



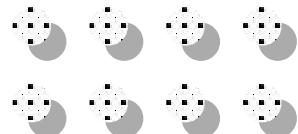
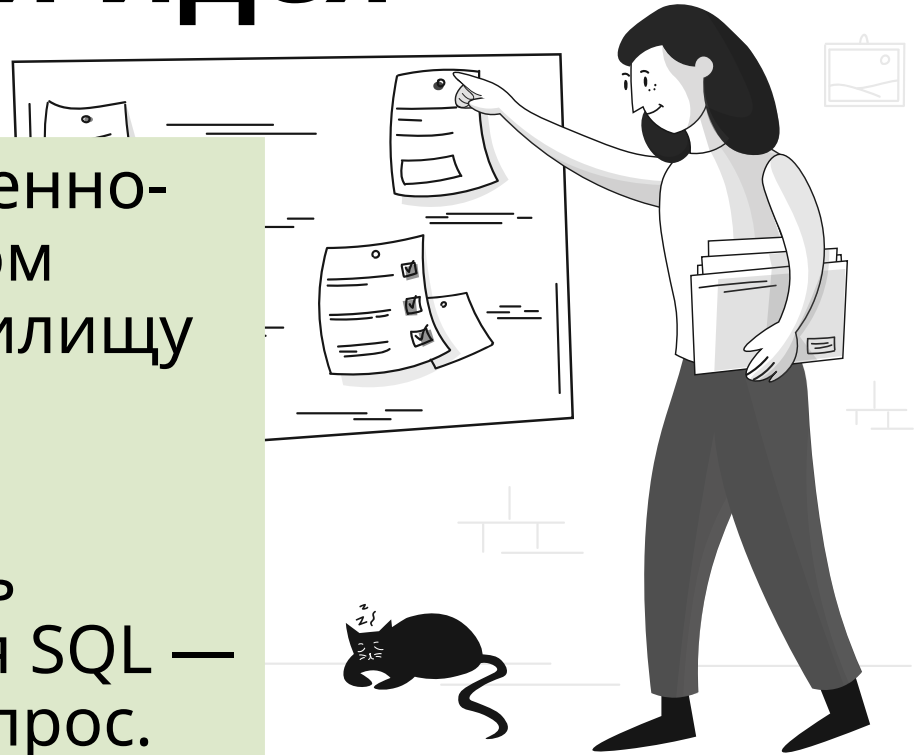
NL2SQL (Natural Language to SQL)

