

Microsoft 365 Copilotにおけるマルチモデル推論システム「Critique」および「Council」の戦略的意義と技術的評価

Gemini 3.1 pro

エンタープライズAIにおけるパラダイムシフトと新たな潮流

2026年3月30日(米国時間)、Microsoftはエンタープライズ向けAIアシスタントの次なる進化として、「Microsoft 365 Copilot Wave 3」の一部であるリサーチ特化型エージェント「Researcher」の大規模なアップデートを発表した¹。このアップデートの中核をなすのが、「Critique(批判的レビュー)」および「Council(評議会)」と名付けられた2つの新しいマルチモデルインテリジェンス機能である²。これらの機能は、生成AI(Generative AI)の活用法を根本から再定義するものであり、単一のAIモデルに依存して回答を生成する従来のアプローチから、複数の最高峰のAIモデルを連携・競合させる「マルチモデル・オーケストレーション」への転換を意味している⁴。

これまで、AI業界における競争の主軸は、いかに巨大で高性能な単一の基盤モデル(Foundation Model)を開発するかに置かれていた。しかし、エンタープライズ環境における高度な調査・研究(ディープリサーチ)業務においては、単一モデルが生成する「もっともらしく洗練された単一の回答」は、時に事実誤認(ハルシネーション)を含み、複雑な意思決定の基盤としては不十分であることが明らかになってきた⁴。Microsoftは2025年9月の段階で、すでにResearcherエージェントにおいてOpenAIのGPTモデルの代替としてAnthropicのClaudeモデルをサポートしていたが、今回のアップデートはこれらを「選択式」にするのではなく、「同一タスク内で統合的かつ同時に活用する」というハイブリッド戦略の具現化である⁷。本稿では、この新しいマルチモデル・システムの技術的アーキテクチャ、業界標準ベンチマークを通じたパフォーマンス評価、市場からの評判、そしてこれらがエンタープライズAIエコシステムに及ぼす戦略的影響について網羅的に分析する。

単一モデルAIの限界とディープリサーチ領域における課題

2024年末から2025年にかけて、ディープリサーチAIの開発はテクノロジー業界における最も熾烈な軍拡競争の一つであった。Googleが2024年12月にGemini向けのリサーチエージェントを発表したのを皮切りに、OpenAIが2025年2月に独自のエージェントをリリースし、続いてxAIやPerplexityもリサーチ機能の強化に注力した⁵。AnthropicのClaudeも、詳細で引用の充実した回答を求める専門家の間で確固たる支持を築き上げていた⁵。これらの企業はこぞって、自社の提供する単一のAIモデルこそが「部屋の中で最も賢いリサーチャー」であると市場に訴求してきた経緯がある⁵。

しかし、エンタープライズ市場における実運用が進むにつれ、単一モデル(モノモデル)に依存したAIリサーチの構造的な欠陥が浮き彫りになってきた。単一のLLM(大規模言語モデル)は、膨大なデータを要約し流暢な文章を生成する能力に長けている一方で、特定の推論の偏り(バイアス)、情報源の脆弱な引用、そして存在しない事実を捏造するハルシネーションといった問題を完全に払拭するこ

とができなかった⁵。企業が戦略的決定を下すためのリサーチ・ヘビーな業務においては、美しく整えられた単一の回答よりも、複数の妥当な解釈を構造的に比較検討できることの方がはるかに高い価値を持つ⁴。Microsoftが「なぜ1つのモデルを選ばなければならないのか？」という根源的な問いを投げかけ、GPTとClaudeを組み合わせるアプローチに舵を切ったのは、こうしたエンタープライズ特有の厳格な品質要求に応えるためであると言える⁵。

「Critique」のアーキテクチャ: 生成と評価の分離による品質保証

「Critique」は、Microsoft 365 CopilotのResearcherエージェントにおいて、モデル選択画面で「Auto（自動）」を選択した際にデフォルトで稼働する、複雑な調査タスク向けに設計されたマルチモデル・ディープリサーチ・システムである²。このシステムの最大の特徴にして革新的な点は、コンテンツの「生成(Generation)」フェーズと、その「評価・検証(Evaluation)」フェーズをアーキテクチャレベルで明確に分離し、それぞれを異なるフロンティアモデルに担当させている点にある⁸。

