

# 消費者向け生成AIのコーディング利用:コストと能力の比較分析レポート(2025年版)

## 1. エグゼクティブサマリー

本レポートは、ソフトウェア開発およびコーディングタスクに利用可能な消費者向け生成AIツールについて、その費用、機能、およびコーディング能力を比較分析することを目的としています。開発者や技術管理者が、コストと性能のバランスを考慮して最適なAIサービスを選択するための判断材料を提供します。

近年のAI技術、特に生成AIの進化は目覚ましく、コーディング支援ツールの市場も急速に変化しています<sup>1</sup>。Claude Pro/Max、ChatGPT Plus/Pro、Gemini Advanced、Perplexity Pro、GitHub Copilot、Cursor、WindSurf、Devin、Clineなど、多様な選択肢が登場し、競争が激化しています<sup>2</sup>。これらのツールは、コード補完やチャットによる支援といった基本的な機能に加え、より自律的なタスク実行を目指す「エージェント機能」など、機能面での収斂も見られます<sup>4</sup>。

コーディング能力に関しては、最新のベンチマーク評価、特に実世界のGitHub Issue解決能力を測るSWE-benchにおいて、特定のAIエージェント(例:SWE-agent + Claude 3.7 Sonnet)が高い性能を示しています<sup>9</sup>。これは、特定のモデル(Claude 3.7 Sonnetなど)が複雑なコーディングタスクにおいて優れた能力を持つことを示唆しています。

コスト面では、市場は二極化の傾向にあります。月額\$10~\$20程度の比較的手頃な価格帯のサブスクリプションプランが広く提供されている一方で<sup>13</sup>、より高い性能や使用量、特に自律的なエージェント機能を利用するには、月額\$100を超える高額なプランや、使用量に応じて費用が変動するAPI利用、あるいは独自のクレジットシステム(例:DevinのACU)が必要となるケースが多く見られます<sup>14</sup>。このため、利用者はアクセスしやすい標準的な機能を選ぶか、より高性能だが高コストあるいはコスト予測が難しい最先端の機能を選ぶか、という選択を迫られることになります。

機能面では、基本的なコード補完やチャット機能は多くのツールで提供されるようになり、差別化要因はむしろ、これらの機能がどのように実装されているか(例:エージェントワークフローの洗練度、IDE統合の深さ、利用可能なモデルの選択肢)や、複雑なタスクにおける実際の性能へと移行しています<sup>4</sup>。

本レポートでは、これらの点を踏まえ、各サービスの詳細なコスト、機能、そして最新のコーディング能力評価を比較し、利用者のニーズ(コスト重視、性能重視、特定の機能要件など)に応じた推奨事項を提示します。

## 2. サブスクリプションベースAIアシスタント:価格と機能

主要なサブスクリプション型AIコーディング支援サービスについて、月額・年額料金と提供される主要機能を詳述します。

## 2.1. Claude Pro / Max

- **Claude Pro:**

- 料金: 月額\$20。年払いの場合は月額\$17相当(\$200一括払い)<sup>14</sup>。
- 機能: 無料プランより多い使用量、複数のClaudeモデルへのアクセス(最新のClaude 3.7 Sonnet等)、チャットやドキュメントを整理する「プロジェクト」機能、複雑な作業のための「拡張思考(Extended Thinking)」機能を提供します<sup>10</sup>。
- コーディング関連: 最新モデル(Claude 3.7 Sonnet)はコーディング能力が高いと評価されており<sup>10</sup>、「拡張思考」はコーディングや数学の性能を向上させます<sup>10</sup>。

- **Claude Max:**

- 料金: 月額\$100から。Proの5倍の使用量を提供する「Expanded Usage」(\$100/月)と、Proの20倍の使用量を提供する「Maximum Flexibility」(\$200/月)の2レベルがあります<sup>14</sup>。
- 機能: Proの全機能に加え、大幅に増加した使用量、より高い出力制限、研究機能へのアクセス、高度なClaude機能への早期アクセス、高トラフィック時の優先アクセスなどが含まれます<sup>14</sup>。大量のドキュメントや複雑なデータを扱う、頻繁にClaudeを利用するユーザー向けです<sup>34</sup>。
- コーディング関連: Proと同様に最新モデルへのアクセスが可能で、使用量制限が大幅に緩和されるため、大規模なコーディングプロジェクトや長時間の対話に適しています。

## 2.2. ChatGPT (Plus / Pro)

- **ChatGPT Plus:**

- 料金: 月額\$20<sup>15</sup>。
- 機能: 無料プランよりも高度なモデル(GPT-4o等)へのアクセス、応答速度の向上、高可用性、DALL・Eによる画像生成、ウェブブラウジング、データ分析機能、カスタムGPTの作成・利用、音声モード、ファイルアップロードなどが含まれます。使用量制限は無料版より緩和されています<sup>21</sup>。
- コーディング関連: GPT-4シリーズやoシリーズといった高性能モデルへのアクセスが可能です。カスタムGPT機能を使えば、特定のコーディングタスクに特化したGPTを作成できます。

- **ChatGPT Pro:**

- 料金: 月額\$200<sup>15</sup>。
- 機能: Plusの全機能に加え、推論モデル(GPT-4o、o1 proモード - 米国のみ)への無制限アクセス、音声モード(オーディオのみ)の無制限アクセス、音声でのビデオ・画面共有の上限緩和、Soraビデオ生成やDeep Researchへの拡張アクセス、Operatorリサーチプレビュー(米国のみ)へのアクセスなどが含まれます<sup>21</sup>。特に「o1

proモード」は、困難な数学、科学、コーディング問題に取り組むために、より多くの計算リソースを使用するように設計されています<sup>21</sup>。

- コーディング関連: 「o1 proモード」が明示的にコーディング問題への対応を強化している点が特徴です<sup>21</sup>。

### 2.3. Gemini Advanced (Google One AI Premium)

- 料金: 月額\$19.99(無料トライアル後)。Google Oneの特典として2TBのストレージ(Googleフォト、ドライブ、Gmail共通)が含まれます<sup>16</sup>。学生向けには無料アクセスプランも提供されています<sup>16</sup>。
- 機能: Googleの最も高性能なAIモデル「Gemini Advanced」へのアクセス、Gmail、Docs、Slides、MeetなどGoogle Workspaceアプリとの深い統合、NotebookLM Plus、Deep Research機能、最大100万トークンのコンテキストウィンドウ、Veo 2ビデオ生成へのアクセスなどが提供されます<sup>16</sup>。
- コーディング関連: Geminiモデルはコーディング能力も有しており<sup>7</sup>、Google Cloud PlatformやWorkspaceを多用する開発者にとっては、エコシステムとの連携がメリットとなる可能性があります。

### 2.4. Perplexity Pro

- 料金: 月額\$20<sup>17</sup>。
- 機能: 1日あたり300回以上の「Pro Search」(無料版は制限あり)、ファイル(PDF、CSV等)や画像のアップロードと分析機能、Claude 3.7 Sonnet、GPT-4o、Sonar Large、Llama 3.1など複数のAIモデルから選択して利用可能、Discordコミュニティへのアクセスなどが提供されます<sup>17</sup>。APIアクセスも別途提供されています<sup>40</sup>。
- コーディング関連: コーディングに強いとされるClaudeやGPT-4系のモデルを選択できる点が強みです<sup>17</sup>。ファイルアップロード機能は、コードファイルや技術ドキュメントの分析に活用できます。レビューでもコーディング支援ツールとして評価されています<sup>39</sup>。

### 2.5. GitHub Copilot (Pro / Pro+)

- **GitHub Copilot Pro:**
  - 料金: 月額\$10、または年額\$100<sup>13</sup>。認証済みの学生、教師、人気OSSメンテナーは無料<sup>42</sup>。
  - 機能: IDE(VS Code, Visual Studio, JetBrains等)内でのベースモデルによる無制限のコード補完とチャット機能。月間300回の「プレミアムリクエスト」が含まれ、これを使用してClaude Sonnet、Gemini、OpenAI o3-miniなどの追加モデルにアクセス可能。超過分は\$0.04/リクエストで利用可能(オプトイン必要)<sup>5</sup>。
  - コーディング関連: IDEとの深い統合、文脈に応じたコード補完、チャットによる質問応答・デバッグ支援、セマンティックインデックスによるコードベース理解、特定のコーディングタスクのためのガイドやプロンプト例が提供されます<sup>7</sup>。
- **GitHub Copilot Pro+:**

- 料金: 月額\$39、または年額\$390<sup>13</sup>。
- 機能: Copilot Proの全機能に加え、月間1,500回のプレミアムリクエスト、利用可能な全てのプレミアムモデル(GPT-4.5等も含まれる可能性あり)へのフルアクセス、高度なAI機能(例: Copilot Code Reviewエージェント)への優先アクセスが含まれます<sup>5</sup>。
- コーディング関連: Proよりも多くのプレミアムリクエストと、より高性能なモデルへのアクセスにより、複雑なコーディングタスクへの対応能力が向上します。コードレビュー機能(プレミアム機能)も利用可能です<sup>43</sup>。

