

Apple Foundation ModelsとPrivate Cloud Compute (PCC)で作る、 一歩先のAI体験

初心者でも簡単！プライバシーを守りながら
「もっと賢いAI」をアプリに組み込もう

優しい参考書シリーズ



「もっと賢いAI」を使いたいけれど、 こんな壁を感じていませんか？

「デバイス内のAIは便利
だけど、長い文章の処理や
複雑な推論は少し苦手…」

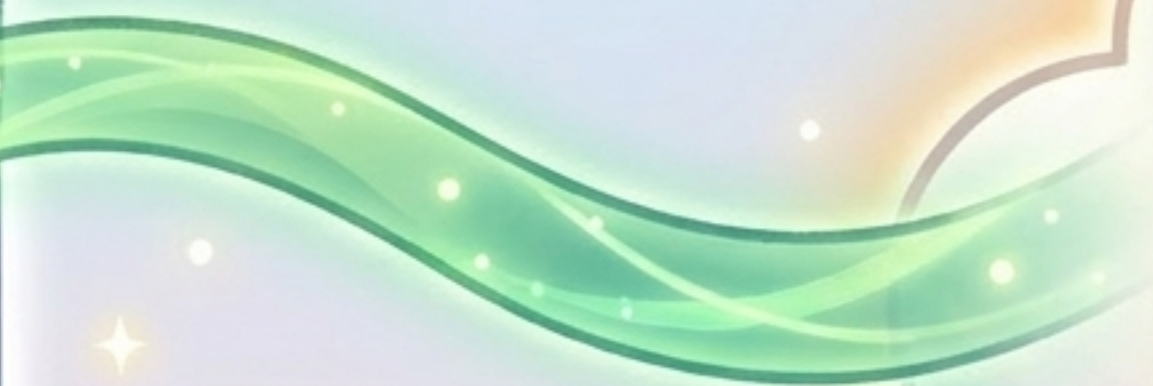


「サーバー側のAIを使いたい
けど、APIキーの管理や
サーバー構築が難しそう…」

「ユーザーのプライバシーデータを
外部に送るのは不安…」

そんな初心者の悩みを一気に解決するのが、
Appleの新しい仕組み「Private Cloud Compute (PCC)」です！

Private Cloud Compute (PCC) ってなに？



Bigger Brain



Privacy

圧倒的な賢さ

複雑な推論や、非常に長い文章（32Kトークン）の処理が得意なサーバーモデルです。

究極のプライバシー

データはリクエスト時のみ使用されます。保存されたり、AppleのAI学習に使われたりすることは絶対にありません（外部検証済み）。

面倒な設定はゼロ

AppleのOSとiCloudに深く統合合されているため、開発者による面倒な認証や「APIキー」の管理は一切不要です！

どっちを使う？オンデバイス vs PCC



特徴	SystemLanguageModel (オンデバイス)	PrivateCloudComputeLanguageModel (PCC)
プライバシー	✓ 完璧（デバイス内で完結）	✓ 完璧（外部検証された安全なサーバー）
オフライン動作	✓ ネットなしで動く	⊘ ネットワーク接続が必須
記憶力（コンテキスト枠）	4Kトークン（短め）	32Kトークン（超長文もOK!）
推論力	標準的	複数のレベルで深く考えられる
利用制限	無制限	1日あたりのユーザー制限あり

スピードやオフライン重視ならオンデバイス、複雑な処理や長文重視ならPCCを使い分けましょう。

導入は魔法のように簡単！「統一API」の威力

Foundation Modelsフレームワークは「統一API」を採用しています。プロンプトの記述やツール呼び出しのコードは一切書き直す必要がありません。モデルの指定を1行変えるだけです。



「えっ、これだけ！？」——
はい、これだけで巨大なサ
ーバーモデルに繋がります！



変更前 - On-Device

```
let session = LanguageModelSession(  
  model: SystemLanguageModel()  
)
```