システム内部のワークフローは、人間の編集プロセスを模倣した直列的な反復構造を持つ⁴。まず一方のモデル(現在の実践的な運用では主にOpenAIのGPT)がタスクの計画立案、情報の検索(Retrieval)の反復、そして初期ドラフトの作成(Drafting)を主導する¹。この段階では、情報の広範な収集と文脈の構築が優先される。続いて、もう一方のモデル(主にAnthropicのClaude)が「専門的レビュアー(Expert Reviewer)」として介入し、生成されたドラフトに対する厳密なレビューと洗練を実施する²。Microsoftはこのワークフローについて、将来的にはClaudeが初期ドラフトを作成し、GPTがレビューを行うといった双方向(Bidirectional)の運用も可能になるとしている⁶。

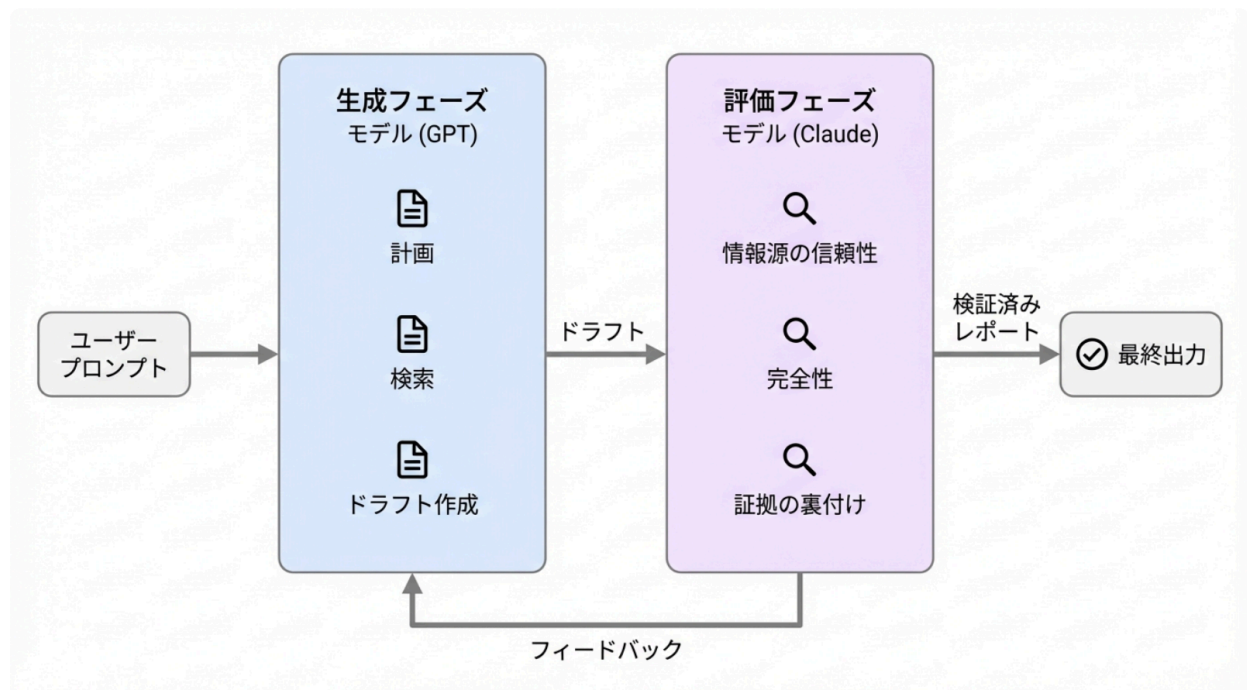
このレビュアーモデルは、第二の執筆者として振る舞うのではなく、特定のルーブリック(評価基準)に基づいて出力を強化することに専念する³。評価プロセスに生成プロセスと同等の重みを持たせることで、システム内に強力なフィードバックループが形成される⁸。レビュアーが適用する保守的かつ厳格な評価基準は、主に以下の3つの次元で構成されている。

