

Qwen3 Coder Next fürs Coding

Genug Infos für eine gute Einschätzung. Hier ist die Zusammenfassung:

Qwen3-Coder-Next – Kurzeinschätzung

Was es ist: Ein Open-Weight-Modell von Alibaba (Release: 3. Februar 2026), speziell für **agentisches Coding und lokale Entwicklung** gebaut – basierend auf Qwen3-Next-80B-A3B mit hybrider Attention/MoE-Architektur.

Kennzahlen

Eigenschaft	Wert
Gesamtparameter	80B (MoE)
Aktive Parameter	nur 3B
Kontextlänge	256K Tokens
Modus	nur "non-thinking" (kein <code><think></code> -Block)

Benchmark-Ergebnisse

- **SWE-Bench Verified:** ~70,6 % (mit SWE-Agent-Scaffold) — starkes Ergebnis
- **SWE-Bench Pro:** 44,3 %
- **Terminal-Bench 2.0:** 36,2 %
- Laut Reddit-Community aktuell **Top-Modell im SWE-rebench @ Pass 5**

Stärken

- **Extrem effizient:** Mit nur 3B aktiven Parametern erreicht es Leistung, die sonst 10-20x größere Modelle brauchen – läuft z. B. auf einem 64 GB MacBook lokal
- **Sehr gut bei "Fehler beheben & aus Terminal-Output iterieren"** – wichtig für echte Coding-Workflows
- **Lange Task-Ketten (long-horizon):** Skaliert gut mit mehr Agent-Turns bei komplexen Aufgaben
- Gute Tool-Calling-Fähigkeiten, kompatibel mit Claude Code, Cline, Qwen Code, Kilo, Trae, OpenClaw etc.
- Sehr günstig: ca. 0,11 \$ pro Mio. Input-Tokens auf OpenRouter

Schwächen ⚠

- **Kein "one-shot"-Löser:** Nutzer berichten, dass es selten Aufgaben sofort perfekt löst, aber in Kombination mit einem guten Agenten (z. B. Claude Code als Orchestrator) gute Ergebnisse liefert
- Bei sehr schweren Aufgaben (SWE-Bench Pro, Terminal-Bench) klar hinter den Top-Closed-Source-Modellen wie Claude Opus oder GPT-5.x
- Community-Tests zeigen: gewinnt bei **Qualität** gegen gpt-oss-120b, verliert aber bei **Geschwindigkeit**

Fazit

Für ein **lokal betreibbares, extrem effizientes** Coding-Modell ist Qwen3-Coder-Next momentan wohl **das beste Open-Source-Modell in seiner Gewichtsklasse** – es kommt bei einfachen bis mittelschweren Aufgaben nahe an Claude/GPT heran, fällt aber bei sehr komplexen, langen Multi-Step-Aufgaben klar zurück. Als Ergänzung zu einem stärkeren Cloud-Modell (z. B. für Routineaufgaben oder als lokaler "erster Entwurf") ist es sehr empfehlenswert – als vollständiger Ersatz für Claude Code/GPT bei anspruchsvollen Projekten (noch) nicht ganz auf Augenhöhe.

Revision #1

Created 2026-07-08 02:23:11 UTC by art10m

Updated 2026-07-08 02:23:37 UTC by art10m