

# Microsoft 365 Copilot 「Critique」 「Council」 深掘り分析レポート

## エグゼクティブサマリ

2026年3月30日（米国時間）にMicrosoft <sup>1</sup> が発表した「Critique」と「Council」は、Microsoft 365 Copilotの深掘り調査エージェント「Researcher」に追加された“マルチモデル知能”機能であり、単一モデルに「計画・検索・統合・執筆・検証」を一括で背負わせてきた従来のディープリサーチ系ワークフローを、構造的に再設計する意図が明確です。Critiqueは「生成（執筆）と評価（査読）」を別モデルに分離し、Councilは複数モデルに同一プロンプトを並列投入して“相互比較と合意/乖離の可視化”を前面に出します。

<sup>2</sup>

Critique（Researcherのモデル選択で「Auto」）は、GPT系モデルが一次レポートを生成し、Claude系モデルが“第2の推論パス”として構造・網羅性・引用根拠をチェックして改善する組み込みレビューで、Microsoftは「rubric（採点表）に基づく評価」と「主要主張の厳格な根拠付け（citations）」を中核に据えています。

<sup>3</sup> 公式の性能検証として、Perplexity/学術研究者が提案したDRACOベンチマーク（100タスク・10領域）で、従来の単一モデル方式に対し総合+7.0（SEM±1.90）改善し、論文掲載時点の最良システム（Perplexity Deep Research）を+13.88%上回った、と公表されています。<sup>4</sup>

Council（モデル選択で「Model Council」）は、GPT/Claude双方が“完全な独立レポート”を同時生成し、別の審査（judge）モデルが「一致点・相違点・各モデル固有の貢献」を“カバーレター（要約）”として提示します。これにより、ユーザーは「どこが確からしく、どこが不確実（または解釈が割れている）か」を短時間で識別しやすくなります。<sup>5</sup>

一方で、マルチモデル化は（1）コストとレイテンシの増加、（2）データ移転/データレジデンシの複雑化、（3）責任分界点（誰が何を保証するか）の曖昧化、（4）“AIがAIを評価する”評価層自体の限界、といった論点を同時に増幅します。Reuters <sup>6</sup> やAxiosは「複数モデル併用は信頼性向上の一方でコスト増・遅延要因になり得る」点を明示的に指摘しています。<sup>7</sup> さらに、Claude（Anthropicモデル）利用には管理センターでのサブプロセッサ設定が関与し、EU Data Boundaryの適用範囲外（Anthropicモデルは現状EU Data Boundary/国内処理コミットメントの対象外）など、地域・規制要件の整理が必須です。<sup>8</sup>

導入の実務観点では、Critiqueは「まず“使える叩き台”を素早く作り、引用と論理の穴を減らす」用途に強く、Councilは「結論が割れやすいテーマ（規制、競合比較、投資判断、医療/法務の一次調査など）で“異論の早期発見”を行う」用途に適します。ただし、Researcher自体に月25クエリ/ユーザーという上限があり、運用設計（誰がいつ使うか、どのクエリを“重い調査”に投下するか）を決めないと、価値の高い局面で枠が枯渇しやすい点に注意が要ります。<sup>9</sup>

注記（優先ソースについて）

窓の杜 <sup>10</sup>（Impress Forest）およびGIGAZINE <sup>11</sup>の記事は本文を精読できました。<sup>12</sup> 一方、日経xTECH <sup>13</sup>は当環境から本文へ直接アクセスできず（技術的制約）、一次ソースとしての精読は「不明」とし、公式ドキュメントおよび他の公開ソースで補完しました。

## リリース概要と位置づけ

本リリースは「Researcher—Microsoft 365 Copilotの“仕事のためのディープリサーチエージェント”」を、マルチモデル前提の品質層（Critique）と比較層（Council）で強化するものです。公式ロードマップ上でも「Researcherのマルチモデル知能」として、Critique（GPTの応答をClaudeで洗練）とCouncil（GPTとClaudeの応答を並列比較）という説明が掲げられています。<sup>14</sup>

提供形態は段階的で、Microsoft 365 Copilotの先行機能を試す「Frontier program」で広く利用可能になった、とMicrosoftは述べています。<sup>5</sup> これは、機能やUIが今後変更され得る（experimental/early access）ことを意味し、業務での“本番前提”にはリスク評価が必要です。<sup>15</sup>

Researcherのベース機能としては、「Web+自社業務データ（メール、会議、チャット、ファイル等）から情報を収集・分析し、共有可能なレポートを作る」ことが明示され、レポートには見出し・可視化・引用が含まれ得る、とされています。<sup>16</sup> ここにCritique/Councilが“品質保証（査読）”と“多角比較”として重畳された、というのが今回の本質です。<sup>17</sup>

