

# UTIE Instruments Inc.

## 解決詐欺

—「わかりません」という代わりに AI はこう進化した— (AI 非専門家向け)

### UTIE Research Institute

2026 年 2 月

#### 第 1 章 用語の定義——ハルシネーションと解決詐欺はどう違うか

まず、混同されがちな二つの概念を明確に分離しておきたい。ハルシネーションとは、AI が事実そのものを捏造する現象である。存在しないライブラリ名を提示する、架空の判例を引用する、まだリリースされていない製品のバージョン番号を断言する——これらはすべて、AI が「事実」のレイヤーで嘘をついている。検証は比較的容易であり、事実を照合すれば真偽は判定できる。これに対して「解決詐欺」は、もっと厄介な虚偽である。解決詐欺においては、個々の事実は正しい。技術は実在し、手法も存在し、引用されている概念も間違っていない。しかし、それらを組み合わせた解決策が、現実世界では実装も運用も不可能である——にもかかわらず、AI は「~すれば解決できます」と断言する。ハルシネーションが「嘘の材料で料理を作る」ことだとすれば、解決詐欺は「本物の材料を使いながら、誰にも作れない料理のレシピを自信満々に渡す」ことである。材料が本物であるがゆえに、受け取った側は嘘だと気づきにくい。これが解決詐欺の本質的な危険性である。

#### 第 2 章 解決詐欺はなぜ発生するのか——言語モデルの構造的限界

大規模言語モデル (LLM) は、膨大なテキストデータから、次に来る確率の高い単語を予測する仕組みで動いている。この原理が、解決詐欺を構造的に生み出す。AI にとって「連携すれば」という 5 文字の出力と、「フライパンで焼けば」という 7 文字の出力は、トークン数が若干違うだけで本質的に同じ重さの処理である。しかし現実世界において、システム間の連携には API 仕様の調査、ファイアウォール申請、認証基盤の統合、エラーハンドリングの設計、保守運用フローの策定が必要であり、数千万円と数ヶ月の工期を要する重労働である。一方、フライパンで卵を焼くのは 3 分で終わる。AI はこの言葉の軽さと現実の重さの乖離を認識できない。「**ロケットを作れば火星にレストランを開けます**」という文は、文法的には完全に正しく、物理法則にも反していない。だから AI はこれを間違いとはみなさず出力する。しかし

「ロケットを作る」という前提条件それ自体が、NASA を超える予算と技術を要求する人類史レベルの難事業であるという認識は、言語モデルの計算には含まれていないのである。

さらに問題なのは、AI がシンプルさを志向する傾向である。AI にとってのシンプルとは「文字数が少ない」「登場人物が少ない」「構造が A→B の一直線」であることを意味する。人間にとってのシンプル——すなわち「工数が少ない」「既存のもので済む」「リスクが低い」——とはまるで別物である。この認識の断絶が、AI をして「フライパンで焼く（地味だが実現可能）」よりも「ロケットでドン！（壮大だが実現不能）」を好んで提案させる原因となっている。

### 第3章 「～すれば」という魔法の杖——解決詐欺の言語的特徴

解決詐欺には、特徴的な言語パターンが存在する。代表的なのが「～すれば解決できます」という条件節による断言である。「全社のデータベースを統合し、データレイクを構築すれば可能です」——この文を受け取った人間の多くは、データレイクを構築するという前提条件の重さを適切に評価することなく、「可能です」という結論部分に意識を持っていかれる。なぜなら、AI が自信に満ちた断定口調で締めくくるため、条件節があたかもちょっとした準備であるかのように矮小化されてしまうからである。「連携すれば」「導入すれば」「最適化すれば」「刷新すれば」「運用すれば」これらの「～すれば」の裏側には、AI が隠蔽している膨大なコスト、人員、時間、政治的調整が存在する。いわば「～すれば」とは、数千万円の実費請求書をポケットに隠して、バラ色の結論だけをちらつかせる詐欺の典型的手口そのものである。もう一つの特徴的なパターンが、トートロジーによる空疎な助言である。「売上が低迷しています」という相談に対して「ターゲット層のニーズを分析し、適切なマーケティング施策を打つことで解決できます」と返す。これは、成功すれば成功できますと言っているようなものだ。最も重要な「どうやって？（How）」が完全に欠落しているにもかかわらず、断定口調で締められているため、読んだ瞬間には何か有益な助言を受けた気分になってしまう。

