

第1 本ガイドラインの目的

本ガイドラインは、AIで生成したバーチャルヒューマン（以下「AIモデル」といいます）を、広告・販促、接客・カスタマーサポート（以下「CS」といいます）、SNS運用等の事業活動に利用する事業者（以下「利用者」といいます）が、法的・倫理的なリスクを低減し、安心・安全で信頼性のある活用を行うために、利用者が遵守すべき事項および留意すべき事項を整理することを目的とします。

本ガイドラインは、AIモデルの利活用を過度に萎縮させることを意図するものではありません。他方で、AIモデルの利活用が社会に受け入れられ、市場が健全に発展していくためには、権利侵害、誤認を誘う表示、虚偽・誇大な訴求、なりすまし、フェイクレビュー等の不適切な利用が抑制され、説明責任を果たせる運用が確立される必要があります。利用者は、その責任主体として、適正な利用を自ら設計し、継続的に改善しなければなりません。

【解説】

利用者は「AIを使う側」とすると同時に、広告や発信の主体であり、受け手（消費者・視聴者・顧客等）に対して最終的な説明責任を負います。「AIが生成した」「AIがそう言った」という事実は、責任の所在を移転するものではありません。本ガイドラインは、その前提に立ち、利用者が日常運用として実装すべき行為を文章の形で示すものです。

第2 本ガイドラインの対象と構成

本ガイドラインが対象とするAIモデルとは、AIによって生成された、実在しない人物表現（外見、声、人格設定、発話、所作等）を含むものをいいます。画像・動画・音声・テキスト等の形式を問いません。

本ガイドラインは、利用者がAIモデルを「利用する」局面に焦点を当て、(1)全体に共通する遵守事項、(2)想定用途別の留意事項、の順に構成します。想定用途とは、広告・販促、接客・CS、SNS運用を指します。

【解説】

AIモデルの制作・提供の形態（フルスクラッチ、外部モデル利用、追加学習等）により、開発・提供側の義務や技術的対策の範囲は異なります。しかし、利用者にとって重要なのは、「自らの利用が、第三者の権利侵害や消費者の誤認・被害につながらない状態」を、契約・運用・表示・監視によって担保することです。本ガイドラインは、利用者が提供事業者の形態差を踏まえつつも、利用の現場で必要な統制を外さないことを重視します。

第3 前提条件（利用者が満たすべき前提）

1 責任主体の明確化

利用者は、AIモデルを利用して発信・表示・応対を行う場合、当該発信・表示・応対の責任主体が誰であるかを明確にしなければなりません。具体的には、利用者内部において、

最終承認責任者、運用責任者、緊急停止責任者を定め、承認・差し止め・訂正・削除を実行できる体制を整備するものとします。

2 利用環境の把握と提供事業者との連携

利用者は、利用する AI モデルサービスの仕様、禁止事項、出力物の取り扱い条件、ログやデータの保存方針、ならびに権利侵害の疑義が生じた場合の連絡・対応手順を把握し、必要に応じて提供事業者と共有・連携しなければなりません。運用環境の変更、利用範囲の拡大、配信面の変更等がある場合には、リスクが増大する可能性があるため、利用者は提供事業者に情報共有し、必要な更新・再審査を行うことに努めなければなりません。

3 検証可能性（説明できる状態）の確保

利用者は、AI モデルの利用が社会的に問題となった場合に、当該利用が適正であったかを検証できる状態を確保しなければなりません。具体的には、制作・利用の経緯（企画意図、生成指示、参照素材、編集履歴、掲出面、掲出期間、承認者、修正履歴等）を、合理的な範囲で記録・保存し、必要に応じて提示できるようにするものとします。

【解説】

「検証可能性」は、法的紛争に備えるだけでなく、炎上時の迅速な原因切り分けと再発防止に直結します。広告、SNS、CS はいずれも、問題が表面化してから対応を始めると被害が拡大します。運用開始時点で「後から説明できる」設計にしておくことが、事業継続の観点でも重要です。

第4 全体に共通する遵守事項（利用者が守らなければならないこと）

本章では、想定用途にかかわらず、利用者が共通して遵守すべき事項を定めます。

1 第三者の権利侵害を目的とする利用の禁止

利用者は、第三者の著作物、第三者のキャラクター、第三者の商標・ブランド、ならびに特定の実在人物（著名人を含みます）の肖像や声、特徴、社会的信用等に依拠し、当該第三者の権利・利益を侵害することを目的として AI モデルを利用してはなりません。

