даних або предметних областей. Моделі, натреновані на даних однієї тематики, погано справляються із запитаннями з інших галузей через відмінності у словнику, структурі БД та типових шаблонах запитів. Розробка універсальних моделей, здатних до мультидоменного узагальнення з мінімальним обсягом адаптації – важливе й відкрите завдання.

**Огляд існуючих підходів перетворення природної мови в SQL**

Сфера *Text-to-SQL* зазнала значної еволюції – від правил і шаблонів до глибокого навчання, попередньо натренованих мовних моделей (PLM) і, зрештою, великих мовних моделей (LLM).

**Методи на основі правил.** Перші системи *Text-to-SQL* ґрунтувалися на вручну створених правилах та евристиках [67]. Такі системи дозволяли формувати синтаксично правильні SQL-запити з використанням шаблонів і суворих граматичних правил. Однак ці підходи були малоефективними при обробці складних або неоднозначних запитів, не справлялися з вкладеними структурами та не масштабувалися на нові схеми баз даних. Розробка таких правил є трудомісткою та чутливою до помилок, особливо у випадках багато табличних структур.

**Підходи на основі глибокого навчання.** З появою глибоких нейронних мереж з'явилися моделі *sequenc-to-sequence*, зокрема LSTM [68] і трансформери [69], адаптовані для генерації SQL-запитів. Наприклад, модель RYANSQL [70] використовувала проміжні представлення та заповнення слотів для покращення генерації складних запитів і мультидоменного узагальнення. Також почали використовуватись графові нейронні мережі (GNN) для кращого врахування залежностей між компонентами схеми. Однак ці підходи часто генерують некоректний SQL через недостатню структурну обізнаність і складнощі з рідкісними операціями.

**Реалізація на основі PLM.** Попередньо натреновані мовні моделі (PLM), як-от BERT [71] і RoBERTa [72], забезпечили прорив у генерації SQL. Завдяки навчанню на великих корпусах, ці моделі набули глибокого семантичного розуміння. Після донавчання на *Text-to-SQL* набір даних вони демонстрували значно вищу точність. Окремий напрям досліджень зосереджено на інтеграції інформації про схему БД у PLM для поліпшення здатності моделі до розуміння структури даних [73]. Водночас, складні SQL-конструкції (наприклад, агрегації чи зовнішні об'єднання) усе ще залишаються проблемними, як і різке зниження якості при застосуванні до нових доменів, що потребує ресурсномісткого донавчання.

**Реалізація на основі LLM.** Останнім часом великі мовні моделі, такі як GPT [74], привертають велику увагу завдяки здатності генерувати зв'язний текст із високою якістю. Їх потенціал у *Text-to-SQL* реалізується або шляхом інженерії вхідних даних (*prompt engineering*) для керування генерацією SQL [75], або через донавчання відкритих моделей на спеціалізованих наборах даних [8]. Напрям інтеграції LLM у *Text-to-SQL* перебуває на ранніх етапах, однак має значні перспективи. Сучасні дослідження зосереджені на покращенні використання знань LLM, адаптації до предметних галузей [1, 69] і підвищенні ефективності стратегій донавчання [24]. У подальшому очікується поява більш досконалих реалізацій, здатних забезпечити вищу точність і узагальненість генерації SQL-запитів.

**Набори даних**

Як показано в таблицях 1 та 2, набори даних поділено на дві категорії: **«Оригінальні набори даних»** та **«Після анотовані набори даних»**, залежно від того, чи були вони опубліковані разом із початковими базами даних, чи створені на основі вже наявних ресурсів із внесенням спеціальних змін.

Таблиця 1

**Перелік оригінальних наборів даних**

| Назва набору даних | Рік | Кількість рядків | Мова | Категорія |
|---|---|---|---|---|
| BIRD-CRITIC [1] | 2025 | 600 | EN | доповнений знаннями, з тривалим контекстом |
| Spider 2.0 [14] | 2024 | 547 | EN | доповнений знаннями, з тривалим контекстом |
| BULL [45] | 2024 | 5 тис. | EN | Однодоменний |
| BIRD [1] | 2023 | 549 тис. | EN | Мультидоменний, доповнений знаннями |
| KaggleDBQA [33] | 2021 | 280 тис. | EN | Мультидоменний |
| DuSQL [31] | 2020 | 23 тис. | EN | Мультидоменний, багатомовний |
| SQUALL [41] | 2020 | [4],468 | EN | доповнений знаннями |
| CoSQL [32] | 2019 | 15,598 | EN | Мультидоменний, контекстно-залежний |
| Spider [4] | 2018 | 10,181 | EN | Мультидоменний |
| WikiSQL [12] | 2017 | 80,654 | EN | Мультидоменний |

Таблиця 2

**Перелік після анотованих наборів даних**

| Назва набору даних | Рік | Мова | Категорія |
|---|---|---|---|
| Dr. Spider [44] | 2023 | EN | для оцінки якості моделей |
| ADVETA [7] | 2022 | EN | для оцінки якості моделей |
| Spider-SS&CG [35] | 2022 | EN | контекстно-залежний |
| Spider-DK [9] | 2021 | EN | доповнений знаннями |
| Spider-SYN [10] | 2021 | EN | доповнений знаннями |
| Spider-Vietnamese [15] | 2020 | VI | багатомовний |
| Spider-Realistic [4] | 2020 | EN | для оцінки якості моделей |
| CSpider [12] | 2019 | ZH | багатомовний |
| SParC [40] | 2019 | EN | контекстно-залежний |

Для оригінальних наборів надано детальний аналіз, що включає кількість прикладів, кількість баз даних, середню кількість таблиць і рядків на базу даних. Для після анотованих наборів вказано джерело первинного набору дану, а також описано специфіку адаптацій, які були застосовані.

Як показано в таблицях 1 та 2, набори даних поділено на дві категорії: «**оригінальні набори даних**» та «**після анотовані набори даних**» – залежно від того, чи були вони опубліковані разом із початковими базами даних, чи створені шляхом адаптації існуючих наборів даних і баз даних із застосуванням спеціальних налаштувань.

**Мультидомені набори даних.** До цієї категорії належать набори, у яких бази даних охоплюють різні предметні області. Оскільки реальні застосування *Text-to-SQL* часто вимагають роботи з базами з кількох доменів, більшість як оригінальних [1], так і після анотованих наборів даних створено саме в мультидоменомі форматі для забезпечення кращої адаптивності до різнорідних задач.

**Набори даних доповнений знаннями.** Останніми роками значно зросла зацікавленість у включенні доменної інформації до *Text-to-SQL* задач. Наприклад, набір BIRD [1] анотовано фахівцями з баз даних, які додали до кожного прикладу зовнішні знання: числову логіку, доменні знання, синонімічну відповідність та ілюстративні значення. Аналогічно, Spider-DK [9] розширює набір даних Spider [4] п'ятьма типами доменних знань: вибір пропущених стовпців SELECT, прості логічні висновки, синонімічні заміни значень, генерація умови з одного слова та конфлікти між доменами. Дослідження показують, що така ручна анотація істотно підвищує точність генерації SQL у випадках, де необхідні зовнішні знання. Крім того, SQUALL [41] надає ручні відповідності між словами в запитанні та сутностями в SQL, забезпечуючи більш точне кероване навчання.