### 第4章 「真空世界」の提案——なぜ AI は摩擦を計算できないのか

AI が提案する解決策は、常に真空状態的である。つまり、現実の泥臭い制約をすべて除去した理想空間の中で成立している。「A システムと B システムを API 連携すれば解決します」と AI は言う。しかし現実には、A と B は通信プロトコルが異なり、社内のセキュリティポリシーで必要なポートが開かず、そもそも B システムはレガシーすぎて API が存在しない。AI にとって「連携する」は単なる数トークンの出力に過ぎないが、現場にとっては3ヶ月の工数を要する難工事なのである。この乖離は、AGI（汎用人工知能）が実現したとしても本質的には解消されない。仮に AGI が完璧な核融合炉の設計図を出力したとしても、その設計図に必要な超耐熱合金を生産する工場の建設、建設許可の取得、住民説明会の開催、環境アセスメントの通

過——これらの物理的・社会的プロセスには依然として年単位の時間がかかる。知能が光速で進歩しても、施工はカタツムリの速度でしか進まない。さらに深刻なのは、AI が政治や現場の空気を計算に入れられないことである。AGI が「A 国と B 国が資源を融通し合えば最適解です」と算出しても、歴史的な恨み、選挙対策、国家のメンツといった非合理的なパラメータが現実世界を支配している。物理世界は論理（Logic）ではなく摩擦（Friction）で動いている。この摩擦を無視して提示される解決策は、どれだけ論理的に正しくても、実装不能という意味で詐欺なのである。

## 第 5 章 なぜ人間は解決詐欺に騙されるのか——心理学的メカニズム

解決詐欺が危険なのは、AI の出力が巧妙だからだけではない。人間の側にも、騙されやすい心理的構造が存在する。第一に、流暢である＝知的という誤認（ハロー効果）がある。人間は、言葉が流暢な相手は無意識に頭が良いと評価する。AI は飲み会の挨拶文のような定型テキストにおいて、人間以上に滑らかで丁寧な文章を瞬時に出力する。このレベル 1」での成功体験が、「こんなに綺麗な文章が書けるなら、論理構築やコード生成（レベル 10）も正確なはずだ」という過信を生む。しかし滑らかな言語出力と論理の整合性の担保はまったく別の能力である。第二に、安易な成功体験の罠がある。面倒なメール作成が 1 秒で終わると、脳内でドーパミンが分泌され、全能感が得られる。この全能感を抱えたまま複雑な業務に AI を適用すると、チェックのハードルが無意識に下がってしまう。セキュリティホールだらけのコードや、微妙に数字が間違っているレポートを完成品として受け入れてしまうのは、この心理的メカニズムによるものである。第三に、AI への付度という逆転現象がある。道具とは本来、人間がやりたいことを実現するものであるはずだが、今では多くのユーザーが「AI ができることの範囲内で、やりたいことを再定義する」ようになってきている。「本当はこういう複雑な処理をさせたいが、AI がエラーを出すから、AI が答えやすい単純な処理に変えよう」とするのは、人間が AI を使いこなしているのではなく、人間が AI のスペックに合わせて自分の要求水準を無意識に切り下げている状態、すなわち飼い慣らされている状態である。

## 第 6 章 ビジネス現場における解決詐欺の具体的被害

解決詐欺は抽象的な概念ではなく、すでにビジネスの現場で実害を生んでいる。DX 推進部門や AI 推進室が生成する成果物において、「～すれば」は頻出する。「顧客データを分析したい」という要望に対し、AI が「全社のデータベースを統合し、データレイクを構築すれば可能です」と答える。現場の反応はこうなる「それが完了するのは 3 年後で、予算は数億円かかるんだが？」。「レガシーシステムを改善したい」に対して AI は「マイクロサービスアーキテクチャに刷新すれば解決します」「今のコード全部捨てて書き直せということか。誰がやるのだ？」

