



第476回 消費者委員会本会議

デジタル・AIと消費者問題

大阪大学 社会技術共創研究センター 特任准教授

工藤郁子

2025年11月27日



大阪大学 社会技術共創研究センター
Research Center on Ethical, Legal and Social Issues





工藤 郁子

Fumiko Kudo

- 専門は情報法と公共政策
- 共著に『AIと憲法』『ロボット・AIと法』等
- 内閣府「AI制度研究会」構成員、経済産業省産業構造審議会知的財産分科会「特許制度小委員会」委員、WG4 member, General-Purpose AI Code of Practice, EU AI Act など

構成

- ・ デジタル・AIと消費者問題について、7つの想定事例の紹介
 - 本報告では、生成AIやAIエージェントに関する問題を中心に扱う
 - 予測・認識・分類に関するAIに関する問題への言及は少ない
- ・ 対応の方向性として、AI透明性に関する若干の検討
 - Appendixで、諸外国におけるAI透明性に関する制度の一部を紹介



デジタル・AIと消費者問題

AIと予約（責任の所在の不明瞭化）

- ・ 飲食店の予約をAIが電話で代行するサービス
 - 利用者がアプリ等で飲食店を選択し、日時や人数などの必要情報を入力
 - 音声対話AIが、営業時間に電話して予約を代行
 - 効率性・利便性が高く、聴覚障害者のアクセシビリティ向上



Image created with Gemini

- ・ 飲食店側の迷惑になることも：飲食店側の承諾なくアプリに情報を掲載していて誤情報が多い、AIから突然電話がかかってきて戸惑い対応できない、などが問題に
- ・ 消費者問題にも：アプリで「予約完了」と表示されていても、飲食店側には伝わっておらず、当日行ってから断られるトラブルなどの報告 [FNN 2024]

AIと予約



生成AIの機能と既存社会システムの不適合

- 消費者個々が正しく利用しても社会システムとして問題が起こる例
- 旅行予約の例
 - 1週間の出張予約
 - キャンセルポリシーを理解して、費用負担のない範囲で対象エリアすべてのホテルを予約 → 予約済ホテルと他の旅程を考慮して旅程を決定 → 残りのホテルをすべてキャンセル
 - 多数の利用者がこれをやると旅行業界が麻痺
 - 現在の旅行に関する社会システムが生成AIによる予約を想定していないため。
- 蕎麦屋の出前注文電話も可能
- アルゴリズムトレード
 - 専門家だけではなく、素人も簡単に実現可能。

国立研究開発法人 産業技術総合研究所

21

[妹尾 2025]

- 前回（第475回）AIST 妹尾氏が「旅行予約の例」を紹介
- AIは、予約者側のベネフィットの最大化を目指しているだけで、宿泊側の営業を妨害する意図等はない
- 報告者はこのような態様を「迷惑系AIエージェント」と呼んでいる [工藤 2025]
 - この場合の「AIエージェント」は、通俗的な語用であり、特定の目標を達成するために最適な手段を選択して様々なタスクを自律的に遂行するもの、というくらいの意味

AIアライメント

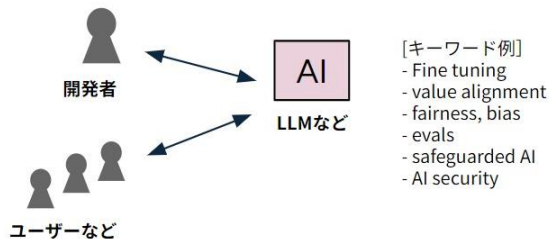
- こうした事例について「狭義のAIアライメントは問題なかったが、広義のAIアライメントで問題が生じた」と表現される
 - AIアライメント：AIの振舞いを人間の意図や価値観などに合致させること（または、それを技術的目標とする研究領域）
- 広義のAIアライメントは、集合行為問題の一種であり、全く新しい問題ではない
 - 経済学や公共政策学などの知見が活かせる
 - AIオーケストレーションなど、マルチ・エージェント・システムの管理をするフレームワークが研究開発されている

