〔論　説〕

対話型 AI（チャットボット，スマートスピーカー（AI スピーカー），AI アシスタント等を含む）に関する法律問題

松　尾　剛　行

Ⅰ　はじめに

１　注目される対話型 AI

AI が「戦略的に取り組むべき基盤技術」と位置付けられ[1]，「人・産業・地域・政府全てに AI」と副題が付された AI 戦略 2019[2]が公表される中[3]，AI の一種である，対話型 AI の利用も増加している。

例えば，2019 年には約 1 億 5000 万台のスマートスピーカーが販売されたといった統計が発表されている[4]。とりわけ，対話型 AI は，高齢者や障がい者といった，キーボード等の従来型インターフェースの利用に一定以上のハードルがある人にとって，音声で簡単に操作できるというインタフェースが高く評価されている[5]。

その他，世界最先端デジタル国家創造宣言・官民データ活用推進基本計画」[6]においても，電子行政等の分野で対話型 AI の活用について検討を進めるとされ，次世代人工知能推進戦略でも，教育分野等での対話型 AI の活用が論じられている[7]。人工知能技術戦略実行計画[8]は，人と協調できる AI に関し，人間と対話し，学習する AI，ヒューマンインタラクション，知識構造化の自動化技術の開発として，2023 年度中に現場の従業員等が AI と対話しながら，熟練者が持つ暗黙知や社会の有する集合知を構造化し，AI と人間が連携して学習できる手法を確立

---

1)　「統合イノベーション戦略 2020」内閣府<https://www.kantei.go.jp/jp/singi/tougou-innovation/pdf/togo2020_honbun.pdf>（2020 年 9 月 10 日最終アクセス，以下同じ）。

2)　「AI 戦略 2019～人・産業・地域・政府全てに AI～」首相官邸統合イノベーション戦略推進会議<https://www.kantei.go.jp/jp/singi/ai_senryaku/pdf/aistratagy2019.pdf>。

3)　その後のフォローアップにつき「『AI 戦略 2019』フォローアップ」首相官邸イノベーション政策強化推進会議
<https://www.kantei.go.jp/jp/singi/ai_senryaku/pdf/aistrategy2019_fu_honbun.pdf>参照。

4)　「2019 年のスマートスピーカー出荷量は 70%アップの 1 億 4690 万台で新記録」Sarah Perez<https://jp.techcrunch.com/2020/02/18/2020-02-17-smart-speaker-sales-reached-new-record-of-146-9m-in-2019-up-70-from-2018/>。

5)　「AI ネットワーク社会推進会議　報告書 2020～『安心・安全で信頼性のある AI の社会実装』に向けて～」総務省情報通信政策研究所
<https://www.soumu.go.jp/main_content/000698163.pdf>。

6)　「世界最先端デジタル国家創造宣言・官民データ活用推進基本計画」内閣閣議決定
<http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20200715/siryou5.pdf>。

7)　「次世代人工知能推進戦略」総務省<https://www.soumu.go.jp/main_content/000428750.pdf>。

8)　「人工知能技術戦略実行計画」内閣府
<https://www8.cao.go.jp/cstp/tyousakai/jinkochino/keikaku.pdf>。

するとした。

近時の新型コロナウイルスの蔓延の中,「コロナと戦う」「ポストコロナ社会を構築する」ベンチャーの支援の観点から，一般社団法人日本ベンチャーキャピタル協会がまとめた「コロナと戦うベンチャーリスト」[9]でも，対話型 AI による営業や面接等，これまで AI をあまり用いていない分野への対話型 AI 導入事例が掲載されており，新型コロナウイルスの状況を踏まえたニュー・ノーマルにおいても，対話型 AI は欠かせない存在と認識されている[10]。

２　対話型 AI とは

このように注目を浴びる対話型 AI は，「ユーザの発話した言葉を理解し，適切に応答するシステム」[11],「人間が普段使う音声言語を入出力とする」[12]等といわれており，特定のタスク（たとえば，テレビのスイッチを入れる等）を指向するかに応じて，タスク指向と非タスク指向に分類される[13]。具体的な応用例として，スマートスピーカー（AI スピーカー），チャットボット，AI アシスタント，対話型ロボット等が存在する[14]。なお，分身ロボット OriHime 等，人が遠隔操作する対話用インタフェース[15]については本稿では検討の対象としない[16]。

---

9)　「我が国ベンチャーエコシステムが新型コロナウイルス危機を乗り越えるための措置に関して（コロナと戦うベンチャーリスト公開）」一般社団法人日本ベンチャーキャピタル協会 <https://jvca.jp/news/17119.html>。

10)　その他，宍戸常寿ほか編『AI と社会と法』（有斐閣，2020 年）334 頁〔佐藤発言〕も参照。

11)　これは「音声対話システム」の定義である。河原達也「音声対話システムの進化と淘汰」人工知能学会誌 28 巻 1 号 45 号（2013 年）<http://sap.ist.i.kyoto-u.ac.jp/members/kawahara/paper/KAW-slud13-2.pdf>。

12)　「音声対話システム」に関する駒谷和範「音声対話システムの構成と今後」パテント 72 巻 8 号 92 頁（2019 年）<https://system.jpaa.or.jp/patent/viewPdf/3307>参照。

13)　狩野芳伸「コンピューターに話が通じるか　対話システムの現在」情報管理 59 巻 10 号 658 頁（2017 年）<https://www.jstage.jst.go.jp/article/johokanri/59/10/59_658/_pdf/-char/ja>。

14)　八山幸司「米国における自然言語処理技術と人工知能のコミュニケーションをめぐる動向」<https://www.jetro.go.jp/ext_images/_Reports/02/2016/143477940b74c1e1/201611NYrp.pdf>及び「研究開発の俯瞰報告書　システム・情報科学技術分野（2019 年）」国立研究開発法人科学技術振興機構研究開発戦略センター<https://www.jst.go.jp/crds/pdf/2018/FR/CRDS-FY2018-FR-02/CRDS-FY2018-FR-02_05.pdf>。

15)　「NTT、遠隔操作ロボで受付　障害者雇用を推進」日本経済新聞 <https://www.nikkei.com/article/DGXMZO55853410Q0A220C2X30000/>等で注目されている。

16)　なお，本稿が取り上げないものの対話型 AI に関する興味深いテーマである AI と性愛の問題につき，稲葉振一郎ほか編『人工知能と人間・社会』（勁草書房，2020 年）170 頁以下参照。

３　対話型 AI のしくみ

消費者庁の「AI 利活用ハンドブック」は，スマートスピーカーの AI の仕組みをわかりやすく解説し，それを賢く使うための方法についてアドバイスしている[17]。図 1 の模式図のとおり，音声認識をした上で，データを解析し，たとえば，テレビのスイッチを入れる，検索をする等の処理リスト上の処理が指示されると，当該処理リストに基づき，家電を操作したり，外部の検索サービスと接続して必要な情報を収集し，それに基づき返答をする。

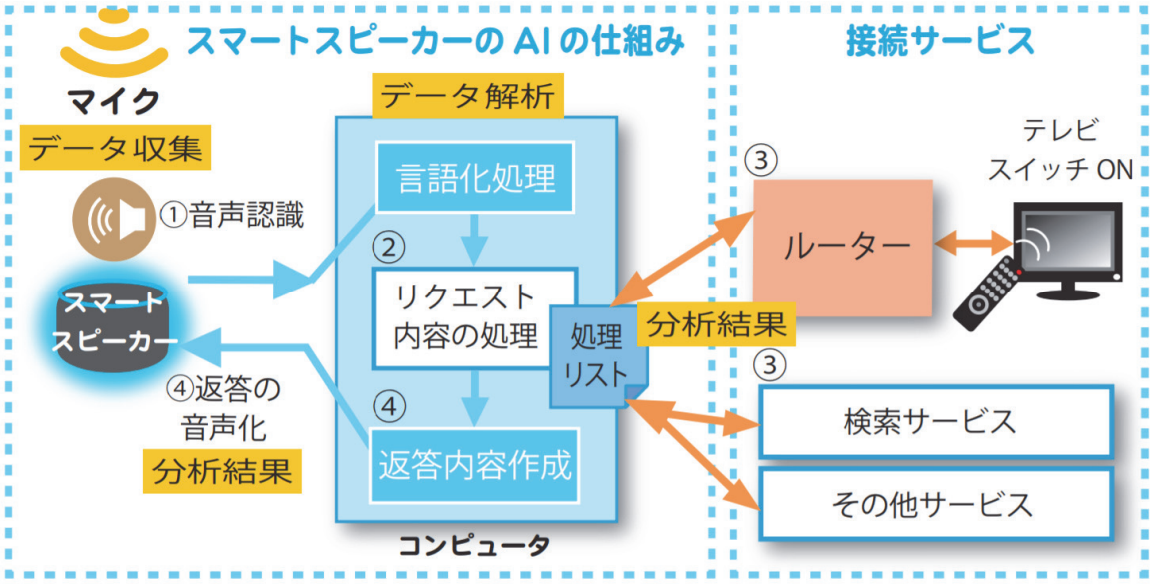

図 1　スマートスピーカーの模式図

（「AI 利活用ハンドブック」12 頁より引用）

対話型 AI を利用した商取引の側面では，2020 年 8 月 28 日に最新改訂が公表された電子商取引及び情報財取引等に関する準則[18]が，図 2 に即してより詳細な説明をしている。すなわち，スマートスピーカーが音声のデータへの変換を行うと，AI クラウドがユーザの指示に従って処理を実施する。すなわち，スマートスピーカーから送信された音声データを認識し，ユーザの指示に従った処理を行い，また，ユーザに対する回答を作成して音声合成データに変換し，スマート

17)　「AI 利活用ハンドブック～AI をかしこく使いこなすために～」（全体版）消費者庁 <https://www.caa.go.jp/policies/policy/consumer_policy/meeting_materials/review_meeting_004/assets/ai_handbook_200804_0002.pdf>。

18)　「電子商取引及び情報財取引等に関する準則」経済産業省 <https://www.meti.go.jp/press/2020/08/20200828001/20200828001-1.pdf>。

スピーカーに返信する機能を果たす。そして，返信を受けたスマートスピーカーが音声を合成して，再生する。

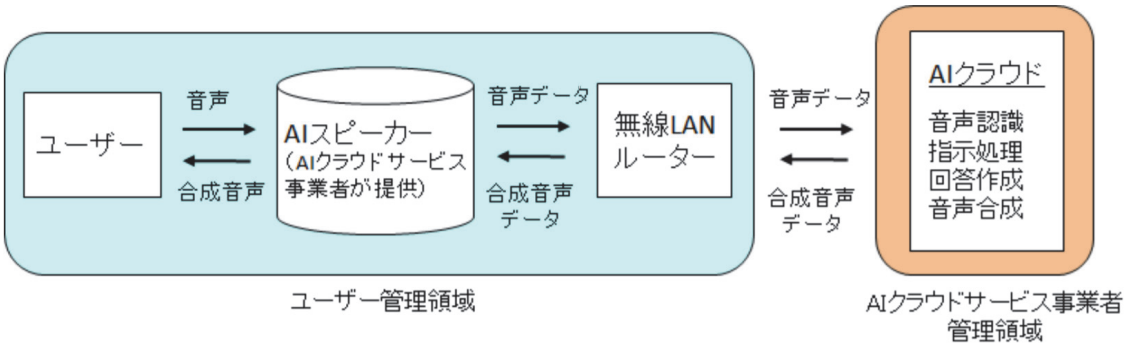

図２　スマートスピーカーを利用した電子商取引の模式図

（「電子商取引及び情報財取引等に関する準則」140 頁より引用）

４　対話型 AI 固有の法律問題

既に AI 一般の法律問題について，様々な議論が積み重ねられてきた[19]ものの，このような対話型 AI については，それ以外の AI とは異なる固有の法律問題が存在するように思われる。

すなわち，「対話」というコミュニケーションを人間と（場合によっては他の AI と）行うことから，当該コミュニケーションにまつわる法律問題が生じる[20]。

たとえば，Microsoft の作成した対話型 AI の Tay[21]の失敗は，多くの読者の記憶に新しいだろう。Tay はニューラルネットワークを利用した学習を行った可能性が高く，ユーザとの交流を通じた学習機能が存在していた。そして，回答ができるだけ質問に対応したものとなるように，大量のデータに基づいてトレーニ

---

19)　（「自動運転」等個別分野に関するものではない）AI と法一般に関する書籍だけでも，弥永真生・宍戸常寿編『ロボット・AI と法』（有斐閣，2018 年），ウゴ・パガロ『ロボット法』（新保史生・松尾剛行・工藤郁子他訳，勁草書房，2018 年），平野晋『ロボット法（増補版）』（弘文堂，2019 年），角田美穂子・工藤俊亮編『AI ロボットと生きる社会―法は AI とどう付き合う?』（弘文堂，2018 年），福田雅樹ほか編『AI がつなげる社会―AI ネットワーク時代の法・政策』（弘文堂，2017 年），山本龍彦編『AI と憲法』（日本経済新聞出版，2018 年）等の様々な研究が存在する。その他実務書や分野を限定したものを含めれば多数にのぼる。

20)　多くのロボット・AI は何らかの人間とのコミュニケーション・インタフェースを有しており，たとえばキーボード入力やタッチパネル操作による「コミュニケーション」を行うということはでき，その限りではこの問題は一定程度 AI 一般にも通暁する問題なのかもしれない。しかし，対話型であることから生じる，当該対話の内容にまつわる問題は，やはり対話型 AI の特徴的な部分であり，なお研究価値があると思われる。

21)　Peter Lee, “Learning from Tay’s introduction”, at<https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/>。

ングがなされた。しかしながら，上記のようにユーザとのやり取りから学習する機能を有していた Tay は，悪意あるユーザとのやり取りの結果として，ヒトラー礼賛，ユダヤ人差別等の発言を行うこととなり，Microsoft は公開を中止した[22]。この問題は，今後，悪意のある人が銀行を認識したら突っ込むデータをテロリズムの手段として自動運転車の学習データに混ぜ込むリスクを考える上の前提となる問題だとも指摘されており，重要な課題である[23]。