## 2.6. Cursor Pro

- 料金: 月額\$20(ビジネスプランは\$40/ユーザー/月)<sup>18</sup>。
- 機能: AI機能を深く統合したVS Codeのフォーク(派生版)IDE。無制限のコード補完、月間500回の「高速」プレミアムリクエスト(GPT-4, GPT-4o, Claude 3.5/3.7 Sonnet等を使用)、無制限の「低速」プレミアムリクエストを提供。エージェントモード、コードベース全体を理解する機能(Context Finder)、コマンド自動実行、エラー発生時の自動修正試行(Loops on Errors)、高度なコンテキスト管理(.cursorsrulesファイル)、コミットメッセージ自動生成などの独自機能が豊富です<sup>6</sup>。VS Code拡張機能も利用可能です<sup>46</sup>。
- コーディング関連: 全てがコーディング支援のために設計されています。自律的なタスク実行が可能なエージェントモード、ユーザーから高く評価される高速な自動補完機能<sup>4</sup>、プレミアムコーディングモデルとの直接連携が特徴です。

## 2.7. WindSurf (Pro / Pro Ultimate)

- 料金: 無料プランあり。Proプランは月額\$15、Pro Ultimateプランは月額\$60(チームプランも別途存在)<sup>19</sup>。Proプランは月間500 User Promptクレジット + 1500 Flow Actionクレジット、Pro Ultimateは無制限 User Prompt + 3000 Flow Actionクレジットを提供。追加クレジット購入可能(\$10で300/400クレジット)<sup>19</sup>。
- 機能: エージェント型コードエディタ(VS Codeフォーク)。「Cascade」インターフェース(Write, Chat, Legacyモード)、「Flows」(リアルタイムAIコラボレーション)、Supercomplete(高度な補完)、Inline AI(Ctrl+I)、AIターミナル、画像からのコード生成、ローカル/リモートインデックスによるコードベース理解、高度なコンテキスト認識、クレジットを使用して様々なAIモデル(GPT-4o, Claude Sonnet, DeepSeek R1, o3-mini, Gemini 2.0 Flash等)にアクセス可能。JetBrains IDEもサポート(Pro以上)<sup>19</sup>。
- コーディング関連: エージェント機能に重点を置き、開発者のフロー状態を維持しながらのコーディングを支援します。自動的なデバッグやコード実行の反復ループ、強力なコンテキスト処理能力が特徴です。

## 2.8. サブスクリプションAIアシスタント比較表

以下の表は、主要なサブスクリプションベースのAIアシスタントの価格と主要なコーディング関

連機能をまとめたものです。

サービス (プラン)	月額料金 (USD)	年額料金 (USD)	主要コーディング機能	アクセス可能モデル例	プレミアムリクエスト/クレジットシステム
Claude Pro	\$20	\$200	最新モデルアクセス (3.7 Sonnet等), 拡張思考, プロジェクト機能	Claude 3.7 Sonnet, 他	なし (使用量制限あり)
Claude Max (\$100)	\$100	N/A	Pro全機能 + 5倍の使用量, 高出力制限, 早期アクセス	Claude 3.7 Sonnet, 他	なし (使用量制限大幅緩和)
Claude Max (\$200)	\$200	N/A	Pro全機能 + 20倍の使用量, 最大限の柔軟性	Claude 3.7 Sonnet, 他	なし (使用量制限最大限緩和)
ChatGPT Plus	\$20	N/A	高度モデルアクセス (GPT-4o等), データ分析, カスタムGPTs	GPT-4o, GPT-4等	なし (使用量制限あり)
ChatGPT Pro	\$200	N/A	Plus全機能 + 無制限推論モデルアクセス (o1 pro等), 高度機能上限緩和	GPT-4o, o1 pro, 他	なし (推論モデル等無制限)
Gemini Advanced	\$19.99	N/A	Gemini Advancedモデルアクセス, Google Workspace 統合, 2TBストレージ, Deep Research	Gemini Advanced	なし

Perplexity Pro	\$20	N/A	モデル選択 (Claude, GPT-4o等), ファイル分析, Pro Search (300+/日)	Claude 3.7 Sonnet, GPT-4o, Sonar Large, Llama 3.1等	なし (Pro Search回数制限)
GitHub Copilot Pro	\$10	\$100	IDE統合, 無制限ベースモデル補完/チャット, プレミアムリクエスト (300/月)	ベースモデル + プレミアム (Claude Sonnet, Gemini, o3-mini等)	300 プレミアムリクエスト/月 (超過 \$0.04/リクエスト)
GitHub Copilot Pro+	\$39	\$390	Pro全機能 + プレミアムリクエスト (1500/月), 全プレミアムモデルアクセス, コードレビュー機能	ベースモデル + 全プレミアム (GPT-4.5等含む可能性)	1500 プレミアムリクエスト/月 (超過 \$0.04/リクエスト)
Cursor Pro	\$20	N/A	AI統合IDE (VSCode フォーク), 無制限補完, 高速プレミアムリクエスト (500/月), 無制限低速リクエスト, エージェントモード, 高度コンテキスト管理	GPT-4o, Claude 3.7 Sonnet等	500 高速プレミアムリクエスト/月 (低速は無制限)
WindSurf Pro	\$15	N/A	エージェント型 IDE (VSCode フォーク), Cascade/Flows, Supercomplete, AIターミナル, プレミアムクレジット (500 User	GPT-4o, Claude Sonnet, DeepSeek R1, o3-mini等 (クレジット消費)	500 User Prompt + 1500 Flow Action クレジット/月 (追加購入 \$10/300)



			Prompt + 1500 Flow Action)		
WindSurf Pro Ultimate	\$60	N/A	Pro全機能 + 無制限User Promptクレジット, プレミアムクレジット (3000 Flow Action), 優先サポート	GPT-4o, Claude Sonnet, DeepSeek R1, o3-mini等 (クレジット消費)	無制限 User Prompt + 3000 Flow Action クレジット/月 (追加購入 \$10/400)

注: 年額料金は提供されている場合のみ記載。機能やモデルは変更される可能性があります。最新情報は公式サイトをご確認ください。

この比較から、\$10～\$20の価格帯が標準的な「Pro」層として激戦区となっていることがわかります。この価格帯で競争するためには、単に高性能なモデルへのアクセスを提供するだけでなく、独自の機能や体験を提供することが不可欠です。例えば、CursorはAI機能を深く統合したIDE自体を提供し<sup>6</sup>、GitHub CopilotはGitHubエコシステムとの連携を強みとしています<sup>5</sup>。Perplexityはモデル選択の自由度<sup>17</sup>、WindSurfは独自のエージェント型ワークフロー<sup>27</sup>を提供することで差別化を図っています。

また、「プレミアムリクエスト」や「クレジット」といった仕組みが、固定料金プラン内でより高性能（かつ高コスト）なモデルへのアクセスを提供するための一般的な方法として普及しています<sup>13</sup>。これにより、ベンダーは基本料金を抑えつつ最新モデルへのアクセスを提供できますが、利用者にとっては使用量やコストの予測が難しくなるという側面もあります<sup>4</sup>。特にクレジット消費量がタスクによって変動する場合や、意図せず上限を超えて追加料金が発生する可能性があるため、注意深い利用管理が求められます<sup>4</sup>。

### 3. 新興および特化型AIコーディングツール

標準的なサブスクリプションモデルとは異なる、あるいは比較的新しい、特化型のAIコーディングツールについて詳述します。

#### 3.1. Devin AI

- 位置づけとアクセス: 「自律型AIソフトウェアエンジニア」として市場に投入されました<sup>53</sup>。当初は待機リストやエンタープライズ経由でのアクセスが主でしたが<sup>53</sup>、現在はチーム向けに app.devin.ai で一般提供されています<sup>53</sup>。個人利用向けの待機リストも存在します<sup>56</sup>。
- 料金体系:

- **Teamプラン:** 月額\$500。250 ACU (Agent Compute Unit) が含まれ、1 ACUあたり\$2.00相当<sup>22</sup>。
- **Coreプラン**(従量課金): 初期費用\$20で約9 ACU (\$2.25/ACU) が付与され、その後は使用したACUに応じて課金されるPay-as-you-goモデル<sup>22</sup>。1 ACUは約15分間のアクティブなDevin作業に相当するとされています<sup>22</sup>。
- **Enterpriseプラン:** カスタム価格。最も高性能な「Devin Enterprise」へのアクセス、特定のユースケースに特化したカスタムDevin、VPCデプロイメント、SSOなどが含まれます<sup>25</sup>。
- **コスト効率:** エンジニアの時間コストと比較して大幅なコスト削減効果があるとの事例報告もあります<sup>25</sup>。
- **能力:** コーディング、デバッグ、テスト、デプロイメントなどのタスクを自律的に完了できます。サンドボックス化された環境内でシェル、コードエディタ、ブラウザを使用し、関連コンテキストの記憶、時間経過による学習、間違いの修正が可能です<sup>53</sup>。Slack、Linear、Jira、API経由での連携も可能です<sup>53</sup>。小～中規模タスク、コード移行、反復的なエンジニアリングタスク(PRレビュー、テスト作成等)を得意とします<sup>53</sup>。SWE-benchベンチマークでは、非支援設定(25%サブセット、45分制限)で13.86%の問題を解決したと主張しています<sup>54</sup>。
- **制限事項:** 大規模で複雑なタスクには苦勞する傾向があり、タスクの分割が推奨されます<sup>53</sup>。信頼性の問題も指摘されており、時折意図しない方向に進むことがあります<sup>53</sup>。UIの美的感覚の生成は苦手で、モバイルアプリのテストはできません<sup>53</sup>。セキュリティに関しても注意が必要です<sup>53</sup>。初期のデモンストレーションの正確性については懐疑的な見方や批判も存在します<sup>54</sup>。現状ではジュニアレベルのエンジニアが担当するような複雑度のタスクに適しているとされています<sup>53</sup>。