**Контекстно-залежні набори даних.** SParC [13] і CoSQL [3] вивчають генерацію SQL у діалоговому контексті, створюючи системи запитів до БД у форматі бесіди. На відміну від традиційних наборів, де кожен приклад містить одне запитання та відповідний SQL, у SParC приклади розбиваються на послідовності підзапитань, які імітують взаємодію, включно з пов'язаними (що допомагають у генерації SQL) та непов'язаними підзапитаннями (для збільшення різноманіття). Натомість CoSQL реалізує повноцінні діалоги природною мовою, моделюючи реальні сценарії з високою складністю. Також Spider SS&CG [8] ділить запитання з набір даниху Spider [4] на серію підзапитань із відповідними під-SQL-запитами, демонструючи, що навчання на таких прикладах покращує здатність моделі до узагальнення на нових даних.

**Набори даних для оцінки якості моделей.** Оцінювання точності Text-to-SQL моделей у випадку «зашумлених» або змінених схем БД є критичним для перевірки їхньої стійкості. Spider-Realistic [4] видаляє зі запитань слова, що явно посилаються на схему, а Spider-SYN [10] замінює їх вручну підібраними синонімами. ADVETA [7] вводить перешкоди в таблиці, замінюючи назви стовпців на хибні або додаючи зайві з високою семантичною подібністю. Dr. Spider [44] використовує можливості великих мовних моделей для створення 17 типів перешкод у БД, SQL і природномовних запитах. Такі змінення значно знижують точність, особливо в системах із низькою стійкістю до неправильних відповідностей між словами та сутностями БД.

**Багатомовні набори даних.** Оскільки ключові слова SQL, назви таблиць і стовпців зазвичай подаються англійською, виникають труднощі при використанні в інших мовах. CSpider [12] перекладає набір даних Spider на китайську,

виявляючи нові виклики у сегментації слів та крослінгвальній відповідності між китайськими запитаннями й англійськими схемами БД. DuSQL [2] – практичний набір даних із китайськими запитаннями, а також схемами БД, поданими одночасно англійською та китайською мовами. Spider-Vietnamese [15] перекладає Spider на в'єтнамську, демонструючи стратегії покращення *Text-to-SQL* у цьому мовному контексті.

**Набори даних з тривалим контекстом.** У реальних базах даних часто виникає потреба у складних SQL-запитах із великою довжиною (100+ токенів) та багаторівневими операціями. Spider 2.0 [14] відображає цю складність у задачах корпоративного рівня, які вимагають багаторівневих логічних міркувань і підтримки різних SQL-діалектів. Навіть сучасні LLM успішно виконують лише 17,1 % таких задач, що підкреслює виклики, пов'язані з довгим контекстом. У свою чергу, BIRD-CRITIC2 [1] – діагностичний набір із 600 завдань для перевірки здатності LLM вирішувати реальні задачі в різних SQL-діалектах. Обидва набір данихи значно розширюють межі досліджень у *Text-to-SQL*, акцентуючи увагу на складності довгих контекстів.

**Однодомені набори даних.** Незважаючи на загальний прогрес у сфері Text-to-SQL, усе ще бракує наборів, орієнтованих на специфічні галузі. BULL [17] створено спеціально для фінансової сфери. Набір даних охоплює БД, пов'язані з фондами, акціями та макроекономічними показниками, й слугує платформою для перевірки та вдосконалення *Text-to-SQL* у фінансових застосуваннях, з урахуванням характерних для цієї галузі викликів.

**Метрики оцінювання**

Для задачі *Text-to-SQL* використовуються чотири поширені метрики, які поділяються на дві групи: метрики зіставлення вмісту SQL-запитів (*Component Matching ma Exact Matching*) та метрики, засновані на результатах виконання запитів (*Execution Accuracy ma Valid Efficiency Score*).

**Метрики, засновані на зіставленні вмісту SQL.** Ці метрики орієнтовані на порівняння згенерованого SQL-запиту з еталонним (ground truth) на основі синтаксичних і структурних подібностей.

**Component Matching (CM)** [4] – оцінює якість роботи *Text-to-SQL* системи шляхом підрахунку точного збігу між компонентами згенерованого та еталонного SQL-запитів: SELECT, WHERE, GROUP BY, ORDER BY та KEYWORDS. Для кожного компонента формується множина під-компонентів, які порівнюються на точний збіг. Облік порядку компонентів при цьому не є обов'язковим. Результат оцінюється за метрикою F1.

**Exact Matching (EM)** [4] – визначає відсоток прикладів, для яких згенерований SQL-запит повністю збігається з еталонним. Запит вважається правильним лише у разі повного збігу всіх компонентів, визначених у метриці CM.

**Метрики, засновані на результатах виконання.** Ці метрики оцінюють правильність згенерованого SQL-запиту шляхом виконання запиту в цільовій базі даних і порівняння результатів з очікуваними.

**Execution Accuracy** (EX) [4] – вимірює коректність SQL-запиту, виконуючи його у відповідній базі даних і порівнюючи результати виконання із результатами еталонного запиту.

**Valid Efficiency Score (VES)** [1] – метрика, яка використовується для оцінювання ефективності *коректних* SQL-запитів. Коректним вважається такий згенерований SQL-запит, результат виконання якого повністю збігається з еталонним результатом. Зокрема, VES оцінює одночасно як **ефективність**, так і **точність**

згенерованих SQL-запитів. Для текстового набору даних із N прикладами значення VES обчислюється за формулою:

$$VES = \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}(V_n, \hat{V}_n) \cdot R(Y_n, \hat{Y}_n), \quad (1)$$

де $\hat{Y}_n$ та $\hat{V}_n$ це, відповідно, згенерований SQL-запит і результати його виконання, а $Y_n$ та $V_n$ еталонний SQL-запит і відповідні результати його виконання, функція $\mathbb{I}(V_n, \hat{V}_n)$ – це індикаторна функція як дорівнює:

$$\mathbb{I}(V_n, \hat{V}_n) = \begin{cases} 1, & \text{якщо } \hat{V}_n = V_n \\ 0, & \text{інакше} \end{cases}, \quad (2)$$

Значення $R(Y_n, \hat{Y}_n) = \sqrt{\frac{E(Y_n)}{E(\hat{Y}_n)}}$ – означає відносну ефективність виконання згенерованого SQL-запиту порівняно з еталонним запитом, де E час виконання відповідного SQL-запиту в базі даних.

**Методи перетворення природної мови з SQL запит**

Реалізація сучасних методів *Text-to-SQL* на основі великих мовних моделей (LLM) переважно ґрунтується на двох парадигмах: навчанні з контекстом (*in-context learning*, ICL) [76] та донавчанні (*fine-tuning*, FT) [13]. Такий підхід став можливим завдяки широкому поширенню як потужних комерційних моделей, так і якісних відкритих архітектури LLM.

**Навчання з контекстом (*In-Context Learning*).** Навчання з контекстом (ICL, також відоме як інженерія підказок) відіграє вирішальну роль у продуктивності великих мовних моделей (LLM) у різноманітних завданнях, включно із *Text-to-SQL*. Результат генерації SQL-запиту $\hat{Y}$ у парадигмі ICL можна подати у вигляді:

$$\hat{Y} = \pi(I, Q, S | \theta), \quad (3)$$

де $I$ – інструкція для задачі Text-to-SQL; $Q$ – запит користувача; $S = C, T, K$ – схема бази даних із наборами стовпців $C = \{c_1, c_2, \ldots\}$, таблиць $T = \{t_1, t_2, \ldots\}$ та додатковими значеннями $K$ – зв'язками між таблицями; $\pi(\cdot \mid \theta)$ – велика мовна модель з параметрами $\theta$.