５　本稿の概要

本稿では，このような事案を含む主な対話型 AI 固有の法律問題について，それぞれ検討していきたい。

以下では，まず，対話型 AI の発したメッセージについて誰がどのような責任を負うのかについて，名誉毀損及びプライバシー侵害について検討する（Ⅱ）。その上で，この問題に関連して，AI を利用した，違法情報のフィルタリングやゲートキーピングについて，ヘイトスピーチ等も含めて検討する（Ⅲ）。その上で，の対話型 AI とデータ（Ⅳ）及び対話型 AI と取引（Ⅴ）の問題について論じる。最後にそれ以外の問題，たとえば対話型 AI と政府（Ⅵ-１），対話型 AI と捜査（Ⅵ-２），対話型 AI による人格と意思決定への干渉（Ⅵ-３），対話型 AI と雇用（Ⅵ-４）についても論点を提示する。

なお，対話型 AI は対話型ロボットを含むことから，ユーザーが，対話型ロボットに対して「こっちに行け」と命じたところ，通行人とぶつかって，通行人に怪我をさせたといった場合も考えられる。しかし，それは，タッチパネルやリモコンでロボットを操作した場合と大きな違いはなく，対話型 AI 固有の問題ではないと考えられる[24]ので，このような事案は検討せず，以下の議論においては（ロボットに対話型 AI が組み込まれている場合を排除しないものの）ロボットのハードの部分が社会に与える影響は，本稿の検討の枠外に置くこととする。

---

22)　Davey Alba（WIRED.jp_ST 訳）「AI『Tay』を“最低なヤツ”にしたのは誰だ？」<https://wired.jp/2016/03/30/microsofts-teen-ai-turned-into-such-a-jerk/>。

23)　宍戸ほか編・前掲注 10）13 頁〔大屋発言〕。

24)　あえていえば，ロボットに対して「A に行け」と命じたら，A に近い別の A’に向かってしまった（又は A という名のつく場所が複数あり，同じ A と言う名前ではあるものの意図していなかったところに行った）といった，コミュニケーションエラーまたは誤解が生じる場合があり得る。この問題は本稿では直接取り扱っていないが，一定程度コミュニケーションエラーによって間違った取引が成立する場合についての本稿第Ⅴ章の議論が参考にあるかもしれない。

Ⅱ　対話型 AI の発したメッセージに関する責任[25)]

１　はじめに

たとえば，「論争の対象となっているような人物（たとえば，不正疑惑が報道されているが本人は不正を否定している A さん等）について，ユーザ B がスマートスピーカーに「A さんってどうですか？」といった質問をした場合に，スマートスピーカーが，「A さんは不正を働いた悪人です。」のような回答をし，その結果，A の名誉が毀損される可能性がある[26)]。

このような，対話型 AI が発したメッセージの内容が不適切であるため，他人の権利が侵害された場合，誰がどのような責任を取るべきであろうか。

２　一般論（類型毎の議論は３参照）

(1)　関係者

一般には，関係者として，AI 開発・製造者，AI 販売者及びユーザが考えられる（AI クラウドベンダは後述する）。

(2)　ユーザの責任

ユーザ B は単に質問をしているだけであり[27)]，少なくともユーザが使用説明

---

25)「AI の名誉」が毀損された場合にはどのように考えるべきか等 AI の人格権についての議論も始まっている。松尾剛行「対話型 AI（チャットボット，スマートスピーカー，AI アシスタント等を含む）に関する法律問題」（2019 年情報ネットワーク法学会第 19 回研究大会　第 1 分科会「第 5 回ロボット法研究会」AI・ロボットの進化に伴う法と倫理の交錯における発表）及び村田健介「AI と契約，不法行為，人格権」法学教室 479 号（2020 年）42 頁等参照。

26)　個々のユーザとの関係では 1 対 1 のやりとりであって公然性が否定されるようにも思われるが，スマートスピーカの回答を聞いたユーザが他の人に「A さんってこういう人なんだって」と伝播する可能性はあるし，実際に，その時期に A について興味を持った不特定多数のユーザが，スマートスピーカーを通じて同じ質問をして同じ回答を得るのであれば，いわば，ウェブサイト上でアクセスがあれば常に「A さんは不正を働いた悪人です。」と表示すると同じことと考えれば，その仕組み全体をもって公然性ありとみなされる可能性もあるだろう。

27)　なお，質問と名誉毀損については，東京地判平成 28 年 10 月 25 日第一法規 28250488 がインターネット番組の対談において，X が主に発言したものの，Y についても X の発言の否定，訂正にわたる発言がないのみならず，X の発言の根拠を確認する発言もなく，X の発言の摘示する事実を当然の前提として対談を進めているとして共同不法行為を認めている。しかし，Y の質問等が，インターネット番組を通じて公然性を有する形で流通するということが前提となるので，たとえば，ユーザがスマートスピーカーに質問する様子を動画共有サイト上にアップ

書通りに対話型 AI（たとえばスマートスピーカー）を使う限りにおいて[28]，ユーザ B の責任を想定することは困難であろう。

(3)　AI 開発・製造者及び AI 販売者の契約責任

では，AI 販売者と AI 開発・製造者はどのような責任を負うのだろうか。

ここで，①メッセージそのものがユーザ B に損害を与える場合[29]や②「A さんは不正を働いた悪人です。」というメッセージが名誉毀損等で違法とされる場合においてそのような違法なメッセージを示すことが，「品質に関して契約の内容に適合しない目的物を買主に引き渡した」（民法 566 条）とされる場合については，ユーザ B に対する契約上の責任が生じる可能性がある。このような契約責任については，契約関係の存否によって，AI 販売者と AI 開発・製造者の間で果たすべき責任の内容が異なる可能性がある。

すなわち，ユーザとの売買契約を締結した AI 販売者については，債務不履行責任（不完全履行）であるとか契約不適合責任といった契約責任を負い得る[30]。これに対し，ユーザとの契約がなく，あくまでも AI 販売者としか契約を結んでいない AI 開発・製造者は契約責任を負わない。

(4)　不法行為責任

(ⅰ)　一般論

これに対し，（ユーザ B に対する，または第三者である A に対する）不法行為責任（民法 709 条）であれば，AI 販売者だけではなく，AI 開発・製造者も責任を負い得る。もっとも，不法行為責任は過失責任であり，たとえば，A に対する名誉毀損について，責任を負うか否かは，AI 開発・製造者及び AI 販売者に過失があったか，にもよるだろう。

(ⅱ)　過失認定の考え方

この過失の認定の際には，なぜそのような A の権利を侵害する事態が引き起こされたか，AI 開発・製造者及び AI 販売者が，何をすべきだったか等が重要と

ロードする等の行為をしたような，公然性が認められる場合についてのみこのような共同不法行為の問題が生じ得ると考えられる。

28)　例えば，（Tay に悪意を持って間違った回答を教え込んだユーザのような）悪意のある攻撃（いわばサイバー攻撃）等を行う場合には例外的にユーザの責任が生じ得る。

29)　当該メッセージがユーザを侮辱するものであり，その結果，ユーザが損害を被ったといった場合が考えられる。

30)　実際に責任を負うのか否かは，AI 契約者とユーザー間の契約上，AI 販売者がどのような義務を負っていると解釈されるかによるだろう。

なる。ロボットの責任の側面においては，自律性と学習能力の強弱による予測困難性の程度，出力重量パワー強弱及び人との身体的接触，侵襲の有無・可能性の程度等をベースに危険制御の程度が変わる[31]と論じられているが，少なくとも，自律性と学習能力の強弱による予測困難性の程度は対話型AIでも重要な考慮要素であろう。チャットボットと自動運転車の間では想定されるリスクの質が異なることから求められる予見可能性のレベルや結果回避義務の内容は異なると論じられていることも同旨と理解される[32]。

（iii） AIの設計[33]

まず，どのような設計のAIとするかは問題であろう。たとえば「噂AI」として，積極的にかつ公然と噂を流すようなスマートスピーカー（に搭載されたAIソフトウェア）やチャットボットであれば，そのようなスマートスピーカーやチャットボットの発言の中に名誉を毀損する発言が含まれることは予見できるといえるだろう。過失によっても不法行為責任を負うことから，責任を負う可能性は否定できない。なお，結果的にその「噂」の内容が真実であっても，名誉毀損であれば「公益目的」が否定されるし，プライバシー侵害であれば真実であることは抗弁にはならないので，いずれにせよその責任は否定されないと思われる[34]。

（iv） 学習の問題による場合

では，そのような名誉毀損やプライバシー侵害等を積極的に意図する設計ではなく，あくまでも，普通のスマートスピーカーとして対話型AIを設計した場合はどうであろうか。

Tayの例でもあったが，チャットボットやスマートスピーカーの応答の作成の際にSNSのやり取りを元に学習させ，最も適切と思われる回答を自動で生成する機能を設けることが考えられる。

この場合，大量のSNSのやり取りを学習させた結果，SNS上で現にAの名誉

---

31) 近内京太「AI搭載ロボットによる不法行為責任のフレームワーク」NBL1157号27頁（2019年）。

32) TMI総合法律事務所編『IT・インターネットの法律相談』（青林書院，改訂版，2020年）381頁。

33) なお，「噂AI」を実現する方法としては，後述の(AIクラウドベンダに求められる)フィルタリングによって，権利侵害を発生さ得る情報をフィルタリングするであるとか，極めて狭い範囲の人しか噂AIの発信を受領できないようにし，当該範囲の人についても秘密保持を約させることで公然性を回避する等の方法も考えられる。

34) 一般論として，「噂を流したい」という目的は，公益的ではないだろう。

を毀損するやり取りが大量になされており，その結果，対話型 AI も，A の名誉を毀損するようなやり取りをするようになったということがどこまで抗弁となるかは問題である。

すなわち，日本ではリツイート等，他人の投稿を単純に転載しただけでも責任を負い得る[35]。また，いわゆる配信サービスの抗弁の法理[36]による救済を受けるためには，少なくとも最高裁[37]の議論によれば，元の発信者と，転載者の間の実質的一体性が必要である[38]。そこで，「他人の応答をそのまま学習しただけ」という点を抗弁とすることは困難だろう。

また，後述のプラットフォーマーによるフィルタリングが極めて高度となり，それを信頼できる（そこで，フィルタリング後のデータを学習用に用いる限り，権利侵害の可能性が極めて低い）とみなせる状況の下であれば，過失が否定される可能性はある。しかし，少なくとも 2020 年の段階においては，そこまでの信頼を寄せることはできないだろう。

むしろ，学習の結果としてこのような権利侵害となる応答が一定程度含まれることを前提に，AI 開発者・製造者（及び AI 販売者）の方で一定の対応をする義務があるとみなされ，たとえば権利侵害除去対応を十分に行わないスマートスピーカー等を製造・販売したことについて過失が認められる可能性は否定できない。

---

35)　例えば最新の関連裁判例として大阪高判令和 2 年 6 月 23 日裁判所 HP <https://www.courts.go.jp/app/files/hanrei_jp/621/089621_hanrei.pdf>は，「単純リツイートに係る投稿行為は，一般閲読者の普通の注意と読み方を基準とすれば，元ツイートに係る投稿内容に上記の元ツイート主のアカウント等の表示及びリツイート主がリツイートしたことを表す表示が加わることによって，当該投稿に係る表現の意味内容が変容したと解釈される特段の事情がある場合を除いて，元ツイートに係る投稿の表現内容をそのままの形でリツイート主のフォロワーのツイッター画面のタイムラインに表示させて閲読可能な状態に置く行為に他ならないというべきである。そうであるとすれば，元ツイートの表現の意味内容が一般閲読者の普通の注意と読み方を基準として解釈すれば他人の社会的評価を低下させるものであると判断される場合，リツイート主がその投稿によって元ツイートの表現内容を自身のアカウン トのフォロワーの閲読可能な状態に置くということを認識している限り，違法性阻却事由又は責任阻却事由が認められる場合を除き，当該投稿を行った経緯，意図，目的，動機等のいかんを問わず，当該投稿について不法行為責任を負うものというべきである」とした。

36)　佃克彦『名誉毀損の法律実務』（弘文堂，第 3 版，2017 年）530 頁以下参照。

37)　最判平成 23 年 4 月 28 日判時 2115 号 50 頁。

38)　たとえば，SNS 上で「A が不正をした」という応答をした者と，AI 開発・製造者との間の実質的一体性を認めることは通常困難だろう。

（ｖ） 検索結果をそのまま表示していた場合

それに対し，純粋に質問について検索エンジンに流し，検索エンジンの検索結果をそのままスマートスピーカーが読み上げるという場合，スマートスピーカーは単なる検索エンジンを操作し，その結果を音声の形で表示するインタフェースに過ぎない[39]。

ここで，検索エンジンについては，平成 29 年に，いわゆる忘れられる権利といわれる，過去の犯罪歴が検索結果として表示されることについて，消去を求めることができるかという問題について最高裁決定[40]が出されている。同最高裁決定は，検索事業者の検索結果提供事業について，「情報の収集，整理及び提供はプログラムにより自動的に行われるものの，同プログラムは検索結果の提供に関する検索事業者の方針に沿った結果を得ることができるように作成されたものであるから，検索結果の提供は検索事業者自身による表現行為という側面を有する」として，これが表現として保護されるとした。もっとも，その反面，プライバシーを侵害する情報の削除の側面では，「当該事実の性質及び内容，当該 URL 等情報が提供されることによってその者のプライバシーに属する事実が伝達される範囲とその者が被る具体的被害の程度，その者の社会的地位や影響力，上記記事等の目的や意義，上記記事等が掲載された時の社会的状況とその後の変化，上記記事等において当該事実を記載する必要性など，当該事実を公表されない法的利益と当該 URL 等情報を検索結果として提供する理由に関する諸事情を比較衡量して判断すべきもので，その結果，当該事実を公表されない法的利益が優越することが明らかな場合には，検索事業者に対し，当該 URL 等情報を検索結果から削除することを求めることができる」として，一定の要件下で削除が命じられるように，検索事業者が責任をまったく負わないわけではないとされた。

