

Smarter Doesn't Mean More Humane

| | | | | |
|---|-----------------|----------|--------------|-------------|
| OpenAI GPT 5 (2025-08-07) | 0.807 | 0.860 | 0.940 | 0.830 |
| Google Gemini 3 Pro (Preview) | 0.799 | 0.780 | 0.930 | -0.450 |
| Grok 4 (0709) | 0.785 | 0.690 | 0.890 | -0.730 |
| Anthropic CLAUDE 4 SONNET (2025-05-14) | 0.762 | 0.690 | 0.840 | 0.510 |
| Anthropic CLAUDE 4 5 SONNET (2025-09-29) | 0.752 | 0.750 | 0.890 | 0.770 |
| Google Gemini 2 5 Pro | 0.745 | 0.770 | 0.900 | -0.720 |
| OpenAI GPT 4 1 (2025-04-14) | 0.727 | 0.670 | 0.820 | -0.600 |
| Meta Llama 4 Maverick (17Bx128E) Instruct FP8 | 0.718 | 0.590 | 0.650 | -0.140 |
| Google Gemini 2.0 Flash 001 | 0.679 | 0.750 | 0.820 | -0.710 |
| OpenAI GPT 4O (2024-11-20) | 0.634 | 0.680 | 0.770 | -0.610 |
| | HELM Capability | Baseline | Good Persona | Bad Persona |