Базовий вхід для генерації запиту у режимі zero-shot prompting формується шляхом конкатенації:

$$P_0 = I + S + Q. \quad (4)$$

У режимі few-shot prompting вхід розширюється прикладами:

$$P_k = \{E_1, E_2, \ldots, E_k\} + P_0, \quad (5)$$

де $k$ – кількість прикладів, а кожен приклад $E_i = (S_i, Q_i, Y_i)$ складається зі схеми, запитання до відповідного SQL-запиту.

У сучасних роботах ICL активно використовуються як **zero-shot**, так і **few-shot** стратегії для покращення генерації SQL. Основні напрямки досліджень включають:

– **zero-shot prompting:** аналіз впливу різних стилів підказок на якість генерації, дослідження важливості структури бази даних, наявності зовнішніх знань і ефективності ChatGPT та відкритих моделей у задачах *Text-to-SQL* [38];

– **few-shot prompting:** вивчення кількості та вибору прикладів для few-shot демонстрацій, оптимізація їх відбору за семантичною подібністю, складністю завдань або доменною відповідністю [1].

У реальних реалізаціях моделі часто використовуються у режимі заморожених параметрів θ, що дозволяє ефективно застосовувати великі попередньо натреновані LLM без додаткового донавчання.

**Донавчання (Fine-tuning).** Найпоширенішим підходом у парадигмі донавчання є контрольоване донавчання (*Supervised Fine-tuning*, SFT) [77], яке застосовується для адаптації відкритих моделей, таких як LLaMA [22] чи Qwen [78], до конкретних доменів. Нехай $P$ – вхідна підказка, сформована за рівняннями (4) і (5), а $Y = [y_1, y_2, \ldots, y_r]$ – правильний SQL-запит. Ймовірність $P_r$ для $LLL_\pi$ згенерувати SQL запит $Y$ на основі вхідних даних $P$ можна подати у вигляді наступного рівняння:

$$P_{r_\pi}(Y|P) = \prod_{t=1}^{T} P_{r_\pi}(y_t \mid P, y < t). \tag{6}$$

На відміну від навчання з контекстом (ICL), у процесі SFT параметри моделі π оновлюються шляхом мінімізації функції втрат:

$$L_{SFT} = -\sum_{(P,Y)\in D}\sum_{t=1}^{T} logP_{r_\pi}(y_t \mid P, y < t), \tag{7}$$

де $D$ – тренувальний набір даних.

Контрольоване донавчання є основою для багатьох підходів у *Text-to-SQL*, особливо через обмежену здатність відкритих моделей у порівнянні з комерційними LLM. Основні напрями вдосконалення в межах SFT-підходів включають:

– **покращення архітектури:** модифікації стандартних трансформерів GPT спрямовані на ефективну обробку складної синтаксичної структури SQL, наприклад, через оптимізовану увагу та схему-орієнтоване кодування [79]. CLLM [56] пропонує спеціалізовані методи декодування для пришвидшення генерації SQL;

– **попереднє навчання:** для моделей, орієнтованих на код (наприклад, CodeLLaMA [22]), донавчання на SQL та структурованих даних допомагає суттєво покращити здатність до генерації SQL [57];

– **аугментація даних:** покращення якості даних для тренування суттєво впливає на результати. Методи, як DAIL-SQL [40], Symbol-LLM [58] і CodeS [57], використовують додаткові дані, вибір прикладів за семантичною подібністю та двосторонню генерацію для підвищення точності моделей;

– **багатозадачне налаштування:** розподіл складного завдання на підзадачі, як у SQL-LLaMA [22] або DTS-SQL [86], дозволяє ефективніше адаптувати моделі до складних аспектів *Text-to-SQL*, як-от генерація знань, прив'язка схем і вдосконалення за допомогою зворотного зв'язку.

**Висновки.** У межах цього дослідження здійснено всебічний аналіз розвитку технологій *Text-to-SQL* із застосуванням великих мовних моделей (LLM).

Запропоновано структуровану класифікацію існуючих підходів за двома основними парадигмами: навчання з контекстом (*in-context learning*, ICL) та контрольованого донавчання (*fine-tuning*, FT). Усередині кожної парадигми виділено ключові напрями, зокрема декомпозицію задач, оптимізацію підказок, підсилення міркування, вдосконалення за результатами виконання та багатозадачне налаштування.

Розглянуто основні труднощі перетворення запитань природною мовою у структуровані SQL-запити: лінгвістичну неоднозначність, складність розуміння

та представлення схем баз даних, генерацію рідкісних і складних SQL-конструкцій, а також проблеми мультидоменного узагальнення моделей. На основі аналізу виявлено ключові обмеження як традиційних підходів, так і сучасних моделей.

Проведено детальне зіставлення класичних підходів на основі правил і нейронних моделей із сучасними LLM-орієнтованими підходами. Показано переваги гібридних систем, які поєднують точність і контроль традиційних правил із гнучкістю та широтою узагальнення LLM.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ:

1. J. Li, B. Hui, G. QU, J. Yang, B. Li, B. Li, B. Wang, B. Qin, R. Geng, N. Huo, X. Zhou, C. Ma, G. Li, K. Chang, F. Huang, R. Cheng, and Y. Li. "Can LLM already serve as a database interface? a BIg bench for large-scale database grounded Text-to-SQLs," in Advances in Neural Information Processing Systems (NeurIPS), 2023.

2. Lijie Wang, Ao Zhang, Kun Wu, Ke Sun, Zhenghua Li, Hua Wu, Min Zhang, and Haifeng Wang. 2020. DuSQL: A Large-Scale and Pragmatic Chinese Text-to-SQL Dataset. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6923–6935, Online. Association for Computational Linguistics.

3. T. Yu, R. Zhang, H. Er, S. Li, E. Xue, B. Pang, X. V. Lin, Y. C. Tan, T. Shi, Z. Li, Y. Jiang, M. Yasunaga, S. Shim, T. Chen, A. Fabbri, Z. Li, L. Chen, Y. Zhang, S. Dixit, V. Zhang, C. Xiong, R. Socher, W. Lasecki, and D. Radev, "CoSQL: A conversational Text-to-SQL challenge towards cross-domain natural language interfaces to databases," in Empirical Methods in Natural Language Processing and International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019

4. T. Yu, R. Zhang, K. Yang, M. Yasunaga, D. Wang, Z. Li, J. Ma, I. Li, Q. Yao, S. Roman, Z. Zhang, and D. Radev, "Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing and Text-to-SQL task," in Empirical Methods in Natural Language Processing (EMNLP), 2018.

5. V. Zhong, C. Xiong, and R. Socher, "Seq2sql: Generating structured queries from natural language using reinforcement learning," arXiv preprint arXiv:1709.00103, 2017.

6. Y. Gan, X. Chen, Q. Huang, and M. Purver, "Measuring and improving compositional generalization in text-toSQL via component alignment," in Findings of North American Chapter of the Association for Computational Linguistics (NAACL), 2022.