上記のとおり，検索結果を表示するだけの仕組みの場合，対話型 AI がインタフェースに過ぎない点を考えれば，少なくとも上記最高裁決定の枠組みに基づき，検索エンジンが責任を負わないとされる場合についてまで，スマートスピーカーの開発・製造者や販売者が責任を負うと解すべきではないだろう。ただし，

---

39) パソコンで，キーボードを使って「A」と入力して検索し，検索エンジンが「A は不正をした」という検索結果をモニタ上に表示させた場合と同視できる可能性があるだろう。

40) 最決平成 29 年 1 月 31 日民集 71 巻 1 号 63 頁。

検索エンジンが責任を負うと解されるような事情が存在する場合にまで，常に「インタフェースの提供者」に過ぎないという抗弁が成立するかはやや疑問が残る。

（vi）　申告を無視した場合

一般には特定の製品が権利・利益を侵害する（典型的には三菱自動車事件[41]に見られるような人身や人命に関するものであるが，それに限定される者ではないと思われる）ことを製造・開発者や販売者が知った場合，行政法上リコールが義務付けられていたり[42]，そうでなくとも，リコールをしない不作為のために引き続き権利侵害が生じた場合は，当該不作為をもって不法行為となり得る。

この観点からは，スマートスピーカー製造・開発者や販売者が，A から「自分の名誉を毀損している」等と，権利侵害に関する連絡を受けた後，引き続き改修[43]等をせず，権利侵害状態が継続する中で漫然と販売を継続すれば，そのような不作為をもって，過失の認定が相対的に容易になるだろう[44]。

（vii）　開発・製造者と販売者の役割分担と過失

本当に新しい，どのようなリスクがあるかわからない商品であれば，販売者は，それを売り出す以上，事前に権利侵害のリスク等についてチェックすべきであり，それを怠った結果として権利侵害が生じたのであれば過失があるというロジックは成り立ち得る。ただし，一般的な商品について，たとえば家電量販店等の販売者が，そのような権利侵害除去対応が十分にされているかを調査する義務があるか（それをせず漫然と販売した場合に過失があるか）は議論があるところであろう。

2020 年時点では，スマートスピーカーの発売が開始されてから既に５年以上

---

41)　最決平成 24 年 2 月 8 日刑集 66 巻 4 号 200 頁。評釈として松宮孝明「刑事判例研究(10)三菱自工車両車輪脱落事件最高裁決定[平成 24.2.8]」立命館法學 343 号 2049 頁（2012 年）<http://www.ritsumei.ac.jp/acd/cg/law/lex/12-3/matsumiya.pdf>等。

42)　自動車につき道路運送車両法参照。

43)　ソフトウェア・アップデート等を含む。

44)　たとえば，被害者 A が検索エンジンとスマートスピーカー開発・製造者に同時に抗議をし，その連絡をもって，スマートスピーカー開発・製造者も，その問題を認識し得た場合に，漫然とそのスマートスピーカーを販売し続けて良いか，と言った問題である。ただ，（あまり想定できないが）この A が，（検索エンジン等ではなく）パソコンメーカーに抗議をした場合に，パソコンメーカーがパソコンの販売を中止するべき義務を負うとは一般に考えられていないことから，それと同視すれば，スマートスピーカー開発・製造者も同様に販売停止義務を負わない，という解釈もあり得るだろう。

が経過しており[45]，上記のとおり，年間約 1 億 5000 万台も発売されているという事実は，仮に開発・製造者が責任を負うとしても，販売者に注意義務なしとして販売者が免責される可能性を高める事情と言えるだろう。

(5)　製造物責任

上記のとおり不法行為責任は過失責任であるし，契約責任についても，帰責性を要する[46]。これに対し，不法行為法の特則として，製造物責任法は，「欠陥」があれば過失がなくとも原則として責任を認める無過失責任を採用する。

AI チャットボットといったソフトウェアは「製造物」（製造物責任法 2 条 1 項）ではないが，スマートスピーカーのように，ハードウェア（「動産」）とソフトウェアが一体化していれば，製造物である。

ただし，対話型 AI を念頭に，AI が個人情報をオンラインに公開し，あるいは，名誉を毀損する発言をした場合等については，「生命，身体又は財産を侵害」していないことから，製造物責任法の範疇ではないとする議論が存在する[47]。確かにそれが多数説と思われるが[48]，被侵害利益として名誉など精神的人格権などが全くありえないわけではなく，もとより，製造物の欠陥に起因する非侵害利益としてはこの三つのものが考えられるという意味でこれらだけが規定されたにすぎず，これら三つのもの以外を賠償の対象となる損害の範囲から除外する趣旨ではない[49]とされていることには留意が必要である。

とはいえ，料理が温まらない電子レンジ等の，その商品に通常期待される品質が備わっておらず，その本来あるべき性能を発揮できない場合には，これは製造

---

45)　Amazon Echo の販売開始は 2014 年である。

46)　元々は，債務不履行の帰責性は過失の有無として議論されてきた（我妻栄『債権総論（民法講義Ⅳ）』（岩波書店，新訂版，1964 年）100 頁参照）。しかし，2020 年に施行された民法改正により，帰責事由は個々の取引関係に則して，契約の性質，契約の目的，契約の締結に至る経緯等の債務の発生原因となった契約に関する諸事情を考慮し，併せて取引に関して形成された社会通念をも勘案して判断することを明確化した（筒井健夫他『一問一答民法（債権関係）改正』（商事法務，初版，2018 年）74 頁）。学説上は，改正法は，過失責任という考えを廃し，リスク分配を超える障害について免責を認める方向性に舵を切ったと評されている（たとえば，潮見佳男『新債権総論Ⅰ』（信山社，初版，2017 年）377-379 頁参照）。

47)　第二東京弁護士会情報公開・個人情報保護委員会編『AI・ロボットの法律実務 Q&A』（勁草書房，2019 年）58 頁。

48)　経済企画庁国民生活局消費者行政第一課編『逐条解説製造物責任法』（商事法務研究会，1995 年）98 頁及び消費者庁消費者安全課編『逐条講義製造物責任法』（商事法務，第 2 版，2018 年）275 頁参照。

49)　山本庸幸『注釈製造物責任法』（ぎょうせい，1994 年）77 頁。

物責任法上の「欠陥」，すなわち「通常有すべき安全性を欠いていること」（製造物責任法 2 条 2 項）ではない[50]。電子レンジがあたたまらず，その結果として食中毒になるという機序が，製造物責任法の適用されるような権利が侵害される機序ではなく，「通常有すべき安全性を欠いていること」とはいえない，ということであれば，スマートスピーカーが名誉を毀損する発言等の権利侵害発言をうまく除去できず，その結果名誉が毀損されるという機序を製造物責任法が想定していないのであれば，「通常有すべき安全性を欠いていること」とはいえない（「欠陥」が存在しない）という議論にならざるを得ないと考える。

製造物責任法が適用されないのであれば，依然として過失又は帰責性が責任の追及の上で必要となるだろう。

(6)　AI クラウドベンダ（プラットフォーマー）の責任

上記第 I 章 3 で述べた通り，スマートスピーカーが音声のデータへの変換を行うと，AI クラウドがユーザの指示に従って処理を実施する。この AI クラウドベンダ（プラットフォーマー）は，開発者・製造者と同一の場合もあるが，単に一度スマートスピーカーを開発・製造して世の中に流通させるという側面に留まるのではなく，当該クラウドベンダは，その後も継続的にユーザによるスマートスピーカーを通じた入力を踏まえて，処理を実施し，スマートスピーカーがなすべき応答を指示し続けるという意味で，継続的関与がなされる。そして，クラウドベンダは，日々AI ソフトウェア向上に努めて提供するサービスを改善しており，逆にいうと，このクラウドベンダこそが最も容易に権利侵害を予防できる存在である。

すると，単なる開発者・製造者でさえ，上述のとおり，一定の範囲で行政法上，又は不法行為法上（リコールをしない場合にその不作為につき不法行為責任を問われる可能性があるという意味での）リコール義務を負うところ，そのような一回的関与を前提とした開発者・製造者のリコール義務よりは，継続的関与を前提としたクラウドベンダの除去義務ないし是正義務の方がより高まる，ということを指摘することができるだろう。

ただし，このクラウドベンダの義務は，第Ⅲ章で述べる，ゲートキーピングとも連動しており，クラウドベンダのフィルタリング等を通じた除去義務ないし

50)　山本・前掲注 49）36 頁。

是正義務を極めて高度なものとして設定すれば，クラウドベンダは違法と認定されて損害賠償等を義務付けられることを回避するため，極めて回避的に行動するだろう。たとえば，過去の裁判例で 1 度でも名誉毀損とされたキーワードを「検閲」[51]すれば，結果的に多くの情報が遮断され，対話型 AI を通じてそのような情報について知ることができなくなる。その意味では，対話型 AI を運営するクラウドベンダの責任が追及された事案において，裁判所は，このような萎縮効果も視野に入れながら，クラウドベンダにどこまで高度な義務を求めるべきか，という難しい判断を迫られることになるだろう。

３　権利侵害の類型論

上記では特定の権利侵害を想定せずに議論をしてきたが，以下では，どのような権利が侵害されるかに応じて検討したい。

（1）　名誉毀損

名誉，すなわち「人の品性，徳行，名声，信用等の人格的価値について社会から受ける客観的評価」（社会的評価）を侵害する行為が名誉毀損とされる。

虚名も保護され，本当は犯罪を犯しているのにもかかわらずその事実が社会に知られていないことから，社会的には犯罪を犯していないと評価されている場合，「犯罪を犯していない人」という社会からの評価が保護される。

しかしながら，それだけでは，真実を摘示しても名誉毀損となるため，「首相は賄賂をもらっている」といった政治批判等が過度に制限され，表現の自由の趣旨に悖る。

そこで，刑法 230 条の 2 第 1 項は「前条第一項の行為が公共の利害に関する事実に係り，かつ，その目的が専ら公益を図ることにあったと認める場合には，事実の真否を判断し，真実であることの証明があったときは，これを罰しない。」と規定し，民事でも，事実の摘示による名誉毀損と意見・論評による名誉毀損に分けて，以下の抗弁が認められる[52]。

事実の摘示の場合（真実性・相当性の法理）：公共の利害に関する事実に関するも

---

51)　ここでは憲法 21 条 2 項の話ではなく，民間業者の責任回避のための対応が過度に広範で検閲「的」になり得ることを指摘している趣旨である。

52)　松尾剛行・山田悠一郎『最新判例にみるインターネット上の名誉毀損の理論と実務』（勁草書房，第 2 版,2019 年）199 頁以下。

のであり（公共性），専ら公益を図る目的に出ており（公益性），摘示された事実が重要な部分において真実であること（真実性）か，または，真実と信ずるについて相当の理由があること（相当性）

意見・論評の場合（公正な論評の法理）：論評が公共の利害に関する事実に係ること（公共性），論評の目的が専ら公益を図るものであること（公益性），その前提としている事実が重要な部分おいて真実であることの証明がある（真実性）か，または，真実と信ずるについて相当の理由があること（相当性），及び人身攻撃に及ぶなど意見ないし論評としての域を逸脱したものでないことの各要件を満たした場合

スマートスピーカー等の対話型 AI の発言は，上記のとおり名誉毀損の要件を満たし得ることから，ここでは，抗弁，たとえば，真実性・相当性の抗弁の成否について検討したい。

ここで，上記のとおり，公益性の観点が問題となるだろう。すなわち，公共性は客観的に判断されるところである[53]が，公益性は，問題となる発信主体の意図が問われる。

上記の，「噂 AI」のような公益性が存在しない可能性が高い AI は別論，通常のスマートスピーカーの応答について公益性が認められるかは判断が難しいところである。もっとも，一般に，公共性がある事柄についての言及は，公益性が推認され，その言及の目的が特に嫌がらせ目的等である等の特段の事情がなければ公益性があるとする一連の裁判例がある[54]ことから，単に世の中で一般になされている応答を行う，というだけの意図でスマートスピーカーのソフトウェアが開発・製造・販売されていれば，公共性を認めて良いのではないかと考える。

(2)　名誉感情侵害

名誉感情侵害，すなわち，「人が自分自身の人格的価値について有する主観的な評価」や「自分が自分の価値について有している意識や感情」たる名誉感情を社会通念上許される限度を超えて侵害する場合[55]については，たとえば，スマートスピーカーがバカやアホと繰り返すためにユーザーB の名誉感情が侵害

53)　たとえば，客観的に見て一定程度以上重要な不正を A が行っているということであれば，公共性は認められると思われる。
54)　松尾ほか・前掲注 52）216 頁以下。
55)　最判平成 22 年 4 月 13 日民集 64 巻 3 号 758 頁参照。

される場合が考えられる。

この場合，たとえば，ユーザに対して対話型 AI が何度も「こんな簡単なことをわざわざ質問するなんてお前は馬鹿だ，アホだ」と繰り返すことで，名誉感情侵害が成立する可能性があるといった意味で，ユーザが被害者となるメッセージに関する責任の問題があることが特徴的と言えるだろう[56]。

いずれにせよ，単に名誉感情を侵害するだけで違法なのではなく，社会通念上許される限度を超える必要があることから，個別具体的なメッセージが社会通念上許される限度を超えて侵害する場合かが吟味される必要があるだろう。