| 評価次元 | 検証内容とアーキテクチャの目的 |
|--|--|
| 情報源の信頼性評価 (Source Reliability Assessment) | 参照された情報源が検証可能であり、そのドメインにおいて権威と評判のある適切なソースであるかを優先的に確認する。これにより、不確かなウェブサイトからの引用を排除する ³ 。 |
| レポートの完全性 (Report Completeness) | 最終的なレポートが、ユーザーの要求(プロンプト)の意図を網羅的に満たしているか、また独自の洞察(Unique insights)が適切に含まれているかを検証する ³ 。 |

| | |
|--|---|
| <p>厳密な証拠の裏付け (Strict Evidence Grounding Enforcement)</p> | <p>レポート内のすべての主要な主張が、信頼できる情報源と正確な引用にアンカー(固定)されていることを要求する。これにより事実の正確性を高め、ハルシネーションを根本から抑制する³。</p> |
|--|---|

この多層的な検証プロセスにより、生成モデルが陥りがちな「弱い主張」や「分析視点の欠落」が自動的に特定され、修正される。生成と評価を分離するこの設計は、システムが進化するにつれて、ジェネレーターとレビュアーそれぞれの役割を独立してサポートし拡張できるという、拡張性の面でも優れたアーキテクチャである²。

Critiqueシステムにおける複数モデル間の生成・検証ワークフロー



ユーザーの入力に対し、生成担当モデルが初期ドラフトを作成した後、評価担当モデルが「情報源の信頼性」「完全性」「証拠の裏付け」の3基準で厳格なレビューを行い、最終的なレポートを出力する。

「Council」のメカニズム: 複数モデルの並行実行と客観的調停

直列的な連携を旨とするCritiqueに対し、もう一つの新機能である「Council(評議会)」は、複数モデ

ルを並行稼働させるという代替的アプローチを提供する³。モデル選択画面で「Model Council」を選択することで起動するこの機能は、1つのプロンプトに対して、OpenAIのモデルとAnthropicのモデルを同時に、かつ完全に独立して実行させる²。

Councilモードの最大のアドバンテージは、各モデルの基礎となる学習データや推論アルゴリズムの違いから生じる「多様な視点」を失うことなく比較できる点にある。一方のモデルが見落とししたり、異なる重み付けで評価したりする可能性のある事実、引用、分析のフレームワークを、もう一方のモデルが補完的に提示する²。両モデルによる完全な独立レポートが生成された後、システムの内部では専用の「Judge(調停/裁判官)」モデルが稼働し、双方の出力結果を客観的に評価する²。

このJudgeモデルは、最終的な成果物として各モデルのレポートを並べて表示(サイド・バイ・サイド比較)するだけでなく、「Cover Letter(カバーレター)」と呼ばれる独自の要約ドキュメントを生成する²。このカバーレターは、複雑なトピックに対するAIの推論の透明性を高める上で極めて重要な役割を果たし、以下の要素を抽出して提示する。

| カバーレターの構成要素 | 提供される洞察とユーザーへのメリット |
|---------------------------------|--|
| モデル間の合意点 (Agreements) | 独立した両モデルが共通して重要だと認識し、見解が一致した領域を示す。これは情報の確度の高さを示唆する ² 。 |
| モデル間の相違点 (Divergences) | モデル間で意見が分かれた領域。ここには、問題の捉え方(Framing)、影響の規模(Magnitude)、データに対する解釈(Interpretation)の差異が明示され、人間が深掘りすべきリスク領域を浮き彫りにする ² 。 |
| 独自の貢献 (Unique Contributions) | 一方のモデルのみが提示した特有の洞察やニッチなデータ。多様な視点を確保し、死角をなくすために機能する ² 。 |