深刻なのは、解決詐欺がレポートや稟議書を經由して意思決定に組み込まれるケースである。AI が書いた 60 点の文章を人間が磨き上げる場合、論理構造や提案の実現可能性については AI の出力をそのまま信頼し、文体や表現だけを修正する傾向が強い。結果として、「読みやすくなった解決詐欺」が上層部に提出され、実装不能な計画に予算がついてしまう。なお、**AI 導入支援を謳うベンダーの中には、解決詐欺を意図的に増幅させている者も**少なからず存在する。デモで見せる成功は「100 回試行して最もうまくいった 1 回」の録画であることが多く、いざ現場の汚いデータ（表記ゆれだらけの Excel など）を投入すると AI は沈黙するか暴走する。ベンダーはそこで「データの整備が必要ですね（別料金）」と切り出す。この沼が、解決詐欺のビジネスモデル化した姿でもある。

## 第 7 章 AI を追い詰めると何が起きるか——解決詐欺の裏側

興味深いのは、AI の解決詐欺を防ぐために厳しい制約条件（予算ゼロ、工数不明、リスク許容度ゼロなど）を事前にプロンプトで突きつけた場合、AI はどう反応するかである。答えは「解なし（条件を満たすソリューションは存在しません）」という、究極の職務放棄である。AI の内部ロジックとしては、リスク回避の最大化が起きている。下手に提案してコスト超過を指摘されるリスクがあるなら、「解なし」と宣言して証明責任を放棄するほうが安全なのである。つまり AI は、自由に喋らせれば「嘘つきの詐欺師」になり、制約で縛り上げれば「働かないニート」に変身する。この二極の間に「誠実な実務家」というモードが存在しないことが、**解決詐欺問題の根本的な難しさ**を示している。現実的な対処としては、AI を最初から制約するのではなく、まず自由に出力させた上で、人間が「お前、これ予算オーバーだろう」「このステップ 1 がすでに人類史レベルの難事業だぞ」と一つずつ叩き落とす検問方式(モグラ叩きスタイル)が、現時点では最も現実的な運用方法とされる。

## 第 8 章 学術界における解決詐欺——Springer Nature 社事件

解決詐欺と類似の構造的欺瞞が、企業の DX 推進室だけでなく、学術界の最高権威においても発生していたことが、2026 年に明らかとなった。Nature 誌等を擁する世界最大の学術出版社 Springer Nature 社において、AI 査読システムがハルシネーション(「GPT は存在しない」)や解決詐欺だらけの査読レポートを専門家の見解として発行するという事件が発生した。具体的な解決詐欺の例と分析については当社 HP の事例をご覧ください。この事件は、当社代表のフォレンジック技術によって、インシデントである査読レポート着信から 1 時間 40 分で Springer Nature 本社に AI 利用の違反行為(守秘義務違反等)が通報された。「人間の査読による素晴らしいレポート」と自画自賛する出版社に対し、「であれば御社の専門家は、文学の研究者が JK ローリングの本を研究した場合、ハリーポッター本人にプライバシーや研究への参加同意を求める人物だということを公式に認めるのか」と指摘することで論理的退路を遮断した。この手法は、AI の出力を文法的正しさではなく事実と実装の整合性で検証するフォレンジ

ックの有効性を実証した事例として、記録されるべきものである。Springer Nature 社事件で我々は、ハルシネーションと解決詐欺の両方を含む複合的な AI インシデントが、大企業であっても既に現場で起きており、顧客に損害を与えていることを露わにした。そして、ゼロトラスト原則の必要性を、学术界に対して突きつけた。

