Claude 3.5 Sonnet

Claude 3.5 Sonnet sets new industry benchmarks for graduate-level reasoning (GPQA), undergraduate-level knowledge (MMLU), and coding proficiency (HumanEval). It shows marked improvement in grasping nuance, humor, and complex instructions, and is exceptional at writing high-quality content with a natural, relatable tone.

Claude 3.5 Sonnet operates at twice the speed of Claude 3 Opus. This performance boost, combined with cost-effective pricing, makes Claude 3.5 Sonnet ideal for complex tasks such as context-sensitive customer support and orchestrating multi-step workflows.

In an internal agentic coding evaluation, Claude 3.5 Sonnet solved 64% of problems, outperforming Claude 3 Opus which solved 38%. Our evaluation tests the model’s ability to fix a bug or add functionality to an open source codebase, given a natural language description of the desired improvement. When instructed and provided with the relevant tools, Claude 3.5 Sonnet can independently write, edit, and execute code with sophisticated reasoning and troubleshooting capabilities. It handles code translations with ease, making it particularly effective for updating legacy applications and migrating codebases.

