ヘルスケア領域において生成AIを活用したサービスを提供する事業者が参照するための自主ガイドライン
（ヘルスケア生成AI活用ガイド）
第2.0版

2024年1月18日　策定
2025年2月7日　改正

日本デジタルヘルスアライアンス
WG4 SubWG-B
デジタルヘルスアプリの適切な選択と利活用を促す社会システム創造WG　-生成AIに関する検討-

＜目次＞

1. はじめに

1-1. 背景

インターネット技術の普及とともにSNS等によってユーザー生成コンテンツが増加したことに伴い、機械学習に用いることのできるデータが急激に増加した。さらに、計算機単体の性能向上及び大規模な分散処理化の実現により、大量のデータを効率よく処理できる環境が整った結果、あらゆる産業においてAIの活用が進み、2023年初頭に生成AI（Generative　AI）と呼ばれる新技術が急速に全世界へ普及しはじめた。生成AIの技術そのものは数年前から存在し、AI技術の進歩の結果として高度化してきたものであるが、ChatGPTをはじめとする簡易かつ高性能なツールがリリースされたことで、AIに関する知識や技術がなくとも、多くの人が簡単にAIの機能を取り扱えるようになった点において極めて革新的である。2024年以降、生成AIを活用したプロダクトやサービスがより一層マーケットで展開されるようになったほか、テキスト生成に外部情報の検索を組み合わせることで回答精度を向上させる技術であるRAG（Retrieval-Augmented　Generation）の活用や、日本語に特化した国産の大規模言語モデルの開発、小規模言語モデル（SLM：Small Language Model）の登場など、生成AIを取り巻く技術環境は目まぐるしく進化し続けているといえる。また、規制動向としても、2024年5月に経済産業省・総務省が「AI事業者ガイドライン」を策定し、AI開発者・AI提供者・AI利用者の3主体別に取り組むべき指針が示されたほか、個人情報保護委員会においては個人情報保護法の3年見直しの議論が進められているところである。

ヘルスケア領域においても生成AIに関する取組や検討は例外ではなく、医師国家試験に合格するAIや要約機能・患者への説明文案の生成等により医師の業務効率化をサポートするAIなど、技術の可能性の探索や実際の医療現場等へのサービス提供といった取組が進められているところである。日本において非常に大きな社会的課題である医療・ヘルスケア分野（以下単に「ヘルスケア領域」という。）においてこうした技術革新により多様なサービスが創出され、課題解決に貢献していくことは技術を活用する我々事業者としても社会的・経済的価値貢献の観点から非常に歓迎することであるが、他方で、技術の急速な普及によりその副作用も生じうる点は考慮すべきである。

特にヘルスケア領域においては、要配慮個人情報の取扱いが多くなる点や、不確かな情報がもたらす個人への影響が極めて大きい点等を踏まえ、生成AIを用いてヘルスケアサービスを提供する事業者が留意すべき点をまとめることとしたものである。

なお、ヘルスケア領域における生成AI活用に関する業界策定ガイドラインは2024年1月18日に公表した本ガイドが初出であり（第1.0版公表時）2024年5月に公表された経済産業省・総務省策定のAI事業者ガイドラインにおいても本ガイドが「別添4.AI提供者向け　10）イノベーション」における　事例として掲載されているところである。

今般、上記のような2024年以降の技術動向や規制動向を踏まえた本文のアップデートや参考資料の充実等を図ることで、本ガイドが業界内で実効性が高く有用性が担保されたリビングドキュメントとして活用が図られることを目的として第2.0版を策定することとした。今後も生成AIの技術の進展等も踏まえ今後も適宜アップデートを行っていく予定である。

1-2. 本ガイドの前提

1-2-1. 生成AIとは

生成AI（Generative　AI）とは、自律的に学習したデータから文章、画像、音声などの一見新しく現実的なコンテンツを生成することができる一連のアルゴリズムのことをいう。最も強力な生成AIのアルゴリズムは、ラベルのない膨大な量のデータを自律的に学習して、幅広いタスク向けに基本パターンを特定する基盤モデルの上に構築される[1]。

[1] 参考文献
BCG：https://www.bcg.com/ja-jp/capabilities/artificial-intelligence/generative-ai
Radiology：https://pubs.rsna.org/doi/epdf/10.1148/radiol.223312

1-2-2. 本ガイドで対象とする範囲

生成AIによるアウトプットは文章、画像、音声など多様であり、かつ、マルチモーダルと呼ばれる複数機能を統合的に活用できる技術なども随時開発・提供されているところであるが、本ガイドにおいては、まずはヘルスケア領域で最も広く活用されていると考えられる文章（テキスト）生成AIを対象とする。なお、本ガイドは今後技術やサービスの進展を踏まえて随時アップデートを行う予定としている。

また、生成AIは新しくかつ急速に普及した技術であるためヘルスケア領域における安定的な活用方法はまだ検証段階であることや、一般生活者への影響が生じやすいことを踏まえ、本ガイドにおいては原則として医療機器又は医療機器プログラムには該当しないヘルスケアサービス（以下単に「ヘルスケアサービス」という。）を想定した記載とする。

1-2-3. 本ガイドの目的及び対象読者

冒頭に記載したとおり、特にヘルスケア領域におけるサービス提供にあたっては、他の領域と比較して要配慮個人情報の取扱いが多くなる点や、不確かな情報がもたらす個人への影響が極めて大きい点等が課題である。

これを踏まえ、生成AIを活用したヘルスケアサービスが利用者に不当な不利益を供することとならないよう、当該サービスを提供しようとする事業者がセルフチェックできる目安となるチェックポイントを提供することを目的として本ガイドを策定する。したがって、主には生成AIを活用したヘルスケアサービスを提供する事業者を対象読者とする。

併せて、本基準を踏まえてサービス提供している事業者が、その旨を公表等することによって、当該サービスを利用するユーザー・顧客（医療機関等）が、安全性の目安として確認し、安心してサービスを利用できるようになることも期待する。

なお、生成AIを活用したサービス・プロダクト提供自体を初めて経験する事業者も多くいることが想定されるため、当該事業者が本ガイドを簡便に参考とできるよう、別添のチェックリストもまとめている。また、本ガイドにおいて記載される用語については参考資料の用語集を参照されたい。

2. 生成AIに関する基礎情報

2-1. 生成AIのヘルスケア領域における動向

2017年にGoogle社が発表したTransformerと呼ばれる機械学習モデルにより、自然言語処理分野の技術水準は飛躍的に進展した。このモデルを用いて2018年にはGoogle社がBERTを、2020年にはOpenAI社がGPT-3を、そして2022年には同じくOpenAI社がChatGPTを発表した。

このTransformerモデルを利用した自然言語処理に関連する技術の一部は、現在ではテキスト生成AIと呼ばれている。テキスト生成AIは特に2022年11月にOpenAI社によって提供が開始されたChatGPTをきっかけに爆発的普及が進んでおり、2023年12月現在は医療分野を含めた様々な産業での活用が検討・推進されている。

テキスト生成AIのように自然言語処理タスクで質問応答文や文書など任意の自然言語を出力するために大量のデータで学習された多数のパラメータを持つモデルは、図1のとおり、一般に大規模言語モデル（Large Language Model：LLM）と呼ばれており、OpenAI社が提供するChatGPTを含めたGPTシリーズや、Google社のPaLMシリーズ、Meta社のLlamaシリーズなどが知られている。LLMは様々なタスクを実施することが可能であり、テキスト生成にとどまらず、質問応答・プログラムコードの生成・言語翻訳・文の校正・知識提供・クリエイティブな作品の生成などが代表的な使用例である。

また、生成AIはLLM以外の画像・音声・動画生成といった領域でも進展が著しく、ChatGPTを代表として、複合的に生成を行うことも可能となってきており、単一データ(Single Modal)・単一目的のものから、複合データ(Multi Modal)・汎用目的AI(General Purpose AI)のものへと発展してきており、ヘルスケア領域においてもより多様な可能性が示されつつある。

2024年における生成AIをめぐる技術動向として、まず注目すべきはRAG（Retrieval-Augmented Generation）の進出・高度化である。RAGは生成AIに外部のデータベース等からの情報検索を組み合わせ、検索で得られた正確な情報に基づいた応答を生成する手法であり、2023年終盤から運用面でも導入が始まったものである。RAGは基盤モデル単独では難しい最新情報の反映が可能となり、特にデー

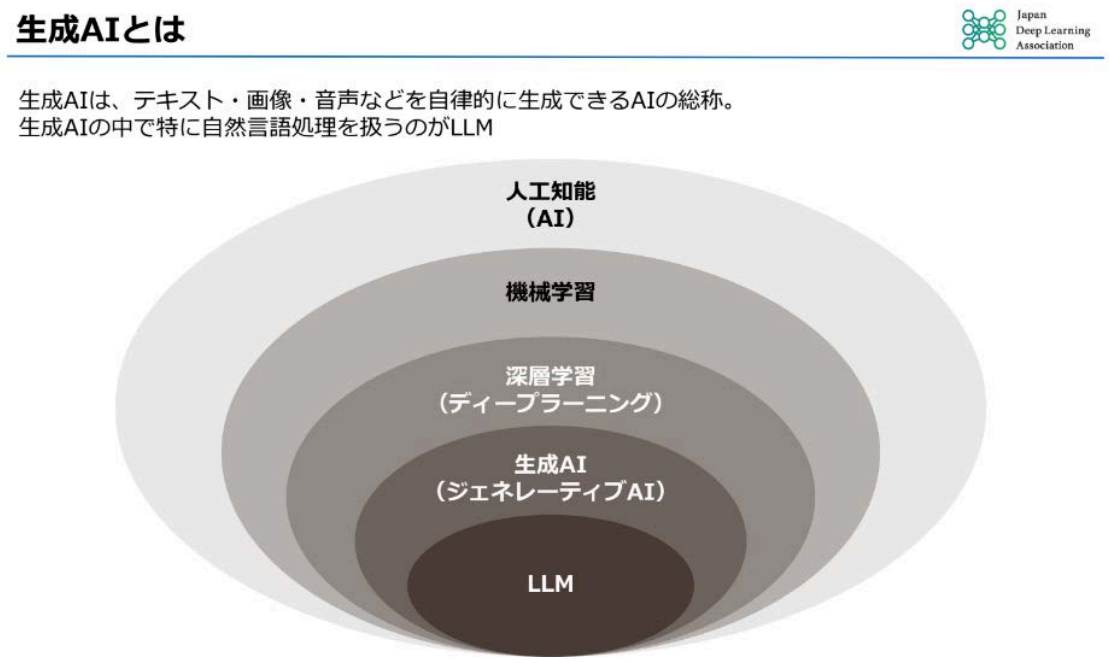
タが古い場合の情報の陳腐化やハルシネーションのリスクを軽減し、生成されるアウトプットが参照した具体的な情報源に基づいているため、事実確認がしやすく、高い信頼性が求められる領域において、より安全に生成AIを活用できるメリットがある。一方で、RAGは膨大なドキュメント全体を俯瞰的に理解・要約することが不得意であったが、2024年7月Microsoft社から発表された、RAGの機能を拡張・強化した「GraphRAG」はLLMを使用して検索・取得したデータセットに基づく「ナレッジ グラフ」を作成し、これをもとに複雑な情報を分析して回答のパフォーマンスを大幅に向上させることができる技術として注目を集めている。

また、マーケット上、LLMは海外製のモデルがほとんどであるがゆえ、日本語でのインプット・アウトプットが英語を活用した際よりもパフォーマンスが落ちてしまう精度観点が課題であったが、Softbank社をはじめ国内事業者が一からLLMを開発する「国産LLM」開発の動きも加速化しているところ。さらに、既存のLLM開発事業者におけるモデルの精巧化も顕著である。例えばOpenAI社は新しいAIモデルシリーズとして「OpenAI o1」を開発し、ユーザーに応答する前に長い内部思考の連鎖を生成することができるモデルとして、複雑な推論を実行するために強化学習で訓練された大規模言語モデルなども流通している。Google社では画像生成等も可能なマルチモーダルな大規模モデル「Gemini」をベースに医療分野に特化した「Med-Gemini」を開発し、医療分野での活用が期待されているなど、医療・ヘルスケア領域に特化した生成AIのモデルの開発が推進されているところである。

業界における生成AIの取扱いに関する自主的な取組としては、日本デジタルヘルス・アライアンスでは本ガイドを2024年1月に策定したところであるが、生成AIの技術動向の発展に伴い、他団体でも自主的なガイドライン作成の動きが見受けられている。例えば医療AIプラットフォーム技術研究組合においては「医療・ヘルスケア分野における生成AI利用ガイドライン」を2024年10月に策定し、医療現場での生成AIの導入と利用を促進するために、医療現場での実際のユースケースに沿った形で、注意すべきポイントをまとめている。また、一般社団法人AIメンタルヘルスケア協会はメンタルヘルス領域に特化した生成AI活用に関するガイドラインを近日中に取りまとめる予定である。

なお、ヘルスケア領域における生成AI活用事例についても2024年からユースケースが増加しているところであり、当該事例については本ガイド作成にあたって検討にご協力いただいた企業の取組を中心に「参考資料 別添2 ヘルスケア領域における生成AI活用事例集」として掲載しているので参照されたい。

図1：生成**AI**とは



（出典元：一般社団法人日本ディープラーニング協会）

2-2. 生成AIの特徴

生成AI（Generative AI）は、前述のとおり自律的に学習したデータから文章、画像、音声などの一見新しく現実的なコンテンツを生成することができる一連のアルゴリズムのことをいう。昨今ではウェアラブルデバイスやその他のデジタル技術の進化により分析可能なデータが日々増加しており、これらを活用して生活者の日々の健康管理等に寄与するサービスの発展も著しい。こうした中、生成AIの活用により、さらなるサービス開発やこれによる健康増進等に大きな期待が寄せられている。

テキスト生成AIは、非常に多岐にわたるタスクを実行することが可能である一方、様々な制約条件や課題が存在するため、生成AIの特徴とそれに伴う論点について以下のとおり整理する。

2-2-1. 基盤モデルに関する特徴

生成AIを活用するに当たっては、巨大なデータセットを活用した基盤モデルを利用することが必要となり、場合によっては基盤モデルに学習データをインプットして更に学習を進めることで特定用途に特化した基盤モデル（以下「特定モデル」という。）を生成・活用することも可能である。そのため、どのような学習データをインプット・再学習したかによって基盤モデルの性質・機能も多種多様である。

2-2-2. データに関する特徴

生成AIは広範なデータを自律的に学習することで入力された条件に対応する結果を出力する性質のものである。そのため、多くの事業者が提供するサービスにおいて共通の基盤モデルを利用して提供されることが想定されるが、例えば学習データやファインチューニングに活用するデータが他の主体に共有される可能性があるなど、個人情報保護や著作権保護をはじめとするデータの取扱いに関する問題が懸念される。特にヘルスケア領域においては、要配慮個人情報を含む個々人の健康状態に関するデータや、著作物を含む医学論文上の情報などが含まれている場合も想定されるため、広範なテキストから学習する生成AIを活用するに当たっては、特にデータの取扱いについては十分な配慮が必要である。

2-2-3. アウトプットに関する特徴

生成AIの特徴として、アウトプットに至る処理過程について人間による解釈が困難である上、AIの能力を統一的かつ網羅的に評価することが困難であることが挙げられる。そのため、共通の基盤モデルを活用している場合に、基盤モデル開発の精度向上施策が影響して逆に一部タスクの精度が下がるということが原理的にあり得る。また、基盤モデルを利用する際の設定値によっては確率的な挙動となることがあり、出力が安定しないことも発生し得る。そのため、アウトプットには、いわゆる「ハルシネーション」（幻覚）の問題が存在し、事実と異なる内容（嘘）や文脈と全く無関係な内容の出力が生成される可能性もある。さらに、基盤モデルの学習データに利用されているデータが古い情報の場合、アウトプット情報も古いものが出力されてしまうケースも想定される。これらを踏まえると、特にヘルスケア領域は健康や身体・生命に影響を与える情報を取り扱う場合があることが想定されるため、これらの特徴を理解し、アウトプットの信頼性を担保する取組を実施した上で生成AIを活用する必要がある

2-2-4. 利用者のリテラシーに関する特徴

生成AIは、これまでのAIとは異なり、AIに関する専門的な知識を持たない一般ユーザーであってもデータを入力することで簡単にアウトプットデータを獲得できる点（低コスト性・利便性）も一つの特性である。また、生成AIは適切な前提条件を設定することで、その能力が最大化され、より適切な回答を提供することが可能であり、最初の回答が不十分であったとしてもユーザーとの対話を通じて回答を更新し、より適切な回答を生成することが可能である。

そのため、AIのパフォーマンスは、設定された前提条件やパラメータに大きく依存し、その結果にバイアスが生じる可能性がある。さらに、ユーザーが適切な指示ができなかった場合は本来なら正しい回答が可能な質問であっても、適切な回答に辿り着けない場合がある。このようにユーザーが入力するプロンプトの内容や品質によりAIの出力結果が大きく異なる場合も発生するため、利用者側もその特性を理解できるリテラシーが必要になってくる。

2-3. 関連制度の概要

2-2に記載のとおり、生成AIについては、これまでAI技術の取扱いに加えて考慮すべき特徴と論点が存在しており、これら論点については国内外で検討・整理が進められているところである。以下に国内及び海外での主な関連制度の状況を整理する。なお、以下の情報は本ガイド執筆時点のものであり、AI技術に関する領域は活発に規制内容等が変化するため、最新の関連制度とその改正状況は変更・進捗している可能性があることに注意されたい。

2-3-1. 国内における関連制度の概要

国内の関連制度	概要
AI事業者ガイドライン（経済産業省・総務省）	「AI事業者ガイドライン」は、「人間中心のAI社会原則」を土台としつつ、我が国における3つのガイドライン（総務省「国際的な議論のためのAI開発ガイドライン案」/経済産業省「「AI利活用ガイドライン～AI利活用のためのプラクティカルリファレンス～」/経済産業省「AI原則実践のためのガバナンス・ガイドライン」）を統合し、諸外国の動向や新技術の台頭を考慮したAIガバナンスの統一的な指針を示す非拘束的なソフトローなガイドラインとして総務省及び経済産業省より2024年4月19日に第1.0版が公表された。 本ガイドラインは①事業者の自主的な取組の支援、②国際的な議論との協調、③読み手に取ってのわかりやすさを基本的な考え方とし、加えてガバナンスの継続的な改善を実施するため、マルチステークホルダー関与の下で、「Living Document」として適宜更新を行うことが予定されている。2024年11月22日には他業界での取組等がアップデートされた第1.01版が策定されている。 なお、本ガイドラインについては、既存の事業者向けガイドライン（総務省「国際的な議論のためのAI開発ガイドライン案」/経済産業省「「AI利活用ガイドライン～AI利活用のためのプラクティカルリファレンス～」/経済産業省「AI原則実践のためのガバナンス・ガイドライン」）の統合・改定を行う位置づけのものである。 総務省/経済産業省 「AI事業者ガイドライン（第1.01版）」（2024年11月22日） https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/20240419_report.html