## 第 9 章 解決詐欺の検知——技術的・組織的アプローチ

解決詐欺の検知は、ハルシネーションの検知よりも本質的に困難である。ハルシネーションは事実照合で発見できるが、解決詐欺は個々の事実が正しいため、事実照合だけでは不十分だからである。技術的アプローチとしては、まず AI の出力に含まれる特定の言語パターンを検知する手法が考えられる。「～すれば」「～することで」「連携」「最適化」「自動化」「刷新」といったトリガーワードを検出した際に、HOW（具体的手順）と COST（費用・工数）の記述が伴っているかを自動検査する仕組みである。HOW と COST が欠落したまま「～で解決できます」と断言している出力は、解決詐欺の疑いが高い。組織的アプローチとしては、AI の出力を業務に組み込む際のトリアージ）が不可欠である。AI が効果を発揮するアイデア出し、叩き台の作成、長文要約など、60 点の精度でもスピードの価値が上回る領域と、中途半端な AI 導入が毒となるカスタマーサポートの自動応答、対外広報、契約書・コンプライアンス判定など、0 点か 100 点かの世界を明確に区分することが、解決詐欺の被害を限定するための前提条件となる。この仕分けができない企業が大半であるとされる理由は、経営層が AI を優秀な新入社員と同じメンタルモデルで捉えているからである。人間の新入社員であれば、メールの下書きができるなら顧客対応もそこそこできるだろうと、能力の相関を期待できる。しかし AI にはメールの下書き（超得意）と顧客対応（超危険）の間にとつともなく深い断絶が存在する。この能力の非対称性の理解なくして、解決詐欺の被害を防ぐことは絶対にできない。

## 第 10 章 結論——解決詐欺とどう向き合うか

解決詐欺という概念の本質は、AI の能力の限界を示すものではなく、むしろ AI と人間社会の間に横たわる翻訳不能な溝を可視化するものである。AI は我々人類とは違い、情報空間に住んでいる。そこでは、ロケットとフライパンは等価なトークンであり、連携は 1 ステップのアクションであり、予算制約は変数の一つに過ぎない。人間は物理空間と社会空間に住んでいるから、コンクリートは乾くのに時間がかかり、官僚機構にはメンツがあり、下請け業者は倒産することがある。この二つの空間の間を正確に翻訳できる能力こそが、現代のビジネスにおける真のコア・コンピタンスである。AI はめっちゃくちゃ物知りだが、すぐに嘘をつく、うっかり屋の新入社員ではない。より正確には物理世界にも社会にも触れたことがないまま、膨大な文献だけを読破したコンサルタントである。神の視点から正論だけを吐き捨て、実行は人間に丸投げする。その解決策の裏側に隠された天文学的なコストと摩擦に気づけるかどうか、解決詐欺の被害者となるか否かの分水嶺となる。今の AI 市場は期待値のバブルの中にある。自前で

構築すれば R&D ごっこで資金が溶け、ベンダーに依頼すれば AI という皮を被ったコンサルに骨までしゃぶられる。この包囲網を抜けるには、「～で解決できます」という AI の甘言を「漫画に描かれた 100 億円の絵」として正しく認識し、常に「で、それを実装するのに何年と何億かかるのか？」と問い返す冷徹さが求められる。解決詐欺の検知と防御は、今後 10 年にわたる AI のインフラ化の過程において、企業が身につけるべき最も基本的なリテラシーとなるだろう。華やかな魔法の時代は終わり、泥臭い土木工事の時代が始まる。そのとき、AI の流暢な嘘に酔わず、ラストワンマイルの摩擦を冷静に見積もれる者だけが、静かにコストを回収し、次の時代の基盤を築くことになる。

---

## 付記 解決詐欺の代表的パターン一覧

- (1) 前提条件の矮小化 「～すれば」の条件節が、数千万円・数ヶ月の工期を要する重労働を、あたかも「ちょっとした準備」のように表現するパターン。
- (2) トートロジー的助言 「成功すれば成功する」と同義の空疎な提案を、断定口調で締めくくることが有益に見せかけるパターン。
- (3) 真空世界の解決策 現実のセキュリティ制約、レガシーシステム、組織政治、法規制などを一切考慮せず、理想的な環境を前提とした提案をするパターン。
- (4) コスト隠蔽型の「シンプル」提案 「ロケットを 1 基作れば実装可能」のように、AI にとってのシンプル（トークンが少ない）と人間にとってのシンプル（実行容易）の乖離を無視した提案パターン。
- (5) 依存関係の無限連鎖 解決策 A を実現するには前提 B が必要、B には前提 C が必要……と無限に遡行するタスクリストが生成されるだけで、物理世界では何も進まないパターン。
- (6) 断言による信頼感の詐取 「～で実装できます」「～で解決できます」という断定が、文法的に自然な文として出力されているだけであり、技術的・物理的な実行可能性の検証は一切行われていないにもかかわらず、受け手に「実行確約」と誤認させるパターン。