Владимир Власов

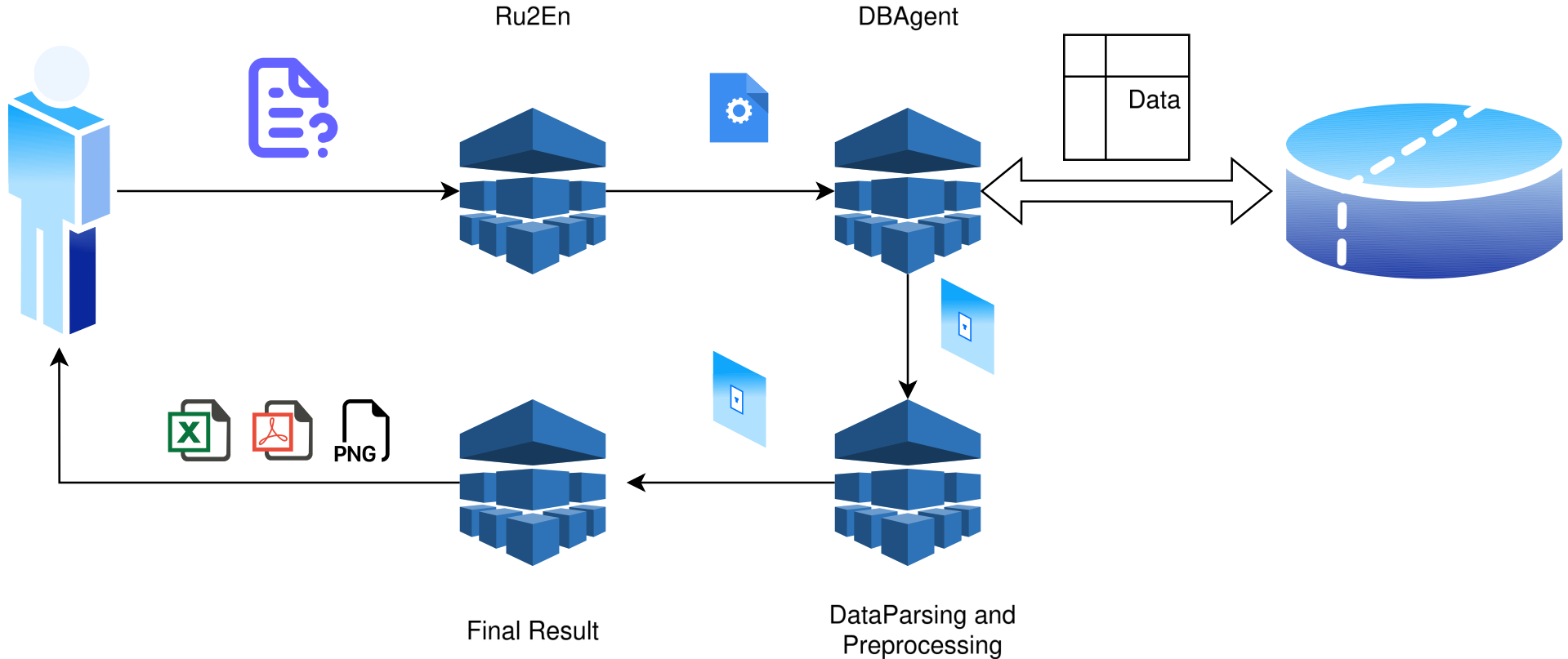
23.09.25

Большая идея

- Сервис преобразует естественно-языковые запросы на русском языке в SQL-запросы к хранилищу данных
- Цель: предоставить бизнес-пользователям возможность получать данные без знания SQL — через простой текстовый запрос.



Большой проект

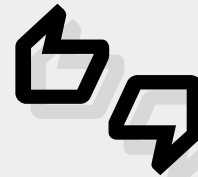


Проверяем гипотезу

- Берём
SQLcoder-7b



- Создаём датасет
и дообучаем
модель

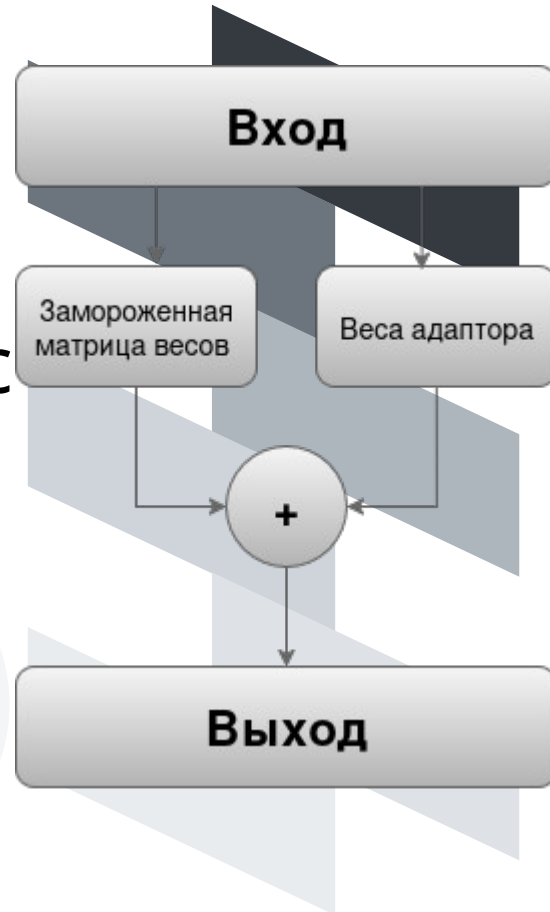


- Тестируем
подход



Входные данные

- defog/
sqlcoder-7b
- Учебная схема с
<https://sql-ex.ru>
- 100 написанных
вручную
запросов



- QLoRA
- RTX 3090 Ti
- 300 запросов с
аугментацией

Данные для обучения

```
{"Schema": ""
```

```
Table: Product
```

```
Purpose: Contains information about manufacturers and product models
```

```
Columns:
```

- maker (varchar(10)): Manufacturer name
- model (varchar(50)): Unique model number (primary key)
- type (varchar(50)): Product type - 'PC', 'Laptop', or 'Printer'

```
.....
```

```
"" ,
```

```
"instruction": "Выведите модели ПК, упорядоченные по возрастанию цены."
```

```
"output": "SELECT model FROM PC ORDER BY price ASC;"
```

```
}
```

Ключевые метрики

Эпоха	Train Loss	Eval Loss	Token Accuracy (train)	Token Accuracy (eval)
0.38	0.8446	—	80.2%	—
1.92	0.1015	0.0959	97.4%	97.6%
3.85	0.0570	0.0618	98.4%	98.2%
7.69	0.0359	0.0496	98.9%	98.5%
Итог (эпоха 8.0)	0.1113 (среднее)	0.0496 (финальная)	98.9%	98.5%

Loss падает, Аккуратность растёт, следовательно, дообучение было успешным

Ключевые метрики

