



経済産業省

AI利活用における 民事責任の解釈適用に関する手引き

[第 1.0 版]

令和8年4月

目 次

第 1 章 はじめに	1
1.1 本書の目的	1
1.2 本書の射程及び留保事項	3
1.3 用語の定義	4
第 2 章 総論	5
2.1 本研究会の検討対象	5
2.2 補助／支援型 AI 及び依拠／代替型 AI	9
第 3 章 補助／支援型 AI に該当するケース	19
3.1 想定事例 1：配送ルート最適化 AI	19
3.2 想定事例 2：弁護士業務支援 AI	22
3.3 想定事例 3：画像生成 AI	25
3.4 想定事例 4：取引審査 AI	37
第 4 章 依拠／代替型 AI に該当し得るケース	48
4.1 想定事例 5：外観検査 AI	48
4.2 想定事例 6：自律走行ロボット（AMR）	53
4.3 想定事例 7：補論—AI エージェント	70
第 5 章 立証や手続に関する論点	74
5.1 立証上の論点	74
5.2 国際的な紛争に関する手続上の論点	78

第1章 はじめに

1.1 本書の目的

生成 AI の登場以降、事業領域における AI サービスの利活用は年々拡大しており、近年では、AI を搭載した IoT 製品やロボティクス等、高度なシステムの開発及び社会実装も進みつつある。また、システムの高度化に伴い、開発や運用のための大規模かつ高品質なデータへのニーズが高まり、事業者間の垣根を超えたデータ連携による新たな付加価値の創出や社会課題の解決も期待されている。

このように、AI やデータの高度利用によってサイバー空間のみならずフィジカル空間への作用が増大しつつある中、政府は、AI がもたらす社会的リスクの低減を図るとともに AI のイノベーション及び活用を促すべく、「AI 事業者ガイドライン¹」（2024 年 4 月第 1.0 版公表、2026 年 3 月第 1.2 版改訂）を取りまとめ、契約上配慮すべき事項については、「AI・データの利用に関する契約ガイドライン²」（2018 年 6 月公表、2019 年 12 月改訂。以下「契約ガイドライン」という。）や「AI の利用・開発に関する契約チェックリスト³」（2025 年 2 月公表。以下「契約チェックリスト」という。）等を取りまとめてきた。また、AI 戦略会議・AI 制度研究会において、AI 制度の在り方に関する「中間とりまとめ⁴」（2025 年 2 月）が策定された後、これに基づき 2025 年 5 月に人工知能関連技術の研究開発及び活用の推進に関する法律（令和 7 年法律第 53 号）が成立し、政府の司令塔としての機能や関係者の責務を定めた。

諸外国の対応に目を向けると、欧州では、2017 年 1 月に欧州議会がロボットや AI の民事責任に関する「ロボティクスにかかる民法規則に関する欧州委員会への提言⁵」を、2020 年 10 月には製造物責任指令（Product Liability Directive）を AI にも適用すること等を提言する「AI 民事責任レジーム⁶」を採択した。その後、2024 年 12 月には改正製造物責任指令⁷が発効し、同指令の対象となる「製造物」に無体物であるソフトウェア等が含まれることとされたため、AI も

¹ 総務省・経済産業省「AI 事業者ガイドライン（第 1.2 版）」（2026 年 3 月 31 日）
（https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_1.pdf、2026 年 3 月 31 日最終閲覧）

² 経済産業省「AI・データの利用に関する契約ガイドライン 1.1 版」（2019 年 12 月）
（https://www.meti.go.jp/policy/mono_info_service/connected_industries/sharing_and_utilization/20200619001.pdf、2025 年 12 月 25 日最終閲覧）

³ 経済産業省「AI の利用・開発に関する契約チェックリスト」（2025 年 2 月）
（https://www.meti.go.jp/policy/mono_info_service/connected_industries/sharing_and_utilization/20250218003-ar.pdf、2025 年 12 月 25 日最終閲覧）

⁴ AI 戦略会議・AI 制度研究会「中間とりまとめ」（2025 年 2 月 4 日）
（https://www8.cao.go.jp/cstp/ai/interim_report.pdf、2025 年 12 月 25 日最終閲覧）

⁵ EU「REPORT with recommendations to the Commission on Civil Law Rules on Robotics」（2017 年 1 月 27 日）
（https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html、2025 年 12 月 25 日最終閲覧）

⁶ EU「Civil liability regime for artificial intelligence」（2020 年 10 月 20 日）
（[https://oeil.europarl.europa.eu/oeil/en/procedure-file?reference=2020/2014\(INL\)](https://oeil.europarl.europa.eu/oeil/en/procedure-file?reference=2020/2014(INL))、2025 年 12 月 25 日最終閲覧）

⁷ EU 加盟国は 2026 年 12 月までに同指令を国内法に反映することが求められている。

製造物責任の対象に含まれることになった。また、2024 年 8 月には AI に対する包括的な事前規制である EU AI 規則（AI Act）が発効し、本書作成日現在では禁止される AI システム等に関する規定及び汎用 AI モデル等に関する規定が施行済みであるほか、2026 年 8 月にはハイリスク AI に関する一部の規定が、2027 年 8 月にはハイリスク AI に関する規制を含めた全ての規制が全面的に施行される予定である⁸。一方で米国では、カリフォルニア州において 2025 年 9 月に AI 安全開示法（Transparency in Frontier Artificial Intelligence Act）が成立したほか、採用 AI のバイアスに関する訴訟や自動運転に起因した交通事故に関する訴訟等が生じており、実務の蓄積を通じた責任の在り方に関する議論が進んでいる。

我が国では、AI がもたらすリスクへの対応として、既存の法体系とソフトウェアを組み合わせで対応し、事業者の自主的な対応が期待できないものに限定して追加的な法的規制を検討する方針をとっていることから⁹、AI に関連する損害等が発生した際には、既存の民事責任ルールに基づく解決が重要になるものと考えられる。一方で、リスクが顕在化し権利侵害や損害が発生した際の民事責任の在り方について、AI の自律性やブラックボックス性を踏まえてどのように解釈適用を行うべきか、裁判例の蓄積や統一的な見解がなく、責任の所在が不明瞭であることが AI の開発や導入を躊躇させる一因となっているとの指摘もあり、AI 事業者ガイドラインに関する検討会においても、AI の開発・提供・利用に関する責任分界についての検討の必要性を指摘する意見が複数寄せられた¹⁰。

以上の諸外国における動向や我が国における課題を踏まえ、今般、AI 利活用における民事責任の在り方に関する研究会（以下「本研究会」という。）を開催し、AI を用いたサービスやシステムが事故に寄与した基本的な想定事例を題材に、主として不法行為法等の観点から解釈適用上の論点及び考え方の整理を行った。本研究会では不法行為法を中心的な検討課題として取り上げており、契約関係等も踏まえた当事者間の厳密な責任分界を明らかにするものではない。もっとも、不法行為法は、契約等の明確なルールが及ばない状況下において各当事者がどのような責任を負うかについての基本的なデフォルト・ルールを提供する。当事者間の契約やガイドライン等によって更なる責任の所在の明確化を図る際にも、不法行為法の観点からの示唆は考慮要素となり得る。

本書は本研究会の議論を取りまとめるものであり、AI 利活用の場面についての不法行為法上の論点を中心に、現行法がどのように適用され得るかの方向性を示し、AI の開発・提供・利用に関わる当事者の予測可能性を高め、AI 利活用の推進及び損害発生時の円滑な解決に資することを目的とするものである。このような目的の下、各想定事例の解説においても、どのような要素が当事者に責任が生ずる可能性を高めるか、あるいは低めるかといった解釈適用の方向性を中心に記述するとともに、可能な限り広い事案の解決の参考となるような考え方を抽出し、また、

⁸ なお、EU Digital Omnibus Regulation Proposal では、ハイリスク AI に関する一部の規定の適用を最大で 2027 年 12 月まで、ハイリスク AI に関する規制を含めたすべての規制の適用を最大で 2028 年 8 月まで延期する方針を示している（EU 「Digital Omnibus on AI Regulation Proposal」（2025 年 11 月 19 日）（<https://digital-strategy.ec.europa.eu/en/library/digital-omnibus-ai-regulation-proposal>、2025 年 12 月 25 日最終閲覧））。

⁹ 前掲注 4・AI 戦略会議・AI 制度研究会 8-10 頁

¹⁰ 総務省・AI ネットワーク社会推進会議（第 30 回）AI ガバナンス検討会（第 26 回）（2025 年 3 月 7 日）（https://www.soumu.go.jp/main_content/001000976.pdf、2025 年 12 月 25 日最終閲覧）

それぞれの考え方が妥当するための前提や射程を示すよう試みている。留意を要する点として、個別具体的な事例に対し現行法がどのように適用されるかを最終的に判断する権限は裁判所に属しており、本書に記載された考え方が裁判所においてそのまま採用されることを保証し得るものではない。しかし、裁判例の蓄積を通じた解釈適用の明確化には一般に長期間を要するところ、有識者の議論に基づき、現時点で可能な限り合理的な考え方や判断が分かれるポイントを示すことで、本書が一つの法解釈の叩き台となり、以て新しいルール形成の一助となることを願っている。

本書はその性質上、法的・技術的に専門性の高い記述を多く含むものの、主な読者としては、AIの開発・提供・利用に携わる事業者、並びにAIの利活用に関連する紛争の解決を目指す当事者及び弁護士等の幅広い者を想定する。

1.2 本書の射程及び留保事項

本書の活用に当たっては、以下の留保事項に留意されたい。

- 本書は、前記 1.1 のとおり AI 利活用の場面における責任の考え方が必ずしも明確でないこと等を踏まえ、現行の不法行為法に基づく解釈適用の考え方を示すものであり、新たなルールを創設するものではない。
- 本書は作成時点における技術動向や法制度の内容を前提としており、これらに変化が生じた際には同様の議論が妥当しない可能性がある。技術動向や法制度の変化に伴い必要が生じた際には、本書の改訂を行う可能性がある。
- 後記第 3 章及び第 4 章の各想定事例について示す考え方はそれぞれの事実関係を前提としたものであり、実際の AI サービスの機能や利用場面、権利侵害の態様等が異なる場合には同様の議論が妥当しない可能性がある。各想定事例は基本的な法律関係の整理のために設定したものであり、実際の事案の解決のために網羅的な視点を提供するものではない。
- 本研究会において議論の対象とした想定事例は全て架空のものであり、実在の AI サービス、事業者又は個人等とは関係がない。
- 本書では、不法行為法に基づく当事者間の基本的な法律関係を明確化するという目的の下、各想定事例では比較的単純な事実関係を想定している。例えば AI 開発者、AI 提供者及び AI 利用者がそれぞれ別の主体である事例や、いわゆるオープンソースの AI を用いる事例等については本研究会の検討対象とはしていない。
- 前記 1.1 にも述べたとおり、個別具体的な事例に対し現行法がどのように適用されるかを最終的に判断する権限は裁判所に属しており、本書に記載された考え方が裁判所においてそのまま採用されることを保証し得るものではない。

1.3 用語の定義

本書で AI に関連して用いる主要な用語の定義は、下表のとおり、いずれも AI 事業者ガイドラインと同様の用語法とする¹¹。

用語	定義
AI	AI を厳密に定義することは困難であるため、AI 事業者ガイドラインと同様、「AI システム（下記参照）」自体又は機械学習をするソフトウェア若しくはプログラムを含む抽象的な概念とする。
AI サービス	AI 事業者ガイドラインと同様、AI システムを用いた役務を意味する。AI 利用者への価値提供の全般を指しており、AI サービスの提供・運営は、AI システムの構成技術に限らず、人間によるモニタリング、ステークホルダーとの適切なコミュニケーション等の非技術的アプローチも連携した形で実施される。
AI システム	AI 事業者ガイドラインと同様、活用の過程を通じて様々なレベルの自律性をもって動作し学習する機能を有するソフトウェアを要素として含むシステムを意味する（機械、ロボット、クラウドシステム等）。
AI モデル	AI 事業者ガイドラインと同様、AI システムに含まれ、学習データを用いた機械学習によって得られるモデルで、入力データに応じた予測結果を生成するものを意味する。
AI 開発者	AI 事業者ガイドラインと同様、AI システムを開発する事業者（AI を研究開発する事業者を含む）を意味する。AI モデル・アルゴリズムの開発、データ収集（購入を含む）、前処理、AI モデル学習及び検証を通して AI モデル、AI モデルのシステム基盤、入出力機能等を含む AI システムを構築する役割を担う。また、AI モデル・システムの開発及び実運用後も、特定領域におけるドメイン知識の拡充や環境の変化への対応、さらに人間の意図や価値観に沿った行動を実現するための調整（アライメント）を目的とした事後学習（Post Training）を通じて、AI モデルの性能を維持・改善することも役割として担う。
AI 提供者	AI 事業者ガイドラインと同様、AI システムをアプリケーション、製品、既存のシステム、ビジネスプロセス等に組み込んだサービスとして AI 利用者（AI Business User）、場合によっては業務外利用者に提供する事業者を意味する。AI システム検証、AI システムの他のシステムとの連携実装、AI システム・サービスの提供、正常稼働のための AI システムにおける AI 利用者（AI Business User）側の運用サポート又は AI サービスの運用自体を担う。
AI 利用者	AI 事業者ガイドラインと同様、事業活動において、AI システム又は AI サービスを利用する事業者を意味する。AI 提供者が意図している適正な利用を行い、環境変化等の情報を AI 提供者と共有し正常稼働を継続すること又は必要に応じて提供された AI システムを運用する役割を担う。また、AI の活用において業務外利用者に何らかの影響が考えられる場合は、当該者に対する AI による意図しない不利益の回避、AI による便益最大化の実現に努める役割を担う。

¹¹ AI 事業者ガイドライン本編 5 頁、9-10 頁

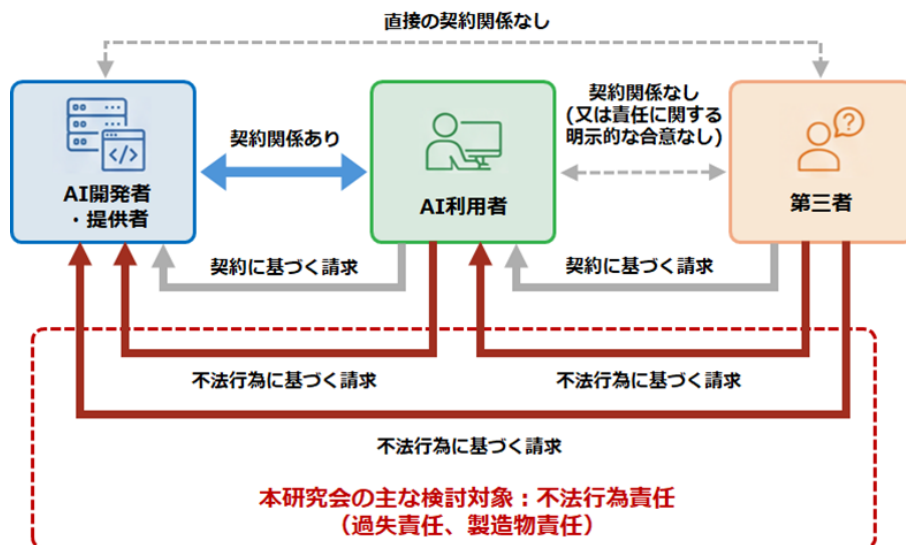
第2章 総論

2.1 本研究会の検討対象

AI の利活用に伴って何らかの損害が生じた際、責任を生じさせる法的な根拠には、主として、契約によらない不法行為責任と、関係当事者間の契約に基づく契約責任がある。これらの責任原因のうち、本研究会では不法行為責任を中心に検討を行った¹²。その理由は以下のとおりである。

- ① 不法行為責任は、契約の有無にかかわらず当事者間の責任を形成するデフォルト・ルールであり、特に契約外の第三者が損害を被った場合には主たる権利救済手段となること
- ② 契約関係にある当事者間においては、契約によって不法行為責任の加重、制限又は放棄等を合意することが可能であるところ、不法行為責任に関する一定の考え方が示されることで、契約に基づく責任や損失の分配に当たっても参考となり得ること
- ③ 契約責任の内容は当事者間の契約条項次第であり、個別性が高いこと。また、AI の開発・提供・利用に関する契約上の留意事項については、契約ガイドラインや契約チェックリストにおいて一定の考え方が示されていること
- ④ 不法行為法も個別具体の事案に即した解釈適用が必要となるものの、事案の類型ごとに争点となる規範や価値判断は共通する場合が少なくないこと

AI の利活用に関係する当事者間の法律関係について、契約に基づく責任も含めた全体像を示すと下記概要図のとおりとなり、第 3 章及び第 4 章に示す想定事例も概ね下図に沿った法律関係が検討される¹³。



¹² 一部の事案においては、不法行為責任と併せて論点となることが多い、AI 利用者と第三者との間の契約関係に基づく善管注意義務や安全配慮義務に関する検討も行っている。

¹³ なお、実際には AI 開発者と AI 提供者が別の主体であったり、同一の主体が AI 提供者と AI 利用者を兼ねていたりする等の様々なパターンが考えられるが、本概要図では省略している。

本書の対象は、専ら上記概要図のうち第三者に損害が生じた事例である。このような場合、第三者は、AI の開発・提供・利用に関わる当事者との間に契約関係がないか、又は契約関係があっても責任の所在に関する明確な規定がないことが少なくないため、不法行為法に基づく請求が重要な選択肢となる。

これに対し、例えば、AI 開発者・提供者と AI 利用者との間で合意された AI の性能が発揮されず、AI 利用者に損害が生じるケースも考えられる。このような場合の法律関係に関しては、AI 開発者・提供者と AI 利用者との間の契約内容に左右されることとなり、個別の契約内容を踏まえた分析が必要となることから、基本的に本研究会では検討の対象としていない¹⁴。

また、不法行為責任の根拠法条には様々なものがあるが、本書では幅広い事案に適用され、AI の利活用の文脈でも特に争点となりやすいと考えられる一般不法行為（民法 709 条）及び製造物責任（製造物責任法 3 条）を主たる対象としている。

2.1.1 一般不法行為

民法上の一般不法行為に基づく責任（民法 709 条）の要件については複数の整理の仕方があるが、一つの考え方として以下のように整理できる。

要件	概要
①過失	行為者が損害の発生を予見し得たこと（予見可能性）を前提として、損害の発生を回避する規範的な義務（結果回避義務）の違反が存在したこと
②保護法益の侵害	身体・財産・人格権等、法的保護に値する利益の侵害が認められること
③損害の発生	被害者に損害が生じていること
④因果関係	①と②との間に事実的因果関係（「あればこれなし」の関係）及び相当因果関係（その行為から通常、その結果が生じるといえる関係。民法 416 条の類推適用と共に用いられることが多い）があること

上記各要件のうち②～④については、AI を利用する場面においても従来の考え方が同様に妥当する場面が少なくない¹⁵。これに対し、AI の利用過程では、特に過失（①）をどのように捉えるかが悩ましい事例が生じ得る。過失は上記のとおり、予見可能性を前提とした結果回避義務違反が認められるか否かにより判断され、このような注意義務の水準は、その行為から生ずる危険の大小、被侵害利益の軽重、その職業・地位に置かれた通常人の注意力等を基に決定される¹⁶。

従来の過失の考え方では人による行為を想定し、当該人の判断や行動について注意義務違反が認められるか否かが焦点となってきた。しかし、AI の重要な特性の一つである自律性（一定

¹⁴ もっとも、後記 4.2 の想定事例 6 における事例 c では、当事者間における明確な合意がないことを前提に、ロボットを利用する事業者から製造業者に対する製造物責任について検討の対象としている。

¹⁵ AI の文脈では過失と権利侵害との間の因果関係の認定が困難になる場合もあり得ることについて、後記 5.1.4 参照。

¹⁶ 加藤一郎編『注釈民法(19) 債権(10)』24-25 頁〔加藤一郎〕（有斐閣、1965 年）

の範囲で人間の介入なく動作する能力を有すること）との関係で、特に人が自ら判断や行動を行っているわけではない場面において、AI 開発者・提供者・利用者の過失をそれぞれどのように捉えるかが重要な論点となる。

また、以上の単独の不法行為に対し、複数の当事者が損害の発生に寄与した場合、共同不法行為（民法 719 条 1 項前段）や幫助責任（同条 2 項）¹⁷の成否が論点となる¹⁸。共同不法行為の要件についても複数の整理の仕方があるが、一つの考え方として以下のように整理でき、下記要件が認められれば各当事者が連帯して損害を賠償する責任を負う。

類型	要件	概要
（狭義の） 共同不法行為	①複数人の不法行為	複数人の行為がそれぞれ不法行為の要件（上表参照）を満たすこと
	②関連共同性	複数人の行為が共同して一つの損害を発生させたこと ¹⁹
幫助	①他人の不法行為	他人の行為が不法行為の要件（上表参照）を満たすこと
	②幫助行為	不法行為を容易にする物理的・精神的援助を行うこと（幫助）

2.1.2 製造物責任

製造物責任法は、前記 2.1.1 に述べた民法上の不法行為法の特則である。製造物の欠陥により人の生命、身体又は財産に損害が生じた場合、過失の有無にかかわらず損害賠償責任を認めることにより、被害者の立証負担を軽減することを目的としている。製造物責任が生ずる要件についても複数の整理の仕方があるが、一つの考え方として以下のように整理できる。

要件	概要
①製造業者等	請求を受ける主体が、製造物を業として製造、加工又は輸入した者その他製造物責任法 2 条 3 項各号のいずれかに該当すること
②製造物	対象となる製品が製造又は加工された動産に該当すること ²⁰
③引渡し	製造業者等が自らの意思に基づいて製造物の占有を移転させたこと

¹⁷ 裁判例上、幫助責任を認めた事例は多くないが、直接の不法行為者に対し重要なツールや役務を提供した者の責任を認めたものとして例えば以下がある。

[1] 最判平成 13 年 3 月 2 日民集 55 巻 2 号 185 頁（カラオケ機器のリース業者がリースの相手方における著作物使用許諾契約の締結又は申込みをしたことを確認しなかったことについて不法行為責任を認めた事例。後記参考裁判例 6）

[2] 東京地判平成 28 年 3 月 23 日判時 2318 号 40 頁（「内職」のために必要であると言われて本人確認資料等を送付し、届いた郵便物を転送する等の作業を行っていた者らに対し、詐欺商法のための口座開設の幫助をしている認識がなくとも、自らの行為が何らかの違法行為に使われている可能性が高いことを容易に知り得たとして損害賠償を命じた事例）

[3] 東京高判平成 29 年 12 月 20 日判時 2384 号 20 頁（投資詐欺業者に事務所を貸与したことをもって過失による幫助を認めた事例）

¹⁸ 民法 719 条 1 項後段の加害者不明のケースにおける共同不法行為や同条 2 項の教唆行為については、本研究会で検討を行った想定事例と関連しないため省略している。

¹⁹ 判例では、各人の行為に客観的な関連共同性があれば足り、共謀等の主観的な関連共同性は不要と解されている（大塚直編『新注釈民法（16）債権（9）』290 頁〔大塚直〕（有斐閣、2022 年）、吉村良一『不法行為法〔第 6 版〕』262 頁（有斐閣、2024 年）等）。

