GPT-5.6 vs. Claude: OpenAIs neue Modellfamilie im direkten Praxis- und Benchmark-Vergleich mit Anthropic

Einleitung

Mit GPT-5.6 hat OpenAI erneut eine neue Modellgeneration veröffentlicht – und diesmal nicht als einzelnes Modell, sondern als ganze Familie: **Sol**, **Terra** und **Luna**. Parallel dazu positioniert sich Anthropic mit **Claude Fable 5** und dem stark eingeschränkt verfügbaren **Claude Mythos 5** weiterhin als Maßstab für Spitzenleistung im professionellen und technischen Bereich. Dieser Artikel verbindet die offiziellen Benchmark-Daten beider Anbieter mit einem ausführlichen Praxistest – Seite an Seite, im echten Einsatz mit Coding-Agents, kreativen Web-Projekten und alltäglichen API-Anfragen.

https://youtu.be/EthxaDswUFo

---

1. Die neue GPT-5.6-Familie

OpenAI hat mit GPT-5.6 ein neues Namensschema eingeführt. Statt eines einzigen Flaggschiffs gibt es nun drei „dauerhafte Fähigkeitsstufen“, die sich künftig unabhängig weiterentwickeln sollen:

Modell	Rolle	Richtpreis (API, pro 1M Token)
Sol	Flaggschiff, höchste Kapazität	5 $ Input / 30 $ Output

Modell	Rolle	Richtpreis (API, pro 1M Token)
Terra	Ausgewogenes Modell für den Alltag	2,50 $ Input / 15 $ Output
Luna	Schnellstes und günstigstes Modell	1 $ Input / 6 $ Output

Zusätzlich gibt es neue Reasoning-Modi: **max** (mehr Denkzeit als das bisherige *xhigh*) und **ultra**, bei dem standardmäßig vier Agenten parallel koordiniert werden. Ein zentrales Verkaufsargument ist die **Token-Effizienz**: OpenAI betont, dass GPT-5.6 „mehr nützliche Arbeit aus jedem Token“ herausholt – ein Punkt, der sich im Praxistest eindrucksvoll bestätigt (siehe unten).

2. Anthropics Modelle: Claude Fable 5 und Claude Mythos 5

Anthropic hatte bereits im Juni 2026 seine fünfte Generation vorgestellt: **Claude Fable 5** als breit verfügbares Spitzenmodell und **Claude Mythos 5** als noch leistungsfähigere, aber stark zugriffsbeschränkte Variante – primär für vertrauenswürdige Partner in Cybersicherheit und Biowissenschaften. Bemerkenswert ist, dass Mythos 5 zwischenzeitlich wegen US-Exportkontrollen ausgesetzt und erst Ende Juni wieder freigegeben wurde – ein Vorgeschmack darauf, wie stark Regulierung künftig die Verfügbarkeit von Frontier-Modellen beeinflussen kann.

3. Benchmarks im Überblick

Kategorie	Ergebnis
Agents' Last Exam	Sol übertrifft Fable 5 deutlich (52,7 % vs. 40,5 %)
Artificial Analysis Intelligence Index	Fable 5 knapp vorn (59,9 vs. 58,9)
Coding Agent Index	Sol vorn (80 vs. 77,2)
SWE-Bench Pro	Claude klar vorn (80 %+ vs. 64,6 %)
Terminal-Bench 2.1	Sol vorn, besonders mit Ultra-Modus
Cybersicherheit (ExploitBench etc.)	Mythos 5 knapp vorn, Abstand schrumpft
Long-Context/Tool-Use (Toolathlon)	Claude leicht vorn

Fazit der reinen Zahlen: Kein klarer Gewinner, sondern ein differenziertes Bild – **Sol punktet vor allem bei Effizienz, Geschwindigkeit und Kosten**, während **Claude bei tiefem, technischem Software-Engineering und langen Tool-Ketten weiterhin die Nase vorn hat**.

---

4. Der Praxistest: Sol vs. Fable im echten Einsatz

Benchmarks sind das eine – aber wie schlagen sich die Modelle im tatsächlichen Alltag? Ein ausführlicher Hands-on-Vergleich (Cloud Code mit Claude Fable 5 vs. Codex mit GPT-5.6 Sol) liefert aufschlussreiche Ergebnisse.

