

**Claude Sonnet 4とGPT-5のコーディング性能評価の妥当性**

**現状の主なベンチマーク結果**

* Claude Sonnet 4のSWE-bench Verified（GitHubの実課題修正ベンチマーク）スコアは**約72.7%**。[[1]](#fn1)[[2]](#fn2)[[3]](#fn3)
* GPT-5の同ベンチマークスコアは\*\*74.9%\*\*であり、Sonnet 4より若干高い数字。[[3]](#fn3)[[4]](#fn4)[[5]](#fn5)
* Aider Polyglot（多言語コード編集ベンチ）でもGPT-5は88%と最先端水準。[[4]](#fn4)[[5]](#fn5)

**実務的な観点と補足**

* Claude Sonnet 4は速度・直接性・コスト効率に優れ、単純な修正や高速なイテレーションには強み。[[6]](#fn6)
* GPT-5は精度重視、複雑なリファクタやクロスファイル依存解決、大規模案件などでより信頼される傾向がある。[[4]](#fn4)[[6]](#fn6)
* 実務ユーザーの体感評価は「好み」や「用途」により分かれ、どちらか一方が常に上回るとは言えない。[[6]](#fn6)

**結論**

* 現行の客観的ベンチマーク（SWE-benchなど）では、**GPT-5の方がSonnet 4よりコーディング性能が僅かに高い**のが事実です。[[5]](#fn5)[[3]](#fn3)[[4]](#fn4)[[6]](#fn6)
* Sonnet 4の方が性能が高いという評価は現状の数値・業界比較からは**妥当ではありません**。
* ただし、速度重視・コスト重視・用途特化ではSonnet 4が選ばれる場面も多く、性能の「質」や「ターゲット」によって最適モデルを使い分けるべきです。[[3]](#fn3)[[6]](#fn6)

**参考：主要ベンチマーク値比較表**

|  |  |  |  |
| --- | --- | --- | --- |
| モデル | SWE-bench Verified | Aider Polyglot | 特記事項 |
| Claude Sonnet 4 | 約72.7%[[2]](#fn2)[[1]](#fn1)[[3]](#fn3) | 約77%[[7]](#fn7) | 高速・コスト効率 |
| GPT-5 | 74.9%[[4]](#fn4)[[5]](#fn5)[[3]](#fn3) | 88%[[4]](#fn4)[[5]](#fn5) | 精度・クロスファイル・大規模対応 |

**妥当性まとめ**  
Sonnet 4がGPT-5よりコーディング性能で高評価されるのは、現状の客観的ベンチマークおよび業界評価の観点から**妥当ではありません**。どちらもトップ性能だが、僅差で**GPT-5優位**です。用途別・ユーザー好みで使い分けるのが最適解です。

⁂

1. <https://zenn.dev/saan/articles/265d063eb90f31>

1. <https://blog.scuti.jp/claude-4-advanced-coding-reasoning-overview/>

1. <https://blog.getbind.co/2025/08/04/openai-gpt-5-vs-claude-4-feature-comparison/>

1. <https://openai.com/index/introducing-gpt-5-for-developers/>

1. <https://www.vellum.ai/blog/gpt-5-benchmarks>

1. <https://www.augmentcode.com/blog/gpt-5-is-here-and-we-now-have-a-model-picker>

1. <https://www.qt.io/ja-jp/blog/improving-qml-coding-by-up-to-11-with-sonnet-4-qt-ai-assistant-0.9.3-released>