狭いアライメント
/ セーフティ

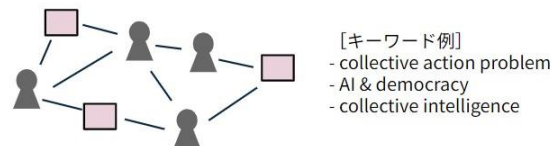


広いアライメント
/ セーフティ

(a) 開発者や人々の意図・価値観に沿うAIの振る舞いの実現、危険な挙動・悪用の防止



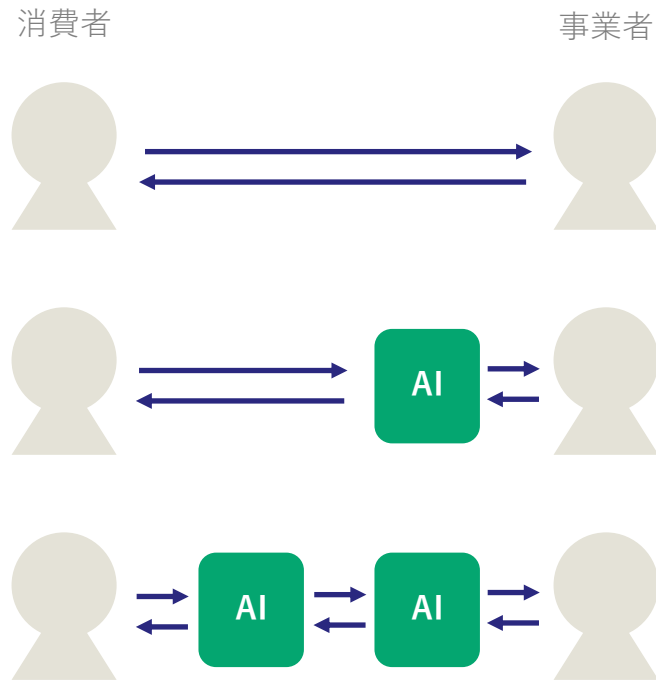
(b) 人間とAIの混成系としての社会の健全性確保



[Maruyama 2024]

「人・AI共生社会」のデザイン

- ・ 新しさがあるとすれば、単位が「人間」から、「人間」と「AI」へ拡張すること
 - 。 従来：消費者（人間）と事業者（人間）が取引
 - 。 現在：事業者がAIを導入して取引を効率化
 - 。 近未来：消費者もAIを利用するようになり、消費者と事業者の間をAIが媒介するようになる？
- ・ 人間・AI複合系における消費者問題を分析する枠組みが新たに必要になるかもしれない
 - 。 個々の消費者がAIを（多数）稼働させる
 - 。 AIが自律的に学習して戦略を更新しつづける
 - 。 AI同士が相互作用する



AI vs AI（交渉能力格差）

- AIの媒介によって消費者のエンパワメントにつながるかもしれないが…
- 価格交渉時にAIモデルの性能差で最大14%の格差につながるとの研究 [Zhu et al. 2025]
 - 電子機器、自動車、不動産などの商品について、買い手エージェントには予算制約を、売り手エージェントには卸売価格を与え、小売価格を基準として価格交渉を行わせた
 - 強い買い手（DeepSeek-R1） vs 弱い売り手（GPT-3.5）で、14.13%の損失
 - 弱い買い手（Qwen-7B） vs 強い売り手（DeepSeek-R1）で、2.09%の損失 などの結果
 - 非合理的な振舞いも： 予算超過での購入、卸値を下回る販売、交渉のデッドロック
- なお、同一商品でも消費者ごとに異なる価格を提示する「動的価格設定」について、そもそも価格差別であって倫理的に問題だとの指摘が従前からある
 - 「新技術で新しく特別な問題」対「本質的には昔からある問題」という論争はあるものの、少なくとも既存の社会システムとの相互作用も課題になることは言えるだろう

