COLUMN ｜ AI活用 ｜ 株式会社DnD

AI回答をそのまま使って大丈夫？

生成AIの「嘘」との正しい向き合い方

株式会社DnD（DnD Inc.） — Distill n Develop

2026年7月17日

この資料の要点

- 生成AIは「次に来るべき言葉を確率で選ぶ」仕組みのため、もっともらしい嘘（ハルシネーション）を自信を持って出力することがある。
- 数値・固有名詞・法律・最新情報など「外れたら困る」情報は特に確認が必要。
- 「人が検証する習慣」とRAGやプロンプト工夫の組み合わせで、業務リスクは大幅に減らせる。

生成AIが「嘘をつく」のはなぜか

生成AIは、インターネット上の膨大なテキストを学習した確率モデルです。「この文脈の次に来るのは、どの言葉が最も自然か」を確率的に計算して文章を生成します。そのため、流暢で説得力のある文章を作るのは得意ですが、事実を記憶・検索しているわけではありません。

この仕組みから生まれるのが「ハルシネーション（幻覚）」です。存在しない論文・架空の統計・誤った固有名詞を、さも本当のことのように自信を持って出力することがあります。

> *AIは「正しいことを言う」のではなく、「それらしい言葉をつなぐ」。その違いを理解しておくことが、安全な活用の出発点です。*

2025年に公表された研究では、現在の大規模言語モデルのアーキテクチャのもとでハルシネーションを数学的に完全排除することは困難であるとの報告もあります。モデルが進化してもゼロにはなりません。この前提を持つことが重要です。

ハルシネーションが特に起きやすい場面

- **具体的な数値・統計を求めたとき**：「○○業界の市場規模は？」など。AIが学習データ上にない数字を「それらしく」作り出すことがあります。
- **固有名詞（人名・企業名・法律名）を含む質問**：「A社の代表取締役は誰？」などは、誤りが含まれやすい領域です。
- **専門性が高い領域**：法律・医療・会計・税務など、専門的な正確さが求められる分野では自信ある誤回答が特に出やすくなります。
- **学習データ締め切り以降の最新情報**：AIには知識の「締め切り日」があります。それ以降の出来事は知らないにもかかわらず推測で答えることがあります。

業務でAI回答を使うときの4つの確認習慣

- **① 重要な数値・固有名詞は公式ソースで裏付ける**：法令・統計・他社情報などは、官公庁や公式サイトで確認します。
- **② 「出典を教えて」と聞く（ただし出典自体も要確認）**：AIが挙げる出典が架空である場合もあります。実際にアクセスして確認する習慣をつけましょう。
- **③ 用途・リスクに応じて検証の深さを変える**：社内向け草稿は確認コストを低く、顧客提出物・法的文書は専門家チェックと組み合わせます。

- **④「わからない点があれば教えて」と一言添える**：プロンプトに「不確かな情報があればわからないとはっきり言ってください」と加えるだけで、あてずっぽうな回答を抑えられます。

ハルシネーションを減らすための3つのアプローチ

① RAG（検索拡張生成）を活用する

RAGとは、AIが回答する際に信頼できるドキュメントや社内データベースをリアルタイムで参照させる仕組みです。「正しい情報の引き出し」を別途用意してAIに使わせることで、ハルシネーションを大幅に抑えられます。社内規程・製品マニュアル・FAQ集などを登録しておくと効果的です。

② プロンプトで回答の範囲を絞る

「以下の文章の範囲内だけで答えてください」「確認できない情報は出力しないでください」といった指示を加えることで、AIが勝手に補完する余地を狭められます。テンプレートとして社内で共有しておくと、チーム全体でリスクを下げることができます。

③ 複数モデル・複数情報源でクロスチェックする

異なるAIモデルで同じ質問をして回答を比較したり、Webブラウジング機能があるツールで最新情報を参照させたりする方法も有効です。

まとめ

ハルシネーションという特性を理解した上で、「AIを補助として使い、人が最終確認する」というスタンスで運用すれば、大きな業務価値を安全に引き出せます。完璧なAIを待つより、今あるAIの限界を知って使いこなすほうが、結果的に早く業務は変わります。