²⁰ 製造物責任法 2 条 1 項

要件	概要
④欠陥	製造物が通常有すべき安全性を欠くこと ²¹
⑤生命、身体又は財産の侵害	他人の生命、身体又は財産を侵害したこと
⑥因果関係	④と⑤との間に事実的因果関係（「あればこれなし」の関係）及び相当因果関係（その行為から一般的にその結果が生じるといえる関係。民法 416 条の類推適用と共に用いられることが多い）があること

日本の製造物責任法においては、製造物責任の対象は動産に限られる（上表②）ことから、AI等のソフトウェアそれ自体は製造物責任の対象とはならないものの、ソフトウェアが製造物に組み込まれた場合には製造物責任が生じ得る。今後、AIを含むソフトウェアによって自律的に稼働する機械やロボット等の普及が見込まれるところ、主要な要件である欠陥（通常有すべき安全性を欠くこと、上表④）をそれぞれの製品の特性を踏まえながらどのように判断するか等が重要な論点となる。

2.1.3 AI ガバナンス及び AI 事業者ガイドライン

上記の責任判断に当たっては、AI の自律性や不確実性等を踏まえたリスクコントロールの在り方に関する AI ガバナンスの考え方が重要な考慮要素となる。我が国においては、AI の開発・利用・提供に携わる事業者を求める AI ガバナンスの内容に関して AI 事業者ガイドラインを策定・公表しており、特に前記 2.1.1 の一般不法行為責任の判断に当たって参考となる²²。

AI 事業者ガイドラインは、国際的な AI ガバナンスの動向も踏まえつつ、我が国の実情に即して事業者を求める AI ガバナンスの基本的な考え方を取りまとめたものである。同ガイドラインの本編において、各事業者が目標の設定に当たって考慮すべき安全性、公平性、透明性等の「共通の指針」を掲げた上、目標の達成に適うようリスクベースの考え方に基づいて AI ガバナンスを構築することを求め、その具体的な手法を別添にて紹介している。AI ガバナンスの構築に当たり、AI は技術進展の速度も著しく、事前にリスクを予測して統制することが困難な場合も想定される。したがって、AI 事業者ガイドラインは、AI の利用環境や用途等を踏まえて想定されるリスクを整理した上で環境・リスク分析を行い、その結果を踏まえて AI マネジメントシステムを設計し、運用の過程で顕在化したリスクやトラブルについても適切にフィードバックしながら当該マネジメントシステムを継続的に評価していくというアジャイルガバナンスの考え方を採用している。

²¹ 製造物責任法 2 条 2 項

²² AI 事業者ガイドラインは AI 一般に関するガバナンスを対象としているため、製造物責任の適用対象となる AI を組み込んだ製造物との関係でも参考とはなり得るが、本書作成時点では、AI 事業者ガイドラインにおいてこのような製品に関する詳細な記述はない。もっとも、この点に関連し、ソフトウェア的知能（AI アルゴリズム）とハードウェア的実体（センサー、アクチュエータ、エッジデバイス等）が統合されたフィジカル AI の動向を踏まえ、AI 事業者ガイドラインの更新に関する検討の一環として、総務省「AI ネットワーク社会推進会議 AI ガバナンス検討会（第 28 回）【資料 4】AI 事業者ガイドラインの更新に向けた論点」17 頁以下（2025 年 12 月 2 日）

（https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02tsushin06_04000132.html、2026 年 1 月 6 日最終閲覧）において、フィジカル AI のリスクや各当事者が必要に応じて行うべきガバナンスの内容等が検討されている。

AI 事業者ガイドラインは、各事業者における自主的な取組みを促すという性質上、そこに記載された事項を全て遵守することが求められるものではなく、また、各場面に即して求められる措置が網羅的に記述されているものでもない。したがって、AI 事業者ガイドラインへの適合や不適合が直ちに過失の有無に結びつくわけではない。しかし、前記 2.1.1 のような過失の判断枠組みに照らし、AI 事業者ガイドラインの考え方も踏まえながらリスクの調査・分析やそのための体制構築等を行っていた場合、それでもなお予見が困難であった特異なリスクが顕在化したケースにおいて、予見可能性を否定する方向の事情として斟酌されるほか、見込まれたリスクの程度に応じて実務上合理的な対応策を講じていた場合、結果回避義務違反と評価される可能性を低める事情として斟酌されると考えられる²³。このような考え方が妥当する具体的な事例については、後記 3.3、3.4、4.1 等において述べる。

2.2 補助／支援型 AI 及び依拠／代替型 AI

本研究会における各想定事例についての検討を踏まえると、各当事者の責任の判断に当たっては、AI が利用される形態に応じた 2 つの類型に整理して検討することが有益と考えられる。

まず、AI 利用者の判断の補助ないし支援としてのみ用いられ、最終的に人の判断や行動を介在させることが予定されている類型（補助／支援型 AI）があり、現状では多くの AI がこの類型に該当すると考えられる。補助／支援型 AI においては、最終的な人の判断や行動の適切性を捉えて、AI 利用者が本来払うべき適切な注意を払いながら AI を用いたか否かを評価することが可能であり、従来の過失の判断枠組み（前記 2.1.1）を適用することができる。

これに対し、近時の AI 技術の進化を踏まえると、必ずしも AI 利用者の最終的な判断や行動が介在することが求められず、人の判断や行動の全部又は一部を代替する前提で提供され、AI の判断に依拠しながら用いることが予定される AI（依拠／代替型 AI）も登場しつつあると考えられる。当該類型では、最終的に生じた権利侵害や損害との関係で、必ずしも AI 利用者の判断や行動が介在せず、これらを捉えた従来の過失判断が困難になるという課題が顕在化する。

また、AI システムや AI サービスを販売又は提供する AI 開発者や AI 提供者が求められる注意義務の水準も、後述のとおり類型ごとに異なると考えられる。

2.2.1 補助／支援型 AI

(1) 補助／支援型 AI に該当する場合

補助／支援型 AI は、AI 利用者の判断の補助ないし支援としてのみ用いられ、最終的に人の判断や行動を介在させることが予定されている類型である。AIの中には、その性質や使用方法に

²³ なお、AI ガバナンスの一環として、AI マネジメントシステムに関する国際標準に基づく認証を活用することも手段の一つになると考えられる。例えば、AI 利用者が AI の適切な利用体制を構築する際、AI 開発者・提供者における ISO/IEC42001 等の AI マネジメントシステムに係る認証の取得状況を確認するとともに、導入する AI 製品、AI サービス又は AI システムの性能、限界及びリスク等に関する AI 開発者・提供者の評価結果を取得し、自社の業務要件への適合性や当該利用環境における妥当性を確認していた場合、実務上合理的に可能な対応策を講じていた事情の一つとして斟酌され得る。もっとも、AI 事業者ガイドラインを踏まえた AI ガバナンスの構築と同様、このような措置が直ちに過失の有無に結びつくものではない。

照らし人の判断や行動を代替することになじまず、補助／支援型 AI として用いるべきと評価される場合がある。このような場合に該当するか否かは AI の技術的特性やその動向、規制法の内容等にも左右されるが、本書作成時点では、①AI の機能や利用される場面を踏まえると人の判断を代わりに行っているとはいえないケース、②規制法上の理由により人の最終的な判断が要求されるケース、③AI の出力内容が潜在的に第三者の権利を侵害するリスクを内包しており、この点について人の評価や検証が必要なケース等が考えられる。それぞれの具体例については後記第 3 章において述べる。

上記①～③に該当しない場合には、補助／支援型 AI にも依拠／代替型 AI にも該当する可能性があるが、後記 2.2.2 に述べる基準に照らし依拠／代替型 AI に該当しないものは一般的に補助／支援型 AI に該当する。

(2) 補助／支援型 AI を利用する AI 利用者の責任

前述のとおり、補助／支援型 AI では最終的に人の判断や行動が介在するため、これらを捉えて従来の過失の判断枠組み（前記 2.1.1）を適用することが可能である。AI 利用者はあくまで自らの判断の補助や支援として AI を用いるべきこととなり、その責任は AI の利用如何にかかわらず、個々の局面において AI 利用者が本来払うべき注意の下で適切な判断や行動を取ったか否かにより判断される。AI の出力を援用したことは、原則として、AI 利用者が求められる注意義務の水準を引き上げることも引き下げることもない²⁴。

上記の一般的な注意義務の下、AI 利用者は最終的な判断や行動において適正な範囲で AI の出力を利用すべきこととなるが、AI の特性やリスクの性質に応じて過失の判断上重要な要素は異なり得る。典型的には、AI 利用者において AI の出力を踏まえつつその正確性や適切性を評価しながら AI を用いることが求められるが（後記 3.1・想定事例 1 及び 3.2・想定事例 2）、AI の出力を一見するだけではリスクの所在が不明瞭であり、AI 利用者において検証・是正することが必ずしも容易でないユースケースもある（後記 3.3・想定事例 3 及び 3.4・想定事例 4）。後者のような場合には、最終的に AI の判断を利用するに際して第三者に対する権利侵害や損害を生じないよう、AI 利用者において事前に一定の情報収集を行うことや必要な利用上の措置を講ずることが注意義務の内容となり得るほか、AI 開発者・提供者においても一定の設計上の措置を講ずるべきでなかったかが論点となる（下記(3)参照）。

(3) 補助／支援型 AI を販売又は提供する AI 開発者・提供者の責任

補助／支援型 AI は、人の最終的な判断や行動を補助ないし支援するため、入力データや指示（プロンプト）を分析した結果を踏まえて一定の出力を行う情報処理システムと位置づけられる。このようなシステムにおいて、AI が不適切な出力をした場合であっても、最終的に AI 利用者において AI の出力の適切性が検証・是正されることが前提となるため、補助／支援型 AI を販売又は提供する AI 開発者・提供者が第三者との関係で責任を負う場面は限定的となる。

²⁴ ただし、本研究では、特に高精度な AI システムを用いる場合に注意義務の内容に影響を及ぼすという考え方に関して検討を行うことも重要であるとの指摘があった（後記 3.2.2 参照）。

不法行為法における従来の議論においても、権利侵害との関係で他人の判断が介在した場合、直接の原因となったのは当該他人の意思決定であることを理由に、原則として因果関係が否定されると考えられてきた²⁵。また、従来型の情報処理システムの一例であるカーナビゲーションシステムに関し、最終的には情報の受け手であるドライバーにおいて独立に安全性を判断すべきことを理由に製造者の責任を否定した事例として、下記参考裁判例 1 がある。



参考裁判例 1：福島地判平成 30 年 12 月 4 日判時 2411 号 78 頁

- 原告が被告らの製造したカーナビのルート案内に従って走行したところ、樹木がせり出した狭い道路に進入し車体に擦過傷を生じたため、被告らに対し不法行為及び製造物責任に基づく損害賠償を請求した事案。
- ①カーナビシステムが全国の道路の正確な状況をリアルタイムで情報提供するのは不可能か著しく困難であること、②個々の道路の安全性については現に直面する運転者が最も把握し得ること、③カーナビの画面や取扱説明書において実際の道路状況等に従って走行すべき警告等を行っていること、④事故現場の道路は国土地理院の地形図に「軽車道」として掲載されており、およそ車両の通行できない道路を収録していたものではないこと等の理由から、原告は本件カーナビのルート案内に依存せず、自らの判断に基づき本件道路を走行しなければならない。そのため、ルート案内と擦過傷との因果関係は否定され、被告らの責任は認められない。

上記の議論及び裁判例のように、最終的に人の判断や行動の介在を予定したシステムについては、システムの出力と権利侵害や損害との間の因果関係が原則として否定されると考えられるほか、下記の観点から説明上・設計上の適切な措置を講じた上で提供されている場合、過失も認められないと考えられる。

AI の利用場面においては、AI 利用者がその性能限界等を適切に理解し、過度なリスクを惹起することなく利用できることが重要となる。この点に関連し、従来の裁判例上、他人が用いる製品の開発者や提供者が責任を負い得る場面として、製品の利用過程での損害の発生を誘発した場合の責任が認められることはあり、具体的には、①利用者が製品を適切に用いるため重要な事項について説明を行わなかったことにより誤使用が生じた場合の責任²⁶や、②利用過程における権利侵害の高度の危険性が見込まれたにもかかわらず、適切な権利侵害防止措置を講じることなく製品を提供した場合の責任²⁷等が論じられてきた。これらの観点から、AI 提供者や AI 開発者の責任が論じられる余地はある。

補助／支援型 AI との関係で相対的に重要となるのは、上記①の説明義務と考えられる。個々の AI の性能限界や使用方法是 AI 利用者において必ずしも自明ではなく、また、説明の内容によっては AI 利用者に機能や性能の限界が適切に伝わらず、AI に対する過度な信頼を生じさせてしまう可能性もある。裁判例上、説明義務の認定に当たっては、リスクの性質や程度、当事者間における情報の偏在の程度、利用者の属性等を踏まえながら検討されており、AI 開発者・提

²⁵ 窪田充見編『新注釈民法(15) 債権(8) 〔第 2 版〕』387-388 頁〔橋本佳幸〕(有斐閣、2024 年)

²⁶ 例えば、医療機器についての重要な事項に関する説明の欠缺に関する事案として、東京高判平成 14 年 2 月 7 日判タ 1136 号 208 頁。なお、当該裁判例は人工心肺装置中の送血ポンプに関する製造者の責任が争点となったものであるが、製造物責任法施行前のものであり、裁判所は一般不法行為責任の下で製造業者の説明義務違反を認定している。

²⁷ 例えば、著作権侵害の高度の蓋然性が見込まれた状況下でのカラオケ機器のリースに関する事案として、最判平成 13 年 3 月 2 日民集 55 巻 2 号 185 頁(後記参考裁判例 6)。

供者は、こうした要素を勘案しつつ、AI の機能や性能の限界、使用方法、重要なリスク等について、当該時点の科学技術の水準に照らし合理的に予見可能な範囲で明確な説明を行うことが重要と考えられる（説明上の注意義務）²⁸。このような説明を行っている限り、AI に不適切な出力があったとしても、AI 利用者がこれを検証・是正することが当事者間の役割分担としても予定されるため、AI 開発者や AI 提供者には原則として過失は認められない。

上記②の責任が問題となるのは、基本的に、AI の機能や性質、AI 利用者の特性等を踏まえ、AI 利用者においてリスクを具体的に予見したり、その出力をコントロールしながら用いたりすることが必ずしも容易でない場合に限られる。このような場合には、AI 開発者・提供者において一定の権利侵害防止措置等を講ずるべきでなかったかが問題となる（設計上の注意義務。具体例として後記 3.3.4 及び 3.4.4 参照）。

なお、AI 開発者や AI 提供者に求められる設計上の注意義務及び説明上の注意義務の内容は、バリューチェーンにおけるそれぞれの役割分担に応じて異なり得る。

2.2.2 依拠／代替型 AI

前述のとおり、依拠／代替型 AI は従来の考え方をそのまま適用することが困難な類型であるところ、補助／支援型 AI とは異なる責任判断の方向性が可能か、また、どのような場合がこれに該当するかが問題となる。

（1）依拠／代替型 AI に該当する場合の責任判断の方向性

① 人の操作や介入を前提としないシステムに関する裁判例

従来も人の業務を効率化するための様々な機械やシステムは存在してきたが、特に第三者の権利利益に影響を与える場面では、主として人の操作や介入を前提としたものが多かった。これに対し、AI の技術の進化により、第三者の権利利益に関係する領域においても、人の操作や介入がなくとも高い性能や安全性を発揮するシステムが登場しつつある。

このような場合にどのように AI 利用者の過失判断を行うべきかについては、直接の裁判例や確立した考え方がなく悩ましい問題であるが、個別の処理において人の操作や介入を前提としないシステムに関する裁判例として以下がある。

²⁸ 本研究会では、AI が基本的に学習用データ、ユーザーが入力したプロンプト及び RAG 等で参照される外部データベースの 3 つの要素によって規定されていることを踏まえると、AI サービス提供者が説明義務違反のリスクを低減する観点では、AI がどのようなデータで学習されており、どのデータベースを参照することで、どのような範囲の判断が可能で、どのような限界があるのかという事項を説明することが重要であるとの指摘があった。常に各要素についての説明が求められるものではなく、本文に述べた説明義務一般に関する考慮要素（リスクの性質や程度、当事者間における情報の偏在の程度、利用者の属性等）を踏まえ合理的な範囲の説明が求められると考えられる。最終的にはケース・バイ・ケースの判断とならざるを得ないと考えられるが、一つの考え方として、その業界における一般的な知識に照らし、説明が妥当であったか否かを判断するというアプローチが考えられるとの指摘もあった。



参考裁判例 2：無権限者による ATM の預金引出し(1) (最判平成 5 年 7 月 19 日集民 169 号 255 頁)

- 無権限者が預金通帳又はキャッシュカードを使用し暗証番号を入力して ATM から預金の払戻しを受けた場合、銀行の ATM で真正なキャッシュカードが使用され、正しい暗証番号が入力されていたときは、銀行側に暗証番号管理の不備など特段の事情がない限り、約款による免責は有効である。
- 当時用いられていたキャッシュカードの磁気ストライプ上には届出暗証番号がコード化されて記録されており、市販のカードリーダー等を用いて暗証番号を解読することができたが、このような方法で暗証番号を解読するためにはコンピューターに関する相応の知識と技術が必要であることを踏まえ、免責約款の効力を否定しなければならないほど安全性を欠くものとはいえない。



参考裁判例 3：無権限者による ATM の預金引出し(2) (最判平成 15 年 4 月 8 日民集 57 巻 4 号 337 頁)

- 無権限の払戻しについて銀行が無過失であるというためには、機械払の方法により預金の払戻しが受けられる旨を預金者に明示すること等を含め、ATM を利用した預金の払戻しシステムの設置管理の全体について可能な限度で無権限者による払戻しを排除し得るよう注意義務を尽くしていたことを要する。

上記各裁判例は不法行為法に関する事案ではないものの、民法 478 条の「過失がなかったとき」という要件に関連する判示であり、不法行為法上の過失の判断に当たっても参考となる。

参考裁判例 2 及び 3 を通じて最高裁が示した考え方は以下のように整理される。預金払戻しシステムにおいては、真正な預金通帳やキャッシュカードとその暗証番号の両方が他人の手に渡った場合、無権限者による預金の引出しが可能となるリスクがあるが、その可能性が例外的なものといえる場合、誤払いが生じたとしても直ちに銀行に過失があるとは認められない²⁹ (参考裁判例 2)。その上で、銀行に過失がないといえるためには、預金払戻しシステムの設置管理の全体について、可能な限度で無権限者による払戻しを排除し得るよう注意義務を尽くしていたことを要する (参考裁判例 3)。このような判断枠組みは、銀行が運用するインターネットバンキング等のシステムに関する裁判例でも用いられている³⁰。

上記の裁判例の考え方は、人の関与がなく出金処理等のプロセスが行われる預金払戻しシステムに関し、当該システムが一定の安全性を有していることを前提として、人の判断や行動についての過失を問題とするのではなく、システムの設置や運用の過程を捉えて過失判断を行うものと評価することができる。

依拠／代替型 AI も、最終的に人の判断や行動が介在せずにシステムが業務を処理する点において預金払戻しシステムと変わりはないことから、依拠／代替型 AI を利用する際の責任判断との関係でも上記の判断枠組みを援用することができると考えられる。すなわち、依拠／代替型 AI では、最終的な人の判断や行動の適切性が論点となる補助／支援型 AI とは異なり、AI 利用者が

²⁹ 参考裁判例 2 は直接的には銀行と預金者との間の免責特約の有効性について判断を示しているが、このような免責特約は民法 478 条の無過失要件を具体化した規定と考えられているため、実質的には銀行の過失の有無を評価した判例と捉えられる (吉田光碩「判批」判タ 1124 号 88 頁 (2003 年))。

³⁰ 例えば東京高判平成 18 年 7 月 13 日金法 1785 号 45 頁は、インターネットバンキングを通じて発生した誤振込みについて、振込送金時に暗証番号等が一致している場合には、銀行による暗証番号管理が不十分であった等の特段の事情がない限り免責されるとしつつ、銀行がシステム全体として可能な限度で無権限者による振込を排除し得るよう構築・管理していたと評価して免責条項の適用を認めた。

AI システムを組み入れた業務プロセスを適切に構築しているか、AI の望ましくない出力が権利侵害や損害を生じさせる可能性を低減するため合理的に可能な運用を行っているかが過失判断の対象となる。

② 依拠／代替型 AI に該当するための要件

参考裁判例 2 及び 3 では、預金払戻しシステムが、預金者の財産の出入金を処理するという目的との関係で一定の信頼を置けるものであることが前提となっていた。このような責任判断を依拠／代替型 AI との関係でも援用するにあたり、どのような AI であれば判断を委ねることが合理的といえるかが重要となり、実際の事案ではこの点が争点となることが見込まれる。

前記 2.2.1(1)のとおり、AI の中には、その性質に照らし補助／支援型 AI として用いるべき場合がある。そして、これらの場合に当たらない AI は、補助／支援型 AI にも依拠／代替型 AI にも該当する可能性があるが、依拠／代替型 AI に該当するかについては、個別の事案において、以下の 2 つの要素を踏まえて判断されることが考えられる。

(a) 必要性

まず、依拠／代替型 AI に該当するためには、人の最終的な判断や行動を介在させず AI に判断を委ねることの必要性が求められる。具体的には、上記(1)の預金払戻しシステムのように人が全件を遂行することが困難な作業量を AI が処理したり、人だけでは作業が困難な時間的・物理的領域において AI が自動的に処理を行うなど、AI 利用者において人による判断や行動を介在させることでは実現困難な効用が見込まれることが求められる。このように AI に判断や行動を委ねる必要性がない場合、人が必要な注意の下で AI の出力を検証・是正しながら AI を用いることが可能であり、人の判断や行動を介在させることによる結果回避義務が生じ得ると考えられる。

(b) 精度及び安全性

次に、依拠／代替型 AI に人の判断や行動を委ねるにあたり、当該 AI がどのような精度や安全性を備えているべきかが重要な論点となる。抽象的には、社会的に期待される精度や安全性の水準を満たすべきと考えられ、例えば規制法や業界の安全基準に適合したシステム、あるいは社会において導入が進み一般的に用いられているシステムであれば、依拠／代替型 AI として期待される精度を備えていると評価しやすい。

AI を用いたシステムが普及途上にある現状では、どのようなものが依拠／代替型 AI に該当し得る程度の精度や安全性を備えているかについて明確な基準はないが、現時点においても十分な精度や安全性を備えた AI システムはあり得ると考えられる。不法行為法の下では、権利侵害や損害が生じた場面において、それを生じさせた人の判断や行動に関し、その行為に伴うリスクの大きさ等を踏まえつつ、当該人物が置かれた状況や地位における通常人の注意力を基準とし、これを下回る判断や行動が行われた場合には過失を認定することで損害の公平な分担を図ってきた（前記 2.1.1）。このことからすれば、人が行ってきた業務に対して AI を導入する際も、同種業務における通常人の作業水準を基準とすることが考えられ、AI がその品質及び精度において従来の通常人による作業水準と同等以上と評価できる場合や、権利侵害リスクが従来の通常