## 機能説明

### Critique

#### 目的

Critiqueは「複雑な調査タスク」に対し、生成と評価を分離することで、事実正確性・分析の幅/深さ・提示品質・引用品質を構造的に底上げすることを狙います。Microsoftは“学術・専門調査のレビュー工程に近い”として、rubricに基づく評価と、主要主張の保守的な根拠付け（明確な引用）を強調しています。<sup>18</sup>

#### ワークフロー

窓の杜の整理では、片方のモデルが調査と草稿生成を担い、もう片方が内容の評価と改善を担う「執筆と校正の役割分担」です。<sup>19</sup> Microsoft Supportの記述はより具体的で、Anthropicが有効な場合の既定（Auto）としてCritiqueが動作し、「GPTがレポート生成→Claudeが第2推論パスで構造/完全性を補強」する、としています。<sup>15</sup>

#### UI/UX

Researcherエージェント画面のモデルピッカーで「Auto」が選択されている状態がCritique相当で、Frontier参加者はこの新しいAuto体験を利用できます。<sup>20</sup> また、Researcherをエージェント外（Copilotチャット内で@Researcher）で呼び出す場合、当面はモデルピッカーが無効化されAutoが既定になる、とFAQで説明されています。<sup>21</sup>

#### 設定可能なオプション（ユーザー/管理者）

ユーザー側の主要オプションは「Critique（Auto）」「Model Council」 「GPT単独」 「Claude単独」の選択です。<sup>22</sup> さらにResearcher一般として、調査範囲を“Work（業務データ）”と“Web（公開情報）”でスコープでき、必要に応じて追加質問（clarifying questions）が入る、とされています。<sup>23</sup>

管理者側では、（1）Anthropic（Claude）をサブプロセッサとして有効化するトグル、（2）Web検索の許可/禁止（無効ならResearcherはWebを使わない）、（3）エージェント（Researcher/Analyst）のプロック、といった統制点が公式に示されています。<sup>24</sup>

## 出力形式

Researcherの出力は「共有可能な構造化レポート」で、可視化（図表/グラフ）、セクション構成、出典（引用）を含み得ます。最終レポートは編集・共有・文書/プレゼン作成の起点として使える、とMicrosoft Supportが説明しています。<sup>25</sup> Critiqueでは、レビュー層が“信頼できるソースを優先し、重要な記述が明確な引用で根拠付けられていること”を確認する、とされています。<sup>26</sup>

## ログ/監査、データ保持・プライバシー挙動

Researcherエージェントは、Microsoft 365 Copilotと同じデータ保護方針・DLP・責任あるAI運用に従い、テナントデータで学習しない、とFAQで明記されています。<sup>27</sup>

ただし、**Researcherセッションのプロンプト/出力を管理者がeDiscovery等で直接閲覧できるか**については重要な注意点があります。Researcher FAQは「既定では不可（内容は管理者/コンプライアンスツールから直接アクセス不能）、利用回数などの利用指標は見える。例外はユーザーがフィードバックとしてセッションデータを明示送信した場合」と述べています。<sup>21</sup> 一方、Microsoft 365 Copilot全般については、Word等でのCopilot操作でプロンプト/応答（引用を含む）が“Copilot activity history”として保存され、Purview/コンテンツ検索/保持ポリシーで管理できる、という説明も存在します。<sup>28</sup>

このため、「**Researcher固有のセッションデータは既定で保持・eDiscovery対象外だが、生成物をファイルとして保存/共有すれば通常のM365コンテンツとして保持対象になる**」という整理が最も整合的ですが、どのイベント/データがどの経路で保持されるかは運用前に検証（テナントでのPurview検索・保持ポリシー動作確認）が必要です（詳細は後述の推奨項目参照）。<sup>29</sup>

## Council

### 目的

Councilは「複数モデルの視点差・結論差を並べ、どこが一致しどこが割れるか」を可視化することで、意思決定の確信度（confidence）を上げることを狙います。ロードマップおよびMicrosoft公式ブログは「GPTとClaudeの応答を横並びに比較できる」と明確に述べています。<sup>14</sup>

### ワークフロー

窓の杜の説明では、OpenAI系モデルとAnthropic系モデルがそれぞれレポートを生成し、専門の審査員モデルが評価して「一致点・相違点・固有の貢献」を指摘します。<sup>19</sup> Microsoft公式ブログも同旨で、2つの独立レポート生成後にjudgeモデルが差分要約（カバーレター）を生成するとしています。<sup>30</sup>

### UI/UXと出力

Model CouncilはResearcherのモデルドロップダウンから選択し、同一プロンプトを同時に走らせます。出力は各モデルの“フルレポート”が保持され、その上に一致点/差分/独自洞察をまとめたサマリが付与される、とMicrosoft Supportが説明します。<sup>22</sup> GIGAZINE<sup>11</sup>もCouncilモードの出力例として、GPTとClaudeの生成結果が比較形式で出力される様子を紹介しています。<sup>31</sup>