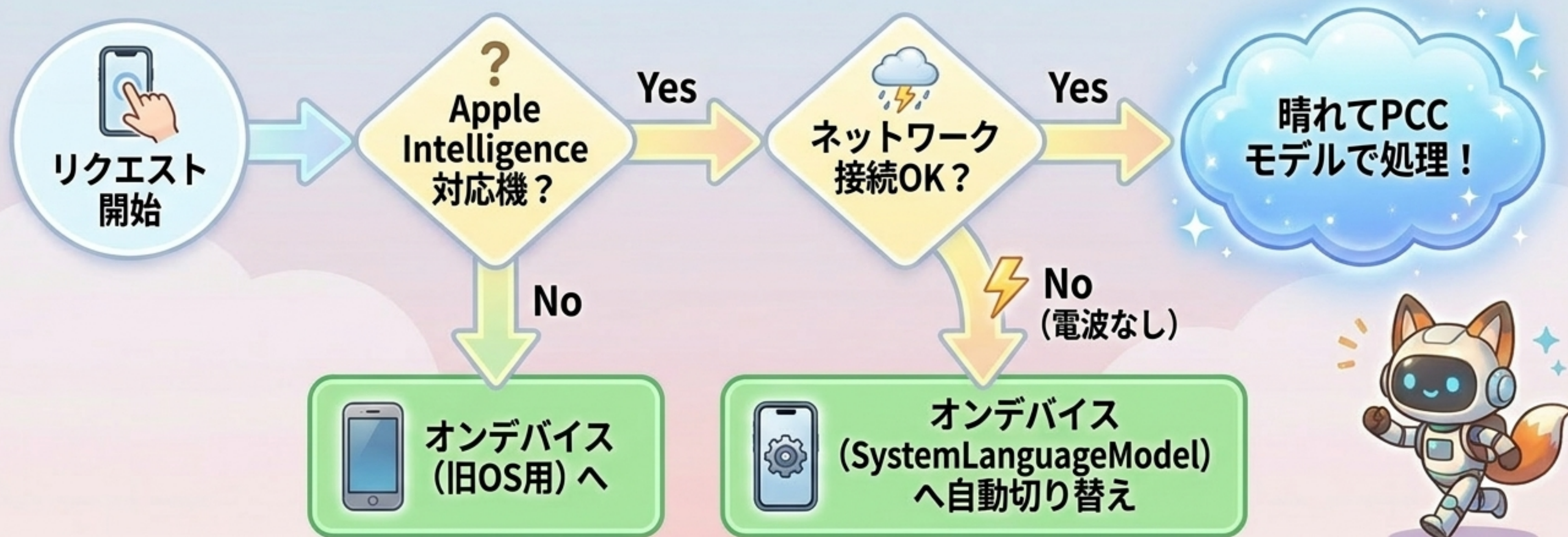


変更後 - PCC

```
let session = LanguageModelSession(  
  model: PrivateCloudComputeLanguageModel()  
)
```

電波がない！PCCが使えない時の「優しいフォールバック」

PCCは強力ですが、オフライン時やネットワークエラー時に備える必要があります。リクエストが失敗した場合は、自動的にデバイス内モデル (SystemLanguageModel) で再試行する安全な設計 (フォールバック) を実装しましょう。



PCCの真骨頂：「考える力」をコントロールする

PCCは答えを出す前に「じっくり考える」ことができます。
タスクの複雑さに合わせて、コードから3段階の賢さを選びましょう。

Light



待ち時間を短くしたい時に。

Moderate



標準的な賢さ。まずはこの
レベルから評価をスタート！

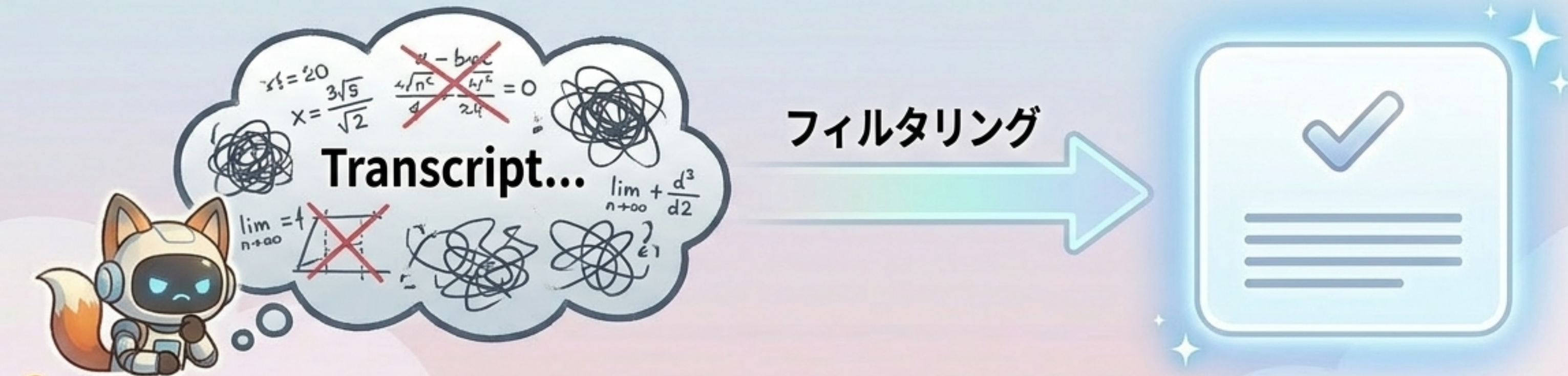
Deep



複雑な制約が絡むタスクに。
時間はかかりますが、見
落としが減ります。

AIの「心の声」を覗いてデバッグしよう

AIがじっくり考えている間（推論中）、裏側では「推論テキスト（Transcript）」が生成されています。この「心の声」はユーザーの最終的な回答には表示されませんが、開発者が確認することで「なぜこの回答になったのか？」のデバッグに役立ちます。