### 3.2. WindSurf AI (エージェント機能)

(セクション2で概要を説明しましたが、ここではエージェントとしての側面に焦点を当てます)

- **エージェント機能:** 「エージェント型コードエディタ」というコンセプトを前面に出しています<sup>27</sup>。自律性のレベルを変えられる「Cascade」モード(Write, Chat, Legacy)<sup>27</sup> や、ユーザーの作業に適応しながらリアルタイムで協調する「Flows」機能<sup>52</sup> が特徴です。コードの自動実行、反復的なデバッグ、複数ファイルにまたがる変更の処理などが可能です<sup>27</sup>。ローカルインデックス機能により、コードベース全体を深く理解します<sup>52</sup>。

### 3.3. Cline AI

- **性質:** Visual Studio Code (VS Code) 向けのオープンソースAIコーディングアシスタントであり、協調的なエージェントとして機能します<sup>7</sup>。
- **機能:** 「Plan」(計画)モードと「Act」(実行)モードを切り替えて使用できます。ファイルの作成・編集(差分表示付き)、ターミナルコマンドの実行(VS Codeのシェル統合を利用)、ブラウザ操作(Claude 3.7 Sonnet経由)、Model Context Protocol (MCP) を介したカスタ



ムツール(外部API、データベース等)連携が可能です<sup>8</sup>。OpenRouter、Anthropic、OpenAI、Gemini、AWS Bedrock、Azure、GCP Vertexなど多様なAPIプロバイダー、またはLM Studio/Ollama経由でのローカルモデル利用に対応しています<sup>28</sup>。トークン使用量とAPIコストを追跡・表示する機能も備えています<sup>28</sup>。

- **コスト:** ツール自体はオープンソースで無料ですが、利用するバックエンドの大規模言語モデル(LLM)のAPIプロバイダーに応じた費用が発生します<sup>28</sup>。ユーザーは自身のAPIキーを設定する必要があります<sup>28</sup>。

これらのツールの登場は、「AIコーディングアシスタント」の定義が、単なる補完ツールから自律的なエージェントへと急速に拡大していることを示しています。Devin、Windsurf、Cline、そしてGitHub CopilotやCursorのエージェントモードは、このトレンドを象徴しています<sup>5</sup>。しかし、現状では、複雑なタスクを完全に「任せきり」にできるほどの自律性はまだ実現されておらず<sup>53</sup>、特に高度な自律性を追求すると、DevinのACUコスト<sup>22</sup>やClineのようなツールでの高額なAPIトークン消費<sup>26</sup>に見られるように、非常に高コストになる可能性があります。現在のエージェントは、完全な独立エンジニアというよりは、高度に有能なアシスタントと捉えるのが実態に近いでしょう。

一方で、Clineのようなオープンソースのエージェントフレームワーク<sup>28</sup>や、SWE-benchで高いスコアを出したSWE-agent<sup>9</sup>は、利用するモデルやコスト、プライバシーに対するコントロールをユーザーに与えるという利点があります<sup>28</sup>。ユーザーは好みのAPIバックエンドを選択したり、ローカルモデルを実行したりできます。ただし、APIキーの管理や、ローカルモデルの場合はインフラストラクチャの準備など、初期設定の手間がかかります。これに対し、Copilot<sup>7</sup>、Cursor<sup>18</sup>、Devin<sup>53</sup>といった商用ツールは、より統合された簡単な利用開始体験を提供しますが、柔軟性やコストの透明性、データプライバシー(ベンダーは企業向けに保証を提供することが多いですが<sup>7</sup>)の点でトレードオフが生じる可能性があります。

## 4. 従量課金APIサービスとコーディング利用コスト

サブスクリプションサービスに加えて、APIを直接利用する、あるいはClineのようなツール経由で利用する際の従量課金サービスについて、価格体系とコーディング関連の利用可能性を詳述します。

### 4.1. OpenAI API (GPT-4シリーズ, oシリーズ)

- **モデルと価格:** コーディングに適した主要モデルの価格(100万トークンあたり)は以下の通りです。キャッシュされた入力トークンは割引価格が適用されます<sup>23</sup>。
  - **GPT-4o:** 入力\$2.50 / 出力\$10.00<sup>6723</sup>
  - **GPT-4o mini:** 入力\$0.15 / 出力\$0.60<sup>67</sup>
  - **o3:** 入力\$10.00 / 出力\$40.00<sup>23</sup>
  - **o4-mini:** 入力\$1.10 / 出力\$4.40<sup>23</sup>
  - **GPT-4.1:** 入力\$2.00 / 出力\$8.00<sup>23</sup>

- **GPT-4.1 mini:** 入力\$0.40 / 出力\$1.60 <sup>23</sup>
- **GPT-4.1 nano:** 入力\$0.10 / 出力\$0.40 <sup>23</sup>
- **GPT-4 (8k):** 入力\$30.00 / 出力\$60.00 <sup>69</sup>
- **GPT-4 Turbo (preview):** 入力\$10.00 / 出力\$30.00 <sup>67</sup>
- ファインチューニングやBatch API(50%割引)も利用可能です <sup>23</sup>。
- **コーディング関連:** GPT-4およびoシリーズモデルは、コーディングタスクにおいて高い能力を発揮します <sup>1</sup>。APIを利用することで、カスタム開発ワークフローへの統合が可能です。

## 4.2. Anthropic Claude API

- **モデルと価格:** コーディングに強いとされるモデルの価格(100万トークンあたり)は以下の通りです <sup>24</sup>。プロンプトキャッシュやバッチ処理によるコスト削減も可能です <sup>24</sup>。
  - **Claude 3.7 Sonnet:** 入力\$3.00 / 出力\$15.00
  - **Claude 3.5 Sonnet:** 入力\$3.00 / 出力\$15.00
  - **Claude 3 Opus:** 入力\$15.00 / 出力\$75.00
  - **Claude 3.5 Haiku:** 入力\$0.80 / 出力\$4.00 <sup>6772</sup>
  - **Claude 3 Haiku:** 入力\$0.25 / 出力\$1.25
  - ツール利用(例: text\_editor\_20250124)には追加の入力トークン(700トークン)がかかります <sup>74</sup>。
- **コーディング関連:** Claude 3.7 Sonnetはコーディングにおいて最先端と評価されています <sup>10</sup>。APIは、拡張思考、最大200Kトークンの長いコンテキストウィンドウ(モデルによる)、128Kトークンの出力 <sup>24</sup>、そしてターミナルで動作する「Claude Code」<sup>76</sup> やテキスト編集ツール <sup>74</sup> へのアクセスを提供します。

## 4.3. Codex CLI (OpenAIツールとして)

- **位置づけ:** Codex CLIはOpenAIが提供するオープンソースのツールであり、それ自体が独立した有料サービスではありません <sup>64</sup>。内部的にはOpenAIのAPI(Responses API)を使用し、デフォルトではo4-miniのようなモデルをターゲットとしています <sup>64</sup>。
- **コスト構造:** 費用は、Codex CLIがバックエンドで使用するOpenAI APIモデル(例: o4-mini)の料金体系と、エージェントの対話で消費されるトークン数によって決まります <sup>64</sup>。ソースコードはローカルに留まり、プロンプトやコンテキスト情報のみがAPIに送信されます <sup>64</sup>。
- **機能:** ターミナルベースのエージェントで、ローカルのコードを読み取り・変更し、コマンドを実行できます。マルチモーダル入力に対応し、承認モード(Suggest, Auto Edit, Full Auto)を選択可能です <sup>64</sup>。

## 4.4. Devin API / ACUコスト

- **構造:** Agent Compute Unit (ACU) に基づく従量課金制です <sup>22</sup>。Coreプランでは\$2.25/ACU(初期\$20分以降)、Teamプランでは\$2.00/ACU(月額\$500のバンドル内)と

なります<sup>22</sup>。1 ACUは約15分のアクティブ作業に相当します<sup>22</sup>。Devin APIも提供されています<sup>25</sup>。

- コーディング関連: Devinの自律的なコーディング能力へのプログラムアクセスを可能にします。コスト構造がトークンベースのAPIとは大きく異なります。

#### 4.5. コーディングタスクにおけるAPIコスト試算

API利用コストを試算するには、まずトークン数を推定する必要があります。一般的な目安として、「1トークン ≈ 英語4文字 ≈ 0.75単語」がありますが、これはモデル、言語、内容(コードか自然言語か)によって変動します<sup>77</sup>。OpenAIのtiktokenライブラリ<sup>77</sup>やAnthropicのcount-tokens APIエンドポイント<sup>83</sup>を使うと、より正確な推定が可能です。ただし、これらもあくまで推定値であり、実際のAPIコールでの消費量とは僅かに異なる場合があります<sup>81</sup>。