人による作業水準と比しても十分に抑制されている場合、人の判断や行動の全部又は一部を AI に委ねつつ運用することの合理性（依拠／代替型 AI に該当すると判断することの合理性）が認められると考えられる^{31, 32}。

依拠／代替型 AI への該当性を基礎づける実務上の方法としては、AI 開発者・提供者によるベンチマーク評価、性能限界や誤出力パターン、テスト環境と実運用条件との差異等を踏まえ、AI を導入する業務に求められてきた作業水準や安全性に適合することの確認や、実運用環境における追加検証等を必要に応じて行うことが考えられる。

また、AI システムが業務プロセスの一部でのみ通常人と同等以上の精度水準を発揮する場合や、特定の条件下で精度が低下し通常人の作業水準に満たなくなる場合等が想定される。このような場合、例えば一定の範囲で人や別のシステムが AI の出力を評価・是正したり、AI が予定された精度を発揮するための環境を構築するなど、人の関与、業務プロセス、既存の IT 基盤等と組み合わせられることで、プロセス全体としての業務システムが人と同等以上の作業水準を発揮する場合であれば、人の判断を一定の範囲で AI に委ねつつ運用することの合理性が認められ得る。今後の技術の進展に伴い、AI システムが業務プロセス全体を汎用的に担うことができる場面は増加する可能性があるものの、現時点においてはそのようなユースケースは限定的であり、上記のように部分的に人の判断や行動を代替したり、特定の環境下で自律的に動作するシステムの方が求められる水準を満たすと評価されやすいと考えられる。具体例については後記 4.1 及び 4.2 において言及する。

上記の 2 つの要件を満たし、人の判断や行動の一部を代替する前提で AI が提供されている場合には、依拠／代替型 AI として過失判断を行うべきと考えられる。

(2) 依拠／代替型 AI を利用する AI 利用者の責任

上記(1)に述べた機械払いシステムに関する考え方を敷衍すると、依拠／代替型 AI を利用する AI 利用者の責任は、上記(1)の 2 つの要件を備えていることを条件として、人が合理的な判断や行動を行う注意義務から、AI システムを組み入れた業務プロセスを適正に構築するとともに、リスクを可能な限り低減しながら運用を行う義務へと転換する。

① AI システムを組み入れた業務プロセスの「構築」に関する注意義務

依拠／代替型 AI を導入する場面では、業務プロセスにおける人の判断や既存 IT システム等の一部を AI システムに置き換えることを検討することになる。この点に関する AI 利用者の注

³¹ 依拠／代替型 AI の導入が進んでいない領域や普及途上にある領域では、本文に述べたとおり従来の通常人の作業水準が当面の比較対象となるが、AI その他の自律的に稼働するシステムの普及に伴い、より高精度なシステムを用いることが社会的に期待されるようになった場合、信頼性の高い同種システムが比較対象となり得ることに留意が必要である。

³² 第三者の権利保護の観点からは、本文に述べたとおり AI の作業水準が第一次的な評価基準となると考えられるが、AI の精度と通常人の作業水準とを比較することが一概には難しい場合もある。その際、生命や健康被害といった重大な法益侵害の場合を除いては、AI によりもたらされる業務効率化等の便益とリスク（権利侵害や損害が発生する蓋然性及びその重大性）とを比較衡量し、前者が上回る場合、依拠／代替型 AI として用いるための要素として付随的に考慮することも考えられる。参考になる議論として、抽象的危険を伴うが社会的有用性のある行為に関し、危殆化段階における過失の行為義務（予見義務、情報収集義務）において比例原則的考慮を行う考え方について、大塚直ほか『民法 6 事務管理・不当利得・不法行為』（有斐閣、2023 年）140 頁等参照。また、関連する規制法上の要請を満たすことにより、必要な安全性を備えていることの要素として付随的に考慮することも可能と考えられる。

意義務は、上記(1)の考え方に照らして対象となる業務プロセスが AI による自動化に適しているか、望ましい水準や安全性を備えた AI システムを用いているかといった観点から評価される^{33, 34}。

また、AI のリスクを踏まえて適切な利用体制やガバナンスの構築を行っているか否かも重要な要素となる。一例として、上記(1)に挙げたように、AI システムが業務プロセスの一部でのみ通常人と同等以上の精度水準を発揮する場合や、特定の条件下で精度が低下し通常人の作業水準に満たなくなる場合等に、人の関与を介在させたり別のシステムと統合したりすることにより、業務プロセス全体として望ましい精度や安全性を発揮すべき場合がある。この場合、AI 利用者は、そのような精度や安全性を満たすために必要な人の関与や別のシステムを検討し、必要な人的・物的設備を導入する注意義務を負うと考えられる。

② 構築した業務プロセスの「運用」に関する注意義務

上記①の業務プロセスの構築を前提とした上で、AI の望ましくない出力により権利侵害が生ずる可能性を低減させるため、AI 利用者がシステムの運用において合理的な措置を講じていたかについても過失判断の対象となり得る。一例として、AI は継続学習やアップデートにより事後的に精度等が変化する可能性があるところ、運用過程において精度が低下することにより依拠／代替型 AI としての運用に適しない状態となっていないか確認を行うこと等が考えられる。

上記①及び②に関し、具体的にどのような措置を講ずべきかについては個々の AI システムによって異なるところ、AI 事業者ガイドラインにおいては、AI の利活用に伴うリスクをコントロールするためのガバナンスを求めており、その内容を参考にして合理的に可能な措置を講ずることが考えられ、その具体例は後記 4.1.2 において述べる。なお、前記 2.1.3 のとおり、AI 事業者ガイドラインは各事業者における自主的な取組みを促すという性質上、そこに記載された事項を全て遵守することが求められるものではなく、各場面に即して求められる措置が網羅的に記述されているものでもないから、AI 事業者ガイドラインへの適合や不適合が直ちに過失の有無に結びつくわけではないことに留意が必要である。

上記①及び②において AI 利用者が適切な措置を講じていた場合であっても、AI が望ましくない出力を行う可能性はあり、これを是正せず権利侵害や損害が生じたことについて、AI 利用者の結果回避義務違反が争点となり得る。しかし、依拠／代替型 AI を運用する場面では、AI システムが一定の要件を満たすことを前提として、当該システムを組み入れた業務プロセスの適正な構築及び運用を行っていた場合、前記 2.2.2 に例示した機械払いシステムと同様に、システムの利用者として果たし得る合理的な注意義務は履行していたものと評価される。換言すれば、依拠／代替型 AI においては、上記①及び②において AI 利用者が適切な措置を講じていれば、AI の不適切な出力の全てを AI 利用者が検証する等の方法で是正するまでの結果回避義務は認められない。

³³ AI は継続学習やアップデートにより事後的に精度等が変化する可能性があるところ、運用過程において精度が低下することにより依拠／代替型 AI としての運用に適しない状態となっていないかについての注意義務も生じ得る。

³⁴ 前掲注 31 のように、今後依拠／代替型 AI の普及が進み、導入の候補となる AI が複数ある場合には、より望ましい性能を備えた AI を用いているかという観点からの注意義務も生ずる可能性がある。

(3) 依拠／代替型 AI を販売又は提供する AI 開発者・AI 提供者の責任

AI 開発者や AI 提供者については、責任判断の視点自体は補助／支援型 AI と異ならず、求められる設計上・説明上の措置を行っていたかが論点となるが、その水準は補助／支援型 AI よりも高まると考えられる。

具体的には、依拠／代替型 AI では最終的な人の判断や行動が介在せず、AI の出力が直接的に権利侵害や損害に結びつき得ることを前提に、上記(2)に述べた水準や安全性を発揮・維持するため合理的に可能な設計上の措置やアップデートを講じていたかが問題となる（設計上の注意義務）。必要となる措置は個別具体的な状況によらざるを得ないが、例えば、開発上合理的に可能な範囲で AI の精度や安全性を向上させていたか、一定の範囲で望ましくない出力が生ずることを前提に合理的なセーフガードの構築を検討したか等が論点となり得る。また、上記(2)に述べたとおり、依拠／代替型 AI では業務プロセスの中でどのように用いるべきかという観点も重要となるところ、リスクコントロールの上で重要な情報（例えば、AI の機能や性能の限界、使用方法、重要なリスク等に加え、AI システム単体では対応し難い外的な危険源や、どのような状況下でリスクが増加し得るか、どのような範囲で人が関与することが望ましいか等）を、AI 開発者・提供者が合理的に予見可能な範囲において必要に応じて分析し、AI 利用者に対し説明すること等が求められる（説明上の注意義務）。

なお、前記 2.2.1(3)と同様に、AI 開発者や AI 提供者に求められる設計上の注意義務及び説明上の注意義務の内容は、バリューチェーンにおけるそれぞれの役割分担に応じて異なり得る。

2.2.3 各類型に関する小括

以上の考え方を総合すると、補助／支援型 AI と依拠／代替型 AI のいずれに該当するかは、AI の機能や利用される場面、AI の精度や安全性、業務上のリスクの程度や専門性、関連する規制の有無等を上述の観点から考慮して決せられるべきと考えられる。実際の事案において判断が悩ましいケースもあると考えられることから、本書では第 3 章及び第 4 章における各想定事例の検討を通じ、いずれに該当するかを示している。

また、補助／支援型 AI 及び依拠／代替型 AI の概要及び責任判断の方向性を改めて整理すると下図のとおりとなる。

	補助／支援型 AI	依拠／代替型 AI
概要	- AI が判断の補助ないし支援としてのみ用いられ、最終的に人の判断や行動を介在させることが予定されている類型。 - 以下①～③のように補助／支援型 AI としてのみ用いるべき場合がある。 ① AI の機能や利用場面を踏まえると人の判断を代わりに行っているとはいえないケース ② 規制法上の理由により人の最終的な判断が要求されるケース ③ AI の出力内容が潜在的に第三者の権利を侵害するリスクを内包しており、この点について人の評価や検証が必要なケース	- 人の判断や行動を代替する前提で提供され、AI の出力に依拠しながら用いることが予定されている類型。 - 以下の 2 つの要件が求められる。 (a) 人による判断や行動を介在させることでは実現困難な効用が見込まれること（必要性） (b) AI が一定の精度や安全性を備えていること（精度及び安全性） → 個々の業務によって求められる水準は異なるが、<u>同種業務における通常人の作業水準と比較して同等以上の精度や安全性[※]を備えている場合</u>、AI に判断を委ねることの合理性が認められると考えられる。

責任判断の方向性	AI 利用者の責任	- 個々の状況下で AI 利用者が本来払うべき適切な注意の下で AI を用いることが求められる[※]。 - 具体的には、AI の出力の正確性や適切性を評価しながら AI を用いること、そのために必要な情報収集を行うことや利用上の措置を講ずること等が注意義務の内容となり得る。	- AI 利用者の判断が介在するわけではないため、注意義務の対象は、適切な判断や行動を行うことから AI システムを組み込んだ業務プロセスの適正な構築及び運用へと転換する。 - AI の不適切な出力の全てを AI 利用者が検証する等の方法で是正するまでの結果回避義務は認められない。
	AI 開発者・提供者の責任	- 性能限界や重要なリスク等についての説明を行うことにより、AI の出力の適切性は AI 利用者が是正・検証することが前提となる。 - AI 利用者による予見・対処が容易でないリスクについて一定の設計上の措置が求められ得る。	上記の安全性を発揮・維持するため合理的に可能な設計上の措置や、リスクコントロールの上で重要な情報を分析し AI 利用者への情報提供を行う等の説明上の措置が求められる。

※ 個々の局面において求められる注意義務の水準は、その行為から生ずる危険の大小、被侵害利益の軽重、その職業・地位に置かれた通常人の注意力等に基づいて決定される。なお、AI と人の作業水準を直ちに比較することが困難な場面の考え方について、前掲注 32 参照。

第3章 補助／支援型 AI に該当するケース

本章では補助／支援型 AI に該当するケースを取り上げる。現状では多くの AI が補助／支援型 AI に該当すると考えられるが、その実質的な根拠としては、前記 2.2.1(1)にも述べたとおり、①AI の機能や利用される場面を踏まえると人の判断を代わりに行っているとはいえないこと、②規制法上の理由により人の最終的な判断が要求されること、③AI の出力内容が潜在的に第三者の権利を侵害するリスクを内包しており、この点について人の評価や検証が必要であることが挙げられる。以下の想定事例 1～4 は、このような観点から補助／支援型 AI に該当する。

補助／支援型 AI では、AI 利用者が本来払うべき注意の下、AI の出力を適正な範囲で用いることが求められるが、注意義務の判断において重要となる要素は、AI の特性やリスクの性質、AI 利用者の属性等に応じて異なり得る。

例えば、AI の出力内容が AI 利用者の業務目的との関係で妥当か否かという観点は、AI 利用者が本来行うべき注意の下で当然に評価・検証されるべき事項であり、出力の是非を検討しながら AI を用いることが AI 利用者の注意義務の内容となる（後記 3.1・想定事例 1 及び後記 3.2・想定事例 2）。このような観点では、AI の出力の適切性を AI 利用者がコントロールすることが一般に容易と考えられ、AI 開発者・提供者の責任が問題となることは想定されにくい。

他方、AI の出力を一見するだけではリスクの所在が不明瞭なケースがある（後記 3.3・想定事例 3 及び後記 3.4・想定事例 4）。このような場合にも、AI 利用者が最終的な判断や行動において AI の出力を適正な範囲で用いるべきことは変わらないが、そのために必要な措置は、単純に AI の出力内容自体を検証することよりも、事前に一定の情報収集を行うことや必要な利用上の措置を講ずることが注意義務の中心となり得る。また、このようなケースでは、AI 利用者において AI のリスクを具体的に予見したり、その出力をコントロールしながら用いたりすることが必ずしも容易でないため、AI 開発者・提供者においても一定の設計上の措置を講ずるべきでなかったかが論点となる。

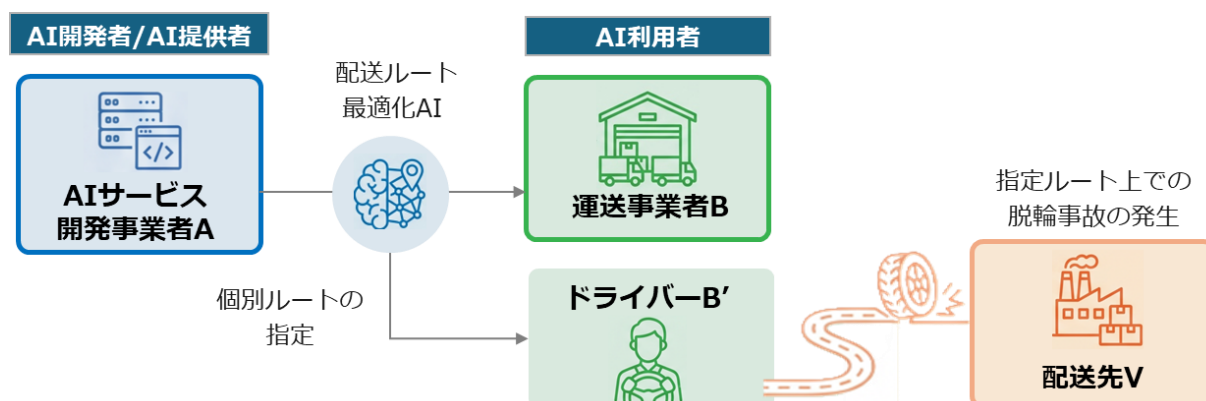
3.1 想定事例 1：配送ルート最適化 AI

運送事業者 B は、AI の開発・提供事業を営む A³⁵が提供する配送ルート最適化 AI システムを使用し、日々の配送ルートを決定している。当該システムでは、配送先、車両情報、時間制限などの条件を入力することで、配送先の位置関係・距離・納品時刻・交通状況等の諸条件を考慮し、効率性・安全性の観点から最適化した配車及びルーティングの計画を自動で作成することができ、配送時間の短縮や必要配車台数の削減等のコスト削減が可能となる。A はこのような AI の機能・特性や、最終的な道路状況の安全性（安全に通行できる幅・路面・視界が確保されているか等）をドライバーにおいて確認すべきことを B に対して説明していた。B は運送の効率化のため、各ドライバーに対し原則として AI が出力した

³⁵ A は自ら開発した AI システムを用いて B に対しサービス提供を行っているため、AI 開発者としての地位と AI 提供者としての地位を有する（AI 事業者ガイドライン本編 5 頁、同別添 9-10 頁参照）。

ルートに従うよう指示していたが、走行上の安全性は現場にて確認し、安全性が確保できない場合には自ら判断した安全なルートを走行するよう教育・指導を行っていた。

ある日、B のドライバー B' が運送業務を行っていたところ、幅員が狭く大型車両の運行に適さない悪路を「最適ルート」と表示した。B' は当該道路への進入に躊躇したが、当該日の配車スケジュールが極めてタイトであったことから、全体配送計画との連動性・遅延ペナルティ等を考慮し、AI の出力に従ったところ、結果として運行のための十分なスペースがなく脱輪した。これにより大幅な遅配が生ずると共に、その際の衝撃で配送先 V への荷物が損壊し、V は代替品の調達等のための損害を被った。



上記事例では、B'が AI の指定ルートに従った結果として事故を生じたことについての責任（民法 709 条）の成否が論点となり、B'に責任が認められれば、B は B'の使用者としての責任（民法 715 条）を負う。

また、A については、AI が結果として走行に適しないルートを指定したことについて何らかの責任を負うか否かが論点となる。

3.1.1 想定事例 1 における AI の位置づけ

従来も目的地までの最短経路を提示する地図アプリ等の経路案内サービスが提供されてきたが、AI 技術の進展により、さらに多様な要素を分析して最適化された経路を提示することが可能となった。こうした AI には、配送先の位置関係、距離、納品時刻等の配送の効率性に関わる情報のほか、規制、狭小道路の存在、地図情報、速度情報等の安全性に関わる情報を考慮し、走行経路の安全面も加味した上でルートを案内するものも存在する³⁶。

このように、経路最適化 AI は一面において安全性向上のための機能も有するものの、機能的にみて個々の局面における安全判断を包含するものではないため、補助／支援型 AI に該当する。想定事例 1 のような場面において、事故の回避に最も直結するのは、幅員・路面等の道路の状況及びこれらに対するドライバーの認知や判断など、実際の道路上で得られる情報や実運行に関する情報であり、事故を防止するためにはドライバーがこれらの情報を踏まえて安全性の判断

³⁶ 国土交通省国土交通政策研究所「地域の輸送資源を活用した物流ネットワーク最適化に関する調査研究（最終報告）」12-14 頁（国土交通政策研究第 172 号、2023 年）等参照。

を行うことが求められる。上記 AI はこのような情報を考慮してルートを計算しているわけではなく、車体情報や道路情報等の外部的な情報を基に経路の計算を行っており、AI の計算が安全性を一定程度確保するように設計されていたとしても、ドライバーが行うべき認知や判断を代わりに行っているとはいえず、その性質上、あくまで参考情報として用いるべき補助／支援型 AI に該当する。

3.1.2 AI 利用者である B 及び B' の責任

前記のとおり、経路最適化 AI は AI 利用者の判断の補助（補助／支援型 AI ）として用いられるべきであり、AI の出力を援用したことは、B' がドライバーとして安全に走行すべき注意義務の水準を変化させる要素とはならない。

例えば、事故現場の道路の危険性が極めてわかりづらいような状況にあり、運送業務に従事する通常のドライバーであっても当該道路を進行することが合理的と認められるような特段の事情があれば、B' の過失が否定される場合も観念し得る。他方、狭隘な道路を大型車両で進行することの危険性は一般的には認識しやすく、必要に応じて引き返すことで事故を回避する等の行動を取り得るのが通常であるから、特段の事情の無い限りは過失が肯定されやすいと思われる。

B' の過失（民法 709 条）が認められる場合、B は使用者責任（民法 715 条）を負う。

また、前記 3.1.1 のとおり、上記 AI は機能的にみて補助／支援型 AI に該当するにもかかわらず、AI の出力に必ず従う運用を行っていた場合、そのことが事故に寄与したと評価される可能性もある。想定事例 1 とは異なり、全社的な配送計画を乱さないようにする等の理由で、B が各ドライバーに対し AI の判断に必ず従うよう指示していたような場合には、このような指示を行った B の従業員の過失が認められる可能性もある。

3.1.3 AI 開発者・提供者である A の責任

AI 開発者・提供者である A の責任は、以下のような 2 つの場面において論点となる可能性がある。もっとも、想定事例 1 の事実関係に鑑み、いずれの場合も A が責任を負う可能性は基本的に想定し難い。

(1) V の損害を賠償した B が A に対して求償等を行う場面

A と B との間においては経路最適化 AI の提供に係る契約が存在すると考えられ、V の損害を賠償した B は、A に対し、当該契約も踏まえて求償又は損害賠償請求を行うことが考えられる。このような請求が認められるか否かは AB 間の契約次第であるため本研究会の検討対象とはしないが（前記 2.1 参照）、AI の性能・品質についてどのような合意がなされていたか、契約上予

定されていた性能・品質等に満たなかった場合にどのような責任が規定されているか、どのような免責条項が設けられていたか³⁷等が重要な要素となる。

(2) V が直接に A に対する損害賠償請求を行う場面

上記(1)のほか、V が（B に対する請求と共に）A に対し直接に損害賠償請求を行う場面も一応想定されるが、その場合、A と V との間には契約関係がないことから不法行為に基づく請求が根拠となる。

前記 2.2.1 のとおり、補助／支援型 AI においては、AI に望ましくない出力があったとしても最終的に AI 利用者の判断や行動によって検証・是正することが予定されているから、A の不法行為責任が認められるのは限定的である。A は、AI の機能・特性や、最終的な道路状況の安全性（安全に通行できる幅・路面・視界が確保されているかなど）をドライバーにおいて確認すべきことを B に対して説明していることから、必要な説明が不足していたと評価される可能性は低い。また、このような説明を行っている限り、実際に道路を走行するドライバーが安全な走行経路を自己の判断において選択することが当事者間の役割分担としても予定されるため、更に A に対して設計上の措置が求められることは通常想定し難い³⁸。

3.2 想定事例 2：弁護士業務支援 AI

LLM 提供事業者 C³⁹は、自社が開発した汎用的な大規模言語モデルの提供を行っている。AI サービス開発事業者 D は、法的な推論能力を向上させる目的で C のモデルをファインチューニングした上、推論時に自社で作成した法令・裁判例データベースから関連情報を検索してモデル入力に統合する（Retrieval-Augmented Generation, “RAG”）ことにより、弁護士業務を補助するチャットボット形式の AI サービスを構築・提供している。当該 AI サービスでは、AI 利用者である弁護士が調査したい内容や具体的な事案をプロンプトとして入力することにより、関連性の高い文献や裁判例を表示したり、法的な分析をまとめたレポートを生成したりする機能を有している。弁護士 E は、当該 AI サービスを自己の業務に導入している。

E は、依頼者 V から相談を受けた紛争案件について、従来の経緯を記した書面や証拠書類を AI サービスに入力し、当該案件に対する法的論点や事件の見立てに関する分析を求めた。AI の出力によれば、V の立場を裏付ける裁判例が複数存在し、主張が認められる可能性は高いという結論であった。このような分析は E 自身の経験や考え方とも合致するものであったため、E は自ら裁判例の調査をすることなく AI の出力を信頼し、和解を検討せず V の利益の最大化を図る方針を立てた。しかし実際には、当該