⚠ トークン消費に注意: 深く推論（Deep）するほど、32Kのコンテキスト枠（記憶領域）を消費します。推論レベルと入力文字数のバランスを取りましょう。

使いすぎへの配慮：1日の「利用制限 (Quota)」

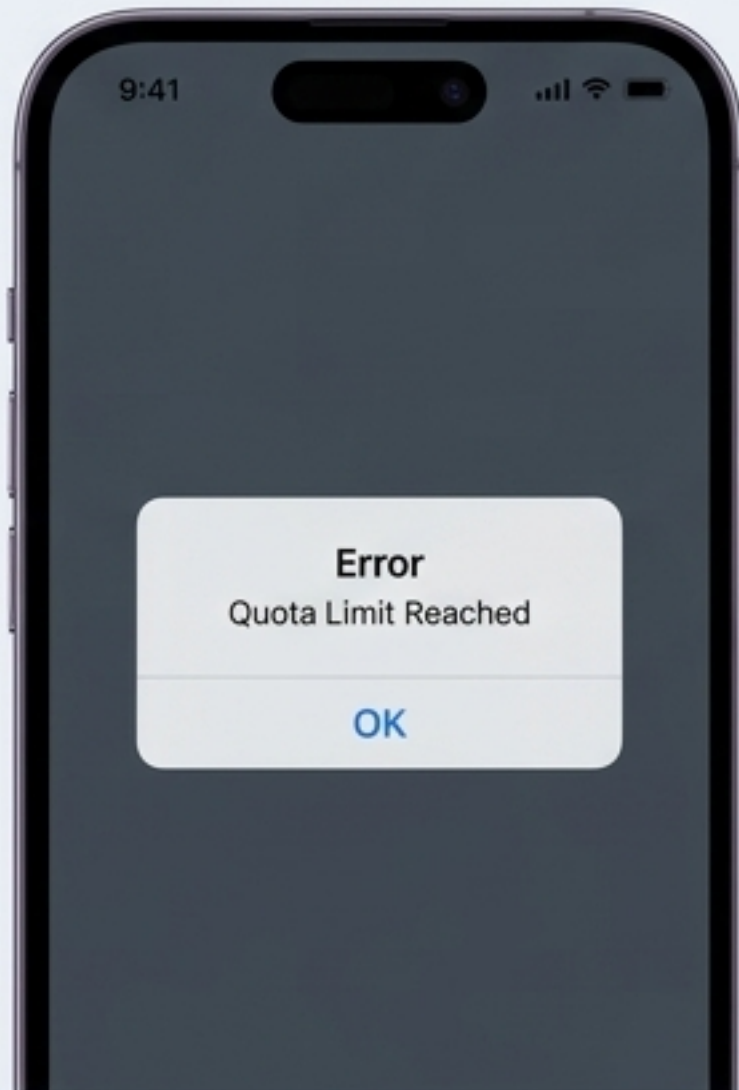


無料で強力なサーバーAIを使える代わりに、ユーザーごとに1日あたりのリクエスト上限が設定されています。上限に達すると `quotaLimitReached` エラーが発生します。ユーザーは翌日のリセット時間を待つか、iCloud+ にアップグレードすることで枠を増やすことができます。