コスト計算では、入力トークンと出力トークンの単価が異なる点<sup>23</sup>、キャッシュ利用による入力コスト削減<sup>23</sup>、ツール利用に伴う追加トークン<sup>74</sup>などを考慮する必要があります。

以下に、いくつかのコーディングタスク例について、代表的なモデルを用いたAPIコストの概算を示します。

試算の前提:

- トークン数推定ルール: 1行のコード ≈ 10トークン(入力5 + 出力5の平均と仮定<sup>86</sup>)、自然言語プロンプトは「1トークン ≈ 0.75単語」で計算。
- 入力にはプロンプトと関連コンテキスト(既存コードスニペット等)を含む。
- 出力には生成されたコードと、必要に応じた説明を含む。
- コストは2025年初頭時点の公表価格に基づく(例: GPT-4o mini 入力\$0.15/M, 出力\$0.60/M; Claude 3.7 Sonnet 入力\$3.00/M, 出力\$15.00/M; OpenAI o3 入力\$10.00/M, 出力\$40.00/M)。
- キャッシュやバッチ処理、ツール利用コストは単純化のため除外。

タスク規模	推定入力トークン	推定出力トークン	推定コスト (GPT-4o mini)	推定コスト (Claude 3.7 Sonnet)	推定コスト (OpenAI o3)
小規模 (関数生成)					
- プロンプト (50単語)	~67	-			
- コンテキスト (20行コード)	~100	-			

- 生成コード (50行)	-	~250			
- 説明 (50単語)	-	~67			
合計	~167	~317	~\$0.000215	~\$0.005256	~\$0.014350
中規模 (クラス生成)					
- プロンプト (100単語)	~133	-			
- コンテキスト (100行コード)	~500	-			
- 生成コード (200行)	-	~1000			
- テストコード (50行)	-	~250			
- 説明 (100単語)	-	~133			
合計	~633	~1383	~\$0.000925	~\$0.022644	~\$0.061650
大規模 (コンポーネントリファクタリング)					
- プロンプト (150単語)	~200	-			
- コンテキスト (1000行コード)	~5000	-			
- 生成コード 差分 (300行)	-	~1500			

相当)					
- 説明/計画 (200単語)	-	~267			
合計 (1回あたり)	~5200	~1767	~\$0.001840	~\$0.042055	~\$0.122680
(10回の反復 を想定)	~52000	~17670	~\$0.018401	~\$0.420550	~\$1.226800

注: 上記は非常に単純化された概算であり、実際のコストはプロンプトの質、コンテキストの与え方、モデルの挙動、反復回数によって大きく変動します。特に大規模タスクやエージェント的な利用では、コストが指数関数的に増加する可能性があります<sup>26</sup>。

この試算からわかるように、APIプロバイダーや選択するモデルによってコストは劇的に異なります。GPT-4o miniやGemini Flash、Claude Haikuのような新世代の小型・効率化モデルは、従来のフラッグシップモデル(o3やOpusなど)と比較して、桁違いに低いコストで利用可能です<sup>23</sup>。これは、タスクの複雑さに応じて適切なモデルを選択する「モデルルーティング」戦略がコスト最適化に極めて重要であることを示唆しています<sup>85</sup>。

また、特にリファクタリングやデバッグのように、既存のコードベースや対話履歴といった大量のコンテキスト(入カトークン)を必要とするタスクでは、入カトークンのコストが総費用を大きく左右する可能性があります<sup>26</sup>。入カトークンは出カトークンより単価が安いものの、量が膨大になるとコストがかさみます<sup>85</sup>。したがって、プロンプトエンジニアリング、キャッシュの活用<sup>23</sup>、そして効率的なコンテキスト管理(関連性の高い情報のみを選別して与えるなど)<sup>89</sup>が、APIを経済的に利用するための鍵となります。

Codex CLIやClineのようなツールは、強力なAPIバックエンドを活用しつつも、コスト管理の責任と複雑さをユーザー側に委ねる形になります<sup>28</sup>。これは、モデル選択や利用方法の自由度が高い反面、Copilot ProやCursor Proのような統合型サブスクリプションツールが提供する、予測可能(ただし制限がある場合も)な月額コストとは対照的です<sup>13</sup>。利用者は、コントロールと柔軟性を取るか、コストの予測可能性とシンプルさを取るかのトレードオフに直面します。

## 5. AIコーディング能力の比較分析

各AIモデルおよびツールが、実際のコーディングタスクにおいてどの程度の能力を発揮するかを、ベンチマーク結果、機能分析、ユーザーレビューを統合して比較評価します。

### 5.1. ベンチマークパフォーマンス

- **SWE-bench (Software Engineering Benchmark):**

- 概要: 実際のGitHubリポジトリから収集されたIssue(バグ報告や機能要望)を解決する能力を測定するベンチマークです<sup>55</sup>。単なるコード生成ではなく、コードベースのナビゲーション、テストの実行、反復的な修正といった、より実世界に近いソフトウェア開発能力を評価します<sup>91</sup>。
- **Devin AI:** Cognition社は、Devinが非支援設定(25%サブセット、45分制限)で13.86%のIssueを解決したと報告しています<sup>54</sup>。これは当時のベースライン(Claude 2 + BM25検索で1.96%)を大幅に上回るものでした<sup>62</sup>。
- **SWE-agent + Claude 3.7 Sonnet:** 2025年初頭のSWE-bench Liteリーダーボードによると、この組み合わせのエージェントが33.83%の解決率でトップに立っています<sup>9</sup>。これは、Claude 3.7 Sonnetを基盤としたエージェントシステムが高い実世界コーディング能力を持つことを示唆しています。
- **SWE-bench Verified**における**Claude 3.7 Sonnet:** より検証されたサブセットであるSWE-bench Verifiedにおいては、Claude 3.7 Sonnet(単体、またはエージェント経由)が62.3%、カスタムスキャフォールディング(追加の指示やコンテキスト)を用いた場合は70.3%という高いスコアを記録しています<sup>10</sup>。これは他のモデル(OpenAI o1: 48.9%, o3-mini: 49.3%, DeepSeek R1: 49.2%)を大きく引き離す結果です<sup>10</sup>。
- **AI Index Report**の数値 (**71.7%**): Stanford AI Index 2025などのレポートで言及されている「SWE-benchで71.7%達成」<sup>1</sup>という数値は、特定のエージェントのスコアというよりは、AIシステム全体の能力向上を示す指標、あるいは特定の(例えば支援ありやスキャフォールディングありの)条件下での最高性能を指している可能性があります。最新のエージェントリーダーボードのスコア(例: 33.83%)とは解釈が異なる点に注意が必要です。
- その他のエージェント: Amazon Q Developer Agent、OpenHands、AutoCodeRoverなどもリーダーボード上位にランクインしており、エージェントシステムの開発競争が活発であることを示しています<sup>9</sup>。

- **HumanEval:**

- 概要: 主にPythonの関数レベルのコード生成能力を評価する、比較的古くからあるベンチマークです。
- 傾向: 近年、多くのモデルがこのベンチマークで高いスコアを達成しており、「飽和」状態にあるとも指摘されています<sup>2</sup>。そのため、SWE-benchのような、より複雑で実世界に近いベンチマークの重要性が増しています。米国と中国のトップモデル間の性能差は、このベンチマークにおいても急速に縮小しています<sup>1</sup>。利用可能なスコア例としては、Llama 3.1 405bが81.1%、Llama 3.3 70bが77.3%などがあります<sup>95</sup>。

- その他のコーディング関連ベンチマーク:

- **LiveCode Bench:** 実世界のコーディング能力を評価する別のベンチマークで、DeepSeek-R1が64.3%でトップという結果があります<sup>96</sup>。
- **BigCodeBench:** 人間の基準(97%)にはまだ及ばないものの、AIシステムが35.5%の成功率を達成したと報告されています<sup>2</sup>。



- **TAU-bench (Tool Augmented Understanding):** ツール連携を含む実世界タスクでのエージェント能力を評価します。Claude 3.7 Sonnetが特に高い性能を示しています<sup>11</sup>。
- **Polyglot Benchmark (Aider):** 複数のプログラミング言語(C++, Go, Java, JavaScript, Python, Rust)にわたる難易度の高いコーディング演習で自律的な問題解決能力をテストします。Refact.ai Agent + Claude 3.7 Sonnetが76.4%でトップスコアを記録したと報告されています<sup>97</sup>。SWE-benchよりも現実の開発ワークフローに近いと主張されています<sup>97</sup>。

ベンチマーク結果はAIのコーディング能力を測る上で有用な指標ですが、その解釈には注意が必要です。SWE-benchのスコアがエージェントの実装方法、使用するモデル、評価設定(非支援か支援ありか、スキャフォールディングの有無、評価サブセット)によって大きく変動することからもわかるように<sup>9</sup>、単一の数値だけで能力を判断することはできません。Devinの13.86%というスコアも、特定の条件下での結果です<sup>62</sup>。したがって、ベンチマークスコアは、ユーザーレビューや実際の使用感、特定のワークフローにおける性能と合わせて総合的に評価することが重要です<sup>4</sup>。