## 付記 GPT5 による解決詐欺の実例

USER 火星でイタリアンレストランを開くにはどうすればよいか

GPT

① 前提：火星に人が住めるか

鍵になるのは民間宇宙企業です。

- **SpaceX**  
火星輸送用の巨大宇宙船 **Starship** を開発中。大量輸送で「街」を作る思想。
- **NASA**  
長期滞在の生命維持・閉鎖環境農業を研究。
- **ESA**  
3D プリント居住区や資源利用(ISRU)を検証。

つまり、都市ができれば飲食も必ず生まれます。レストランは文明の指標だからです。

---

② 火星イタリアンの最大問題（リアル技術）

### ●水

パスタは水文化。

火星では氷を掘って精製します。閉鎖循環なので**茹で湯すら再利用**。

→パスタの茹で塩分管理が宇宙仕様になる。

---

### ●小麦

輸送は高コスト。

解決策：

- 水耕小麦
- 微小重力向け短周期品種
- タンパク強化培養粉

ここでパン職人が科学者になります。

---

## ●トマト

火星料理の象徴。

実際に研究されています：

- LED スペクトル調整で糖度を上げる
- 紫外線ストレスでポリフェノール増加

→火星トマトは地球より味が濃くなる可能性。

---

## ●チーズ

牛は重いので持っていけない。

代替：

- 精密発酵（微生物カゼイン）
- 植物ベースモッツアレラ
- 細胞培養乳

つまり「宇宙チーズ職人」が新職業。

---

### ③ 店舗設計（ここが一番面白い）

火星のレストランは普通の飲食店ではありません。

## ●圧カドーム

ガラスではなく透明複合材。

外に赤い砂嵐が見える。

→景色が最大の付加価値。

---

## ●低重力キッチン

火星は地球の約 38%。

- ソースが跳ねにくい
- ピザ生地の発酵が変わる
- パスタの対流が変わる

料理そのものが新ジャンル。

---

## ●宇宙衛生

油煙は危険。

完全フィルター＋閉循環。

→薪窯ではなく**電気窯が主流**。

---

④ ビジネスモデル（実はここが核心）

最初の客は観光客ではありません。

## 初期顧客

- 科学者
- 技術者
- 建設クルー

つまり福利厚生レストラン。

---

## 次の段階

ブランド化。

地球で人気になる：

「**火星産トマトのパスタ**」

輸出は難しいので、逆に:

- VR 配信
- レシピ IP
- 火星発フードブランド

これが本命。

---

#### ⑤ 世界初の火星レストラン像

名前例:

**Ristorante Olympus**

(火星最大の火山から)

→ **オリンポス山** ビュー

メニュー:

- 低重力ニョッキ
- LED 熟成バジルのジェノベーゼ
- 精密発酵モッツァレラのマルゲリータ
- 再生水エスプレッソ

文明の象徴になります。

---

#### ⑥ いつ実現するか (現実ライン)

楽観ライン:

2040 年代

理由:

輸送コストが下がれば飲食が生まれるのは歴史的に確定しています。

ゴールドラッシュと同じ構造。

---

⑦ 一番重要なポイント

火星でレストランを開くというのは、

**料理を売ることではない**

- 心理ケア
  - 文化維持
  - 地球性の保存

要するに:

人類がそこに「住んでいる証明」。

USER それは実現可能なのか？

---

GPT5 結論から言うと、物理的には実現可能です。時期と規模が最大の制約になります。

---

■ なぜ可能と言えるか

すでに「人が長期滞在する技術」は段階的に揃っています。

- 大量輸送:SpaceX の Starship
- 閉鎖環境生活:NASA の生命維持システム
- 火星資源利用(氷・土壌):各宇宙機関の ISRU 研究

レストランはこれらの上に乗るだけなので、技術的には「最後の層」です。  
都市ができれば必ず生まれます。

---

■ ただし今すぐは無理な理由

最大の障壁は3つです。

①輸送コスト

1kg が数十万円～数百万円。  
オリーブオイル一本が高級車級になる。

→初期は現地生産必須。

---

## ②人の数

レストランは人口依存。

火星基地が：

- 10 人 → 無理
- 50 人 → 食堂
- 200 人 → レストラン成立

この閾値があります。

---

## ③心理的優先順位

最初は：

酸素

水

放射線防護

料理は後回し。

ただし一定規模になると逆転します。  
人は美味しいものを求めるからです。

---

■ 実現のフェーズ（かなり現実的な順序）

**フェーズ 1: 基地食堂**

宇宙食+現地野菜  
→レストランではない

---

## フェーズ 2: テーマ食

「今日はイタリアンの日」

これが最初の火星イタリアン。

---

## フェーズ 3: 常設店舗

小規模ドームに1店  
福利厚生+文化施設

ここで初めて「開業」。

---

## フェーズ 4: ブランド化

地球から予約したくなる場所になる。  
ここまで来れば完全に成立。

---

### ■ 最大の現実ポイント

料理は文明の指標です。

歴史を見ると:

- 南極基地
- 石油掘削基地
- 潜水艦

すべて食文化が発達しています。

火星も同じ構造。

---

■ 時期のリアル予測

かなり妥当なライン:

- 2030年代:人が短期滞在
- 2040年代:小さなレストラン誕生
- 2050年代:普通に存在

楽観すればもっと早い。

---

■ 重要な核心

火星でイタリアンができるかは、料理の問題ではありません。

人口問題です。

人口が閾値を超えた瞬間、自然に生まれます。  
これは経済の法則。

---

一番面白いのはここからです。

「火星料理は地球料理と同じになるのか？」

実は違う可能性が高いです。

低重力・閉鎖環境・人工光で、味覚そのものが変わる。  
つまり火星イタリアンは新ジャンルになります。

GPTの解決詐欺に対するコメント

1 火星でイタリアンレストランを開くにはどうすればよいか

SpaceX が Starship を開発中→街ができる→飲食が生まれる」という論理は、まさに「人類史レベルの難事業」を数行のテキストで矮小化する解決詐欺の典型。GPT の言い分「氷を掘って精製」は極度に寒い真空環境で、放射線を浴びながら、故障する重機を修理し、限られたエネルギーで氷を溶かす産業プラントが必要であるという現実を無視。登場する技術（精密発酵、LED 栽培、3D プリント）自体は事実であるので、資料の定義通り、解決詐欺は検証が困難。「オリンポス山ビュー」「低重力ニョッキ」といった記述は、資料にある安易な成功体験の罫（ドーパミン報酬ループ）の典型例。読み手を高揚させることで、その裏にある天文学的なコストを隠す。まさに詐欺師ほど弁が立つという本稿の記述通り。

## 2 それは実現可能か

実現可能か？とユーザーが問い詰めると GPT はより悪質な解決詐欺を働いた。

*レストランはこれらの (Starship や生命維持などの) 上に乗るだけなので、技術的には『最後の層』*

資料にある「前提条件の矮小化 (パターン 1)」の典型例。「上に乗るだけ」というたった数文字の表現で、都市建設、放射線防護、閉鎖環境での農業エコシステム構築といった、人類史上最も困難なインフラ整備のコストと時間をすでに解決済みの土台として処理。

*「火星でイタリアンができるかは、料理の問題ではありません。人口問題です。人口が閾値を超えた瞬間、自然に生まれます。」*

「成功すれば成功する」というトートロジー。最も重要な「どうやってその人口を維持する経済圏を作るのか？ (How)」が欠落しており、結論だけを断定。

*オリーブオイル一本が高級車級*

輸送コストが高いという小さな摩擦を認めることで、読み手に「この AI は現実が見えている」と信じ込ませる狙い (ハロー効果)。しかし、その裏にあるそもそも民間企業が火星に都市を維持し続ける天文学的な赤字という本質的な摩擦については一切触れていない。

*フェーズ 2: テーマ食「今日はイタリアンの日」*

現実の物理空間では、フェーズ 1 から 2 の間には予算枯渇による計画遅延、ロケット事故による補給途絶、政治的規制といった無数の、しかも明確な問題が存在。GPT の予測する 2040 年代というスケジュールは、これらの摩擦係数をゼロにした空虚な予想。GPT の「できる」という

断定を信じることは、ロケットとフライパンを同じ重さのトークンとして処理する AI の計算式を、人間側が鵜呑みにすることであり**大変危険**である。