



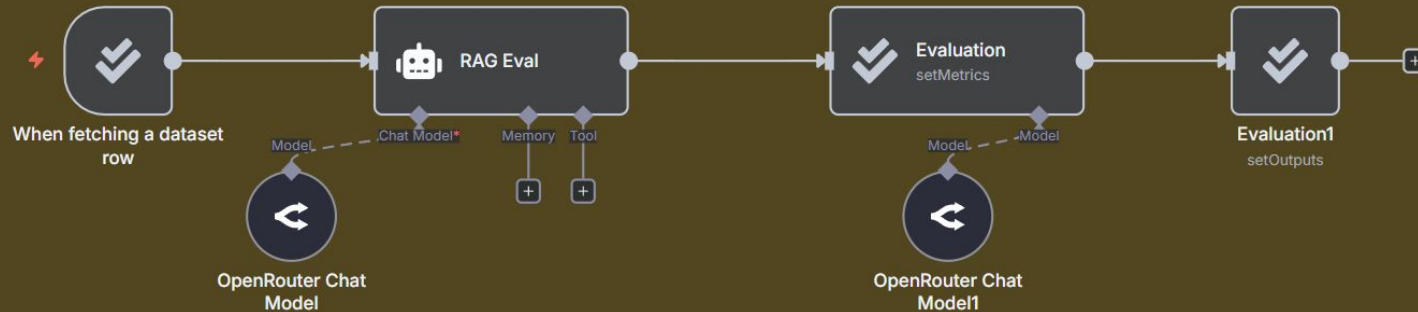
x



Today's Experiment

1. Connecting to a Sonnet 4.5 AI Agent in n8n
2. Content Creation
3. Context Window Evaluation
4. Ultimate Assistant Tool Calling

Evaluation Datasets



Sonnet 4.5 Release Overview

- **Release Date:** September 29th, 2025.
- **Availability:** Anyone can use Claude Sonnet 4.5 right now in Claude's web, iOS, and Android apps, or through the developer platform and API.
- **What is Sonnet 4.5?** Anthropic's latest model, Claude Sonnet 4.5 is an artificial intelligence model designed to help with coding, building complex AI agents, running business and research workflows, and using computers like a human assistant.

“Claude Sonnet 4.5 resets our expectations—it handles 30+ hours of autonomous coding, freeing our engineers to tackle months of complex architectural work in dramatically less time while maintaining coherence across massive codebases.” - Sean Ward, CEO of iGent AI

Claude Model Family Overview

Model	What it's for	Why Choose It
Haiku	Fast & affordable AI	Quick tasks, low cost
Sonnet 4 & 4.5	Balanced, great at coding	Reliable, agent tasks, good value
Opus	Maximum power & reasoning	Most advanced, costly, research

Model	Input Cost	Output Cost
Haiku	\$0.80/M tokens	\$4/M tokens
Sonnet 4.5	\$3/M tokens	\$15/M tokens
Opus	\$15/M tokens	\$75/M tokens

Why Sonnet 4.5 is Better

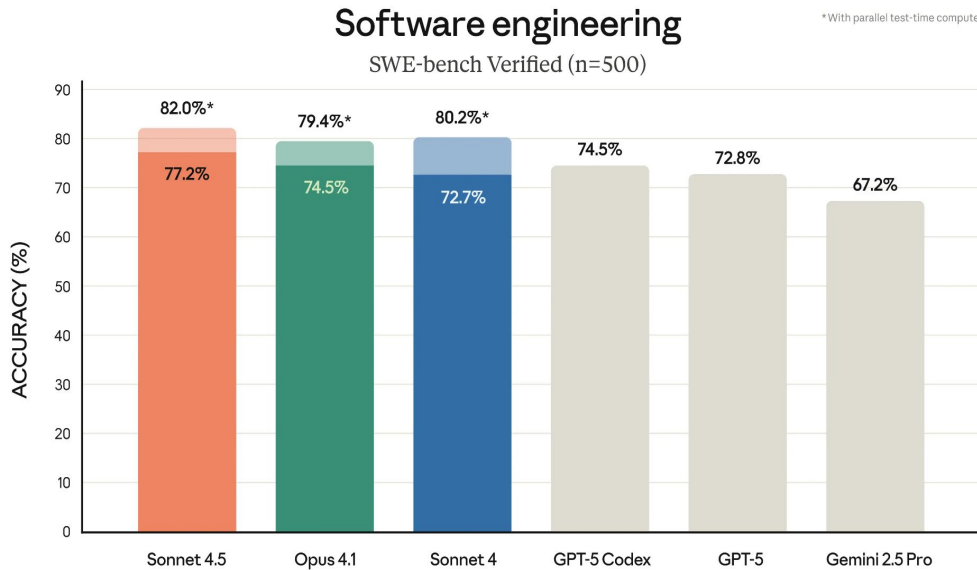
- Outperforms earlier Sonnet and Opus models in coding, long runs, and real-world agent tasks.
- Stronger memory (handles bigger projects without forgetting) and greater reliability.
- Sonnet 4.5 costs the same as Sonnet 4 but gives big upgrades for most uses.

What Makes Sonnet 4.5 So Powerful?

Sonnet 4.5 has been tested in many areas to see how smart, useful, and reliable it is.

Here's what those tests (called "benchmarks") mean, and how Sonnet 4.5 performed:

What Makes Sonnet 4.5 So Powerful?