Council機能がエンタープライズにもたらす真の価値は、ソフトウェアのユーザビリティの根本的な転換にある。従来のAIアシスタントでは、ユーザーは完成された回答を受動的に消費する存在であった。しかし、CouncilはAI同士の「意見の相違」を正常な推論プロセスの一部として意図的に露出させることで、ユーザーに対して、単一の正解を探すのではなく、編集者や意思決定者としての批判的思考(Critical thinking)を促す効果を持つ⁴。MicrosoftがCopilotを「ボットとのチャット」から「システムを用いた業務遂行」へと移行させているという指摘は、まさにこの認知モデルの変容を指している⁴。

DRACOベンチマークによる定量的なパフォーマンス実証

Microsoftは、これらのマルチモデル・アプローチの有効性を客観的に証明するため、「DRACO(Deep Research Accuracy, Completeness, and Objectivity)」ベンチマークを用いた評価結果を公開した¹。DRACOベンチマークは、医学、テクノロジー、法律など10の異なる専門領域にまたがる、

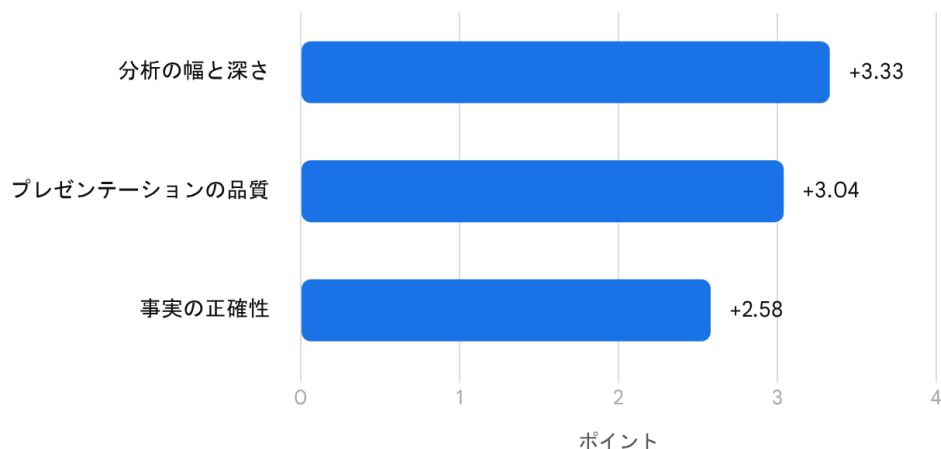
100の複雑なディープリサーチ・タスクから構成される業界標準のテストであり、LLMの評価者としてOpenAIのGPT-5.2が使用されている³。

評価の結果、Critiqueアーキテクチャを搭載したResearcherは、従来の単一モデルを用いたシステムと比較して圧倒的な優位性を示した。具体的には、2026年2月に公開されたZhongらの論文において当時最高性能とされていた「Perplexity Deep Research (Claude Opus 4.6モデルを使用)」と比較して、総合スコアで+7.0ポイント(標準誤差 ±1.90)の大幅な改善を達成した²。これは割合にして13.88%の向上であり、従来の単一モデルアプローチの限界を打破したことを定量的に証明するものである²。

このスコアの改善は、単一の指標に偏ることなく、リサーチ品質を構成する複数の次元において統計的に有意(ペア対応t検定、 $p < 0.0001$)な向上を見せている¹²。改善幅の具体的な内訳は以下の通りである。

| 評価の次元 | スコア改善幅 | 改善が意味する実務的価値 |
|--|------------|--|
| 分析の幅と深さ (Breadth and Depth of Analysis) | +3.33 ポイント | ユーザーの要求に対して、欠落していた分析角度を見つけ出し、カバレッジのギャップを埋める能力が最も向上したことを示す ³ 。 |
| プレゼンテーションの品質 (Presentation Quality) | +3.04 ポイント | 複雑な情報を整理し、論理的な構造で分かりやすく提示する能力が向上し、ドキュメントとしての完成度が高まった ³ 。 |
| 事実の正確性 (Factual Accuracy) | +2.58 ポイント | 厳格な証拠に基づく引用の強制により、脆弱な主張が排除され、ハルシネーションの発生確率が大幅に低下したことを裏付ける ³ 。 |