利用者は、生成指示（プロンプト等）において、特定の権利者名、特定の作品名、特定のキャラクター名、特定の著名人名、特定の実在人物名等を用いて「それに寄せる」「それを再現する」趣旨の指示を行うべきではなく、そのような指示が実務上混入しない運用を整備しなければなりません。

【解説】

「意図していないが似た」ケースも起こり得ますが、そもそも「寄せる意図」を示す指示や制作メモが残っていると、社会的にも法的にも不利になりやすいことに注意が必要です。利用者は、制作会社・代理店・外部委託先を含め、指示文やクリエイティブブリーフの段階から統制する必要があります。

2 類似性の抑制（直接侵害・間接侵害の回避）

利用者は、生成された AI モデルまたはその生成物が、既存の著作物、特定の著名人の肖

像、特定の実在人物の肖像等と同一または高度に類似する可能性がある場合には、当該生成物を利用（掲出、投稿、配布、販売、アップロード等）してはなりません。

利用者は、制作・入稿・公開の前に、合理的な範囲で類似性確認を行い、疑義がある場合には差し止め・差し替えを行うものとします。また、利用者が提供するサービスやキャンペーン等において第三者が生成機能を利用できる場合には、利用者は、当該第三者による権利侵害（間接侵害を含む）を誘発しないよう、入力制限、利用規約、モデレーション等の措置を講じなければなりません。

【解説】

利用者が自社で生成する場合だけでなく、ユーザー参加型企画（「あなたも AI モデルを作れる」等）では、利用者が設計した仕組み自体が問題視されやすくなります。入力制限や規約の整備を「提供事業者の責任」として放置せず、利用者の企画責任として実装することが重要です。

3 違法・有害コンテンツへの不関与（特に児童・性的領域）

利用者は、児童に関する違法コンテンツ、搾取を助長する表現、同意なき性的表現、またはそれらを想起させる素材の収集・保管・利用・拡散に関与してはなりません。

利用者は、AI モデルを用いた表現を取り扱うにあたり、企画・制作・配信・コメント運用等の各段階で、違法・有害コンテンツが生成・流通しないよう、禁止事項、審査、監視、通報、削除、再発防止の手順を整備するものとします。

【解説】

違法・有害領域は、意図の有無にかかわらず社会的許容度が極めて低く、被害回復が困難です。運用上は「起こさない」設計（入力制限、審査、監視）と「起きたら止める」設計（即時停止・削除・通報）を分けて整備することが有効です。

4 誤認の防止（AI であること、広告であること等）

利用者は、受け手が、AI モデルを実在人物であると誤認すること、または、実在人物が推奨・監修・出演していると誤認することがないように、表示・演出・文言・UI 設計を行わなければなりません。

利用者は、広告・PR・タイアップ等の商業的コミュニケーションにおいては、それが広告であることが受け手に分かるように表示し、AI モデルの利用が誤認を生みやすい態様である場合には、AI による表現であることを適切に示すことに努めなければなりません。

【解説】

誤認は「小さい注釈を入れたか」だけで決まりません。全体の印象、文脈、見出し、サムネイル、口調、権威付けの演出によって、注釈があっても誤認が残る場合があります。利用者は、注釈を“免罪符”としてではなく、誤認を起こしにくい設計の一部として位置づける必要があります。

5 虚偽・誇大な表示の禁止（AI ウォッシングの抑制）

利用者は、AI という語の印象や先端性を利用して、実態が伴わないにもかかわらず、性

能・効果・安全性・優位性等を過度に強調し、受け手に誤認を与える表示（いわゆる AI ウォッシングを含みます）を行ってはなりません。

利用者は、効果・実績・比較・No.1等の主張を行う場合には、根拠となるデータ・条件・範囲・例外を整理し、必要に応じて受け手が確認できる形で提示しなければなりません。また、投資・副業・健康・医療・金融等の分野において、確実性を断定する表現、著名人や専門家を装う表現、フェイクレビュー等による誘引は、被害を拡大させやすいことを踏まえ、特に慎重でなければなりません。

