**calendar\_SMT\_DeepSeek-V3 (92% correct within 5 iterations of evaluation)**

|  |  |
| --- | --- |
| **Number of Iterations** | **Count** |
| 1 | 75 |
| 2 | 12 |
| 3 | 4 |
| 4 | 2 |
| 5 | 7 |
| **Grand Total** | **100** |

Error analysis:

pass 1:

17 errors (syntax), 8 wrong plan (semantics)

Pass 2:

8 error (syntax), 5 wrong plan (semantics)

Pass 3:

5 error (syntax), 4 wrong plan (semantics)

Pass 4:

3 error (syntax), 4 wrong plan (semantics)

Pass 5:

4 error (syntax), 2 wrong plan (semantics)

**meeting\_SMT\_DeepSeek-V3 (52% correct within 5 iterations of evaluation)**

|  |  |
| --- | --- |
| **Number of Iterations** | **Count** |
| 1 | 22 |
| 2 | 12 |
| 3 | 7 |
| 4 | 5 |
| 5 | 54 |
| **Grand Total** | **100** |

Error analysis:

pass 1:

41 errors (syntax), 37 wrong plan (semantics)

Pass 2:

15 error (syntax), 51 wrong plan (semantics)

Pass 3:

11 error (syntax), 48 wrong plan (semantics)

Pass 4:

11 error (syntax), 43 wrong plan (semantics)

Pass 5:

11 error (syntax), 37 wrong plan (semantics)

**trip\_SMT\_DeepSeek-V3 (29% correct within 5 iterations of evaluation)**

|  |  |
| --- | --- |
| **Number of Iterations** | **Count** |
| 1 | 18 |
| 2 | 5 |
| 3 | 4 |
| 4 | 1 |
| 5 | 72 |
| **Grand Total** | **100** |

Error analysis:

pass 1:

42 errors (syntax), 40 wrong plan (semantics)

pass 2:

29 errors (syntax), 48 wrong plan (semantics)

pass 3:

23 errors (syntax), 50 wrong plan (semantics)

pass 4:

26 errors (syntax), 46 wrong plan (semantics)

pass 5:

24 errors (syntax), 47 wrong plan (semantics)