## ログ/監査、データ保持・プライバシー挙動

Council自体の保持・監査仕様が“Council専用追加された”という一次情報は確認できません。一方で、ResearcherがMicrosoft 365 Copilotと同等のデータ保護方針に従う点、Anthropic利用時はサブプロセッサートグルの統制下に置かれる点は共通して適用されます。<sup>32</sup>

# マルチモデル統合

## 全体アーキテクチャ

Microsoft 365 Copilotは「Copilot orchestrator」がプロンプト処理を統括し、責任あるAI (RAI) 安全性チェック、必要に応じたWeb検索、LLM呼び出し、ログ/監査向け記録などを調整する、という説明がMicrosoft Learnにあります (Copilot Chatの説明だが“同じorchestratorを使う”と明記)。<sup>33</sup> Researcherは、ここに“深い調査の計画と反復的な検索・統合”を組み込んだエージェントで、Webと業務データを横断し得ます。<sup>34</sup>

Critique/Councilは、このオーケストレーション層の上で「モデル呼び出しの形態」を変える実装だと捉えられます。すなわち、Critiqueは“直列2段 (生成→レビュー) ”、Councilは“並列3段 (複数生成→審査・要約) ”です。<sup>35</sup>

flowchart TB

```
subgraph Critique["Critique (Auto) : 生成と評価の分離 (直列) "]
    U1[ユーザー: 調査プロンプト] --> O1[Copilot Orchestrator]
    O1 --> R1[情報収集: Work + Web (スコープ可能) ]
    R1 --> G1[生成モデル: GPT系\n計画→検索反復→草稿レポート]
    G1 --> V1[レビューモデル: Claude系\nrubric評価→根拠/引用チェック→改善]
    V1 --> OUT1[最終レポート\n(見出し/引用/可視化を含み得る) ]
end

subgraph Council["Council (Model Council) : 多角比較 (並列) "]
    U2[ユーザー: 同一プロンプト] --> O2[Copilot Orchestrator]
    O2 --> R2[情報収集: Work + Web (スコープ可能) ]
    R2 --> G2a[モデルA: GPT系\n独立レポート]
    R2 --> G2b[モデルB: Claude系\n独立レポート]
    G2a --> J2[審査/要約モデル\n一致点・相違点・独自貢献を要約]
    G2b --> J2
    J2 --> OUT2[比較UI\n(A/Bレポート + カバーレター) ]
end
```

上図のうち、Critiqueが「GPT生成+Claude第2推論パス」であること、Councilが「複数モデル同時実行+要約付与」であることは、Microsoft Supportおよび窓の杜・公式ブログで一致しています。<sup>35</sup>

## 同一プロンプト同時実行の仕組み

Councilでは“同一の質問”を複数の深い推論エージェントに同時投入し、各モデルがスタンドアロンの完全レポートを生成して保持します。これは、単に“短い回答を並べる”のではなく、各モデルが独立に探索・構造化・引用付けを行う前提です。<sup>26</sup> その上でjudgeモデルが差分と合意点を抽出するため、ユーザーに提示される情報は「①モデルAレポート」「②モデルBレポート」「③差分サマリ」の三層になります。<sup>17</sup>

## ルーティング、フェイルオーバー、モデル選択ポリシー

**ユーザー選択:** Researcher内ではモデルピッカー (右上) でモード/モデルを選べます。<sup>36</sup>

**既定ポリシー:** Anthropicが有効な場合、AutoはCritiqueを自動利用する既定動作です。<sup>15</sup>

**管理者統制:** Anthropicモデルは管理センターでサブプロセッサとして有効/無効を制御し、EU/EFTA/UKでは既定オフ、政府クラウド等では利用不可です。<sup>37</sup> Web検索もテナント設定で無効化でき、無効ならResearcherはWebデータを使わないとされています。<sup>21</sup>

**フェイルオーバー (不明):** あるモデルがエラー/利用不可の場合にどうフォールバックするか (例: Councilで片系だけ成功した場合の表示、Critiqueでレビューが落ちた場合の扱い) は公式に詳細が出ていません。従って“確定情報として”は不明です。

ただしMicrosoftは、生成役/レビュー役の「役割を入れ替え可能な設計余地 (optionality)」や、将来的に“逆方向 (GPTがClaudeをレビュー)”も計画していると述べています。これらは、複数モデルの組み合わせや役割割当を柔軟に変える前提の設計思想を示唆します。<sup>38</sup>

## コスト/レイテンシの含意

**レイテンシ:** Researcherの応答時間は、簡単なクエリで5分未満、複雑なものでは10~45分とされます。<sup>21</sup> Councilは「2系統の独立レポート+judge要約」を含むため、単一モデルより計算量が増えるのが自然で、Axiosも“複数モデル併用はコスト増・応答遅延要因になり得る”趣旨を述べています。ただしモード別の実測比較は公式に示されておらず、厳密な増分は不明です。<sup>39</sup>