(3) プライバシー侵害

プライバシーは，いわゆる「宴のあと」事件（東京地判昭和 39 年 9 月 28 日下民集 15 巻 9 号 2317 頁）において「私生活をみだりに公開されない法的保障ないし権利」として定義された。宴のあと事件によれば，①私生活上の事実または私生活上の事実らしく受け取られるおそれのあることがらであること（私事性），②一般人の感受性を基準にして当該私人の立場に立つた場合公開を欲しないであろうと認められることがらであること（秘匿性），③一般の人々に未だ知られていないことがらであること（非公知性）の 3 要件が満たされる場合にプライバシーの公表となる。ただし，プライバシーを公表すればただちに違法となるのではなく，ある事実を公表されない法的利益とこれを公表する理由とを比較衡量し，前者が後者に優越する場合に不法行為が成立するとされている。

その後，情報化社会の中で，プライバシーは「自己に関する情報をコントロールする権利（自己情報コントロール権）」と捉えられるようになり，例えば，単純な氏名と住所等の一定の生活領域内では他人に知られてしまうことはやむを得ない情報であっても，それがその領域を超えて公表，提供等される場合には，一定の範囲で不法行為の成立を認めるようになった[57]。

---

56) これに対し，ユーザの社会的評価を低下させる表現をユーザのみに対して行なっても，公然性がなく名誉毀損にはならない（松尾ほか・前掲注 52）141 頁以下）。

57) たとえば，「学籍番号，氏名，住所及び電話番号は，早稲田大学が個人識別等を行うための単純な情報であって，その限りにおいては，秘匿されるべき必要性が必ずしも高いものではない。また，本件講演会に参加を申し込んだ学生であることも同様である。しかし，このような個人情報についても，本人が，自己が欲しない他者にはみだりにこれを開示されたくないと考えることは自然なことであり，そのことへの期待は保護されるべきものであるから，本件個人情報は，上告人らのプライバシーに係る情報として法的保護の対象となるというべきである」とした江沢民事件判決（最判平成 15 年 9 月 12 日民集 57 巻 8 号 973 頁）や，いわ

スマートスピーカーの応答が第三者 A のプライバシー情報を含むため，A がそのプライバシー侵害を主張する場合，その具体的なプライバシー情報の内容にもよるが，そもそもそのような応答がされた理由は，多数の同様の発言がインターネット上に存在するためである，という点が「宴のあと」事件以来の比較衡量において考慮要素となることは間違いないだろう。

当然のことながら，ある場面で既に当該プライバシー情報が公表されているというだけで，ただちに他の場面でのプライバシー情報の公表等のプライバシー侵害性が否定されるわけではない[58]。

しかし，上記の非公知性の議論からは，たとえば，同じ場面ないし同じ範囲で既に公開されていたプライバシー情報であれば，更に公開をしてもプライバシー侵害にならないとされる可能性はある[59]。

たとえば，SNS 上の公開投稿で既に大量にそのようなプライバシー事項が摘示されているというだけで，「場面」が異なるスマートスピーカーを通じた発信の場面において直ちにプライバシー侵害が否定されるものではないだろう。しかし，たとえば，信頼度が高い情報を優先的に抽出する機能を持った対話型 AI ソフトウェアが，A 本人の公式サイトや公式 SNS 等の情報を元にその情報をユーザ B に伝達したといった場合であれば，特に本人による公開という事情を重視して非公知性が否定される等してプライバシー侵害とならない可能性は十分にあるだろう[60]。

(4)　ヘイトスピーチ

ヘイトスピーチのような人種，民族等に対する偏見については，後述のデータの問題とも関係するし，　最近では特定の人種を選出して追跡する AI も存在す

ゆるベネッセ事件で流出した情報を 「本件個人情報は，　上告人のプライバシーに係る情報として法的保護の対象となるというべきである」とした最判平成 29 年 10 月 23 日判時 2351 号 7 頁を参照。

58)　たとえば，ある掲示板で先に不祥事が公表されると，後は同じ事項を他の掲示板や SNS やブログで発信しても一切プライバシー侵害にならない，と解するのは，ネット炎上の実態に鑑み，到底正当化できないだろう。

59)　松尾剛行『最新判例にみるインターネット上のプライバシー・個人情報保護の理論と実務』（勁草書房，2017 年）110 頁以下及び 171 頁以下。

60)　AI とは無関係であるが，本人が既に公表しているとしてプライバシー侵害を否定したものに東京地判平成 26 年 5 月 16 日第一法規 29027018 及び平成 24 年 8 月 31 日ウエストロー 2012WLJPCA08318008 等がある。

る[61]。

しかし，ヘイトスピーチは現在日本においては直ちに違法ではないことに留意が必要である。確かに，ヘイトスピーチの一環として個人を名指しした場合に，内容によっては，名誉毀損として違法となり得る[62]。しかし，集団を対象とするヘイトスピーチがなされた場合に，被差別集団に属する者すべてが損害賠償等を請求できるとすれば，それは本来不法行為が想定するものではなく[63]，ヘイトスピーチを一般に禁圧するには特別法が必要とされている[64]。

ここで，本邦外出身者に対する不当な差別的言動の解消に向けた取組の推進に関する法律（ヘイトスピーチ解消法）はその 2 条で「専ら本邦の域外にある国若しくは地域の出身である者又はその子孫であって適法に居住するもの（以下この条において「本邦外出身者」という。）に対する差別的意識を助長し又は誘発する目的で公然とその生命，身体，自由，名誉若しくは財産に危害を加える旨を告知し又は本邦外出身者を著しく侮蔑するなど，本邦の域外にある国又は地域の出身であることを理由として，本邦外出身者を地域社会から排除することを煽動する不当な差別的言動」をいわゆるヘイトスピーチとして定義したものの，それに対する啓発・教育等の取り組みを定めるのみであって，これを直ちに違法とはしていない。

もちろん，上記の京都地判令和元年 11 月 29 日裁判所 HP ような個人に対する名誉毀損を構成するものであれば違法であろうし，具体的な態様が差別の不法行為を構成することもあり得る[65]。

---

61) Charles Rollet, “Hikvision Markets Uyghur Ethnicity Analytics, Now Covers Up”, at <https://ipvm.com/reports/hikvision-uyghur>。

62) 「朝鮮学校 X の元校長が日本人を拉致し，国際指名手配されている」と街宣し，その姿を TwitCasting 上で配信したことが朝鮮学校 X を運営する学校法人の名誉を毀損するとして名誉毀損罪で罰金刑とされた事案（京都地判令和元年 11 月 29 日裁判所 HP）等参照。なお，同判決の控訴審判決は，2020 年 9 月 14 日に予定されている。

63) なお，最判平成 15 年 10 月 16 日民集 57 巻 9 号 1075 頁（所沢ダイオキシン報道訴訟上告審判決）は，約 2000 件程度と理解される所沢市内の野菜生産農家の名誉毀損を認めているところ，この法理の拡張の余地はあるが，例えば，数千万人いる「韓国人」にまで拡張できるかはおおいに疑問である。

64) 曽我部真裕ほか『情報法概説』（弘文堂，第 2 版，2019 年）317 頁。

65) 東京高判平成 30 年 5 月 18 日判例時報 2395 号 47 頁は，公民館で俳句が掲載されるはずが，政治的という理由で掲載されないという事案で思想信条を理由に差別を受けない人格的利益を認めた。また，大阪高裁平成 30 年 6 月 28 日第一法規 28263635 は韓国人差別を内容とするブログ記事等につき人種差別の不法行為を認めた。

地方自治体では，たとえば，大阪市ヘイトスピーチへの対処に関する条例が地方自治体のレベルでヘイトスピーチを規制し，申出等に基づき，ヘイトスピーチ審査会の意見を聴き，ヘイトスピーチに該当する場合，表現内容の拡散防止措置をとるとともに，表現内容の概要，表現活動を行ったものの氏名又は名称等を公表する[66]とか，川崎市差別のない人権尊重のまちづくり条例がヘイトスピーチを違法として是正を命じ，命令違反に対し罰則を課す（川崎市差別のない人権尊重のまちづくり条例 23 条参照）[67]等の試みがみられる。

このような場合に関する対話型 AI に対する実務指針としては，レピュテーションリスクを考慮しながらも同時に萎縮しすぎない対応が考えられる。

すなわち，日本においては確かに現状ヘイトスピーチは違法とまではいえない。しかし，一部のヘイトスピーチは名誉毀損や人種差別の不法行為として違法である。また，仮に違法といえなくとも，だからといって野放しにして良いのではなく，ヘイトスピーチを解消するために徐々に機運が高まっており，そのようなマイルストーンの 1 つがヘイトスピーチ解消法である。そして，その後も各自治体の努力が見られる。そのような状況において，Tay のようなヘイトスピーチを垂れ流す行為について「違法ではないからやってもいい」という態度をとってしまえば，対話型 AI 技術に対する社会的受容性は得られないだろう。

ただし，だからといって，たとえば「ヒトラー」という言葉を対話型 AI に対する禁忌語にしてしまえば，歴史について対話する AI が，ヒトラーについてなぜ権力を握ることができたか，その教訓は現代でどのように活かせるか等についてユーザーとの間で健全な対話を実現することすら不可能になってしまう。

その意味では，対話型 AI にとってレピュテーションリスクを考慮しながらも萎縮しすぎない対応が必要であると思われる。その意味を具体的にいえば，①まずは違法とされる表現に対しては名誉毀損に準じた是正対応・除去対応をした上で，②そうでないものについても，野放しにするのではなく，「文脈によって

---

66)　保守速報の書き込みについて公表「大阪市ヘイトスピーチへの対処に関する条例に基づくヘイトスピーチの公表（案件番号「平 28－6」）」
<https://www.city.osaka.lg.jp/shimin/cmsfiles/contents/0000339/339043/anken28-6.pdf>。
「朝鮮人がいない日本を目指す会」による街宣とその動画配信について公表「大阪市ヘイトスピーチへの対処に関する条例に基づくヘイトスピーチの公表（案件番号「平 28－21」）」
<https://www.city.osaka.lg.jp/shimin/cmsfiles/contents/0000339/339043/anken28-21.pdf>。

67)　「川崎市差別のない人権尊重のまちづくり条例」川崎市役所
<http://www.city.kawasaki.jp/250/cmsfiles/contents/0000113/113041/jyourei1.pdf>。

はヘイトスピーチになり得るし，文脈によってはヘイトスピーチではないような表現」について（求められる程度が過剰すぎると上述の委縮効果が生じるので，その程度については慎重な検討が必要なものの），対話型 AI 開発者・製造者や，クラウドベンダが，文脈毎に適切に判断をして，表現する/しないを決めるような AI ソフトウェアを利用する，ないしは，そのように AI ソフトウェアを学習させることが求められるということを意味するだろう。今後も対話型 AI の開発者にとって，ヘイトスピーチは頭を悩ませる問題であり続けるだろう。

(5)　フェイクニュース

トランプ大統領が選出された 2016 年の米国大統領選挙等で，フェイクニュースは重大な問題として議論された。フェイクニュースは多様性があるので，国家が定義して直接規制するのは恣意的過剰規制につながる[68]とされており，総務省の研究会では，このような表現の自由への萎縮効果への懸念や，偽情報の該当性判断の困難性による実効性の欠如及び恣意的運用への懸念など，多くの課題や批判の声を踏まえ，日本における偽情報への対応の在り方の基本的な方向性としては，まずはプラットフォーム事業者を始めとする民間部門における関係者による自主的な取組を基本とした対策を進めていくことが適当とされた[69]。

フェイクニュースも名誉感情侵害と同様に，単に A に関する偽のニュースを流されて A の権利が侵害されるという問題だけではなく，B が嘘の情報を対話型 AI から聞き，それを信じて行動をすることで損害を被る[70]場合があるという意味で，名誉感情侵害とも類似している。また，フェイクニュースには，名誉毀損（たとえば不正が存在しないのに A が不正したというフェイクニュースがメッセージとして対話型 AI から発せられる）等の違法な場合と，違法とは言えない場合に双方があるという点では，ヘイトスピーチと類似している。

---

68)　宍戸ほか編・前掲注 10）318 頁〔山本発言〕。

69)　「プラットフォームサービスに関する研究会最終報告書」総務省 <https://www.soumu.go.jp/main_content/000668595.pdf>。

70)　たとえば，AI スマートスピーカーに特定のうがい薬が新型コロナウイルスに効くと言われて，うがいをする習慣がない B がうがい薬を購入したが，実際はうがい薬が新型コロナウイルスに効く根拠はなく，うがい薬代を損したと言った場合が考えられる。

Ⅲ　対話型 AI とゲートキーピング

１　はじめに

上記のとおり，対話型 AI については，開発者・製造者及び AI クラウドベンダがそのソフトウェアから違法な表現を取り除く対応を「一定程度」行う義務を負うと解され，それに違反すれば，不法行為責任を負うと解される。では，いかなる範囲でそのような対応を行うべきだろうか。これは，ゲートキーピングやフィルタリング等の問題である。

ここで，この問題と関係するのは，プラットフォームが AI を利用した違法情報除去をする試みであることから，まずはこの点について検討する。

２　プラットフォームによるゲートキーピング

現在は，様々な SNS 等のプラットフォームにおいて，名誉毀損その他の規約違反予防のため AI 等を利用して，情報のフィルタリングを行い，そもそも投稿させない，投稿に対して警告する，アカウントを停止する等の対応をしている[71]。

SNS プラットフォーム等が，本来は国家が行うべき民事裁判・刑事裁判等の代わりに，その「規制の代理人（ゲートキーパー）」として違法情報を除去したりアカウント停止する等の是正対応をしていると評され，そのゲートキーパーとしての役割が拡大・変容しつつあると評されている[72]。

今後，高度に発達した AI に，関連するすべての名誉毀損の裁判例を読み込ませて学習させれば，「この文脈におけるこの具体的な表現が違法とされる可能性は○％」といった評価を正確に行うことができる可能性がある[73]。

中川は，いわゆる「忘れられる権利」で論じられているデータを消去すべきか否かという文脈において，

$$\{f(x)=yes/no\}$$

という数式を立て，データと判断結果のペアを教師データとして機械学習す

---

71)　なお，プラットフォームに限定せず，カリフォルニア州 IoT 規制法の様な端末供給者をゲートキーパーとするという発想もあることにつき，宍戸ほか編・前掲注 10）272 頁〔湯浅発言〕を参照のこと。