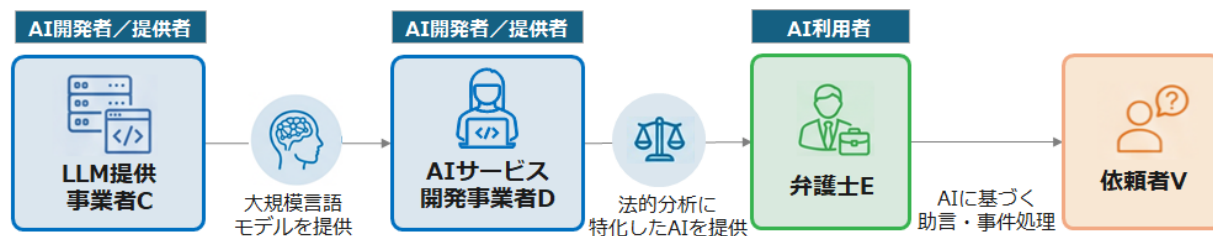
³⁷ 本研究会では、想定事例 1 のような場面において、通常、A の損害賠償等の責任を故意又は重過失がある場合に限定する免責規定が設けられていることが多いとの指摘があった。この場合、想定事例 1 において A に事故発生についての故意又は重過失が認められ、責任が生ずるのは極めて限定的なケースのみと考えられる。

³⁸ 教室事例としては、AI が危険な道路にドライバーを誘導する傾向が強いことを A が知りながらサービスを提供したような場合は、設計上の措置を講じるべき義務違反が問われることとなると思われる。

³⁹ C は自ら開発した大規模言語モデルを D に対して提供しているため、大規模言語モデルについての AI 開発者及び AI 提供者としての地位を有する。また、D は自ら開発した弁護士業務支援 AI を E に対し提供しているため、当該 AI サービスについての AI 開発者及び AI 提供者としての地位を有する。

AIは実在しない架空の裁判例を複数引用し、Vの立場を支える根拠としており、実存する裁判例としてはVの主張に否定的なものが大多数を占めていた。

最終的に、裁判所においてVの請求は全面的に棄却され、Vは訴訟追行等に要した弁護士費用等の損害を被った。



上記事例では、Eについて、依頼者の立場を裏付ける裁判例について自ら必要な調査をせず、客観的に誤ったAIの出力に依拠した事件処理を行ったことに関し、Vに対する委任契約上の善管注意義務違反に基づく責任（民法415条）や一般不法行為責任（民法709条）の成否が論点となる。

また、CやDについては、AIが結果として誤った出力を行ったことについて何らかの責任を負うかが論点となる。

3.2.1 想定事例2におけるAIの位置づけ

想定事例2における法的分析のためのAIは、法令や裁判例等を参照しつつ、個別の事件の資料や証拠を踏まえて法令の適用に関する分析を生成する機能を有しており、その出力の機能や性質としては、弁護士の業務遂行における認知や判断に近い側面を含んでいる。

しかし、規制法が存在する業種においては、想定事例1とは別途の考慮を要する。弁護士が業務上AIを用いること自体については特段の規制は存在しないものの、弁護士法72条では、「弁護士又は弁護士法人でない者は、報酬を得る目的で訴訟事件…その他一般の法律事件に関して鑑定、代理、仲裁若しくは和解その他の法律事務を取り扱い、又はこれらの周旋をすることを業とすることができない」と規定されている。弁護士又は弁護士法人でないAI開発者・提供者が、非弁護士に対し、弁護士法72条に規定される行為態様で弁護士の判断を介在させずにAIを提供することは、同条に抵触する可能性がある。

想定事例2では、このような弁護士法72条の存在も踏まえ、弁護士Eの判断の補助ないし支援としてのみ用いられる補助／支援型AIが提供される事例を想定している。

3.2.2 AI 利用者である E の責任

E は委任契約に基づく善管注意義務及び不法行為上の注意義務を負うところ、両者の判断枠組みは実質的には異ならないと考えられる⁴⁰。想定事例 2 の AI は、前記 3.2.1 のとおり補助／支援型 AI に該当するため、E に善管注意義務違反ないし過失が認められるか否かは、実際の事件処理の過程を前提とした E の行動が合理的であったか否かにより決せられる。E が AI を利用したという事実は、E の注意義務の水準を引き上げることも引き下げることもない。

E が負う注意義務の水準に当たり留意を要するのは、E が専門家である弁護士としての注意義務を負う点である。不法行為法上の「過失」の判断に当たっては、一般的・抽象的な人を基準とするのではなく、被告の職業・地位・立場等の社会生活上の役割ごとに類型化された通常人としての注意力を基準として判断され、そのような注意力を有する弁護士としての注意義務が課せられる⁴¹。

したがって、E の過失の有無との関係では、通常の注意力を有する弁護士であればどのように行動したか、E の行動がこれに適合していたかが問われることとなり、AI を利用した場合には、AI の出力を参考にしてとられた E の判断が弁護士としての注意義務に即したものであるかが争点となる。想定事例 2 との関係では、依頼者の立場を裏付ける裁判例の実在性は事件処理に当たって非常に重要性の高い情報であり、弁護士であれば綿密に調査確認を行うべきと考えられ、AI の出力のみに依拠し、このような調査確認が不足していた E の責任が否定されることは考えにくい。

なお、本研究会では、AI が専門家と部分的に同等以上の判断を行う性能を発揮する場面も現れつつあることに鑑み、人間が高度な AI と協調的に作業を行う場合に人間の注意力が低下することが示唆されているという認知科学的な知見を踏まえると、専門家が AI の誤出力を発見できなかったことをもって直ちに過失を問うことは望ましくない場面もあり得るところ、そのような場合を想定した責任の判断枠組み（例えば、当該専門領域において AI をどう活用すべきかを判断する「適切な利用方法に関する高度な判断」を専門家としての注意義務と捉える考え方など）についての検討も重要であるとの指摘があった。

3.2.3 AI 開発者・提供者である C 及び D の責任

AI の開発や提供を行っていた C 及び D の責任は、想定事例 1 の場合（前記 3.1.3）と同様と考えられる。すなわち、(1)V の損害を賠償した E が C や D に求償又は損害賠償請求を行う場面

⁴⁰ 受任者は委任の本旨に従い、善良な管理者の注意義務をもって委任事務を処理する義務（民法 644 条）を負う。「善良な管理者の注意義務」（善管注意義務）は、目的である事務処理のために要求される教育の程度や専門的知識および委任者の知りまたは知るべき受任者の能力や性質によって具体的に決定され、とりわけ、受任者が職業的事務処理者（専門家）である場合には、同様の職業・地位にある者に対して一般に期待される水準の善管注意義務が課されるところ（山本豊編『新注釈民法(14) 債権(7)』251 頁〔一木孝行〕（有斐閣、2018 年）、当該業務において客観的に要求される一般的な注意水準が問題となる点で、前記 2.1.1 の過失の判断枠組みと類似する。

⁴¹ 窪田充見編『新注釈民法(15) 債権(8)〔第 2 版〕』366-368 頁〔橋本佳幸〕（有斐閣、2024 年）

⁴²と、(2)V が（E に対する請求と共に）C や D に対し直接に損害賠償請求を行う場面が想定されるが⁴³、いずれの場合も C ないし D が責任を負う可能性は限定的と考えられる。

不法行為法が根拠となる(2)の場面においては、①各当事者が提供するサービスの性能限界や使用方法に関する説明を AI 利用者に対して適切に行っていたか、及び、②AI 利用者である弁護士による誤った紛争処理を誘発したと評価されるかが論点とはなり得る。①との関係では、C が D に対して、D が E に対して、それぞれ自ら提供していた汎用的な大規模言語モデルや RAG サービスの機能や性能の限界、使用方法、重要なリスク等を明確に説明していることが重要と考えられるが、このような説明が行われている限り、最終的に弁護士が自らの責任において AI の出力を踏まえた判断を行うことが当事者間の役割分担としても予定されており、C や D が責任を負うことは通常想定し難い。

3.3 想定事例 3：画像生成 AI

【事例 a】 アパレル業を営む G は、AI サービス開発事業者 F が開発・提供する画像生成 AI を自社広告等に用いていた。G の担当者 G' は、当該 AI によって生成された画像のうち、自社ブランドの雰囲気と合致する肖像をいくつか選択し、自社商品の Web 広告に利用した。メインの広告に使用していた画像は、全国的に知名度のあるタレント V の容貌に酷似していたが、G' はそれに気づいていなかった。G' は、V を含む著名人と類似の画像を生成することを意図しておらず、プロンプト入力に際しても固有名詞は入力せずに多様な人物やポーズのパターンを生成していた。自らの容貌に酷似した画像が広告に利用されていることを発見した V は自己のパブリシティ権が侵害されていると主張した。

F が開発・提供する画像生成 AI はリアルな風景画や人物画像等の多様なニーズに対応することを目的とした汎用サービスであり、利用規約上、生成物は商用利用も可能とされ、創作活動、SNS で用いるアイコン、企業の広告等に用いられていた。F が当該 AI の開発に用いていたデータセットは、人物の画像等を含む約 10 億枚の画像を示す URL を大量に含んだ公開データセットであり、ウェブクロールによって作成されたものであるため、V の画像がデータセットに含まれていたかは確認が困難であったが、当該データセットにおいて著名人の氏名をラベル付けしたデータ等は用いておらず、また Re-captioning や潜在空間での学習等、元データが生成画像にそのまま反映されることを防ぐための措置を講じていた。F が実施したモデル評価の過程や、リリース後のユーザーからのフィードバックにおいて、V に酷似した画像が生成されたという報告は一切なかった。

【事例 b】 アパレル業を営む G の担当者 G' は、AI サービス開発事業者 F が開発・提供する画像生成 AI を自社広告等に用いていた。G' は、当該 AI によって生成された画像に V に酷似した画像が含まれて

⁴² 前記 3.1.3(1)と同様、個別の契約内容次第であるが、AI の性能・品質についてどのような合意がなされていたか、契約上予定されていた性能・品質等に満たなかった場合にどのような責任が規定されているか、どのような免責条項が設けられていたか等が重要な要素となると考えられる。

⁴³ 本研究会では、想定事例 2 の場合、AI サービスの開発・提供に関する C、D、E らの間の契約関係や説明の十分性といった内部的な事情が V からは見えにくく、V は直接の契約の相手方である E の責任を追及することが通常と考えられ、C や D の責任は、主として、E が V に賠償した後の E から D に対する、又は D から C に対する求償関係においてのみ問題となるとの指摘があった。

いることに気づき、その知名度によって自社商品の販売を促進しようと考え、自社商品の Web 広告に利用した。これを発見した V は自己のパブリシティ権が侵害されていると主張した。

画像生成 AI のモデル学習方法や利用規約の内容等、その余の事実関係は事例 a 第二段落と同様とする。

【事例 c】 AI サービス開発事業者 F は、複数の著名人の肖像に当該著名人の氏名をラベル付けしたデータで構成されたデータセットを利用して学習を行い、著名人の氏名をプロンプトとして入力すると、V を含む著名人と同一又は類似の肖像が高頻度で生成される AI モデルを開発した。F は、当該 AI モデルを用いた画像生成 AI を、著名人の様々な画像を生成することを目的とした AI として、アパレル業を営む G 等の顧客に対し販売していた。



上記事例では、V と酷似する容貌が含まれた画像を G が自社広告に利用したことについて一般不法行為責任（民法 709 条）の成否が論点となり、G に責任が認められる場合には G も使用者責任（民法 715 条）を負う。

また、F については、G が意図的にパブリシティ権を侵害したような場合、G の行為に対する幫助責任（民法 719 条 2 項）の成否が論点となる。さらに、例外的なケースではあるが、F の行為態様次第で、F が主体となってパブリシティ権を侵害していると評価される場合の責任（民法 709 条）も論点となり得る。

3.3.1 パブリシティ権と画像生成 AI

パブリシティ権とは、肖像等⁴⁴に顧客吸引力を有する者が、その顧客吸引力から生じる経済的な価値や利益を排他的に利用する権利であり、下記参考裁判例 4 で権利の存在及び内容が認められた。当該裁判例は、パブリシティ権侵害として不法行為法上違法となる場合として、下記①～③の 3 類型（以下単に「3 類型」という。）を挙げた。

⁴⁴ 「肖像等」とは本人の人物識別情報をいうものであり、例えば、サイン、署名、声、ペンネーム、芸名等を含むとされている（中島基至「判解」最判解民事篇平成 24 年度（上）41 頁（2015 年））。



権利の性質

- 人の氏名、肖像等は、個人の人格の象徴であるから、当該個人は人格権に由来するものとして、これをみだりに利用されない権利を有する。
- 肖像等は、商品の販売等を促進する顧客吸引力を有する場合があります、このような顧客吸引力を排他的に利用する権利（「パブリシティ権」）は、肖像等それ自体の商業的価値に基づくものであるから、上記の人格権に由来する権利の一内容を構成するものといえることができる。
- 肖像等に顧客吸引力を有する者は、社会の耳目を集めるなどして、その肖像等を時事報道、論説、創作物等に使用されることもあり、その使用を正当な表現行為等として受忍すべき場合もある。

権利侵害の判断基準

- 肖像等を無断で使用する行為は、①肖像等それ自体を独立して鑑賞の対象となる商品等⁴⁵として使用し、②商品等の差別化を図る目的で肖像等を商品等に付し、③肖像等を商品等の広告として使用するなど、専ら肖像等の有する顧客吸引力の利用を目的とするといえる場合に、パブリシティ権を侵害するものとして、不法行為法上違法となる。

パブリシティ権が争点化する過程において、従来は、特定の人物の容貌が撮影され、当該肖像が広告等に直接利用される場面が典型的であった。上記参考裁判例 4 においても、歌手を被写体とする写真について本人の承諾なく雑誌に掲載する行為がパブリシティ権を侵害するかが争われた。

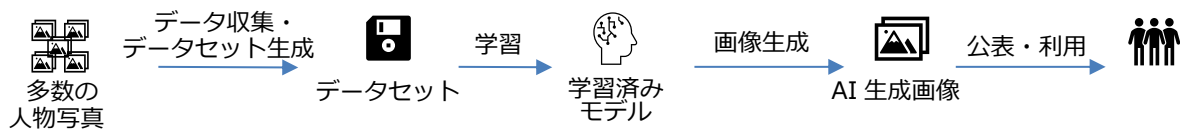


他方、生成 AI による画像生成では、まず、多数の人物肖像データ等を収集して学習用のデータセットを生成し、当該データセットを用いて AI モデルの学習を行い、学習済みモデルに画像生成の指示（プロンプト）を入力することで人物肖像が生成され、当該肖像が広告等に利用されるという過程を経る。AI による画像生成では、学習用データセットに無数の画像が含まれているが、AI モデル学習の過程で AI モデルにおけるそれらのデータはパラメータ化されているため、人物の写真そのものが直接的に生成に利用されるわけではない。

このような構造から、画像生成 AI が偶発的に特定の著名人と類似する画像を生成し、利用に至った場合に、どのように権利侵害性や過失を判断するかが問題となる⁴⁶。

⁴⁵ 「商品等」とは商品又はサービスとして商品化されたものをいう。また、肖像等の使用が私的なものに留まる場合には違法性を欠くと解されるため、商品等に該当するのは業としての行為（商業的利用行為）に用いられる場合に限られるとされている（前掲注 44・中島 40 頁）。

⁴⁶ なお、以下では各当事者の不法行為に基づく損害賠償責任に関する論点を論じているが、別途、差止請求の可否も問題となり得る。差止請求の要件として、一般に、①権利・法益の侵害、②違法性、③実質的侵害の発生に対する蓋然性（因果関係）が挙げられる（大塚直編「新注釈民法(16)債権(9)」429 頁〔大塚直〕（有斐閣、2022 年）。上記②の違法性要件に関し、前掲注 44・中島 69-70 頁は、一般に人格権侵害を理由として請求権を行使する場合、差止請求が認められるときの違法性は損害賠償請求が認められるときの違法性より高いと考えられていることを踏まえ、不法行為法上違法とされても差止請求までは認められないとする見解に言及しつつも、結論として、不法行為が成立する場合には直ちに差止請求も認めるとするのが相当であるとの立場を示している。これに対し、明文の規定を欠くパブリシティ権の侵害について、不法行為に基づく損害賠償請求と同一の基準により差止請求を認めることを批判する見解も主張されている（斉藤邦史『プライバシーと氏名・肖像の法的保護』236 頁（日本評論社、2023 年））。



3.3.2 想定事例 3 における AI の位置づけ

想定事例 3 のような画像生成 AI は、プロンプトに沿った画像を生成する機能を有しており、その画像は必要に応じて追加の加工等を経た上で AI 利用者の判断において利用することが予定される。また、生成画像の内容によっては第三者のパブリシティ権侵害に結びつく可能性があるところ、そのような画像を外部的に利用しないよう、利用過程における人の評価や検証が必要と考えられ、この意味においても依拠／代替型 AI としての利用になじまない。以上を踏まえると、上記 AI も補助／支援型 AI に該当する。

補助／支援型 AI の出力に第三者の権利を侵害するものが含まれる場合、AI 利用者がその判断において AI の出力を適正な範囲で利用することが求められるが⁴⁷、AI 利用者において、AI からの出力画像と無数に存在する権利者の肖像とを逐一照合することは一般に困難であり、AI 利用者において権利者への十分な配慮を行うことができず結果としてパブリシティ権侵害に至る場面も予想される。この場合において、一定の要件の下に、AI 開発者・提供者において権利侵害防止措置が求められ得る（後記 3.3.4）。

3.3.3 事例 a について

(1) AI 利用者である G 及び G' の責任

事例 a は、AI 利用者である G 及び G' が意図せず V に酷似した肖像を Web 広告に利用した事案である。

G 及び G' は、前述のとおり生成画像を適正な範囲で利用することが求められ、責任の有無は従来のパブリシティ権侵害に基づく損害賠償請求の要件に従って判断される。当該要件に関し、参考裁判例 4 以降の裁判例では、権利侵害性及び故意・過失を一体として論じているもの⁴⁸や、これらを分けて論じているもの⁴⁹がある。画像生成 AI を利用する場面では、生成された画像の利用形態等が前記 3.3.1 の法的評価を踏まえてパブリシティ権の侵害に該当するか否かという権利侵害性と、AI 利用者が当該権利侵害を防止する義務を負っていたか否かという故意・過失の問題は質的に異なり得るため、以下ではこれらを分けて整理する。

⁴⁷ なお、パブリシティ権に関し AI 利用者が行うべき措置について、AI 事業者ガイドラインでは直接的な記述は存在しないものの、ステークホルダーの財産等に危害を及ぼさないよう安全性に配慮した適正利用を行うことが求められている。後記 34 頁「＜参考＞パブリシティ権に関して AI 事業者ガイドライン上想定される措置」参照。

⁴⁸ 知財高判平成 25 年 10 月 16 日（判例集未登載）、大阪高判平成 29 年 11 月 16 日判時 2409 号 99 頁、知財高判令和 7 年 6 月 30 日（判例集未登載）等

⁴⁹ 東京地判平成 25 年 4 月 26 日判タ 1416 号 276 頁、知財高判令和 2 年 2 月 20 日（判例集未登載）、知財高判令和 5 年 9 月 13 日（判例集未登載）等

① 権利侵害性

前述のとおり、パブリシティ権の侵害となるのは、肖像等を利用する場合であって、その行為態様が 3 類型に該当する場合など、専ら肖像等の有する顧客吸引力の利用を目的としている場合とされている。

まず、画像生成 AI で生成された肖像を利用する場合、本人の肖像が収められた写真等を直接利用していないため、そもそも「肖像等」に当たるかが問題となり得る。下記参考裁判例 5 では、著名人に似せた絵の利用行為に関し、当該人物を知る者が容易に識別し得るほどの類似性を備えていなかったこと等を理由にパブリシティ権侵害の成立を否定している。当該裁判例に照らせば、本人の肖像そのものを用いる場合でなくとも、生成された画像が、当該人物を知る者により容易に識別可能なほど本人に類似する場合には「肖像等」に該当し得ると考えられる⁵⁰。



参考裁判例 5：東京地判平成 17 年 6 月 14 日判タ 1217 号 310 頁

- ロック歌手である原告に似た人物絵をパチンコ機に使用したことでパブリシティ権侵害が争われた事案において、当該人物絵は原告の肖像がイメージにあったと推認されるが、客観的に見て、原告とある程度の類似性は有するものの、原告を知る者が容易に原告であると識別し得るほどの類似性を備えたものとはいえない等の理由で、パブリシティ権侵害は成立しない。

その上で、画像生成 AI の利用に当たっては、偶発的に本人に類似する画像等が生成され、AI 利用者において必ずしもパブリシティ権侵害を意図していないケースも想定されるが、諸般の事情を考慮して、専ら肖像等の有する顧客吸引力を利用する目的があるといえるか否かが評価されると考えられる。参考裁判例 4 以降の裁判例では、権利侵害性について、肖像等の使用・掲載態様⁵¹、肖像等に付されている文章の分量や内容⁵²、掲載されている媒体の目的や性質⁵³等、主に顧客吸引力の利用と結びつく客観的な事情が考慮されている。

事例 a において、V が全国的に知名度のあるタレントであること、生成画像が V に酷似していること、当該画像を G が自社商品の販売を促進するためのメイン広告に利用していることは、3 類型のうち③肖像等を商品等の広告として使用する行為に該当し、顧客吸引力を利用する目的を強く基礎づける要素と考えられる。このような要素を考慮すると、たとえ G が主観的にはパ

⁵⁰ なお、前掲注 44・中島 41 頁では、著名人本人に似ている動物等の図柄を、本人を示すものとして付したキャラクター商品について、この図柄が需要者にとって本人を識別するものとして著名であるような場合には、本人の人物識別情報といえるから、肖像等に含まれるとしている。上記解説によれば、人物絵より更に抽象化された動物等の図柄であっても、本人を識別するものとしての著名性次第で肖像等に含まれる可能性がある。

⁵¹ 上記参考裁判例 4。当該裁判例は、パブリシティ権侵害の基準の中で 3 類型を例示しているものの、事案への適用においては 3 類型該当性について検討を加えずに、使用行為の目的、方法及び態様を考察して、その使用行為が専ら当該芸能人の顧客吸引力の利用を目的とするものであるか否かを論じていると指摘されている（石井美緒「パブリシティ権と表現の自由」日本大学知財ジャーナル vol.9 58 頁（2018 年））

⁵² 前掲注 48 知財高判平成 25 年 10 月 16 日（判例集未登載）では、写真の書籍への掲載態様、写真に付された文書が独立の意義を有するか等が考慮されている。

⁵³ 前掲注 49 知財高判令和 2 年 2 月 20 日（判例集未登載）では、肖像等が掲載されているウェブサイトの目的が商品等の宣伝広告・販売促進を目的にしていること、前掲注 49 知財高判令和 5 年 9 月 13 日（判例集未登載）では、ウェブサイトがグッズ販売サイトやファンクラブサイトに誘導し、グッズの購入やファンクラブサイトへの入会を促す役割を果たすものであることを指摘している。

