

なぜ AI は思い通りの回答をくれないのか？

～AIに「考えさせる力」は人間が握っている～

ライフデザイン研究部 主席研究員/テクノロジーリサーチャー 柏村 祐

1. AIが的外れな回答を返すとき、欠けているのは“問い”である

生成AIは、かつての検索エンジンのように「知識を取りに行く」ツールではなく、「思考を返す」パートナーとして期待されている。しかし、多くの利用者が抱える実感は異なる。「期待した答えが返ってこない」「内容が薄い」「見当違いだ」といった期待外れの応答はどこから生まれるのだろうか。

この現象の原因は、人間の質問力の欠如にある。AIは「問いを構造化して返す鏡」である。つまり、AIに考えさせるには、明確な文脈(コンテキスト)に基づいて構造化された問いを渡さなければならない。だが実際、多くの現場では「背景を省いたまま依頼する」ことが常態化しているのではないだろうか。「提案を出して」「まとめておいて」という抽象的な指示ではAIは何を基準に判断すべきか分からず、曖昧な出力しか返せない。人間でいえば、「目的地も地図もないまま“走っておいて”と言われてる」ようなものである。

AIが思い通りの回答をくれないのは、AIの能力不足ではなく、人間が思考を十分に言語化していない点にある。本レポートでは、AIの出力品質を左右する要因と、その具体的な対策について論じていく。

2.構造化された指示がAIの思考を導く

第1章で提示した「AIの出力品質は、人間の問いの質に依存する」という仮説は、実証研究によって裏付けられている。ハンブルク大学が2025年に発表した論文「大規模言語モデルのフィードバック提供者としての可能性と落とし穴」では、プロンプト(AIへの指示)の品質が、AIの生成するフィードバックの品質にどれほど影響を与えるかが詳細に検証された。

1)AIの思考を起動させる「良質な指示」の構成要素

この研究では、まず高品質なプロンプトを開発するための評価基準が策定された。これは、AIに思考の「足場」を与えるために、人間が何を言語化すべきかを示している。具体的には、「コンテキスト(背景)」「ミッション(任務)」「明確性と具体性」といった大項目の中に、AIの役割設定や対象者の指定など、8つの評価軸が設けられている(図表1)。この基準は、第1章で述べた「目的地も地図もないまま走らせる」

ような曖昧な指示とは対極にある。AI に質の高い思考を求めるには、これほど構造化された情報提供が必要不可欠なのである。

図表 1 高品質なプロンプトを構成する 8 つの要素

カテゴリー	サブカテゴリー	内容
コンテキスト	役割 (Role)	AI と質問者の役割を明確に説明する
	対象者 (Target audience)	回答を受け取る相手を明確に定義・描写する
	媒体 (Medium/channel)	回答が使用される媒体(メール、スライド等)を明記する
ミッション	任務/質問 (Mission/question)	AI が達成すべき任務を明確に記述する
明確性と具体性	形式と制約 (Format and constraints)	文体や文字数などの仕様を記述する
	簡潔さ (Conciseness)	任務に直接関連する情報のみを含み、簡潔である
	専門性 (Domain specificity)	専門用語を正しく使用し、AI がそれを参照できるようにする
	論理性 (Logic)	指示内容に論理的な一貫性があり、理解しやすい構成である

資料: Jacobsen, L.J. & Weber, K.E. (2025) "The Promises and Pitfalls of Large Language Models as Feedback Providers" を基に筆者作成

2) 指示の質が AI の回答の質を劇的に変える

次に、この研究では実際に「低品質」「中品質」「高品質」の 3 種類のプロンプトを AI (ChatGPT-4) に与え、生成されたフィードバックの品質を比較評価した。評価は、教育学研究の手法に基づき、訓練を受けた 3 名の評価者が Prilop ら (2019) と Prins ら (2006) の基準をもとに 9 項目を 3 段階 (0~2 点) で採点したものである。その結果、ごく一部の項目を除き、高品質プロンプトは統計的に優位に高得点だった。

特に、フィードバックの根拠となる「評価基準の的確さ」では、低品質プロンプトのスコアが 0.45 だったのに対し、高品質プロンプトでは 1.95 と、4 倍以上の差があった。このデータは、曖昧な指示からは曖昧な回答しか生まれず、明確で構造化された指示こそが AI の思考精度を最大化させることを明確に示している (図表 2)。

図表2 プロンプトの品質によるAIの回答品質の変化

評価項目	低品質プロンプト (Prompt 1)	高品質プロンプト (Prompt 3)
評価基準的確かさ (Assessment criteria)	0.45	1.95
説明の質 (Explanation)	0.25	1
質問の質 (Questions)	1.2	1.9
具体性 (Specificity)	0.1	1.35