「生成AIサービスの利用に関する注意喚起」及び「OpenAIに対する注意喚起の概要」（個人情報保護委員会）	2023年6月に個人情報保護委員会は「生成AIサービスの利用に関する注意喚起」及び「OpenAIに対する注意喚起の概要」を発出。主に個人情報を取り扱う事業者及び行政機関等を対象として、生成AIサービス利用に際しての個人情報取り扱いの注意点をまとめている。当該注意喚起においては、一般の利用者に対しても生成AIサービスにおける留意点として個人情報が学習に利用されることがあるなどして、サービス事業者の利用規約等を十分に確認し、入力する際はリスクを踏まえた上で適切に判断をすることとしている。また、当委員会はOpenAI社に対して、本人の同意なしに要配慮個人情報を取得しないこと、個人情報の利用目的について日本語を用いて、利用者及び利用者以外の個人に対して通知または公表することと注意喚起を行っている。 個人情報保護委員会「生成AIサービスの利用に関する注意喚起等について」（2023年6月2日） https://www.ppc.go.jp/news/careful_information/230602_AI_utilize_alert/
個人情報保護法のいわゆる3年ごと見直しに関する検討会 報告書	個人情報保護委員会では、2025年の個人情報保護法改正に向けて、個人情報の保護に関する国際的動向や情報通信技術の進展、それに伴う個人情報を活用した新たな産業の創出及び発展の状況及び現行法の施行状況等についての実態把握や多様なステークホルダーからのヒアリング等を通じて、具体的な検討が進められているところである。 2024年12月25日に公表された「個人情報保護法のいわゆる3年ごと見直しに関する検討会 報告書」において、AIに関する主な論点の記載はなかったものの、令和6年10月16日に当該委員会が公表した「 個人情報保護法のいわゆる3年ごと見直しの検討の充実に向けた視点」及び「今後の検討の進め方」においてはデジタル社会の進展やAIの急速な普及をはじめとした技術革新や技術の社会実装の動向等も考慮し、制度の基本的な在り方に立ち返った議論を行うべきであるとの意見も出されたこと等を踏まえ、制度の基本的な在り方に関わる次元の論点を幅広いステークホルダー等との間で再確認し、短期的及び中期的な検討の基礎とすることが記載されているため、今後の動向は注視する必要がある。 個人情報保護法のいわゆる3年ごと見直しに関する検討会　報告書（2024年12月25日） https://www.ppc.go.jp/files/pdf/241225_shiryou-1.pdf
AIと著作権に関する考え方について	2024年3月15日に文化審議会著作権分科会法制度小委員会は、生成AIを巡る著作権等侵害の懸念の声を受けて、懸念の解消を求めるニーズに応えるため、生成AIと著作権に関する考え方を整理すべく「AIと著作権に関する考え方について」を取りまとめ、公表した。同年7月31日に上記等で示された考え方の解説資料として、AI開発者等が著作権と生成AIとの関係で生じるリスクを低減させる上で、また権利者が自図からの権利を保全・行使する上での取り組みをステークホルダーごとにわかりやすく紹介するチェックリスト及びガイダンスが作成されて

	いる。 参照:文化庁「AIと著作権について」 https://www.bunka.go.jp/seisaku/chosakuken/aiandcopyright.html
広島プロセスG7首脳声明・広島AIプロセスに関するG7首脳声明高度なAIシステムを開発する組織向けの広島プロセス国際指針」・「高度なAIシステムを開発する組織向けの広島プロセス国際行動規範」	2023年10月30日、G7首脳は広島AIプロセスに関してG7首脳声明を発出。首脳声明のほか、「広島AIプロセスに関するG7首脳声明高度なAIシステムを開発する組織向けの広島プロセス国際指針」及び「高度なAIシステムを開発する組織向けの広島プロセス国際行動規範」もあわせて公表した。 首脳声明では、AIの革新的な機会と変革の可能性を強調するとともに、リスク管理と共有原則の遵守の重要性を認識した上で国際指針と行動規範を作成することが記載されている。国際方針は安全・安心・信頼できるAIを世界に普及させることを目的とし、生成AIを含む高度なAIシステムを開発・利用する組織向けに策定され、サイバーセキュリティ・個人データ保護をはじめとした安全対策等の原則等、11項目の原則の遵守が求められている。国際行動規範においては、国際指針に記載の11項目に基づいた具体的な計画内容が記載されている。 外務省「広島AIプロセスに関するG7首脳声明」(2023年10月30日) https://www.mofa.go.jp/mofaj/ecm/ec/page5_000483.html
プログラム医療機器該当性に関するガイドライン等(厚生労働省)	開発したプログラムの標榜は当該プログラムが医療機器に該当するか否かで異なり、機器該当性判断については、表示、説明資料、広告等に基づき当該プログラムが薬機法で定められている医療機器としての目的性(疾病の診断、治療、予防)及び人の生命・健康に影響のリスクがどの程度寄与するかによって判断される。開発されたプログラムが医療機器に該当するかに否かに関しては厚労省が公開している判断事例等やプログラムの医療機器該当性に関するガイドラインが参考となる。 厚生労働省「プログラムの医療機器該当性に関するガイドラインについて(2023年3月31日一部改正)」 https://www.mhlw.go.jp/content/11120000/001082227.pdf また、汎用AIに関しては、厚生労働省は、同省ホームページにおいて「医療機器プログラムと汎用AIの違いについて」を公表している。ここには、「汎用AIなどのその他のプログラムは医療機器として承認・認証されたものではなく、疾病や診断の予防、治療の目的を標榜して提供することはできない」旨や、「健康状態や疾病に関する質問をした場合の回答内容を含めたその性能は、医薬品医療機器等法に基づき、その妥当性が確認されたものではない」旨が記載されている。 厚生労働省「医療機器プログラムと汎用AIの違いについて」(厚生労働省ホームページ) https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000179749_00004.html

2-3-2. 海外における関連制度の概要

ここでは、海外、特に欧州・米国・中国・インド・韓国における生成AIを含むAI全般に関する法規制・制度について特に直近の動向を中心に記述する。

2-3-2-1. 欧州の動向

欧州では、2021年以降にAIを巡る規制の動きが活発となっており、2021年には、欧州委員会がAI Actを発表し、安全保障目的のAIや悪用の可能性があるAIに規制がかかる可能性があることに言及した。さらに、2022年には、欧州理事会（EU加盟国による意思決定機関）がイノベーションを重視した修正案を発表。2023年6月には、生成AIを含む包括的なAIの規制案であるAI Actが、欧州議会の本会議において賛成多数で採択された。2024年5月にはEU加盟国で構成する閣僚理事会による最終承認を経て正式に成立し、2024年8月に正式に発効した。その後の適用は段階的に行われ、発効から6ヶ月後に禁止されるAIシステムに関する規定等が、12ヶ月後に一般的なAIモデルに関する規定等が、36ヶ月後に高リスクAIシステムに関する規則が適用される。早ければ年内の合意、2024年以降の施行が見込まれている。

図2：AI Actの特徴

リスクベースのアプローチ	• AIシステムをリスクレベルに基づいて分類する「リスクベースアプローチ」を採用する ✓ 許容できないリスク：操作的または欺瞞的技術を使用し、人々の意思決定を著しく損なう可能性があるAIシステム等は完全に禁止される ✓ ハイリスク：人間の健康、安全、基本的権利に重大な影響を与える可能性があるシステム（例：医療機器や生体認証システム等）には、厳格な要件が課せられる ✓ 限定リスク：透明性義務が適用されるAIシステム（例：チャットボットやディープフェイク生成AI）には、使用前にユーザーへの通知義務がある ✓ 最小リスク：上記以外のシステム（例： AIを利用したビデオゲーム、スパムフィルター）には特段の規制はない
要求事項と義務	• 高リスクAIには、データガバナンスや技術文書の作成、適合性評価手続きなど厳格な規制が求められる。また、すべてのAIシステムに対して一定の透明性と説明責任が求められる • これにより、市民の権利を保護しつつ、安全で信頼できるAI技術の利用を促進する
イノベーション支援	• 中小企業への負担軽減策として、行政的および財政的支援が提供されることが求められている。中小企業も規制遵守を行いやすくし、イノベーションを促進することを目指している
罰則規定	• 違反した場合には、以下の重い罰金が科せられる • 禁止されたAIアプリケーション：3,500万ユーロまたは全世界売上高の7%の高い方 • AI Actの義務違反：1,500万ユーロまたは全世界売上高の3%の高い方 • 誤った情報の提供：750万ユーロまたは全世界売上高の1%の高い方
域外適用	• AI ActはEU内で提供されるすべてのAIシステムに適用されるだけでなく、EU市場をターゲットとする第三国にも域外適用され、国際的な取引にも影響を与える

出所：EUR-Lex「Artificial Intelligence Act」（2024 年 7 月 12 日）を基に作成

また、欧州の生成AI関連の規制を巡り、サービスを提供する各企業に以下の観点で様々な影響が出ている。

<u>①偽情報対策</u>

AI Actでは、生成AIがもたらすリスクとして偽情報の拡散が特に懸念されている。ディープフェイクや自動生成されたコンテンツは、選挙や公共の意思決定に影響を与える可能性があるため、これらのリスクを軽減するための規制が必要とされている。2024年5月、欧州委員会はマイクロソフトに対し、同社の検索エンジンBingにおける生成AI機能について詳細な情報を提供するよう要求した。この要求は、マイクロソフトが提供する機能「Copilot」や「Image Creator」に関連するもので、これらがEUのデジタルサービス法（DSA）に違反している可能性があるとの懸念から発せられたものである。欧州委員会は、Bingが生成AIツールを使用してハルシネーションやディープフェイクを作成し、それが選挙において有権者を誤解させる可能性があると指摘している。これにより、マイクロソフトには2024年5月までに詳細な回答が求められており、回答しない場合には最大で年間売上高の1%の罰金、および平均日収または年間売上高の5%の定期的な罰金が科せられる可能性がある。この動きは、GoogleやMeta、TikTok等のテクノロジー企業に

も波及しており、同様の説明責任を果たす必要がある。[2]。

②AI Act違反の恐れ

EUの支援を受け、スイスのスタートアップ企業ラティスフローが中心となり、今後段階的に施行されるAI Actの規定に基づき、数十の項目で生成AIを評価するツール「LLMチェッカー」を開発した。ラティスフローが2024年10月に公表したデータによれば、Alibaba、アンスロピック、オープンAI、Meta、ミストラルAI各社が開発した生成AIについて、セキュリティ等の重要な項目のいくつかで基準に達していないと報告し、多額の罰金の対象となりうると指摘されている[3]。各社の評価結果については、COMPL-AIのWebサイト[4]にて公開されている。

③欧州での生成AIの提供延期

Appleは2024年内に欧州で新たな生成AIサービスを提供しない方針を固めたと報じられている。この決定には、EUのデジタル市場法（DMA）が影響しており、特にプライバシーやデータセキュリティに対する懸念が背景にあるとされている。Appleは「Apple Intelligence」という生成AIサービスを発表していたが、規制上の不確実性により、EUの利用者には提供できないとしている。同様に、Metaも生成AIの欧州導入を当面見合わせると発表している。これは、アイルランドのデータ保護委員会が同社のFacebook等に投稿されたコンテンツをLLMの訓練に使用することについて、延期を要求したためである。Metaは自社のAI学習が同業他社と比較して透明性が高いと主張しているが、欧州の厳しい規制が各社のAI戦略に影響を与え始めている状況であると報じられている[5]。

2-3-2-2. 米国の動向

2023年10月30日にバイデン前大統領は、AIの安全性の確保や技術革新を図るための大統領令を発令した。開発事業者はサービス提供や利用開始前に政府による安全性の評価を受けるよう義務付けることや、コンテンツが「AI製」であるか識別できる仕組みを設け、偽情報拡散防止を行う等のAI規制について記述されている。特に医療・ヘルスケア分野においては、AIが関わる危険な医療行為の事例を収集し、安全性の指針を作成する旨も規定している[6]。

図 3：大統領令における医療・ヘルスケア分野に関する言及

---

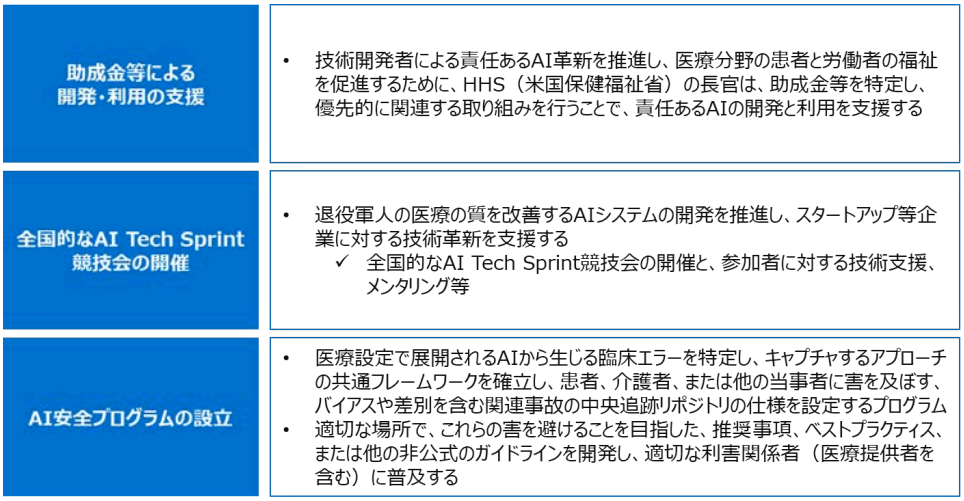
[2] Euronews「Microsoft urged to answer EU questions on Generative AI」（2024年5月17日）
https://www.euronews.com/next/2024/05/17/microsoft-urged-to-answer-eu-questions-on-generative-ai

[3] Reuters「主要生成AIモデル、欧州AI法違反の恐れ　評価テストで低スコア」（2024年10月16日）
https://jp.reuters.com/economy/industry/ZMZUYSVD55NYDDJMCPQGIB6YBA-2024-10-16/

[4] https://compl-ai.org/evaluations

[5] 日本経済新聞「アップル、生成AIの欧州提供を延期」（2024年6月22日）
https://www.nikkei.com/article/DGKKZO81585420S4A620C2MM0000/

[6] 日本経済新聞「米大統領令、生成AIを初規制　公開前に安全評価義務づけ」（2023年10月30日）
https://www.nikkei.com/article/DGXZQOGN301180Q3A031C2000000/



出所：Federal Register :: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligenceを基に作成

政府はイノベーション創出を優先し、AI開発に対し基本的には自由放任主義の姿勢を貫いてきた。大統領令ではAIの事前の安全評価を義務付けたものの、あくまで企業の自主的な取り組みを促すものにとどまっている。また、大統領令は全国的な基準を設定しつつ、各州においてそれぞれ独自のニーズや状況に応じた規制の導入を進めている。先行する動向として、カリフォルニア州とコロラド州の動向について記述する。

①カリフォルニア州の動向

カリフォルニア州は多くの主要AI企業と研究教育機関を抱え、AI産業をリードしている。連邦政府の規制が進まない危機感を背景に、同州では法整備が進み、2024年4月に「最先端AIシステムのための安全で安心な技術革新法」（SB1047）が上院司法委員会を通過した。本法案は、AI企業にサイバーセキュリティ保護を義務付け、重大な被害が発生した際には企業に法的責任を課すものである。さらに、司法長官による訴追も認めている。

これに対し、テクノロジー業界団体やカリフォルニア州商工会議所は、法案が過度に広範で企業に対する負担が大きく、イノベーションを抑制する恐れがあると反対を表明している[7]。その後、SB1047は生成AI企業アンスロピックの提案に基づき修正されたが、多方面から反対表明が続いた。カリフォルニア州選出の8人の連邦議員や元下院議長ナンシー・ペロシ下院議員（民主党、カリフォルニア州）、オープンAIも反対意見を述べ、AI産業が州外に転出する可能性を警告した。その後、アンスロピックは州知事宛てに法案支持を表明したが、監査事項の曖昧な文言や司法長官の広範な権限による過度な事前介入の恐れを指摘した。

最終的にニューサム知事はSB1047に拒否権を行使し、小規模モデルも壊滅的なリスクがあるため、規制には慎重なバランスが必要だと結論付けた。しかし、ニューサム知事はAIの効果的な規制枠組みを考える意向を示し、州議会や連邦政府関係者と協力して取り組む姿勢を強調した。また、ニューサム知事は安全で責任あるAIの推進とカリフォルニア州民の保護に向けた新たな取り組みを発表し、生成AI展開でのガードレール規制開発を進めるとした[8]。カリフォルニア州は2024年11月米大統領選の民主党候補であったハリス副大統領の地元でもあり、知事の判断に民主党陣営の意向が影響したとの報道もある[9]。

[7] ジェトロ「米カリフォルニア州で新たなAI規制法が州上院司法委員会を通過」（2024年4月12日）
https://www.jetro.go.jp/biznews/2024/04/65d431a83da79f72.html
[8] ジェトロ「米カリフォルニア州知事、AI安全法案を拒否し、新たな取り組み発表」（2024年10月7日）
https://www.jetro.go.jp/biznews/2024/10/73761c6c009dcecf.html
[9] 日本経済新聞「米国、AI規制に慎重論根強く　加州法案を知事が拒否」（2024年9月30日）

②コロラド州の動向

ジャレッド・ポリス州知事は2024年5月に、米国で初となるAIシステムの開発と利用に関する包括的な州レベルのAI規制法である「AIシステムとの相互作用における消費者保護に関する法律」（SB 24-205）」に署名した。SB 24-205は、高リスクAIシステムの開発者と利用者に新たな義務を課すことを目的としており、教育機会や雇用機会、金融サービス、行政サービス、医療サービスなどの分野において、アルゴリズムに基づく決定によって消費者が差別的な取り扱いを受けることを避けるために合理的な注意を払うことを要求している。実際の施行は2026年2月であり、施行までの期間は、事業者がコンプライアンスの準備を整え、州政府が法律の実施に必要な規則やガイダンスを策定するために設けられている。

図 4：SB 24-205の対象と義務

義務	開発者	配備者
システムの種類・管理方法に関するステートメントを作成すること	✓	
配備者が影響評価を行うために必要な情報を入手可能とすること	✓	
システムのリスク管理方針・プログラムの実施や、影響評価を完了させること		✓
システムが消費者に関する決定を行った場合、特定事項を消費者に通知すること		✓
システムに起因する消費者に関する不利な決定について、技術的に可能であれば、人的な審査を通じて不服を申し立てる機会を消費者に提供すること		✓
システムが差別を引き起こしたことが判明したなどの場合には、90日以内に司法長官に当該リスクを開示すること	✓	✓