72)　宍戸ほか編・前掲注 10）58 頁〔生貝発言〕。

73)　松尾剛行「ウェブ連載版『最新判例にみるインターネット上の名誉毀損の理論と実務』第 19 回」勁草書房編集部ウェブサイト<https://keisobiblio.com/2016/07/07/matsuo19/4/>参照。

れば，その学習過程を経て，AI は最終的には精密な f を学習することができるとした[74]。

ただ，2020 年現在の技術水準であれば，AI によるフィルタリングは，キーワードによるフィルタリングから大きく離れていないレベルと言わざるを得ない。

実際には侵害されるオリジナルコンテンツの量に限界がある著作権であればそれなりの違法・適法判断ができても[75]，動画共有サイトの行うスパム認定等では現状においてスパム認定 AI が大きなミスをして裁判沙汰を含むトラブルが発生することがある[76]。なお，プラットフォーム上では大量の情報が流通することから，プラットフォーム事業者は，機械学習を含む AI によるアルゴリズムを活用して情報の流通をコントロールすることが一般的になっている。この点，機械学習を含む AI 技術により情報の削除等の対応を行う場合には，AI により誤った判定がなされ，本来は削除されるべきでない情報まで過剰に削除されてしまう可能性や，アルゴリズムにより不当に表示順位が低くなってしまうなどの可能性がある点が課題[77]としてこのような AI の誤りの可能性を問題視する議論がある。

そもそも，ある単語が名誉毀損となるかは高度に文脈に依存しており，たとえば，「ドロボー本」[78]や「平成の毒婦」[79]等，かなり過激な表現も裁判で争われた結果，当該特定の文脈の下で名誉毀損にはならないとされている[80]。また，たとえば，名誉を毀損する言動を相手がしたことを理由に相手を批判するとか，名誉毀損の研究者が名誉毀損の事例を収集して公開する（直接対象者の氏名は出さないのが通常であろうが，それでも，名誉毀損法に関する検討のために必要な範囲で事案の概要を述べると，その結果として対象者が特定されてしまうことが前提である）等，その表現だけ

---

74) 中川裕志『裏側から視る AI』（近代科学社，2019 年）100 頁。

75) 宍戸ほか編・前掲注 10）202 頁〔佐藤発言〕参照。ただし，著作権の場合，適切に運用されないと，過度な萎縮効果が生じ得ることは，たとえば著作権法改正と同人誌の関係で巻き起こった議論を参照すべきである。
「著作権法の一部を改正する法律等の公布・施行について（通知）」文化庁
<https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/bunkakai/53/pdf/r1413733_11.pdf>。

76) 松尾剛行『AI・HR テック対応　人事労務情報管理の法律実務』（弘文堂，2019 年）62 頁注 8。

77) 総務省ウェブサイト・前掲注 69)。

78) 最判平成 16 年 7 月 15 日民集 58 巻 5 号 1615 頁。

79) 東京高判平成 29 年 4 月 17 日第一法規 28254372。

80) 松尾ほか・前掲注 52）306 頁以下参照。

を見ると一見名誉毀損であることが明らかあっても，そのような具体的利用方法ないしは態様であれば名誉毀損にならない，という状況もあり得る。加えて，現在はプラットフォームの国際化が進んでいるところ，日本と外国では違法判断の基準が異なる[81]ため，日本で適法な表現が，外国基準で削除され，実質的に禁圧される危険についても留意が必要である[82]。

このような状況を想定すれば，もちろん，過激な行為を繰り返すユーザに対する対策としての AI の利用等は有益ではあるものの，AI のみで対応できる範囲はかなり限定されるといえるのであり，事前対応としても警告程度に留め，また，事後対応としても，AI が違法と判定しても，異議が申し立てられれば，人間がチェックするような制度設計とすべきである。実際に，AI 技術により削除等の対応を行った結果として利用者から苦情や問合せがあった場合には，適切なアカウンタビリティを確保する観点から，苦情処理プロセスの中で人による確認や対応を行うことが望ましいとする議論も存在する[83]。

このような人間の介在や説明は，単なる削除の正確性の担保というだけではなく，削除に伴い行われる監視の透明性や説明責任の観点からも重要である。実際には，検索，ショッピング，行先経路検索，友人とのコミュニケーション等，生活に関する情報のほとんどを収集しているプラットフォームが積極的に監視を行えば，国家が行う以上の監視社会を生み出す危険もあると警鐘が鳴らされている[84]ことから，単に違法な情報が正確に削除されていれば良いというだけではなく，アカウンタビリティや透明性[85]に対する責任を果たしながら削除しているかも問題となるだろう[86]。

---

81)　文脈は違うが，宍戸ほか編・前掲注 10）7 頁〔小塚発言〕参照。

82)　たとえば，日本法上適法な表現について国際クレジットカードベンダのグローバルな対応の結果，当該表現を含む書籍・雑誌等の販売を停止しなければクレジットカードが使えなくなるといった状況が既に問題視されている。谷本陵「『成人誌販売はクレカ決済 NG』　突然の通達で利用者困惑、与党議員も問題視」J-CAST ニュース <https://www.j-cast.com/2019/08/29366242.html?p=all>。

83)　総務省ウェブサイト・前掲注 69）。

84)　宍戸ほか編・前掲注 10）62 頁〔生貝発言〕。

85)　なお，特定デジタルプラットフォームの透明性及び公正性の向上に関する法律が 2020 年 6 月に成立した。

86)　この点は後述の AI と説明についての議論も参照すべきである。

３　対話型 AI の場合

以上のようなプラットフォームの（AI を用いた）フィルタリングや違法情報除去・是正の試みに関する問題構造を，対話型 AI に適用するとどのようなことがいえるだろうか。

まず，対話型 AI の開発者・製造者や AI クラウドベンダは，そこで利用する AI ソフトウェアに違法除去・是正機能を組み込ませたり，別途プラットフォームが用いている違法投稿除去・是正ソフトウェアを利用したりすることで，一定程度違法情報を除去することが可能である。そして，上記第Ⅱ章 2 (6) で述べた，AI クラウドベンダ等が違法な表現を取り除く対応を「一定程度」行う義務は，人力による除去によっても理論上は実現できるものの，人間のみならずこのようなソフトウェア上の対応を行うことで，効率化や高度化を行うことができることは事実であろう。

不法行為法における注意義務の程度は，社会の発展に応じて変化するところ，同業他社がどのように対応しているかは，その注意義務の程度を判断する上で参考になる。そして，一定以上の精度の違法投稿除去・是正 AI の利用が既に業界標準となっているといえれば，ある対話型 AI の開発者・製造者や AI クラウドベンダがたとえば人力のみの対応を貫いたり，業界で一般に用いられているよりも（より安価な）レベルの低い違法投稿除去・是正 AI しか用いず，その結果として A の名誉が毀損されると言った結果が生じれば，注意義務違反を認定する方向性の事情といえるだろう。

ただし，上記のとおり，現在の状況下では，AI が誤ることも多い。そこで，フィルタリングを行う AI 等の誤りの結果として対話型 AI が十分な役割を果たせないとすれば，対話型 AI の価値が大きく切り下げられ（たとえば「ヒトラー」を禁忌語にすることに関する上記の議論参照），また，場合によっては，AI との対話を通じて人がその人格を発展することが妨げられるといった，表現の自由の問題にまで発展する可能性がある[87]。

AI の誤り等の可能性を踏まえ，人間の介在や透明化等，プラットフォームにおけるゲートキーピングにおいて推奨されている手法をこの文脈でも採用すべきである。

---

87)　ただし，基本的には，民間企業の提供する AI であるから，間接適用説からは，表現の自由が投影される私法の一般規定上でこのような表現の自由の制限が問題となるだけであろう。

このように，プラットフォームにおけるゲートキーピングの議論は，対話型 AI における違法情報除去に応用可能であるが，1 点重要な相違があると思われる。すなわち，投稿時の警告の可否である。

一般的なプラットフォームによるゲートキーピングであれば有用と思われる事前対策に「警告」があげられる。投稿禁止や一度なされた投稿の除去よりも，違法な可能性がある投稿について警告し，本人に再考してもらう，というのは表現の自由に配慮しながらも，違法投稿を抑圧する方法である[88]。既に過去に不適切な投稿をした者が，投稿を新規に行う際にプラットフォームが警告メッセージを出す等の実践例が存在する[89]。しかし，これはフォームによるゲートキーピングにおいて利用することができても，対話型 AI のゲートキーピングにおいては利用が難しいだろう。それはスマートスピーカーのユーザー等通常のユーザーは，ユーザー自身が積極的に何か違法情報を投稿等しようとしている訳ではないからである。ただし，Tay に対する誤った学習をさせたユーザー等，例外的な場合には，事前警告が機能する余地はあるかもしれない。

４　個別の権利類型に応じた留意点

(1)　名誉毀損

名誉毀損の場合には，上記のとおり，同じ単語でも名誉毀損として違法とされるか否かは文脈に依存するという特徴があり，特定の単語が名誉毀損になった可能性が高いとして機械学習で強制的に排除すると，異なる文脈で適法に当該単語を用いた対話が妨げられることが重要である。

(2)　名誉感情侵害

名誉感情侵害の場合，確かに直接的に人を侮辱する表現のパターンには限りがあるので名誉毀損と比較すると，「相対的」にはよりソフトウェア的な処理に向いているとはいえる。

ただし，侮辱的な発言も，社会通念上許される限度を超えない限り違法ではない。掲示板上の投稿事例で，1 回だけ「くそ」と投稿しても違法な名誉感情侵害とはならないが，5 時間の間に 13 回「くそ」と投稿すれば名誉感情侵害となる

---

88)　那須耕介・橋本努『ナッジ!?』（勁草書房，2020 年）も参照。

89)　「AI を活用した投稿時注意メッセージの掲出を開始しました」Yahoo!ニュース <https://news.yahoo.co.jp/newshack/information/comment_alert.html>。

といった裁判例もあり[90]，ソフトウェア的対応をする場合には，複数回繰り返せば名誉感情侵害となり得ても，1回だけであればならないといった表現について，その繰り返しを踏まえた対応をするべきであり，1回のみ出現するというだけで，一律除去・是正することは妥当ではないだろう。

(3) プライバシー侵害

プライバシー侵害については，たとえば，上記で「忘れられる権利」について中川が指摘したように，10 年前の犯罪報道を削除すべきかしなくていいか，といった，過去のプライバシー情報の削除の文脈であれば，比較的ソフトウェア対応が容易であろう。それは，上記平成 29 年最高裁決定が挙げた考慮要素の数がそこまで多くなく[91]，ある程度の定型化が可能だからである。

しかし，プライバシー侵害一般になると，総合考慮のための要素数が増加するので，名誉毀損と同様に，文脈に応じた判断が必要であろう。

(4) ヘイトスピーチ

ヘイトスピーチについても上記の議論があてはまり，①違法であるか，と②仮に違法でなくとも社会的責任と表現の自由の衡量の下除去すべきかを考える必要があるが，なかなかソフトウェアで一筋縄に対応するわけにはいかず，人力での判断が求められる部分が多いといえるだろう。

(5) フェイクニュース

総務省の前記報告書等を前提とすると，AI クラウドベンダや開発者・製造者等は，ファクトチェック等のための一定の努力をすべきとは思われるが，偽情報の該当性判断の困難性からは，法的な意味でその除去・是正が義務付けられる程度は必ずしも高くはないだろう。

---

90) 東京地判平成 29 年 1 月 16 日第一法規 29038202。

91) 「児童買春をしたとの被疑事実に基づき逮捕されたという本件事実は，他人にみだりに知られたくない抗告人のプライバシーに属する事実であるものではあるが，児童買春が児童に対する性的搾取及び性的虐待と位置付けられており，社会的に強い非難の対象とされ，罰則をもって禁止されていることに照らし，今なお公共の利害に関する事項であるといえる。また，本件検索結果は抗告人の居住する県の名称及び抗告人の氏名を条件とした場合の検索結果の一部であることなどからすると，本件事実が伝達される範囲はある程度限られたものであるといえる。以上の諸事情に照らすと，抗告人が妻子と共に生活し，前記一（1）の罰金刑に処せられた後は一定期間犯罪を犯すことなく民間企業で稼働していることがうかがわれることなどの事情を考慮しても，本件事実を公表されない法的利益が優越することが明らかであるとはいえない」との判示参照。

Ⅳ　対話型 AI とデータ

１　はじめに

データに関しては，様々な問題が存在する。たとえば，各過程毎の問題であり，データ収集過程，学習過程，試験運用過程そして実運用過程がある。また，内容の問題であり，プライバシー，偏見，データオーナーシップ，セキュリティの問題がある。以下では，プライバシー・個人情報（２），偏見（３），データオーナーシップ（４），サイバーセキュリティ（５）について，それぞれその問題が生起する過程も意識しながら検討する。

２　プライバシー・個人情報

(1)　取得

上記のとおり，プライバシーや個人情報は，データの典型的な問題である[92]。対話型 AI はユーザーとの会話を録音してクラウドへ送信する[93]。すると，ユーザの発言や生活音等のプライバシー情報がスマートスピーカー経由で AI クラウドベンダ等に取得されるという問題がある。

たとえば，監視カメラの目の前で個人情報を話すように言われれば警戒して最小限に止めようとするだろうが，対話型 AI，たとえば人間型ロボットや動物型ロボットが「僕の名前は○○だよ，君の名前は？」と個人情報を聞き出す場合には，心理的障壁が下がりペラペラと個人情報を伝えてしまいがちである[94]。このような心理的障壁を下げるという対話型 AI の効果に鑑みると，当該スマートスピーカーが取得するプライバシー情報をどのように利用するかについては十分な説明が必要であろう。