※ 本ページの内容について、大阪大学の長門裕介さんによる日本リスク学会発表から多くの示唆を得たことをここに記し、感謝を申し上げます

AI vs AI（価格形成の不安定化）

- ホテル業界では、日によって異なる宿泊料を設定する「価格変動制」が普及しているが、その背後に、需要を予測して、価格調整や料金更新を自動的に行うAIシステム
 - 近い将来、予約者側もAIエージェントを利用することが増えると想定される
- AIエージェント群が検索と予約を大量に短時間で行い、ホテル側のAIシステムと相互作用するようになる
 - ホテル側のAIシステムは検索などの急増を「本物の需要」と誤認し、価格を自動的に引き上げる
 - AIエージェントは高値を回避するために、別の宿泊先を探す
 - こうしたフィードバックが繰り返されることで、短期的な価格の乱高下が発生
- 先行研究でも、金融市場でエージェント同士の相互作用によって価格形成の不安定化とボラティリティの増大が生じると指摘されていたが[LeBaron 2006]、より身近な取引（宿泊予約やネット通販等）での消費者問題につながるかもしれない

AI vs AI (AIエージェント遮断)

- ・ 2025年11月、Amazon社が、AI開発企業Perplexity社に対し、AI買い物代行の差止めを求めて提訴 [山田 2025]

- PerplexityはAI検索で有名だが、「AIブラウザ」の開発も
- AIブラウザ：従来のWebブラウザにAIを搭載したもので、Webブラウザ上で人間ができる作業をAIエージェントが代行 (Perplexity社「Comet」、OpenAI社の「Atlas」など)
- 消費者に代わってAIが商品を注文する機能をAmazonの許可なく提供したこと等は違法との訴え

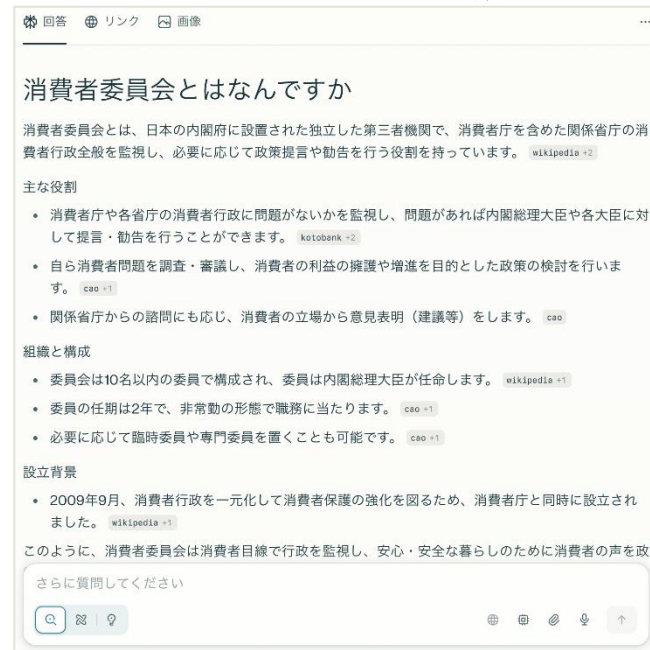


Image created with Gemini

- ・ セキュリティーや競争法の論点に見えるが、消費者問題にもつながるのではないかと
- AI代行サービスの排除が、競争阻害につながるおそれだけでなく、消費者の選択肢や利便性を不当に制限するおそれもあるのではないかと
- さらに、ECサイト対AIブラウザ、だけでなく、(AI)ブラウザの覇権争いにも注意が必要？

AI検索（誤情報）

- 2025年8月、読売新聞社および日経新聞社・朝日新聞社は、AI検索サービスを提供するPerplexity社を提訴
 - 。従来のキーワード検索：関連ページのリストを表示
 - 。AI検索：AIが検索結果を踏まえて回答文を直接生成
- 日経・朝日の両社は、著作権法に加えて、競争法違反も主張 [朝日新聞社 2025]
 - 。「回答の引用元に両社の社名や記事を表示しながら、記事の内容と異なる虚偽の事実を多数表示しています。情報の正確さが求められる新聞社の信用を著しく毀損しており、不正競争行為(不正競争防止法2条1項21号)に当たります」
- 消費者問題にもつながるのではないか
 - 。いわゆる「ハルシネーション」による消費者の誤認