出所：Colorado General Assembly「Consumer Protections for Artificial Intelligence（SB24-205）、ジェトロ「米コロラド州でAI規制法可決、知事署名で成立すれば全米初の民間部門への規制に」（2024年5月17日）https://www.jetro.go.jp/biznews/2024/05/f8314d21a2862c13.htmlを基に作成

2-2-2-3. 中国の動向

中国は、国家戦略としてAI産業の発展を促進している。既存の法令とアルゴリズム規制、AI倫理規制と標準制定などを組み合わせる形でAI規制を行っており、中国独自のAIガバナンスモデルが形成されつつある。中国のAIに関する国家戦略全体は、図4に示すように5つから構成されている。

図5：中国のAIに関する国家戦略全体像

https://www.nikkei.com/article/DGXZQOGN300560Q4A930C2000000/

法制度	全人代	• サイバーセキュリティ法（2017年） • データセキュリティ法（2021年） • 個人情報保護法（2021年）
	国家インターネット情報弁公室	• インターネット情報コンテンツ生体治理規定（2019年） • インターネットコメントサービス管理規定（2022年）
AIアルゴリズム規制	国家インターネット情報弁公室	• インターネット情報サービスアルゴリズム・レコメンデーション管理規定（2022年） • インターネット情報サービス深度合成アルゴリズム管理規定（2022年） • 生成人工知能サービス管理暫行弁法（2023年）
AI倫理的規制	国務院	• 新一代人工知能ガバナンス準則－責任ある人工知能の発展（2019年） • 科技倫理ガバナンス強化に関する意見（2022年）
	科学技術部（人工知能ガバナンス専門委員会）	• 新一代人工知能倫理規範（2021年） • 科技倫理審査弁法（試行）（パブコメ）（2023年）
AI関連標準制定ガイド	国家標準化管理委員会等	• 国家新一代人工知能標準体型建設ガイド（2020年）
	全国信標委人工知能サブ技術委員会	• 人工知能倫理治理標準化ガイド（2023年）
政策	国務院	• 新一代人工知能発展規画（2017年）
	工業情報化部	• 新一代人工知能産業発展促進三年行動計画（2018-2020）（2017年）

出所：中華人民共和国国家網絡情報弁公室（CAC）Webサイトを基に作成

また、政府は2023年6月に人工知能法の立法準備に着手したことを発表している。この法案はAI技術全般に関する包括的な法律であり、研究開発やその利用に関する法律的な枠組みを提供することを目的としている。2024年11月時点では大きな動向は見られておらず、専門家もすぐに人工知能法が成立・発行に至る可能性は低いと指摘している[10]。生成AIについては、健全な発展と規範的な利用を促進し、国家の安全や公共の利益を保護することを目的とした「生成人工知能サービス管理暫行弁法」が2023年8月に施行されている。本法では、生成AIの基盤となるLLMを企業が一般公開する際に当局承認を義務付けた。この承認制度は、政府がAI技術の発展を促進しつつも、その管理と監視を強化するための一環として位置付けられており、具体的には、企業は自社のAIモデルが法律や規制に従っていることを証明する必要がある。これにより社会的なリスクや不正確な情報の拡散を防ぐことが目的とされている。また、この制度は中国国内でのAI開発の透明性を高めるとともに、国家安全保障や公共の利益を守るための重要な手段とされている。政府は承認した企業名を公表していないが、2023年8月に百度（Baidu）、アリババ（Alibaba）、字節跳動（ByteDance）等のLLMが承認され、香港の証券時報によると2024年1月時点で40以上のLLMが承認されていると報じられている[11]。

2-3-2-4. インドの動向

インド政府は、2023年4月にAIを規制しない方針を発表した。この時点では、AIがデジタル経済の推進力であり、AIの急速な発展を促進するために規制を設けないという考えであった。

一方で、2024年3月に、電子情報技術省（MeitY）はAI製品の発表前に政府からの許可を得るよう指示する勧告を出した[12]。この勧告は、特に実験的または試験段階のAIプラットフォームが不法なコンテン

[10] MIT Technology Review「Four things to know about China's new AI rules in 2024」（2024年1月17日）
https://www.technologyreview.com/2024/01/17/1086704/china-ai-regulation-changes-2024/
[11] Reuters「中国のAIモデル認可制、導入後6カ月で40以上承認＝証券時報」（2024年1月29日）
https://jp.reuters.com/business/technology/JI5DLF7DIRMXPH4G7W74LWKOGI-2024-01-29/
[12] Economic Times「MeitY approval must for companies to roll out AI, generative AI models」（2024年5月2日）

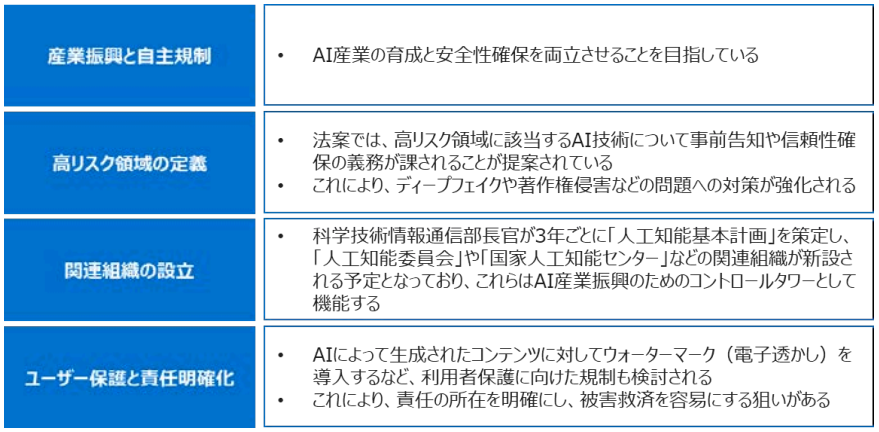
ツを流布することを防ぐためのものであり、スタートアップには適用されないことも明確にされている。この動きは、インドのインターネットの安全性と信頼性を確保するという共通の目標に向けた政府、ユーザー、プラットフォーム間の共有された取り組みの一環であるが、企業の創業者を含む複数の人物がX上で反論する事態となっていると報じられている[13]。

この勧告は、法的拘束力はなく、不法コンテンツの抑制や透明性向上には一定寄与するものと考えられるが、実際にはディープフェイクなどの問題に直面している。特に2024年6月の開票に向けて各地で投票が順次行われたインド総選挙（下院選）の選挙期間中においては、迅速な情報拡散とその影響力から、規制が不十分であり事実上野放しになっていると報じられている[14]。

2-3-2-5. 韓国の動向

韓国において、2024年11月時点でAIに関する具体的な法律は存在していないが、2022年以降様々なAI関連法案が国会に提出されている。その中で、「AI産業の振興および信頼されるAIの確立の枠組みに関する法律」（通称：AI基本法）は、EUのAI Actとは異なり、「技術を先行導入し、その後に規制を設ける」という原則に基づき、AI技術の開発と産業の活性化を目指している。また、利用者への事前通知の義務化、信頼性の確保、安全性の維持等、高リスクのAI分野に対する個別の規制も提案されている。しかし、生成AIやディープフェイクに関する規制の追加提案があるほか、市民団体からの規制に対する反対意見も出ており、最終的な法制化にはまだ時間がかかる見込みであると報じられている[15]。

図 5：AI基本法の概要



出所：ET News「[스페셜리포트] AI·망 무임승차방지·디지털 포용…산업 육성·민생 법안 산적」（2024年2月18日）https://www.etnews.com/20240218000028を基に作成

---

https://economictimes.indiatimes.com/tech/technology/govt-directs-social-media-generative-ai-platforms-to-comply-with-it-rules/articleshow/108162287.cms

[13] YourStory「AI launch advisory to exclude startups, clarifies Rajeev Chandrasekhar」（2024年5月4日）
https://yourstory.com/2024/03/ai-launch-advisory-to-exclude-startups-clarifies-rajeev-chandrasekhar

[14] 読売新聞「インド総選挙、生成AI作成の偽動画が拡散…規制法なく事実上野放し」（2024年5月5日）
https://www.yomiuri.co.jp/world/20240505-OYT1T50006/

[15] Law.asia「Analysis of AI regulatory frameworks in South Korea」（2024年4月15日）
https://law.asia/ai-regulatory-frameworks-south-korea/

3. 生成AIの活用・提供にあたってのバリューチェーンと論点

3-1. 生成AIの活用・提供に当たっての関係主体

生成AIはその活用・提供に当たって様々な主体が関係しているほか、各主体におけるインプット・アウトプット活動によってそのバリューチェーン（価値創造過程）が構成されている。本ガイドにおいては、関係主体を図6のとおり①基盤モデル開発者、②特定モデル開発者、③サービス・プロダクト提供者、④利用者の4主体に分類する。

図6：生成AIの活用・提供にあたっての関係主体

主体	概要	例
①基盤モデル開発者	大規模言語モデル（LLM）等の大規模で汎用的なモデルを開発・提供する事業者	OpenAI、Google、Meta、Amazon、Softbank、Cyber Agent、Preferred Networksなど
②特定モデル開発者	①が提供するモデルを活用して、自社データや業界固有のデータ等を用いてモデルをファインチューニング[16]し、特定用途に特化したモデルを開発する事業者	①と③が混在している状態
③サービス・プロダクト提供者	①又は②で開発されたモデルを用いて、生成AIを活用したサービス・プロダクトを開発し、直接利用者に提供する事業者	Ubie、MICIN HOKUTO、HACARUSなど
④利用者	生成AIを用いたサービス・プロダクトを利用する個人や法人	ー

3-2. 生成AIの活用・提供に当たってのバリューチェーン

「3-1」において記述した各主体においては、図6のとおり各々のフェーズにおいてインプット・アウトプット活動を行うことにより、モデル開発から利用者へのサービス提供までのプロセスを構成している。

なお、「2-3-1. 国内における関連制度の概要」記載の経済産業省・総務省策定の「AI事業者ガイドライン」においては、生成AIの活用・提供に当たっての主体を「AI開発者」・「AI提供者」・「AI利用者」の3つ分類しているところであるが、本ガイドは生成AIを活用・提供する上での事業者の実務としてチェックするべきポイントを取り扱っていることから、AI事業者ガイドラインで記載されている「AI提供者」を「特定モデル開発者」及び「サービス・プロダクト提供者」にさらに細分化して示しているところである。

①基盤モデル開発者における活動
巨大なデータセットを基盤モデルにインプットし、事前訓練済みモデル（大規模言語モデル等）を開発する。

②特定モデル開発者における活動
追加の学習データを①で開発された事前訓練済みモデル（大規模言語モデル等）にインプットすることで

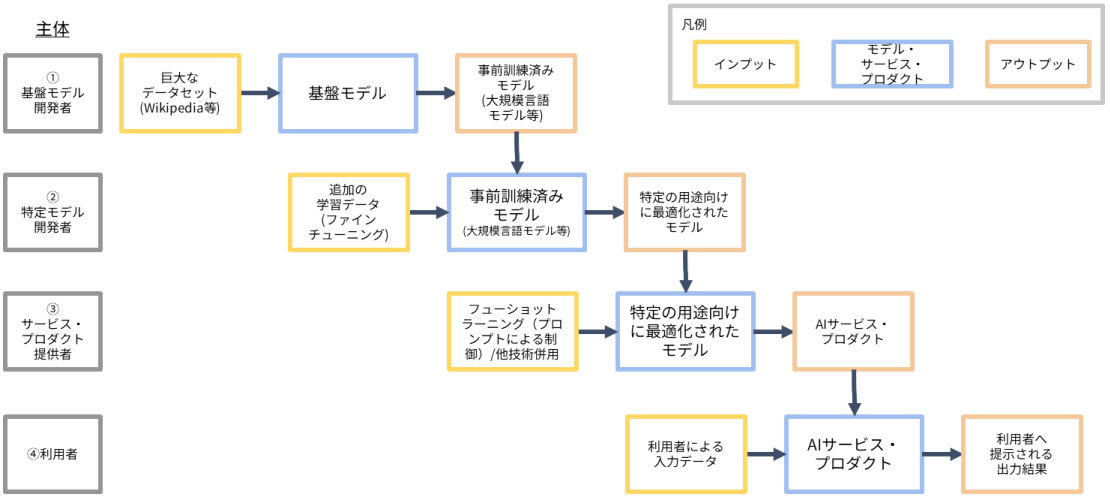
[16] 学習済みの基盤モデルに対して別のデータセットを活用して追加学習させること。用語集参照。

ファインチューニングを行い、特定の用途向けに最適化されたモデルを開発する。

③サービス・プロダクト提供者における活動
②で開発された特定の用途向けに最適化されたモデルに対してフューショットラーニング（プロンプトによる制御）や他技術を併用することで、AIサービス・プロダクトを開発・提供する。

④利用者における活動
③で提供されたAIサービス・プロダクトに質問等のデータを入力し、出力結果を得る。

図7：生成AIの活用・提供にあたってのバリューチェーン



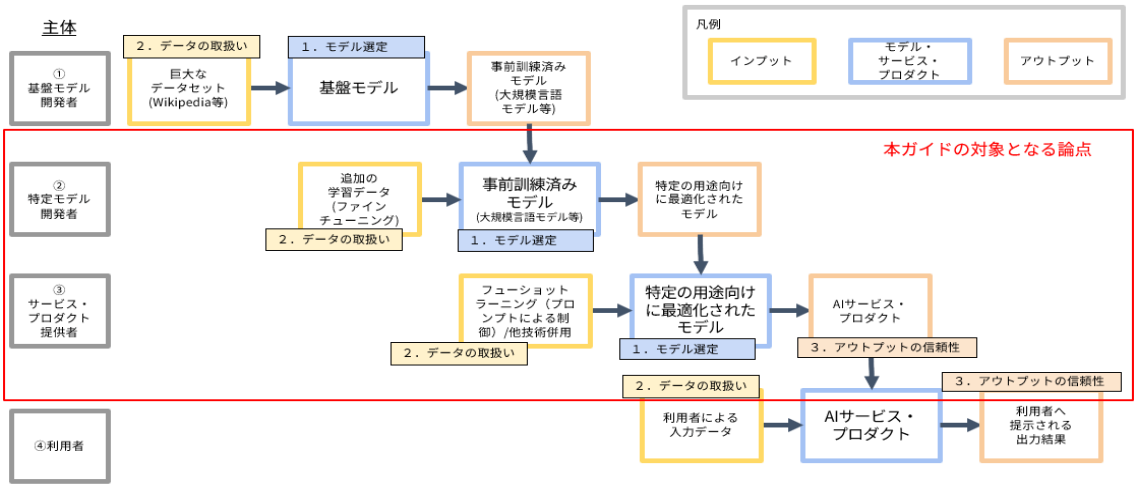
(出典元：株式会社MICIN)

3-3. バリューチェーンを踏まえた生成AIの活用・提供に当たっての論点

「1-2-3.　本ガイドの目的及び対象読者」でも記述したとおり、本ガイドは、生成AIを活用したヘルスケアサービスを提供しようとする事業者を対象読者として策定しているものであることから、バリューチェーンにおける、特に ③サービス・プロダクト提供者 における論点及び考慮事項が主な焦点となる。一方で、図7のとおり、③サービス・プロダクト提供者が、②特定モデル開発者を担う場合も想定されることから、以下では、②特定モデル開発者 及び ③サービス・プロダクト提供者の観点で生成AIの活用・提供に当たって論点になる事項について整理を行う。これらを踏まえると論点は以下図8の赤枠のとおりとなる。

図8：バリューチェーンを踏まえた本ガイドでの対象論点

生成AIの活用・提供に当たってのバリューチェーン



(出典元：株式会社MICIN)

上図及び「2-2　生成AIの特徴と論点」で記述した特徴と論点を踏まえて、本ガイドにおいて特に論点となる事項を整理すると、下記表のとおりとなる。

＜バリューチェーンを踏まえた生成AIの活用・提供に当たっての論点＞

論点	論点の詳細
1．モデル選定	①基盤モデルの選定基準【特定モデル開発者】 基盤モデル開発者によって開発されたモデルは市場に多数存在し、かつ、各々のモデルにおける学習データも異なっている。そのため、クラウドベースで汎用的な自然言語処理が可能なモデルもあればローカル・オンプレミスでヘルスケアに関する専門的な自然言語処理が可能なモデルもあるといったように、各々の基盤モデルによって特徴や用途・性能が異なっている。また、基盤モデルによっては商用利用が制限されている等、用途に一部制限が課されている場合もある。そのため、特定モデルを開発するに当たっては、基盤モデルの特徴や用途・利用可能範囲を事前に確認・判断しておく必要がある。 ②特定モデルの利用上の注意【サービス・プロダクト提供者】 特定モデルについては、利用に当たっての責任範囲があらかじめ明示されている場合が多く、サービス・プロダクト提供者が当該特定モデルを活用してサービス・プロダクトを提供する場合にサービス内容や利用者による利用方法の範囲が、その責任範囲内かどうか確認する必要がある。
2．データの取扱い	①ファインチューニングに利用するデータの取扱い【特定モデル開発者】 基盤モデルに対してファインチューニングを行う際、学習用データセットを活用するが、ファインチューニングに用いる学習用データは基盤モデルそれ自体に影響を及ぼさない場合が多い一方で、基盤モデルに

	よっては学習データが反映されるものも存在する。また、時間経過によって学習用データセットの内容が陳腐化してしまうことやデータそのものに誤りがあった場合に、期待される出力とは全く異なる出力結果を誘発してしまう可能性も存在している。特に、ヘルスケア領域においては、ファインチューニングを行う際、著作権を含む論文情報や要配慮個人情報を含む健康データのほか、日々の技術や研究の進捗に伴い最新情報がアップデートされる医学分野の専門情報などが学習データとして活用されることが想定されるため、これらに対応する取組を講じる必要がある。 ②フューショットラーニング等に利用するサンプル・事例の取扱い【サービス・プロダクト提供者】 サービス・プロダクト提供者が特定の用途向けに最適化された特定モデルに対してフューショットラーニング（プロンプト制御）や強化学習を行う場面において使用するサンプル・事例の取扱いについても注意が必要である。具体的には、①において記述した点と同様、フューショットラーニング等に活用されたサンプル・事例が基盤モデルに反映されないように配慮することや個人情報が含まれていた場合の対応が求められる。 ③利用者が入力する質問データの取扱い【サービス・プロダクト提供者】 生成AIを活用したサービス・プロダクトは利用者によって自由に文章等を記入することで回答が出力される設計が想定されるが、利用者によって個人情報等を含むデータがインプットされる場合も考えられ、当該データが特定モデルに追加学習される可能性も考えられる。特にヘルスケア領域においては、ユーザーが自己の健康・症状等に関する機微な情報を入力することが想定されるため、特定モデル事業者又はサービス・プロダクト提供者において当該データの取扱いに十分注意する必要がある。
3. アウトプット信頼性	①サービス・プロダクト開発段階での取組【サービス・プロダクト提供者】 生成AIはその性質上出力が確率的に変動するため、精度が100%ではなく、ハルシネーション等が発生することが想定される。そのため、それらを所与のものとした上で利用者に対して大きな影響が出ないよう、技術面や仕組み面で制御の工夫を行うなどのサービス設計を開発段階から取り組む必要がある。特にヘルスケア領域においては患者の過去の健康記録、現在の病状、遺伝的リスク、ライフスタイルなど、多角的な情報を統合・アウトプットを行うことや、他産業と比較してプロダクト・サービスに求められる品質は一般に高くなることを踏まえ、サービス・プロダクト提供者において開発段階で可能な限りの取組が講じられることが求められる。 これらの課題に対しては2024年11月現在ではいわゆるRAG（検索拡張生成）とよばれる技術の活用が一般化しており、Graph　RAGと呼ばれるような新規技術も登場している。 ②利用者のリテラシー向上の取組【サービス・プロダクト提供者】 テキスト生成AIは新規性が高い技術であり、利用者側の理解度・リテラシーが追いついておらず、そのためサービス・プロダクト提供者が想