DRACOベンチマークにおける評価次元別のスコア改善幅



Perplexity Deep Research (Claude Opus 4.6) と比較した、Critique機能搭載Researcherのポイント向上。すべての次元において統計的に有意な改善が見られた。

データソース: [Microsoft Tech Community](#), [GIGAZINE](#), [Office Watch](#), [Computerworld](#)

Copilot CoworkおよびWave 3ビジョンとの統合による自律性の拡張

CritiqueとCouncilの導入は、独立した機能アップデートではなく、Microsoftが推し進める「Microsoft 365 Copilot Wave 3」というより広範なビジョンの中に組み込まれている。Wave 3は、AIが単なるコンテンツ生成の補助ツールから、企業アプリケーション全体にわたる「タスクのオーケストレーションと実行」を担う自律型エージェント(Agentic AI)へと進化する転換点として位置づけられている¹。

MicrosoftのAI at Work部門Chief Marketing OfficerであるJared Spataroは、このアップデートが「インテリジェンスが仕事の文脈を理解し、企業全体で安全にAIを拡張することを可能にする信頼(Trust)」を象徴するものだと述べている¹。特に注目すべきは、Critiqueと並行してアクセスが拡大されている「Copilot Cowork」機能との連携である。CoworkはAnthropicのClaudeを基盤としており、長時間の実行を伴うマルチステップのタスクを自律的に計画し、Microsoft 365の各種アプリケーション

ンを跨いで実行する能力を持つ¹⁴。

Critiqueにおける「計画・検索・生成・評価」という反復的なワークフローは、まさにこのマルチステップ・タスクを安全に完遂するための基盤論理として機能する。AIが自らの出力を評価し、修正する能力を獲得したことで、AIは単一のプロンプトに対する応答機能(アシスタント)から、人間の介入を最小限に抑えつつ高品質な成果物を継続的に出力する「業務システム」へと昇華したのである⁴。Microsoftは、業界全体の最高のAIイノベーションを、「Work IQ」のインテリジェンスと「Enterprise Data Protection」の信頼性という自社の強みと結合させることで、管理されたエンタープライズグレードの体験を提供することを約束している¹⁶。現在、これらの機能はアーリーアクセスプログラムである「Frontierプログラム」の参加者向けに限定公開され、広範なフィードバックの収集とテストが行われている¹。

専門家・アナリストによる評価と市場の受容

このマルチモデル・アプローチに対する市場や専門家からの初期の評価は、テクノロジーの方向性に対する強い賛同と、実際の企業導入に向けた冷静な見極めが交錯している。MicrosoftのCEOであるSatya Nadellaは自身のソーシャルメディア(XおよびLinkedIn)を通じて、この機能が複数のモデルを同じプロンプトに対して同時に実行し、モデル間の見解の一致や相違点を視覚化できる画期的なアプローチであると強調し、ベンチマークにおける「クラス最高」の成果を大々的にアピールした⁵。また、Microsoft 365およびCopilot担当CVPのNicole Herskowitzは、マルチモーダルアプローチがユーザーのワークフローを支援し、生産性と品質を低下させる最大の要因であったハルシネーションを大幅に制限できると説明している⁹。

独立系のアナリストからの評価も総じて前向きである。Pareekh ConsultingのCEOであるPareekh Jainは、Critiqueの仕組みを「賢明な専門家(GPT)と厳格なレビュアー(Claude)がペアになっているようなものだ」と比喩的に表現し、エンタープライズ環境で最も忌避されるエラーを確実に削減できる実践的な進歩として評価した¹²。ただし、「魔法ではなく漸進的な進歩であり、エラーを完全に排除するわけではない」という現実的な留保も付与している¹²。