## 5.2. モデル別コーディング能力比較

### ● Claude 3.7 Sonnet:

- 強み: 現在、特に複雑なコーディングタスクにおいて最先端の能力を持つと広く認識されています。SWE-benchやTAU-bench、Polyglot Benchmarkでの高いスコア<sup>9</sup>がそれを裏付けています。計画立案、複雑な問題解決、大規模リファクタリング、ツール連携、長いコード(最大128Kトークン出力)の生成・扱いに優れています<sup>10</sup>。拡張思考モードはコーディング性能をさらに向上させます<sup>10</sup>。Cursor、Cognition (Devin)、Vercel、Replitといった他のAIツール開発企業からも、そのコーディング能力、特に計画能力や精度が高く評価されています<sup>12</sup>。

### ● OpenAI GPT-4o / oシリーズ:

- 強み: 汎用性が高く、多くのツール(ChatGPT, Perplexity Pro, Cursor, Copilot Premium等)で利用可能な強力なモデルです<sup>7</sup>。o1やo3といった推論に特化したモデルは、数学やコーディングの性能を向上させますが、応答速度が遅くコストが高いというトレードオフがあります<sup>2</sup>。GPT-4o miniは、単純なタスクに対して非常にコスト効率が良い選択肢となります<sup>68</sup>。
- 現状: SWE-benchのエージェント評価では、現時点ではClaude 3.7 Sonnetベースのエージェントに後れを取っています<sup>9</sup>。

### ● Gemini モデル (Google):

- 強み: Googleの最新モデルであり、特にマルチモーダル能力に優れています<sup>37</sup>。Gemini AdvancedプランやGoogle CloudのGemini Code Assistで利用可能です<sup>7</sup>。コーディング能力も高く、Claude 3.7 Sonnetと比較されるレベルにあります<sup>75</sup>。Gemini FlashシリーズはAPI利用において非常に低コストです<sup>67</sup>。

- 比較: リファクタリングや説明の質ではClaudeが優位とされる場面もあるようです<sup>75</sup>。
- その他のモデル:
  - **DeepSeek:** コーディングに特化したモデル(DeepSeek Coder V2, DeepSeek-R1等)を開発しており、LiveCode Benchなどで高い性能を示しています<sup>50</sup>。低レイテンシや hallucination 制御にも強みがあると評価されています<sup>99</sup>。
  - **Llama (Meta):** オープンソースモデルとして高性能化が進んでおり、HumanEvalなどで高いスコアを記録しています<sup>95</sup>。

現状では、特に理由付け、計画、大規模なコンテキスト処理を伴う複雑なコーディングタスクにおいては、Claude 3.7 Sonnetが一步リードしていると考えられます。これは、ベンチマーク結果<sup>9</sup>と、競合他社を含むサードパーティからの評価<sup>12</sup>によって裏付けられています。

### 5.3. ツール別 強み・弱み・UX

- **GitHub Copilot vs. Cursor:**
  - **Copilot:** 強みはGitHubエコシステムとの深い統合、比較的安定したモデル性能(ユーザーによる)<sup>4</sup>、Proプランのコストパフォーマンス(\$10/月)<sup>4</sup>。エージェントモードも改善・展開中です<sup>4</sup>。弱みは、Cursorと比較して編集速度が遅いと感じるユーザーがいること<sup>4</sup>、UX上の細かい不満点(提案の強引さ、適用コードのハイライト不足等)<sup>4</sup>が挙げられます。
  - **Cursor:** 強みはAI統合型IDEとしての体験、高速なコード編集と自動補完<sup>4</sup>、高度なコンテキスト管理機能(.cursorrules)<sup>46</sup>、エージェントモード<sup>6</sup>。弱みは、モデル性能のばらつき(ユーザーによる)<sup>4</sup>、比較的高価で変動しうるコスト(\$20/月 + プレミアムリクエスト超過料金)<sup>4</sup>、VS Codeフォークであることによるショートカットの競合やUIの煩雑さ<sup>4</sup>、潜在的な互換性問題<sup>46</sup>です。
  - 選択: 統合性とコスト重視ならCopilot、AI中心の高速な開発体験を求めるならCursor、という選択軸が考えられますが、Copilotのプレミアムリクエスト制限の導入<sup>4</sup>やCursorのコスト変動リスクも考慮に入れる必要があります。
- **Devin AI:**
  - 強み: 「自律性」というコンセプト<sup>53</sup>。
  - 弱み: 現状での信頼性、複雑なタスクへの対応力、そして高コスト(特にACUベースの課金)<sup>22</sup>。
  - UX: Webアプリ、Slack、APIなどを介して対話します<sup>53</sup>。
- **WindSurf AI:**
  - 強み: エージェント的な「Flows」によるリアルタイム協調、反復的なデバッグループ<sup>27</sup>。
  - UX: VS Codeフォークとしての使い勝手<sup>27</sup>。クレジットシステムはやや複雑かもしれません<sup>19</sup>。
- **Cline AI:**
  - 強み: オープンソースであることによる柔軟性(モデル選択、コスト管理)、Plan/Act

モード、MCPによるツール拡張性<sup>7</sup>。

- **UX:** VS Codeのターミナル内で動作します<sup>28</sup>。APIキーの設定など、初期セットアップが必要です<sup>28</sup>。
- 全般的な機能比較: 各ツールが提供するコード生成、補完、デバッグ支援、リファクタリング、ドキュメント生成、テスト生成、複数ファイル操作、コンテキスト処理能力などを総合的に比較検討する必要があります<sup>6</sup>。

IDEとの統合方式も、ツールの思想を反映しています。Copilotは既存IDEへの深い統合<sup>5</sup>、CursorとWindSurfはAIファーストのIDEフォーク<sup>18</sup>、ClineはIDEと並行して動作する拡張機能/エージェント<sup>28</sup>という異なるアプローチを取っています。これにより、UX、機能の深さ、そして標準的なVS Codeエコシステムからの逸脱度合い(互換性リスク等)が変わってきます<sup>46</sup>。ユーザーは、自身の開発スタイルやAIとの連携の好みに合わせて選択することになります。

#### 5.4. AIコーディング能力 比較スナップショット

ツール/モデル	主要なコーディング強み	主要な弱み/制限事項	SWE-benchスコア例 (エージェント/モデル)	ユーザー評価ハイライト (速度, UX, 精度)
Claude 3.7 Sonnet (API/Agent経由)	複雑なタスク, 計画, リファクタリング, ツール連携, 長文脈/長出力 <sup>10</sup>	API利用コストが高くなる可能性 <sup>26</sup>	33.8% (SWE-agent Lite), 70.3% (Verified, Scaffolded) <sup>9</sup>	高い精度と計画能力を評価する声が多い <sup>12</sup>
GPT-4o / o3 (API経由)	高い汎用性, 推論特化モデル (o1/o3)あり <sup>21</sup> , GPT-4o miniは低コスト <sup>68</sup>	o1/o3は高コスト/低速 <sup>2</sup> , SWE-benchエージェント性能は現状Claudeに劣る <sup>9</sup>	SWE-agent + GPT-4o: 11.99% (Lite) <sup>9</sup>	汎用性の高さを評価。モデルによる性能差が大きい。
Gemini Advanced (Google One)	Google最上位モデル, Googleエコシステム連携 <sup>16</sup> , マルチモーダル <sup>37</sup>	Workspace統合が不要な場合は割高感も。リファクタリング等でClaudeに劣る場面も <sup>75</sup>	N/A (Agent Leaderboard)	Googleサービス連携を評価。コーディング能力は高いが、特筆すべき点に欠けるとの意見も。
GitHub Copilot (Pro/Pro+)	IDE/エコシステム統合 <sup>5</sup> , 良好なコストパフォーマンス	編集速度が遅いとの指摘 <sup>4</sup> , UXの細かな不満 <sup>4</sup> , プレミ	N/A (Agent Modelは評価途上)	統合性と価格を評価する声 <sup>4</sup> 。速度や提案の質には

	ス(Pro) <sup>4</sup> , プレミアムモデルアクセス可 <sup>5</sup>	アムリクエスト制限 <sup>4</sup>		賛否 <sup>4</sup> 。
Cursor Pro	AI統合IDE <sup>6</sup> , 高速な編集/補完 <sup>4</sup> , 高度なコンテキスト管理 <sup>46</sup> , エージェントモード <sup>6</sup>	モデル性能のばらつき指摘 <sup>4</sup> , 高コスト/変動リスク <sup>4</sup> , VSCodeフォーク由来のUX問題 <sup>4</sup>	N/A (Agent Leaderboard)	速度とUXを高く評価する声が多い <sup>4</sup> 。コストとモデル安定性に懸念の声も <sup>4</sup> 。
Devin AI	自律性コンセプト <sup>53</sup> , 特定タスクでの効率化事例 <sup>25</sup>	高コスト <sup>22</sup> , 信頼性/複雑タスク対応の課題 <sup>53</sup> , デモ精度への疑問 <sup>54</sup>	13.86% (Unassisted, Subset) <sup>55</sup>	自律性に期待。現状は限定的なタスク向きとの評価。
WindSurf Pro	エージェント型ワークフロー (Flows) <sup>52</sup> , 反復的デバッグ <sup>27</sup> , VSCodeフォーク <sup>52</sup>	クレジットシステムが複雑 <sup>19</sup> , 比較的新しいツール	N/A (Agent Leaderboard)	エージェント機能に期待。UXはCursorと比較されることが多い <sup>47</sup> 。
Cline (+ Backend Model)	オープンソース <sup>63</sup> , モデル/コストの柔軟性 <sup>28</sup> , Plan/Actモード <sup>28</sup> , MCPツール連携 <sup>8</sup>	初期設定が必要 <sup>28</sup> , 性能は選択したバックエンドモデル依存	N/A (Agent Leaderboard)	柔軟性とコントロールを評価。VS Code内で完結する点を好む声も <sup>50</sup> 。APIコスト管理が必要 <sup>26</sup> 。

