# Table D: Performance by Model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Paradigm | Model | Questions | Semantic Similarity | Faithfulness | Answer Correctness | Context Precision | Total Tokens | Avg Cost ($) | Avg Latency (s) |
| Naive | GPT-4 | 60 | 0.898 | 0.862 | 0.551 | 0.743 | 868.0 | 0.0286 | 3.7 |
| Naive | DeepSeek | 60 | 0.901 | 0.9 | 0.604 | 0.762 | 913.0 | 0.0004 | 10.8 |
| Advanced | GPT-4 | 60 | 0.899 | 0.881 | 0.611 | 0.753 | 931.0 | 0.0307 | 14.8 |
| Advanced | DeepSeek | 60 | 0.901 | 0.92 | 0.614 | 0.733 | 973.0 | 0.0004 | 23.2 |
| Agentic | GPT-4 | 60 | 0.904 | 0.873 | 0.626 | 0.757 | 1688.0 | 0.0556 | 8.0 |
| Agentic | DeepSeek | 60 | 0.903 | 0.893 | 0.591 | 0.712 | 1643.0 | 0.0006 | 19.0 |