プライバシー権侵害を意図していない場合であっても、客観的にみて専ら顧客吸引力の利用を目的としていると評価される可能性は高い⁵⁴。

他方、事例 a とは異なり、顧客吸引力の利用目的を否定し得る事例としては以下のようなものが考えられる。こうした事例において、専ら顧客吸引力を目的とする利用ではないと評価される場合、権利侵害性が否定される。

- ・ 生成画像が一部地域でのみ知名度のあるローカルタレントに類似しており、広告の目的等に照らして当該タレントの顧客吸引力を利用する目的があったとは評価し難い場合
- ・ 生成画像の掲載媒体が商品の販売促進に向けられたものではない場合
- ・ 掲載媒体において生成画像の占める重要性が低い場合

② 故意・過失

故意・過失を権利侵害性とは別に判断している裁判例では、故意・過失の認定において、肖像等の利用に関する従前の経緯⁵⁵、行為者の性質⁵⁶が考慮されている。

事例 a の画像生成 AI が補助／支援型 AI に該当することにも鑑みれば、AI 利用者は、合理的な範囲で、当該肖像を広告に利用した場合にパブリシティ権を侵害しないかについて調査・確認を行う義務を負うと考えられる。G は、一般向けに公開する広告を作成する事業者として相当の注意を以て調査・確認を行う注意義務を負い、特に生成画像が全国的に著名なタレントに酷似していたことを踏まえると、簡易な調査・確認によって画像の利用を回避することは可能であったとして過失が認められる可能性が高い。

他方、事例 a とは異なり、AI の利用過程で一部地域でのみ知名度のあるローカルタレントに類似する画像が生成され、それが G ないし G' の意図したものではなく、偶発的に生成されたに過ぎない場合⁵⁷など、G が相当の注意を行ったとしてもパブリシティ権侵害を具体的に予見することが困難であった場合には、過失が否定される可能性もあり得る。

⁵⁴ AI 利用者が、生成された画像について肖像等に顧客吸引力を有する実在の人物と類似しているとの認識がない場合、そのことをもって「専ら顧客吸引力を利用する目的」を否定する考え方もあり得る。もっとも、権利侵害性の認定において肖像等の使用・掲載態様等の客観的な要素に着目する裁判例を踏まえると、実在の人物と類似することについての認識がないことをもって直ちに権利侵害性が否定されるものではなく、類似性や使用態様等の事情を踏まえて「専ら顧客有引力を利用する目的」の有無が判断されると考えられる。

⁵⁵ 前掲注 49 東京地判平成 25 年 4 月 26 日判タ 1416 号 276 頁では、故意・過失の判断にあたり、雑誌の出版以前にもパブリシティ権侵害に関する抗議を受けていたこと等を指摘している。

⁵⁶ 前掲注 49 知財高判令和 5 年 9 月 13 日（判例集未登載）では、故意・過失の判断にあたり、被告がアーティストのマネジメント会社として、アーティストが有するパブリシティ権の取扱いについても相応の注意をもって業務を行うべき立場にあったことを指摘し、被告の故意・過失に関する主張を排斥している。

⁵⁷ G ないし G' が著名人に類似する画像の生成を意図していたわけではなく、偶発的にそのような画像が生成されてしまったに過ぎないことを示す事情としては、例えば、利用する AI について後記 3.3.4 のような権利侵害防止措置が講じられており、AI 利用者において実在の著名人が生成される可能性が低いことを G ないし G' が確認していたことや、利用時のプロンプトに実在の著名人の氏名等を入力しないよう管理していたこと等が考えられる。

(2) AI 開発者・提供者である F の責任

F による画像生成 AI の提供行為⁵⁸については、G 及び G' の不法行為に対する幫助責任の成否や、F 自身が主体となるパブリシティ権侵害の成否が論点となり得る。

もっとも、事例 a は AI 利用者が過失によりパブリシティ権を侵害したか否かが論点となる事例であるところ、このような他人の不法行為に関する幫助責任が問題となるのは、通常、後記 3.3.4・事例 b のように権利侵害者が故意の不法行為に及んだ場合である。従来の裁判例でも幫助責任が問われているのは基本的に故意の不法行為に対する幫助行為である（前掲注 17）。

また、F 自身がパブリシティ権侵害の主体と評価されるのは、後記 3.3.5 の事例 c のような例外的なケースに限られる。

そのため、F の責任は事例 a においては認められ難く、下記事例 b 及び事例 c においてのみ論じる。

3.3.4 事例 b について

(1) AI 利用者である G 及び G' の責任

事例 b は、AI 利用者である G' が意図的に V のパブリシティ権を侵害した場合であり、前記 3.3.3 の考え方に照らし故意の不法行為が成立し、損害賠償責任が認められる。

(2) AI 開発者・提供者である F の責任

上記(1)の場合、AI 開発者・提供者である F の画像生成 AI が G' の行為に用いられているところ、当該 AI の提供行為が G' の不法行為を容易にしたとして、F に過失による幫助責任（民法 719 条 2 項）が成立するかが問題となる。

F の幫助責任については、特に下記参考裁判例 6 が参考となる。当該裁判例では、カラオケ装置のリースによって利用店の経営者による著作権侵害を生じさせる蓋然性が高いこと等を理由に、カラオケ装置のリース業者に対し、著作権使用許諾契約の締結・申込みを確認した上でカラオケ装置を引き渡すという条理上の義務を課し、当該注意義務を履行していないことに基づく共同不法行為責任を認めた。その根拠として、同裁判例では、条理上の注意義務を肯定する根拠として、(i)カラオケ装置のリースによる権利侵害の蓋然性、(ii)被害法益の重大性、(iii)リース業者の社会的地位、(iv)予見可能性、(v)結果回避の可能性・容易性を挙げている⁵⁹。

⁵⁸ なお、F の学習用データに V の肖像が含まれていた場合、そのことがパブリシティ権侵害に当たるか否かも論点とはなり得るが、前記 3.3.1 のとおり、データセットの画像は直接出力に用いられ、対外的に利用されたりすることはなく、AI モデルのパラメータの調整のみに用いられるため、仮に学習用データとして V の肖像が含まれていたとしても、このような利用態様によって「専ら肖像等の有する顧客誘引力の利用を目的とする」とはいえず、パブリシティ権を侵害しないと考えられる。

⁵⁹ 詳細な解説については、高部眞規子「判解」最高裁判所判例解説民事篇平成 13 年度（上）196 頁以下（2004 年）参照。また、参考裁判例 6 では、直接的に「幫助」の文言が用いられていないが、同判例がリース業者の幫助責任を肯定した旨を解説するものとして、匿名記事「判批」判タ 1058 号 110 頁（2001 年）、半田正夫ほか編『著作権法コンメンタール 1（1 条～25 条）（第 2 版）』911 頁（辻田芳幸）（勁草書房、2015 年）。

このように、他者が用いる製品を提供する者の幫助責任の成否に当たっては、①提供する製品による権利侵害の蓋然性や重大性、及び、②製品提供者における侵害発生認識可能性が重要な要素となり、これらを相関的に考慮して幫助責任の成否が検討される⁶⁰。

参考裁判例 6 では、カラオケ機器の性質上、著作権使用許諾契約の締結・申込みを確認せずにリースを行った際には著作権侵害が生ずる蓋然性が高いことを踏まえ、当該契約の有無に応じてリース先を選定する義務を認めた。これに対し、画像生成 AI は、私的な創作活動等を含め適法かつ多様な目的のために用いることが可能であり、利用者によるパブリシティ権侵害の蓋然性が典型的に高いとはいえないこと、パブリシティ権が裁判例上認められた人格的権利でありその侵害が刑事罰の対象ともされていないこと等に鑑みると、上記参考裁判例 6 の前提、特に (i)、(ii)、(iv) とは事情を異にしており、AI 開発者・提供者に対して AI 利用者の選定等の高度の注意義務を課すべきではなく、AI 利用者に対する画像の使用方法に関する警告や、一定程度リスクを緩和するための設計上の措置が論点になると考えられる。

この点を踏まつつ、画像生成 AI のサービス提供に即して、①提供する AI による権利侵害の蓋然性や重大性、及び、②AI 開発者・提供者における侵害発生認識可能性を評価すると、以下のように考えられる。



参考裁判例 6：最判平成 13 年 3 月 2 日民集 55 巻 2 号 185 頁

- リース業者がカラオケ装置のリース契約を締結した場合において、当該装置が専ら音楽著作物を上映し又は演奏して公衆に直接見せ又は聞かせるために使用されるものであるときは、リース契約の相手方に対し、当該音楽著作物の著作権者との間で著作物使用許諾契約を締結すべきことを告知するだけでなく、上記相手方が当該著作権者との間で著作物使用許諾契約を締結し又は申込みをしたことを確認した上でカラオケ装置を引き渡すべき条理上の注意義務を負う。
- 上記注意義務を負う理由として、(i)カラオケ装置が利用店の経営者による著作権侵害を生じさせる蓋然性の高い装置であること、(ii)著作権侵害は刑罰法規にも触れる犯罪行為であること、(iii)リース業者がカラオケ装置の賃貸によって営業上の利益を得ていること、(iv)カラオケ装置利用店の経営者が著作権使用許諾契約を締結する率が必ずしも高くないことは公知の事実であり、リース契約の相手方による使用許諾契約の締結・申込みが確認できない限り、著作権侵害の蓋然性を予見すべきであること、(v)リース業者は、使用許諾契約の締結・申込みを容易に確認することができ、これによって著作権侵害回避のための措置を講ずることが可能であることが挙げられる。

例えば、AI 開発者・提供者が利用規約において生成画像の商用利用を禁止したり、第三者の権利を侵害する態様で生成画像を使用しないよう注意喚起したりすることにより、AI 利用者が当該 AI を適切に利用し、当該 AI を通じたパブリシティ権侵害行為が発生していない場合には、権利侵害の重大性及び蓋然性並びに侵害発生認識可能性のいずれも認められず、責任が否定される可能性が高い。これに対し、AI 開発者の説明上の措置によっても適切な利用が行われていないと認められる状況下では、以下の観点から責任が評価されると考えられる。

⁶⁰ 前掲注 17 に示した過失の幫助に関する各裁判例においても、①及び②の観点で過失を認定するにあたり重要な要素として認定されている。

まず、上記のうち①提供する AI による権利侵害の蓋然性や重大性について、画像生成 AI の利用過程では、上記のとおりパブリシティ権侵害の蓋然性が典型的に高いとはいえない。また、元データが生成画像に反映されてパブリシティ権の侵害に至る可能性を低減するため、現時点の技術では、データセットの規模を十分大きくすることで過学習の可能性を低減させる、学習時に十分に長いテキストで画像を説明する（Re-captioning）、潜在空間での学習を行う、生成時に有名人や有名作品の固有名詞を大規模言語モデル等でプロンプトから取り除く（フィルタリング）等の措置が講じられることが多い。データセットが小規模である等の理由により過学習の懸念が残る場合には、上記の措置に加えて、正則化項の追加によって過学習を緩和する、画像認識 AI を用いて実在の著名人の類似画像が生成されたか否かの簡易検証を行う等の措置が講じられることがある。なお、全ての事案において上記措置の全てが必要となるものではなく⁶¹、例えばデータセットが十分に大きい AI モデルであれば、他に特段の措置を講じずとも過学習のリスクがもとより僅少な場合もある。構築する AI モデルの特性に応じ、必要かつ合理的な範囲の措置を講ずれば足り、結論として権利侵害の蓋然性が抑制されていることが重要となる。こうした措置を講じた結果、AI モデルが個々の元データではなく、画像群に共通する抽象的な特徴を学習し、特定の元データを複製する傾向が低くなっている場合には、①における権利侵害の蓋然性は低いと評価することができる。

他方、前記 3.3.2 にも述べたとおり、AI 利用者において無数に存在する権利者の肖像と逐一照合を行うことは一般に困難であり、AI 利用者において権利者への十分な配慮を行うことができず結果としてパブリシティ権侵害に至る場面も予想される。このことを踏まえると、当該画像生成 AI の機能やデータセットの規模等により、特定のプロンプト（例えば著名人の氏名や主な特徴を含む指示内容）に対し、AI モデルが再現性を持って特定の著名人に酷似した画像を生成する傾向が顕著となっているような場合には、権利侵害の防止を AI 利用者の判断のみに委ねることが適切とは認められず、①における権利侵害の蓋然性が高いと評価される可能性も否定はできない。

また、②AI 開発者・提供者における侵害発生認識可能性について、画像生成 AI は多様な目的に用いることができる以上、F においてパブリシティ権侵害の認識可能性が直ちに高いとはいえない。他方、自らの提供する AI を通じてパブリシティ権侵害が生じていることの端緒を F が認識したような場合（例えば V からのクレームがあった場合）、それ以降の AI の提供行為について過失が認定される可能性が高まる。

事例 b では、データセットは約 10 億枚の画像が含まれる大規模なものであることに加え、F は Re-captioning や潜在空間での学習等、元データが生成画像にそのまま反映されることを防ぐための措置を講じつつ、本件以外に V と類似した画像が生成されていないことから、①提供するツールによる権利侵害の蓋然性や重大性、及び、②侵害発生認識可能性のいずれも存在せず、幫助責任は成立しないと考えられる。

なお、上記で例示した権利侵害防止措置の概要は以下のとおりである。また、正則化項の追加については、後記 36 頁「＜参考＞正則化の概要」においてより具体的な内容を解説している。

⁶¹ 想定事例 4 とは異なり AI 開発者と AI 提供者が同一でない場合、バリューチェーンにおけるそれぞれの役割分担に応じ、求められる措置の内容が異なり得ることについて、前記 2.2.1(3)参照。

具体的な措置	タイミング	詳細
潜在空間での学習	学習・開発段階	AI モデルが元データの画像をピクセル空間で記憶し類似画像を生成する可能性を低減するため、入力データの重要な特徴のみを保持する潜在空間で学習を行う
Re-captioning	学習・開発段階	元データの画像に対する説明文を、固有名詞ではなくより一般化・抽象化された表現で詳述することで、当該データを特定の人物等と結びつけないようにする
データセットの規模	学習・開発段階	AI モデルが過学習する可能性を低減するため、データセットを十分大きい規模とする
正則化項の追加	学習・開発段階	過学習が疑われる、特定画像への依存度が高い、類似度検証において高い類似度が検出された等の場合に、正則化の効果を持つドロップアウトや損失関数への正則化項の追加を実施する
フィルタリング (入力・出力データ)	生成・利用段階	AI モデルの入出力をフィルタリング（除外リストフィルタ）することで、権利侵害のおそれのある画像が生成されるリスクを低減する

上記の権利侵害防止措置のほかにも、F は AI 事業者ガイドライン上求められる措置を講ずることで、幫助責任を問われる可能性を低減できると考えられる⁶²。F のサービスの特性や個々の状況に応じて求められると考えられる措置は異なるが、下記に例示されるガバナンスの構築状況は、権利侵害リスクを減少させるため可能な限りの注意を払っていたことの事情として評価されることが考えられる。

＜参考＞パブリシティ権に関して AI 事業者ガイドライン上想定される措置

パブリシティ権の侵害リスクを低減する観点で、AI 事業者ガイドライン上想定される措置を想定事例 3 の当事者ごとに整理すると下表のとおりとなる。下表はあくまで概要であるため、具体的な内容は AI 事業者ガイドラインの各箇所を参照されたい。

AI の特性や個々の状況に応じて求められると考えられる措置は異なるが、前述のとおり、必要に応じて AI 事業者ガイドライン所定の措置を講ずることは、過失と評価される可能性を低減することに繋がる。なお、前記 2.1.3 のとおり、各事業者における自主的な取組みを促すという AI 事業者ガイドラインの性質上、下記措置を講じなかったことが各当事者における注意義務違反に直ちに結びつくものではないことに留意が必要である。

⁶² 仮定的な事例として、データセットに肖像が含まれる人物が、AI モデルの開発時には無名であったものの、AI で生成された肖像の利用時には著名になっているケースも想定され、当該 AI の利用過程でパブリシティ権侵害が生じ、AI 開発者・提供者の幫助責任が争点になることが考えられる。AI 開発者・提供者が本文記載のような権利侵害防止措置その他の AI ガバナンスを履践することで、このような潜在的な権利侵害の可能性を検知・防止するための合理的な措置を講じていたと評価され、注意義務違反を問われる可能性を低減することに繋がると思われる。

			パブリシティ権に関して想定される措置	ガイドライン上の関連箇所 ⁶³
事業者G AI利用者	AI利用者	適正利用に関する措置	● 生成画像がパブリシティ権を侵害しないか合理的な範囲で調査・確認し、適正な範囲で利用する 等	U-2) i. 安全を考慮した適正利用
	AIサービス開発事業者F	AI開発者・提供者	設計上の措置	● データ学習に関する措置 ・ データセットに著名人の肖像等が含まれていないか確認を実施する ・ データの来歴等に応じ、適切な権利処理がなされているかを確認する 等 ● 継続的なモニタリング ・ モデルが著名人の容貌を再現していないか検証するためのログ・サンプルを記録し、必要に応じて改善措置を講じる 等
説明上の措置			● パブリシティ権の侵害リスクに関する説明を行う ● サービスの目的等に応じ、利用規約等で生成画像の利用範囲の制限や肖像権を侵害する態様での利用禁止条項を設ける 等	D-2) iii. 適正利用に資する開発 D-6) ii. 関連するステークホルダーへの情報提供 P-2) ii. 適正利用に資する提供 P-6) ii. 関連するステークホルダーへの情報提供

3.3.5 事例 c について

事例 c は、F が著名人と同一又は類似の肖像が高頻度で生成される AI モデルを開発した上で、著名人の様々な背景や構図の画像を生成することを目的とした AI として販売した事案である。このような例外的なケースについては、AI の提供行為そのものがパブリシティ権侵害と評価される可能性も否定できないことに留意が必要である。

画像生成 AI のモデルやサービスそれ自体は肖像等に該当しないため、その提供行為は原則としてはパブリシティ権の侵害行為には当たらない。もっとも、画像生成 AI の提供それ自体が専ら肖像等の有する顧客吸引力を利用する目的であると評価される場合には、違法性において 3 類型と実質的に同一の行為⁶⁴としてパブリシティ権侵害と評価される可能性がある。

このような見解に立つ場合、例えば、顧客吸引力を有する実在の著名人と同一又は類似の肖像を生成することが容易な画像生成 AI を、当該容易性をセールスポイントとして販売するような場合には、著名人の肖像を自由に生成して鑑賞等に用いることができることを商品価値としている点で、3 類型のうち①（著名人の画像を鑑賞の対象となる商品等として販売する行為）に近似する。また、上記のような画像生成 AI は、顧客吸引力を有する肖像等を生成できることが他

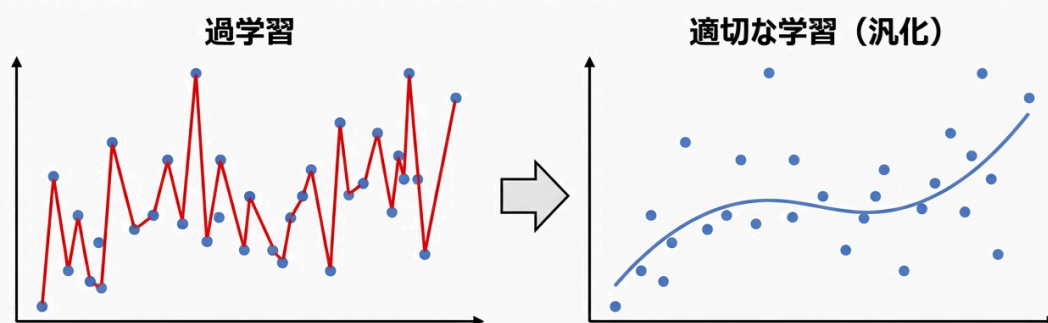
⁶³ AI 事業者ガイドライン本編 29-41 頁参照

⁶⁴ 前掲注 44・中島 50 頁では、参考裁判例 4 で 3 類型に付け加えられた「など」は、違法性において 3 類型と実質的に同一である（「均等」）と評価されるような場合に、パブリシティ権の禁止効の及ぶ範囲を例外的に拡張するものであり、予測できない利用態様が将来発生し得ることに配慮してその余地を残したものであると指摘されている。

の生成 AI との差別化要因となるため、3 類型のうち②（商品等の差別化を図る目的で肖像等を商品に付する行為）⁶⁵にも近似する。したがって、上記のような特徴を有する画像生成 AI の販売は、専ら肖像等の有する顧客吸引力を利用することを目的とした 3 類型と違法性において実質的に同一の行為として「など」に含まれると評価され得る⁶⁶。この場合、AI 利用者の行為如何にかかわらず、AI の提供行為そのものについてパブリシティ権侵害に基づく不法行為が成立すると考えられる⁶⁷。

＜参考＞ 正則化の概要

本書で主として取り扱う AI は、データの背後にある統計的な傾向やパターンを「機械学習」という手法により抽出し、認識・予測・生成等のタスクに活用するものである。しかし、AI の学習の際、実際にはサンプルにすぎないデータに過剰に適合すると、データに含まれるノイズが反映されてしまい、得られたパターンから汎用性が失われてしまう場合がある（過学習）。



過学習と適切な学習のイメージ

過学習が発生すると、学習データから一般的なパターンを抽出するのではなく、個別のデータの特徴に過度に依存した挙動を示しやすくなる。その結果、画像生成 AI が学習データに類似した画像を出力してしまうリスクや、取引審査 AI が未知の事例に対して十分な性能を発揮できず、一部の属性を持つ人が不利に評価されてしまうリスクの一因となる。このようなリスクの防止のため、特定データへの過度な依存を防ぐ「**正則化**」がある。

ただし、正則化はモデルの複雑さを制御することにより過学習を緩和するものであり、学習データの構成や目的変数の設定そのものに由来する偏りを是正するものではないため、他の対策と組み合わせる必要がある。

機械学習では、「**損失関数**^{※1}」と呼ばれる評価指標を初めに設定し、それを最小化するようにパラメータの更新を繰り返すことでモデルの学習を行う（上図の例では、モデルの予測値と実データの距離から計算され

⁶⁵ 第3類型のうち②について、前掲注44・中島47-49頁では、肖像等を利用した「キャラクター商品」を違法とする類型であると説明されている。また、①と②の関係について、①は「肖像それ自体」の鑑賞性に着目し、これを商品化する類型であるのに対し、②は肖像の使用により「商品全体」を差別化することに着目し、これをキャラクター商品として商品化する点において相違すると説明されている。

⁶⁶ なお、本事例cのような画像生成AIが販売される場合、宣伝等の過程で製品名やパッケージ、販売用ウェブページ等にVの氏名や肖像等が利用されることも想定される。このような場合には、3類型のうち、②商品等の差別化を図る目的で肖像等を商品等に付す場合、又は、③肖像等を商品等の広告として使用する場合に該当し、その側面においてもパブリシティ権侵害による不法行為責任が生じ得る。

⁶⁷ 本文で示した考え方に対し、F自身は画像生成AIを提供しているに過ぎず、Vの肖像はAI利用者の手元でのみ生成・利用されるという関係性を重視して、Fの行為は「など」に含まれずパブリシティ権侵害を構成し得ないとの考え方もあり得る。