7. X. Pi, B. Wang, Y. Gao, J. Guo, Z. Li, and J.-G. Lou, "Towards robustness of Text-to-SQL models against natural and realistic adversarial table perturbation," in Association for Computational Linguistics (ACL), 2022.

8. Y. Gan, X. Chen, Q. Huang, and M. Purver, "Measuring and improving compositional generalization in text-toSQL via component alignment," in Findings of North American Chapter of the Association for Computational Linguistics (NAACL), 2022.

9. Y. Gan, X. Chen, and M. Purver, "Exploring underexplored limitations of cross-domain Text-to-SQL generalization," in Empirical Methods in Natural Language Processing (EMNLP), 2021.

10. Y. Gan, X. Chen, Q. Huang, M. Purver, J. R. Woodward, J. Xie, and P. Huang, "Towards robustness of Text-to-SQL models against synonym substitution," in Association for Computational Linguistics and International Joint Conference on Natural Language Processing (ACLIJCNLP), 2021.

11. X. Deng, A. H. Awadallah, C. Meek, O. Polozov, H. Sun, and M. Richardson, "Structure-grounded pretraining for Text-to-SQL," in North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 2021.

12. Q. Min, Y. Shi, and Y. Zhang, "A pilot study for Chinese SQL semantic parsing," in Empirical Methods in Natural Language Processing and International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019.

13. T. Yu, R. Zhang, M. Yasunaga, Y. C. Tan, X. V. Lin, S. Li, H. Er, I. Li, B. Pang, T. Chen, E. Ji, S. Dixit, D. Proctor, S. Shim, J. Kraft, V. Zhang, C. Xiong, R. Socher, and D. Radev, "SParC: Cross-domain semantic parsing in context," in Association for Computational Linguistics (ACL), 2019.

14. F. Lei, J. Chen, Y. Ye, R. Cao, D. Shin, H. SU, Z. SUO, H. Gao, W. Hu, P. Yin, V. Zhong, C. Xiong, R. Sun, Q. Liu, S. Wang, and T. Yu, "Spider 2.0: Evaluating language models on real-world enterprise Text-to-SQL workflows," in International Conference on Learning Representations (ICLR), 2025.

15. A. Tuan Nguyen, M. H. Dao, and D. Q. Nguyen, "A pilot study of Text-to-SQL semantic parsing for Vietnamese," in Findings of Empirical Methods in Natural Language Processing (EMNLP), 2020.

16. S. Chang, J. Wang, M. Dong, L. Pan, H. Zhu, A. H. Li, W. Lan, S. Zhang, J. Jiang, J. Lilien et al., "Dr. spider: A diagnostic evaluation benchmark towards Text-to-SQL robustness," in International Conference on Learning Representations (ICLR), 2023.

17. C. Zhang, Y. Mao, Y. Fan, Y. Mi, Y. Gao, L. Chen, D. Lou, and J. Lin, "Finsql: model-agnostic llmsbased Text-to-SQL framework for financial analysis," in Conference on Management of Data (SIGMOD), 2024.

18. T. Zhang, T. Yu, T. B. Hashimoto, M. Lewis, W. tau Yih, D. Fried, and S. I. Wang, "Coder reviewer reranking for code generation," in International Conference on Machine Learning (ICML), 2023.

19. M. Pourreza and D. Rafiei, "DIN-SQL: Decomposed in-context learning of Text-to-SQL with self-correction," in Advances in Neural Information Processing Systems (NeurIPS), 2023.

20. C.-Y. Tai, Z. Chen, T. Zhang, X. Deng, and H. Sun, "Exploring chain of thought style prompting for textto-SQL," in Empirical Methods in Natural Language Processing (EMNLP), 2023.

21. X. Dong, C. Zhang, Y. Ge, Y. Mao, Y. Gao, J. Lin, D. Lou et al., "C3: Zero-shot Text-to-SQL with chatgpt," arXiv preprint arXiv:2307.07306, 2023.

22. B. Wang, C. Ren, J. Yang, X. Liang, J. Bai, L. Chai, Z. Yan, Q.-W. Zhang, D. Yin, X. Sun, and Z. Li, "Macsql: A multi-agent collaborative framework for textto-sql," in International Conference on Computational Linguistics (COLING), 2024.

23. Y. Xie, X. Jin, T. Xie, M. Lin, L. Chen, C. Yu, L. Cheng, C. Zhuo, B. Hu, and Z. Li, "Decomposition for enhancing attention: Improving llm-based Text-to-SQL through workflow paradigm," in Findings of Association for Computational Linguistics (ACL), 2024.

24. Q. Zhang, J. Dong, H. Chen, W. Li, F. Huang, and X. Huang, "Structure guided large language model for sql generation," arXiv preprint arXiv:2402.13284, 2024.

25. Y. Fan, Z. He, T. Ren, C. Huang, Y. Jing, K. Zhang, and X. S. Wang, "Metasql: A generate-then-rank framework for natural language to sql translation," in International Conference on Data Engineering (ICDE), 2024.

26. Z. Li, X. Wang, J. Zhao, S. Yang, G. Du, X. Hu, B. Zhang, Y. Ye, Z. Li, R. Zhao, and H. Mao, "Pet-sql: A prompt-enhanced two-stage Text-to-SQL framework with cross-consistency," arXiv preprint arXiv:2403.09732, 2024.

27. T. Ren, Y. Fan, Z. He, R. Huang, J. Dai, C. Huang, Y. Jing, K. Zhang, Y. Yang, and X. S. Wang, "Purple: Making a large language model a better sql writer," in International Conference on Data Engineering (ICDE), 2024

28. G. Qu, J. Li, B. Li, B. Qin, N. Huo, C. Ma, and R. Cheng, "Before generation, align it! a novel and effective strategy for mitigating hallucinations in Text-to-SQL generation," in Findings of Association for Computational Linguistics (ACL), 2024.

29. D. Lee, C. Park, J. Kim, and H. Park, "MCS-SQL: Leveraging multiple prompts and multiple-choice selection for Text-to-SQL generation," in International Conference on Computational Linguistics (COLING), 2025.

30. S. Talaei, M. Pourreza, Y.-C. Chang, A. Mirhoseini, and A. Saberi, "Chess: Contextual harnessing for efficient sql synthesis," arXiv preprint arXiv:2405.16755, 2024.

31. B. Li, Y. Luo, C. Chai, G. Li, and N. Tang, "The dawn of natural language to sql: Are we fully ready?" in International Conference on Very Large Data Bases (VLDB), 2024.

32. K. Maamari, F. Abubaker, D. Jaroslawicz, and A. Mhedhbi, "The death of schema linking? Text-to-SQL in the age of well-reasoned language models," in NeurIPS 2024 Third Table Representation Learning Workshop (NeurIPS), 2024.

33. Z. Cao, Y. Zheng, Z. Fan, X. Zhang, and W. Chen, "Rslsql: Robust schema linking in Text-to-SQL generation," arXiv preprint arXiv:24[4].00073, 2024.

34. J. Shi, B. Xu, J. Liang, Y. Xiao, J. Chen, C. Xie, P. Wang, and W. Wang, "Gen-SQL: Efficient Text-to-SQL by bridging natural language question and database schema with pseudo-schema," in International Conference on Computational Linguistics (COLING), 2025.

35. M. Deng, A. Ramachandran, C. Xu, L. Hu, Z. Yao, A. Datta, and H. Zhang, "Reforce: A Text-to-SQL agent with self-refinement, format restriction, and column exploration," arXiv preprint arXiv:2502.00675, 2025.

