# Table D: Performance by Model

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Paradigm | gpt-4 | gpt-4 (SD) | gpt-4 (n) | deepseek-chat | deepseek-chat (SD) | deepseek-chat (n) | Model Agreement (r) | CV Across Models |
| Naive | 0.898 | 0.034 | 60 | 0.901 | 0.026 | 60 | 0.76 | 0.002 |
| Advanced | 0.899 | 0.034 | 60 | 0.901 | 0.029 | 60 | 0.762 | 0.001 |
| Agentic | 0.904 | 0.032 | 60 | 0.903 | 0.025 | 60 | 0.839 | 0.001 |