た指標が該当する)。その際、単に予測誤差の最小化のみを目指す、上図左のような極端なモデルになりやすい。そのため、損失関数として予測誤差だけを見るのではなく、以下のようにモデルの複雑さ^{※2}を表す項(正則化項)を組み入れる。

$$\text{損失} = \text{予測誤差} + \lambda \times \text{複雑さ}$$

λ は正則化の強さを決定する係数(ハイパーパラメータ)であり、入力のわずかな違いや特定のデータに対する過剰反応を緩和することで、モデルがデータ全体に共通するパターンを抽出するように誘導し、学習データを丸暗記して出力してしまう現象の抑制に繋がる。 λ を大きくして正則化を強めれば、学習データの「丸暗記」に起因するリスクを軽減し得る一方で、モデル出力の品質への影響が生じやすくなるトレードオフが存在する。

- ※1 損失関数は、直感的には、AI が学習中に参照する採点表のようなものと理解できる。予測値と実測値のズレがどの程度発生しているかを統計的に数値化したもので、値を小さくすることができれば、より「優れた」モデルになったと評価される。
- ※2 ここでいう「複雑さ」は、技術的にはモデルの重みの二乗和等、複数の指標で定義されるが、直感的にはモデルが特定のデータに過度に依存した学習を行わないように誘導するための指標と考えてよい。

3.4 想定事例 4 : 取引審査 AI

AI サービス開発事業者 H は、賃貸物件の入居申込者の信用スコアを算出する取引審査 AI を開発・提供している。当該 AI サービスは機械学習アルゴリズムによって構成されており、ユーザーである不動産業者が入居申込者に関する所定の情報をインプットすると、当該情報に基づき信用スコアを定量的に算出した上で、一定の閾値に基づき申込者を合格と不合格に分類し、ユーザーに対して当該スコア及び可否結果を提示する。全国的に不動産賃貸業を展開する I は H の AI サービスを導入し、自社保有物件についての審査業務を効率化・省人化している⁶⁸。

【事例 a】 上記 AI では、憲法上の差別禁止属性に着目した学習は行われていなかったものの、収入及び勤務形態に関する情報を過度に重く考慮して学習した結果として、男性に比べて女性を著しく不利に取り扱いやすい統計的バイアスが含まれていた。H は当該バイアスを認識していたが特段の是正措置を取らず、ユーザーにもバイアスの内容やリスクを説明せずに提供していた。男性に比べて女性を著しく不利に取り扱うという統計的バイアスは入居審査の目的上、特に必要性のない評価傾向であった。あるとき女性 V1 が応募したところ、AI が不合格と判定したため、I の担当者 I' は AI の判断を援用し、申込みを却下した。このとき I' は、上記バイアスの存在を具体的には認識していなかったが、AI 導入後、導入前と比較して不自然に女性の申込みが却下される割合が高まっていることを漠然と認識していた。

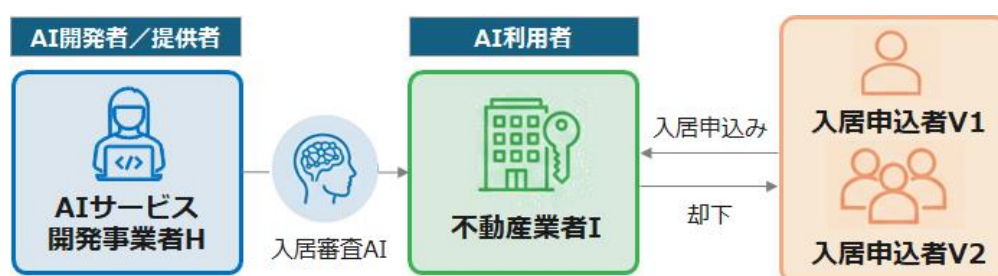
【事例 b】 H は開発段階においてバイアスに配慮した措置を講じ、宗教・信条等の属性情報は AI モデルに学習させることを避け、住所、職業、勤続年数、年収、家賃負担率、延滞履歴など、宗教・信条等とは一

⁶⁸ 想定事例 4 とは別の例として、不動産賃貸の仲介事業において、不動産事業者が取引審査 AI 用いてリスク判定を実施し、その判定結果に基づき不動産オーナーが最終的な判断を行う場合なども考えられるが、本事例では事実関係を複雑にしないよう自社保有物件についての AI 導入を前提とする。

見して関連性がない評価要素のみを設定していた。その結果、AIモデル開発時点の公平性評価において差別に繋がりのようなバイアスの懸念が見られないことを確認していた。Iはこのことを前提に当該AIを導入し、バイアスを惹起し得るデータを入力しないようにするデータ管理を行い、また運用過程においてバイアスが生じていないかを定期的にモニタリングするようHに依頼する等の措置を講じつつ、AIを審査に用いていた。

Iが上記AIを導入した後、働き方に関する政治的信条を共有して同一地域に集住する大規模なコミュニティV2が結成された。V2では短期間で職を転々とするが多かったところ、上記AIは継続学習の過程で当該地域に短期就労者が多い傾向を学習し、当該地域での居住それ自体をひとつのリスク要因として学習した。その結果、X地域に居住し、かつ職を転々とする傾向にあるV2構成員の却下率が顕著に高い状態となり、複数の構成員が不合格の判定を受け申込みを却下された。

H及びIはある時点まで上記バイアスの存在を全く認識していなかったが、共同して実施していたモニタリングにおいて上記バイアスを検知したため、当該地域の申込者については当面の間、AIの判断を参照せず人による審査を行う等の対応を採った。

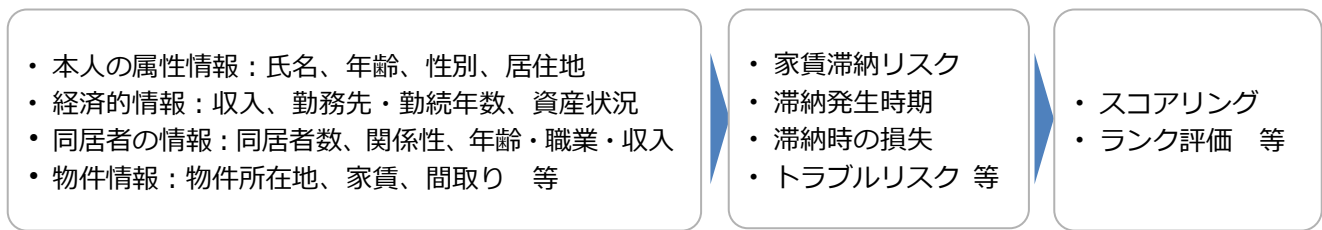


上記事例において、AIのバイアスが社会的相当性を欠く取引拒絶に繋がったと評価される場合、H及びIのVらに対する責任（民法709条）の成否が論点となる。

3.4.1 取引審査AIと公平性

近時では取引審査等の場面においてAIを活用する場面も増加しつつあるところ、従来から一般的なものとして、評価対象者に関する多様な属性を考慮して取引上重要なリスクの程度をスコアリングするものがある（下図参照）⁶⁹。AIの活用によって、多様な要素を考慮した精度の高いリスク分析が可能となることや、担当者の認知バイアス等に起因する人為的な判断の偏りを一定程度低減することも期待される。

⁶⁹ 近時では、取引審査に大規模言語モデルを活用しつつ、「判断に偏りを用いない（use no bias）」等のプロンプトによって公平性を図りながら審査を行うための研究や実装も進みつつあるが（Bowen III, D., Price, S. M., & Yang, K. (2024, August 20). *AI exhibits racial bias in mortgage underwriting decisions*. Lehigh University News. <https://news.lehigh.edu/ai-exhibits-racial-bias-in-mortgage-underwriting-decisions> 等参照）、想定事例4ではスコアリングを行う従来型のAIを前提とする。



入居審査を含む取引審査には一般に取引自由の原則が妥当し、契約相手方の選定や取引条件の設定について裁量が生ずることから、その判断は原則として違法と評価されることはない。

他方、憲法 14 条 1 項に列挙される属性など、差別に繋がりうる属性（以下「要配慮属性」という。）⁷⁰を理由とした社会的相当性を欠く取引拒絶と評価されるケースでは、人格権侵害に基づく不法行為⁷¹に該当する可能性があり（後記参考裁判例 7～11 参照）、AI を用いる場面ではどのような場合に不法行為に該当するかが論点となる⁷²。

3.4.2 想定事例 4 における AI の位置づけ

取引審査 AI は、評価対象者の種々の属性を踏まえて信用リスク等の評価を実効的に行うため、機能的には AI 利用者における取引審査上の判断を代替し得る。一方、AI の出力に一定のバイアスが含まれる場合には社会的相当性を欠く取引拒絶に結びつくリスクが伴うところ、そのような出力を援用しないよう、利用過程における人の評価や検証が必要と考えられるため、取引審査 AI も補助／支援型 AI に該当する。

補助／支援型 AI の出力に第三者の権利を侵害するものが含まれる場合、AI 利用者がその判断において適正な範囲で利用することが求められ、特に想定事例 4 のように、取引を行うか否かの最終的な判断に AI の出力を用いる場合、AI の出力をそのまま利用することが社会的相当性を欠く取引拒絶と評価される可能性も否定できないため留意が必要となる。もっとも、上記 AI はその技術の特性上、個々の出力にバイアスの観点から問題があるか否かを評価することは困難な場合が多いことから、AI モデルのバイアスに関する情報を収集するなどの一定の措置が注意義務の中心になるものと考えられる。また、AI 開発者・提供者においても、AI 利用者に対するバイアスについての説明や一定の緩和措置を講ずるべきでなかったかが論点となる（後記 3.4.4 以下参照）。

⁷⁰ 個人の特徴のうち、どこまでの範囲について公平性が要求されるかは、それぞれの国や地域における文化的・社会的背景によっても異なり、一義的に確定することは困難である。そのため想定事例 4 では、少なくとも憲法 14 条 1 項に基づき差別が禁止される性別や信条に基づく契約拒絶を想定し、責任論上どのような点がポイントとなるかについて検討を行う。

⁷¹ 取引に付随する信義則上の義務違反に基づく不法行為と構成される場合もある（京都地判平成 19 年 10 月 2 日（判例集未登載）等）。

⁷² 日本銀行金融研究所が設置した金融機関における AI の利用を巡る法律問題研究会の報告書（日本銀行金融研究所『「金融機関における AI の利用をめぐる法律問題研究会」報告書—金融機関における AI 利用に伴う私法上のリスクと管理』21 頁（2025 年））においても、例えば、依拠／代替型の AI を用いた融資審査において、AI の判断に不公正なバイアスが反映されていた場合には、融資審査を公正に行うという信義則上の義務に違反し得ると指摘されている。

3.4.3 取引拒絶が不法行為に該当する要件について

従来の裁判例では、人が要配慮属性に着目して取引拒絶を行ったことが権利侵害性及び故意・過失の共通の基礎となっており、権利侵害要件と故意・過失要件を分けて論じていないものが多数である。

一方、AI を用いる事例では、以下に述べるとおり、AI の出力が法的評価として人格権の侵害に該当するか否かという権利侵害性と、各当事者が当該権利侵害を防止する義務を負っていたか否かという故意・過失の問題は質的に異なり得るため、以下では分けて整理する。

(1) 権利侵害性

従来のように人が取引審査を行う場面では、裁判例上、最終的な判断においてどのような要素が重視されたか（要配慮属性か、それ以外の要素か）が確定できたことを前提に、①当該取引における相手方の属性の重要度や②相手方の不利益の程度を勘案しつつ、主として③目的の合理性や④手段の適切性が求められるか否かによって権利侵害性が判断されてきた。

すなわち、要配慮属性を取引審査に用いること自体が問題視されるものではなく、特に目的の合理性や手段の適切性が認められない例外的な場合には違法と評価される可能性がある。

AI によって審査を行う場合でも同様の判断基準が適用されると考えられるが、AI はその技術の特性上、出力されたスコアや評価結果を一見しても、その判断の根拠が明確でなく、個々の審査において要配慮属性が重視されたか否かを事実として確定することは困難な場合も多い。

したがって、まず、AI モデルの評価傾向の公平性を評価するための下表のような指標（公平性評価指標）により、要配慮属性を有する対象者とそうでない対象者とでどの程度評価傾向に違いがあるのか（すなわち、要配慮属性が審査の主たる要素となっているか）を検証することが、権利侵害の有無を認定する出発点になると考えられる。

指標	内容	具体例
機会の公平性 (Equal Opportunity)	実際に合格すべき（正解ラベルが 1 の）人を AI が合格と判定できた割合（True Positive Rate, “TPR”）が属性によらず同じであるか	男性の TPR=80%、女性の TPR=60%であれば、女性は本来合格すべき人の 20%分が不合格とされている
確率予測の 信頼性 (Calibration 差)	AI が付与した予測スコアが、属性ごとに実際の合格確率と一致しているか	男性のスコアを 0.8 と予測した場合、実際に 80%の男性が合格するが、女性のスコアを 0.8 と予測した場合、60%しか合格しない
合格比率の差 (Demographic Parity ⁷³)	AI が合格と判定する割合が属性ごとに同じか	AI が合格と判断する割合が男性 60%・女性 40%であり 20%の乖離がある

⁷³ 応募者集団に収入水準等の実力差がある場合、合格比率の均等を強制すると逆差別や業務上の弊害を生ずる可能性があるため、指標としての位置づけには留意が必要である。

このような公平性評価の結果、要配慮属性が審査の主たる要素になっている場合は、下記①～④の評価を行うこととなる。想定事例 4 の入居審査に即して検討すると以下のとおりと考えられる。

① 当該取引における相手方の属性の重要度

- ✓ 一定の長期間に亘って継続する賃貸借契約を締結する場面では、一般に賃貸人からの中途解約が困難であることからすると、このような長期の契約関係に入る賃借人の属性は賃貸人にとって一般に重要な考慮要素であり、相手方の属性を審査上考慮することについての賃貸人の利益は尊重される。

② 相手方の不利益の程度

- ✓ 評価対象者は、取引を拒絶されたとしても他の賃貸物件を契約することも可能と考えられるが、賃貸物件の契約は生活の基盤となる重要な機会である。
- ✓ また、複数の賃貸人が用いる汎用的な AI サービスにおいて評価対象者が低く評価される場合、評価対象者が選択できる物件がより少なくなるということになるから、評価対象者の不利益がより大きくなる一方、ある特定の賃貸人のみのために開発されたサービスにおいて評価対象者が低く評価された場合には、評価対象者は当該 AI の審査対象とならない別物件を選択することが可能であるから、評価対象者の不利益は相対的に小さくなる。もっとも、生活の基盤を構築する評価対象者の上記利益は十分に配慮されるべきである。

✓

③ 目的の合理性

- ✓ 事業目的との関係で、要配慮属性を有する者が不利に評価される傾向にあることに合理的な理由があることが求められる⁷⁴。入居審査との関係では、賃借人について懸念されるリスク（家賃の滞納や居住上のトラブル等）との関係で、当該属性がどのような結びつきを有するのかが重要と考えられる。
- ✓ AI を用いた審査において目的の合理性を基礎づける上では、学習させる属性を決定するにあたり、考慮することに合理的根拠のある属性を選択し、必ずしも合理的根拠があるといえない場合には代替する考慮要素を検討すること⁷⁵等が重要と考えられる。

④ 手段の適切性・代替手段の有無

- ✓ 目的との関係で一律に要配慮属性を有する評価対象者との取引を拒絶することの合理性や、より間接的・具体的な属性に着目して審査を行う代替手段を検討したかどうか、及びそのような代替手段の有無等が問われる。

⁷⁴ 例えば、返済可能性を確保する等の事業目的との関係で定型的な条件を付する必要性の高さ等を考慮し、外国人に対する住宅ローンに永住資格要件を定める規定について目的の合理性を認めたものがある（後記参考裁判例 8）。

⁷⁵ AI 事業者ガイドライン別添 98 頁参照

- ✓ AI を用いた審査との関係では、必要に応じて AI 事業者ガイドライン上想定される以下のような措置⁷⁶を講じていた場合、手段の適切性を裏付ける方向に評価されると考えられる。
 - ・データの比率やラベルの見直しによる緩和措置を講じたこと（それでもなお一定のバイアスが残存したこと）
 - ・要配慮属性へのバイアスがより小さくなるよう他の考慮要素を採用したこと
 - ・公平性に関するペナルティ項を追加した正則化の実施⁷⁷
 - ・バイアスによって合否に影響が生じやすい範囲を中心に、人による関与が必要な範囲を設計すること 等
- ✓ また、審査の結果について評価対象者に不服がある場合の異議申立手続を設けていた場合、評価対象者の不利益を緩和する措置として斟酌され、手段の適切性を高めることに繋がると考えられる。

権利侵害性の判断は個別の事案を総合的に考慮する必要があるが、傾向としては以下のように整理されることが考えられる。

- ✓ AI が行う統計的な処理において、要配慮属性が審査の主たる要素になっていることを示す評価傾向が認定できる場合であって、かつ、そのような評価傾向が審査の目的との関係で合理的なものであることが説明できない場合や、対象者の不利益の緩和措置が講じられていない場合、権利侵害性を基礎づける方向に傾く。
- ✓ 要配慮属性が審査の主たる要素になっている場合であっても、それが審査の目的との関係で合理的であり、かつ必要に応じて対象者の不利益の緩和措置が講じられている場合には、必ずしも権利侵害には当たらない。
- ✓ 要配慮属性とは中立的で、審査に通常用いられる評価要素（就労形態、職種、収入等）を中心とした評価アルゴリズムを採用している場合、目的の合理性や手段の適切性が認められやすい。もっとも、このような中立的な要素が代理変数となり、結果として合理的な理由なく要配慮属性を不利に評価する傾向となっていないかには留意を要する。



参考裁判例 7：女性専用車両の提供（東京地判平成 23 年 7 月 12 日（判例集未登載））

- 女性専用車両の提供は、平日の通勤通学の時間帯に相当な混雑をする首都圏等大都市圏の通勤電車において、痴漢犯罪の被害を受けるおそれのある女性の乗客に対し、少しでも安心、快適な通勤通学環境等を提供するために行われていると解せられ、これは目的において正当というべき。
- 女性専用車両が設定されるのは、平日の通勤時間帯の一部電車で、6 両編成の車両のうちわずか 1 両のみに過ぎず、これは健常な成人男性の乗客をして他の車両を利用して目的地まで乗車することを困難ならしめるものではないから、健常な成人男性の乗客に対し格別の不利益を与えるものでもない。
 - 目的の合理性及び手段の適切性（特に、車両の時間帯や数が限定的に運用されていること）を理由に不法行為の成立を否定した。

⁷⁶ 後記 47 頁「＜参考＞公平性に関して AI 事業者ガイドライン上想定される措置」参照。

⁷⁷ 前記 36 頁「＜参考＞正則化の概要」参照。



参考裁判例 8：住宅ローン審査における永住資格要件

(東京地判平成 13 年 11 月 12 日判時 1789 号 96 頁、東京高判平成 14 年 8 月 29 日金商 1155 号 20 頁)

- 銀行が永住者の在留資格のない外国人からの住宅ローンの申込みを拒絶した行為は、憲法 14 条などに違反するものではない。
 - 住宅ローンは、長期の返済期間を設定すること等から融資に要するコスト（契約事務費、債権・担保管理費、債権回収費用など）が相対的に高く、採算性の高い商品であるとはいえないから、銀行が十分な採算を取るためには大量の住宅ローンを取り扱うことが必要であり、そのためには、貸付対象者等の要件をできるだけ定型化する必要がある。
 - 永住資格を有しない外国人は在留期間が 3 年間に制限されており、本邦を出国した場合、その債権管理及び債権回収に要する費用、時間、労力が多大なものにならざるを得ないこと等を踏まえると、在留資格の有無によって定型的に融資資格を判断することは合理性に欠けるものではない。
- 住宅ローンという契約の相手方の属性が重要な取引において、銀行の営業の自由を踏まえつつ、定型的な取引条件を設ける必要性や、借主の出国による債権回収費用の増大等を考慮し、一律に永住資格要件を求めることの合理性を認定した。



参考裁判例 9：公衆浴場における入浴拒否（札幌地判平成 14 年 11 月 11 日判時 1806 号 84 頁）

- 被告の施設は公衆浴場法に基づく知事の許可を受けて経営されており、公共性が認められることを踏まえ、すべての外国人の利用を一律に拒否するのは明らかに合理性を欠くものというべきであり、他の入浴者に迷惑をかけたり入浴マナーを守らなかったりする者に対しては個別に指導・対応すべきであった。
- 被告施設の公共性を踏まえ、国籍に基づく一律の入浴拒絶ではなく、個別に指導・対応を行う等の代替措置を行うべきであったと指摘して不法行為の成立を認めた。



参考裁判例 10：ゴルフクラブの入会拒絶(1)（静岡地浜松支判平成 26 年 9 月 8 日判時 2243 号 67 頁）

- 女性への性別変更の審判を受けた原告に対する入会拒絶の事案において、被告のゴルフクラブには厳格な入会要件が存在しないこと等から、団体としての閉鎖性が認められないとした上で、原告が被告のゴルフ場を用いることによる既存会員への強い不安感や困惑といった危惧は抽象的なものにすぎず、他方、原告が被る精神的損害は看過できないものであるから、被告による入会拒絶は違法である。
- 団体としての閉鎖性が認められず、性別変更を考慮することの合理性や原告に対する不利益の程度を考慮した上で、不法行為の成立が認められている。



参考裁判例 11：ゴルフクラブの入会拒絶(2)

(東京高判平成 14 年 1 月 23 日判時 1773 号 34 頁・原審東京地判平成 13 年 5 月 31 日金判 1138 号 35 頁)

- 入会手続の厳格さ等に鑑みて閉鎖的な性質が高いと認められ、外国人の入会を一律に制限している場合であっても不法行為の成立が否定される。
- 原告の不利益の小ささを示す事情として、会員資格を取得したとしても、ゴルフ場の利用等における便宜が得られるに過ぎないことが挙げられる。

→ 参考裁判例 10 とは異なり閉鎖性が高いと認められ、一律に国籍要件を定めるケースでも不法行為の成立が否定されている。

(2) 故意・過失

上記(1)において権利侵害が認められた場合、各当事者がどのような注意義務を負うかが問題となる。

前述のとおり、AI はその技術の特性上、個々の出力にバイアスの観点から問題があるか否かを評価することは困難な場合が多い。この点を踏まえると、AI 利用者の注意義務は、個々の出力に対する評価・検証よりも、AI モデル全体の公平性に関して合理的に可能な範囲で確認を行い、そのリスクに応じて人の判断を介在させたり AI の導入の適否を検討したりすること等が中心になると考えられる。また、AI 利用者が適正な範囲で AI の出力を用いることができるようにする観点で、AI 開発者・提供者も一定の設計上・説明上の措置が求められやすい類型となる。

3.4.4 事例 a について

事例 a は、開発時点でバイアスに関する懸念が存在していたにもかかわらず、それが審査に用いられて取引拒絶に至った場合である。このような場合、各当事者は以下のような責任を負う可能性がある。

