**Ontology of Generated SQL Errors**

**Report on SQL Query Mutation Analysis**

**1. Introduction**

This report analyzes the ontology of errors generated by an AI-based SQL mutation system. The system takes correct SQL queries as input and produces syntactically valid but logically incorrect variations. The goal is to categorize and quantify the types of errors introduced, providing insights into common SQL mistakes.

**2. Error Classification**

The generated errors fall into distinct categories, each affecting query correctness differently. Below is a taxonomy of observed error types:

**2.1. Join Condition Errors (30%)**

* **Inequality Joins (≠, <, >)**
  + Example: **INNER JOIN T2 ON T1.id != T2.id** (instead of **=**)
* **Wrong Join Columns**
  + Example: Joining on **T1.constructorId = T2.nationality** (invalid relationship)

**2.2. Filter Condition Errors (25%)**

* **Incorrect Comparison Operators**
  + Example: **WHERE name = 'John'** → **WHERE name != 'John'**
* **Wrong Literal Values**
  + Example: **WHERE region = 'east Bohemia'** → **WHERE region = 'wrong value'**

**2.3. Logical Operator Errors (20%)**

* **AND/OR Swaps**
  + Example: **WHERE (A > 5 AND B < 10)** → **WHERE (A > 5 OR B < 10)**
* **Added False Conditions**
  + Example: Appending **AND 1=0** to WHERE clauses

**2.4. Aggregation & Sorting Errors (15%)**

* **Wrong ORDER BY Direction**
  + Example: **ORDER BY score ASC** → **ORDER BY score DESC**
* **Incorrect LIMIT Usage**
  + Example: **LIMIT 10** → **LIMIT 1**

**2.5. Structural Errors (10%)**

* **Missing/Extra Clauses**
  + Example: Omitting **GROUP BY** when using aggregates
* **Subquery Errors**
  + Example: Correlated subqueries with incorrect references

**3. Distribution of Errors (Figure 1)**

*Figure 1: Pie chart showing the prevalence of different SQL error types in generated mutations.*A pie chart with text on it

AI-generated content may be incorrect.