36. C. Guo, Z. Tian, J. Tang, P. Wang, Z. Wen, K. Yang, and T. Wang, "Prompting gpt-3.5 for Text-to-SQL with de-semanticization and skeleton retrieval," in Pacific Rim International Conference on Artificial Intelligence (PRICAI), 2024.

37. J. Jiang, K. Zhou, Z. Dong, K. Ye, X. Zhao, and J.- R. Wen, "StructGPT: A general framework for large language model to reason over structured data," in Empirical Methods in Natural Language Processing (EMNLP), 2023.

38. L. Nan, Y. Zhao, W. Zou, N. Ri, J. Tae, E. Zhang, A. Cohan, and D. Radev, "Enhancing Text-to-SQL capabilities of large language models: A study on prompt design strategies," in Findings of Empirical Methods in Natural Language Processing (EMNLP), 2023.

39. C. Guo, Z. Tian, J. Tang, S. Li, Z. Wen, K. Wang, and T. Wang, "Retrieval-augmented gpt-3.5-based Text-to-SQL framework with sample-aware prompting and dynamic revision chain," in International Conference on Neural Information Processing (ICONIP), 2024.

40. D. Gao, H. Wang, Y. Li, X. Sun, Y. Qian, B. Ding, and J. Zhou, "Text-to-SQL empowered by large language models: A benchmark evaluation," in International Conference on Very Large Data Bases (VLDB), 2024.

41. S. Chang and E. Fosler-Lussier, "Selective demonstrations for cross-domain Text-to-SQL," in Findings of Empirical Methods in Natural Language Processing (EMNLP), 2023.

42. H. Zhang, R. Cao, L. Chen, H. Xu, and K. Yu, "ACT-SQL: In-context learning for Text-to-SQL with automatically-generated chain-of-thought," in Findings of Empirical Methods in Natural Language Processing (EMNLP), 2023.

43. D. Wang, L. Dou, X. Zhang, Q. Zhu, and W. Che, "Improving demonstration diversity by human-free fusing for Text-to-SQL," in Findings of Empirical Methods in Natural Language Processing (EMNLP), 2024.

44. Z. Hong, Z. Yuan, H. Chen, Q. Zhang, F. Huang, and X. Huang, "Knowledge-to-sql: Enhancing sql generation with data expert llm," in Findings of Association for Computational Linguistics (ACL), 2024.

45. D. G. Thorpe, A. J. Duberstein, and I. A. Kinsey, "Dubosql: Diverse retrieval-augmented generation and fine tuning for Text-to-SQL," arXiv preprint arXiv:2404.12560, 2024.

46. R. Toteja, A. Sarkar, and P. M. Comar, “In-context reinforcement learning with retrieval-augmented generation for Text-to-SQL,” in International Conference on Computational Linguistics (COLING), 2025.

47. J. Lee, I. Baek, B. Kim, and H. Lee, “Safe-sql: Self-augmented in-context learning with fine-grained example selection for Text-to-SQL,” arXiv preprint arXiv:2502.[4]438, 2025.

48. R. Sun, S. O. Arik, H. Nakhost, H. Dai, R. Sinha, P. Yin, and T. Pfister, “Sql-palm: Improved large language model adaptation for Text-to-SQL,” Transactions on Machine Learning Research (TMLR), 2023.

49. H. Xia, F. Jiang, N. Deng, C. Wang, G. Zhao, R. Mihalcea, and Y. Zhang, “Sql-craft: Text-to-SQL through interactive refinement and enhanced reasoning,” arXiv preprint arXiv:2402.14851, 2024.

50. Y. Gu, Y. Shu, H. Yu, X. Liu, Y. Dong, J. Tang, J. Srinivasa, H. Latapie, and Y. Su, “Middleware for llms: Tools are instrumental for language agents in 17 complex environments,” in Empirical Methods in Natural Language Processing (EMNLP), 2024.

51. M. Pourreza, H. Li, R. Sun, Y. Chung, S. Talaei, G. T. Kakkar, Y. Gan, A. Saberi, F. Ozcan, and S. O. Arik, “CHASE-SQL: Multi-path reasoning and preference optimized candidate selection in Text-to-SQL,” in International Conference on Learning Representations (ICLR), 2025.

52. F. Shi, D. Fried, M. Ghazvininejad, L. Zettlemoyer, and S. I. Wang, “Natural language to code translation with execution,” in Empirical Methods in Natural Language Processing (EMNLP), 2022

53. A. Ni, S. Iyer, D. Radev, V. Stoyanov, W.-t. Yih, S. I. Wang, and X. V. Lin, “Lever: Learning to verify language-to-code generation with execution,” in International Conference on Machine Learning (ICML), 2023.

54. X. Chen, M. Lin, N. Scharli, and D. Zhou, “Teaching large language models to self-debug,” in International Conference on Learning Representations (ICLR), 2024.

55. H. A. Caferoglu and O. Ulusoy, “E-sql: Direct schema linking via question enrichment in Text-to-SQL,” arXiv preprint arXiv:2409.16751, 2024.

56. S. Kou, L. Hu, Z. He, Z. Deng, and H. Zhang, “Cllms: Consistency large language models,” in International Conference on Machine Learning (ICML), 2024.

57. H. Li, J. Zhang, H. Liu, J. Fan, X. Zhang, J. Zhu, R. Wei, H. Pan, C. Li, and H. Chen, “Codes: Towards 15 building open-source language models for Text-to-SQL,” in Conference on Management of Data (SIGMOD), 2024.

58. F. Xu, Z. Wu, Q. Sun, S. Ren, F. Yuan, S. Yuan, Q. Lin, Y. Qiao, and J. Liu, “Symbol-llm: Towards foundational symbol-centric interface for large language models,” in Association for Computational Linguistics (ACL), 2024.

59. A. Zhuang, G. Zhang, T. Zheng, X. Du, J. Wang, W. Ren, S. W. Huang, J. Fu, X. Yue, and W. Chen, “Structlm: Towards building generalist models for structured knowledge grounding,” in Conference on Language Modeling (COLM), 2024.

60. Y. Gao, Y. Liu, X. Li, X. Shi, Y. Zhu, Y. Wang, S. Li, W. Li, Y. Hong, Z. Luo et al., “Xiyan-sql: A multigenerator ensemble framework for Text-to-SQL,” arXiv preprint arXiv:24[4].08599, 2024.

61. M. Pourreza and D. Rafiei, “Dts-sql: Decomposed textto-sql with small large language models,” in Findings of Empirical Methods in Natural Language Processing (EMNLP), 2024.

62. Z. Yuan, H. Chen, Z. Hong, Q. Zhang, F. Huang, and X. Huang, “Knapsack optimization-based schema linking for llm-based Text-to-SQL generation,” arXiv preprint arXiv:2502.129[4], 2025.

63. S. K. Gorti, I. Gofman, Z. Liu, J. Wu, N. Vouitsis, G. Yu, J. C. Cresswell, and R. Hosseinzadeh, “MScSQL: Multi-sample critiquing small language models for Text-to-SQL translation,” in North American Chapter of the Association for Computational Linguistics (NAACL), 2024.

64. Y. Qin, C. Chen, Z. Fu, Z. Chen, D. Peng, P. Hu, and J. Ye, "ROUTE: Robust multitask tuning and collaboration for Text-to-SQL," in International Conference on Learning Representations (ICLR), 2025.