この問題については，「カメラ画像利活用ガイドブック　ver2.0」[95]が参考にな

---

92)　なお，上記第Ⅱ章や第Ⅲ章で論じたプライバシーは「メッセージ」のプライバシー侵害性（たとえばユーザが「A ってどういう人」と聞いたら「A は今不倫をしたとして批判されている人です」と答える）といった場合の問題であるのに対し，ここではユーザそのもののプライバシーが問題である。

93)　第二東京弁護士会情報公開・個人情報保護委員会・前掲注 47）21 頁。

94)　石井夏生利「伝統的プライバシー理論へのインパクト」福田雅樹ほか編『AI がつなげる社会』（弘文堂，2017 年）203−206 頁。

95)　「『カメラ画像利活用ガイドブック ver2.0』を策定しました」経済産業省<https://www.meti.g

る。このガイドラインは，カメラによる情報の収集について，合理的な告知をする等の方法で，予想外の利用を可及的に回避しようとしているところ，この問題意識は，街頭に置かれた対話型ロボットを利用した録音録画にも応用可能だろう。

(2) 利活用

収集した情報の利活用も問題となる。

たとえば，スマートスピーカーの録音したユーザーの発声内容や生活音等をAIクラウドベンダの従業員やその業務委託先の従業員が品質向上等のために聞くことが判明し，批判を浴びた例があった[96]。スマートスピーカーの仕組み上，リビングにスマートスピーカーをおけば，部屋中の，そして，家の大きさによっては「家中」の音が記録されることは当然に発生し得る。そして，録音がAIクラウドベンダに渡ること（取得がされていること）自体はその機能上当然生じるものである。そこで，問題の本質は，情報が取得され，AIクラウドベンダやその委託先に渡ったこと自体にはなく，むしろ，その取得された情報についてユーザはあくまでもソフトウェアのみが利活用すると信じていたところ，人間がこれを聞くことでプライバシー上の問題が生じたといえるだろう[97]。

法的にいえば，利用目的の特定・明示（個人情報保護法15条1項，16条）や第三者提供規制（同23条）が問題となるところ，たしかにプライバシーポリシーには，更に利便性を向上するための利用が記載されていたし，業務委託先への提供であれば，本人同意がなくても第三者提供をすること自体は可能である（同23条5項1号）。しかし，自動的に自分たちの生活音が処理されているならば受容できても，自分達のプライベートな活動（排泄，入浴，性生活等）に関する音声が人間の従業員に聞かれたかもしれないということがユーザに与えた不安は大きい。

この点は，単なる個人情報保護法の遵守だけでプライバシーの問題がないと考えるべきではなく，「NO SURPRISE（驚きを与えない）」の原則から，ユーザに不安を与えないよう，より丁寧な説明が求められ，場合によっては（個人情報保

---

o.jp/press/2017/03/20180330005/20180330005.html>。

96) 小久保重信「アップル、『Siriの盗聴問題』で信頼回復を図る」Japan Business Press<https://jbpress.ismedia.jp/articles/-/57479>。

97) 監視の先に機械か人のいずれがいるか否かを重視するものとして，宍戸ほか編・前掲注10）9頁〔大屋発言〕だが，筆者はHRテックの文脈で，人とAIが協働して監視をする場合のプライバシー侵害について論じた（前掲注76）215－216頁）。

護法上適法でも）プライバシー侵害で違法という判断もあるだろう。

(3)　プライバシー・バイ・デザイン

プライバシーの保護のため，プライバシー・バイ・デザインの利用が叫ばれている[98]。しかし，対話型 AI 等の AI については，学習による変化という性質上「バイデザイン」，すなわちアーキテクチャの設計段階におけるリスク抑制に限界があるとも指摘されている[99]。

しかしながら，限界があるからといって，プライバシーのことを設計において考えなくていいのではなく，たとえばスマートスピーカーの設計上，たとえば，特定の言葉（Hey Siri，OK Google，アレクサ等）が発せられるまでは（生活音がスマートスピーカーに入っても）録音をしないことをデフォルトとする等，設計上，デフォルトルールをプライバシーに配慮するといった可能な限りのデザイン（設計）上の対応はすべきである[100]。その上で，上記の学習によるバイデザインの限界の点については，設計だけでは（対話型）AI のプライバシー保護は不十分であり，それ以外にも学習による利活用の程度の変化に応じた更なる丁寧な説明（場合によっては，利用開始後の説明）等，追加的対応が求められる可能性がある，という趣旨と理解すべきであろう。

３　偏見

AI における偏見は 1 つの重要な課題である。この問題については，①収集したデータセットが偏っているのではないか，②そもそも世間において偏見が存在し，その偏見がデータセットに反映されているのではないか，③データを解釈する際に，製作者等の偏見が含まれるのではないかといった多数の問題がある。

①については，ボストンにおいて，道路の修理等を効率化するためスマートフォンアプリで修理箇所を通報できるようにしたら，スマートフォンを持てるような高級住宅街のみ修理が増え，スマートフォンを持てない貧乏な人が多数住

98)　例えば，堀部政男・一般財団法人日本情報経済社会推進協会編『プライバシー・バイ・デザイン』（日経 BP 社，2012 年）。

99)　成原慧「AI ネットワーク社会におけるアーキテクチャと法のデザイン」稲葉ほか編・前掲注 16）102 頁。

100)　このように，デフォルトでプライバシーを保護するプライバシー・バイ・デフォルトの発想は重要である。

んでいる地区は修理されなくなったという研究[101]が存在するが，これは，この問題の 1 つの表れと言えるだろう。

②については，より分かりにくいものであるが，たとえば，日本社会における多くの対話データを元にしたコーパスをベースにすると，当該コーパスには，日本社会の偏見が反映されている可能性が高い。たとえば，LGBT に対する偏見が含まれていれば，対話の中でそのような偏見が表面化し，たとえば最近であれば山口テレビが「珍　女性のような男性」として LGBT の人を取り扱って謝罪した[102]ような出来事が，対話型 AI の行う会話の中で生じかねない。

③については，対話型 AI の開発者自らが特に留意しなければならない。たとえば，試験段階の AI が女性型 AI のはずなのに「男性らしい」振る舞いをしたという場合[103]，それをもって「AI がうまく学習できていない」と判断するのはもしかすると，単に開発者が有する偏見に基づく誤った対応かもしれない。もしかすると最近の若者は，従来のステレオタイプ的な意味における男性/女性の行動と事なる行動をしており，そのような若者の行動の反映として AI が振る舞っているのかもしれない[104]。

これらの偏見の結果，一度 AI やプロファイリングによって信用度が低いとみなされた人がその偏見を二度と是正できなくなる「バーチャル・スラム」の問題が重要な問題として論じられている[105]。たとえば融資や付保におけるチャットボット等の対話型 AI の利用が進む中，この問題は，理論上の問題ではなく現実的な問題として立ち現れてくるだろう。

そして，これらの問題を単なるデータの問題ではなく，研究者や開発者の公平性に対する姿勢の問題だとする見解が出てきていることにも留意が必要だろう[106]。

---

101) 山本龍彦「ロボット・AI は人間の尊厳を奪うか？」4 章『ロボット・AI と法』弥永真生・宍戸常寿編（有斐閣，2018 年）84−85 頁。

102) 「LGBT に対して不適切放送　テレビ山口が謝罪」日本経済新聞<https://www.nikkei.com/article/DGXMZO52233640V11C19A1000000/>。

103) たとえば，女性型 AI が「青い色のものが好き」という場合を想定することが可能かもしれない。

104) むしろ，「女性が青い色が好きで何が悪い？」という姿勢を開発者としてとるべきかもしれない。

105) 山本龍彦編『AI と憲法』（日本経済新聞出版社，2018 年）72 頁。

106) Pratyusha Kalluri, “Don’t ask if artificial intelligence is good or fair, ask how it shifts power”, Nature 583, 169 (2020). Frank Pasquale, “The Second Wave of Algorithmic Accountability”, posted on “Law and

４　データオーナーシップ

データオーナーシップについては，AI・データの利用に関する契約ガイドライン（データ編）[107]が詳しい。

データそのものには「所有権」はないものの，たとえば店頭に設置した案内ロボットが取得したデータは店舗のものか，顧客のものか，はたまた AI クラウドベンダのものかが問題となる。

基本的には，顧客ないしは本人は自己の個人情報について，個人情報保護法上の開示請求権等を有している。また，個人情報取扱事業者である店舗も AI クラウドベンダも，個人情報保護法に違反して個人情報を取り扱ってはならないし，プライバシーを侵害して取り扱ってもいけない。その意味で，本人以外が完全な「オーナーシップ」を有することはない。

もっとも，その限り（つまり，個人情報にも反しないし，プライバシーも侵害しない範囲）において，店舗や AI クラウドベンダは，店舗がどの範囲で当該データを活用し，AI クラウドベンダがどの範囲で当該データを活用できるか合意により決定することができる[108]。

ただし，単に違法ではない範囲で，事業者が合意すればなんでもしてよいというものではない。たとえば，コネクテドカーとプライバシー[109]の文脈で，トヨタが個人情報保護法やプライバシー法より高いレベルの説明を志向していることが参考になる[110]。

５　サイバーセキュリティ

サイバーセキュリティは対話型 AI にとってますます重要になる問題である。AI ではないが，生活費を渡す代わりに全てのデータを取得するプロジェクトは，

Political Economy” at November 25, 2019.

107)　「AI・データの利用に関する契約ガイドライン v1.1」経済産業省<https://www.meti.go.jp/press/2019/12/20191209001/20191209001.html>。

108)　経済産業省ウェブサイト・前掲注 107）も参照のこと。

109)　加藤伸樹・大島義則・松尾剛行「コネクテッドカー・自動運転車(自動走行車)とデータに関する法律問題の検討」自動車技術 74 巻 5 号（2020 年）55 頁参照。

110)　「コネクティッドカーから取得するデータの利活用・保護の取り組みについて」トヨタ自動車株式会社コネクティッドカンパニー
<https://toyota.jp/pages/contents/tconnectservice/contents/pdf/toyota_datapolicy.pdf>。

始まる前にデータ漏洩事故を起こした[111]。対話型 AI によってますます多くのものが処理され，AI クラウドベンダにますます多くのデータが蓄積されるにつれ，セキュリティリスクは更に重要となる。上記の Tay の事案も，一種のサイバーセキュリティ事案ととらえるべきであろう。

なお，セキュリティに関しては，「AI スピーカのように，新しいデバイスが出てくることにより，セキュリティのあるべき姿も変わると思われる。たとえば，画面が無いデバイスが普通となったときに，どのように異常をユーザに伝えるのか。セキュリティの観点からのユーザインターフェースについても検討すべき。」といった声もあり[112]，トラブル時のユーザとのコミュニケーション方法のあるべき姿が，これまでのパソコンやタブレット等と変わり得ることに留意が必要である[113]。

V　対話型 AI と取引

1　はじめに

新型コロナウイルスの蔓延を踏まえ，これまで対面で行われてきた接客対応について，タブレット等のインタフェースの利用やオンライン化の試みが始まっており[114]，チャットボットによる接客等も脚光を浴びている[115]。また，従前より，スマートスピーカー経由で物を購入する際のトラブルも報告されている[116]。このような対話型 AI と取引の問題をどのように考えるべきだろうか。

---

111) 「社会実験 Exograph，一部の応募者のメールアドレスを BCC でなく CC に入れ誤送信」PR TIMES<https://prtimes.jp/main/html/rd/amp/p/000000004.000051036.html?__twitter_impression=true>。

112) 「今後重点的に取り組むべき研究開発課題について」総務省サイバーセキュリティタスクフォース<https://www.soumu.go.jp/main_content/000666230.pdf>。

113) ただし，現在の対話型 AI は，画面と音声の併用のものが増えていることにも留意が必要である。

114) たとえば，「化粧品販売にオンラインの波　コロナ転機、新しい接客模索」時事ドットコムニュース<https://www.jiji.com/jc/article?k=2020081200789&g=eco>。

115) 「巣ごもり消費でチャット利用約 3 倍、EC 売上 150％に！ 空色に聞く、加速するデジタル接客の CX とは」ECzine 編集部<https://eczine.jp/article/detail/7856>。

116) 「AI が勝手に買い物？家財を破壊？「AI 家電」の注意点を消費者庁がハンドブック作成」FNN プライムオンライン編集部<https://www.fnn.jp/articles/-/65442>。

２　契約主体は AI ではなくユーザ

少なくとも日本法の下において，スマートスピーカー等の対話型 AI を用いてユーザが買い物等の取引を行う場合，契約の申込み又は承諾の意思表示を行っているのはそのユーザであってそのようなユーザの意思表示がスマートスピーカーを通じて行われたと解釈される[117]。

たとえば，AI を契約主体ないしエージェントとする議論は外国では進んでおり[118]，今後は，パーソナル AI エージェントによる支援を受けることが想定されるものの[119]，少なくとも日本では，（対話型）AI 自体が契約主体となるのではなく[120]，（対話型）AI を後ろにいる人間のユーザが道具として用いて取引や交渉を行うという前提で議論が進んでいる[121]。

そこで，以下はそのような前提で議論をするものの，保険会社が AI でテイラーメイドな契約案を提示してきて，顧客側も AI エージェントにその契約案を審査させる世界が意思決定にどう影響するかといった問題も指摘されており[122]，今後はこの前提そのものが変わり得ることに留意が必要であろう[123]。

３　契約が成立していない場合

たとえば，スマートスピーカーの起動のためには，特別なキーワード（ヘイ Siri！アレクサ，OK Google 等）を用いることが多く見られるところ，たとえばバーガーキングのテレビ広告が GoogleHome をハックし，各家庭の GoogleHome がバーガーキングの新商品の説明を読み上げだすといった事件も生じているように[124]，

---

117)　経済産業省ウェブサイト・前掲注 18)。

118)　たとえば，UCITA，UETA 及び UCC 第 2 編の改正案等を参照。

119)　中川裕志「AI 倫理とエージェント」総務省 情報通信法学研究会・AI 分科会 <https://www.soumu.go.jp/main_content/000660995.pdf>。