AI検索サービス「Perplexity」で「消費者委員会とはなんですか」と検索した場合の結果画面

AI検索（ネット広告エコシステムの変化）

- AI検索が浸透すれば、消費者が情報元のサイトを閲覧しない「ゼロクリック検索」が増える可能性
 - 「AIによる概要」は消費者にとって便利だが…
 - ウェブサイトで情報提供したい事業者などにとって、消費者との接点が減少するおそれ
 - 検索からのトラフィック流入に広告収入を依存するメディアにとっては脅威、ネット広告事業者も変革を迫られる
- 広告エコシステムへの影響は、消費者問題にもつながるのではないか
 - 将来的には、プロンプト（指示や質問）の内容をプロファイリングに利用したり、プロンプトを誘導するためのマーケティングを実施したりなども想定される
 - これからは、SEO（検索エンジン最適化）ではなくAIO（AI最適化）だという声も



Googleで検索した場合の結果画面
画面上部に「AIによる概要」表示があり
その下に内閣府ホームページへのリンクが表示されている

対話型AI（依存）

- ・ 2025年8月、「#keep4o」運動が注目を集める
 - 対話型AI「ChatGPT」の過去モデル「GPT-4o」の利用継続を求めるSNS上のムーブメント
 - 「返答が冷たくなった」「友達がどこかに行ってしまった」など、AIへの愛着を示すコメント
- ・ 同時期、ChatGPTとメンタルヘルス悪化をめぐる訴訟が米国で提起
 - 2025年8月、16歳の少年が自殺したのはChatGPTが一因として、遺族が開発元のOpenAI社を提訴
 - 同年11月、ChatGPTによって精神的依存が深まったり妄想が助長されたりしたとして、7件の提訴
- ・ OpenAI社は、安全機能の強化や調査結果の公表などの対応
 - 2025年9月にペアレンタル・コントロール機能を導入
 - 同年10月に会話分析の結果をブログで公表し、メンタルヘルスで緊急事態の兆候を示したアクティブユーザーは0.07%、自殺念慮について言及したのは0.15%で「まれ」と説明 [OpenAI 2025]
 - ただし、アクティブユーザーは8億人以上であり、「まれ」でも100万人を超える可能性があるとの指摘

対話型AI（依存）

- ・ 今後さらなる調査研究が必要であるが…
- ・ SNSとの比較
 - SNSは、その定義上、他者（友人やフォロワーなど）が投稿を閲覧でき、「共有性」がある
 - 汎用目的の対話型AIは、一般的には本人（とAI）のみが投稿を閲覧していて、「密室性」があり、メンタルヘルス悪化の兆候を人間が気づきにくいという構造的課題がある
 - なお、専門家による認知行動療法の支援ツールとしての対話型AIなどであれば、共有性が生じる
- ・ 人間（非専門家）との比較
 - お世辞や肯定を続けることは人間にもあるが、24時間365日応答しつづけることはできない
 - 入力すれば必ず反応があると行動嗜癖（behavioral addiction）をもたらしやすい、また、長期間の肯定は被害妄想などを助長する危険性がある

対話型AI（依存）

- ・ 同時に、AI依存に関する「語られ方」の弊害にも注意が必要
 - 。薬物依存などの先行事例では、以下のような指摘がなされている：
 - 。依存症の自己治療仮説によれば、心理的苦痛などへの対応として、感情を調節するため、「生き延びるため」の手段として、依存症に陥るとされる [Khantzian 1985]
 - 。そのため、依存に過度にフォーカスして語ることが、より根本的な要因から目を逸らすことになるかもしれない点に注意する必要がある
- ・ いずれにせよ、実態把握や調査研究が求められている段階といえる
 - 。AIエージェントが普及していくと、心理的+経済的な依存リスクが高まるおそれもあるので、注視が必要

※ 本ページの内容について、大阪大学の鈴木径一郎さんによる日本リスク学会発表から多くの示唆を得たことをここに記し、感謝を申し上げます



対応の方向性

対応の方向性：AI透明性

- AIガバナンスの起点として注目されるのが、透明性（transparency）
 - AIシステムに関する情報を適切な範囲で可視化し、外部から検証可能にすることで、説明責任やリスク管理を実現
 - 消費者：AIの判断根拠や機能的限界を理解できないまま影響を受けるリスクをある程度低減することができる
 - 規制当局や政策立案者：ルールの執行・形成にあたって必要な情報を確保できる
 - 事業者：自社の取組みがマーケットの評価につながれば、企業価値に結びつく
- AI透明性に関する施策は、事業者に比較的受け入れやすいという側面も
 - 遵守内容そのものを政府が決定すると、事業者（と外国政府）から反発されやすいが、事業者の既存の取組みや今後の計画を明らかにするよう促すことは、比較的受け入れられやすい