【解説】

海外では、生成 AI を利用した虚偽・誇大広告、著名人になりすました勧誘、フェイクレビュー等が社会問題化し、監視・規制・法執行の強化が進められています。利用者は「自社は善意だから大丈夫」と考えるのではなく、受け手の視点で誤認可能性を評価し、断定・煽り・権威付けを抑制した表現設計にする必要があります。

6 利用規約・運用ルールの整備（第三者利用がある場合）

利用者が、AI モデル生成機能または類似の生成機能を第三者に提供する場合、利用者は、サービス利用規約または運用ルールにおいて、権利侵害を目的とする入力、特定の人物や作品に類似させる生成、類似物が生成された場合の利用（アップロード、頒布、販売等）、ならびに不適切用途（アダルト目的等）を禁止事項として定め、これを実効的に運用するものとします。

利用者は、禁止事項を定めるだけで足りると考えるべきではなく、技術的制限、審査、監視、通報対応等を組み合わせ、実効性を確保しなければなりません。

【解説】

規約だけを置いて技術的に無制限の生成を可能にすると、「禁止しているのに大量に生成される」状態が起り得ます。利用者は、規約・技術・運用の三点セットで実効性を担保する必要があります。

7 情報提供とインシデント対応

利用者は、AI モデルを用いたサービスやコミュニケーションにおいて、受け手が適切に判断できるよう、必要な範囲で、AI の利用事実、適切な利用方法、制約や限界、不具合や誤りが生じ得ること、ならびに問い合わせ先等の情報を提供するよう努めなければなりません。

また、権利侵害の疑義、誤認を招く表示、虚偽・誇大の疑義、個人情報の問題、違法・有害コンテンツの流通等のインシデントが発生した場合には、利用者は、速やかに停止・差し止め・訂正・削除等の措置を講じ、原因分析及再発防止策を実施しなければなりません。

【解説】

「速やかに」とは、理想論ではなく運用要件です。停止権限や連絡網が整備されていないと、最初の数時間で被害が拡大します。平時から、緊急停止手順と責任者を固定し、外部

委託先にも共有しておくことが重要です。

第5 想定用途別の留意事項（用途ごとに追加で守るべきこと）

本章では、AIモデルの利用において特に問題が生じやすい局面を想定し、用途ごとの留意事項を定めます。利用者は、前章の共通遵守事項に加え、本章の内容を用途に応じて適用しなければなりません。

第5-1 広告・販促における留意事項

広告・販促においては、AIモデルの表現が、受け手に対して強い影響を与え、購入・申込・契約等の意思決定に直結します。そのため、誤認防止と虚偽・誇大の抑制を、他の用途以上に厳格に運用しなければなりません。

まず利用者は、AIモデルの外見・声・肩書・語り口を用いて、専門性や権威を演出する場合、当該演出が虚偽となっていないか、また受け手に不当な確実性を感じさせていないかを点検しなければなりません。AIモデルに「専門家」「医師」「金融のプロ」「元〇〇」等の属性を付与することは、誤認のリスクを高めるため、事実に基づく裏付けのない設定をしてはなりません。仮にキャラクター設定として行う場合であっても、受け手が現実の資格や職業と誤認しない表現・導線にする必要があります。

次に利用者は、効果・実績・比較・No.1等の主張を行う広告表現を採用する場合には、その根拠を整理し、条件・範囲・例外を含めて説明できる状態にしなければなりません。特に、投資・副業・健康・医療・金融等の領域では、断定的な成果保証や、著名人・専門家の推薦に見せかける手法、体験談を装う表現は、消費者被害と結びつきやすいことを踏まえ、抑制的に運用すべきです。

さらに利用者は、広告制作プロセスにおいて、生成指示、参照素材、編集履歴、差し替え履歴等を記録・保存し、掲出後に指摘があった場合に迅速に原因を特定できるようにするものとします。広告は配信量が大きく拡散力が高いため、問題が起きた場合の損害が大きく、停止の遅れがそのまま被害拡大につながります。ゆえに、広告配信においては、緊急停止の実行権限をあらかじめ確保し、外部代理店・媒体運用者とも停止条件を共有しなければなりません。