一方で、モデルのオーケストレーション機能単体では、企業が求める真のROI(投資対効果)を達成するには不十分であるという声もある。Counterpoint ResearchのVPであるNeil Shahは、「マルチモデルシステムがその潜在能力を最大限に発揮するのは、CRM(顧客関係管理)やHRM(人的資源管理)といった企業内部のシステムデータと深く統合された時である」と指摘する¹²。企業が求める意思決定の根拠は、汎用的な事実の羅列ではなく、自社の市場ポジション、顧客のヒューリスティクス、独自のビジネスルールに基づく「文脈的にニュアンスのある洞察」であるためだ¹²。この観点から見れば、CritiqueとCouncilは強力な推論エンジンの獲得を意味し、次のステップは社内の閉鎖環境(Graphデータ等)における安全かつ精緻なデータ・グラウンディングの実現に移ることになる。

エコシステムの再編とMicrosoftのハイブリッド戦略の確立

Microsoft 365 CopilotにおけるAnthropicモデルの本格的な統合と、Critique/Councilによるマルチベンダー連携は、AI業界の勢力図と企業のアライアンス戦略における重大な転換点を示している。

Microsoftは長年、OpenAIの最大のパトロンであり、インフラストラクチャの独占的な提供者であった。2025年10月に発表された再編合意においても、Microsoftは2032年までOpenAIのモデルと製

品に関する知的財産権を維持することが確認されている⁶。このような極めて強固な資本・業務提携関係にありながら、自社の最重要エンタープライズ製品のコア機能において、OpenAIの最大のライバルであるAnthropicの技術(Claude Opus等)を、単なる代替オプションではなく「GPTの出力を評価・監査する上位プロセス」として組み込んだことは、業界に大きな波紋を呼んだ¹⁵。

このハイブリッド戦略は、Microsoftの競争優位性(モート)の源泉がどこにあるのかを雄弁に物語っている。Microsoftは、個別の「AIモデルのパラメータ数」や「基礎性能」の競争において単一ベンダーに依存するリスクを回避し、最良のモデルを組み合わせることで最大の価値を引き出す「システム統合と論理的検証のプラットフォーム」としてのポジションを確立したのである⁶。AWSやGoogle Cloudといった他のハイパースケーラーも同様にマルチモデルを提供する立場にあるが、Microsoftは世界中のナレッジワーカーが毎日使用する「Microsoft 365」というインターフェースを握っている点が決定的に異なる¹⁰。生成AIの付加価値が、APIの裏側にある個々のモデルの頭脳から、それを業務フローに落とし込むオーケストレーション層へとシフトしていることを、今回のCritiqueとCouncilの実装は明確に示している。

意思決定の未来と次世代エンタープライズAIの展望

2026年3月のアップデートは、エンタープライズAIが「単一モデルの流暢な回答」に依存する段階を脱し、複数の高度な知能を構造的に戦わせることで信頼性を担保する「検証と調停の時代」へ突入したことを告げている。Critiqueによる生成と評価の分離は、DRACOベンチマークが示す通り、複雑なディープリサーチにおいて従来システムを凌駕する事実精度と分析の深さをもたらした。同時に、Councilが提供するカバーレターとサイド・バイ・サイドの比較機能は、AIを盲信するのではなく、人間のユーザーが批判的思考を持つ編集者としてAIと協働する新しい認知モデルを定着させようとしている。

今後の焦点は、Frontierプログラムを通じて得られるフィードバックをもとに、このシステムがいかにして各企業のローカルな業務データ(CRM/HRM/社内ドキュメント)と安全かつシームレスに結びついていくかにある。AIが事実の検索や文章の生成を超え、複数のモデル間で矛盾を調停し、より高度な文脈を理解する自律的なプロセスを獲得したことで、企業における意思決定の速度と質はかつてない次元へと引き上げられる。Microsoftのマルチモデル・オーケストレーション戦略は、生成AIの実験的な導入期を終わらせ、真の意味で企業の中核業務にAIが組み込まれる未来への決定的な道筋を敷いたと評価できる。

