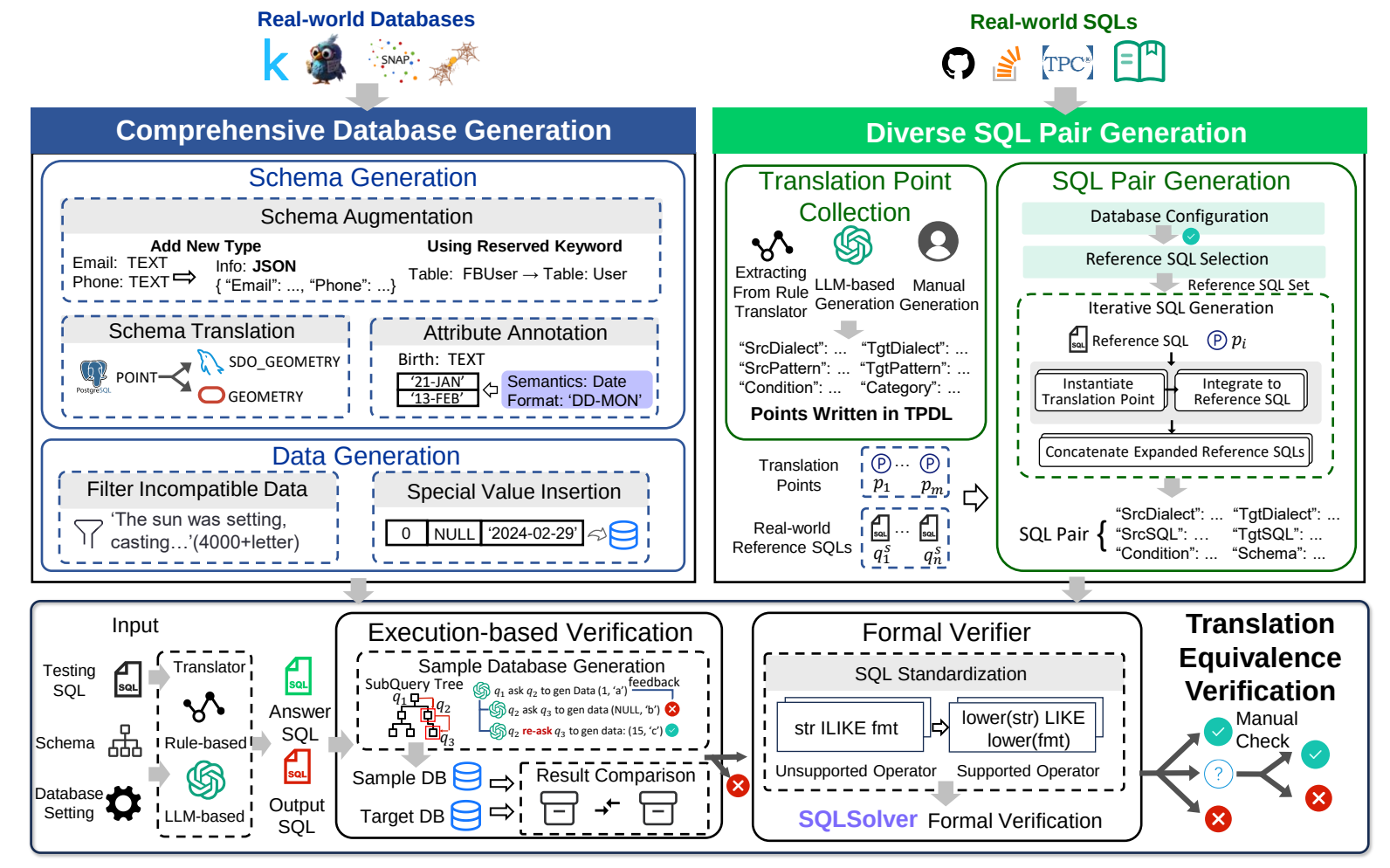
Real-world Databases
Real-world SQLs
Comprehensive Database Generation
Diverse SQL Pair Generation
Schema Generation
Schema Augmentation
Add New Type
Email: TEXT
Phone: TEXT
Info: JSON
{ "Email": ..., "Phone": ...}
Using Reserved Keyword
Table: FBUser → Table: User
Schema Translation
POINT
SDO_GEOMETRY
GEOMETRY
Attribute Annotation
Birth: TEXT
'21-JAN'
'13-FEB'
Semantics: Date
Format: 'DD-MON'
Data Generation
Filter Incompatible Data
'The sun was setting, casting…'(4000+letter)
Special Value Insertion
0
NULL
'2024-02-29'
Translation Point Collection
Extracting From Rule Translator
LLM-based Generation
Manual Generation
"SrcDialect": ... "TgtDialect": ...
"SrcPattern": ... "TgtPattern": ...
"Condition": ... "Category": ...
Points Written in TPDL
Translation Points
Real-world Reference SQLs
SQL Pair Generation
Database Configuration
Reference SQL Selection
Reference SQL Set
Iterative SQL Generation
Reference SQL
Instantiate Translation Point
Integrate to Reference SQL
Concatenate Expanded Reference SQLs
SQL Pair
"SrcDialect": ... "TgtDialect": ...
"SrcSQL": ... "TgtSQL": ...
"Condition": ... "Schema": ...
Input
Testing SQL
Schema
Database Setting
Translator
Rule-based
LLM-based
Answer SQL
Output SQL
Execution-based Verification
Sample Database Generation
SubQuery Tree
feedback
re-ask
Sample DB
Target DB
Result Comparison
Formal Verifier
SQL Standardization
str ILIKE fmt
lower(str) LIKE lower(fmt)
Unsupported Operator
Supported Operator
SQLSolver Formal Verification
Translation Equivalence Verification
Manual Check