第5-2 接客・CSにおける留意事項

接客・CSにおいては、AIモデルの対応が、顧客の権利義務（契約、解約、返金、支払等）や、安全・健康・信用（医療、金融、法務等）に影響し得ます。そのため利用者は、AIモデルの役割を適切に限定し、誤案内・誤約束・なりすまし等のリスクを抑制しなければなりません。

利用者は、顧客がAI対応であることを理解できるよう、対応の開始時点で、AIモデルが対応している旨を示し、必要に応じて人の担当者へ切り替えられる手段を用意するものと

します。重要手続き（解約、返金、支払い、本人確認を伴う変更等）については、AI の案内のみで完結させることが事故につながりやすいため、確認画面、同意、担当者への引き継ぎ等を含む安全設計を行う必要があります。

また利用者は、AI モデルの対応が、事実として確定していない事項を「完了した」「保証する」「必ずできる」などと断定しないよう、応答テンプレートと禁止表現を整備し、誤案内が起きた場合の修正手順を定めなければなりません。AI モデルが推測で回答する領域と、システム上の真実（在庫、契約状態、配送状況等）に基づいて回答すべき領域を分離し、後者は必ず正規データに基づく回答となるよう設計することが望まれます。

加えて、接客・CS では個人情報を取り扱いやすいことから、利用者は収集最小化を徹底し、公開コメント等で個人情報を求める運用を行ってはなりません。本人確認が必要な手続きは専用フローに分離し、第三者によるなりすましや不正要求に応じない統制を組み込む必要があります。あわせて、対応ログを合理的な範囲で保存し、誤案内の検知、品質改善、インシデント対応に活用できる状態を確保するものとします。

第 5-3 SNS 運用における留意事項

SNS 運用においては、AI モデルが“日常的な人格”として受け手に接触し続けるため、誤認、炎上、二次拡散、なりすまし、誘導型詐欺の温床化等のリスクが高まります。利用者は、アカウント設計、投稿設計、コメント・DM 運用、危機対応の各段階で、共通遵守事項をより実務的に適用しなければなりません。

利用者はまず、アカウントのプロフィール等において、AI モデル（バーチャル）であること、運営主体、問い合わせ先を明確にし、受け手が実在人物アカウントと誤認しないようにするものとします。加えて、タイアップ投稿や販促目的の投稿では、それが広告・PR であることを分かりやすく示し、受け手の誤認を抑制しなければなりません。AI モデルの人格が強いほど、広告であることが見落とされやすいことに留意が必要です。

投稿内容について、利用者は、短尺で断定しがちな SNS の特性を踏まえ、比較煽り、断定的成果保証、権威を装う語り、体験談の捏造に該当し得る表現を抑制しなければなりません。特に、投資・副業・健康・医療・金融等は、誤情報が被害に直結するため、一般論の案内に留め、条件・例外・注意事項を添え、必要に応じて公式情報や窓口へ誘導する運用とすることが望まれます。

コメントや DM の運用においては、利用者は、個人情報の収集を安易に行わず、センシティブな相談（健康、金銭トラブル、法的紛争等）では、AI モデルが断定的に助言することを避け、人の担当者または適切な専門窓口につなぐべきです。また、SNS では偽アカウントが生じやすいため、利用者は「公式以外から金銭や個人情報を求めない」等の注意喚起を行い、偽アカウントの通報・告知のフローを整備するものとします。

最後に、SNS は炎上時の初動が被害を左右します。利用者は、投稿停止・非公開化・削除・訂正・説明の判断基準、責任者、実行手順を平時から定め、発生時には躊躇なく措置

を講じなければなりません。AI モデルの発言は切り抜き拡散されやすいことを前提に、「誤解されない表現」を目指すだけでなく、「誤解されたときに被害を抑える設計」まで含めて運用することが必要です。

付記（本ガイドラインの運用）

利用者は、本ガイドラインを一度整備して終わりとせず、実際の運用で生じた問題、受け手の反応、法規制・業界動向の変化、提供事業者の仕様変更等を踏まえて、継続的に見直しを行うものとします。特に、AI モデルの利用が拡大し、用途が多様化するほど、誤認・誇大・権利侵害・悪用のリスクは複合的になります。利用者は、ガイドラインを社内規程、制作フロー、入稿審査、CS 手順、SNS 運用ルールに落とし込み、実効性を確保しなければなりません。