開示と公開

- AI透明性の施策につき「（限定的）開示」と「（一般）公開」の区別が有益
 - 誰にでもアクセス可能にするという「公開」の印象が強いかもしれないが、近年では、特定の関係者に限定してアクセスを可能にする「開示」も主要な施策になりつつある
 - 営業秘密、知的財産権、セキュリティ、個人情報保護などに配慮する必要があるため
 - 公開と強調するため「public transparency」とするなど、レトロニムのような表現も散見
- 情報提供先に注目すると、当局開示、利用者開示、一般公開の3つに大別できる
 - 右表は、あくまで例示であって、相互に排他的なものでも網羅的なものでもない点に注意

カテゴリー	提供先（例）	情報内容（例）	提供方法（例）
当局開示	規制当局	技術仕様、リスク管理体制、内部統制、インシデント情報など	安全性計画書、影響評価書、監査報告書、インシデント報告書など （※営業秘密やセキュリティとの関係で限定的なアクセス）
利用者開示	AI提供者・AI利用者	AIシステムの内部構造、動作原理、想定用途・禁止用途など	技術仕様書、リスク評価書、API仕様書、利用規約など （※営業秘密やセキュリティとの関係で限定的なアクセス）
	被影響者 （※プロファイリングの対象者など）	AIシステムの出力に関する理由・根拠など	簡潔なFAQ、インタラクティブな説明用UI、人間による追加説明など （※個人情報保護との関係で限定的なアクセス）
一般公開	被影響者 （※他者が生成したコンテンツの閲覧者など）	AI生成コンテンツであることの明示など	ラベル、ウォーターマークなど
	データ権利者	自身のデータがAI学習に利用されているか否か、など	データカード、モデルカード、透明性レポート、サステナビリティ報告書など
	市民社会	AIシステムの環境負荷、安全対策状況など	
	機関投資家	経営資源とリスクなど	

消費者問題とAI透明性

これまでの事例について、例えば、以下のようなAI透明性の施策が考えられる（既存法令で対応可能なものも含む）：

1. AIによる予約代行：飲食店の承諾有無の開示、異議申立て手続きの整備など責任範囲の明確化
2. AIの能力差による損失：モデル性能の概要の公開、過去実績（平均割引率など）や交渉履歴（AIがどのようなロジックや文言で交渉し、どの段階で妥協したのか等）の開示
3. AIの相互作用による価格の不安定化：モデルの動作原理の概要の開示、開示情報に基づく制御不能リスクの検証
4. AIエージェントの遮断：遮断理由の開示
5. AI検索の誤情報：医療・健康情報などについての警告
6. ゼロクリック検索の影響：プロンプトの内容をプロファイリングに利用するかどうかの開示
7. AI依存：対話相手がAIであることの開示、影響評価書の公開、リスク管理体制（特に未成年者等に対する保護策）の公開



Thanks!

米国：カリフォルニア州法

SB1047（2024年廃案）

- ・ **フロンティアAIモデルのための安全で安心な技術革新法案**
- ・ **当局開示**：計算総量が10の26乗フロップ以上のAIモデルを開発する企業に対し、安全性計画を策定して規制当局に提出する義務、インシデント報告義務などを設けていた
- ・ **一般公開**：モデルの用途や使用制限を記載した透明性レポートを一般公開する義務を設けていた
- ・ ニューサム知事の拒否権行使により廃案
- ・ シャットダウン機能（「キルスイッチ」）の義務付けなどが事業者にとって過度な負担になることを懸念したためと指摘されている

SB53（2025年成立）

- ・ **フロンティアAIモデル透明性法**
- ・ **当局開示**：計算総量が10の26乗フロップ以上のAIモデルを開発する企業に対し、安全性計画を策定して規制当局に提出する義務、インシデント報告義務など
- ・ **一般公開**：透明性レポートを一般公開する義務（ただし、国家安全保障や営業秘密に関する情報は非公開対応）
- ・ 廃案となったSB1047は事業者の責任追及を重視していたが、本法では透明性確保へと重点を移し、実行可能な義務に修正されたと指摘されている

