|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Phase 1 | | | | | | | | | | | | | | | | | |
| GPT -3.5 | | | | | Github COPILOT 3.5 Turbo | | | | | GPT -4 | | | | GPT -4o | | | |
| Metric No | Chat 1 | Chat 2 | Chat 3 | Chat 4 | Chat 1 | Chat 2 | Chat 3 | Chat 4 | Chat 5 | Chat 1 | Chat 2 | Chat 3 | Chat 4 | Chat 1 | Chat 2 | Chat 3 |
| 1 | Yes | Yes | No | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | No | Yes | Yes | No |
| 2 | No | Yes | Yes | Yes | No | Yes | No | No | Yes | No | Yes | No | Yes | No | Yes | No |
| 3 | 13 | 10 | 11 | 11 | 8 | 11 | 7 | 9 | 8 | 9 | 8 | 8 | 8 | 5 | 7 | 5 |
| 4 | 1 \* ( 3 -0)= 3 | 1 \* (2 – 1)=1 | 1\*(2-0)= 2 | 1\*(1.0-0)=1 | 0 | 0 | 1\*(1-0)=1 | 1\*(3-0)=3 | 1\*(3-0)=3 | 0.5\*(2-1)=0.5 | 1\*(3-0)=3 | 1\*(1-0)=1 | 1\*(2-0)=2 | 1\*(2-0)=2 | 1\*(2-0)=2 | 1\*(2-0)=2 |
| 5 | 1 \* (3-0)= 3 | 1\*(3-0) 3 | 1\*(2-0)=2 | 1\*(3-0)=3 | 1\*(3-0)=3 | 1\*(1-0)= 1 | 1\*(3-0)=3 | 1\*(3-0)=3 | 1\*(4-0)=4 | 1\*(3-0)=3 | 1\*(3-0)=3 | 1\*(3-0)=3 | 1\*(4-0)=4 | 1\*(3-0)=3 | 1\*(3-0)=3 | 1\*(4-0)=4 |
| 6 | 28.57 % | 42.85% | 28.57% | 42.85% | 28.57% | 42.85% | 42.85 % | 42.85 % | 42.85 % | 28.57% | 42.85 % | 28.57% | 28.57% | 28.57% | 28.57% | 28.57% |
| 7 | 27.08% | 37.5% | 22.91% | 25% |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 | 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 9 | The system used George Red as a persona in the first feature, but it didn’t use them at all the other features, as well as the items as hard coded variables | The system used George Red as a persona, but not in all the steps, but didn’t understand the items as hard coded variables | The system didn’t understand personas or hard code variables at all | The system didn’t understand personas or hard code variables at all | The Ai partly used personas and hard coded variables, in not that many step definitions | The Ai used Personas like George Red in all the step definitions, as well as for the items , so it understood good what to do with these hard coded variables | The Ai used Personas like George Red in all the step definitions, as well as for the items , so it understood good what to do with these hard coded variables | The Ai used Personas like George Red in all the step definitions, as well as for the items , so it understood good what to do with these hard coded variables | The Ai used Personas like George Red in all the step definitions, as well as for the items , so it understood good what to do with these hard coded variables | The system used only George Red as a persona but it didn’t use the other hard coded variables like the item names | The Ai didn’t understand personas or hard coded variables at all | The Ai didn’t understand personas or hard coded variables at all | The Ai didn’t understand personas or hard coded variables at | The Ai only understood the personas of George Red | The Ai somewhat understood personas, but didn’t use them in all the step definitions | The Ai didn’t understand personas or hard coded variables at |
| 10 | (3/7 + 0/3 +4/4)= 50% | (3/7 + 3/7 + 3/4) = 50% | (2/7 + 0/3)= 20% | (2/7+ 0/3 + 2/4)= 28.57% | (2/7 + 0/3 + 2/4 ) 28.57% | (2/7 + 0/3 + 4/4 )= 42.85% | (2/7 + 0/3 + 4/4)= 42.85% | (2/7 + 0/3 + 4/4) = 42.85% | (2/7+ 0/3 + 4/4)= 42.85% | (2/7+ 0/3)=20% | (2/7 +1/3+  4/4)= 50% | (2/7 + 0/3)= 20% | (1/7+0/3)= 10% | (2/7 + 0/3)= 20% | (2/7+1/7)= 21.42% | (2/7 + 0/3)=20% |
| 11 | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |  |  |  |
| 13 | 4 | 4 | 4 | 4 | 4 | 5 | 2 | 3 | 3 | 5 | 2 | 2 | 2 | 0 | 1 | 0 |
| 14 | 28 | 7 | 15 | 11 | 23 | 29 | 11 | 19 | 8 | 10 | 4 | 2 | 2 | 0 | 10 | 0 |