65. P. Ma and S. Wang, "Mt-teql: evaluating and augmenting neural nlidb on real-world linguistic and schema variations," in International Conference on Very Large Data Bases (VLDB), 2021.

66. P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "SQuAD: 100,000+ questions for machine comprehension of text," in Empirical Methods in Natural Language Processing (EMNLP), 2016.

67. F. Li and H. V. Jagadish, "Constructing an interactive natural language interface for relational databases," in International Conference on Very Large Data Bases (VLDB), 2014.

68. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, 1997.

69. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems (NeurIPS), 2017.

70. D. Choi, M. C. Shin, E. Kim, and D. R. Shin, "Ryansql: Recursively applying sketch-based slot fillings for complex Text-to-SQL in cross-domain databases," Computational Linguistics, 2021.

71. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 2019.

72. Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," arXiv preprint arXiv:1907.[4]692, 2019.

73. H. Li, J. Zhang, C. Li, and H. Chen, "Resdsql: Decoupling schema linking and skeleton parsing for Text-to-SQL," in Conference on Artificial Intelligence (AAAI), 2023.

74. A. Radford, K. Narasimhan, T. Salimans, I. Sutskever et al., "Improving language understanding by generative pre-training," OpenAI Blog, 2018.

75. J. Wang, E. Shi, S. Yu, Z. Wu, C. Ma, H. Dai, Q. Yang, Y. Kang, J. Wu, H. Hu et al., "Prompt engineering for healthcare: Methodologies and applications," arXiv preprint arXiv:2304.14670, 2023.

76. P. Sahoo, A. K. Singh, S. Saha, V. Jain, S. Mondal, and A. Chadha, "A systematic survey of prompt engineering in large language models: Techniques and applications," arXiv preprint arXiv:2402.07927, 2024.

77. W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong et al., "A survey of large language models," arXiv preprint arXiv:2303.18223, 2023.

78. J. Bai, S. Bai, Y. Chu, Z. Cui, K. Dang, X. Deng, Y. Fan, W. Ge, Y. Han, F. Huang et al., "Qwen technical report," arXiv preprint arXiv:2309.16609, 2023.

79. J. Li, B. Hui, R. Cheng, B. Qin, C. Ma, N. Huo, F. Huang, W. Du, L. Si, and Y. Li, "Graphix-t5: Mixing pre-trained transformers with graph-aware layers for Text-to-SQL parsing," in Conference on Artificial Intelligence (AAAI), 2023.

## REFERENCES:

1. J. Li, B. Hui, G. QU, J. Yang, B. Li, B. Li, B. Wang, B. Qin, R. Geng, N. Huo, X. Zhou, C. Ma, G. Li, K. Chang, F. Huang, R. Cheng, and Y. Li. "Can LLM already serve as a database interface? a BIg bench for large-scale database grounded Text-to-SQLs," in Advances in Neural Information Processing Systems (NeurIPS), 2023.

2. Lijie Wang, Ao Zhang, Kun Wu, Ke Sun, Zhenghua Li, Hua Wu, Min Zhang, and Haifeng Wang. 2020. DuSQL: A Large-Scale and Pragmatic Chinese Text-

to-SQL Dataset. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6923–6935, Online. Association for Computational Linguistics.

3. T. Yu, R. Zhang, H. Er, S. Li, E. Xue, B. Pang, X. V. Lin, Y. C. Tan, T. Shi, Z. Li, Y. Jiang, M. Yasunaga, S. Shim, T. Chen, A. Fabbri, Z. Li, L. Chen, Y. Zhang, S. Dixit, V. Zhang, C. Xiong, R. Socher, W. Lasecki, and D. Radev, "CoSQL: A conversational Text-to-SQL challenge towards cross-domain natural language interfaces to databases," in Empirical Methods in Natural Language Processing and International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019

4. T. Yu, R. Zhang, K. Yang, M. Yasunaga, D. Wang, Z. Li, J. Ma, I. Li, Q. Yao, S. Roman, Z. Zhang, and D. Radev, "Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing and Text-to-SQL task," in Empirical Methods in Natural Language Processing (EMNLP), 2018.

5. V. Zhong, C. Xiong, and R. Socher, "Seq2sql: Generating structured queries from natural language using reinforcement learning," arXiv preprint arXiv:1709.00103, 2017.

6. Y. Gan, X. Chen, Q. Huang, and M. Purver, "Measuring and improving compositional generalization in text-toSQL via component alignment," in Findings of North American Chapter of the Association for Computational Linguistics (NAACL), 2022.

7. X. Pi, B. Wang, Y. Gao, J. Guo, Z. Li, and J.-G. Lou, "Towards robustness of Text-to-SQL models against natural and realistic adversarial table perturbation," in Association for Computational Linguistics (ACL), 2022.

8. Y. Gan, X. Chen, Q. Huang, and M. Purver, "Measuring and improving compositional generalization in text-toSQL via component alignment," in Findings of North American Chapter of the Association for Computational Linguistics (NAACL), 2022.

9. Y. Gan, X. Chen, and M. Purver, "Exploring underexplored limitations of cross-domain Text-to-SQL generalization," in Empirical Methods in Natural Language Processing (EMNLP), 2021.

10. Y. Gan, X. Chen, Q. Huang, M. Purver, J. R. Woodward, J. Xie, and P. Huang, "Towards robustness of Text-to-SQL models against synonym substitution," in Association for Computational Linguistics and International Joint Conference on Natural Language Processing (ACLIJCNLP), 2021.

11. X. Deng, A. H. Awadallah, C. Meek, O. Polozov, H. Sun, and M. Richardson, "Structure-grounded pretraining for Text-to-SQL," in North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 2021.

12. Q. Min, Y. Shi, and Y. Zhang, "A pilot study for Chinese SQL semantic parsing," in Empirical Methods in Natural Language Processing and International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019.

13. T. Yu, R. Zhang, M. Yasunaga, Y. C. Tan, X. V. Lin, S. Li, H. Er, I. Li, B. Pang, T. Chen, E. Ji, S. Dixit, D. Proctor, S. Shim, J. Kraft, V. Zhang, C. Xiong, R. Socher, and D. Radev, "SParC: Cross-domain semantic parsing in context," in Association for Computational Linguistics (ACL), 2019.

14. F. Lei, J. Chen, Y. Ye, R. Cao, D. Shin, H. SU, Z. SUO, H. Gao, W. Hu, P. Yin, V. Zhong, C. Xiong, R. Sun, Q. Liu, S. Wang, and T. Yu, "Spider 2.0: Evaluating language models on real-world enterprise Text-to-SQL workflows," in International Conference on Learning Representations (ICLR), 2025.

15. A. Tuan Nguyen, M. H. Dao, and D. Q. Nguyen, "A pilot study of Text-to-SQL semantic parsing for Vietnamese," in Findings of Empirical Methods in Natural Language Processing (EMNLP), 2020.

16. S. Chang, J. Wang, M. Dong, L. Pan, H. Zhu, A. H. Li, W. Lan, S. Zhang, J. Jiang, J. Lilien et al., "Dr. spider: A diagnostic evaluation benchmark towards Text-

to-SQL robustness," in International Conference on Learning Representations (ICLR), 2023.