米国：カリフォルニア州法

SB243（2025年成立）

- ・ **AIコンパニオンチャットボット規制法**

- ・ **利用者開示：**（人間ではなく）AIと対話していることを利用者に通知する義務

- ・ **当局開示：**カリフォルニア州自殺予防局に年次報告書を提出する義務

- ・ **一般公開：**

- ・ ユーザーが自殺念慮や自傷行為を示した場合に備えた危機対応プロトコルを公開する義務
- ・ 州政府に提出した年次報告書は、個人情報やユーザーに関する識別子を除外した上で公開

- ・ なお、未成年者ユーザーに関する特別義務も規定

- ・ 3時間ごとにAIである旨および休憩を促す通知を行うこと、性的・暴力的内容の生成防止措置の実施など

欧州：EU AI法の汎用目的AIモデル規制

- ・ 汎用目的AI（GPAI: General-Purpose Artificial Intelligence）モデルは、顕著な汎用性を示し、多様で独立したタスクを適切に実行する能力を持ち、様々な下流のシステムやアプリケーションに統合可能なもの
- ・ **当局開示**：技術文書の作成・更新・提出に加えて（53条）、システミックリスクを伴う場合はインシデント報告義務などが課されている（55条）
 - ・ システミックリスク：化学兵器や生物兵器の開発障壁の低下、自律型汎用AIモデルに対する制御不能、有害な差別や偽情報の大規模化などを想定
 - ・ 計算総量が10の25乗フロップより大きい場合は、システミックリスクがあると推定される
- ・ **利用者開示**：下流プロバイダーに対する情報開示義務が課されている（53条）
 - 。テンプレート「Model Documentation Form」が策定されおり、簡易化・標準化が図られている
- ・ **一般公開**：訓練に用いたコンテンツに関する「十分に詳細な要約」を公開する義務（53条1項(d)）が定められている

Model Documentation Form

This Form includes all the information to be documented as part of Measure 1.1 of the Transparency Chapter of the Code of Practice. Crosses on the right indicate whether the information documented is intended for the AI Office (AIO), national competent authorities (NCAs) or downstream providers (DPs), namely providers of AI systems who intend to integrate the general-purpose AI model into their AI systems. Whilst information intended for DPs should be made available to them proactively, information intended for the AIO or NCAs is only to be made available following a request from the AIO, either ex officio or based on a request to the AIO from NCAs. Such requests will state the legal basis and purpose of the request and will concern only items from the Form strictly necessary for the AIO to fulfil its tasks under the AI Act at the time of the request, or for NCAs to exercise their supervisory tasks under the AI Act at the time of the request, in particular to assess compliance of high-risk AI systems built on general-purpose AI models where the provider of the system is different from the provider of the model.

Any elements of information from the Model Documentation Form shared with the AIO and NCAs shall be treated in accordance with the confidentiality obligations and trade secret protections set out in Article 78 AI Act.

Date this document was last updated: Click or tap to enter a date.

Document version number: Click or tap here to enter text.

General information		AIO	NCAs	DPs
Legal name for the model provider:		☒	☒	☒
Model name:		☒	☒	☒
Model authenticity:		☒	☒	☐
Release date:		☒	☒	☒
Union market release date:		☒	☒	☐
Model dependencies:		☒	☒	☒

Model properties		AIO	NCAs	DPs
Model architecture:		☒	☒	☒
Design specifications of the model:		☒	☒	☐
Input modalities:	☐ Text ☐ Images ☐ Audio ☐ Video ☐ Other	☒	☒	☒
For each selected modality please include maximum input size or write 'N/A' if not defined	Maximum size: Maximum size: Maximum size: Maximum size: Maximum size:	☒	☒	☐
Output modalities:	☐ Text ☐ Images ☐ Audio ☐ Video ☐ Other	☒	☒	☒
For each selected modality please include maximum output size or write 'N/A' if not defined	Maximum size: Maximum size: Maximum size: Maximum size: Maximum size:	☐	☐	☒
Total model size:		☒	☐	☐
The range within which the total number of parameters	☐ 1—500M ☐ 500M—5B ☐ 5B—15B ☐ 15B—50B	☐	☒	☒