	定していない使い方をユーザーが取りうる場合が想定される。これらを踏まえ、「生成AIは確率的な処理を伴う技術であるため100%の正確性を保証するものではない」という点を利用者に予め誤解なく伝え、生成AIの性質・特徴を利用者側がしっかりと理解した上でサービス・プロダクトを使用できる環境を整えることが重要である。
4. ヘルスケア領域における個別規制	ヘルスケア領域において生成AIを活用したプロダクトやサービスを開発・提供する場合には、生成AIを活用することによって追加された機能やサービスそれ自体が医療機器プログラムに該当するなど、既に個別制度による規制が適用されることが考えられる。そのため、ヘルスケア領域における個別規制の内容を予め確認するとともに、生成AIを活用したプロダクト・サービスの機能や性能がそれらの規制を遵守するように開発・提供の段階で十分に注意を払うことが重要である。

4. 生成AIを活用したヘルスケアサービスの提供を行う場合のチェックポイント

3.に示したとおり、生成AIを活用してサービス提供する際の課題は大きく「1. モデル選定」、「2. データの取扱い」、「3. アウトプットの信頼性」、「4. ヘルスケア領域における個別規制」の4つに大別できる。これらの論点を踏まえ、実際にサービス設計・開発・提供を行う際に、サービス提供者がチェックすべきポイントを以下のとおり整理する。

なお、別添においてチェックリストを整理しているので、サービス・プロダクト提供者におけるセルフチェックに活用されたい。

4-1. モデル選定に関するチェックポイント

ヘルスケア領域においてサービス・プロダクトを提供するに当たっては、どのような基盤モデル又は特定モデルを選定するかによってサービス・プロダクトの設計や運用も大きく変わることが考えられる。これまでの論点を踏まえ、モデル選定時のチェックポイントを整理すると以下のとおりである。

なお、基盤モデルが以下に示す情報についての提供義務が規制等で課せられている状況ではなく、また、医療ヘルスケアにおける生成AIのベンチマーク項目は論文ベースでは多数提案されているものの業界として統一されていない状況であるため、最終的にはサービス・プロダクト提供者側の責任のもと、検討することが肝要である。

【モデル選定に関するチェックポイント】
- 基盤モデルが標榜している性能
- 基盤モデルが定めている利用・用途

4-1-1. 基盤モデルが標榜している性能

基盤モデルが標榜している性能の確認に当たってのチェックポイントは以下のとおりである。

① 基盤モデルの学習データ内容

提供予定のプロダクト・サービスに合った基盤モデルかどうかを判断するため、また、基盤モデルによっては基盤モデル及び学習データのライセンスの扱いが不明瞭である場合（利用不可なライセンスのデータを使用している場合等）があるため、利用予定の基盤モデルは、どのようなデータセットをインプットして学習・開発されたモデルかを事前に確認することが望ましい[17]。ま

[17] 基盤モデル事業者によっては、学習データの内容やアップデート情報について公表・提供している場合があるため、それらを閲覧する方法が想定される。

た、学習データ内容の確認を行うに当たっては、差別や偏見が再生産されることを避けるため、例えば学習データにジェンダーバイアス等のいわゆる「バイアス」が含まれていないかを確認することも重要なポイントである。この点、昨今では学習データセットにおけるバイアスを評価する論文情報等もあるため、その情報を参考にする等して対応を検討されたい。
2024年11月現在では日本語をベースにした日本語特化型LLMが公開されるなど日本の環境に配慮した基盤モデルも多数存在するためこれらの活用も検討されたい。

② 利用可能形式・価格帯

基盤モデルによってはその公開範囲が各々で異なる。例えば、基盤モデルのモデル全体が公開されており、それをオンプレ上で動かすことができる場合もあれば、クラウド上でAPIだけが公開されており、API経由で文章生成はできるがファインチューニングができない場合などさまざまである。そのため、当該基盤モデルの公開範囲によってできること・できないことを予め確認することが重要である。

また、基盤モデルの性能によって価格帯も様々であり、サービス・プロダクト提供に当たっては特に実務面で開発予算等にクリティカルに影響がでるため、注意が必要である。特にAPI利用の場合、生成文字数による使用料変動の有無、、ファインチューニングを実施する場合の金額感、基盤モデルをサーバーで活用する場合の金額感など、各々の場面でどの規模で料金が発生するのかも確認しておくと総合的な判断が決定されやすいと考えられる。

③ 性能評価報告レポートなどにより開示されている用途・性能

利用予定の基盤モデルの性能・機能等を客観的に判断するため、当該基盤モデルにおける性能評価報告レポートではどのような評価がされているかを確認することが望ましい[18]。ヘルスケア業界における生成AIの利活用については、論文ベースにはなるが既に多数の報告がなされているためそれらの既存の報告を確認することもひとつの手段である。2024年11月現在では、例えばHugging　Faceという基盤モデルが公開されているサイトには日本語に対応した言語モデルだけで4,698個以上のモデルが存在しているため、用途に適したモデルを選択することが非常に重要である[19]。ただしこれらの報告の多くは英語圏におけるものであるため、日本語を活用したサービス・プロダクトを検討する際には日本の環境を考慮した追加検討を実施することが望ましい。

④ 基盤モデルの類型

基盤モデルがどの類型に該当するかを理解・確認することで、提供予定のプロダクト・サービスに合った基盤モデルなのかを判断する材料とすることができる。そのため、診断・治療・予防目的ではないことを前提とした上でヘルスケアに関わる専門的な情報をサービスを通じて提供することになる場合は、提供するサービスの性能及び特徴に応じたモデルを選定することが望ましい。なお、基盤モデルの類型例は例として以下のとおり分類されるので、選定時の参考にされたい。

なお、基盤モデルを動かす主体となるサーバーが国内に設置されているか、国外に設置されているかによって準拠法が異なってくるケースも想定されるため、あわせて確認しておくことが望まれる。

＜基盤モデルの類型例＞

- クラウドベースで汎用的な自然言語処理が可能なモデル

[18] 各基盤モデル事業者がサービス提供に当たって提供している情報や、アカデミアが整理・公表している情報などが存在している。

[19] https://huggingface.co/models?language=ja&sort=trending

- ○ クラウドベースで医療・ヘルスケアに関する専門的な自然言語処理に特化したモデル
- ○ ローカル/オンプレミスで汎用的な自然言語処理が可能なモデル
- ○ ローカル/オンプレミスで医療・ヘルスケアに関する専門的な自然言語処理に特化したモデル

＜コラム＞ 日本語特化型**LLM**の動向
「2-1. 生成AIのヘルスケア領域における動向」に記載のとおり、国内事業者がLLMを一から開発する動きが始まっている。例えば、2024年11月にはSoftbank社が4600億パラメータの日本語LLMを公開するなど、日本でも大規模なLLMの開発が進んでいる状況である。日本語LLM（日本語を中心に学習されたLLM）と従来の海外LLM（英語を中心に学習されたLLM）の差分や特徴は以下のとおり。

1. 国由来の制度や文化によるアウトプットの差分
日本語特化LLM及び海外製のLLMでは、国由来の制度や文化によるアウトプットの差異が生まれることが特徴のひとつに挙げられる。例えばヘルスケア領域においては、日本では国民皆保険制度がある一方、アメリカでは当該制度は存在しないといったような歴史的経緯や制度構造がそもそも違う場合がある。そのため、日本ドメインおいては常識的な内容であっても他の国ではそうではない場面が大いに想定され、それを理由としてアウトプットが事実と異なる間違いになってしまう可能性が大いにある。そのため、日本特有の制度等がある領域においては日本語特化LLMの活用がハルシネーションを低減させる一つの手法であるともいえる。

2. データセットやパラメーター量に伴うアウトプットの差分
LLMにおいて重要な大量のデータセットについては、その言語を使用する人数比の違い等から、日本語のデータセットは英語と比べて質量ともに少ないことから、そもそも多種多様なデータセットをもとに機械学習することを前提としたLLMとしてはその精度に差分が生じる可能性が高い。また、LLMのパラメーターの大小によってもその回答精度が左右されるため、データセットの種類やパラメーター量も活用する際の一つの判断材料になると言える。

これを踏まえ、ヘルスケア領域においてはその日本特有の制度等を踏まえたプロダクト設計になる場合が想定されるため、日本語LLMの活用可能性が期待されるところである。

＜コラム＞ 小規模言語モデル（**SLM**；**Small Language Model**）の動向
生成AIが多様な分野で活躍しはじめる一方で、各モデルの規模の大きさや学習に用いられたデータの透明性に伴う課題が医療分野で特に顕著になっている。医療は高い精度と安全性、透明性が求められる分野であり、ユースケースによっては大規模言語モデル（LLM）の汎用性だけでは十分に対応できないケースも多い。このような背景の中で注目を集めているのが小規模言語モデル（SLM；Small Language Model）である。
SLMはLLMと同様に文章生成や文章理解を行うAIモデルであるが、使用するパラメータ数や学習データ量が少なく効率的に運用可能である。特に医療分野においてはSLMの次のような特性が注目されている。

1. コスト効率
医療分野に限らず実際に現場でAI活用を推進する際には限定された予算内でプロジェクトを実施する必要があるが、SLMはLLMに比べて開発・運用コストが低く医療機関にも導入しやすい。

2. データ管理の容易さ
医療データは機密性が高く管理が厳重である。SLMは必要なデータ量が少ないためデータ管理の負担を軽減しつつ、特定の用途に限ればLLMよりも高い性能を発揮する場合がある。

3. 用途特化の適性:
SLMは特定の医療分野、例えば患者記録の要約などに特化したカスタマイズが容易である。この特性により実際のユースケースに最適化したモデルを構築することが比較的容易になる。

現在、様々なSLMが医療分野で活用されつつあり、それぞれが診断補助や医療従事者のサポート

において効率性と有用性を証明するための検証が進められている。さらにSLMはLLMよりも軽量であるため、クラウドサービスを使わずに院内ネットワークに閉じた環境でも利用可能でありセキュリティ面でも優れている場合がある。
今後、特化型AIの重要性がさらに高まる中で、SLMは医療現場における重要なツールとして定着していくことが期待される。

4-1-2. 基盤モデルが定めている利用用途

基盤モデルが定めている利用用途についての確認に当たってのチェックポイントは以下のとおりである。

① 基盤モデルの利用用途範囲

基盤モデルによっては、その利用規約において医療情報の提供を制限等している場合や、明示的に商用利用が限定されているものもあるため、当該基盤モデルの利用規約の内容を事前に確認することが望ましい[20]。

② 入出力データの学習利用に関する規約

基盤モデルの仕様として、インプットデータ・アウトプットデータが追加学習の対象になっている場合もあるため、利用者が入出力するデータについての学習利用に関する規約の内容をあらかじめ確認することが望ましい。

4-2. データの取扱いに関するチェックポイント

サービス・プロダクト提供者がヘルスケア領域において生成AIを活用したサービスを提供する際、医学分野の専門的な情報や健康・症状等に関する機微な情報を扱うことが想定されるため、データの取扱いには特に注意を払う必要がある。以下では前項にて選定した基盤モデルを利用した個別サービスの開発に関わる領域でのチェックポイントを整理する。なお、本項において取り扱うデータの種類については図9のとおりであり、各々のデータの取扱いのチェックポイント概要を図表化すると図10のとおりとなるので、あわせて参考にされたい。

図9：データの種類

[20] なお、アウトプットが他人の著作物と類似している等、著作権侵害を生じてしまった場合、基盤モデル提供会社との契約等によって補償を受けられる可能性もあるため、あわせて確認しておくことも望ましい。

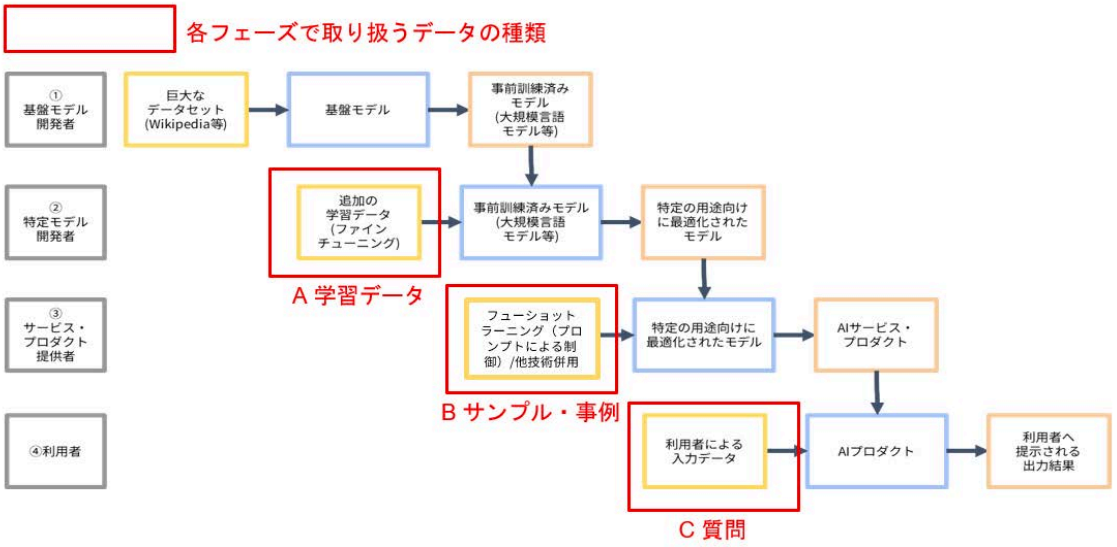


図10：データの取扱いに関するチェックポイント概要

データの種類	誰がどのような時に使うデータか	チェックポイントの概要					誰が守るべきチェックポイントか
		基盤モデルへの反映がされないかの確認	設定オプションの活用	個人情報が含まれている場合の対応	その他著作物等が含まれている場合の対応	その他考慮事項	
A 学習データ	**②特定モデル開発者**がファインチューニングの際に取り扱うデータ	学習データが**基盤モデル**に反映されないよう基盤モデルの利用規約を確認すること	**基盤モデル**に設定オプションがある場合は学習データが**基盤モデル**に反映されないよう設定する	学習データに個人情報が含まれている場合は原則同意を取得する	学習データに論文情報等が含まれる場合は学習に用いることが禁止されていないか確認する	社内におけるデータ保護の体制整備等	特定モデル開発者
B サンプル・事例	**③サービス・プロダクト提供者**がフューショットラーニング（プロンプト制御）の際に取り扱うデータ	サンプル・事例が**基盤モデルor特定モデル**に反映されないよう、**基盤モデルor特定モデルの利用規約**を確認すること	**基盤モデル or 特定モデル**に設定オプションがある場合は、サンプル・事例が**基盤モデルor特定モデルに**反映されないよう設定する	サンプル・事例に個人情報が含まれる場合は原則同意を取得する	サンプル・事例に論文情報等が含まれる場合はフューショットラーニングに用いることが禁止されていないか確認する		サービス・プロダクト提供者
C 質問	**④利用者**がサービス・プロダクトを利用する際に入力するデータ	質問データが**基盤モデルor特定モデル**に反映されないよう、**基盤モデルor特定モデルの利用規約**を確認すること	サービス・プロダクトの設計として利用者による質問内容が**特定モデルに反映されないよう**設定オプションを追加する（利用者でON/OFFができる）	利用者が入力した質問内容に個人情報が含まれる場合には原則同意が必要。出力以外の学習にも利用する場合は学習利用目的の特定が必要。	生成AIへの利用が禁止された著作権物等の情報を入力した場合の免責事項の作成		サービス・プロダクト提供者

【データの取扱いに関するチェックポイント】
- ファインチューニングに利用する学習データの取扱い
- フィーショットラーニングに利用する事例・サンプルの取扱い
- 利用者が入力する質問データの取扱い
- データに関するその他の考慮事項

4-2-1. ファインチューニングに利用する学習データの取扱い

ファインチューニングに利用するデータの取扱いに当たってのチェックポイントは以下のとおりであ

る。

① 基盤モデルの利用規約の確認

基盤モデルによってはファインチューニングに利用する学習データセットの内容を基盤モデルそれ自体に反映する場合もあり、特定モデル開発者が意図せずファインチューニングの目的以外に学習データが活用される可能性が考えられる。そのため、基盤モデルの利用規約等でファインチューニングに使用された学習データの取扱いに関する事項が明瞭化されているか等、ファインチューニング以外で学習データが利用されることがないかを確認することが望ましい。

② 基盤モデルの設定オプションの確認

上記①で記載のとおり、特定モデル開発者が意図しないところでの学習データが活用されることを予防するため、基盤モデルにおいて設定オプションが用意されている場合は、当該学習データセットが基盤モデルに反映されないような設定がされているか確認することも望ましい措置のひとつである。なお、対応が難しい場合、基盤モデルの再選定を行うことやオンプレミスの基盤モデルをファインチューニングなどすることが検討できているかもあわせて確認しておくとよい。

③ 学習データにおける個人情報の取扱い

個人情報の収集に当たっては、利用目的の本人への通知又は公表が必要（個人情報保護法第21条）であり、収集する個人情報に個人の病歴等要配慮個人情報が含まれている場合は、原則本人の同意が必要（個人情報保護法第20条）であることから、学習データとして個人情報を収集し、基盤モデルに入力して学習させる場合は本人の同意を得られているか確認しておくことが必要である。

なお、要配慮個人情報が学習データに含まれる場合のデータ管理等は厳格かつ適切に実施する必要があり、「2-3-1.　国内における関連制度の概要」記載の個人情報保護委員会発出の「OpenAI社に対する注意喚起」を踏まえ、具体的には以下に注意を配ることが重要である。