17. C. Zhang, Y. Mao, Y. Fan, Y. Mi, Y. Gao, L. Chen, D. Lou, and J. Lin, "Finsql: model-agnostic llmsbased Text-to-SQL framework for financial analysis," in Conference on Management of Data (SIGMOD), 2024.

18. T. Zhang, T. Yu, T. B. Hashimoto, M. Lewis, W. tau Yih, D. Fried, and S. I. Wang, "Coder reviewer reranking for code generation," in International Conference on Machine Learning (ICML), 2023.

19. M. Pourreza and D. Rafiei, "DIN-SQL: Decomposed in-context learning of Text-to-SQL with self-correction," in Advances in Neural Information Processing Systems (NeurIPS), 2023.

2. C.-Y. Tai, Z. Chen, T. Zhang, X. Deng, and H. Sun, "Exploring chain of thought style prompting for textto-SQL," in Empirical Methods in Natural Language Processing (EMNLP), 2023.

21. X. Dong, C. Zhang, Y. Ge, Y. Mao, Y. Gao, J. Lin, D. Lou et al., "C3: Zero-shot Text-to-SQL with chatgpt," arXiv preprint arXiv:2307.07306, 2023.

22. B. Wang, C. Ren, J. Yang, X. Liang, J. Bai, L. Chai, Z. Yan, Q.-W. Zhang, D. Yin, X. Sun, and Z. Li, "Macsql: A multi-agent collaborative framework for texttto-sql," in International Conference on Computational Linguistics (COLING), 2024.

23. Y. Xie, X. Jin, T. Xie, M. Lin, L. Chen, C. Yu, L. Cheng, C. Zhuo, B. Hu, and Z. Li, "Decomposition for enhancing attention: Improving llm-based Text-to-SQL through workflow paradigm," in Findings of Association for Computational Linguistics (ACL), 2024.

24. Q. Zhang, J. Dong, H. Chen, W. Li, F. Huang, and X. Huang, "Structure guided large language model for sql generation," arXiv preprint arXiv:2402.13284, 2024.

25. Y. Fan, Z. He, T. Ren, C. Huang, Y. Jing, K. Zhang, and X. S. Wang, "Metasql: A generate-then-rank framework for natural language to sql translation," in International Conference on Data Engineering (ICDE), 2024.

26. Z. Li, X. Wang, J. Zhao, S. Yang, G. Du, X. Hu, B. Zhang, Y. Ye, Z. Li, R. Zhao, and H. Mao, "Pet-sql: A prompt-enhanced two-stage Text-to-SQL framework with cross-consistency," arXiv preprint arXiv:2403.09732, 2024.

27. T. Ren, Y. Fan, Z. He, R. Huang, J. Dai, C. Huang, Y. Jing, K. Zhang, Y. Yang, and X. S. Wang, "Purple: Making a large language model a better sql writer," in International Conference on Data Engineering (ICDE), 2024

28. G. Qu, J. Li, B. Li, B. Qin, N. Huo, C. Ma, and R. Cheng, "Before generation, align it! a novel and effective strategy for mitigating hallucinations in Text-to-SQL generation," in Findings of Association for Computational Linguistics (ACL), 2024

29. D. Lee, C. Park, J. Kim, and H. Park, "MCS-SQL: Leveraging multiple prompts and multiple-choice selection for Text-to-SQL generation," in International Conference on Computational Linguistics (COLING), 2025.

30. S. Talaei, M. Pourreza, Y.-C. Chang, A. Mirhoseini, and A. Saberi, "Chess: Contextual harnessing for efficient sql synthesis," arXiv preprint arXiv:2405.16755, 2024.

31. B. Li, Y. Luo, C. Chai, G. Li, and N. Tang, "The dawn of natural language to sql: Are we fully ready?" in International Conference on Very Large Data Bases (VLDB), 2024.

32. K. Maamari, F. Abubaker, D. Jaroslawicz, and A. Mhedhbi, "The death of schema linking? Text-to-SQL in the age of well-reasoned language models," in NeurIPS 2024 Third Table Representation Learning Workshop (NeurIPS), 2024.

33. Z. Cao, Y. Zheng, Z. Fan, X. Zhang, and W. Chen, "Rslsql: Robust schema linking in Text-to-SQL generation," arXiv preprint arXiv:24[4].00073, 2024.

34. J. Shi, B. Xu, J. Liang, Y. Xiao, J. Chen, C. Xie, P. Wang, and W. Wang, "Gen-SQL: Efficient Text-to-SQL by bridging natural language question and database schema with pseudo-schema," in International Conference on Computational Linguistics (COLING), 2025.

35. M. Deng, A. Ramachandran, C. Xu, L. Hu, Z. Yao, A. Datta, and H. Zhang, “Reforce: A Text-to-SQL agent with self-refinement, format restriction, and column exploration,” arXiv preprint arXiv:2502.00675, 2025.

36. C. Guo, Z. Tian, J. Tang, P. Wang, Z. Wen, K. Yang, and T. Wang, “Prompting gpt-3.5 for Text-to-SQL with de-semanticization and skeleton retrieval,” in Pacific Rim International Conference on Artificial Intelligence (PRICAI), 2024.

37. J. Jiang, K. Zhou, Z. Dong, K. Ye, X. Zhao, and J.- R. Wen, “StructGPT: A general framework for large language model to reason over structured data,” in Empirical Methods in Natural Language Processing (EMNLP), 2023.

38. L. Nan, Y. Zhao, W. Zou, N. Ri, J. Tae, E. Zhang, A. Cohan, and D. Radev, “Enhancing Text-to-SQL capabilities of large language models: A study on prompt design strategies,” in Findings of Empirical Methods in Natural Language Processing (EMNLP), 2023.

39. C. Guo, Z. Tian, J. Tang, S. Li, Z. Wen, K. Wang, and T. Wang, “Retrieval-augmented gpt-3.5-based Text-to-SQL framework with sample-aware prompting and dynamic revision chain,” in International Conference on Neural Information Processing (ICONIP), 2024.

40. D. Gao, H. Wang, Y. Li, X. Sun, Y. Qian, B. Ding, and J. Zhou, “Text-to-SQL empowered by large language models: A benchmark evaluation,” in International Conference on Very Large Data Bases (VLDB), 2024.

41. S. Chang and E. Fosler-Lussier, “Selective demonstrations for cross-domain Text-to-SQL,” in Findings of Empirical Methods in Natural Language Processing (EMNLP), 2023.

42. H. Zhang, R. Cao, L. Chen, H. Xu, and K. Yu, “ACT-SQL: In-context learning for Text-to-SQL with automatically-generated chain-of-thought,” in Findings of Empirical Methods in Natural Language Processing (EMNLP), 2023.

43. D. Wang, L. Dou, X. Zhang, Q. Zhu, and W. Che, “Improving demonstration diversity by human-free fusing for Text-to-SQL,” in Findings of Empirical Methods in Natural Language Processing (EMNLP), 2024.

44. Z. Hong, Z. Yuan, H. Chen, Q. Zhang, F. Huang, and X. Huang, “Knowledge-to-sql: Enhancing sql generation with data expert llm,” in Findings of Association for Computational Linguistics (ACL), 2024.

45. D. G. Thorpe, A. J. Duberstein, and I. A. Kinsey, “Dubosql: Diverse retrieval-augmented generation and fine tuning for Text-to-SQL,” arXiv preprint arXiv:2404.12560, 2024.

