

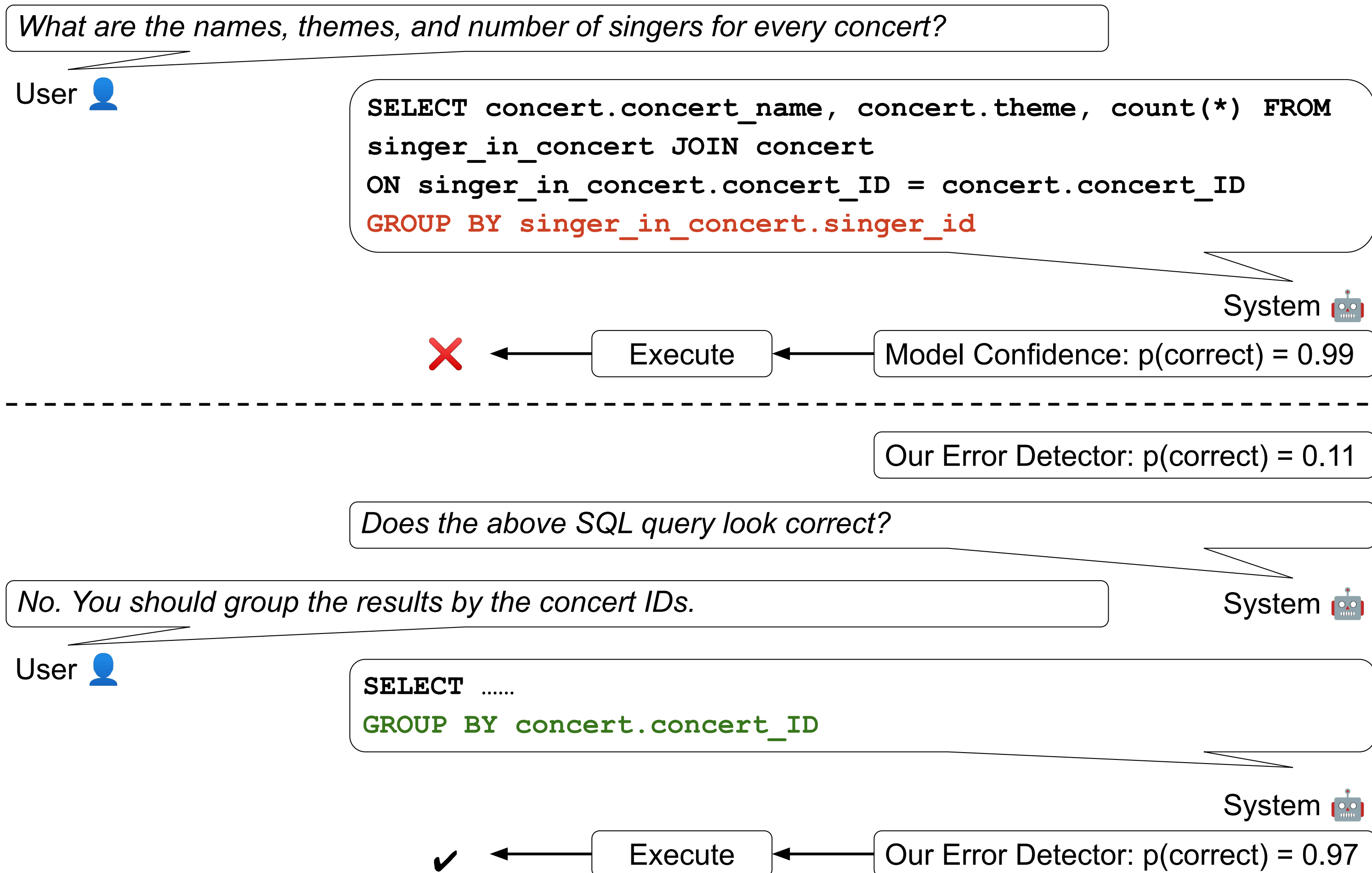
# Error Detection for Text-to-SQL Semantic Parsing