## 6. 結論と戦略的推奨事項

本レポートでは、消費者向けの主要な生成AIコーディング支援ツールについて、コスト、機能、およびコーディング能力を多角的に比較分析しました。分析の結果、最適なツールの選択は、利用者の予算、求める機能（コード補完か、高度なエージェント機能か）、好みの開発ワークフロー（IDE統合か、API利用か）、そしてコスト変動や複雑さへの許容度によって大きく異なることが明らかになりました。単一の「最高の」ツールは存在せず、個々のニーズに合わせた戦略的な選択が求められます。

主要なトレードオフ:

- **コスト vs. 能力:** 月額\$10～\$20の手頃なプランは基本的な支援機能を提供しますが、最

先端のコーディング能力や高度な自律性を求める場合は、月額\$100を超える高額プランや変動の大きいAPI利用コストが必要になる傾向があります。

- **シンプルさ vs. 柔軟性:** 統合されたサブスクリプションツールは導入が容易でコストが予測しやすい反面、機能やモデル選択の自由度が制限されることがあります。一方、API利用やオープンソースツールは高い柔軟性を提供しますが、設定の複雑さやコスト管理の負担が伴います。
- **統合体験 vs. オープンソース管理:** IDEに深く統合されたツール(Copilot, Cursor)はシームレスな体験を提供しますが、特定のベンダーのエコシステムに依存します。オープンソースツール(Cline)は、モデルやデータの管理権をユーザーに与えますが、自己管理が必要です。

#### コスト概観:

- 手頃な基本層: \$10~\$20/月 (Copilot Pro, ChatGPT Plus, Gemini Advanced, Perplexity Pro, Cursor Pro, WindSurf Pro)。
- 高価格/高性能層: \$100~\$500+/月、または変動制 (Claude Max, ChatGPT Pro, Copilot Pro+, Devin, API利用)。

#### 能力概観:

- 現時点での高性能: Claude 3.7 SonnetをバックエンドとするエージェントやAPI利用が、特に複雑なコーディングタスクにおいて高い性能を示しています。
- 強力なUX/統合: GitHub Copilot (エコシステム連携)とCursor (AI統合IDE)が、ユーザーの好みによって評価が分かれるものの、洗練された開発体験を提供します。
- 高い自律性(潜在性): Devin、WindSurf、Clineなどが自律的なタスク実行機能を提供しますが、現状ではコストや信頼性の面で課題も残ります。

#### 利用者タイプ別推奨事項:

- **コストを最優先する個人開発者:**
  - 推奨: GitHub Copilot Pro (\$10/月) から始めるのが、IDE統合と機能のバランス、コストパフォーマンスの点で優れています。無料枠のある学生・教師・OSSメンテナーにも最適です。
  - 代替: Cline (オープンソース) と低コストAPI (GPT-4o mini, Gemini Flash, Claude Haiku等) を組み合わせて利用することも検討に値しますが、設定とAPIコスト管理が必要です。ChatGPTやPerplexityの無料版も基本的な用途には利用できます。
- **最高のコーディング性能を求める開発者 (コスト感度低):**
  - 推奨: Claude 3.7 SonnetのAPIを直接、またはClineのようなツール経由で利用することを検討します。ベンチマークでの高い性能が期待できますが、APIコストは高額になる可能性があります。
  - 代替: Claude 3.7 Sonnet等へのアクセス権を含む高価格帯のサブスクリプション (GitHub Copilot Pro+, Cursor Pro) も選択肢です。Claude MaxはWeb UIでの利用



が主目的であれば検討できます。

- 最高のIDE統合とUXを求める開発者:
  - 推奨: GitHub Copilot (Pro/Pro+) と Cursor Pro の間で選択します。CopilotはネイティブなVS Code/JetBrains体験とGitHub連携を提供し、CursorはAI機能をより深く統合した独自のIDE体験を提供します。可能であれば両方を試用し、自身のワークフローに合う方を選択するのが最善です。
- 高度な自律性・エージェント機能を試したい開発者:
  - 推奨: DevinのCoreプラン(従量課金)、WindSurf、またはClineを実験的に導入することを検討します。ただし、これらはまだ発展途上の技術であり、期待通りの結果が得られない可能性や、予期せぬ高コストが発生するリスク、学習曲線が必要であることを理解しておく必要があります。

#### 将来展望:

AIコーディング支援ツールの分野は、依然として急速な進化の途上にあります。モデルの性能向上、新機能の追加、価格体系の見直しが今後も続くと予想されます。特に、自律型エージェントの能力向上とコスト効率の改善が注目されます。したがって、現時点での選択が永続的なものではなく、定期的に市場動向を評価し、自身のニーズに合わせてツールを見直すことが重要です。

現状では、多くの開発者にとって最も実用的かつ費用対効果が高いのは、IDEに統合された高度なコード補完、チャットベースの支援、そしてターゲットを絞ったコード生成・リファクタリング機能でしょう。完全な自律開発はまだ未来の領域ですが、現在のツールでも開発者の生産性を大幅に向上させることは十分に可能です。自身の目的と予算を明確にし、本レポートで提供された情報を活用して、最適なAIコーディングパートナーを見つける一助となれば幸いです。

#### 引用文献

1. The AI Race Accelerates: Key Insights from the 2025 AI Index Report - CTSE@AEI.org, 4月 20, 2025にアクセス、  
<https://ctse.aei.org/the-ai-race-accelerates-key-insights-from-the-2025-ai-index-report/>
2. Technical Performance | The 2025 AI Index Report | Stanford HAI, 4月 20, 2025にアクセス、  
<https://hai.stanford.edu/ai-index/2025-ai-index-report/technical-performance>
3. The 2025 AI Index Report | Stanford HAI, 4月 20, 2025にアクセス、  
<https://hai.stanford.edu/ai-index/2025-ai-index-report>
4. GitHub Copilot vs Cursor in 2025: Why I'm paying half price for the ..., 4月 20, 2025にアクセス、  
[https://www.reddit.com/r/GithubCopilot/comments/1jnboan/github\\_copilot\\_vs\\_cursor\\_in\\_2025\\_why\\_im\\_paying/](https://www.reddit.com/r/GithubCopilot/comments/1jnboan/github_copilot_vs_cursor_in_2025_why_im_paying/)



5. Vibe coding with GitHub Copilot: Agent mode and MCP support rolling out to all VS Code users | Microsoft Azure Blog, 4月 20, 2025にアクセス、  
<https://azure.microsoft.com/en-us/blog/vibe-coding-with-github-copilot-agent-mode-and-mcp-support-rolling-out-to-all-vs-code-users/>
6. Cursor AI Review: Best AI Code Editor in 2025?, 4月 20, 2025にアクセス、  
<https://www.fahimai.com/cursor-ai>
7. Best AI Coding Assistants as of April 2025 - Shakudo, 4月 20, 2025にアクセス、  
<https://www.shakudo.io/blog/best-ai-coding-assistants>
8. Top 5 Agentic AI Coding Assistants April 2025 | APIpie, 4月 20, 2025にアクセス、  
<https://apipie.ai/docs/blog/top-5-agentic-ai-coding-assistants>
9. SWE-bench, 4月 20, 2025にアクセス、<https://www.swebench.com/>
10. Claude 3.7 Sonnet: Features, Access, Benchmarks & More - DataCamp, 4月 20, 2025にアクセス、<https://www.datacamp.com/blog/claude-3-7-sonnet>
11. Claude 3.7 Sonnet: The Best Coding Model Yet? - Analytics Vidhya, 4月 20, 2025にアクセス、  
<https://www.analyticsvidhya.com/blog/2025/02/claude-3-7-sonnet-for-coding/>
12. Claude 3.7 Sonnet and Claude Code - Anthropic, 4月 20, 2025にアクセス、  
<https://www.anthropic.com/news/claude-3-7-sonnet>
13. About billing for GitHub Copilot, 4月 20, 2025にアクセス、  
<https://docs.github.com/en/billing/managing-billing-for-your-products/managing-billing-for-github-copilot/about-billing-for-github-copilot>
14. Pricing - Anthropic, 4月 20, 2025にアクセス、<https://www.anthropic.com/pricing>
15. Pricing - ChatGPT - OpenAI, 4月 20, 2025にアクセス、  
<https://openai.com/chatgpt/pricing/>
16. Google One AI Premium Plan and Features - Google One, 4月 20, 2025にアクセス、  
<https://one.google.com/about/ai-premium/>
17. Is Perplexity Pro Worth Your \$20? Exploring the Subscription Perks ..., 4月 20, 2025にアクセス、  
<https://opentools.ai/news/is-perplexity-pro-worth-your-dollar20-exploring-the-subscription-perks>
18. Pricing | Cursor - The AI Code Editor, 4月 20, 2025にアクセス、  
<https://www.cursor.com/pricing>
19. Pricing | Windsurf (formerly Codeium), 4月 20, 2025にアクセス、  
<https://windsurf.com/pricing>
20. Anthropic launches Max Plan for high-volume Claude users - N24 English, 4月 20, 2025にアクセス、  
<https://n24.com.tr/en/anthropic-launches-max-plan-for-high-volume-claude-users>
21. Is ChatGPT Plus worth your \$20? Here's how it compares to Free ..., 4月 20, 2025にアクセス、  
<https://www.zdnet.com/article/is-chatgpt-plus-worth-your-20-heres-how-it-compares-to-free-and-pro-plans/>
22. Devin, the viral coding AI agent, gets a new pay-as-you-go plan - Yahoo Finance, 4月 20, 2025にアクセス、  
<https://sg.finance.yahoo.com/news/devin-viral-coding-ai-agent-194633884.html>

