

Claude 4.5

Quick Start Guide

The essential 5-page summary to start with Claude 4.5 today

What is Claude 4.5?

Claude 4.5 is Anthropic's latest AI model family, representing a significant leap forward in speed and capability. It delivers 40-50% faster performance than Claude 3.5, supports 200K to 1M token context windows, achieves world-leading benchmarks for coding tasks, and maintains the same API pricing as Claude 3.5. More power. More speed. No cost increase.

Core Features: 40-50% faster. Up to 1M token context. Leading coding performance. Same pricing.

Key Specifications Comparison

Feature	Claude 3.5 Sonnet	Claude 4.5 Sonnet	Improvement
Context Window	200K tokens	Up to 1M tokens	5x larger
Coding Performance	49% SWE-bench	80.9% SWE-bench	Major gain
Response Speed	Baseline	40-50% faster	Significant
API Pricing (input/output)	\$3 / \$15 per MTok	\$3 / \$15 per MTok	No change
Math Reasoning	Strong	Elite tier	Enhanced

API Pricing (No Change from Claude 3.5)

Input tokens: \$3 per million tokens Output tokens: \$15 per million tokens Prompt caching: 70-90% savings on repeated contexts Batch processing: 50% discount for non-urgent requests Rate limits: 100K tokens/min for standard tier

Prompt Caching for 90% Savings

Cached tokens cost \$0.30 per million instead of \$3.00. Example: A 50K-token context costs \$1.50 instead of \$15.00 per request.

Benchmarks & Performance

Coding Performance: SWE-bench 80.9% - world-leading for coding tasks
Math Reasoning (AIME 2024): 90% accuracy - elite-level problem-solving
General Knowledge: Top-tier performance across standard LLM benchmarks
Safety and Harmlessness: 99.29% safety rating - enterprise-grade security
Context Handling: Maintains accuracy and coherence across 1M-token contexts

How to Get Started (3 Easy Ways)

Web Interface (Claude.ai Pro)

Visit claude.ai - Sign up or log in - Claude 4.5 is the default model

API Integration

Update your API calls to use `model="claude-4.5-sonnet"` (same endpoints, same pricing)

Desktop Application

Download Claude desktop app - Auto-updates to latest Claude 4.5

Sample API Call (Complete & Correct)

```
curl https://api.anthropic.com/v1/messages \
-H "x-api-key: $ANTHROPIC_API_KEY" \
-H "anthropic-version: 2023-06-01" \
-H "content-type: application/json" \
-d '{
  "model": "claude-4.5-sonnet",
  "max_tokens": 1024,
  "messages": [
    {
      "role": "user",
      "content": "Explain Claude 4.5"
    }
  ]
}'
```

Next Steps and Resources

1. Explore Claude 4.5 Capabilities Start with claude.ai to understand Claude 4.5 strengths in your use case. 2. Plan Your API Integration Update your model parameter to `claude-4.5-sonnet` in your code. 3. Optimize with Caching For production workloads, implement prompt caching to save 70-90% on costs. 4. Monitor Performance Track latency and token usage improvements versus Claude 3.5.

Learn More

Official Documentation: docs.anthropic.com API Reference: <https://docs.anthropic.com/v1/reference> Interactive Console: console.anthropic.com 30-Day Claude 4.5 Guide: <https://confidencebuildings.com/claude-4-5-complete-guide>

Frequently Asked Questions

Q: Will my existing Claude 3.5 code break? A: No. Just update `model="claude-3.5-sonnet"` to `model="claude-4.5-sonnet"`. Same API.
Q: Is Claude 4.5 available in all regions? A: Check docs.anthropic.com for regional availability and rollout status.
Q: Sonnet vs Opus - Which should I choose? A: Sonnet is faster and cost-effective. Opus is optimized for complex reasoning. Choose based on your use case.

Quick Reference Card

Model Name: claude-4.5-sonnet Max Tokens: 1M context window Speed: 40-50% faster than Claude 3.5 Pricing: \$3 input / \$15 output per million tokens Availability: Web, API, Desktop app Documentation: docs.anthropic.com

When to Choose Claude 4.5

Claude 4.5 Sonnet: Fast, cost-effective, ideal for most use cases Claude 4.5 Opus: Complex reasoning, research, analysis tasks With Prompt Caching: Production systems with repeated prompts With Batch Processing: Non-urgent requests that can wait 24 hours

Key Takeaways

1. Claude 4.5 is 40-50% faster with no price increase 2. Same API as Claude 3.5 - just change the model name 3. Prompt caching saves up to 90% on repeated contexts 4. Available now via web, API, and desktop app 5. Enterprise-grade safety (99.29% harmfulness rating)

Your Action Plan

Week 1: Explore Claude 4.5 via the web interface to get familiar with its capabilities and performance. Week 2: If using the API, update your integration by changing `model="claude-3.5-sonnet"` to `model="claude-4.5-sonnet"`. Week 3: Implement prompt caching for production workloads. This single change can reduce API costs by up to 90%. Week 4: Monitor performance. Compare latency and token costs versus your previous Claude 3.5 implementation. Measure actual time and cost savings.

Final Notes

Claude 4.5 is a drop-in replacement for Claude 3.5. No breaking changes. Same pricing structure. Same API endpoints. Pure improvement: faster, more capable, same cost. This quick-start guide is your foundation. For deeper technical details, architecture discussions, and advanced use cases, visit the full 30-Day Claude 4.5 Complete Guide on ConfidenceBuildings.com. Questions? Check docs.anthropic.com or the FAQ on page 3 of this guide.