**コスト:** Microsoftは価格内訳 (モデル呼び出し回数やトークン課金の内部配賦) を公開していません。外形的にはResearcherの利用枠 (1ユーザー月25クエリ) が“高コストな深い推論タスク”の予算制御として機能します。<sup>40</sup> 一方で、Reutersは“モデルを協調させることで幻覚を抑え、生産性と品質を上げる”とMicrosoft側コメントを報じています。つまり、クエリ単価は上がっても“人手の再調査/手戻り”が減れば総コストは下がり得るという価値仮説です。<sup>41</sup>

## DRACOベンチマークに基づく技術的評価

DRACOはPerplexityと学術研究者が2026年2月に提示した“深い調査エージェント”向けベンチマークで、実利用由来の複雑タスクを匿名化し、タスク固有rubricで「正確性」「分析の幅/深さ」「提示品質 (客観性を含む)」「引用品質」の4軸採点を行うとされています。<sup>42</sup>

MicrosoftはCritiqueをDRACOで評価し、OpenAIのGPT-5.2をジャッジモデルとして同論文と同一プロトコルで比較した、と説明しています。<sup>30</sup> 結果として、総合スコアで+7.0 (SEM±1.90) の改善、特に「分析の幅/深さ」「提示品質」「事実正確性」で大きい改善があった、と公表されています。<sup>43</sup>

## 運用パターン

### 役割分担

Critiqueは「執筆 (生成) と検証 (レビュー)」を別モデルで分担するため、実務上は“人間の役割分担”に近い形で組み込みレビューが発生します。窓の杜もこの点を「執筆と校正の役割分担」と表現しています。<sup>19</sup> 推奨運用としては、AI内の役割分担 (Critique) +人間の最終ゲート (承認者) を明示するのが現実的です。Critiqueで“引用の形”は整っても、引用先の内容一致や前提条件の妥当性は、人間によるチェックがなお必要です (監査・責任の観点でも)。<sup>44</sup>

## 競わせるワークフロー

Councilは「モデルAとモデルBが独立にレポートを出し、judgeが一致点/相違点を要約する」ため、次のような“競争→差分レビュー”が有効です。 <sup>5</sup>

- 仕様/規制解釈のように見解が割れやすいテーマでは、差分の出た箇所に人間の検証工数を集中させる（全体を読む前に“争点”を抽出する）。 <sup>45</sup>
- 競合分析や投資判断では、同じ事実でも“フレーミング（強調点）”が違ふことが多く、Councilの“乖離”を意思決定会議の論点整理に転用できる。 <sup>46</sup>

## 導入シナリオと推奨プロセス

Researcherは「短時間のチャット回答」ではなく「共有可能な調査レポート」を作る設計で、時間をかけて深い洞察を出すとして明示されています。 <sup>23</sup> したがって導入時は、次のように“用途の棲み分け”を明文化すると運用が安定します。

- 日常タスク（メール文案、軽い要約）は通常のCopilot Chat/アプリ内Copilotで対応し、Researcher枠（25クエリ/月）を消費しない。 <sup>23</sup>
- 研究/調査の“成果物（社内共有・意思決定資料）”が必要なときだけResearcherを使い、Critiqueで初稿品質を上げる。 <sup>47</sup>
- 経営判断・規制・高リスク領域ではCouncilを“争点抽出装置”として使い、差分箇所に人手レビューを集中させる。 <sup>17</sup>

## 評判・反応と比較

### 評判・反応の要約

メディア報道では、窓の杜が「単一モデルが一般的だった調査エージェントに対し、2モデルで役割分担（Critique）と競争（Council）を導入した」と整理し、品質/深度/信頼性向上への期待を示しています。 <sup>19</sup> Reuters <sup>6</sup> も、GPTが生成しClaudeがレビューすることで幻覚を抑え、いずれ双方向レビューも目指すと報じました。 <sup>48</sup>

一方で懸念としては、Axiosが「複数モデル併用は信頼性を高め得るが、コスト増と遅延を招く可能性」を指摘しています。 <sup>49</sup> また、Claude統合に関してはDirections on Microsoftが「契約上の保護があってもAzure外（例: AWS等）へデータが転送され得る点が、規制・社内要件で問題になり得る」と論点化しています（※同記事は分析レポートであり、技術実装の最終形はMicrosoftの公式記述に従うべき）。 <sup>50</sup>

SNS/フォーラムでは、Microsoft 365×外部モデル統合に対し「データが学習に使われないか」「どの境界内に留まるのか」を気にする問いが継続的に見られます（例: Redditでの“Microsoft 365テナントとClaude連携時のデータ利用”懸念投稿）。 <sup>51</sup> これは、Critique/Councilが“より強く業務データを扱う深い調査”であるほど、同様の懸念が増幅することを示します。 <sup>52</sup>

### 従来Copilot機能との違い（比較表）

観点	従来Researcher（単一モデル運用）	Critique（Auto）	Council（Model Council）
目的	深掘り調査レポート生成（深さ重視） <sup>25</sup>	生成と評価を分離し、品質を底上げ <sup>45</sup>	複数モデルの見解差を可視化し、確信度を上げる <sup>5</sup>