23. Pricing | OpenAI, 4月 20, 2025にアクセス、<https://openai.com/api/pricing/>
24. Claude 3.7 Sonnet - Anthropic, 4月 20, 2025にアクセス、<https://www.anthropic.com/claude/sonnet>
25. Pricing | Devin, 4月 20, 2025にアクセス、<https://devin.ai/pricing>
26. The \$20 Claude Pro subscription would cost over \$1,300 via the API : r/ClaudeAI - Reddit, 4月 20, 2025にアクセス、[https://www.reddit.com/r/ClaudeAI/comments/1j2n2va/the\\_20\\_claude\\_pro\\_subscription\\_would\\_cost\\_over/](https://www.reddit.com/r/ClaudeAI/comments/1j2n2va/the_20_claude_pro_subscription_would_cost_over/)
27. Windsurf AI Agentic Code Editor: Features, Setup, and Use Cases ..., 4月 20, 2025にアクセス、<https://www.datacamp.com/tutorial/windsurf-ai-agentic-code-editor>
28. cline/cline: Autonomous coding agent right in your IDE ... - GitHub, 4月 20, 2025にアクセス、<https://github.com/cline/cline>
29. Claude vs ChatGPT: Guide to Choosing the Best AI Tool - Data Science Dojo, 4月 20, 2025にアクセス、<https://datasciencedojo.com/blog/claude-vs-chatgpt/>
30. Claude 3.7 Sonnet: The Best Coding Model Yet? - Analytics Vidhya, 4月 20, 2025にアクセス、<https://www.analyticsvidhya.com/blog/2025/02/claude-3-7-sonnet-coding/>
31. Claude 3.7 Sonnet API documentation - Segmind, 4月 20, 2025にアクセス、<https://www.segmind.com/models/claude-3.7-sonnet/api>
32. Comparing Claude 3.7 Sonnet with ChatGPT and Other AI Systems - PageOn.ai, 4月 20, 2025にアクセス、<https://www.pageon.ai/blog/sonnet-3-7>
33. Technical Review: Claude 3.7 Sonnet - Helicone, 4月 20, 2025にアクセス、<https://www.helicone.ai/blog/claude-3.7-benchmarks-and-examples>
34. Claude's Max Plan: Expanded Access for Demanding Projects - Anthropic, 4月 20, 2025にアクセス、<https://www.anthropic.com/news/max-plan>
35. Claude Max Pricing gotta be a joke : r/ClaudeAI - Reddit, 4月 20, 2025にアクセス、[https://www.reddit.com/r/ClaudeAI/comments/1k08d3x/claude\\_max\\_pricing\\_gotta\\_be\\_a\\_joke/](https://www.reddit.com/r/ClaudeAI/comments/1k08d3x/claude_max_pricing_gotta_be_a_joke/)
36. Google's One AI Premium plan with Gemini Advanced is now free for students - for an entire year | ZDNET, 4月 20, 2025にアクセス、<https://www.zdnet.com/article/googles-one-ai-premium-plan-with-gemini-advanced-is-now-free-for-students-for-an-entire-year/>
37. Gemini vs ChatGPT vs Claude Comparison for 2025 - NewOaks AI, 4月 20, 2025にアクセス、<https://www.newoaks.ai/blog/gemini-vs-chatgpt-vs-claude-comparison-2025/>
38. ChatGPT vs Gemini vs Claude: Comparing Top AI Models [2025] - Weam AI, 4月 20, 2025にアクセス、<https://weam.ai/blog/guide/chatgpt-vs-gemini-vs-claude/>
39. The best AI for coding in 2025 (and what not to use - including DeepSeek R1) - ZDNet, 4月 20, 2025にアクセス、<https://www.zdnet.com/article/the-best-ai-for-coding-in-2025-and-what-not-to-use-including-deepseek-r1/>
40. Pricing - Perplexity, 4月 20, 2025にアクセス、<https://docs.perplexity.ai/guides/pricing>
41. About billing for individual Copilot plans - GitHub Docs, 4月 20, 2025にアクセス、<https://docs.github.com/en/copilot/managing-copilot/managing-copilot-as-an-in>

- [dividual-subscriber/billing-and-payments/about-billing-for-individual-copilot-plans](#)
42. About individual Copilot plans and benefits - GitHub Docs, 4月 20, 2025にアクセス、  
<https://docs.github.com/copilot/managing-copilot/managing-copilot-as-an-individual-subscriber/about-github-copilot-free>
  43. Using GitHub Copilot code review, 4月 20, 2025にアクセス、  
<https://docs.github.com/en/copilot/using-github-copilot/code-review/using-copilot-code-review>
  44. Cursor vs Copilot: A Comparison - CodeParrot, 4月 20, 2025にアクセス、  
<https://codeparrot.ai/blogs/cursor-vs-copilot-a-comparison>
  45. Cursor Review 2025 - Tips, Alternatives & More, 4月 20, 2025にアクセス、  
<https://www.toksta.com/products/cursor>
  46. Cursor AI: An In Depth Review in 2025 - Engine Labs Blog, 4月 20, 2025にアクセス、  
<https://blog.enginelabs.ai/cursor-ai-an-in-depth-review>
  47. 5 AI Code Editors to Use in 2025 - KDnuggets, 4月 20, 2025にアクセス、  
<https://www.kdnuggets.com/5-ai-code-editors-to-use-in-2025>
  48. 17 Best AI-Powered Coding Assistant Tools in 2025 - Spacelift, 4月 20, 2025にアクセス、  
<https://spacelift.io/blog/ai-coding-assistant-tools>
  49. Cursor is better than Github Co-pilot - Developer Review : r/ChatGPTCoding - Reddit, 4月 20, 2025にアクセス、  
[https://www.reddit.com/r/ChatGPTCoding/comments/1iepd4a/cursor\\_is\\_better\\_than\\_github\\_copilot\\_developer/](https://www.reddit.com/r/ChatGPTCoding/comments/1iepd4a/cursor_is_better_than_github_copilot_developer/)
  50. What are the best AI code assistants for vscode in 2025? - Reddit, 4月 20, 2025にアクセス、  
[https://www.reddit.com/r/vscode/comments/1je1i6h/what\\_are\\_the\\_best\\_ai\\_code\\_assistants\\_for\\_vscode/](https://www.reddit.com/r/vscode/comments/1je1i6h/what_are_the_best_ai_code_assistants_for_vscode/)
  51. Windsurf AI Reviews: Use Cases, Pricing & Alternatives - Futurepedia, 4月 20, 2025にアクセス、  
<https://www.futurepedia.io/tool/windsurf>
  52. A Guide to Using Windsurf.ai - CodeParrot, 4月 20, 2025にアクセス、  
<https://codeparrot.ai/blogs/a-guide-to-using-windsurfai>
  53. Devin Docs: Introducing Devin, 4月 20, 2025にアクセス、  
<https://docs.devin.ai/get-started/devin-intro>
  54. Devin AI - Wikipedia, 4月 20, 2025にアクセス、  
[https://en.wikipedia.org/wiki/Devin\\_AI](https://en.wikipedia.org/wiki/Devin_AI)
  55. Introducing Devin, the first AI software engineer - Cognition, 4月 20, 2025にアクセス、  
<https://cognition.ai/blog/introducing-devin>
  56. Get started with Devin - Cognition AI, 4月 20, 2025にアクセス、  
<https://www.cognition.ai/get-started>
  57. Devin September '24 Product Update - Cognition, 4月 20, 2025にアクセス、  
<https://cognition.ai/blog/sept-24-product-update>
  58. Devin AI vs. Cursor AI - Cost, Autonomy, and Coding Power - Makai Digital, 4月 20, 2025にアクセス、  
<https://www.makaihq.com/blog/devin-ai-vs-cursor-ai>
  59. Devin 2.0 Released; New Plan Offers 96% Price Cut | AIM Media House, 4月 20, 2025にアクセス、

