連載

お伴としてのChatGPT

生成AIアシスト 特許の調査＆出願

第3回　生成AIはウソをつく!? AI活用の怖〜い落とし穴7選

深川 栄生　Shigeo Fukagawa

生成AIは非常に便利なツールですが，安全に使うためには，特有のリスクを正しく理解し，適切な対策をとる必要があります．特に，専門知識と最新情報が不可欠な知的財産業務では，リスクを回避することが極めて重要になります．本稿では，生成AIを利用する際に注意すべき7つのリスクについて，プロンプト例を交えながら，わかりやすく解説します．

AI活用の怖い落とし穴－その1「あたかも本当（ハルシネーション）」

● 事実に基づかない，息を吐くようにウソつくAI

生成AIにおけるハルシネーションとは，事実に基づいていない情報をあたかも本当のことのようにもっともらしく生成してしまう現象のことです．実在しない特許番号を提示したり，架空の裁判例について詳細に説明したりするケースがこれにあたります．

ハルシネーションが起こる原因は，生成AIの学習方法と使われるデータにあります．生成AIは自然で人間らしい文章を作ることを重視しており，必ずしも情報の正確さを優先しているわけではありません．その結果，学習に使われたデータの中に誤った情報や古い情報が含まれていると，それらをもとに，もっともらしい内容の誤情報を作り出してしまうことがあります．

また，質問の内容があいまいだったり情報が不足していたりすると，AIが意図を正しく理解できずに，ハルシネーションを引き起こすこともあります．

ハルシネーションによって生成された誤った情報を信じてしまうと，知的財産活動において深刻な影響をおよぼす可能性があります．実際には存在しない先行技術を根拠に開発を中止したり，誤った法解釈に基づいて出願戦略を立てたりするケースが考えられます．

● 便利だとは言え，確認作業の人力が必要だ

最も基本的で重要な対策は，生成AIの回答をうのみにしないという意識をもつことです．

AIが提示する情報は，あくまで参考にとどめ，必ず特許庁のデータベースや法律事務所の公式ウェブ・ページなど，信頼できる複数の情報源で事実確認（ファクトチェック）を行う習慣をつける必要があります．また，AIに質問する際には，前提条件や背景を丁寧に伝え，できるだけ具体的で明確なプロンプトを入力することが，誤解や虚偽の回答を防ぐ上で効果的です．

● ハルシネーション対策を意識したプロンプトの例

「以下の質問に対し，確実な事実に基づいて回答してください．情報源が不明な場合や推測が含まれる場合は，その旨を明記し，わからない場合は『わかりません』と答えてください．日本の特許法における『新規性喪失の例外規定』の適用要件について，具体的な条文を引用しながら説明してください．」

▶プロンプトの解説

AIに対して「わからないときには正直にそう答えてください」というルールを伝えるものです．人間が質問に答える際に，曖昧な知識で取り繕うことがあるように，AIも，もっともらしいうそをつくことがあります．最初に正直な回答を求める指示を与えることで，AIが無理に答えを作り出そうとして起こるハルシネーションを防ぎ，より信頼性の高い回答を引き出せるようになります．

AI活用の怖い落とし穴－その2「バイアス」

● AIも陥りやすい，偏見や固定概念のワナ

生成AIにおけるバイアスとは，学習データに含まれる社会的な偏見や固定観念を反映し，偏った内容を生成する現象のことです．「発明者は男性である」といった無意識の思い込みに基づいて文章を生成したり，特定の人種や国籍に対して否定的な印象を与える内容が生成されたりするケースが考えられます．

バイアスの主な原因は，AIが学習に使うデータにあります．AIはインターネット上の記事や書籍など，人間が作成した膨大な文章を基に学習しています．しかし，その情報の中には歴史的・社会的な偏見や，特

第1回　手間いらずに！エンジニアのお伴に生成AIのススメ（2025年10月号）
第2回　もっと良い回答をAIから引き出すプロンプト入力のコツ（2025年11月号）