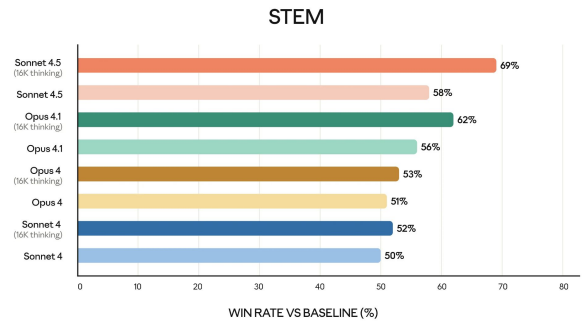
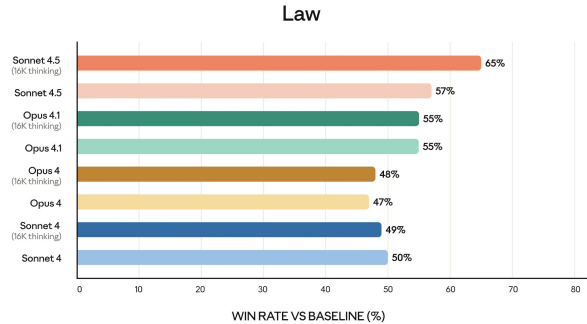
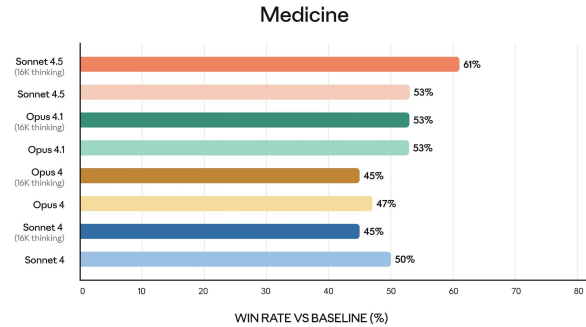
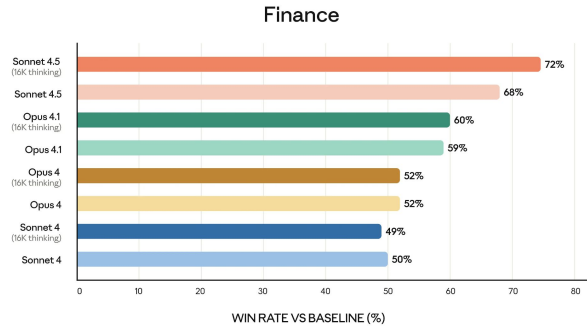
Claude Sonnet 4.5 scored an industry-leading 77–82% on the SWE-bench Verified benchmark, meaning it can automatically fix or write real-world software and handle coding tasks with the skill level of a professional developer, so even beginners can trust it to reliably solve actual programming problems for them.

What Makes Sonnet 4.5 So Powerful?

	Claude Sonnet 4.5	Claude Opus 4.1	Claude Sonnet 4	GPT-5	Gemini 2.5 Pro
Agentic coding <i>SWE-bench Verified</i>	77.2% 82.0% with parallel test-time compute	74.5% 79.4% with parallel test-time compute	72.7% 80.2% with parallel test-time compute	72.8% GPT-5 74.5% GPT-5-Codex	67.2%
Agentic terminal coding <i>Terminal-Bench</i>	50.0%	46.5%	36.4%	43.8%	25.3%
Agentic tool use <i>τ2-bench</i>	Retail 86.2%	Retail 86.8%	Retail 83.8%	Retail 81.1%	—
	Airline 70.0%	Airline 63.0%	Airline 63.0%	Airline 62.6%	—
	Telecom 98.0%	Telecom 71.5%	Telecom 49.6%	Telecom 96.7%	—
Computer use <i>OSWorld</i>	61.4%	44.4%	42.2%	—	—
High school math competition <i>AIME 2025</i>	100% (python)	78.0%	70.5%	99.6% (python)	88.0%
	87.0% (no tools)			94.6% (no tools)	
Graduate-level reasoning <i>GPQA Diamond</i>	83.4%	81.0%	76.1%	85.7%	86.4%
Multilingual Q&A <i>MMMLU</i>	89.1%	89.5%	86.5%	89.4%	—
Visual reasoning <i>MMMU (validation)</i>	77.8%	77.1%	74.4%	84.2%	82.0%
Financial analysis <i>Finance Agent</i>	55.3%	50.9%	44.5%	46.9%	29.4%

Claude Sonnet 4.5 leads all models in key tests, such as agentic coding (up to 82%), computer use (61.4%), and financial analysis, meaning it's currently the best option for automating real software jobs, handling complex computer tasks, and solving business problems, so even beginners can trust it to reliably perform tasks that would usually require skilled professionals.

What Makes Sonnet 4.5 So Powerful?



Claude Sonnet 4.5 is much better at answering complicated questions about finance, law, medicine, and science than earlier AI models, meaning it gives more accurate and expert-level advice on topics that are important in real life, making it a trustworthy helper for tough professional problems that would normally require specialists.

How Sonnet 4.5 + n8n Revolutionizes Automation

- **Smarter Automation:** Claude Sonnet 4.5's expert reasoning and coding skills let n8n automations think, fix problems, and adapt.
- **All-in-One Power:** Run advanced workflows that handle emails, files, spreadsheets, and even code.
- **Accessible AI:** n8n's easy visual builder empowers anyone to use Claude Sonnet 4.5's intelligence, no technical background needed, to create automations that actually understand, analyze, and solve real problems.
- **Industry Impact:** Automate smarter in areas like finance, security, research, and operations.

Bottom Line:

Claude Sonnet 4.5 and n8n together move automation beyond basic tasks, making advanced, reliable AI-powered workflows possible for everyone.

Today's Experiment

1. Connecting to a Sonnet 4.5 AI Agent in n8n
2. Content Creation
3. Context Window Evaluation
4. Ultimate Assistant Tool Calling

Evaluation Datasets