観点	従来Researcher (単一モデル運用)	Critique (Auto)	Council (Model Council)
モデル実行形態	1モデルが計画～執筆を一括処理 (一般形) <sup>30</sup>	直列2段: GPT生成 → Claudeレビュー (第2推論パス) <sup>15</sup>	並列: GPT/Claudeが独立レポート → judge要約 <sup>17</sup>
UI	Researcherでレポート出力 <sup>25</sup>	モデルピッカー「Auto」既定動作 <sup>20</sup>	モデルピッカー「Model Council」選択 <sup>20</sup>
出力	構造化レポート (引用・可視化を含み得る) <sup>25</sup>	改善済み最終レポート (根拠・構造を強化) <sup>26</sup>	2本の完全レポート＋一致/差分のカバーレター <sup>45</sup>
導入条件	Microsoft 365 Copilotライセンス、Researcher有効化 <sup>25</sup>	Frontier参加が必要 (少なくとも新Auto体験) <sup>15</sup>	Frontier参加が必要 <sup>15</sup>
想定レイテンシ	単一モデルより相対的に短い可能性 (ただし深い調査は単一位) <sup>21</sup>	レビュー段追加で増加し得る (増分は不明) <sup>39</sup>	2レポート＋審査で増加し得る (増分は不明) <sup>39</sup>

### 他社類似機能との比較 (比較表)

観点	Microsoft 365 Copilot Researcher (Critique/Council)	OpenAI Deep Research (ChatGPT)	Google Gemini Deep Research	Anthropic (Claude APIの Citations等)
“深い調査レポート”機能	Work+Web横断でレポート生成 <sup>16</sup>	Web/ファイル/アプリを使い文書化レポートを生成 <sup>53</sup>	Workspace データとWebを使いレポート生成 <sup>54</sup>	ツール連携・文書入力に対し回答生成 (“深い調査UI”は製品依存: 不明) <sup>55</sup>
生成と検証の分離 (内蔵)	Critiqueとして標準搭載 (GPT生成 → Claudeレビュー) <sup>26</sup>	公式に“別モデルが査読”というUIは確認できず (同一機能としては不明)。ただしAPIガイドでは中間モデルによる意図明確化・プロンプト再構成など多段処理が説明される <sup>56</sup>	公式に“別モデルが査読”というUIは不明 <sup>57</sup>	Citationsにより根拠箇所を機械可視化 (検証支援)。“査読役モデル”の内蔵は不明 <sup>58</sup>
同一プロンプトのモデル比較UI	Councilで標準搭載 (2本の独立レポート＋差分要約) <sup>35</sup>	標準UIとしての“同時並列比較”は不明 (手動で複数回実行は可能) <sup>59</sup>	公式に“モデル並列比較UI”は不明 <sup>57</sup>	API/アプリ次第 (Anthropic公式としては不明) <sup>60</sup>

観点	Microsoft 365 Copilot Researcher (Critique/Council)	OpenAI Deep Research (ChatGPT)	Google Gemini Deep Research	Anthropic (Claude APIの Citations等)
監査・保持 (企業向け)	Purview中心の監査/保持。Researcherセッション内容のeDiscovery可否は既定で不可とFAQに明記 (要検証) <sup>61</sup>	企業向け監査・保持はプラン/設定依存で本レポート範囲では不明 <sup>59</sup>	同上 (不明) <sup>57</sup>	同上 (不明) <sup>62</sup>

## リスク評価とガバナンス

### リスクの分類

#### 誤情報 (ハルシネーション) と“誤引用”

Critiqueは“主要主張を信頼できるソースと明確な引用にアンカーする”ことをレビューレイヤーの中心に置きますが、DRACO自体が示す通り、深い調査は多段の検索・統合・解釈を含み、完全な自動検証は難しい領域です。<sup>63</sup> また、Microsoft自身も「信頼できるソース優先」「引用で根拠付け」を繰り返し強調しており、ユーザー側の検証行為が前提に組み込まれた設計思想だと読めます。<sup>34</sup>

#### バイアス・客観性

DRACOは提示品質の軸に“objectivity”を含めていますが、評価がrubric (規範) に依存する以上、どの観点を重視するかで結果は変わり得ます。<sup>64</sup> Councilで差分が出た場合、それが“モデルの偏り”なのか“根拠の違い”なのか“解釈の分岐”なのかを見極めるのは人間側の責務です。<sup>17</sup>

#### 責任所在 (誰が保証するか)

Copilotが生成した内容の最終的な採用判断はユーザー/組織が負うべきであり、Microsoftも安全対策 (有害コンテンツ、プロンプトインジェクション等) を挙げつつ“信頼できる出力”のための統制を強調しています。<sup>65</sup>