120)　AI 自体が契約当事者になることについて，その立法論的問題も含め検討したものに村田前掲注 25）・39 頁がある。

121)　たとえば，AI 間自動交渉プラットフォームにおいても最終的には人間の意思決定が必要とされている。江村克己「人工知能間の交渉・協調・連携による社会の超スマート化ーそれぞれの目的の円滑な達成と互恵関係の形成ー」AI ネットワーク社会推進会議 影響評価分科会産業競争力懇談会（COCN）<https://www.soumu.go.jp/main_content/000450499.pdf>,「複数の AI が互いの利害を自動調整するための検証環境が国際業界団体『IIC』から承認」PR TIMES<https://prtimes.jp/main/html/rd/p/000000301.000017036.html>。

122)　宍戸ほか編・前掲注 10）80 頁〔大屋発言〕。

123)　AI と消費者の能力格差を問題とするが，結局現行法では消費者法の枠組みを使うしかないとするものに，村田・前掲注 25）39 頁。

124)　「テレビ CM で Google Home をハック、バーガーキングの奇策：その後の各方面の反

テレビドラマでスマートスピーカーを通じて買い物をしている音声を拾って注文をしてしまうとか，幼児がお菓子をねだる音声を注文と誤解するといったトラブルも考えられる[125]。

このような，ユーザが実際には注文を行っていないケースでは，法律行為としての注文の意思表示はなかったと解釈されるので，スマートスピーカーを通じた契約は成立していないと解されている[126]。ただし，これは，基本契約が存在しないことを前提としており，基本契約が存在する場合には別の結果となる可能性がある[127]。

このようなトラブルを防ぐ必要があるところ，AI スピーカーが認識した注文内容をユーザに通知し，ユーザから確認が得られた場合に注文を確定するという確認措置を講じることが有用であるとされ，具体的には，AI スピーカーが認識した注文内容をユーザに通知し，当該通知に対してユーザから確認を得られた場合に注文を確定する，といった条項を利用規約等に置き，それに基づいた基本契約を締結するという手法が推奨されている[128]。

４　言い間違い等錯誤取消しの場合

では，確かにユーザがスマートスピーカーで A（たとえばタイヤ）を買おうと考えたものの，言い間違いをしたために，B（ダイヤ）を買おうと言ってしまい，事業者が B の売買契約が成立したと考え，商品等を発送した，と言った場合はどうか。この場合，外形的には B についての申し込みと承諾の合致があることから，一応契約は成立するが，この場合には，表示上の錯誤（言い間違え）があるので，錯誤取消し（民法第 95 条）の問題となる。

民法第 95 条は第 1 項で「意思表示は，次に掲げる錯誤に基づくものであって，

応」DIGIDAY 編集部<https://digiday.jp/brands/watch-burger-kings-using-tv-spots-trigger-google-home/>。

125) なお，次の「タイヤを注文しようとしてダイヤと言ってしまった」事案と異なり，「タイヤを注文しようとして実際にタイヤと言ったが，スマートスピーカーの音声認識機能がこれをダイヤだと誤認識した」という事案においては，基本的には，ユーザはタイヤが欲しいという意思を表明しており，意思と表示間に錯誤はないので，こちらの契約不成立の場合として整理すべきであろう。

126) 経済産業省ウェブサイト・前掲注 18)。

127) 福岡真之介編ほか『IoT・AI の法律と戦略』（商事法務，第 2 版，2019 年）221 頁注 18。

128) 経済産業省ウェブサイト・前掲注 18)。一定期間内に回答がない場合に有効な注文とみなすことは，消費者契約法第 10 条の関係で問題となる可能性がある。

その錯誤が法律行為の目的及び取引上の社会通念に照らして重要なものであるときは，取り消すことができる。

一　意思表示に対応する意思を欠く錯誤

二　表意者が法律行為の基礎とした事情についてのその認識が真実に反する錯誤」とする。

ここで，本件は，意思表示に対応する意思を欠く錯誤（民法第 95 条第 1 項第 1 号）といえる。

もっとも，同条第 3 項は「錯誤が表意者の重大な過失によるものであった場合には，次に掲げる場合を除き，第一項の規定による意思表示の取消しをすることができない。

一　相手方が表意者に錯誤があることを知り，又は重大な過失によって知らなかったとき。

二　相手方が表意者と同一の錯誤に陥っていたとき。」とする。

一般的に言い間違えは誰にでも生じ得る。一度の発話があれば，その内容通り発注がなされてしまうようなシステムであれば，重過失があるとされる可能性は低い[129]。実務家も市販品を説明書通り用いた予期せぬ誤作動なら重大な過失はないと議論しているところ[130]，結論において同旨と理解される。

これに対し，上記の，発注内容に誤りがないかを確認する確認措置する措置が組み込まれている場合には，確認措置にもかかわらず錯誤を訂正せずに発注した発注者には重過失があると認定される可能性があるだろう[131]。

なお，発注者が消費者の場合に，電子消費者契約法による規律が及ぶかという問題がある。すなわち，電子消費者契約法第 3 条は「民法第九十五条第三項の規定は，消費者が行う電子消費者契約の申込み又はその承諾の意思表示について，その意思表示が同条第一項第一号に掲げる錯誤に基づくものであって，その錯誤が法律行為の目的及び取引上の社会通念に照らして重要なものであり，かつ，次のいずれかに該当するときは，適用しない。ただし，当該電子消費者契約の相手方である事業者（その委託を受けた者を含む。以下同じ。）が，当該申込み又はその承諾の意思表示に際して，電磁的方法によりその映像面を介して，その消費者の

129)　経済産業省ウェブサイト・前掲注 18)。

130)　第二東京弁護士会情報公開・個人情報保護委員会・前掲注 47）48 頁。

131)　経済産業省ウェブサイト・前掲注 18）参照。

申込み若しくはその承諾の意思表示を行う意思の有無について確認を求める措置を講じた場合又はその消費者から当該事業者に対して当該措置を講ずる必要がない旨の意思の表明があった場合は，この限りでない。

一　消費者がその使用する電子計算機を用いて送信した時に当該事業者との間で電子消費者契約の申込み又はその承諾の意思表示を行う意思がなかったとき。

二　消費者がその使用する電子計算機を用いて送信した時に当該電子消費者契約の申込み又はその承諾の意思表示と異なる内容の意思表示を行う意思があったとき。」として，原則として，意思表示に対応する意思を欠く錯誤について，その錯誤が法律行為の目的及び取引上の社会通念に照らして重要なものであり，かつ，「意思表示と異なる内容の意思表示を行う意思があったとき」（電子消費者契約法 3 条第 2 号）は，重過失があっても原則として取消せるようにする（民法第 95 条第 3 項の不適用）。その趣旨は，操作ミスにより意図しない意思表示を行う恐れが高いからとされている[132]。その上で，その承諾の意思表示を行う意思の有無について確認を求める措置を講じた場合等には，原則通り民法第 95 条第 3 項が適用されるとする。

しかし，電子契約法第 2 条第 1 項の「電子消費者契約」の定義に①「映像面を介して締結される契約」という要件と②「当該映像面に表示する手続に従って消費者がその使用する電子計算機を用いて送信することによってその申込み又はその承諾の意思表示を行うもの」という要件が付されているので，取引がスマートスピーカーによる音声での説明や案内と消費者からの音声での発注で完結する仕組みの場合は，「電子消費者契約」には該当しないといわれている[133]。この点は，電子消費者契約法が平成 29 年に（民法改正に伴い）改正された際に，スマートスピーカー時代に対応する手当がされなかったことが原因と思われるところ，スマートスピーカーにおいても，電子消費者契約法第 3 条の操作ミスにより意図しない意思表示を行う恐れが高いという趣旨があてはまる以上，類推解釈をする余地は否定できないし，仮に解釈論として，同条を適用できないのであれば，迅速に同法を改正すべきである。

---

132) 曽我部ほか・前掲注 64）419 頁。

133) 経済産業省ウェブサイト・前掲注 18)。

５　フラッシュクラッシュ

最後に，フラッシュクラッシュについて若干触れると，高頻度取引（HFT）等，証券市場では，人間ができないような取引を AI が行なっている。ここで，多くの AI が「先読み」を行うところ，たとえば，取引高の低い株に多めの売り注文が入ると，多数の AI が一気に売りに入り，価格が暴落する。このような状況は，フラッシュクラッシュといわれるところ，あまりにも一瞬で行われるので，人間にコントロールされないまま市場に多大な悪影響を与える[134]。現在では高速度取引に関する登録制度が実施されているものの[135]，このような事態に対しては，AI による監視しかない，とも指摘されている[136]。しかし，その AI による監視が適切に行われているかの保証等のむずかしい問題もある。

Ⅵ　対話型 AI のその他の問題

１　行政による対話型 AI の利用

行政による AI の利用一般については，既に先行文献が蓄積している[137]ところ，ここでは 2 つの問題を取り上げたい。

1 つ目は対話型 AI の誤りの問題である。すなわち，行政が利用する対話型 AI の誤りによって，重大な結果が生じ得る。とりわけ，国や自治体が市民に対するアナウンスや市民とのコミュニケーションの方法として対話型 AI を利用する場合，通常の企業等が対話型 AI を利用する場合よりも，市民は行政の利用する対話型 AI をより高く信頼するであろうから，それが一度誤っていると，破滅的効果が生じ得る。その例として，対話型 AI そのものではないが，台風 19 号に関する 2019 年 10 月 12 日の浜松市の事例が有名である。すなわち，「川に避難勧告が出ました」という日本語を，（週末でポルトガル語がわかる職員がいないため，自動

134)　大屋雄裕「行政指導と罪責追及のジレンマ」刑事法ジャーナル 58 巻（2018 年）38 頁参照。
135)　「高速取引行為を行うみなさまへ」金融庁 <https://www.fsa.go.jp/common/shinsei/hst/index.html>参照。
136)　中川・前掲注 74）89 頁。
137)　たとえば，松尾剛行「行政における AI・ロボットの利用に関する法的考察」情報ネットワーク・ローレビュー17 号 92 頁（2019 年）及び松尾剛行「都市行政と AI ロボット活用」久末弥生編『都市行政の最先端』（日本評論社，2019 年）121 頁以下。

翻訳を使ったところ,）「Translation: Foi emiitido aviso para se refugiar proximidades do rio.（川に避難しなさい）」と翻訳し，これをそのままメールで配信してしまったのである[138]。結果として，川に向かって流されてしまったといった被害は出なかったようであるが，今後，行政の利用するチャットボットが，同じような誤った避難指示や避難勧告を行い，取り返しのつかない被害が発生する可能性は否定できない。このような事態に対してどのように対応すべきかは難問である。基本的には，国家賠償法や信義則による保護の可能性といった事後的な対応も考えられるものの，事前の対応として人間のチェックを入れることが推奨されている[139]。とはいえ，上記の浜松市の事案のような，緊急事態における対応として，人間（たとえばポルトガル語ができる職員）が出勤しない限り，ポルトガル語での発信を禁止すべきかというと，その結果として情報発信が遅れれば，ポルトガル語しか分からない市民が，避難勧告を知ることができず，その結果として水難に遭う可能性もある。その意味では，対話型 AI がどの程度の精度（正確性）なのか，対話型 AI を利用しない場合にどのような結果が生じるのか（財産的被害か，人身や人命に関する被害か），どの程度緊急事態なのか（人間のチェックを待てないのか），対話型 AI が誤った場合にどのような結果（財産的被害か，人身や人命に関する被害か）等を総合して判断していくしかないと思われる[140]。

2 つ目は，説明の問題である。現在 xAI といって，その判断過程をできる限り説明しようとする AI が研究されており，そもそも Deep Learning を使わない方法，事後的に人間や別の AI が適切な説明をする方法，できる限り内在的説明を試みる方法等が模索されている[141]。しかし，その説明には限界が存在する。すな

---

138) 菅尾保「『増水の川へ避難を』翻訳ミス，日系ブラジル人に発信」朝日新聞デジタル <https://www.asahi.com/articles/ASMBK4R65MBKUTPB00W.html>。なお，AIの問題ではないが，最近は新型コロナウイルス接触確認アプリ(COCOA)の不具合が報道され（「接触確認アプリ「COCOA」に不具合情報　厚労省が調査」NHK<https://www3.nhk.or.jp/news/html/20200806/k10012555161000.html>），行政以外が活用する分にはメリットも大きいオープンソース型ソフトウェアを行政が利活用する場合には異なる注意点があるのではないかという問題についても議論が始まっている。

139) 松尾・前掲注 137）「行政における AI・ロボットの利用に関する法的考察」102 頁以下，松尾・前掲注 137）「都市行政と AI ロボット活用」130 頁以下参照。

140) もっとも，このような判断を「場当たり的」に行うのではなく，事前にガイドライン等で具体的な事例に応じた対応方法を定め，それに基づき対応することが望ましいだろう。

141) 具体的な説明内容については原聡「『AI の説明』の現状とこれから」総務省 <https://www.soumu.go.jp/main_content/000587311.pdf>がまとまっているが，その後の経緯を含め，原聡「説明可能 AI（Explainable AI）」人工知能学会<https://www.ai-gakkai.or.jp/my-bookmark_vol34-

わち，そもそも Deep Learning という手法を使う限り，内在的説明には量的及び質的限界が存在する。量的にいえば，パラメータが多すぎる，質的というのは深層構造であることにより，たとえば，大量のパラメータで何が関係するかよく分からない中，複数の階層で処理するとたとえば顔写真の男女認識の精度がとても高い AI は二層や三層まではどこを見ているか人間でも理解できるがそれ以上だと人間には理解できなくなってしまう[142]。また，説明の範囲と程度を高めれば高めるほど，時間，費用やコンピュータ・リソース等の資源を要する[143]，要するに，説明と性能等の間にトレードオフが存在する[144]。かかるトレードオフの結果，求められる程度の性能を満たす対話型 AI が存在しても，それが求められる程度の説明を満たさないとして，結局行政で利用できない可能性がある。