制限に達した時に「どのような体験をさせるか」が、アプリの評価を大きく左右します！

ユーザーに寄り添う、優しいUIデザインを

【NG】無機質なエラー



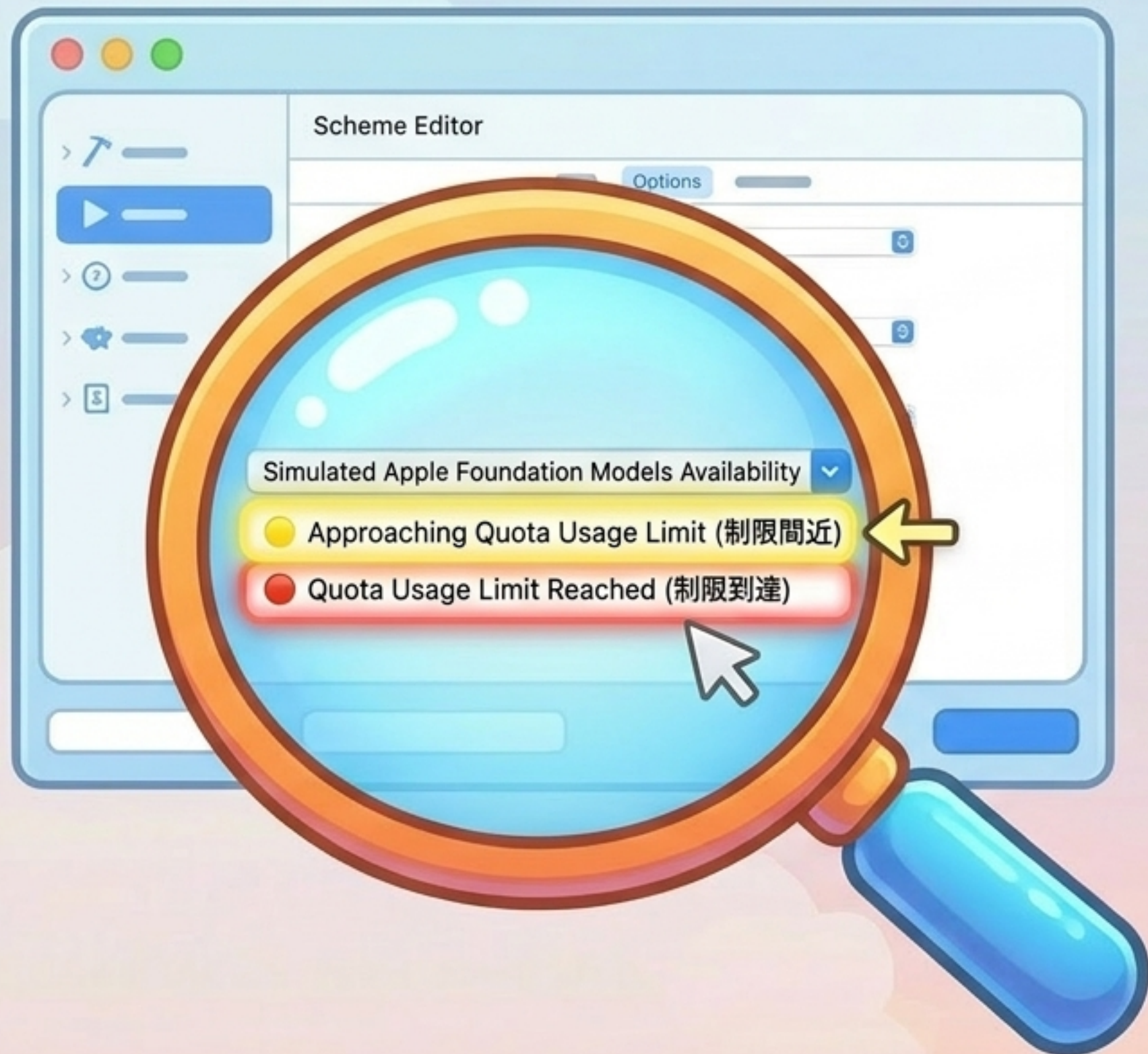
突然のアラート表示。ユーザーは何が起きたか分からず、ただ機能が使えなくなってストレスを感じます。

【OK】自然な案内と解決策



上限に達した理由と、「iCloud+へのアップグレード」という具体的な解決策を親切に提示できます。

Xcodeで「制限到達」のUIを簡単にテスト！



1. Xcodeの Product > Scheme > Edit Scheme を開く。
2. Options タブを選択する。
3. メニューから制限の状態を選ぶだけ！

「実際に1日の上限まで使い切らなくても、ボタン一つでテスト環境を作れるんです。リリース前に必ず両方の状態のUIを確認しましょう！」



開発の黄金ルール：「まずはオンデバイスから」



今日から一歩先のAI体験を作ろう！

📝 PCCは
「プライバシー」と
「賢さ」を両立した
究極のサーバーモデル。

📝 「統一API」の
おかげで、
1行変えるだけで
オンデバイスから
移行可能！

📝 「推論レベル」の
調整や、制限時の
「優しいUI設計」で
最高の体験を。



オンデバイスの軽快さとPCCの圧倒的な賢さを組み合わせて、
あなただけの素晴らしいアプリを作り上げましょう！応援しています！

さらに詳しく学びたい方へ



Apple Developer Documentation:

Adding server-side intelligence
with Private Cloud Compute



WWDC Video:

Build with the new Apple Foundation
Model on Private Cloud Compute

※PCCの開発にはエンタイトルメントのリクエストが必要です。
Developerサイトから申請しましょう。

Thank you! 応援しています！