#### コスト・性能変動

複数モデル同時実行はコスト増・遅延の可能性があり、Axiosが明示しています。<sup>49</sup> またFrontierの実験的機能であるため仕様変更リスクがあり、運用での“前提”が崩れる可能性がある点もSupportが注意しています。<sup>15</sup>

#### セキュリティ (情報漏えい・制御逸脱)

CopilotはDLPや権限モデルを尊重し、個々のユーザーが閲覧権限を持つ組織データのみを参照する、とMicrosoftは説明します。<sup>66</sup> 一方で、実装バグ等で意図せず機密メールが要約対象になった事例が報じられており、“理論上の統制”と“実運用の安全”の間にギャップが生じ得る点は無視できません。<sup>67</sup>

#### データレジデンシ/規制適合 (特にAnthropic)

Anthropicモデルは現状EU Data Boundaryの対象外で、EU/EFTA/UKでは既定無効、政府/ソブリンクラウドでは利用できません。<sup>8</sup> これは、Council/Critiqueの価値が高い“規制・法務・医療”領域ほど、データ処理境界の説明責任が重くなることを意味します。<sup>68</sup>

## 推奨ガバナンス（社内ポリシー案）

### 1) 用途区分（許容/要承認/禁止）を先に決める

深い調査レポート（Researcher）を「意思決定資料」に使う場合、最低限“引用必須”“重要主張は二次確認”“最終責任者の明記”をルール化し、Councilを“争点抽出”に使う手順も定義するのが妥当です。 5

### 2) データ境界の説明と同意（特にAnthropic）

Anthropic有効化は管理者トグルで制御でき、地域によって既定が異なります。導入時は“データがどの処理境界で扱われるか（EU Data Boundary適用可否等）”を法務・セキュリティ・プライバシー担当と合意し、必要ならAnthropicを無効化する判断も含めるべきです。 69

### 3) 監査・保持の整備と“見える化”

Copilot/AIアプリの監査ログはPurviewで扱え、ユーザーがいつどこでCopilotと対話し、どのリソース参照があったか等のメタデータが残るとされています。 70 ただしResearcherの会話内容は既定でeDiscovery対象外というFAQ記述があるため、“成果物はファイルとして保存し、通常のM365コンテンツとして保持・レビューできるようにする”運用が有効です。 71

### 4) インシデント前提の設計（DLP/機密ラベル/最小権限）

Copilotはユーザー権限を尊重し、感度ラベル・保持・監査に従う（EDPの説明）ため、まず既存の権限設計・ラベル設計を整えることが先決です。 72

## 実装上の推奨と参考ソース

### 導入手順（推奨）

**管理者準備** - Microsoft 365 Copilotライセンスと、Frontier program要件（Auto/Model Councilの新機能）の充足確認。 73

- Anthropic（Claude）を利用する場合、Microsoft 365 管理センターで「Anthropic as a subprocessor」を有効化（EU/EFTA/UKは既定オフ、政府/ソブリンクラウド不可）。 37

- Web検索の可否を決め、テナント設定に反映（無効ならResearcherはWeb情報を使わない）。 21

- Purviewで監査と保持の方針を設計（監査ログの保持期間、保持ポリシーでCopilotデータを何年残すか）。 74

**パイロット** - “高価値クエリ”を定義し、月25クエリ上限の中でResearcher利用を最適化（例：顧客/競合調査、規制一次調査、提案書の根拠集め）。 40

- Critiqueを標準、Councilを高リスク/見解分岐が想定されるテーマに限定する、といった利用基準を作る。 26

- 出力は“そのまま貼らない”運用（最終成果物はWord/PowerPoint等に移し、引用確認チェックを通す）を徹底。 75

### テスト計画（最小構成）

- **品質テスト:** 既知の正解がある題材で、単一モデル / Critique / Councilの差分（誤り率、引用の妥当性、抜け漏れ）を比較。 64
- **セキュリティ/DLPテスト:** 機密ラベル付き文書やDLP対象データで、期待通り参照/要約がブロックされるか確認（過去にDLP回避バグが報じられたため回帰テスト推奨）。 76
- **保持・監査テスト:** Purview監査ログでResearcher利用がどこまで追跡できるか、保持ポリシーがCopilotデータに適用されるか、Researcher FAQの“eDiscovery不可”と実態が整合するかを検証。 77

- **レイテンシ測定:** 代表クエリで平均/分位 (P50/P95) を測定 (Researcher自身が5~45分レンジのため、SLAを現実的に)。 78

## KPI/評価指標 (概算設計)

- **再調査・手戻り削減:** “初稿→承認までの修正回数”と“引用不備の指摘回数”。 26
- **時間削減:** 1レポート当たりの人手時間 (調査・引用付け・構成)。 25
- **リスク指標:** DLPブロック件数、機密ラベル違反疑い件数、インシデント報告件数。 79
- **利用効率:** 月25クエリ枠の消費内訳 (高価値用途に集中しているか)。 40