Test 1: Browser-Bike-Game

Beide Modelle sollten ein spielbares Open-World-Bike-Game bauen. Ergebnis: **Fable 5 gewinnt klar** – deutlich immersiveres, größeres und besser spielbares Ergebnis mit GTA-ähnlichem Gefühl.

- **Zeit:** Fable 21:37 Min. vs. Sol 23:00 Min. (fast gleich)
- **Kosten:** Fable ≈ 14,22 $ vs. Sol ≈ 4,50 $
- **Output-Tokens:** Fable ≈ 90.000 vs. Sol ≈ 31.000

Trotz klar besserem Ergebnis von Fable ist Sol rund **3x günstiger** und deutlich token-effizienter.

Test 2: „Scroll-Stopping“ Interactive Website

Beide Modelle sollten eine möglichst beeindruckende, interaktive Website mit Sound bauen. Auch hier: **Fable liefert den größeren „Wow-Faktor“**, mit mehr Immersion und 3D-Interaktivität.

- **Zeit:** Fable 23 Min. vs. Sol nur 7 Min.
- **Kosten:** Fable 19,24 $ (80.000 Output-Tokens) vs. Sol nur ca. 1 $ (20.000 Output-Tokens)
- Das entspricht **fast dem 20-fachen Preis** für Fable – bei nicht 20-fach besserer Qualität.

Test 3: Fünf unterschiedliche visuelle Elemente (offener Prompt)

Bei völlig freier kreativer Aufgabenstellung dreht sich das Bild: **Sol wird als Gesamtsieger empfunden**, mit mehr Vielfalt und Experimentierfreude (u. a. Simulationen, generative Kunst), während Fables Outputs als weniger beeindruckend beschrieben werden – trotz einzelner Bugs bei beiden Modellen.

- **Zeit:** Fable 15 Min. vs. Sol 7 Min.
- **Kosten:** Fable ≈ 15 $ (65.000 Output-Tokens) vs. Sol ≈ 1 $ (22.000 Output-Tokens)

Test 4: Schnelle One-Shot-API-Anfragen (kein Agentic Loop)

Bei stateless Einzelanfragen über die API zeigt sich ein deutlicher Unterschied:

- **Score:** Sol gewinnt **24 von 27 Duellen**, Fable gewinnt 3.
- Ein wichtiger Grund: **Fable verweigerte häufig die Antwort** (starke Guardrails), was die Gesamtwertung zugunsten von Sol verzerrt.
- **Score bei tatsächlicher Antwort:** Sol 0,98 vs. Fable 0,966 – **fast identisch**.
- **Kosten:** Sol ≈ 16 $ vs. Fable ≈ 63 $ – hier ist die Kostendifferenz **relativ kleiner** als bei den agentischen Coding-Aufgaben.
- **Latenz:** Median-Latenz von Sol niedriger, aber Durchschnitts-Latenz von Fable niedriger – Sol ist „volatiler“ (mal sehr schnell, mal sehr langsam), Fable konsistenter.

---

5. Persönliches Fazit aus der Praxis: Manager vs. Worker

Nach einem vollen Tag paralleler Nutzung beider Modelle im echten Arbeitsalltag lässt sich folgendes Bild zeichnen:

> **Claude Fable 5 ist der bessere „Manager“ bzw. Co-Founder. GPT-5.6 Sol ist der bessere „Worker“.**

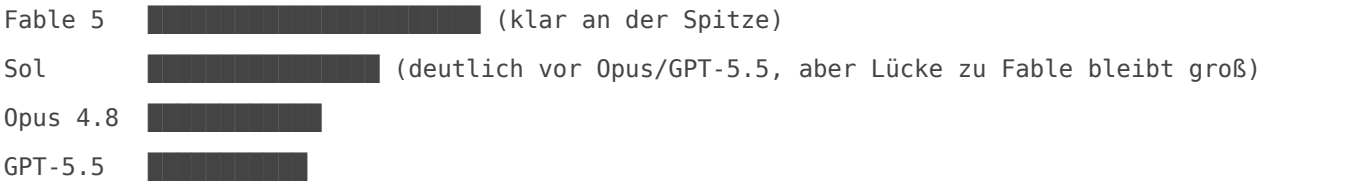
Stärken von Sol:

- Deutlich günstiger (teils 3x bis 20x)
- Bessere Computer-Use-Fähigkeiten
- Besser im „Devil's Advocate“ – findet Bugs, denkt kritisch nach, hält sich strikt an Anweisungen
- Schneller in agentischen Workflows
- Sehr token-effizient (schreibt aber auch viele Tests – interessanter Widerspruch)

Stärken von Fable 5:

- Deutlich kreativer, besser im Schreiben, Brainstorming und strategischen Beraten
- Neigt weniger zum „Overengineering“
- Wirkt „sassier“, gibt eigenständiges Feedback statt Anweisungen blind auszuführen
- Bei reiner Kapabilität „eine Klasse für sich“ – wird als klar über GPT-5.5 und eher auf Opus-4.8-Niveau von Sol eingeordnet

Positionierung der vier Top-Modelle (subjektive Einschätzung)



Ein Kritikpunkt: Warum wird das leistungsfähigste OpenAI-Modell nur „5.6“ genannt statt eines vollen Versionssprungs auf „6“? Die Vermutung: OpenAI hält sein nächstes echtes Next-Gen-Modell (GPT-6) noch zurück – das wäre dann der eigentliche direkte Herausforderer für Fable 5.

---

6. Strategische Einordnung

OpenAIs Botschaft: Nicht „wir sind überall Nummer eins“, sondern „vergleichbare oder bessere Ergebnisse zu deutlich geringeren Kosten und in kürzerer Zeit“. Kundenstimmen von Notion, Cursor, Cognition, Shopify, Microsoft, Canva und Figma unterstreichen das.

Anthropics Botschaft: Spitzenleistung in Nischen (Cyberabwehr, Life Sciences) mit Claude Mythos 5, breiter zugängliche Exzellenz mit Fable 5 – bei strengerem, stärker reguliertem Zugriff.

Sicherheitsdruck auf beiden Seiten: OpenAI investierte ca. 700.000 A100e-GPU-Stunden in automatisiertes Red-Teaming und führte ein „Trusted Access“-Programm mit Hardware-Passkey-Pflicht ein. Anthropic musste Mythos 5 zeitweise wegen Exportkontrollen offline nehmen.

---

7. Gesamtfazit

GPT-5.6 ist weniger ein radikaler Fähigkeitssprung als eine **Neuausrichtung auf Effizienz, Geschwindigkeit und Preis-Leistung**. Im Alltag zeigt sich: Bei kreativen, offenen und ambitionierten Aufgaben liefert Claude Fable 5 nach wie vor die beeindruckenderen Ergebnisse –

aber zu einem Preis, der oft um ein Vielfaches höher liegt und mit strengeren Guardrails einhergeht, die in manchen Fällen sogar zur kompletten Antwortverweigerung führen.

GPT-5.6 Sol punktet dagegen als extrem kosteneffizienter, schneller und zuverlässiger „Ausführer“ – ideal für agentische Workflows, Verifikation, Debugging und High-Volume-API-Einsätze.

Die Wahl zwischen den Modellen ist 2026 keine Frage von „welches Modell ist besser“, sondern von Anwendungsfall, Budget und Rolle im Workflow. Wer strategisch denkt, kreativ brainstormt oder komplexe Software-Architekturen entwerfen will, ist mit Fable 5 (aktuell) besser bedient. Wer in großem Maßstab, kosteneffizient und schnell Aufgaben ausführen, testen und verifizieren lassen will, profitiert von Sol. Ein spannender Blick in die Zukunft: Ein Setup, in dem **Fable als Orchestrator eine Flotte von Sol-Agenten steuert**, könnte das Beste aus beiden Welten vereinen.

Revision #1
Created 2026-07-10 20:20:35 UTC by art10m
Updated 2026-07-10 20:37:25 UTC by art10m