



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][2][3]}
- 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][4][5]}
- 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][5]}

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]
- 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][6]}
- 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]

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][3][4][6]}
- 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][6]}

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][1][3]}	約 77% ^[7]	高速・コスト効率
GPT-5	74.9% ^{[4][5][3]}	88% ^{[4][5]}	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>
2. <https://blog.scuti.jp/claude-4-advanced-coding-reasoning-overview/>
3. <https://blog.getbind.co/2025/08/04/openai-gpt-5-vs-claude-4-feature-comparison/>
4. <https://openai.com/index/introducing-gpt-5-for-developers/>
5. <https://www.vellum.ai/blog/gpt-5-benchmarks>
6. <https://www.augmentcode.com/blog/gpt-5-is-here-and-we-now-have-a-model-picker>
7. <https://www.qt.io/ja-jp/blog/improving-qml-coding-by-up-to-11-with-sonnet-4-qt-ai-assistant-0.9.3-released>