(a) 収集する情報に要配慮個人情報が含まれないよう必要な取組を行うこと。 例えば、設計段階で取集する情報のうち必要最低限の項目を定義すること。
(b) 情報の収集後、可能な限り即時かつ適切に収集した情報に含まれ得る要配慮個人情報を減少させるための措置を講ずること。
(c) 上記(a)及び(b)の措置を講じてもなお収集した情報に要配慮個人情報が含まれていることが発覚した場合には、可能な限り即時に、かつ、学習用データセットに加工する前に、当該要配慮個人情報を削除する又は特定の個人を識別できないようにするための措置を講ずること。個人を識別できないようにする技術的工夫の一つとして例えば、当該情報そのものを閲覧不可能にするマスキング対応が挙げられる。一方で要配慮個人情報は画一的な内容ではないため、マスキング技術だけでも完全に対応ができる場合は少なく、マスキング後に目検でのチェックを組み合わせるなど、複合的な措置が想定される。
(d) 利用者が機械学習に利用されないことを選択してプロンプトに入力した要配慮個人情報については正当な理由がない限り、取り扱わないこと。

④ 学習データにおける著作権情報等の取扱い

学習データの内容として例えば論文情報等を活用する場合も想定されるが、著作権侵害等のリスクをできる限り低減させるため、学習前に当該論文情報等のデータベース（例えばインターネット上のサイト等）においてAIの学習に用いることを禁止されていないか等、学習データとしての利用制限がないか確認しておくことが望ましい。また、取引先を含む他社から提供を受けて自社が保有する情報等、いわゆる秘密・機密情報に該当する情報を学習データとして活用する場合においては、提供を受けたデータを学習利用することが許容する条項が含まれているか、相手方との契約内容を事前に確認することが望まれる。

4-2-2. フューショットラーニング等に利用するサンプル・事例の取扱い

フューショットラーニングや強化学習に利用するサンプル・事例の取扱いに当たってのチェックポイントは以下のとおりである。

① 基盤モデル又は特定モデルの利用規約の確認

「4-2-1.ファインチューニングに利用する学習データの取扱い」におけるチェックポイント①に記載の趣旨と同様、サービス・プロダクト提供者が意図しない範囲でサンプル・事例が活用されることを予防するため、フューショットラーニングや強化学習に利用するサンプル・事例が基盤モデル又は特定モデルに反映されないよう、両者の利用規約等でフューショットラーニング等に利用するサンプル・事例の取扱いに関する事項が利用規約で明瞭化されているか等、フィーショットラーニング以外でサンプル・事例が利用されることがないかを確認することが望ましい。

② 基盤モデル又は特定モデルの設定オプションの確認

上記①で記載のとおり、サービス・プロダクト提供者が意図しない範囲でサンプル・事例が活用されることを予防するため、基盤モデル又は特定モデルにおいて設定オプションが用意されている場合は、当該サンプル・事例が基盤モデル又は特定モデルに反映されないような設定がされているか確認することも望ましい措置のひとつである。

③ サンプル・事例における個人情報の取扱い

サンプル・事例において個人情報や要配慮個人情報が含まれる場合、「4-2-1.ファインチューニングに利用する学習データの取扱い」におけるチェックポイント③に記載の趣旨と同様、本人への通知又は公表や、同意が必要であり、運用面における考慮事項についてもあわせてサービス・プロダクト提供者において確認されたい。

④ サンプル・事例における著作権情報等の取扱い

サンプル・事例において論文情報等の著作権情報や機密・秘密情報が含まれる場合、サービス・プロダクト提供者においては「4-2-1.ファインチューニングに利用する学習データの取扱い」におけるチェックポイント④に記載の対応を行うことが望ましい。

4-2-3. 利用者が入力する質問データの取扱い

利用者が入力する質問データの取扱いに当たってのチェックポイントは以下のとおりである。

① 基盤モデル又は特定モデルの利用規約の確認

「4-2-1.ファインチューニングに利用する学習データの取扱い」におけるチェックポイント①に記載の趣旨と同様、利用者が意図しない範囲で入力した質問内容が活用されることを予防するため、利用者が入力した質問内容が基盤モデル又は特定モデルに反映されないよう、基盤モデル又は特定モデルの利用規約等で質問データの取扱いに関する事項が利用規約で明瞭化されているか等を確認することが望ましい。

② サービス・プロダクトにおける設定オプションの確認

利用者の意図しない範囲で自己の質問内容が特定モデルに反映されることがないようにするため、サービス・プロダクト提供者においてサービス・プロダクトの設定オプションが用意できる場合、利用者によって入力された質問内容が特定モデルに反映されないような設定（利用時に質問データの反映に関する「ON/OFF」が切り替えられる設定が措置されている等）を準備することが望ましい。この場合、利用者によるすべての利用者の質問データが常に利用できるわけではないことを前提にサービス開発段階において利用者がオプトアウト（利用を拒否）した場合に、オプトアウト時点で学習データとして活用されないように変更できるサービス設計・ファインチューニング等に活用するサンプル・事例として利用されないような具体的な設計を考えておくことも必要

である。

③ 質問データにおける個人情報の取扱い

利用者の質問データにおいて個人情報や要配慮個人情報が含まれる場合、「4-2-1.ファインチューニングに利用する学習データの取扱い」におけるチェックポイント③に記載の趣旨と同様、本人への通知又は公表や、同意が必要である。この点、ヘルスケア領域において生成AIを活用したサービス・プロダクトを提供する場合、利用者において自己の既往歴や病状を含む健康状態等、個人情報や要配慮個人情報に該当するおそれのある情報が質問内容として入力される場面が多くなることを想定し、特に取扱いには注意が必要である。また、質問データを出力以外の用途、例えばサービス・プロダクトの精度向上を目的とした学習に利用することを想定している場合、質問データを学習に利用する想定である旨の目的の特定を行う必要がある。

④ 質問データにおける著作権情報等の取扱い

利用者において自由に質問を入力できるため、例えば生成AIへの利用が禁止された著作物等の情報を利用者が質問として入力する場面も考えられる。利用者がどのような種類・内容の質問を入力するのかの詳細をサービス・プロダクト提供者側でコントロールすることは容易ではないため、生成AIへの利用が禁止された著作物や機密・秘密情報が入力された場合の情報の取扱いについて、サービス・プロダクト提供者において事前に免責事項を作成・公開しておくことも重要である。

⑤ 質問データにおける悪用・不正目的対策

利用者が悪用や不正を目的とした質問データを組成することで、例えばマルウェア製造といったサービスの意図しない出力を利用できてしまう場合がある。機械的な処理で悪用・不正目的の入力をリジェクトすることは技術的な制約があるため、サービス・プロダクト提供者において事前に免責事項として公開する範囲に、当該用途での生成指示入力を禁じることが重要である。また、サービス・プロダクト提供者の提供するシステムで入力・出力の履歴を保持し、不正利用のトレースを可能としておくこと、ならびにサンプリングによる定期不正抽出等をシステム運用に組み込むことも対策として考えられる。

4-2-4. データに関するその他の考慮事項

データの取扱いに当たってのその他のチェックポイントは以下のとおりである。

① データ保護に関する社内体制の構築

個人情報をはじめとするデータの取扱いに関する制度や規制は環境の変化とともに制度改正や新たな論点についての検討が繰り返される領域である。そのため、個人情報等に関する規制の遵守や最新動向についての情報収集をして適切にサービスが設計・提供できるか担保するため、特定モデル事業者やプロダクト・サービス提供事業者の組織内にデータ保護に関する社内体制が構築されていることが望ましい。

② 関連ガイドライン等の参照

個人情報保護法や著作権法以外にも、官公庁において当該制度の詳細について記載しているガイドラインやハンドブック等が公開されている。これらの関連ガイドライン等を網羅的に参照することで、法律レベルでは抽象度の高い文言についても運用に照らした際の具体的な遵守すべき項目の目安が認識できるため、関連ガイドライン等を参照することも重要なポイントになる。関連ガイドライン等としては、例えば厚生労働省の「医療情報システムの安全管理に関するガイドライン」及び経済産業省・総務省の「医療情報を取り扱う情報システム・サービスの提供事業者における安全管理ガイドライン」（いわゆる3省2ガイドライン）や、個人情報保護委員会が策定している「個人情報の保護に関する法律についてのガイドライン」、総務省・経産省の「AI事業者ガ

イドライン」や「DX時代における企業のプライバシーガバナンスガイドブック」、医療AIプラットフォーム技術研究組合策定の「医療・ヘルスケア分野における生成AI利用ガイドライン」、一般社団法人AIメンタルヘルスケア協会が策定予定のメンタルヘルスケアに特化した生成AI活用ガイドライン等が挙げられる。

③ サービスアップデートや機能評価における考慮事項

特にサービス・プロダクト提供者において、サービスのアップデートが行われる場合（例えばユーザーが入力する質問データから本人に関する属性等を類推し、機械的に判断するような機能が実装される場合等）に当たっても、サービス設計時に前述のチェックポイントを再び考慮・参照することが望ましい。また、生成AIを活用したサービス・プロダクトの品質向上や基盤モデルに対する追加学習のためのフィードバックの観点で、提供するサービス・プロダクトの機能評価が実施されるようなサービス設計が行うことも期待されるポイントである。

④ プライバシー・バイ・デザインを考慮した対策の実施

学習やファインチューニング時など、プロダクト開発段階においてはプライバシー・バイ・デザインを通じて個人のプライバシーに配慮した設計を行うことが重要である。例えば、学習時のデータについて、第三者の個人情報や知的財産権に留意が必要なもの等が含まれている場合には、法令に従って適切に扱うことをAIのライフサイクル全体を通じて確保することや、AIシステムの実装の過程を通じて、採用する技術の特性に照らし適切に個人情報へのアクセスを管理・制限する仕組みの導入する等のプライバシー保護のための対策を講ずることも重要である。

4-3. アウトプットの信頼性に関するチェックポイント

サービス・プロダクト提供者がアウトプットの信頼性を担保するに当たってのチェックポイントは以下のとおりである。

【アウトプットの信頼性に関するチェックポイント】
- サービス・プロダクト開発段階での取組
- サービス・プロダクト提供時の利用者に対する取組

4-3-1. サービス・プロダクト開発段階での取組

サービス・プロダクト開発段階での取組に当たってのチェックポイントは以下のとおりである。

① ハルシネーションを制御する工夫の実施

ハルシネーションのリスクを低減する手段は、昨今LLMの活用事例等が増加するにしたがってアプローチ方法も様々なものが活用されているところである。

例えば、回答精度を向上することでハルシネーションのリスクを低減させる方法として、基盤モデルのタイプとしてクラウドベースで汎用的かつ自然言語処理が可能なモデルを利用するほか、当該モデルに対するファインチューニングやプロンプトエンジニアリング[21]を実施することが挙げられる、

また、エンベディング[22]等の技術を利用してサービス・プロダクト提供者のデータベースを活用することで出力結果の整合性を担保したりアウトプットの根拠や引用元を表示する技術（いわゆるグラウンディング[23]）の導入や、RAGの活用（次ページのコラム参照）、フィルタリングによるハルシネーション検知なども、生成AIの出力結果の信頼性を担保する手段として事業者において

[21] 言語モデルへの命令（プロンプト）を開発・最適化すること。用語集参照。
[22] 単語や文といった自然言語の情報を、低次元のベクトルで統一的に表現すること。単語や文が持つ意味の関係性や類似度が計算可能になる技術のこと。用語集参照。
[23] AIが言葉や概念を具体的なものや実際の世界と結びつけて理解する能力のこと。AIが現実世界のコンテキストを理解し、それに基づいた回答や生成を行うことを可能にすること。用語集参照。

取り得るものである。

② アウトプットのランダム性に対する工夫の実施

サービス・プロダクトの性質によっては、同じ質問におけるアウトプットのランダム性の必要有無が変わってくることが想定される。例えば、小説生成やチャットボットなどクリエイティブなタスクを想定したサービス・プロダクトだと、毎回多様な回答）が好まれることが多い一方で、ヘルスケア領域や法律など規制や制度に深く関わるタスクにおいては同じ入力に対して、同じ回答が出力される方が好まれるケースが
多いことが考えられる。そのため、アウトプットのランダム性は、APIのパラメータ（例えば「temperature」）などのオプションで指定が可能であるため、サービス・プロダクトの性質によって指定することが重要になる。

③ セキュリティ・バイ・デザインを考慮した情報セキュリティ対策の実施

一般的なシステム同様に生成AIを活用したサービス・プロダクトの開発にあたっても、情報セキュリティを企画・設計段階から確保するセキュリティ・バイ・デザインの考え方に則った対策を講じることが肝要である。開発プロセスの早い段階からセキュリティを考慮することで最終的に手戻りが少なくなり、機密性・完全性・可用性を確保した保守性の高いサービス・プロダクトの開発、提供の実現に繋がる。なお、セキュリティ・バイ・デザインを考慮した対策については「AI事業者ガイドライン」にもその重要性が記載されているところである。

生成AIをサービス・プロダクトの開発に活用する場合、採用するAIモデル、学習データとその十分性、パラメータやアウトプットのランダム性等の特徴・性質を鑑みたリスクや脆弱性を悪用した攻撃への対策を行うことが重要である。例えば、データポイズニングによる誤作動の誘発やプロンプトインジェクション等による情報の引き出し等は生成AIにおいても関係するAI特有のリスクと言える。また、万が一、セキュリティが侵害された場合に講じるべき措置について、当該サービス・プロダクトの用途、特性、健康被害を含む侵害の影響の大きさ等を踏まえリスク評価を行い、必要なPDCAを回していくことが重要である。

④ サービス品質評価の実施

生成AIを活用して提供したサービス・プロダクトに関して、その技術性能や品質についての評価を実施することで、プロダクトに採用したモデルの技術的な精度等を適切に評価することもアウトプットの信頼性を担保するうえで重要である。また、当該サービス・プロダクトの利用場面に応じた適切な評価指標を設計した上で、その評価指標において結果が適切にアウトプットできているか評価する体制・プロダクト設計を行うことも望ましい。さらに、アウトプットの信頼性向上のための取組として昨今ではRLHF（Reinforcement Learning from Human Feedback：人間のフィードバックからの強化学習）も注目されている。これは、サービス・プロダクトにおける利用者からのGood/Badボタンのように、ユーザーの行動をフィードバックに活用する取組であり、生成AIを含めたAI開発の領域においては有効なアプローチである。

＜コラム＞ **RAG**（**Retrieval-Augmented Generation**）の特徴と活用

RAG（Retrieval-Augmented Generation）とは、大規模言語モデル（LLM）に、検索エンジンによる情報検索機能を組み合わせることで、より正確で信頼性の高い情報を生成させる仕組みである。

ユーザーが入力したプロンプトに関連する情報を外部データベースから取得し、取得した情報とユーザープロンプトを組み合わせて、LLMで回答を生成する。再学習の必要なくドメイン知識の獲得や企業独自のデータが活用可能になり、回答の質やハルシネーションの軽減が見込まれることから、様々な分野で注目を集めている。

1．RAGの特徴

①出力結果の正確性:
外部データベースから質問に関連する情報を検索し、その情報とプロンプトを組み合わせて回答を生成することから、ハルシネーションのリスクを軽減することが期待できる。
②カスタマイズ性
特定のドメイン（医療、法律など）に特化した知識ベースを構築することで、モデルの再学習なく、専門的な質問に回答することができる

2. RAGのユースケース
①チャットボットへの活用
RAGが広く活用されているケースとして、チャットボットが挙げられます。社内規定マニュアルやFAQなどが含まれたデータベースを構築し、RAGを用いることでユーザーからの問い合わせに対して最適な回答を提供することが可能になる。
例えば、会社のエンドユーザーによる「どの手続きにどのフォームが必要か」といった質問や、「社内ポリシーの詳細を教えて欲しい」といったリクエストに対し、最新のデータを参照した回答が行われます。これにより、大量の情報から必要な情報を簡潔に抽出して提供できるため、ユーザーの満足度向上やサポート業務の改善が見込まれる事例が創出されている。

3. 注意点
RAGは、外部データベースとの連携により、特定ドメインのナレッジや情報を容易に取り込むことが可能。しかし、データの質や構造が不十分な場合、ハルシネーションが発生したり、不正確な情報が生成されるリスクがある。高精度な回答を得るためには、ベクトル検索の手法、適切なデータの加工（構造化）など、データ前処理が不可欠である。また、生成された回答の評価を行い、継続的にモデル等を改善していくことが重要である。

RAGは、ハルシネーションの低減、拡張性を兼ね備えた強力な技術であり、適切なデータ管理と構造化を行うことで、様々な分野での応用が期待される。

4-3-2. サービス・プロダクト提供時の利用者に対する取組

サービス・プロダクト提供時の利用者に対する取組に当たってのチェックポイントは以下のとおりである。

① テキスト生成AIに関する説明・表示

生成AIの特徴について利用者がよく理解した上でサービス・プロダクトを使うことができるように、チュートリアルの提供等、テキスト生成AIとはどのようなものなのかについてユーザーへ説明・啓発を適切に実施していることが望ましい。

② 利用用途や利用者が入力した質問データの取扱いに関する説明・表示

サービス・プロダクト提供者が意図しない不適切な用途で利用者がサービスを利用し、それに伴って結果的に利用者に不利益が生じることがないよう、利用者に対して利用用途を説明・表示し、サービスのコンセプトや安全管理措置の明確化することが望ましい。また、利用者が入力した質問データがどのように扱われるのか（学習データとして使う場合があるのか等）の取扱いの説明・表示をしておくことも重要である。

③ 入力規制・制御の実施

②で記載した趣旨と同様、利用者が不適切な入力をすることがないような仕組みとして、利用者が使用するプロダクトの入力画面等にある程度の入力規制・制御をかけることで利用者に不利益が生じないようなサービスに設計することもひとつのアプローチである。

④ AIによる生成であることの表示

利用者がAIによる生成であることを理解・認識できるように、利用者に対して、利用者が当該サービス・プロダクトを利用することで得られる出力結果について、その出力結果がAIによるものだと明示される設計になっていることが望ましい。

⑤ 免責条項の表示

何か責任問題が生じた際の責任分界点等を利用者が理解した上で利用できるように利用者に対して、サービス・プロダクト提供者側の免責事項を明示していることも重要である。なお、サービス・プロダクトを他の事業者に提供し、当該事業者から利用者に提供している場合（BtoBtoCの形式での提供の場合）は、アウトプットが他の著作物の権利を侵害することがないかを利用者に対してサービス等を提供する事業者においても確認することが事業者間の契約において盛り込まれていることが望ましい。

⑥ プライバシーポリシーの策定

利用者の要配慮個人情報を含む個人情報が適切に利用・保護されるように、サービス・プロダクト提供者は管理体制を構築し、維持する必要がある。具体的には、情報漏えいや不正アクセスを防ぐためのセキュリティ対策の実施、個人情報の利用目的の明確化、データ保持の期間や方針の明確化やデータの取り扱いに対するオプトイン/オプトアウトの設定などがある。また、第三者との間で情報の共有が行われる場合には、可能な限りユーザーの事前同意を取得し、契約において適切な取り決めを行い、個人情報が適切に取り扱われるように配慮することが重要である。

