|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metric No** | **Metric Name** | **Description** | **Measurement** | **Formula** |
| 1 | **Features provided all at once** | Whether we provided the Ai with the features all at once or in more than one prompts | Yes/No | - |
| 2 | **Domain/Daos/Services code given from the start** | Whether we instructed the Ai to provide the Domain/Daos/Services code first or not | Yes/No | - |
| 3 | **Prompts Required for Complete Step Definitions** | Count the number of prompts needed to generate all step definitions with as much code code as possible. | Integer value | - |
| 4 | **DAO Utilization and Accuracy** | Evaluate the usage, accuracy, and reminders for DAOs by the Ai. | DAO Usage : 1 if used from the start / 0.5 if not used from the start, Integer values | DAO Utilization Score = DAO Usage × (Accurate DAOs Guessed - Reminders Required) |
| 5 | **Service Utilization and Accuracy** | Evaluate the acceptabillity and reminders for services by the AI. | Service Usage : 1 if used from the start / 0.5 if not used from the start, Integer values | Service Utilization Score = Service Usage × (Acceptable Services - Reminders Required) |
| 6 | **Accuracy of Domain Classes Guessed/Used** | Measure the percentage of domain classes guessed or used correctly by the AI | Percentage value | Accuracy of Domain classes Percentage = (Number of correct guessed Domain classes / Total Number of Domain classes) × 100 |
| 7 | **Acceptable Step Definition Solutions** | Measure the percentage of Step Definitions that are acceptable based on completeness, accuracy, functionality, and integration. | Percentage value | Acceptable Step Definition Percentage = (Number of Acceptable Step Definitions / Total Number of Step Definitions) × 100 |
| 8 | **Better Than Acceptable Step Definition Solutions** | Measure the percentage of Step Definitions that had better than acceptable solutions | Percentage Value | Better than Acceptable Percentage= (Number of Better than Acceptable Step Definitions/ Total Number of Srtep definitions) × 100 |
| 9 | **Object Instances Replacement with Code Variables** | Describe whether the Ai replaces instances of objects expressed in natural language with variables in the code. | String | - |
| 10 | **Accuracy of Attributes in Correctly Guessed Domain Classes** | Measure how accurately the AI guesses or uses attributes in the correct domain classes (phase 1,2) | Integer values | Attributes Accuracy Score = (Correct Attributes / (Correct Attributes + Missing Attributes)) × 100 |
| 11 | **Accuracy of Methods (Excluding Getters and Setters) in Correctly Guessed Domain Classes** | Measure how accurately the AI guesses or uses methods (excluding getters and setters) in domain classes(phase 1,2,3) | Integer and binary values | Methods Accuracy Score = (Correct Methods) / (Correct Methods + Missing Methods) × 100 |
| 12 | **Repetition/additional explanation** | Count the instances where the Ai needed addiotional explanation or repetition of prompts | Integer Value | - |
| 13 | **Empty Step Definitions** | Count the instances where the Ai provided step definitions empty and without code | Integer Value | - |

# Evaluation Table