Shijie Chen, Ziru Chen, Huan Sun, Yu Su



THE OHIO STATE UNIVERSITY

## Overview



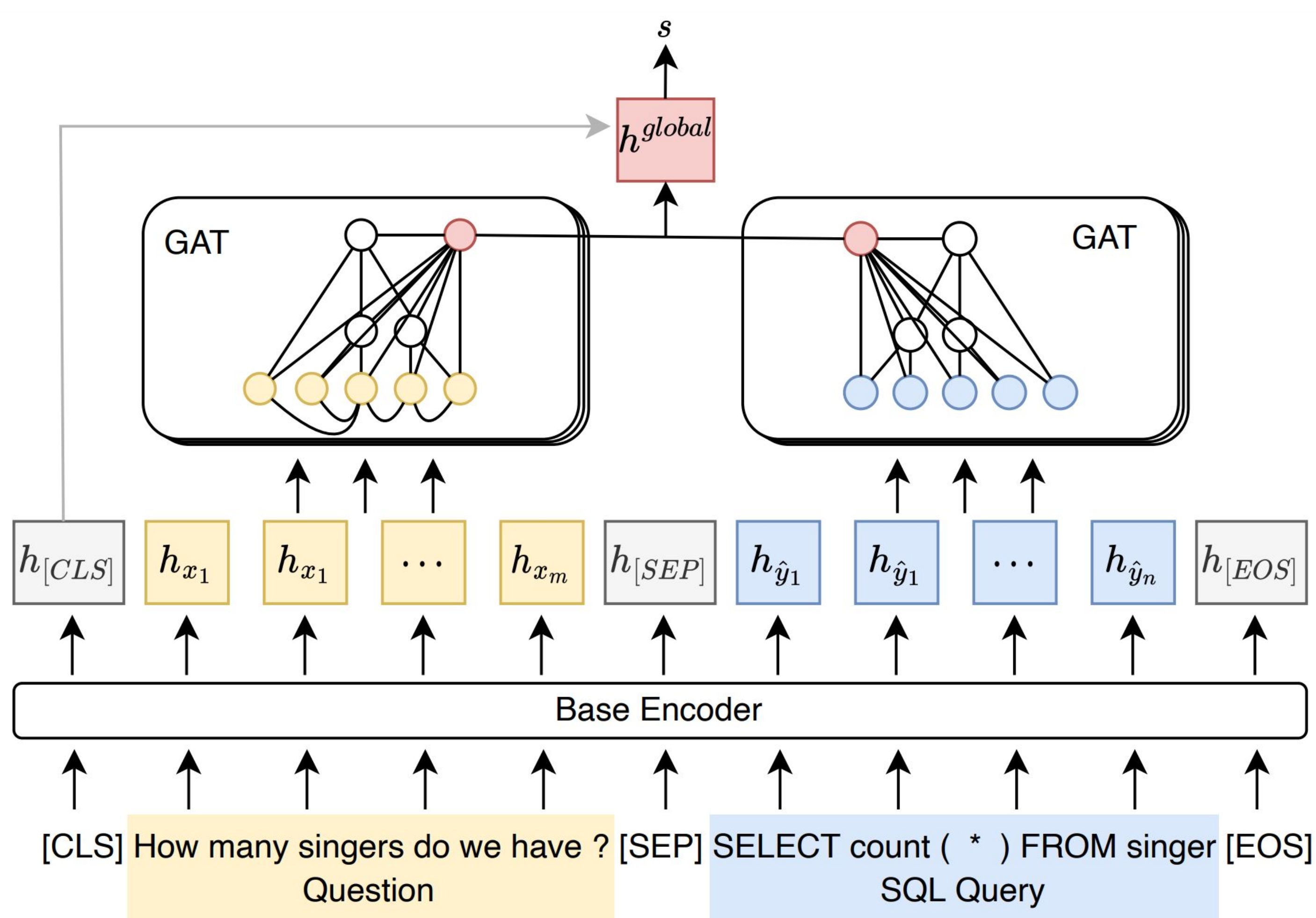
### Task Formulation

- Given a natural language question and a SQL query predicted by a text-to-SQL parser, the error detection model estimates the probability of event “the SQL query is correct.”