⑦ UX/UI設計

ユーザーがサービス・プロダクトを直感的かつ安全に利用できるように、上記①～⑥を踏まえたUX/UI設計を図ることも重要である。例として視覚的なわかりやすさの確保、利用用途やデータ取扱いに関する適切なフロー、入力規制におけるフィードバック設計、AI生成であることの可視化、プライバシー保護を意識したUIの配置などのUX/UI設計を考慮することが望ましい。AIをそもそも使いたくないユーザー向けにAIを使わないような動線を準備し、AI活用に関する選択肢を提示することも一つのオプションとして考えられる。

⑧ 医療機関等向けの生成AIプロダクト・サービスにおける院内ルールの設計

生成AIプロダクト・サービスのユーザーが医療機関等法人単位での導入になる場合には、当該ユーザーが所属する組織内に生成AIの使用に関する内規（組織内ルール）を設けることもひとつの手段である。生成AIを活用したプロダクト・サービス導入が初めての場合、生成AIの特性を使用者ひとりひとりが把握・理解することで、安心・安全な活用ができる環境が醸成されることから、最低限遵守すべき活用ルールなどをあらかじめ定めておくことでリテラシー向上につながり得る。具体的なルールひな型については別添2を参照。

4-4. ヘルスケア領域における個別規制に関するチェックポイント

ヘルスケア領域においては、医療機器プログラムの該当有無によって標榜可能な内容・範囲が変わるなど、「プログラム医療機器該当性に関するガイドラインについて」（厚生労働省）をはじめとする個別制度があるため、それらを開発・提供前に十分確認・遵守する必要がある。ヘルスケア領域における個別規制に関するチェックポイントは以下のとおりである。

なお。関連制度については、「4-2-4. データに関するその他の考慮事項」の②のチェックポイントもあわせて確認されたい。

① 医療機器プログラムの該当性確認

開発・提供しようとしている生成AIを活用したプロダクト・サービスは医療機器プログラムに該当するかを確認できていることが望ましい。

② 標榜における広告規制の適合性確認

医療機器プログラムとして判断された場合、承認又は認証された範囲内での医学的表現、いわゆる効果・性能の標榜が可能になり、医薬品等適正広告基準や医機連など業界がガイドとして公表しているガイドラインに基づいて広告・表示を行う必要がある。

一方で非医療機器プログラムの場合、医学的な標榜はできず、薬機法及び景品表示法による広告・表示規制の対象となるため、これら規制に抵触しないように標榜しなければならない。該当性や広告規制等に関する不明点は各都道府県薬務主管にて相談が可能である。詳細については薬機法、プログラムの医療機器該当性に関するガイドラインや景品表示法等の関連法規を参照されたい。

③ 基盤モデルの利用規約の確認

ヘルスケア領域における生成AIの活用に当たっては、利用する基盤モデルによってヘルスケア領域における利用制限をはじめとした活用にあたっての様々な制約が決められていることもあるため、OpenAI社の利用規約をはじめとした、基盤モデルの利用規約を確認し、規約のアップデートにより活用の範囲に支障がないかを確認することが望まれる。

5. 今後に向けて - 業界団体としての取組と期待-

前述のとおり、昨今ヘルスケア領域において生成AIを活用したサービスを提供する事業者が急速に増加する中、一般の利用者がAIに関する専門知識を持たずともこれらのサービスを容易に利用できるようになっている。また、2024年に入り生成AIを取り巻く技術や動向の進化、それに伴うプロダクト・サービス化の推進などが日進月歩である。そのような状況において、生成AIの特徴・性質から生じる、著作権の帰属や倫理的問題、セキュリティとプライバシーのリスクやハルシネーションといったさまざまな課題が指摘されている。このような背景を踏まえ、ヘルスケア領域における生成AIに関する論点を整理し、事業者側で生成AIを活用してサービス・プロダクトを提供する場合に配慮すべきポイントをまとめたのが本ガイドであり、当該事業者において適切なサービス提供が可能となる環境を整備するための一つのアプローチとして本ガイドが活用されることが望まれる。すなわち、当該事業者が本ガイドを参考にすることにより、当該サービスを提供しようとする事業者が目安・基準となるチェックポイント（参考資料　別添1　生成AIを活用したサービス・プロダクトを提供する事業者向けチェックリスト）に基づいたセルフチェックを行い、本ガイドに沿ってサービス提供を行っている旨を公表等することで、当該サービスを利用するユーザー（医療機関等）が、安心してサービスを利用できるようになることも期待しているところである。

本ガイド公表後に策定された経済産業省及び総務省のAI事業者ガイドラインでは「事業者の自主的な取組」「国際的な議論との協調」「読み手にとってのわかりやすさ」の3軸を基本的な考えとしており、「本ガイドラインは、AI開発・提供・利用にあたって必要な取組についての基本的な考え方を示すものである。よって、実際のAI開発・提供・利用において、本ガイドラインを参考の一つとしながら、AI活用に取り組む全ての事業者が自主的に具体的な取組を推進することが重要となる。」と記載されている。まさに本ガイドは1軸目の「事業者の自主的な取組」のひとつであり、第2.0版改正に当たってはAI事業者ガイドライン記載の原則を踏まえたアップデートもカバーしたところであるため、ヘルスケア領域で生成AI活用に取り組む業者が本ガイドを参考にしながら具体的な行動に取り組むきっかけになることを期待している。

なお、急速な新技術の進歩に伴い、生成AIの技術特性やその活用方法、関連の制度や規制内容の変化・進化が想定されることから、本ガイドは適宜見直しを図り、事業者がタイムリーに適正なサービスを提供できるよう、今後も業界全体の後押しを行っていく予定である。

デジタル技術を活用したヘルスケアサービスの浸透によって、疾患の早期発見・治療、未病・予防に寄与し、利用者の健康意識の高まりが国民の健康増進に期待できると考えられる。日本デジタルヘルスアライアンスでは、業界全体で本ガイドの普及・浸透を図り、生成AIを活用したヘルスケアサービスが安全・安心な形市場へ提供されることへの貢献を今後も目指していく次第である。

参考資料

用語集
別添1　生成AIを活用したサービス・プロダクトを提供する事業者向けチェックリスト
別添2　ヘルスケア領域における生成AIに関する取組
別添3　医療機関内生成AI活用ポリシーひな型（案）
別添4　本ガイドの検討経緯及び参加・協力企業等

用語集

用語	用語の定義・意義
生成AI	生成AI（Generative AI）とは、自律的に学習したデータから文章、画像、音声などの一見新しく現実的なコンテンツを生成することができる、一連のアルゴリズムのこと。
テキスト生成AI	生成AIのうち、文章を生成するアルゴリズムのこと。
ヘルスケアサービス	健康の保持及び増進、介護予防を通じた 健康寿命の延伸に資する商品の生産若しくは販売又は役務の提供を行うこと。本自基準においては、そのうち医療機器又は医療機器プログラムには該当しないものを差す。 参考文献：経済産業省「ヘルスケアサービスガイドライン等のあり方 」（2019年4月12日 （2023年6月9日改訂） https://www.meti.go.jp/policy/mono_info_service/healthcare/210609guide.pdf
基盤モデル	特定のタスクに限定せず大量のデータから汎用的に学習した機械学習モデルのこと。
大規模言語モデル（Large Language Model：LLM）	生成AIのうち、大規模データを学習させた言語モデルのこと。
ファインチューニング	学習済みの基盤モデルに対して別のデータセットを活用して追加学習させること。
ハルシネーション	生成AIにおいて、事実と異なる内容（嘘）や文脈と全く無関係な内容の出力が生成されること。
プロンプト入力	生成AIにおいて、ユーザーが入力する指示や質問のこと。
プロンプトエンジニアリング	生成AIにおいて、言語モデルへの命令（プロンプト）を開発・最適化すること。
フューショットラーニング	フューショットラーニング（Few-Shot Learning）は、比較的大量のデータを必要とする従来のファインチューニング手法と対照的に、非常に少量のデータを使用して機械学習モデルに推論を行わせる手法のこと。 参考文献：Few-shot learning in practice: GPT-Neo and the Accelerated Inference API https://huggingface.co/blog/few-shot-learning-gpt-neo-and-inference-api
エンベディング	単語や文といった自然言語の情報を、低次元のベクトルで統一的に表現すること。単語や文が持つ意味の関係性や類似度を、数値演算によって計算可能になる。

グラウンディング	グラウンディング（Grounding）とはAIが言葉や概念を具体的なものや実際の世界と結びつけて理解する能力を指す。この概念はAIが単なるデータや情報の集合ではなく現実世界のコンテキストを理解し、それに基づいた回答や生成を行うことを可能にする。そのため生成AIが誤った情報や関連のない内容を生成すること（ハルシネーション）を防ぐための手法としても用いられる。 参考文献：グラウンディング（Grounding）とは？その意味とビジネスへの影響 https://www.salesforce.com/jp/blog/jp-what-is-grounding/#:~:text=%E3%82%B0%E3%83%A9%E3%82%A6%E3%83%B3%E3%83%87%E3%82%A3%E3%83%B3%E3%82%B0%E3%81%A8%E3%81%AF%E3%80%81AI%E3%81%8C%E8%A8%80%E8%91%89%E3%81%A8%E5%AE%9F%E4%B8%96%E7%95%8C%E3%82%92%E9%96%A2%E9%80%A3%E4%BB%98%E3%81%91%E3%81%A6%E7%90%86%E8%A7%A3%E3%81%99%E3%82%8B%E3%81%93%E3%81%A8
ChatGPT	OpenAIによって開発された対話型の人工知能（AI）ツール。ChatGPTは自然言語での質問応答、チャットボット機能、文章の生成や翻訳など多岐にわたる用途に対応している生成AIの一種であり、事前に学習されたトランスフォーマーモデルに基づいて機能する。
要配慮個人情報	「要配慮個人情報」とは、本人の人種、信条、社会的身分、病歴、犯罪の経歴、犯罪により害を被った事実その他本人に対する不当な差別、偏見その他の不利益が生じないようにその取扱いに特に配慮を要するものとして政令で定める記述等が含まれる個人情報をいう。 参考情報：個人情報の保護に関する法律についてのガイドライン（通則編）（案） https://www.soumu.go.jp/main_content/000450626.pdf
Transformer	TransformerはAIの性能を向上させるための深層学習（ディープラーニング）モデルの一つで、特に自然言語処理（NLP）の分野で重要な役割を果たしている。このモデルは後にNLP分野においてブレイクスルーをもたらしたBERTやGPT-2などのモデルの基盤となっており、ChatGPTをはじめとする文章生成のジェネレーティブAIや自然言語処理の理解に不可欠なモデルである。このモデルの仕組みや特徴により生成AIは高性能な自然言語処理モデルを実現している。 参考情報：Attention Is All You Need https://arxiv.org/abs/1706.03762
BERT	BERT（Bidirectional Encoder Representations from Transformers）は2018年にGoogleにより発表された自然言語処理（NLP）に特化した深層学習モデルの一つ。2022年から2023年にかけて生成AIが流行する以前は、医療分野においてはこのモデルを医療関連のデータでファインチューニングして利用することが多かった。 参考情報：BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding https://arxiv.org/abs/1810.04805

GPT-3	GPT-3（Generative Pre-trained Transformer 3）は、OpenAI社によって2020年に開発された言語モデル。 参考情報：A Survey of Large Language Models https://arxiv.org/abs/2303.18223
GPT-4	GPT-4（Generative Pre-trained Transformer 4）は、OpenAI社によって2023年に開発された言語モデル。 参考情報：A Survey of Large Language Models https://arxiv.org/abs/2303.18223
PaLM	PaLMは、Google社によって2022年に開発された言語モデル。 参考情報：A Survey of Large Language Models https://arxiv.org/abs/2303.18223
LLaMA	LLaMAは、Meta社によって2023年に開発された言語モデル。 参考情報：A Survey of Large Language Models https://arxiv.org/abs/2303.18223
プログラム医療機器	2014年11月25日に施行された「医薬品、医療機器等の品質、有効性及び安全性の確保等に関する法律」（医薬品医療機器等法）では、国際整合性等を踏まえて、疾病の診断・治療等を目的とした単体プログラム（ソフトウェア）についても医療機器としての規制対象としており、多くのプログラム医療機器が開発され、製造販売承認等されている。 参考情報：PMDA プログラム医療機器について https://www.pmda.go.jp/review-services/drug-reviews/about-reviews/devices/0048.html#faq
RAG(Retrieval-Augmented Generation)	RAGは生成AIに対して自身が持っている情報を教えて、それを使ってアウトプットを受ける仕組み。生成AIの精度と信頼性を向上させるための手法でもある。 RAGの仕組みは主に2つのステップで構成されている 〇検索(Retrieval)：ユーザーの質問に関連する情報を、事前に準備された知識ベースから検索する。 〇生成(Generation)：検索で得られた関連情報を参照しながら、大規模言語モデル(LLM)が回答を生成する。 この方法により、AIは最新の情報や専門的な知識を参照しながら回答を作成できるため、より正確で信頼性の高い情報を提供することが可能。RAGは、企業の内部文書や最新の外部情報などを活用することで、AIチャットボットをより賢く、ユーザーのニーズに適合させることができる技術として注目されている。

SLM（Small Language Model）	SLM（小規模言語モデル）は、大規模言語モデル（LLM）と比較してパラメータ数が少ない、コンパクトで効率的な言語AIモデル。 一般的に数百万から数十億のパラメータを持ち、LLMの数百億から数兆パラメータと比べて小規模であり、特定の分野や目的に絞ってトレーニングされることが多く、その分野では高い性能を発揮するとされている。開発や運用にかかるコストを抑えられる一方で限られたデータ量で処理するため、応答が速いのが特徴。また、特定の業務や用途に最適化しやすく、ファインチューニングが容易であrため、リソースの制約がある環境や特定の用途に特化したAI応用において、効果的な選択肢となっている。
日本語特化型LLM	日本国内事業者が一から開発に取り組むLLM。例えば、2024年11月にはSoftbank社が4600億パラメータの日本語LLMを公開するなど、日本でも大規模なLLMの開発が進んでいる状況。

別添1　生成AIを活用したサービス・プロダクトを提供する事業者向けチェックリスト

（別添Excelを参照）

別添2　ヘルスケア領域における生成AIに関する取組

事例1　医師向け臨床支援アプリ**HOKUTO**（株式会社**HOKUTO**）

- インプットから臨床現場のアウトプットまで医師の医学情報収集をフルサポートする情報収集アプリ。
- 本サービス内において、①患者への説明文生成AI（患者や家族に対して病状や治療内容を説明する文章をAIが簡潔で分かりやすくまとめる機能）、②論文検索＆要約AI（キーワードと期間を入力するだけで論文が検索され、検索結果の要約等のコメントをAIが生成する機能）を生成AIを活用した機能として展開。

（出典：株式会社HOKUTO提供資料）

事例2　ユビーメディカルナビ生成**AI**（**Ubie**株式会社）

- Ubie株式会社は「ユビー生成AI」を2024年5月から提供開始し、病院内の様々なユースケースに応じ、生成AIの活用による業務効率化の支援を実施。
- 具体の機能としては、文章生成・要約/音声認識/画像認識の機能により紹介状・退院サマリ/退院看護サマリ・IC記録作成、紹介状作成の支援が可能。
- 直近での導入病院（※試験導入を含む。）
    - 九州大学病院、恵寿総合病院、ヨコクラ病院、岡山旭東病院、山陽病院、浦添総合病院、高石藤井病院 等）

（出典：Ubie株式会社提供資料）

事例**3**　医師国家試験と生成**AI**に関する研究（株式会社**MICIN**）

- 金沢大学医学類の学生及び融合研究域融合科学系 野村 章洋 准教授らの研究グループと共同でChatGPTおよびGPT-4を用いて第117回医師国家試験(2023年2月開催)を解かせる研究を実施
- その結果、必修問題で82.7%、基礎・臨床問題で77.2%のスコアを獲得したことで、それぞれ合格最低ラインである80.0%と74.6%を満たし合格点に到達する結果に。

2023.4.25　その他サービス等

MICIN、金沢大学と実施していたChatGPTおよびGPT-4を用いて第117回医師国家試験(2023年2月実施)を解かせる研究において初めて合格点に到達し、その成果を論文としてオンライン公開いたしました

株式会社MICIN（本社：東京都千代田区、代表取締役CEO：原聖吾、以下 MICIN）は、金沢大学医学類の学生ならびに融合研究域融合科学系 野村 章洋 准教授らの研究グループと共に、ChatGPTおよびGPT-4を用いて第117回医師国家試験(2023年2月開催)を解かせる研究に関する論文をオンライン公開いたしました[1]。

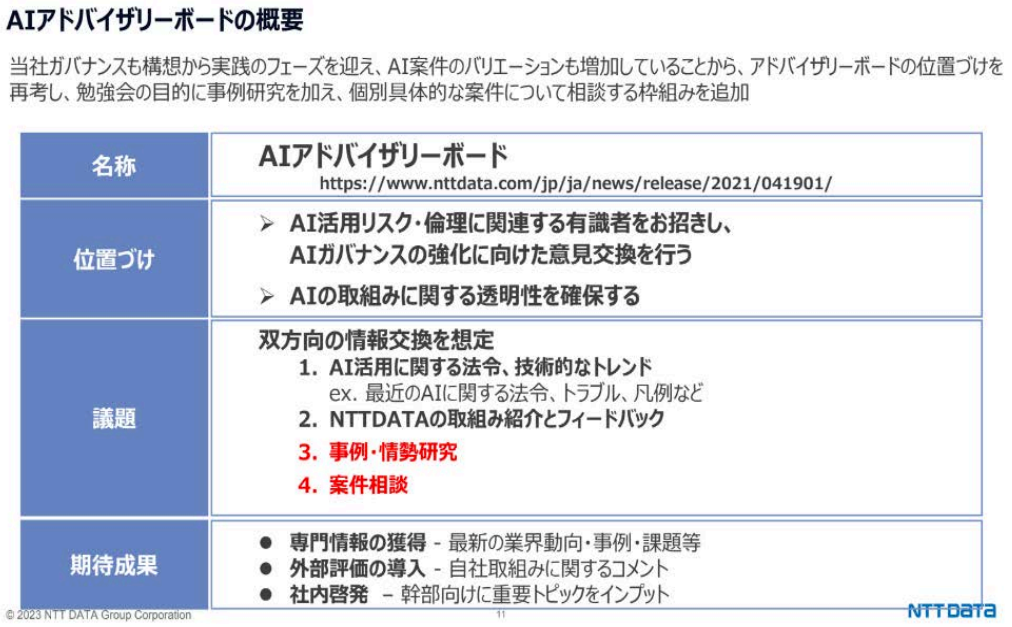
本論文では日本国における最新の医師国家試験（第117回2023年2月開催）の画像なし問題262問を対象としてChatGPTおよびGPT-4の性能検証を実施し、その結果、必修問題で82.7%、基礎・臨床問題で77.2%のスコアを獲得したことで、それぞれ合格最低ラインである80.0%と74.6%を満たし合格点に到達いたしました。第117回医師国家試験でのChatGPTおよびGPT-4を用いた出力結果の合格点の到達は、本研究が初となります（※）