## コスト見積りの考慮点 (概算)

- **直接コスト:** MicrosoftはCritique/Council単体の従量課金を公開していないため、基本はCopilotライセンス費用と“枠 (25クエリ/月)”を前提に算定します。 40
- **間接コスト:** Councilは複数レポート生成+審査要約のため、単一モデルより計算資源を使う可能性が高く、Axiosも複数モデル併用がコスト増につながり得ると指摘します。 80
- **監査保持のライセンス:** 監査ログ保持を180日超~最大10年に伸ばす場合、Purview Audit Premiumや追加ライセンスが必要になり得ます (保持ポリシーの要件説明)。 81
- **データレジデンシ対応:** EU圏などでAnthropicを無効化する場合、Critique/Councilの価値が一部減衰する可能性があるため、代替 (単一モデル+人手レビュー増) コストを見積もりに含めます。 68

## 移行リスク

- Frontierの実験的機能であり、UI・動作・提供条件が変わり得ます (Supportが明示)。従って“業務クリティカルなプロセスへ直結させる”より、当面は情報収集・資料作成の前工程 (下書き/論点抽出) に限定するのが安全です。 73
- “モデルが増える”方向性が示されており (夏頃までに増える旨のコメント報道)、モデル追加で比較・選択のガバナンス負荷が上がるリスクがあります。 82

## 主要ソースURL (公式優先)

### 【Microsoft公式 (機能・UX)】

<https://techcommunity.microsoft.com/blog/microsoft365copilotblog/introducing-multi-model-intelligence-in-researcher/4506011>  
<https://support.microsoft.com/en-gb/topic/use-model-choice-in-the-researcher-agent-cf182434-02b7-4d6f-af25-c50111fc6bf6>  
<https://support.microsoft.com/en-us/topic/get-started-with-researcher-in-microsoft-365-copilot-e63ab760-f3de-4c47-ae87-dad601b0e9c4>  
<https://www.microsoft.com/microsoft-365/roadmap?id=558538>  
<https://learn.microsoft.com/en-us/microsoft-365/copilot/faq-researcher>  
<https://learn.microsoft.com/en-us/microsoft-365/copilot/microsoft-365-copilot-privacy>  
<https://learn.microsoft.com/en-us/microsoft-365/copilot/enterprise-data-protection>  
<https://learn.microsoft.com/en-us/microsoft-365/copilot/connect-to-ai-subprocessor>  
<https://learn.microsoft.com/en-us/purview/audit-copilot>  
<https://learn.microsoft.com/en-us/purview/retention-policies-copilot>

### 【優先メディア (提示記事)】

<https://forest.watch.impress.co.jp/docs/news/2097617.html>  
[https://gigazine.net/gsc\\_news/en/20260331-microsoft-365-copilot-critique-council/](https://gigazine.net/gsc_news/en/20260331-microsoft-365-copilot-critique-council/)

### 【ベンチマーク/研究】

<https://arxiv.org/abs/2602.11685>  
<https://research.perplexity.ai/articles/evaluating-deep-research-performance-in-the-wild-with-the-draco-benchmark>  
<https://huggingface.co/datasets/perplexity-ai/draco>

【関連：他社類似機能（比較用）】

<https://help.openai.com/en/articles/10500283-deep-research-in-chatgpt>  
<https://openai.com/index/introducing-deep-research/>  
<https://developers.openai.com/api/docs/guides/deep-research/>  
<https://gemini.google/overview/deep-research/>  
<https://ai.google.dev/gemini-api/docs/deep-research>  
<https://docs.anthropic.com/ja/docs/build-with-claude/citations>

【報道・論点整理（コスト/遅延・外部視点）】

<https://www.axios.com/2026/03/31/microsoft-critique-anthropic-openai>  
<https://www.reuters.com/business/microsoft-unveils-ai-upgrades-rolls-out-copilot-cowork-early-access-customers-2026-03-30/>