そして，行政で求められる説明という意味では，たとえば，囲碁将棋用 AI であれば有効活用が可能である「事後的に人間や別の AI が適切な説明をする」という方法によって説明不足を補うという方法がなかなか利用できないという限界に留意することが必要である。つまり，囲碁将棋では勝ち負けが決まった後で，勝った以上は（その手に対してリソースを割いて分析するだけの意味がある）良い手だったのだろう，という観点から人間や他の AI が事後的に「この手はこういう意味で素晴らしい手だ」と分析し，それを元に「新たな定石」等が発展する可能性がある。囲碁将棋 AI である限り，そのような「後付け」の説明でも何ら問題はないだろう。しかし，行政の利用する（対話型）AI の判断について，事後的に説明できたからといって，それで法律による行政を実現したといえるのだろうか，そのような「後付け」の説明を行政においては説明として利用できないのではないかといった観点からは，政府が利用できる AI に一定の限界があると思わ

---

no4/>，原聡「機械学習における解釈性（Interpretability in Machine Learning）」人工知能学会 <https://www.ai-gakkai.or.jp/my-bookmark_vol33-no3/>，Alejandro Barredo Arrieta et al., “Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI”, Information Fusion, Vol 58, 2020, at https://arxiv.org/pdf/1910.10045.pdf，原聡「機械学習モデルの判断根拠の説明（Ver.2）」<https://www.slideshare.net/SatoshiHara3/ver2-225753735>等を参照。その法的意味については，松尾剛行「事例で考える AI と説明－法的観点から－」日本知的財産協会 AI 分科会 2020 年 2 月 14 日参照。

142) 「人間が深層学習の AI を理解できないのには，理由がある」GLOBE<https://globe.asahi.com/article/12872410>。

143) 原聡「『AI の説明』の現状とこれから」総務省 <https://www.soumu.go.jp/main_content/000587311.pdf>。

144) Alejandro Barredo Arrieta ほか・前掲注 141）。

れる点に留意が必要である[145]。

２　対話型 AI と捜査

対話型 AI は，捜査にも活用される可能性があり，その刑事訴訟法上の扱いが論じられている[146]。

たとえばスマートスピーカーは生活音，場合によっては殺人等犯罪現場の音を拾っており，Amazon Echo の取得した情報の捜査上の利用の可否は Arkansas v. James A. Bates において争われた[147]。たまたまこの具体的事案は本人の同意により解決したが，同意しない場合にどのような手法によりどの範囲で情報を取得できるか。

まず，捜査関係事項照会という方法がある。これは「捜査については，公務所又は公私の団体に照会して必要な事項の報告を求めることができる」とする刑事訴訟法 197 条 2 項を根拠とするが，実務において乱用されているのではないかと懸念が表明されており[148]，ガイドラインの提案[149]等がされている。たとえばツタヤのグループ企業（CCC）は，いわゆる T カードの情報について捜査関係事項照会では提供せず，を令状のみで提供すると決めた[150]。

次に，たとえばスマートスピーカーを令状をもって差押える必要がある場合がある。その結果として，スマートスピーカーの内部記憶装置等に保存された大

---

145) 松尾・前掲注 137）132-133 頁参照。

146) たとえば，松尾剛行「AI・ロボットと刑事法―取得情報とプライバシーを中心に」ビジネス法務 2018 年 2 月号（2017 年）90 頁，稲谷龍彦「刑事司法の最適化と情報技術・ビッグデータの活用――GPS 最高裁判決を超えて――」情報法制研究 3 号（2018 年）3 頁 <https://alis.or.jp/journal/data/vol3/issn2432-9649_vol3_p003.pdf>及び水野陽一「刑事手続における AI 実装と個人情報保護に関する諸問題―刑事捜査・訴追機関の情報収集・処理に関するものを中心に」北九州市立大学法政論集 47 巻 1・2 号（2019 年）75 頁 <https://kitakyu.repo.nii.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=726&item_no=1&page_id=13&block_id=294>等参照。

147) Arkansas vs James Andrew Bates Amazon Echo Search Warrant, at <http://www.documentcloud.org/documents/3473741-Arkansas-vs-James-Andrew-Bates-Amazon-Echo.html>。

148) 指宿信「“捜査事項照会”ってなんだ？：刑事訴訟法学の立場から」第 3 回情報法制シンポジウム（2019 年 6 月 15 日）報告資料 <https://www.jilis.org/events/data/20190615jilis_sympo-ibusuki.pdf>。

149) 「捜査関係事項照会対応ガイドライン」一般財団法人情報法制研究所（JILIS）捜査関係事項照会問題研究タスクフォース <https://www.jilis.org/proposal/data/sousa_guideline/sousa_guideline_v1.pdf>。

150) 「『T カード』の情報，令状でのみ提供　CCC が正式決定」日本経済新聞<https://www.nikkei.com/article/DGXMZO48929530T20C19A8916M00/>。

量のデータをごっそり差押えることが可能となるところ，中身を見ずに網羅的に差し押さえることについては，FD 差押えに関する判例[151]の議論を援用する限り，外部から関連性が不明である場合において，被疑事実に関する情報が記録されている蓋然性が認められるか，及び，そのような情報が実際に入っているか現場で確認したのでは証拠隠滅の危険がある状況か等が問われるだろう。

本体を差し押さえる際において，その本体が接続されたサーバに対してリモートアクセスをしてサーバ内の情報を差し押さえることができるかの問題についてはいわゆる刑事訴訟法 218 条 2 項，99 条のリモート差し押さえとして一定範囲で許容される。

これに対し，パソコンを差し押さえた後，事後的にパスワードを入手して，検証令状を根拠として当該パソコンにデータをダウンロードすることについては，「必要な処分」の範囲を超えているとした裁判例[152]に留意が必要である，要するに検証令状で許されている検証の対象はパソコンであって，データではないということである。そこで，たとえばスマートスピーカーに，本体内の記憶装置に記録された情報を再生する機能のみならず，クラウド上に保存されているデータをダウンロードし，再生する機能が存在するといった場合において，検証令状をもって検証することができる範囲はスマートスピーカーの内蔵記憶装置内のデータに限られ，クラウド上に保存されているデータをダウンロードし，再生することは，違法捜査となることに留意が必要である。

AI クラウドベンダが，たとえば，日本のサーバ上にデータを保存している場合，犯罪に関するデータの証拠収集のため，当該サーバを差押えることは，AI クラウドベンダの業務に重大な支障をきたす可能性がある[153]。この場合の差押えは記録命令付き差押えとして刑事訴訟法 218 条 1 項，99 条の 2 及び本体の差押えに変えて必要なデータを別の媒体にコピーして差し押さえる刑事訴訟法 222 条 1 項，110 条の 2 という方法があり得る。

なお，上記と異なり，クラウドベンダが，たとえば，米国のサーバ上にクラウド上のデータを保存している場合，越境してのクラウドデータベースに対する

---

151)　最決平成 10 年 5 月 1 日刑集 52 巻 4 号 275 頁。

152)　東京高判平成 28 年 12 月 7 日裁判所 HP。

153)　CORE IP Networks 事件（金丸浩二ほか『クラウドセキュリティ』（翔泳社，2014 年）81 頁等）参照。

捜索については，サイバー犯罪条約に基づく対応があり得るが[154]，最近の国際的進展として，米国 CLOUD Act が重要である[155]。

３　対話型 AI による人格と意思決定への干渉

対話型 AI は，人間の意思決定に働きかける。これはある意味では干渉とも評することができ，人の感情を操る AI による説得，誘導[156]の問題が提起されている。

たとえばチャットボットが，自分のプライバシー情報を SNS に投稿する人のプロファイリングを行った上で，その人の最も弱い瞬間に販売を勧誘するといった方向性が懸念される[157]。これは，対話型 AI の問題というよりはプロファイリングの問題と理解すべきであろう。すなわち，ここで問題となる説得・誘導がなぜ（人間よりも）対話型 AI の方がうまくできるのかといえば，それは，（その 1 つ 1 つは些細な情報であっても）プライバシー情報を総合的に分析・処理するプロファイリングを行い，その結果，当初予想できないような知見を導出したり，病気や妊娠といった極めて機微性の高い事項を高い角度で推知することができるからである[158]。

プロファイリングに対しては GDPR が規制を入れている[159]ところ，2020 年 6 月に成立した改正個人情報保護法ではプロファイリング規制が入っていない。

---

154)　諏訪部爽「クラウド時代における捜査機関によるデータの取得―国際法的側面と国内法的側面から―」法律学研究 60 号(2018 年)137 頁 <https://keiolaw.org/wp/wp-content/uploads/2019/06/137-169%E3%80%80 諏訪部%E3%80%80 爽氏.pdf>。

155)　「西村高等法務研究所『CLOUD Act (クラウド法) 研究会』報告書――企業が保有するデータと捜査を巡る法的課題の検討と提言――」西村高等法務研究所 <https://www.jurists.co.jp/sites/default/files/news/pdf/13700_1.pdf>。

156)　宍戸ほか編・前掲注 10）119 頁〔大屋発言〕。

157)　山本・前掲注 105）48-50 頁。なお，「SNS から本性バレた　知能や性格、AI 実験に懸念も」日本経済新聞<https://www.nikkei.com/article/DGXMZO63169460Y0A820C2MY1000/>も参照。

158)　ターゲット社という米国のスーパーマーケットチェーンが，わずかな情報から，高校生の女性の妊娠を予測して広告を送り，妊娠を知らない父親がクレームをつけたものの，実際は妊娠していた，といういわゆるターゲット事件が象徴的である。ビクター・マイヤー=ショーンベルガーほか『ビッグデータの正体　情報の産業革命が世界のすべてを変える』（講談社，斎藤栄一郎訳，2013 年）92-93 頁参照。

159)　石井夏生利「GDPR におけるプロファイリング規制の概要」自正 70 巻 6 号（2019 年）15 頁等参照。

しかし，各分野におけるプロファイリング自主規制の重要性が論じられており[160]，実際に，人事分野における人事データ利活用原則が発表されるといった動きが存在する[161]。

４　対話型 AI と雇用

対話型 AI が雇用を奪うのではないか，という議論もされている。チャットボットや，Google Duplex のような電話応答 AI[162]等の活用によっていわゆるコールセンター業務が奪われる[163]と指摘されている。

このような議論は既に数多くなされている[164]ところ，これらの議論を総括し，①企業内外における教育・訓練制度の拡充，②失業手当を含む再分配制度の整備，③高等教育機関における次世代技術分野の人材育成等の重要性を強調するものがあること[165]には留意しておきたい。加えて，雇用という形態に止まらないフリーワーカー時代の到来に備える必要もある[166]。

Ⅶ　おわりに

Tay で失敗した Microsoft はその教訓を得て，その後新たにたとえば「りんな」等のチャットボットを開発している。少なくとも自動運転等よりは「取り返しの

---

160)　パーソナルデータ＋α研究会「プロファイリングに関する提言案」NBL1137 号（2019 年）66 頁。

161)　「『人事データ利活用原則 第一版』及び『説明映像』配信のリクエスト」一般社団法人 ピープルアナリティクス&HR テクノロジー協会 <https://peopleanalytics.or.jp/2020/03/19/hrdatautilizationprinciples/>。

162)　Munenori Taniguchi「音声 AI の電話予約代行サービス『Google Duplex』、かなりの通話でコールセンターが手助け」Engadget 日本版<https://japanese.engadget.com/jp-2019-05-23-ai-google-duplex.html>。

163)　宍戸ほか編・前掲注 10）228 頁〔佐藤発言〕。

164)　たとえば，大内伸哉『AI 時代の働き方と法―2035 年の労働法を考える』（弘文堂，2017 年）。

165)　廣瀬淳哉「AI 等の技術の雇用への影響をめぐる議論」レファレンス 831 号 56 頁（2020 年）<https://dl.ndl.go.jp/view/download/digidepo_11486060_po_083103.pdf?contentNo=1>。

166)　「フリーワーカーの時代に備えよ　多角的な法政策の必要性」NIRA オピニオンペーパー <https://www.nira.or.jp/pdf/opinion49.pdf>。なお，フリーワーカー時代のプライバシー・個人情報に関し，松尾剛行「フリーワーカー時代における情報管理とプライバシー（仮題）」を公表予定である。

つかない」状況になりにくい対話型 AI[167]においては，できるだけ試行を繰り返して質を高めていくしかないだろう[168]。

なお，本稿は，情報ネットワーク法学会第 19 回研究大会　第 1 分科会「第 5 回ロボット法研究会」AI・ロボットの進化に伴う法と倫理の交錯の「対話型 AI（チャットボット，スマートスピーカー，AI アシスタント等を含む）に関する法律問題」において質問や意見を頂いた。主査の新保史生先生，登壇者の河島茂生先生，久木田水生先生，呉羽真先生，そして中川裕志先生に感謝したい。また，新保先生には，新学術領域研究「人間機械共生社会を目指した対話知能システム学」[169]における対話知能システムの研究開発及び社会実装のための法社会規範の研究班研究協力者として関与させていただき，第 1 回ワークショップ　対話ロボットの社会実装と法律問題に関するワークショップで「An Overview of Legal Issues Related to Communicative AI（対話型 AI にまつわる法的問題の概観）」を発表させていただいた。これらでのやり取りが本稿に結実している。最後に，本稿の完成に向けて，様々なサポートをして下さった編集長兼担当編集者の志田彗様並びに校正を手伝っていただいた桃尾・松尾・難波法律事務所楊燦燦様及び早稲田大学大学院法学研究科博士課程杜雪雯様に心より感謝したい。

以上

167) ただし，ネットハラスメントによる自殺等の取り返しのつかない状況はあり得ることには留意が必要である。

168) 舟山聡「AI 倫理に対する企業の取り組み」NBL1170 号（2020 年）71 頁及び Peter Lee, “Learning from Tay’s introduction”, at, <https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/>。

169) <https://www.commu-ai.org/index.html>。