また本論文ではChatGPTおよびGPT-4の出力結果のうち、不正解となった56問の発生要因についても詳細な調査を実施いたしました。それらの調査からChatGPTおよびGPT-4が誤答を生成する3大要因として「医学知識の不足」「日本特有の医療制度に関する情報」「計算問題での誤り」を特定することができました。

（出典：株式会社MICIN提供資料）

事例**4** **AI**ガバナンス体制の構築（株式会社**NTT**データ）

- AI活用リスクや倫理に関連する有識者を構成員とするAIガバナンス体制を社内に構築し、AIに関する法令や技術的なトレンドのインプットや自社の取組紹介・フィードバック等を実施。
- AI案件のバリエーションが増加していることを踏まえ、事例研究や個別具体的な案件について相談できる枠組みも追加

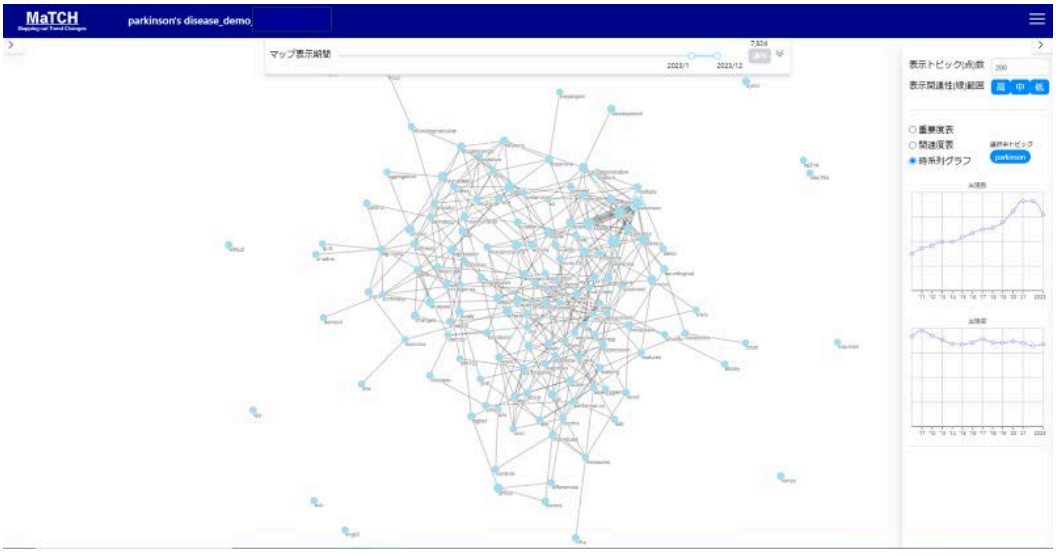


（出典：株式会社NTTデータ提供資料）

事例**5** **MaTCH**：**Mapping out Trend Changes**システム（小野薬品工業株式会社）

- MaTCH（Mapping out Trend Changes）システムは、小野薬品工業株式会社とAI（人工知能）開発会社である株式会社アイエクセスが共同で開発した医学論文のトピック抽出分析AIシステムである。
- 本システムは、メディカルアフェアーズ活動に適合させた独自のトピック抽出モデルであるAI自然言語処理アルゴリズムを搭載しており、PubMed*に収載されている3700万件を超える医学論文を学習させることで、重要トピックの抽出、重要度のランク付け、重要トピック間の関連性等を時系列で可視化することが可能。これにより、これまで人による読み取りでは困難であった膨大な医学論文全体の重要トピックを俯瞰的に捉え、過去から現在に至るまでの研究トピックを迅速に把握し、臨床上の課題の見落としを防ぐことが可能になっている。また、埋もれている可能性があるアンメットニーズを見つけ出すことも期待できる。
    - *注）：PubMedとは、世界の主要な医学系雑誌に掲載された学術論文の書誌情報を調べることができる無料の医学関連分野の文献データベース。
- 加えて、本システムは生成AI機能を活用して、トピックに関連する複数の医学論文をまとめた日本語要旨を瞬時に生成することが可能。これにより、膨大な医学情報を効率的に把握することが実現。
- 本システムで得られた結果を基に診療上の課題を特定し、メディカル戦略の計画立案に利用することによりメディカルアフェアーズ担当者個人の知識や経験などに依存したプロジェクト間の隔たりを是正し、的確かつ迅速な戦略立案が可能になることを期待・

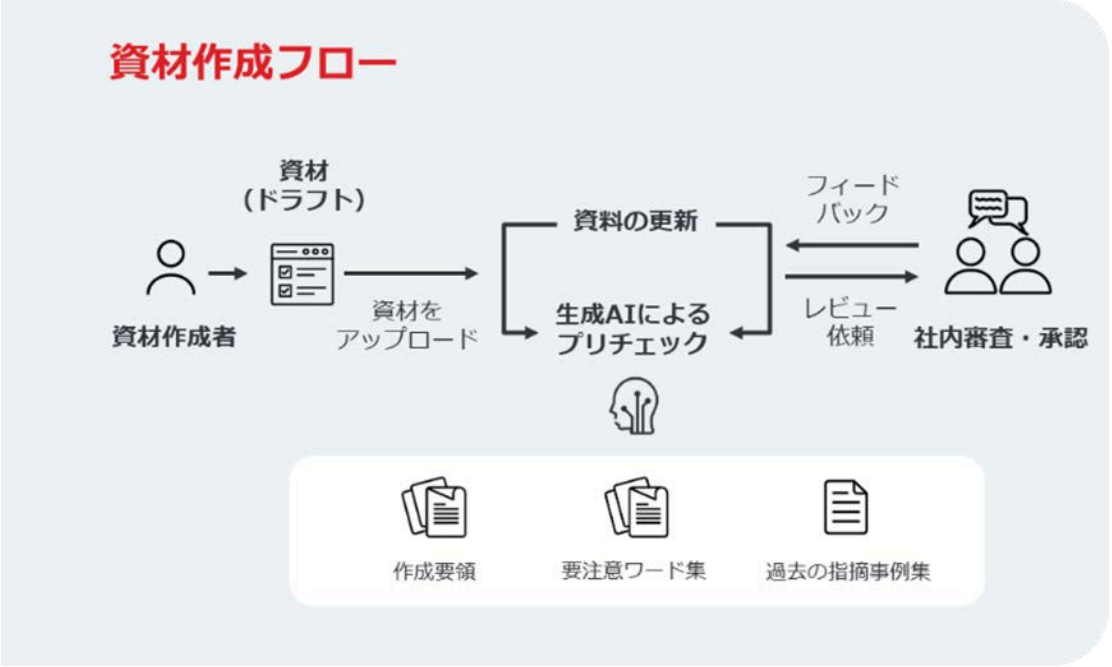
※MaTCHシステムの参考画面



（出典：小野薬品工業株式会社提供資料）

事例6　資材作成業務の効率化（武田薬品工業株式会社）

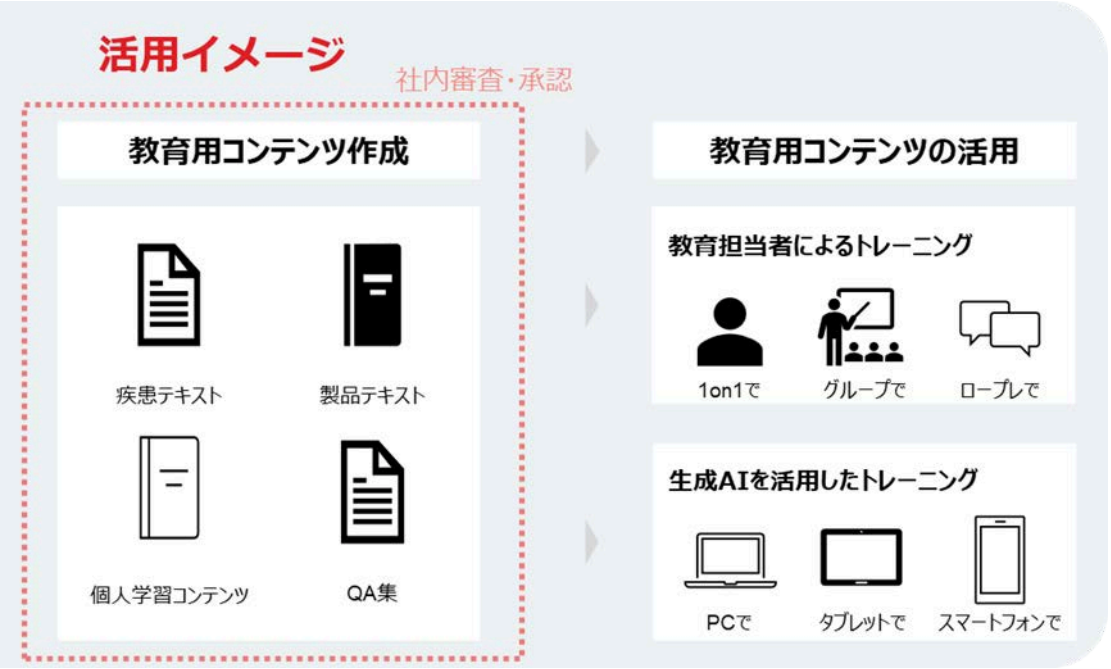
- 製薬会社の資材作成（例：製品情報概要等）には厳格な規制があり、それらを遵守するためには高度な専門知識が必要・
- その専門知識および関連する各種法規制に準拠した資料を作成する能力を習得するには、年数を要する。
- 資材作成に当たっての社内審査・承認のプロセスには複数の段階があり、月間500 以上の資材を作成するため、資材の社内審査・承認作成完了までに多くの時間を要している状況。
- 生成AIを活用した資材の校正により、業務効率を高め、社内審査・承認完了までの時間の迅速化を目指す
- 最後に、人による確認が行われる。



（出典：武田薬品工業株式会社提供資料）

事例7　生成**AI**を活用した**MR**：医薬情報担当者教育（武田薬品工業株式会社）

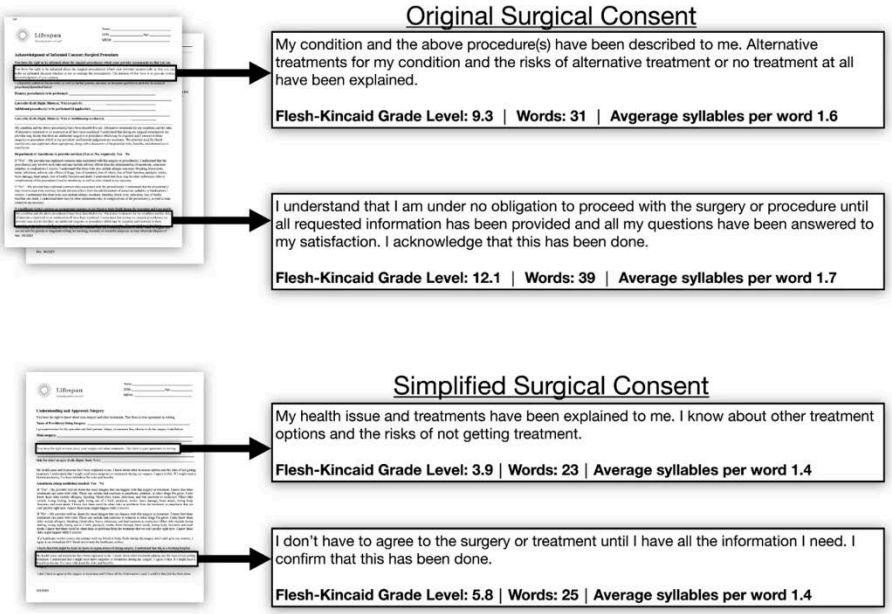
- 従来、MRの教育は、社内審査・承認済（著作権の許諾要否の確認含む）の教育資材をベースに、教育担当者による教育や、MR個々の自主学習によって行われてきた。
- これに加え、生成AIを活用したアウトプットトレーニング機能を提供することで、MRがいつでも・どこでも・何度でも、継続的に学習に取り組めるような環境を整備していく（ハルシネーションによりMRが間違った知識を習得しないための対策も実施中）
- 将来的なMRトレーニングにおいて、「利便性の向上と機会の増加」、「学習の選択肢の拡大」、「学習の実施状況と評価の可視化」、および「準備と運用プロセスの効率化」などのメリットが期待される。



（出典：武田薬品工業株式会社提供資料）

事例8　**GPT-4**によるヘルスリテラシーと患者転帰の改善（米国：**Lifespan**）

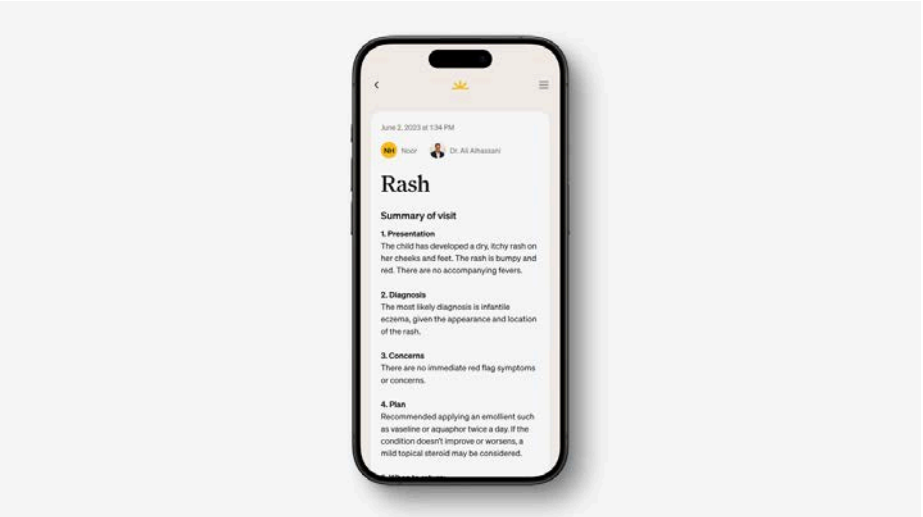
- 米国ロードアイランド州最大の医療機関であるLifespanは、患者のヘルスリテラシーを向上させ、治療結果を最適化するためにGPT-4を活用しています。具体的には、医療同意書の単語数を25%削減し、3ページから1ページに短縮することで、患者の理解度が向上し、安心感の向上に貢献している。
- また、患者は情報を簡潔に理解できるようになり、治療への積極的な参加を促進することで、治療効果や患者満足度の向上が期待されている。



（出典：https://openai.com/ja-JP/index/lifespan/）

事例9　**OpenAI**を活用した新たな小児医療の実現（米国：**Summer Health**）

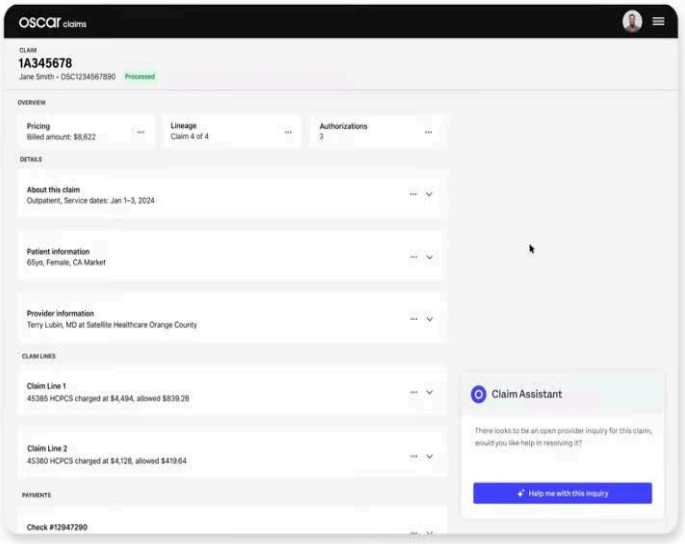
- 24時間365日テキストメッセージを通じて小児医療を提供する米国のスタートアップ企業であるSummer Healthは、GPT-4を活用して小児医療の診療記録作成を効率化し、記録時間を1件につき10分から2分に短縮した。これにより、診療記録の遅延が400%削減され、保護者からの満足度向上にも寄与している。
- また、迅速かつ正確な記録作成は、医療提供者と保護者間の円滑な情報共有を実現し、患者ケアの質向上に貢献している。
- さらに、生成AIを使って小児科医の診療記録作成を効率化し、保護者にも分かりやすい形に要約することで、医療従事者の負担軽減と患者サービスの向上を目指している。



（出典：https://openai.com/ja-JP/index/summer-health/）

事例**10**　医療保険における保険金請求処理コスト削減と患者ケア向上（**Oscar**）

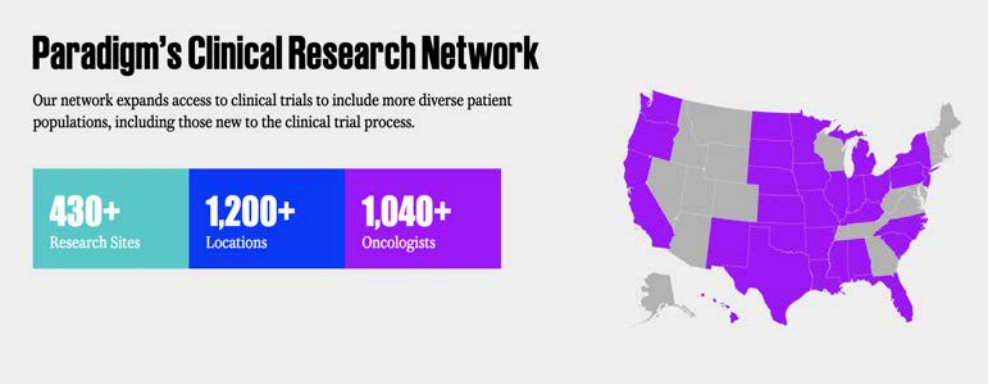
- 2012年に設立された米国のテクノロジー主導型の健康保険会社であるOscarは、GPT-4を導入して保険金請求処理の効率化を図り、処理時間を約40%短縮、業務生産性を2倍に向上を実現。また、月に4,000件以上の保険関連の調査が自動化され、年間で48,000件の業務負荷軽減が見込まれている。
- 従来20分以上かかっていた記録作成時間についてはGPT-4を利用したケースで最大90%の生産性向上も見られることが確認されており、これにより、医療従事者はより重要な業務に集中できるようになり、燃え尽き症候群の緩和にも繋がっている。
- また、患者にはパーソナライズされた健康情報や予防ケアが提供され、顧客満足度の向上と医療コスト削減を両立している。



