

Agentic AI in Action: Deep Research, Claude Code, and Agent Mode

LLM Model Comparisons

- Tested several leading chat models: Claude Sonnet 4.5, Opus 4.1, Gemini 2.5 Pro, Grok 4, DeepSeek
- Claude Sonnet 4.5: strong at nuanced reasoning, long-form writing, code generation, ethical reasoning; made a factual error about real-time info
- Perplexity noted as strong for real-time search, likened to a search engine
- Claude gave a poetic, evocative answer when describing the color blue to someone blind—highlighted as a key strength
- Simple word count question used as a reasoning test; Sonnet 4.5 and Opus 4.1 both initially failed, Opus succeeded with "extended thinking"
- Gemini 2.5 Pro gave a lengthy but incorrect answer to the word count challenge
- Grok 4 Fast answered the word count correctly on the first try
- DeepSeek overthought the word count question, ultimately gave the wrong answer after a long reasoning process

Observations on Model Behavior

- Some models struggle with meta-questions or overthink simple tasks
- "Extended thinking" can improve accuracy for certain models (e.g., Opus 4.1)
- Grok 4 Fast praised for quick, accurate reasoning
- DeepSeek described as powerful but prone to getting stuck in reasoning loops

Features Beyond Chat

- Chat interfaces offer more than just conversation—highlighted "deep research" as a valuable feature
- Deep research lets the LLM perform extended, multi-step analysis and returns with comprehensive answers
- Compared LLM deep research to having a junior analyst do work in the background

Deep Research Demo

- Used ChatGPT5's deep research feature (premium subscription, \$200/month, 250 deep research tasks/month)
- Research prompt: greatest commercial impact of GenAI and agentic AI, most successful commercial applications, revenue, subscription models, and innovations
- Model asked clarifying questions: public companies only, industry focus, global scope, time frame
- Clarified: any company, any industry, global, include earlier examples, all in 2025
- Research process described as robust and thorough, likened to delegating work to an analyst

General Impressions

- LLMs show varied strengths: poetic language, technical reasoning, real-time info, or deep research
- Each model has unique quirks and limitations, especially with reasoning and meta-questions

- Deep research and agentic AI features seen as major value-adds for power users