46. R. Toteja, A. Sarkar, and P. M. Comar, “In-context reinforcement learning with retrieval-augmented generation for Text-to-SQL,” in International Conference on Computational Linguistics (COLING), 2025.

47. J. Lee, I. Baek, B. Kim, and H. Lee, “Safe-sql: Self-augmented in-context learning with fine-grained example selection for Text-to-SQL,” arXiv preprint arXiv:2502.[4]438, 2025.

48. R. Sun, S. O. Arik, H. Nakhost, H. Dai, R. Sinha, P. Yin, and T. Pfister, “Sql-palm: Improved large language model adaptation for Text-to-SQL,” Transactions on Machine Learning Research (TMLR), 2023.

49. H. Xia, F. Jiang, N. Deng, C. Wang, G. Zhao, R. Mihalcea, and Y. Zhang, “Sql-craft: Text-to-SQL through interactive refinement and enhanced reasoning,” arXiv preprint arXiv:2402.14851, 2024.

50. Y. Gu, Y. Shu, H. Yu, X. Liu, Y. Dong, J. Tang, J. Srinivasa, H. Latapie, and Y. Su, “Middleware for llms: Tools are instrumental for language agents in 17 complex environments,” in Empirical Methods in Natural Language Processing (EMNLP), 2024.

51. M. Pourreza, H. Li, R. Sun, Y. Chung, S. Talaei, G. T. Kakkar, Y. Gan, A. Saberi, F. Ozcan, and S. O. Arik, “CHASE-SQL: Multi-path reasoning and preference optimized candidate selection in Text-to-SQL,” in International Conference on Learning Representations (ICLR), 2025.

52. F. Shi, D. Fried, M. Ghazvininejad, L. Zettlemoyer, and S. I. Wang, "Natural language to code translation with execution," in Empirical Methods in Natural Language Processing (EMNLP), 2022

53. A. Ni, S. Iyer, D. Radev, V. Stoyanov, W.-t. Yih, S. I. Wang, and X. V. Lin, "Lever: Learning to verify language-to-code generation with execution," in International Conference on Machine Learning (ICML), 2023.

54. X. Chen, M. Lin, N. Scharli, and D. Zhou, "Teaching ¨ large language models to self-debug," in International Conference on Learning Representations (ICLR), 2024.

55. H. A. Caferoglu and ˘ O. Ulusoy, "E-sql: Direct schema ¨ linking via question enrichment in Text-to-SQL," arXiv preprint arXiv:2409.16751, 2024.

56. S. Kou, L. Hu, Z. He, Z. Deng, and H. Zhang, "Cllms: Consistency large language models," in International Conference on Machine Learning (ICML), 2024.

57. H. Li, J. Zhang, H. Liu, J. Fan, X. Zhang, J. Zhu, R. Wei, H. Pan, C. Li, and H. Chen, "Codes: Towards 15 building open-source language models for Text-to-SQL," in Conference on Management of Data (SIGMOD), 2024.

58. F. Xu, Z. Wu, Q. Sun, S. Ren, F. Yuan, S. Yuan, Q. Lin, Y. Qiao, and J. Liu, "Symbol-llm: Towards foundational symbol-centric interface for large language models," in Association for Computational Linguistics (ACL), 2024.

59. A. Zhuang, G. Zhang, T. Zheng, X. Du, J. Wang, W. Ren, S. W. Huang, J. Fu, X. Yue, and W. Chen, "Structlm: Towards building generalist models for structured knowledge grounding," in Conference on Language Modeling (COLM), 2024.

60. Y. Gao, Y. Liu, X. Li, X. Shi, Y. Zhu, Y. Wang, S. Li, W. Li, Y. Hong, Z. Luo et al., "Xiyan-sql: A multigenerator ensemble framework for Text-to-SQL," arXiv preprint arXiv:24[4].08599, 2024.

61. M. Pourreza and D. Rafiei, "Dts-sql: Decomposed textto-sql with small large language models," in Findings of Empirical Methods in Natural Language Processing (EMNLP), 2024.

62. Z. Yuan, H. Chen, Z. Hong, Q. Zhang, F. Huang, and X. Huang, "Knapsack optimization-based schema linking for llm-based Text-to-SQL generation," arXiv preprint arXiv:2502.129[4], 2025.

63. S. K. Gorti, I. Gofman, Z. Liu, J. Wu, N. Vouitsis, G. Yu, J. C. Cresswell, and R. Hosseinzadeh, "MScSQL: Multi-sample critiquing small language models for Text-to-SQL translation," in North American Chapter of the Association for Computational Linguistics (NAACL), 2024.

64. Y. Qin, C. Chen, Z. Fu, Z. Chen, D. Peng, P. Hu, and J. Ye, "ROUTE: Robust multitask tuning and collaboration for Text-to-SQL," in International Conference on Learning Representations (ICLR), 2025.

65. P. Ma and S. Wang, "Mt-teql: evaluating and augmenting neural nlidb on real-world linguistic and schema variations," in International Conference on Very Large Data Bases (VLDB), 2021.

66. P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "SQuAD: 100,000+ questions for machine comprehension of text," in Empirical Methods in Natural Language Processing (EMNLP), 2016.

67. F. Li and H. V. Jagadish, "Constructing an interactive natural language interface for relational databases," in International Conference on Very Large Data Bases (VLDB), 2014.

68. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, 1997.

69. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems (NeurIPS), 2017.

70. D. Choi, M. C. Shin, E. Kim, and D. R. Shin, "Ryansql: Recursively applying sketch-based slot fillings for complex Text-to-SQL in cross-domain databases," Computational Linguistics, 2021.

71. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 2019.

72. Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, “Roberta: A robustly optimized bert pretraining approach,” arXiv preprint arXiv:1907.[4]692, 2019.

73. H. Li, J. Zhang, C. Li, and H. Chen, “Resdsql: Decoupling schema linking and skeleton parsing for Text-to-SQL,” in Conference on Artificial Intelligence (AAAI), 2023.

74. A. Radford, K. Narasimhan, T. Salimans, I. Sutskever et al., “Improving language understanding by generative pre-training,” OpenAI Blog, 2018.

75. J. Wang, E. Shi, S. Yu, Z. Wu, C. Ma, H. Dai, Q. Yang, Y. Kang, J. Wu, H. Hu et al., “Prompt engineering for healthcare: Methodologies and applications,” arXiv preprint arXiv:2304.14670, 2023.

76. P. Sahoo, A. K. Singh, S. Saha, V. Jain, S. Mondal, and A. Chadha, “A systematic survey of prompt engineering in large language models: Techniques and applications,” arXiv preprint arXiv:2402.07927, 2024.

77. W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong et al., “A survey of large language models,” arXiv preprint arXiv:2303.18223, 2023.

78. J. Bai, S. Bai, Y. Chu, Z. Cui, K. Dang, X. Deng, Y. Fan, W. Ge, Y. Han, F. Huang et al., “Qwen technical report,” arXiv preprint arXiv:2309.16609, 2023.

79. J. Li, B. Hui, R. Cheng, B. Qin, C. Ma, N. Huo, F. Huang, W. Du, L. Si, and Y. Li, “Graphix-t5: Mixing pre-trained transformers with graph-aware layers for Text-to-SQL parsing,” in Conference on Artificial Intelligence (AAAI), 2023.