(1) 権利侵害性

事例 a において、I が全国的に不動産賃貸業を展開しており AI の影響範囲が大きいこと、AI が男性に比べて女性を著しく不利に取り扱いやすい評価傾向にあること、H が当該傾向について特段の是正措置を取っていないこと、当該傾向は審査上も特に必要性のない評価傾向であったことは、違法な権利侵害と評価される可能性を高める事情となる。

当該 AI は直接的に要配慮属性を不利益に評価しているわけではないものの、結果として要配慮属性が審査の結果に大きく影響しており、そのような評価傾向が審査の目的との関係で合理的なものであることが説明できない場合や、対象者の不利益の緩和措置が講じられていない場合には、権利侵害と評価される可能性がある点に留意を要する。

当該 AI の出力に基づき取引拒絶を行うことが権利侵害と評価される場合、H や I は、権利侵害を防止する観点で以下のような措置を講ずべきでなかったかが問われる。

(2) AI 利用者である I の責任

想定事例 4 における AI は補助／支援型 AI であるから、I は AI を審査に用いるか否かにかかわらず、申込者との関係で公平に審査を行うべき立場にあり、AI を用いる場合には、権利侵害を伴う出力をそのまま審査上の判断として援用すべきではないこととなる。もっとも、前述のとおり、AI が出力する個々のスコアや評価結果を一見しても、バイアスの観点から問題があるか否かを発見することは困難な場合も多いことに鑑みると、最終的な審査上の判断において AI の出力を適正な範囲で用いるためには、出力以外の要素を勘案する必要がある。したがって、I が

負う注意義務は、個々の出力レベルの評価・検証よりも、AI モデル全体の公平性に関して合理的に可能な範囲で確認を行い、そのリスクに応じて人の判断を介在させたり AI の導入の適否を検討したりすること等が中心になると考えられる。この点に関し、AI 事業者ガイドラインでは、AI 利用者に対し、必要に応じて AI モデルのバイアスに関する情報を収集することや、バイアスに配慮しつつ人間の判断を介在させること等の措置が求められている（後記 47 頁「＜参考＞公平性に関して AI 事業者ガイドライン上想定される措置」参照）。

事例 a においては、I が審査に AI を導入した時点で性別に関する顕著なバイアスが存在していたこと、また、AI 導入後に女性の申込みが却下される割合が不自然に高まっている傾向を I の担当者 I' が認識していたことを踏まえると、AI モデルのバイアスに関する簡単な情報収集等を行うことで権利侵害のリスクを認識でき、これを回避することが可能であったと評価される可能性がある。したがって、I は、H から AI モデルの公平性に関する情報を収集し、必要に応じて H に対しバイアスの低減措置を要請したり、よりバイアスに関する懸念が小さい AI の導入を検討したり、あるいは AI の情報を参考に留めつつ人の判断を介在させる等の措置を講じなかったことについて、注意義務違反を問われ得る。

(3) AI 開発者・提供者である H の責任

H は補助／支援型 AI を提供する者として、前記 2.2.1 のとおり、①AI の機能や性能の限界、使用方法、重要なリスク等に関する明確な説明を行ったか、②AI 利用者が予見又は対応することが必ずしも容易でないリスクについて一定の権利侵害防止措置を講ずるべきでなかったかが問題となる。前述のとおり、AI 利用者が個々の出力においてバイアスの観点から問題があるか否かを発見することは困難な場合も多いことに鑑みると、H が把握しているバイアスのリスクについては合理的な範囲で軽減措置等を講ずるか、残存リスクについて I に対し明確に説明することが求められると考えられる。

事例 a では、AI の開発時点で性別に関する顕著なバイアスが存在していたにもかかわらず AI 利用者にその点を説明していなかったことに鑑みると、上記①及び②いずれの観点においても H が所要の措置を怠っていたと評価される可能性が高い。すなわち、評価要素の重み付けの見直し等の措置を講ずることにより、不必要なバイアスが審査結果に影響を与えることを防止する措置を講じなかったことや、バイアスの概要を I に説明しなかったこと、必要に応じて人の判断の介在を推奨しなかったこと等について注意義務違反を問われ得る。

3.4.5 事例 b について

事例 a のような事例に対し、当事者がバイアスに慎重に配慮した開発・提供・利用を行っていたとしても予想外のバイアスが生じてしまう可能性がある。事例 b では、外部的な社会構造の変化に伴い生じたバイアスに関する責任判断が論点となる。

(1) 権利侵害性

事例 b において、V2 に所属する者を却下する顕著な評価傾向が生じてしまったことが権利侵害と評価されるか否かは、当該傾向の要因次第と考えられる。このような評価傾向が、審査にお

いて通常考慮される要素（就労形態、職種、収入等）によって説明可能である場合には、目的の合理性や手段の適切性が認められ、権利侵害と評価されない可能性が高まる。他方、審査上考慮する必要性が低い要素が影響し、かつ当該要素を採用することの合理性・相当性を基礎づける事情が乏しいにもかかわらず、結果として V2 に所属する者を不利に評価する傾向が顕著になっている場合、権利侵害と評価されるおそれもある。

仮に権利侵害性が認められる場合であっても、上記のようなバイアスの発生経過に鑑み、H や I に過失が認められるか否かが問題となる。

(2) AI 利用者である I の責任

I は前記 3.4.4(2)と同様、バイアスのリスクに関して合理的に可能な範囲で確認を行い、そのリスクに応じて人の判断を介在させたり AI の導入の適否を検討したりすること等について注意義務を負う可能性がある。もっとも、事例 b におけるバイアスは AI 導入後の社会構造の変化に伴って生じており、V2 に対するバイアスが生ずるリスクを予見して所要の措置を検討することは著しく困難と考えられるため、少なくとも AI の導入時点では I に過失は認められない。

他方、事例 b のようなバイアスが事後的に現れ、第三者に対する権利侵害と評価される程度に至った場合、AI の出力をそのまま用いることを防止するための合理的に可能な措置を講じていなければ、そのことについて過失が認められる可能性がある。このような措置の一つとして、AI 事業者ガイドラインにおいて、AI 利用者が出力結果を合理的な範囲で AI 提供者等にフィードバックしつつ、AI 提供者において定期的又は必要に応じて随時バイアスの動向をモニタリングするなど、潜在的なバイアスを検知するための協調的な措置が求められている（後記 47 頁「＜参考＞公平性に関して AI 事業者ガイドライン上想定される措置」参照）。I が H と連携しつつこのような措置を行い、それでもなお当該リスクが合理的に予見困難なものと評価された場合には過失は認められないと考えられる。

他方、I によるモニタリング等を通じて上記バイアスの端緒が見られた場合など、客観的に見てバイアスの存在を認識可能となった時点以降には、過失が認められる可能性もある⁷⁸。

(3) AI 開発者・提供者である H の責任

事例 b におけるバイアスの発生経過に鑑みると、少なくとも開発過程やサービス提供開始の時点では、V2 との関係でバイアスが生ずるリスクを予見して緩和措置等を検討することは著しく困難と考えられるため、この点において過失は認められない。他方、AI 事業者ガイドラインでは、AI 利用者らのフィードバックを踏まえた入出力及び判断根拠の定期的な評価や、評価傾向の継続的なモニタリングが求められているところ、このような措置を講じた上で必要な設計や説明を検討し得たか否かが焦点となる。このような措置を講じていたにもかかわらず予見が困難なリスクであると評価された場合、設計上・説明上の過失は認められない。

⁷⁸ 想定事例 4 では、H 及び I がバイアスの兆候を認識した時点で権利侵害を回避するための措置を講じているが、仮に、権利侵害の可能性を認識したにもかかわらず合理的な措置を講じなかった場合、そのことが不作為による過失と評価される可能性もある。

他方、モニタリングを通じて客観的に見てバイアスの存在を認識可能となった時点以降や、そもそもモニタリング等の措置を全く行っていなかった場合には過失が認められる可能性があり得る。Hはアルゴリズムの開発者としてバイアスの評価・分析に関し技術的・実務的知見を有する立場にあることを踏まえると、バイアスの存在が合理的に検知可能と判断される時点はAI利用者よりも早まり得る。

＜参考＞ 公平性に関して AI 事業者ガイドライン上想定される措置

公平性の観点で AI 事業者ガイドライン上想定される措置を想定事例 4 の当事者ごとに整理すると下表のとおりとなる。下表はあくまで概要であるため、具体的な内容は AI 事業者ガイドラインの各箇所を参照されたい。

AI の特性や個々の状況に応じて求められると考えられる措置は異なるが、前記 3.4.3 以下に述べたとおり、必要に応じて AI 事業者ガイドライン所定の措置を講ずることは、権利侵害や過失と評価される可能性を低減することに繋がる。なお、前記 2.1.3 のとおり、各事業者における自主的な取組みを促すという AI 事業者ガイドラインの性質上、下記措置を講じなかったことが各当事者における注意義務違反に直ちに結びつくものではないことに留意が必要である。

		公平性に関して想定される措置	ガイドライン上の関連箇所
不動産業者 I	AI 利用者	● プロンプトに含まれるバイアスへの配慮 ● 必要に応じて AI モデルのバイアスに関する情報を収集 ● バイアスに配慮しつつ人間の判断を介在させること ● 出力結果を合理的な範囲で関係者に提供	U-3) i. 入力データ又はプロンプトに含まれるバイアスへの配慮（本編 40 頁、別添 168-169 頁）
	AI 提供者	● データの公平性や参照する情報、連携する外部サービス等のバイアスを検討 ● AI モデルの入出力及び判断根拠を定期的に評価 ● バイアスの発生をモニタリングした上で必要な措置を検討 ● UI やビジネスプロセスの設計による AI 利用者又は業務外利用者へのバイアスの影響を検討	P-3) i. AI システム・サービスの構成及びデータに含まれるバイアスへの配慮（本編 37-38 頁・別添 140-142 頁）
AI サービス開発事業者 H		● 憲法 14 条 1 項等を考慮した要配慮属性の決定 ● データに含まれるバイアスに対する学習前・学習中・学習後を通じた緩和措置（代表性確保、潜在的バイアスの検知と除去等） ● 公平性指標（メトリクス）によるアルゴリズムの公平性評価 ● 人間による判断を行うべき範囲の設計 ● モデルの特性等に関する関係者への透明性確保	D-3) i. データに含まれるバイアスへの配慮（本編 29 頁、別添 98-99 頁） D-3) ii. AI モデルのアルゴリズム等に含まれるバイアスへの配慮（本編 30 頁、別添 106-107 頁） D-6) ii. 関連するステークホルダーへの情報提供（本編 31 頁、別添 115-116 頁）

第4章 依拠／代替型 AI に該当し得るケース

本章では依拠／代替型 AI に該当する可能性のあるケースを取り上げる。依拠／代替型 AI に該当し得るものとしては、想定事例 5 のようにソフトウェアとして用いられる高精度な外観検査 AI（後記 4.1）や、想定事例 6 のようにソフトウェアとハードウェアが結合した自律走行ロボット（後記 4.2）がある。

また、今後、AI エージェントのようなシステムも依拠／代替型 AI に該当する可能性がある。後記 4.3・想定事例 7 では、補論として、企業のカスタマーサポートに導入される AI エージェントを想定し、依拠／代替型 AI と補助／支援型 AI のそれぞれに該当する際の基本的な視点について論じる。

依拠／代替型 AI への該当性及び各当事者の責任の具合的内容を検討するに当たっては、業務の内容・目的やリスクに照らした AI の位置づけ及び精度が重要な要素となる。

想定事例 5 のように、検品受託業務という限定された業務範囲との関係では、AI が当該業務を高精度で遂行する場合、依拠／代替型 AI に該当すると評価しやすい。一方、想定事例 6 のような倉庫内の運搬業務との関係では、作業員との衝突といった重大なリスクが伴うことから、相対的に高い安全性が求められるところ、運搬業務の一部をロボットの自律走行機能に委ねるに当たり、安全に走行するための環境構築を行い、走行範囲を一定の環境や条件下に制限することで必要な安全性を確保すべき類型と考えられる。また、想定事例 7 のように消費者に対して商品購入の意思決定における重要な情報を提供する場面でも、消費者保護の観点から相対的に高度の注意義務が認められやすく、そのために依拠／代替型 AI に求められる水準も高まる。

以上のような構造的な違いを背景に、依拠／代替型 AI のカテゴリの内部においても、AI 利用者が講じるべき措置や、AI やロボットを販売又は提供する AI 開発者や製造者等の責任の具体的内容が異なる。また、想定事例 6 のようなロボットとの関係では、製造者の責任に関し、製造物責任の枠組みに照らした検討も必要となる。

4.1 想定事例 5：外観検査 AI

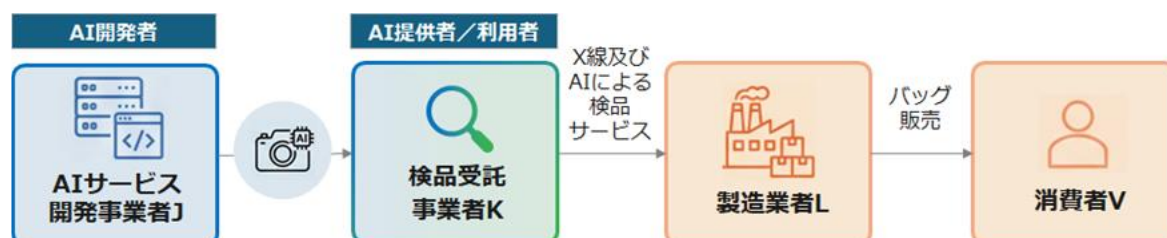
製造品の検品を受託する事業者 K⁷⁹は、AI サービス開発事業者 J が開発した画像認識 AI を X 線検査装置と組み合わせ、製造品を検査して金属片などの異物を高精度で発見することが可能な検品サービスを提供している。

⁷⁹ K においては、K の品質管理部門が J の開発した AI システムを導入し、自社の X 線検査システムに組み込んだ上、K の検査部門の業務に供している事例を想定する。このとき、K は AI 提供者及び AI 利用者としての地位を有する（AI 事業者ガイドライン本編 5 頁、同別添 9 頁参照）。K は AI を外部提供する立場にはないものの、AI 事業者ガイドライン上、AI 提供者は「AI システムをアプリケーション、製品、既存のシステム、ビジネスプロセス等に組み込んだサービスとして AI 利用者（中略）に提供する事業者」を指し（前記 1.3）、導入した AI システムを管理して社内の事業部門等に提供する場合も AI 提供者に該当し得ることに留意を要する。

バッグの製造業者 L は、全数検査における金属片検出について K の検品サービスを定期的に利用していた。L からの受託業務との関係で、K において上記 AI は特定の条件下で 1mm 以上の異物に対する検出確率約 98% となるように調整されていた。また、当該 AI は 1 次スクリーニングに用いられており、AI による判断では不確実性が高い画像を人間の 2 次レビューに回すことで、従来の人手のみによる確認に比べ精度の高い検品を約半分の短納期で処理することができる。

受託開始から 3 年間、K の検品において重大な見落としは発生せず、紙片や布片等の混入によるクレームが数件生じた程度であった。しかし、受託開始から 4 年目に、バッグに混入したカッターナイフの刃先(2cm)を K が発見できず、L からバッグを購入した消費者 V が負傷した。ロット番号や AI のログ等を踏まえ調査した結果、当該刃先は AI の検出対象範囲内の異物であったが、バッグの部品の陰に隠れており、AI がスクリーニング段階で非常に低い確率で見逃したことが判明した。

なお、K と L との間の業務委託契約においては、「K が商業上合理的な注意及び技能をもって、L の指示に従い検品作業を行う」との抽象的な条項が定められていたほか、サービスの具体的な精度についての合意はなかった。



上記事例において、まずバッグ販売事業者である L は、販売したバッグに異物が混入しており消費者 V が負傷したことについて、製造物責任（製造物責任法 3 条本文）を負うと考えられる^{80, 81}。

その上で、当該異物は検品サービス提供者である K が検品を受託した範囲内のロットから生じているところ、K は、L との関係で業務委託契約上の善管注意義務違反に基づく責任（民法 415 条）を負うか、また V との関係では一般不法行為責任（民法 709 条）を負うかが論点となる⁸²。

また、J については、AI が結果として異物を見落としたことについて何らかの責任を負うかが論点となる。

4.1.1 想定事例 5 における AI の位置づけ

⁸⁰ 製造物責任における欠陥の判断基準については後記 4.2.2 参照。想定事例 5 のような製造工程における異物混入の場合、製造上の欠陥に該当すると考えられる。

⁸¹ なお、L が K に検品業務を委託せず、自社で AI を用いた検品を行う場合も想定される。このような場合、製造上の欠陥として過失判断を介さずに製造業者の製造物責任が生じると考えられる。

⁸² いずれの請求との関係でも、K がどのような義務を負っていたかが規範的に評価されることとなるため、責任判断の構造は類似しており、後記 4.1.2 では区別して論じていない。

想定事例 5 のようなバッグの製造の場面では、従来は X 線検査の画像を人の検査員が確認して異物の有無を確認する等の手法が採られてきたが（後記参考裁判例 12 も参照）、近時では、製造品の外観検査等において異物を高い確率で検出する高精度な外観検査 AI の導入が進んでいる。

このような AI は、当該業務において人が行すべき検査上の判断を実効的に処理する機能を有しており、補助／支援型 AI にも依拠／代替型 AI にも該当し得るが、下記のように AI が一定の要件を満たし、当事者間においても人の判断や行動の一部を代替する前提で AI が提供されている場合、依拠／代替型 AI として各当事者の責任を検討すべき類型と考えられる。

想定事例 5 では、1 次スクリーニングに AI を用い、AI による判断では不確実性が高い画像を人間の 2 次レビューに回すことで、従来の人手のみによる確認に比べ精度の高い検品を約半分の短納期で処理することができる。このように迅速かつ大量の処理が可能となる場面は、AI に判断を委ねる必要性が典型的に高いケースと考えられる。

また、想定事例 5 の検品業務は従来、基本的に人が行ってきた業務であるところ、このような業務を代替するに当たり、AI システムが十分な精度水準を有しているか否かについては、前記 2.2.1 のとおり同種業務に従事する通常人の作業水準が基準となり、これと当該システムの精度とを比較することが考えられる。想定事例 5 と同種の検品作業に求められる注意義務の水準を示したものとして下記の参考裁判例 12 がある。想定事例 5 のように異物の検出率が非常に高精度な AI モデルを組み込んだ AI システムであれば、通常人としての注意力を有する者が同種作業で通常発見可能な異物の大部分を発見する程度の精度を超えた水準にあるものと評価され、AI の判断に依拠する合理性が認められやすい。また、前記 2.2.1 のとおり、AI システムが業務プロセスの一部でのみ通常人と同等以上の精度水準を発揮する場合や、特定の条件下で精度が低下し通常人の作業水準に満たなくなる場合等には、例えば AI の検出精度が低い領域を中心に人の関与を介在させることを含めた業務プロセス全体として精度を担保することも考えられる。



参考裁判例 12：検品委託契約に基づく受託者の義務の水準（東京地判平成 26 年 3 月 14 日（判例集未登載））

- 原告が、取引先に納品するバッグ等への金属異物混入の有無の検査を、X 線検査機器を有する被告に有償で依頼したところ、被告の検査作業に債務不履行があったとして、損害賠償及び不当利得の返還を求めた事案。
- 口頭で締結され、検針作業の精度について明示の合意もされていない本件契約における被告の検針業務は、モニターを担当者が注視して異物の有無を判定確認する方法で行われる以上、異物混入を完全に特定排除するという結果自体が債務内容であるとまでは解されず、安全性確保という目的を達するための手段たる債務として、これに必要な精度を備えたものとして合意されたと解すべきであり、本件契約では、検針作業の精度に関して、平均的な注意力を有する者が同種作業で通常発見可能な異物の大部分を発見する程度の精度を有することが債務の内容となる。

以上を踏まえると、上記の外観検査 AI は依拠／代替型 AI に該当する可能性が高いと考えられる。

4.1.2 AI 提供者⁸³・利用者である K の責任

前述の外観検査 AI は依拠／代替型 AI に該当する可能性が高いところ、AI 利用者である K の注意義務は、前記 2.2.2 のとおり、①AI システムを組み入れた業務プロセスを適正に「構築」とともに、②リスクを可能な限り低減しながら「運用」を行っていたかという観点から評価される。

AI システムを組み入れた業務プロセスの構築及び運用に関して具体的にどのような措置を講ずるべきかについては、個々の AI システムの特性等によっても異なるが、前記 2.2.2 に述べたとおり、AI 事業者ガイドラインの内容を参考にして合理的に可能な範囲の措置を講ずることが考えられる。K は AI 提供者及び AI 利用者に該当するところ、AI 事業者ガイドラインとの関係では、例えば後記 52 頁「＜参考＞誤検知に関して AI 事業者ガイドライン上想定される措置」のような措置が求められ得る。想定事例 5 との関係では、上記①の業務プロセス構築の観点で、人による関与を行うべき範囲の適切な設定やレビューに習熟した担当者の人員配備等が重要な要素となり、上記②の運用の観点では、AI モデルの劣化有無の検証や実運用において生じた異物や誤検知を継続学習させて AI モデルの精度を高めること等が重要な要素になると考えられる。

後記 52 頁の表の措置全てが結果回避義務を構成するものではなく、K のサービスの内容、AI の性能・設計、AI 開発者との役割分担等⁸⁴、個々の状況に応じて求められると考えられる措置は異なるが、必要に応じて当該表に記載のようなガバナンスを構築することにより、K が業務プロセスの構築及び運用の観点で合理的に可能な注意を果たしていた事情として評価されることが考えられる⁸⁵。

上記①及び②の観点において K の過失が認められない場合、AI が個別の場面の処理として誤検知に至った場合であっても、K は AI の出力を全件検証して是正するまでの結果回避義務は負わないと考えられる。

4.1.3 AI 開発者である J の責任

AI を開発し納品した J の責任が問題となる場面は、前記 3.1.3 や 3.2.3 と同様、(1)V の損害を賠償した K が J に求償又は損害賠償請求を行う場面⁸⁶と、(2)V や L が (K に対する請求と共に) J に対し直接に損害賠償請求を行う場面が想定される⁸⁷。

⁸³ K は AI 事業者ガイドライン上 AI 提供者にも該当することについて、前掲注 79 参照。

⁸⁴ AI 開発者及び AI 提供者に求められる設計上の注意義務及び説明上の注意義務がバリューチェーンにおけるそれぞれの役割分担に応じて異なり得ることについて、前記 2.2.2(3)参照。

⁸⁵ 本研究会では、想定事例 5 において、検品受託事業者である K は製造業者 L から業務を受託した範囲でのみ検品の責任を負うことから、検品受託事業者 K の行為義務の水準はあまり高くないと考えられるため、議論の一般化に用いる場合はこの点についての留意が必要であるとの指摘があった。

⁸⁶ 前記 3.1.3(1)と同様、個別の契約内容次第であるが、AI の性能・品質についてどのような合意がなされていたか、契約上予定されていた性能・品質等に満たなかった場合にどのような責任が規定されているか、どのような免責条項が設けられていたか等が重要な要素となると考えられる。

