4　頻出ひっかけ問題の整理と対策

⑴　知識拡張アプローチ（RAG 対ファインチューニング）

企業が ChatGPT や Gemini などの汎用的な生成 AI モデルを実務に導入する際、直面する最大の壁が「AI は自社の内部情報や最新情報を知らない」という事実である。汎用モデルに独自の知識や特性を反映させるためのアプローチには、大きく分けて「RAG（検索拡張生成）」と「ファインチューニング（微調整）」の２つが存在する。2026 年のシラバスでは、この両者の仕組みと適用シーンの違いを正確に区別できるかが厳しく問われる。

①　RAG（Retrieval-Augmented Generation：検索拡張生成）

RAG は、AI が回答を生成する前に、外部のデータベースからユーザーの質問に関連する情報を「検索（Retrieval）」し、その検索結果をプロンプトの文脈（Context）に「拡張（Augmented）」して組み込んだ上で、回答を「生成（Generation）」させる仕組みである。

詳しくは、第２時間の「RAG」を参照し振り返っておきたい。おさらいとなるが、RAG の最大のメリットは、ハルシネーションの劇的な抑制と情報更新の容易さにあったのを思い出したい。

②　ファインチューニング（Fine-Tuning：微調整）

ファインチューニングは、プロンプトの工夫や外部データベースの検索によるアプローチとは根本的に異なり、事前学習済みの AI モデルに対し、追加の訓練データを与えてモデル自身の「重み（パラメータ）」を再学習・更新する手法である。

これは、特定の専門用語を AI の内部記憶として深く定着させたり、企業独自の特定の口調や複雑な出力フォーマット（トーン＆マナー）を安定的に再現させたりする場合に有効である。しかし、知識の追加目的でファインチューニングを用いることは現在では推奨されていない。なぜなら、情報が古くなるたびに再学習の計算コストと時間がかかる上、AI が「どの情報源に基づいて回答したか」という根拠を提示することが構造上困難であり、ハルシネーションのリスクを払拭しきれないためである。

比較項目	RAG	ファインチューニング
仕組みの核心	外部 DB から関連情報を検索し、プロンプトに結合して生成する	追加データを用いて、モデル自体のパラメータ（重み）を再学習させる
最適な適用シーン	最新情報の参照、社内規程に基づく回答、ハルシネーションの抑制	文体・口調の統一、特定フォーマットへの適合、応答スタイルの安定化
情報源の明示	検索した引用元（根拠箇所）を提示することが容易	内部化されたパラメータから生成するため、根拠の明示が困難
コストと運用	DB の更新のみで即座に情報反映が可能。計算コストが低い	情報更新のたびに再学習が必要。計算コストが高く専門知識を要する

2026 年改訂シラバス準拠例題

社内マニュアルに関する質問に対応する AI システムにおいて、AI が実在しない業務手順を断定して出力するハルシネーションが発生した。今後の運用において、この問題を根本的に抑制し、信頼性を担保するための再発防止策として最も適切なものはどれか。

① Temperature の数値を 1.0 に近づけ、多様な回答を出させるよう設定を変更する。
② RAG の仕組みを導入し、出力に根拠（引用元）を表示させ、ユーザーが一次資料を確認する運用にする。
③ プロンプトに「嘘をつかないでください」とだけ記載し、AI の倫理観に依存する。
④ 全ての社内文書を用いて、毎日モデルのファインチューニングを実行して内部知識を更新する。

【解説】 **正答：②**

RAG の仕組みを導入し、出力に根拠（引用元）を表示させ、ユーザーが一次資料を確認する運用にする。解説：ハルシネーションを完全にゼロにすることは現在の生成 AI の技術的限界である。したがって、RAG を用いて根拠箇所を提示し、最終的に人間が一次資料でファクトチェックを行う運用（Human-in-the-loop）が、最も現実的かつ安全な対策である。①はハルシネーションを助長し、③は効果が薄く、④はコストと運用の観点から非現実的である。

⑵ データプライバシーと情報セキュリティ

生成 AI を企業で安全に利用するためには、技術的側面に加えて、個人情報保護法やサイバーセキュリティに関する法的・実務的な知識が不可欠である。特にデータの取り扱い区分や、AI を悪用した新たな脅威については、用語の定義を正確に暗記しておく必要がある。

① データの取り扱い区分：「要配慮個人情報」と「匿名加工情報」

企業が保有するデータを AI に読み込ませる際、そのデータの性質によって法的な取り扱いが大きく異なる。試験では、この定義の境界線を狙った問題が頻出する。

日本の個人情報保護法において、本人の人種、信条、社会的身分、**病歴**、犯罪の経歴など、本人の不当な差別や偏見が生じないように特に慎重な取り扱いを要する情報を「要配慮個人情報」と定義している。社内文書を AI に要約させる際、社員の健康診断結果や病歴などが含まれている場合は、原則として本人の明示的な事前同意が必要となる。

上記の「本人の人種、信条、社会的身分、病歴、犯罪の経歴」は要配慮個人情報であり、氏名や住所などの通常の個人情報と区別されることに注意。

一方で、企業がデータを統計分析や AI の学習・検証データとして第三者に提供し、安全に利活用するための枠組みが「匿名加工情報」である。これは、特定の個人を識別することができないように個人情報を加工し、かつ、元の個人情報を復元できないようにした情報を指す。試験における典型的なひっかけとして、「データから氏名と住所を削除すれば、自動的に匿名加工情報として自由に扱ってよい」というものがある。これは誤りである。マイナンバーやパスポート番号、あるいは防犯カメラの顔認証データなど、それ単体で特定の個人を識別できる「個人識別符号」が含まれていれば、それは依然として個人情報に該当するため、より厳格で不可逆的な加工措置が求められる点に注意が必要である。

匿名加工情報は「個人識別性を下げ、復元困難にしたもの」であり、**「マスキング不要」や「同意不要で無制限共有できる」**といった選択肢は誤りなので注意。

⑶　AI を悪用したサイバー脅威とインシデント対応

生成 AI の普及により、サイバー攻撃はより巧妙かつ低コストで実行可能となっている。セキュリティの観点から、以下の用語の正確な理解が求められる。

ディープフェイク（深層偽造）

AI 技術を用いて、実在の人物が実際には行っていない発言や行動をしているかのように見せかける合成音声や動画を指す。株価操作、選挙における世論誘導、または「本人のなりすましによる詐欺」など、深刻な社会的・経済的リスクを引き起こす。

偽情報の分類（ディスインフォメーションとミスインフォメーション）

この２つの違いは明確に問われる。発信者が事実ではないと知りながら、他人を騙す、あるいは社会を混乱させる「悪意を持って」意図的に拡散される偽情報が「ディスインフォメーション」である。一方、発信者に悪意はなく、単なる勘違いや誤認によって拡散されてしまう誤情報が「ミスインフォメーション」である。

ランサムウェアと初動対応

コンピュータやネットワークに感染するとシステムのデータを暗号化し、復旧の対価（身代金）を要求するマルウェアである。試験では感染時の初動対応が問われる。被害拡大（他の端末への水平展開）を防ぐための「ネットワーク（LAN や Wi-Fi）からの物理的・論理的な即時遮断」と「組織のルールに従った報告」が初動の基本原則である。