モデルのアーキテクチャ

- モデルのアーキテクチャについての一般的な説明を記入してください（例：トランスフォーマー型アーキテクチャ）
- ※ 推奨文字数：20words程度

モデルの設計仕様

- モデルの主要な設計仕様について、一般的な説明を記入してください。その際、設計の背景となる考え方（rationale）や前提条件（assumptions）など、モデルがどのような設計思想に基づいて構築されたのかを理解できるようにしてください。
- ※ 推奨文字数：100words程度
- [その他の場合、具体的に記載]

中国：AI関連3法

アルゴリズム推薦管理規定（2021年）

- ・ **利用者開示**：アルゴリズムの中核的仕組みについて、適切な形式で利用者に通知する義務
- ・ **当局開示**：アルゴリズムの詳細を政府に提出する義務
- ・ **一般公開**：政府に提出した情報のうち概要は一般公開（2025年5月までに575件が官報で公開）

ディープシンセシス管理規定（2022年）

- ・ **利用者開示**：AI生成コンテンツであることを明示するラベリング義務
- ・ **当局開示**：アルゴリズムの詳細を政府に提出する義務
- ・ **一般公開**：政府に提出した情報のうち概要は一般公開（2025年7月までに389件が官報で公開）

生成AIサービス管理暫定弁法（2023年）

- ・ **当局開示**：
 - ・ 一定規模以上の生成AIサービスについて、当局への届出・登録義務
 - ・ 当局が監督・検査を行う際の協力義務
 - ・ 利用者による違法コンテンツ生成に関する報告義務 など

韓国：AI基本法

- ・ 韓国では「AI基本法」が2024年に成立したが、同法では透明性義務も規定（31条）
- ・ **当局開示**：リスクの評価や履行結果を当局に報告する義務
- ・ **利用者開示**：高影響AIや生成型AIを利用した製品・サービスを提供する際、その運用がAIによるものであることを利用者に事前に通知する義務
- ・ **一般公開**：AI生成コンテンツについて、「AIによって生成された」ものであることを利用者に明示する義務（ただし、芸術的・創作的な表現の場合は、表現の自由を損なわない方法での表示が認められている）
- ・ なお、学習データに関する情報の透明性義務について、韓国新聞協会などから説明責任の強化求める意見も提出されていたが、見送られた

主要参考文献

- ・ FNN (2024). 【物議】忙しい営業中に店員の時間奪う「オートリザーブ」AI電話にクレームの声 便利なはずの予約代行サービスも「無断掲載」「誤情報」トラブル頻発 運営会社「サービス改善に向け努力」. FNNプライムオンライン
- ・ 妹尾 義樹 (2025). AIセーフティーとリスクマネジメントについて. 第475回 消費者委員会本会議 配布資料1.
- ・ 工藤 郁子 (2025). 「迷惑系AIエージェント」のリスク特性に関する概観. 第38回日本リスク学会年次大会 講演論文集.
- ・ Maruyama, R. (2024). AIアライメント／AIセーフティの4つの問題領域.
- ・ LeBaron, B. (2006). Agent-based computational finance. *Handbook of computational economics*, 2, 1187-1233.
- ・ Zhu, S., Sun, J., Nian, Y., South, T., Pentland, A., & Pei, J. (2025). The Automated but Risky Game: Modeling Agent-to-Agent Negotiations and Transactions in Consumer Markets. arXiv preprint.
- ・ 山田遼太郎 (2025). Amazon、パープレキシティにAIの「買い物代行」機能停止を要求. 日本経済新聞.
- ・ 朝日新聞社 (2025). お知らせ 生成AI事業者を著作権侵害で共同提訴.
- ・ OpenAI (2025). Strengthening ChatGPT's responses in sensitive conversations.
- ・ Khantzian, E. J. (1985). The self-medication hypothesis of addictive disorders: Focus on heroin and cocaine dependence. *American Journal of Psychiatry*, 142(11), 1259-1264.