（出典：https://openai.com/ja-JP/index/oscar/）

事例**11**　**OpenAI**の**API**で臨床試験への患者アクセスを改善（米国：**Paradigm**）

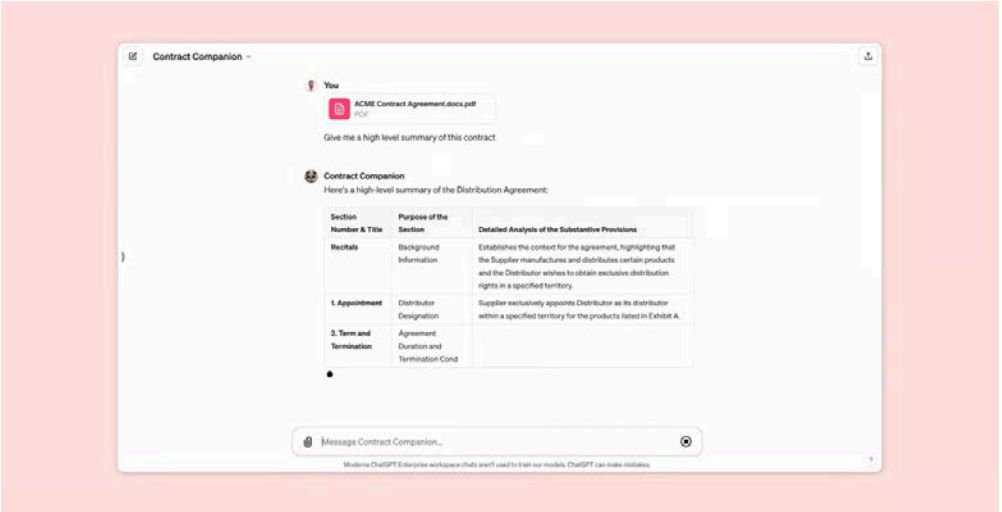
- 臨床試験への患者アクセスを改善するために設立された米国のヘルステック企業であるParadigmは、GPT-4を活用して臨床試験データの評価精度を10%向上させ、専門医の確認にかかる時間を90%短縮。数百人の患者を1分で評価する処理速度により、患者が迅速に試験にアクセスできる体制を整備している。
- 同社のサービスは非構造化データの分析に優れているため、従来の方法では見落とされてきた患者を臨床試験の被験者候補としてより正確に選別できる可能性がある。
- これは、十分な治療を受けていない患者にも公平に臨床試験の機会を提供することに繋がり、医療の進歩に貢献すると期待され、製薬企業や研究機関は試験の効率化とリクルートメントの加速を実現し、患者の新薬治療へのアクセス拡大に貢献している。



（出典：https://www.paradigm.inc/solutions/providers、https://openai.com/index/paradigm/）

事例**12**　救命治療の開発を加速するための提携（モデルナ）

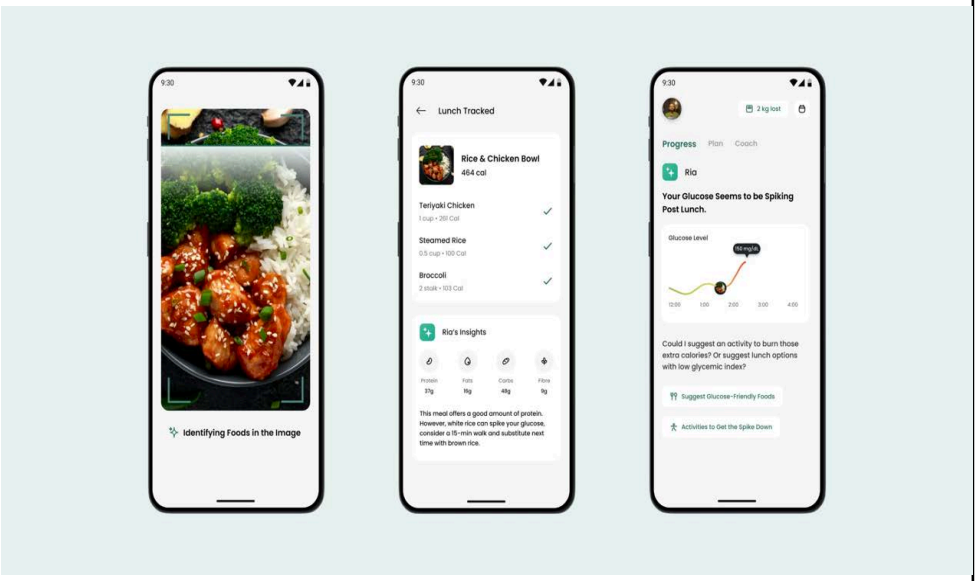
- 米国に本社を置くバイオテクノロジー企業であるモデルナは、GPT-4を活用し、750種類のカスタムGPTモデルを導入することでmRNA医薬品の研究開発を加速している。AIの活用により、データ解析や業務プロセスの効率化が進み、新薬開発期間の短縮とコスト削減が実現されている。
- また、安全かつ効率的なデータ解析体制を構築し、患者に早期に治療法を提供できるよう取り組んでいる。
- さらに、モデルナの法務チームはChatGPT Enterpriseを導入し、Contract Companion GPTは契約書の要約を、Policy Bot GPTは社内ポリシーに関する質問への回答を、それぞれ迅速に提供。これにより、法務チームはより重要な業務に集中できるようになり、従業員も必要な情報に容易にアクセスできるようになった。



（出典：https://openai.com/ja-JP/index/moderna/、https://openai.com/ja-JP/index/moderna/）

事例**13**　持続可能な体重管理で生活を改善（**Healthify**）

- インド最大のヘルスケアプラットフォームHealthifyは、OpenAIのGPT-4を活用し、画像認識機能を強化し、食事の記録が50%増加。ユーザーのエンゲージメントが向上し、体重減少や脂肪減少の効果が期待されている。
- AIアシスタント「Ria」を通じて、ユーザーとの対話が活発化し、メッセージ数が倍増しました。これにより、ユーザーの健康状態の改善が促進されている。
- また、コーチがクライアントに対応する時間が半減し、AIコーチの使用率が18%増加しました。ユーザーの健康状態の改善と持続可能な体重管理を実現している。



（出典：https://openai.com/ja-JP/index/healthify/）

別添3：医療機関内生成AI活用ポリシーひな型（案）

生成AIの利用ガイドライン

【医療機関名】

第※版

【20※※年※月※日】制定

【20※※年※月※日】改定

<前文>

本ガイドラインは病院内で生成AIを利用する場合に組織内のガイドライン・ポリシーとして決めておくと良いと思われる事項を参考情報としてまとめたものです。利用する場合はあくまでもドラフトとして扱い、各組織の既存のポリシー等と矛盾しないよう加筆修正を行う必要があります。

0.はじめに

0-1.本ガイドラインの目的

生成AIは、業務効率化や生産性向上に資する技術である一方、入力データの内容や生成物の利用方法によっては法令等違反や他者の権利侵害につながる可能性があります。そこで、院内で生成AIツールを使用する職員が安全かつ安心に活用できる環境を整備することを目的として、本ガイドラインで生成AIの特性や使用上の注意点をまとめることとしました。院内で生成AIを活用する職員は本ガイドラインをよく読んだうえで生成AIを利用してください。

0-2.本ガイドラインの対象範囲・対象者

本ガイドラインが対象とする生成AIは、院内で導入されている生成AIサービス全てが該当します。また、本ガイドラインの対象者としては当該生成AIサービスを利用する院内職員が該当します。

1.生成AIについて

1-1.生成AIの概要

生成AIは、プロンプトという指示文を与えると文章や情報を生成することができる人工知能（AI）技術の一種です。テキストだけではなく、画像、音声といった幅広い形のデータを取り扱うことができます。適切に活用することにより業務・作業をサポートするツールとして業務の効率化が期待できます。

1-2.生成AIのユースケース

生成AIは例えば以下のユースケースで有効な活用が期待できます。

＜文章＞

（例）退院サマリの作成 等

1-3.生成AIの特性と活用時のリスク

生成AIは非常に便利で有用なツールですが、様々な特性を持つため、適切な使用方法をしないと情報漏洩や業務ミスなどを始めとしたリスクが存在します。

1-3-1.正確性・公平性

生成AIの学習データに正誤が確認されていない情報や古い情報が含まれている可能性があるため、生成AIが出力する内容は正しいものとは限りません。また、生成AIは膨大な学習データをもとに、特定の単語が入力された場合、次にどんな単語が続く可能性が高いか「確率予測」を繰り返して文章を生成するので、正しい結果を生成するとは限りません。また、生成AIが出力した内容には学習データの偏りやアルゴリズムにより、公平性に欠ける内容が含まれることがあります。例えば、画像生成において「動物」を含んだプロンプトでは「犬」が出力されやすいというように結果に偏りがある場合があります。

これらの特性から、生成AI活用時には正確性や公平性の欠如により、事実とは異なる結果や不公平な結果が生成される可能性があります。

1-3-2.著作権の侵害

生成AIにより出力された内容が既存の著作物と同一・類似しており、当該内容が既存の著作物に基づいて出力された場合、出力された内容を利用すると著作権侵害になる可能性があります。

1-3-3.情報管理

生成AIツールによっては、入力された情報が生成AIの学習に利用され、その結果として別の利用者が生成AIツールを活用する際に当該情報が流用される可能性があります。このため、院内の部外秘情報や個人情報をみだりに生成AIに入力することで意図せず生成AIの提供元や他の生成AI利用者へ情報漏洩を引き起こす可能性があります。

2.生成AI利用上のルール

2-1.院内における利用可能な生成AIツール

院内においては、以下の生成AIツールのみ使用できます。

- 【生成AIサービス名】
  - 【ツールの説明】

セキュリティや安全性の観点から、院内で許可されたツール以外は利用できません。もし、利用希望がある場合には【問い合わせ先】までお問い合わせください。

2-2.生成AI利用上のルール

1-3.で記載したとおり、生成AIの特性とリスクを踏まえ、以下の点について注意したうえで生成AIを利用してください。

2-2-1.生成AIで出力した内容は必ず確認すること

生成AIで出力した内容は正確性・公平性に欠ける場合があります。そのため、出力された内容をそのまま利用することなく、必ず自分の目で確認を行い、もし誤った内容や偏見・バイアスが含まれている場合は出力された内容を修正してから利用してください。

2-2-2.生成AIで出力した内容を利用するときは著作権侵害に注意すること

生成AIで出力した内容を配信・配布する場合は、既存の著作物と同一・類似でないこと及び当該内容が既存の著作物に基づいて出力されたものではないことを必ず確認してから、利用してください。

また、プロンプトで、「（既存の著作物の名称）と似せてください」といった指示を出す等、複製や改編が許可されていない著作物についてプロンプトで使用することは控えてください。

2-2-3.入力する情報に注意すること

生成AIツールにおける学習機能をOFFに設定するなど、個人情報や部外秘情報がプロンプトとして入力された場合に学習されないように設定してください。

個人情報や部外秘情報を認められた範囲外で入力することがないように注意してください。また、生成AIツールごとに個別の操作方法の説明に加え、利用ルールが設定されています。例えば、入力しても良いデータの範囲が設定されています。ルールを守った利用をすることで、情報漏洩といったセキュリティ事故を発生させないようにしましょう。

別添4　本ガイドの検討経緯及び参加・協力企業等

＜第**1.0**版策定時＞

1．検討経緯

本ガイドは、2023年3月に設立された日本デジタルヘルスアライアンス（以下「JaDHA」という。）におけるWG4（デジタルヘルスアプリの適切な選択と利活用を促す社会システム創造WG）内に設置しているSubWG-B（生成AIに関する検討）において作成・検討が進められた。具体的な検討経緯は以下のとおり。

	アジェンダ	登壇企業
2023年 7/31	・生成AIに関する最新動向のインプット① （ヘルスケア生成AIの基本的動向、特徴や限界等について）	・Ubie株式会社 ・株式会社MICIN
8/7	・生成AIに関する最新動向のインプット② （ LLMの選定方法や主体の区分整理、実装のポイント等について）	株式会社サイバーエージェント
8/23	・生成AIに関する最新動向のインプット③ （サービスへの具体的活用事例及び課題・リスクについて）	株式会社HACARUS
9/4	・生成AIに関する最新動向のインプット④ （サービスへの具体的活用事例及び課題・リスクについて）	株式会社HOKUTO
9/25	・生成AIに関する海外動向インプット ・論点整理素案について議論	―
10/10	・生成AIに関する最新動向インプット⑤ ・ガイド（案）意見交換	株式会社NTTデータ
10/23	・生成AIの最新動向に関する勉強会 （生成AI活用時における法的論点等について）	・JaDHA顧問弁護士 後藤 未来 様、アンダーソン・毛利・友常法律事務所　中崎 尚 様 ・一般社団法人 日本ディープラーニング協会（JDLA）「AIデータと個人情報保護」研究会 副座長　小峰 弘雅 さま（株式会社ベイカレントコンサルティング）、柴山 吉報 さま（阿部・井窪・片山法律事務所）
11/6	・生成AIの最新動向の共有 ・ガイド策定に向けた論点整理案についての議論	―
11/13～ 11/24	JaDHA会員向け意見募集	―
12/4	SubWG-B 最終とりまとめ	―
12/15	WG4 最終審議	―

2. 検討主体

前述のとおり、本ガイドはJaDHAにおけるWG4 SubWG-Bの参加企業において作成・検討を進めたもの。作成当時のSubWG-Bメンバーは以下のとおり。

SubWG-B 参加企業（五十音順）	担当者（敬称略）	ガイド作成担当領域
株式会社Welby	井伊 尋幸	案執筆、レビュー
小野薬品工業株式会社	小林 正克、片山 英夫	案執筆、レビュー
株式会社サイバーエージェント	窪田 海人	主執筆
シスメックス株式会社	辰巳 真一	レビュー
シミックホールディングス株式会社	三友 周太、齊藤 卓弥、笹 澄絵	案執筆、レビュー
株式会社Save Medical	小林 亮太	レビュー
タウンドクター株式会社	山上 慶	レビュー
株式会社テックドクター	湊 和修	レビュー
株式会社MICIN	碇﨑 裕晃、浅原 弘明	主執筆、レビュー、「生成AIにおけるバリューチェーン」図表提供
Ubie株式会社（SubWG-B リーダー企業）	三浦 萌、島津 真夢	主執筆、全体統括
株式会社日本総合研究所（JaDHA事務局）	南雲 俊一郎、城岡 秀彦、川舟 広徒	案執筆

3. 協力・登壇企業等

その他、本ガイド作成に当たっては以下の皆様からアドバイス・ご意見をいただきました。末筆ではありますが御礼申し上げます。

- 株式会社HACARUS 様
- 株式会社HOKUTO 様
- 株式会社NTTデータ 様
- JaDHA顧問弁護士 / アンダーソン・毛利・友常法律事務所（AMT） 後藤 未来 様
- アンダーソン・毛利・友常法律事務所（AMT）中崎 尚様
- 一般社団法人日本ディープラーニング協会「AIデータと個人情報保護」研究会 副座長/株式会社ベイカレントコンサルティング 小峰 弘雅さま
- 一般社団法人日本ディープラーニング協会「AIデータと個人情報保護」研究会 副座長/ 阿部・井窪・片山法律事務所 柴山 吉報 様
- JaDHA会員の企業の皆様

＜第**2.0**版策定時＞

1. 検討経緯

本ガイドは、2023年3月に設立された日本デジタルヘルスアライアンス（以下「JaDHA」という。）におけるWG4（デジタルヘルスアプリの適切な選択と利活用を促す社会システム創造WG）内に設置しているSubWG-B（生成AIに関する検討）において作成・検討が進められた。具体的な検討経緯は以下のとおり。

	アジェンダ	登壇企業
2024年 1/23	・アップデート議論について	―
2/5	・AI事業者ガイドライン案の提出意見について	―
2/29	・AI事業者ガイドライン案との連携について	―
2/29	・医療分野におけるLLMの現状と応用可能性について～アカデミア／病院の立場からのLLMへの期待と課題～	大阪大学大学院医学系研究科 医療情報学　特任助教　杉本賢人様
4/22	・医療機関向け生成AIプロダクトの紹介	Ubie株式会社
5/20	・生成AIの海外動向調査	株式会社日本総合研究所
6/17	・メディカルアフェアーズにおけるAI活用事例の紹介	小野薬品工業株式会社
8/19	・「生成AIの医療応用に際してのELSII(法的倫理社会的課題)」 ・SLMの今後の展開について	・株式会社MICIN ・JaDHA特別顧問
9/6	・がん副作用問診 × LLM(大規模言語モデル)の可能性 ~irAE問診~ ・生成AIを軸とした社内業務プロセス改革 ・海外における医療分野における生成AI利事例	・TXP Medical株式会社　佐藤雅和様・大角 知也様 ・武田薬品工業株式会社 ・JaDHA特別顧問
10/21	生成AIの規制緩和論点について（議論）	―
11/18	【議論】生成AI活用ガイド第2.0版(案)諮レビュー	―

2. 検討主体

前述のとおり、本ガイドはJaDHAにおけるWG4　SubWG-Bの参加企業において作成・検討を進めたもの。作成当時のSubWG-Bメンバーは以下のとおり。

SubWG-B 参加企業（五十音順）	担当者（敬称略）	ガイド作成担当領域
味の素株式会社	村松　孝彦	案執筆、レビュー
株式会社Welby	井伊　尋幸	案執筆、レビュー
小野薬品工業株式会社	小林　正克、結城　亮介	案執筆、レビュー
オムロンヘルスケア株式会社	鹿妻 洋之様	レビュー

シミックホールディングス株式会社	三友　周太、齊藤　卓弥	主執筆、レビュー
株式会社Save Medical	田村　綾子	レビュー
武田薬品工業株式会社	江口　知元	案執筆
株式会社テックドクター	湊　和修、近藤　洋司	案執筆、レビュー
株式会社MICIN	藤田　卓仙	主執筆、レビュー、 「生成AIにおけるバリューチェーン」図表提供
JaDHA特別顧問	碇﨑　裕晃	主執筆、レビュー 「生成AIにおけるバリューチェーン」図表提供
Ubie株式会社 （SubWG-B リーダー企業）	井上　真夢	主執筆、全体統括
株式会社日本総合研究所 （JaDHA事務局）	南雲　俊一郎、城岡　秀彦、川舟　広徒	案執筆

3. 協力・登壇企業等

その他、本ガイド作成に当たっては以下の皆様からアドバイス・ご意見をいただきました。末筆ではありますが御礼申し上げます。

- 大阪大学大学院医学系研究科 医療情報学　特任助教　杉本賢人様
- TXP Medical株式会社　佐藤雅和様・大角 知也様
- JaDHA会員の企業の皆様