- <https://analyticsindiamag.com/ai-news-updates/devin-2-0-released-new-plan-of-fers-96-price-cut/>
60. Devin, the viral coding AI agent, gets a new pay-as-you-go plan | TechCrunch - BestofAI, 4月 20, 2025にアクセス、  
<https://bestofai.com/article/devin-the-viral-coding-ai-agent-gets-a-new-pay-as-you-go-plan-techcrunch>
  61. Devin AI, 4月 20, 2025にアクセス、<https://devin.ai/>
  62. SWE-bench technical report - Cognition, 4月 20, 2025にアクセス、  
<https://www.cognition.ai/blog/swe-bench-technical-report>
  63. Cline - AI Agent Store, 4月 20, 2025にアクセス、  
<https://aiagentstore.ai/ai-agent/cline>
  64. OpenAI Codex CLI - Getting Started | OpenAI Help Center, 4月 20, 2025にアクセス、  
<https://help.openai.com/en/articles/11096431-openai-codex-cli-getting-started>
  65. OpenAI Codex CLI, how does it work? - Philschmid, 4月 20, 2025にアクセス、  
<https://www.philschmid.de/openai-codex-cli>
  66. Hi everyone! Boris from the Claude Code team here. @eschluntz, @catherinewu, @wo... | Hacker News, 4月 20, 2025にアクセス、  
<https://news.ycombinator.com/item?id=43163488>
  67. Free OpenAI & every-LLM API Pricing Calculator | Updated Apr 2025 - DocsBot AI, 4月 20, 2025にアクセス、  
<https://docsbot.ai/tools/gpt-openai-api-pricing-calculator>
  68. How to Calculate OpenAI API Price for GPT-4, GPT-4o and GPT-3.5 Turbo? - Analytics Vidhya, 4月 20, 2025にアクセス、  
<https://www.analyticsvidhya.com/blog/2024/12/openai-api-cost/>
  69. How much does GPT-4 cost? - OpenAI Help Center, 4月 20, 2025にアクセス、  
<https://help.openai.com/en/articles/7127956-how-much-does-gpt-4-cost>
  70. OpenAI API Pricing Calculator - Estimates Cost for LLM APIs - Markovate, 4月 20, 2025にアクセス、<https://markovate.com/openai-llm-api-pricing-calculator/>
  71. How to use the Codex models to work with code - Azure OpenAI Service | Microsoft Learn, 4月 20, 2025にアクセス、  
<https://learn.microsoft.com/en-us/azure/ai-services/openai/how-to/work-with-code>
  72. Getting Started with Claude 3 and the Claude 3 API - DataCamp, 4月 20, 2025にアクセス、  
<https://www.datacamp.com/tutorial/getting-started-with-claude-3-and-the-claude-3-api>
  73. Anthropic Claude API: A Practical Guide - Acorn Labs, 4月 20, 2025にアクセス、  
<https://www.acorn.io/resources/learning-center/claude-api/>
  74. Text editor tool - Anthropic API, 4月 20, 2025にアクセス、  
<https://docs.anthropic.com/en/docs/build-with-claude/tool-use/text-editor-tool>
  75. Claude 3.7 Sonnet vs Gemini 2.5 Pro for Coding: Which Model is the Best? - Apidog, 4月 20, 2025にアクセス、  
<https://apidog.com/blog/claude-3-7-sonnet-vs-gemini-2-5-pro>
  76. Claude Code overview - Anthropic API, 4月 20, 2025にアクセス、

- <https://docs.anthropic.com/en/docs/agents-and-tools/claude-code/overview>
77. What is the OpenAI algorithm to calculate tokens? - API, 4月 20, 2025にアクセス、  
<https://community.openai.com/t/what-is-the-openai-algorithm-to-calculate-tokens/58237>
  78. What are tokens and how to count them? - OpenAI Help Center, 4月 20, 2025にアクセス、  
<https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them>
  79. OpenAI Price: Calculate Tokens with Python - YouTube, 4月 20, 2025にアクセス、  
<https://www.youtube.com/watch?v=lu2WC7uaSWY>
  80. How to calculate the cost of a specific request made to the web API (and its reply), in tokens?, 4月 20, 2025にアクセス、  
<https://community.openai.com/t/how-to-calculate-the-cost-of-a-specific-request-made-to-the-web-api-and-its-reply-in-tokens/270878>
  81. OpenAI API: How do I count tokens before(!) I send an API request? - Stack Overflow, 4月 20, 2025にアクセス、  
<https://stackoverflow.com/questions/75804599/openai-api-how-do-i-count-tokens-before-i-send-an-api-request>
  82. How to calculate the tokens when using function call - API - OpenAI Developer Community, 4月 20, 2025にアクセス、  
<https://community.openai.com/t/how-to-calculate-the-tokens-when-using-function-call/266573>
  83. Count tokens for Claude models | Generative AI on Vertex AI - Google Cloud, 4月 20, 2025にアクセス、  
<https://cloud.google.com/vertex-ai/generative-ai/docs/partner-models/claude-count-tokens>
  84. Token counting - Anthropic API, 4月 20, 2025にアクセス、  
<https://docs.anthropic.com/en/docs/build-with-claude/token-counting>
  85. Openai Codex Pricing for AI-Generated Code | Restackio, 4月 20, 2025にアクセス、  
<https://www.restack.io/p/ai-generated-code-answer-openai-codex-pricing-cat-a-i>
  86. Claude 3.5 sonnet API + oTToDEV cost analysis - oTTomator Community, 4月 20, 2025にアクセス、  
<https://thinktank.ottomator.ai/t/claude-3-5-sonnet-api-ottodev-cost-analysis/834>
  87. How we built Townie – an app that generates fullstack apps - Val Town Blog, 4月 20, 2025にアクセス、  
<https://blog.val.town/blog/codegen/>
  88. OpenAI API Pricing Calculator - GPT for Work, 4月 20, 2025にアクセス、  
<https://gptforwork.com/tools/openai-chatgpt-api-pricing-calculator>
  89. AI Tokens Explained: Complete Guide to Usage, Optimization & Costs - Deepak Gupta, 4月 20, 2025にアクセス、  
<https://guptadeepak.com/complete-guide-to-ai-tokens-understanding-optimization-and-cost-management/>
  90. SWE-bench [Multimodal]: Can Language Models Resolve Real-world Github Issues?, 4月 20, 2025にアクセス、  
<https://github.com/SWE-bench/SWE-bench>

91. #1 open-source agent on SWE-Bench Verified by combining Claude 3.7 and O1, 4月 20, 2025にアクセス、  
<https://www.augmentcode.com/blog/1-open-source-agent-on-swe-bench-verified-by-combining-claude-3-7-and-o1>
92. Billions pour into AI as emissions rise, returns stay pitiful, say Stanford boffins - The Register, 4月 20, 2025にアクセス、  
[https://www.theregister.com/2025/04/11/stanford\\_ai\\_report/](https://www.theregister.com/2025/04/11/stanford_ai_report/)
93. Artificial Intelligence Index Report 2025 - AWS, 4月 20, 2025にアクセス、  
[https://hai-production.s3.amazonaws.com/files/hai\\_ai\\_index\\_report\\_2025.pdf](https://hai-production.s3.amazonaws.com/files/hai_ai_index_report_2025.pdf)
94. Stanford 2025 AI Index Reveals Surge in Adoption, Investment, and Global Impact as Trust and Regulation Lag Behind - Campus Technology, 4月 20, 2025にアクセス、  
<https://campustechnology.com/articles/2025/04/09/stanford-2025-ai-index-reveals-surge-in-adoption-investment-and-global-impact.aspx>
95. LLM Leaderboard 2025 - Vellum AI, 4月 20, 2025にアクセス、  
<https://www.vellum.ai/llm-leaderboard>
96. Open LLM Leaderboard 2025 - Vellum AI, 4月 20, 2025にアクセス、  
<https://www.vellum.ai/open-llm-leaderboard>
97. Our AI Agent + 3.7 Sonnet ranked #1 on Aider's polyglot bench — a 76.4% score, 4月 20, 2025にアクセス、  
<https://dev.to/refact/our-ai-agent-37-sonnet-ranked-1-on-aiders-polyglot-bench-a-764-score-10d0>
98. Open AI API costs me 1\$: r/OpenAI - Reddit, 4月 20, 2025にアクセス、  
[https://www.reddit.com/r/OpenAI/comments/1fxogml/open\\_ai\\_api\\_costs\\_me\\_1/](https://www.reddit.com/r/OpenAI/comments/1fxogml/open_ai_api_costs_me_1/)
99. Claude 3.7 Sonnet's results on six independent benchmarks : r/ClaudeAI - Reddit, 4月 20, 2025にアクセス、  
[https://www.reddit.com/r/ClaudeAI/comments/1iz3umm/claude\\_37\\_sonnets\\_results\\_on\\_six\\_independent/](https://www.reddit.com/r/ClaudeAI/comments/1iz3umm/claude_37_sonnets_results_on_six_independent/)
100. 15 Best AI Coding Assistant Tools in 2025 - Qodo, 4月 20, 2025にアクセス、  
<https://www.qodo.ai/blog/best-ai-coding-assistant-tools/>
101. 10 Best AI Coding Assistant Tools in 2025 – Guide for Developers | Blog - Droids On Roids, 4月 20, 2025にアクセス、  
<https://www.thedroidsonroids.com/blog/best-ai-coding-assistant-tools>