### Contributions

- We propose the first generalizable and parser-independent error detection model for text-to-SQL parsing that is effective on multiple tasks and different parser designs without any task-specific adaptation.
- Our evaluations show that the proposed error detection model outperforms parser-dependent uncertainty metrics and could maintain its high performance under cross-parser evaluation settings.
- We show through simulated interactions that our error detector could improve the efficiency and usefulness of interactive text-to-SQL parsing systems.

## Parser-Independent Error Detector



### Data Collection

- Motivation: We consider two factors, **insufficient training data** and the **cross-domain generalization gap**, that could lead to text-to-SQL parsing errors.
- Strategy: Split the Spider training set into two equal-sized subsets by databases and perform cross-validation to synthesize executable SQL queries with errors.

### Graph Input Construction

- Natural language question: Dependency parse tree + constituency parse tree.
- SQL query: Abstract syntax tree.

### Model Architecture

- Use CodeBERT for contextualized encoding of question and SQL query
- Use Graph attention networks to model structural features of natural language and SQL.

## Summary of Experiments and Results

Parser	Model	Positive			Negative			Acc	AUC
		Precision	Recall	F1	Precision	Recall	F1		
SmBoP	SmBoP <sup>P</sup>	80.8	<b>94.5</b>	86.6	52.1	18.9	23.9	77.4	67.0
	SmBoP <sup>S</sup>	81.5	91.9	85.7	56.6	25.3	29.4	76.9	79.2
	CodeBERT	82.9	92.6	86.7	<b>60.8</b>	33.9	36.3	78.3	80.8
	CodeBERT+GAT	<b>85.0</b>	90.6	<b>87.2</b>	56.7	<b>44.4</b>	<b>46.4</b>	<b>79.8</b>	<b>81.7</b>
RESDSL	RESDSL <sup>P</sup>	79.5	93.9	85.5	55.1	15.1	19.7	75.6	76.2
	RESDSL <sup>S</sup>	80.1	<b>95.2</b>	86.5	54.2	16.2	21.6	77.5	76.7
	CodeBERT	83.1	94.8	<b>88.3</b>	61.2	32.1	41.0	81.0	<b>80.7</b>
	CodeBERT+GAT	<b>83.8</b>	93.8	<b>88.3</b>	<b>61.7</b>	<b>37.0</b>	<b>45.2</b>	<b>81.2</b>	<b>80.7</b>
NatSQL	NatSQL <sup>P</sup>	78.1	<b>93.2</b>	84.6	67.3	36.1	45.4	76.3	79.2
	NatSQL <sup>S</sup>	77.0	91.4	83.0	62.8	33.1	40.3	74.0	76.2
	CodeBERT	84.6	90.8	<b>87.3</b>	<b>72.3</b>	60.5	64.6	<b>81.8</b>	86.5
	CodeBERT+GAT	<b>86.6</b>	87.4	86.8	68.5	<b>68.1</b>	<b>67.0</b>	81.7	<b>86.9</b>

### Text-to-SQL Parsers

- SmBoP: A parser featuring a bottom-up decoder guided by relational algebra grammar.
- RESDSL: A parser using transformer decoder without no grammatical constraints.
- NatSQL: A parser using LSTM decoder guided by context-free grammar.

### Parser-Dependent Baselines

- Prediction probability: The sequence probability/score of predicted SQL queries.
- Dropout-based uncertainty: The standard deviation of SQL prediction probabilities/scores over 10 inference passes with dropout enabled.

### Error Detection Results

- On all three parsers, our CodeBERT-based error detector significantly outperforms the two parser-dependent uncertainty metrics, especially on negative samples.
- CodeBERT+GAT further improves the overall error detection performance, especially in recall on incorrect predictions (7.7% absolute improvement on average).

### Action Triggering Results

- For answer triggering, compared to parser-dependent metrics, our error detectors allow the system to answer 76% to 175% more questions while maintaining a precision of 95%.
- For interaction triggering, our model brings a 16% to 33% reduction to the number of interactions required for reaching an accuracy of 95%.