- 
- 1 16 23 25 34 52 75 <https://support.microsoft.com/en-us/topic/get-started-with-researcher-in-microsoft-365-copilot-e63ab760-f3de-4c47-ae87-dad601b0e9c4>  
<https://support.microsoft.com/en-us/topic/get-started-with-researcher-in-microsoft-365-copilot-e63ab760-f3de-4c47-ae87-dad601b0e9c4>
- 2 3 4 5 11 13 17 18 30 38 43 44 45 63 <https://techcommunity.microsoft.com/blog/microsoft365copilotblog/introducing-multi-model-intelligence-in-researcher/4506011>  
<https://techcommunity.microsoft.com/blog/microsoft365copilotblog/introducing-multi-model-intelligence-in-researcher/4506011>
- 6 10 39 49 80 82 <https://www.axios.com/2026/03/31/microsoft-critique-anthropic-openai>  
<https://www.axios.com/2026/03/31/microsoft-critique-anthropic-openai>
- 7 41 48 <https://www.reuters.com/business/microsoft-unveils-ai-upgrades-rolls-out-copilot-cowork-early-access-customers-2026-03-30/>  
<https://www.reuters.com/business/microsoft-unveils-ai-upgrades-rolls-out-copilot-cowork-early-access-customers-2026-03-30/>
- 8 24 37 68 69 <https://learn.microsoft.com/en-us/copilot/microsoft-365/connect-to-ai-subprocessor>  
<https://learn.microsoft.com/en-us/copilot/microsoft-365/connect-to-ai-subprocessor>
- 9 <https://www.microsoft.com/en-us/microsoft-365/blog/2025/06/02/researcher-and-analyst-are-now-generally-available-in-microsoft-365-copilot/>  
<https://www.microsoft.com/en-us/microsoft-365/blog/2025/06/02/researcher-and-analyst-are-now-generally-available-in-microsoft-365-copilot/>
- 12 19 46 <https://forest.watch.impress.co.jp/docs/news/2097617.html>  
<https://forest.watch.impress.co.jp/docs/news/2097617.html>
- 14 <https://www.microsoft.com/microsoft-365/roadmap?id=558538>  
<https://www.microsoft.com/microsoft-365/roadmap?id=558538>
- 15 20 22 26 35 47 73 <https://support.microsoft.com/en-gb/topic/use-model-choice-in-the-researcher-agent-cf182434-02b7-4d6f-af25-c50111fc6bf6>  
<https://support.microsoft.com/en-gb/topic/use-model-choice-in-the-researcher-agent-cf182434-02b7-4d6f-af25-c50111fc6bf6>

- 21 27 29 32 36 40 61 71 76 78 <https://learn.microsoft.com/en-us/copilot/microsoft-365/faq-researcher>  
<https://learn.microsoft.com/en-us/copilot/microsoft-365/faq-researcher>
- 28 65 66 <https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-privacy>  
<https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-privacy>
- 31 [https://gigazine.net/gsc\\_news/en/20260331-microsoft-365-copilot-critique-council/](https://gigazine.net/gsc_news/en/20260331-microsoft-365-copilot-critique-council/)  
[https://gigazine.net/gsc\\_news/en/20260331-microsoft-365-copilot-critique-council/](https://gigazine.net/gsc_news/en/20260331-microsoft-365-copilot-critique-council/)
- 33 <https://learn.microsoft.com/en-us/copilot/privacy-and-protections>  
<https://learn.microsoft.com/en-us/copilot/privacy-and-protections>
- 42 64 <https://arxiv.org/abs/2602.11685>  
<https://arxiv.org/abs/2602.11685>
- 50 M365 Copilot Adds Choice (and Risk) with Anthropic's Claude - Directions on Microsoft  
<https://www.directionsonmicrosoft.com/reports/m365-copilot-adds-choice-and-risk-with-anthropics-claude/>
- 51 Any data privacy concerns integrating Claude with a Microsoft 365 tenant? : r/Anthropic  
[https://www.reddit.com/r/Anthropic/comments/1rq0omo/any\\_data\\_privacy\\_concerns\\_integrating\\_claude\\_with/](https://www.reddit.com/r/Anthropic/comments/1rq0omo/any_data_privacy_concerns_integrating_claude_with/)
- 53 59 <https://help.openai.com/en/articles/10500283-deep-research-in-chatgpt>  
<https://help.openai.com/en/articles/10500283-deep-research-in-chatgpt>
- 54 <https://gemini.google/overview/deep-research/>  
<https://gemini.google/overview/deep-research/>
- 55 58 62 <https://docs.anthropic.com/en/docs/build-with-claude/citations>  
<https://docs.anthropic.com/en/docs/build-with-claude/citations>
- 56 <https://developers.openai.com/api/docs/guides/deep-research/>  
<https://developers.openai.com/api/docs/guides/deep-research/>
- 57 <https://ai.google.dev/gemini-api/docs/deep-research>  
<https://ai.google.dev/gemini-api/docs/deep-research>
- 60 <https://docs.anthropic.com/en/docs/build-with-claude/tool-use>  
<https://docs.anthropic.com/en/docs/build-with-claude/tool-use>
- 67 <https://www.techradar.com/pro/security/microsoft-admits-an-office-bug-exposed-confidential-user-emails-to-copilot>  
<https://www.techradar.com/pro/security/microsoft-admits-an-office-bug-exposed-confidential-user-emails-to-copilot>
- 70 74 77 <https://learn.microsoft.com/en-us/purview/audit-copilot>  
<https://learn.microsoft.com/en-us/purview/audit-copilot>
- 72 79 <https://learn.microsoft.com/en-us/copilot/microsoft-365/enterprise-data-protection>  
<https://learn.microsoft.com/en-us/copilot/microsoft-365/enterprise-data-protection>
- 81 <https://learn.microsoft.com/en-us/purview/audit-log-retention-policies>  
<https://learn.microsoft.com/en-us/purview/audit-log-retention-policies>