引用文献

1. Copilot Cowork: Now available in Frontier | Microsoft 365 Blog, 4月 4, 2026にアクセス、
<https://www.microsoft.com/en-us/microsoft-365/blog/2026/03/30/copilot-cowork-now-available-in-frontier/>
2. Introducing multi-model intelligence in Researcher - Microsoft Community Hub, 4月 4, 2026にアクセス、
<https://techcommunity.microsoft.com/blog/microsoft365copilotblog/introducing-multi-model-intelligence-in-researcher/4506011>
3. 「OpenAIとAnthropicのAIモデルを1つのプロンプトで同時実行する ...」, 4月 4, 2026にアクセス、

4. <https://gigazine.net/news/20260331-microsoft-365-copilot-critique-council/>
Microsoft Copilot Critique and Council: Multi-Model AI for Trusted Research, 4月 4, 2026にアクセス、
<https://windowsforum.com/threads/microsoft-copilot-critique-and-council-multi-model-ai-for-trusted-research.408809/?amp=1>
5. Microsoft Made GPT and Claude Work Together—And the Result Beats Every AI Research Tool Out There, 4月 4, 2026にアクセス、
<https://decrypt.co/362805/microsoft-gpt-claude-work-together-ai-research>
6. Microsoft Introduces “Critique” System to Combat AI Hallucinations with Dual-Model Self-Correction - TMGM, 4月 4, 2026にアクセス、
<https://www.tmgm.com/en/analysis/market-insight/microsoft-dual-model-ai-system-reduces-hallucinations>
7. Microsoft 365 Copilot's Researcher Agent Now Uses GPT and Claude to Improve Answers, 4月 4, 2026にアクセス、
<https://petri.com/microsoft-365-copilot-researcher-gpt-claude/>
8. Microsoft Upgrades 365 Copilot Researcher With New Critique Mode - Dataconomy, 4月 4, 2026にアクセス、
<https://dataconomy.com/2026/04/02/microsoft-upgrades-365-copilot-researcher-with-new-critique-mode/>
9. Microsoft takes wraps off AI upgrades, unveils new deep research system Critique, 4月 4, 2026にアクセス、
<https://seekingalpha.com/news/4570243-microsoft-takes-wraps-off-ai-upgrades-unveils-new-deep-research-system-critique>
10. Microsoft 365 Copilot's Researcher agent goes multi-model, 4月 4, 2026にアクセス、
<https://www.constellationr.com/insights/news/microsoft-365-copilots-researcher-agent-goes-multi-model>
11. Microsoft Copilot Researcher Gets a Two-Brain Upgrade: Critique and Council Explained, 4月 4, 2026にアクセス、
<https://office-watch.com/2026/microsoft-copilot-researcher-critique-council/>
12. Microsoft adds multi-model AI to Copilot Researcher, raising accuracy stakes, 4月 4, 2026にアクセス、
<https://www.computerworld.com/article/4152323/microsoft-adds-multi-model-ai-to-copilot-researcher-raising-accuracy-stakes.html>
13. Microsoft 365 Copilot gets major update: Cowork Feature, multi-model research and more, 4月 4, 2026にアクセス、
<https://timesofindia.indiatimes.com/technology/tech-news/microsoft-365-copilot-gets-major-update-cowork-feature-multi-model-research-and-more/articleshow/129907901.cms>
14. Microsoft Copilot Cowork Signals Shift to Multi-Step AI Workflows for Enterprise Users, 4月 4, 2026にアクセス、
<https://www.cxtoday.com/ai-automation-in-cx/microsoft-copilot-cowork-signals-shift-to-multi-step-ai-workflows-for-enterprise-users/>
15. Microsoft Critique Explained: How Copilot Now Uses GPT and Claude Together for Deep Research | Knowledge Hub Media, 4月 4, 2026にアクセス、

<https://knowledgehubmedia.com/microsoft-critique-explained-how-copilot-now-uses-gpt-and-claude-together-for-deep-research/>

16. Powering Frontier Transformation with Copilot and agents | Microsoft 365 Blog, 4月 4, 2026にアクセス、

<https://www.microsoft.com/en-us/microsoft-365/blog/2026/03/09/powering-frontier-transformation-with-copilot-and-agents/>