資料: 図表1に同じ

注: スコアは2点満点の平均値。

3) 最高の問いは、AIに専門家を超越するパフォーマンスをもたらす

さら、高品質なプロンプトを与えられたAIは、人間の専門家と比較しても高いパフォーマンスを示した。この研究では、高品質プロンプトでAIが生成したフィードバックを、初心者(教員志望の学生)と専門家(教員トレーナーや教授)のフィードバックと比較した。結果、一部の項目を除きAIは初心者を上回り、さらに説明の質・質問の質・具体性では専門家よりも有意に高かった(図表3)。

これは、AIが単に情報を整理するだけでなく、質の高いプロンプトを得れば、専門家をも凌駕するレベルで価値あるアウトプットを生み出すポテンシャルをもつことを示唆している。

図表3 初心者・専門家・AI(ChatGPT-4)のフィードバック品質比較

評価項目	初心者 (Peers)	専門家 (Experts)	AI (ChatGPT-4)
説明の質 (Explanation)	0.1	0.55	1
質問の質 (Questions)	0.17	1.36	1.86
具体性 (Specificity)	0.17	0.64	1.6

資料: 図表1に同じ

注: スコアは2点満点の平均値。

ここまで示したデータは、プロンプトの構造化がAIの出力品質を左右することを実証している。しかし、テクノロジーリサーチャーの視点から、さらに強調したいのは、単に「形式を整える」だけでは不十分だという点である。

真に効果的なプロンプト設計には、メタ認知的な視点が求められる。つまり、「自分が何を知らないのか」「どこまで言語化できているのか」を自覚しながら問いを組み立てる能力である。図表2で示された高品質プロンプトと低品質プロンプトの差は、情報量の差ではなく、思考の解像度の差に起因している。AIに「考えてほしいこと」を明確にするには、まず人間が自分の思考のあいまいさに気づき、それを分解する必要がある。

また、プロンプトエンジニアリングは一度限りの作業ではない。AIとの対話を通じ

て出力を検証し、期待とのずれを分析し、問いを再設計するフィードバックループこそが、AI を思考のパートナーへと育てる実践的手法である。技術は進化するが、人間の問いの質が AI の限界を規定するという原則は、今後も変わらないだろう。

3. 質問力とコンテキスト設計力を磨け

AI に成果を出させるには、次の二つが不可欠である。

一つは質問力を磨くことであり、もう一つは背景となるコンテキストを丁寧に与えることである。質問力とは、ただ問いを立てる力ではない。「この質問を通じて何を明らかにしたいのか」「その答えで誰が動くのか」を整理する力である。そしてそれは、一度では身につかない。AI との対話を繰り返し、出力を見て修正し、また問いを立て直す。この試行錯誤の連続が、質問力を磨き上げる唯一の方法である。興味深いのは、AI の時代となっても、いや、AI を最大限機能させるためにこそ、人間側に高いコミュニケーション能力が求められるという点である。技術が進化しても、それを使いこなす人間の基礎能力は依然として重要なのだ。

もう一つの鍵が、コンテキストの設計力である。AI は背景を与えられなければ、思考の方向を定められない。目的、相手、環境、制約を含んだプロンプトを作ること、AI は初めて「理解」したように振る舞う。コンテキスト設計力とは、目的、対象者、前提条件、制約事項といった背景情報を的確に言語化し、構造的に整理して伝える能力を指す。たとえば、「提案書を作って」という指示だけでは、AI は何のための、誰に向けた提案かが分からない。しかし「新規顧客向けの営業提案書を作成してください。相手は製造業の経営層で、コスト削減を最優先課題としています。A4 で 2 ページ以内、専門用語は避けて分かりやすく」と伝えれば、AI は的確な方向性で思考を展開できる。人間でいえば、相手の立場や状況を知らずに議論しても建設的な意見が出ないのと同じである。

図表 4 は、AI への問いかけ方による出力品質の違いを対比的に示している。左側の「BEFORE」では、ユーザーが「提案よろしく」「いい感じにまとめて」といった曖昧な指示を出している。これに対し AI は「何を基準に？」と戸惑い、結果としての的外れで薄い回答しか得られず、思考のパートナーにはなれない。一方、右側の「AFTER」では、目的（若手向けの企画立案）、背景（社内の交流が不足）、制約（予算は 10 万円以内）という具体的なコンテキストが構造化されて提示されている。その結果、AI は「承知しました。3 つの企画案を具体的な予算案と共に提案します」と応答し、共創の相手へと変化する。的確で質の高い思考を返し、生産性を飛躍的に向上させることができるのである。

図表4 AIの知性を引き出す「問いかけ方」



資料:筆者作成

AIは万能ではない。しかし、良い問いと文脈を与えれば、驚くほど知的に考える。AI活用の本質は、自分自身の問いを鍛えることにある。

AIは、与えられた“問い”と“背景”の質に応じて思考する。つまりAIが何も生み出さないとき、それは人間の側がまだ考え切れていないということである。質問力とコンテキスト設計力を磨き続けることこそが、生成AIを“ただのツール”から“共創の相棒”へと変える道なのである。