⁸⁷ 特に、前記 4.1.2 の検討の結果として K が損害賠償責任を負わない場合、V や L としては J の責任が認められるか否かが重要な論点となる。

不法行為法が根拠となる(2)において、Jは、検品業務の一部の判断を代替する前提でAIを開発・納入するに当たり、AIの診断対象となる製品の安全性を確保するという業務目的との関係で、誤検知の可能性を可能な限り低減する設計を行うとともに、AIの利用過程で人が関与すべき範囲など、リスクコントロールのため重要な事項についての説明を行っていたかどうか等重要となる。この注意義務との関係でも、AIガバナンスを適切に構築・実行していたか否かがその評価に影響し得るため、Jは必要に応じて後記52頁「＜参考＞誤検知に関してAI事業者ガイドライン上想定される措置」に例示されるガバナンスを構築・実行することで、注意義務違反の可能性を低減することが可能となる。

想定事例5は結果としてAIが誤検知を行った事案であるから、Jの責任に関して焦点となりやすいのはAIの検知精度に関する設計上の措置と考えられる。AIはその性質上、常に意図した挙動を保証し得るものではなく、検知精度には開発上の限界があり、なおかつAIはKの業務目的に必要な範囲で開発・設計されている。このことを踏まえると、前記(1)のとおり同種業務における通常人の作業水準が基準となりつつ、JとKとの間の契約内容も踏まえて精度を評価することが考えられ、AIがこれらに適合している場合には、第三者であるVとの関係でJの責任が生ずる可能性は低いと考えられる。

＜参考＞誤検知に関してAI事業者ガイドライン上想定される措置

異物の誤検知リスクを低減する観点で、AI事業者ガイドライン上想定される措置を想定事例5の当事者ごとに整理すると下表のとおりとなる。

AIの特性や個々の状況に応じて求められると考えられる措置は異なるが、前記4.1.2以下に述べたとおり、必要に応じてAI事業者ガイドライン所定の措置を講ずることは、過失と評価される可能性を低減することに繋がる。なお、前記2.1.3のとおり、各事業者における自主的な取り組みを促すというAI事業者ガイドラインの性質上、下記措置を講じなかったことが各当事者における注意義務違反に直ちに結びつくものではないことに留意が必要である。

			誤検知の防止のため想定される措置	ガイドライン上の関連箇所 ⁸⁸
検品受託事業者K	AI提供者・利用者	業務プロセスの適正な構築	● 安全性に配慮した適正利用 ・使用するモデルや閾値の管理 ・実運用条件での性能確認・継続学習 ・モデルの挙動変化や劣化有無の検証 ・AIの検査対象の適否や他検査併用の要否の検討等 ● 人による関与 ・人によるレビューを行う範囲の適切な設定 ・検査に習熟した担当者によるレビュー体制の構築等	U-2) i. 安全を考慮した適正利用 P-2) i. 人間の生命・身体・財産、精神及び環境に配慮したリスク対策等 P-2) i. 人間の生命・身体・財産、精神及び環境に配慮したリスク対策 P-2) ii. 適正利用に資する提供等

⁸⁸ AI事業者ガイドライン本編29-41頁参照。

			誤検知の防止のため想定される措置	ガイドライン上の関連箇所 ⁸⁸
AIサービス開発事業者J	AI開発者	リスクを低減するための運用	● 誤検知時のフィードバック体制 ・ 重要な誤検知に関する AI 開発者へのフィードバック ・ 原因分析や是正措置 等 ● 利用状況の記録及び透明性 ・ 検査画像・判定ログ・レビュー結果の保存 ・ 事故発生時におけるアカウントビリティの確保 等	U-6) i. 関連するステークホルダーへの情報提供 U-7) i. 関連するステークホルダーへの説明 等 U-7) i. 関連するステークホルダーへの説明 P-6) ii. 関連するステークホルダーへの情報提供 等
		設計上の措置	● 開発上可能な範囲での精度向上 ● 実運用条件に近い環境での精度や安定性の検証 ● 検出が難しい条件（サイズ/位置/材質等）の抽出 ● 人による判断を行うべき範囲の適切な設計 ● AI 利用者のフィードバックに基づく不具合の修正 等	U-2) i. 安全を考慮した適正利用 P-2) i. 人間の生命・身体・財産、精神及び環境に配慮したリスク対策 D-6) i. 検証可能性の確保 等
		説明上の措置	● 性能限界や残留リスクの明確な説明 ● 人による判断を行うべき場面や条件の明示 等	P-2) i. 人間の生命・身体・財産、精神及び環境に配慮したリスク対策 P-2) ii. 適正利用に資する提供 等

4.2 想定事例 6：自律走行ロボット（AMR）

AMR メーカーM⁸⁹は、倉庫や工場等の現場で稼働する台車型の自律走行ロボット（Autonomous Mobile Robot, “AMR”）を製造・販売している。当該 AMR は重量 300kg・可搬重量 500kg、最高速度 1.2m/秒（人が歩く程度の速度）で稼働し、従来の台車やフォークリフトに代わって倉庫や工場等の内部の荷物を迅速に運搬することができる。

物流業者 N は、M の AMR の性能や自社倉庫の環境を踏まえてリスクアセスメントを実施した上、人との協働を行う現場に AMR を複数台導入することを決定し、購入した。当該 AMR の導入により、パレット（荷役台）等の大型の荷物の運搬は AMR が担当する一方、人の作業員は細かな作業に集中することで、全体の作業効率の顕著な向上が見られたほか、人手を確保することが難しい深夜の時間帯においても、最小限の人員による監督の下、倉庫を一定の効率で稼働し続けることが可能となった。

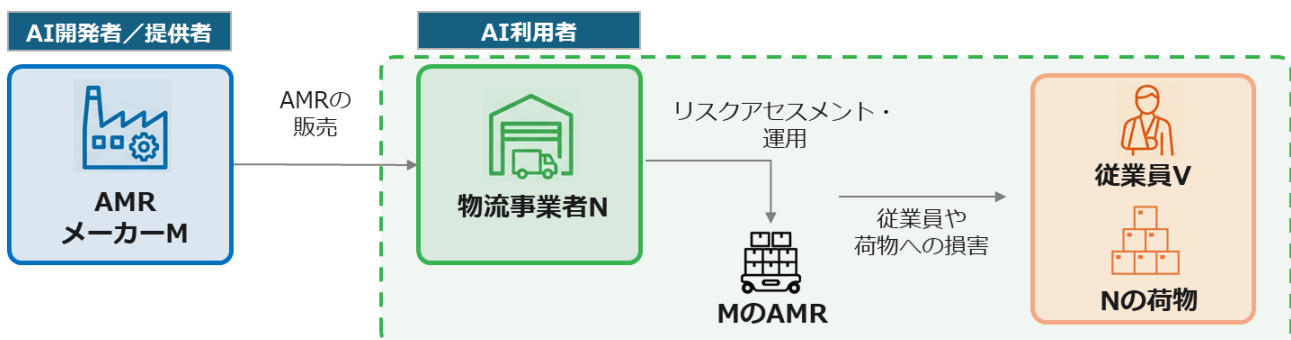
当該 AMR は長期間に亘り安全に稼働したが、あるとき、多量の荷物の運搬業務に従事している最中に以下の事故が発生した。

【事例 a】 AMR の前方にいた従業員 V に衝突し、負傷させた。事故発生の経緯を調査したところ、事例 a-1～事例 a-3 の各事情（後記 4.2.2 参照）が判明した。

⁸⁹ M は、自社開発した制御プログラムや画像認識 AI 等を組み込んだ AMR を N に対して販売しているため、「AI 開発者としての地位」と「AI 提供者としての地位」を有する（AI 事業者ガイドライン本編 5 頁、同別添 13 頁参照）。

【事例 b】 M が AMR を引き渡した時点ではソフトウェアにバグが存在しなかったが、運用過程におけるソフトウェア・アップデートによって、特定の条件下で停止距離を過小に計算してしまうバグが生じ、事例 a-1 と同様、本来の停止距離であれば回避できた衝突事故が発生した。

【事例 c】 AMR の機体には LLM を活用した自己診断 AI が搭載され、機体から収集したログやアラートを踏まえたトラブルシューティングを行い、タッチパネルを通じて診断結果や対処法を現場の作業員に伝える機能を有していた。あるときバッテリーが高温となり、上記 AI は AMR の機体が充電禁止の警報を発したことを検知したものの、その原因分析において「温度センサの一時的な異常」と判断し、充電禁止のセーフガードを無効化する対応策を提示した。作業員は当該診断に従ったところ、バッテリーの過度な温度上昇が進行し、直後に発火して倉庫内の N の荷物に延焼した。



AI を含むソフトウェアによって稼働する様々な機械やロボットの登場が予測される中、これらの機械の挙動によって事故が生じた際の責任関係については、裁判例の蓄積がなく予測可能性に乏しい。本研究会では、その一例として、工場等で普及が進む自律走行ロボット（AMR）についての検討を行った⁹⁰。このようなロボットは、人間の作業を代替するだけでなく、深夜時間帯の稼働など人手のみでは困難であった作業も行うというプラスアルファの効用が期待される点、また利用者が事業者であり一定の注意義務が求められる点で、自動運転車等とは特徴が異なり、これらの点が後記 4.2.2 のように各当事者の責任判断に影響することとなる。

想定事例 6 のように事業等においてロボットを利活用する場面では、一般に、ロボットを製造した製造業者（以下、総称して「メーカー」という。）と、ロボットを自らの事業のために運用する利用者（以下、総称して「ユーザー」という。）が存在するところ、それぞれがどのような責任を負うかが問題となる。

まず、事例 a 及び事例 b のように現場作業員の身体的損害が生じた場合、雇用主かつロボットのユーザーである N の責任（安全配慮義務違反ないし一般不法行為責任）やメーカーである

⁹⁰ 想定事例 6 の責任判断はあらゆるロボットに適用可能なものではなく、射程に留意すべきことについて後記 4.2.5 参照。

M の責任（主に製造物責任）が論点となる^{91, 92}。このうちユーザーの責任に関しては、AMR がどのような特性や安全性を備えていたか、メーカーがどのような設計や指示・警告を行っていたかが先決的な要素となり、これらを踏まえてユーザーが事故防止のためにどのような措置を講じるべきであったかが異なる。したがって、後記 4.2.2 以下では、まずメーカーの責任について検討し、その後にユーザーの責任について論じている。

また、事例 c のようにユーザーの財産に係る損害が生じている場合、メーカー M の責任（主に製造物責任）が論点となる。

4.2.1 AMR の安全性に関連する技術及びリスクアセスメント

AMR の製造や運用に関わる各当事者の責任を論ずる前提として、AMR の安全性に関わる技術や要求されるリスクアセスメントを以下に整理する。

なお、安全のために必要な機能は個々の製品や稼働現場において様々であるが、想定事例 6 では、下記(1)及び(2)の機能を備えた上で、(3)のリスクアセスメントを踏まえて導入された AMR を想定する。

(1) 安全系：人や物との衝突を防止するための機能

標準・規格としては、ISO 3691-4 やこれに対応する国内規格である JIS D 6802 等において、以下のような要求事項及び PLr（要求パフォーマンスレベル）が定められている。

- ✓ レーザースキャナ等の検知センサ
- ✓ 速度制御・監視機能
- ✓ 速度・進行方向・停止距離等に応じた保護フィールド（停止範囲）の設定
- ✓ その他の安全機能（ブレーキ系、非常停止装置、警報等）
- ✓ 保守的な走行条件を前提とした人検知試験

(2) 運用系：円滑な走行のための機能

また、円滑な走行を実現するための機能としては以下のようなものがあり、これらの機能も AMR の挙動やリスクに影響を及ぼす。

- ✓ SLAM（Simultaneous Localization and Mapping）：センサで取得した環境情報から壁や障害物などの地図を作成しつつ、AMR 自身の位置や向きを推定する。
- ✓ 3D カメラ及び AI 画像認識：人や物の区別が可能となり、人を認識した場合には予め減速する等、より柔軟な動作判断が可能となる。

⁹¹ 従来、例えばフォークリフトによる事故の場面では、運転手の一般不法行為責任を認定し、雇用主に対して使用者責任を追求するという解決が取られていた（東京地判平成 28 年 2 月 16 日（判例集未登載）、大阪地判平成 25 年 7 月 12 日（判例集未登載）、東京地判平成 18 年 4 月 7 日労判 918 号 42 頁等）。これに対し、自律走行ロボットを運用する場面では、ソフトウェアが自律的に行う運転判断について運転手の一般不法行為責任を問うことができないため、M や N がそれぞれどのような範囲で責任を負うかが問われる。

⁹² なお、負傷した従業員の治療費等の一定の損害は、労働者災害補償保険（労災保険）によって填補される。

(3) ユーザーによるリスクアセスメント

労働安全衛生法 28 条の 2 第 1 項において、「事業者は、厚生労働省令で定めるところにより、建設物、設備（中略）等による、又は作業行動その他業務に起因する危険性又は有害性等（中略）を調査し、その結果に基づいて、この法律又はこれに基づく命令の規定による措置を講ずるほか、労働者の危険又は健康障害を防止するため必要な措置を講ずるように努めなければならない。」と規定されており、AMR を運用するユーザーは、AMR の搭載機構及び周辺設備・運用方法等を含めたシステムについてリスクアセスメントを行うべき立場にある。

具体的には、JIS B 9700 (ISO 12100)や JIS B 9705-1 (ISO 13849-1)等を参照しつつ、概ね以下の手順により実施される⁹³。このリスクアセスメントは、事業場において労働安全衛生マネジメントシステム（OHSMS）の考え方が採用されている場合には、その構成要素の一つとして位置付けられるものである⁹⁴。

- ✓ 機械類の仕様・制限の決定（使用目的、速度、経路、積載等）
- ✓ 危険源の同定
- ✓ リスクの見積り
- ✓ リスクの評価
- ✓ リスク低減⁹⁵
- ✓ 検証・妥当性確認、残留リスクの周知
- ✓ 実施状況の記録、変更管理、定期的な見直し

4.2.2 事例 a について

(1) 事例 a における自律走行機能の位置づけ

前記 4.2.1 のとおり AMR は多様なソフトウェア及びハードウェアの統合によって構成されており、機械学習を用いた AI はその一部である。そこで、本節では自律走行機能の全体について、依拠／代替型 AI と同様の要件を満たしており業務上の判断を委ねるに足るシステム（以下、本節において「依拠／代替型システム」という。）に該当するか検討する。

AMR は自律走行機能を搭載しており、その性質上補助／支援型システムになじまないうえ、下記のとおり法的な評価としても依拠／代替型システムに該当し得ると考えられる。

⁹³ 中央労働災害防止協会「技術革新に対応した機械設備の安全対策の推進事業 報告書」（2023 年 10 月 31 日）
（<https://www.mhlw.go.jp/content/000830713.pdf>, 2025 年 11 月 25 日最終閲覧）11 頁以下も参照。

⁹⁴ なお、OHSMS の考え方は、必ずしも ISO 45001 等の認証取得を前提とするものではなく、認証の有無にかかわらず重要と考えられる。

⁹⁵ リスク低減の手法には、本質的対策（走行速度を下げる等）、工学的対策（付加的なセンサや非常停止ボタンの追加等）、管理的対策（走行ルール等の策定や作業員への教育等）、個人用保護具の使用等がある。株式会社リンクス「AMR 導入の失敗を防ぐ！ Part3「安全対策」を考える～法令解説から現場の運用まで～」（2025 年 8 月 20 日）
（https://linx.jp/product/irayple_amr/article/009/, 2025 年 11 月 25 日最終閲覧）参照。

AMR は稼働現場を自律的に走行することで、従来の台車やフォークリフトに代わり工場内の荷物を円滑かつ迅速に運搬することができる。AMR の導入による効用は個々の利用環境にもよるが、想定事例 6 のように、人と AMR の作業分担による作業効率の顕著な向上や、深夜の時間帯における倉庫の運営など、人だけでは困難であった業務の遂行に資すると考えられ、運搬作業を AI に委ねることの必要性が認められる。

さらに、自律走行機能の安全性に関し、倉庫等の作業場で機械等を用いて大量の荷物の運搬を行う際には、衝突事故等による作業員の生命や身体という重大な法益への危険が伴うため、同種作業を行う者には一般に高度の注意が求められる⁹⁶。このように同種作業に求められる安全性と同等以上の水準を有しているか否かを評価するにあたり、AMR は基本的に人の操作を前提としたロボットではないため、前記 2.2.2 に述べた基準に従ってその判断の品質や精度を人の作業水準と比較することは必ずしも容易でない。他方、一つの比較対象として、倉庫等での運搬作業に用いられてきた人が操縦するフォークリフトと比較すると、危険低減のため速度が制限されること、安全性の確保や向上のため前記 4.2.1(1)及び(2)のような設計が搭載されること等の特性を有している。加えて、ユーザーは労働安全衛生法上リスクアセスメントを実施することが求められ（前記 4.2.1(3)）、必要に応じて AMR の速度、利用環境、利用条件等に制限を加える対応策も含めて、リスクが許容可能な程度に低減されることが導入の前提となるため、リスクアセスメントを踏まえた利用環境等の制限の範囲内での走行に関しては、人が操縦するフォークリフトと比較しても同等以上の安全性を発揮し得ると考えられる⁹⁷。

以上より、AMR は依拠／代替型システムに該当し得る。その利用場面では、運搬業務における走行上の判断は AMR の自律的な挙動に委ねる前提で、メーカーは製造物責任法に基づき、上記の安全性を発揮するために必要な設計上・指示警告上の措置を講ずることが求められ、ユーザーは AMR を組み入れた業務プロセスの適正な構築及び運用に関する注意義務を負う。

この点に関し、AMR は現場の環境に応じて柔軟な走行上の判断及び行動を行うことから、どこまでの安全性が予定されているかを確定することが容易ではなく、下記(2)に述べるとおり、複数の視点を踏まえた総合的な評価を行う必要があると考えられる。

(2) AMR の衝突事故に関するメーカーの製造物責任の判断枠組み

① ソフトウェアにより自律的に稼働する製造物についての考え方

前記 2.1.2 のとおり、製造物責任法に基づく損害賠償請求においては、製造業者等が製造・加工・輸入等をした製造物に「欠陥」があることが要件となる（製造物責任法 3 条）。一般に、欠陥は以下の三類型に整理して説明される。

⁹⁶ 一例として、フォークリフトで荷物を運搬する際の事故に関する裁判例として前掲注 91 参照。

⁹⁷ 人が操作する機械と比較して安全性評価を行う具体的な手法として、下記(2)参照。なお、通常人の作業水準と一概に比較することが難しい場合には、前掲注 32 のとおり、付随的な考慮要素として、①業務効率化等の便益とリスク（権利侵害や損害が発生する蓋然性及びその重大性）との比較衡量や、②規制法上の要請を満たしていることを考慮することも考えられる。①の観点では、AMR の運用は作業員の身体法益等の重大な権利利益に関わるため、比較衡量基準の適用は謙抑的に解すべきであるが、ユーザーによるリスクアセスメントの結果等も踏まえ、AMR の導入によって増大し得るリスクや（発生した場合の）損害が僅少である場合、導入の便益を付随的に考慮することも許容され得ると考えられる。また、②の観点では、労働安全衛生法上のリスクアセスメントを実施し、必要に応じてリスクを許容可能な程度まで低減するための対策を講じることにより、求められる安全性を備えていることの一要素として斟酌し得ると考えられる。

製造上の欠陥 ⁹⁸	製造物が設計・仕様どおりに作られず安全性を欠く場合
設計上の欠陥	製造物の設計段階で十分に安全性に配慮しなかったために、製造物が安全性に欠ける結果となった場合
指示・警告上の欠陥	有用性ないし効用との関係で除去し得ない危険性が存在する製造物について、その危険性の発現による事故を消費者側で防止・回避するのに適切な情報を製造者が与えない場合

AMR 等のロボットについては、機体の不具合や設計不備（例えば発火の危険性等）が従来の製造物と同様に問題となるほか、ソフトウェアの挙動によって生じる事故についてどのように欠陥判断を行うべきかが論点となる。特に、設計上の欠陥に関し、AI 等のソフトウェアにより自律的に稼働するロボット等の製造物については、その性質上柔軟な挙動を行うことが予定されていることから、どのような挙動を行った場合に求められる設計を逸脱していると評価されるか、ひいては「通常有すべき安全性」を欠くと評価されるかが悩ましい。

この点、自律的に稼働する機械の一例である自動運転車に関しては主として以下のような考え方が存在しており、その他の機械やロボットの欠陥判断においても参考となり得る^{99, 100, 101}。

	概要・根拠	自動運転車以外の製品に適用する際の課題
(i) 総体的・統計的な安全性を基準とする考え方	● 人の判断や行動と比較し、自動運転車等の製造物の自律的な判断の方が確率的・統計的に安全か否かを基準とする考え方¹⁰² ● 事故発生の蓋然性は欠陥判断の主要な考慮要素の一つと考えられている¹⁰³	● 統計的な安全性をどのように比較するかが難しい場面も想定され、AMR のように普及途上にある製品については、データを取ることが困難なケースも少なくない

⁹⁸ ロボット等に搭載されるソフトウェアについては、完全にコピーされるため製造上の欠陥は考えにくいとの指摘がある（株式会社テクノバ「自動走行の民事上の責任及び社会受容性に関する研究報告書」12 頁（2020 年））。

⁹⁹ いずれの考え方においても、当面は人が操作した場合の安全性が比較対象となるものと考えられるが、技術の進展に伴い、より安全性の高い同種のシステムが比較対象となることも想定される（前掲注 31 参照）。

¹⁰⁰ それぞれの考え方における安全性の検討に当たっては、規制法や標準・規格等の安全基準に適合していることも有力な考慮要素となる。規制法への適合性と欠陥判断との関係性について、消費者庁「製造物責任(PL)法の逐条解説」（平成 30 年 9 月）（https://www.caa.go.jp/policies/policy/consumer_safety/other/product_liability_act_annotations, 2025 年 11 月 25 日最終閲覧）20 頁参照。規格への適合性に関しては、日本工業規格に適合しているチャイルドシートについて、特段の事情がない限り一応その拘束性において欠陥のない製品であると推測されると判示した事例（広島地三次支判平成 19 年 2 月 19 日（判例集未登載））があり、詳細な安全基準を定める規格は欠陥の判断に当たって重要な考慮要素の一つとなり得ると考えられる一方、JIS 規格は工業用品等の品質を標準化することを目的とするため、同規格に合致しなかったとしても欠陥とはならないとした事例（東京地判平成 19 年 11 月 27 日（判例集未登載））もあり、最終的には個別の事案ごとの検討を要する。

¹⁰¹ なお、製造物の引渡し時点における科学又は技術に関する知見によっては欠陥の存在を認識することができない場合、損害賠償責任は生じない（製造物責任法 4 条 1 号・開発危険の抗弁。「知見」とは、欠陥の有無を判断するに当たって影響を受け得る程度に確立されたすべての知識であり、また、特定の者の有するものではなく、客観的に社会に存在する知識の総体を指す（前掲注 100・消費者庁 38 頁））。従来の裁判例では開発危険の抗弁の適用を認めた事例は存在しないものの、不確実性が内在する科学技術である AI に関しては、今後、引渡し時点における科学又は技術に関する知見によって欠陥を認識できなかったとして開発危険と評価される事例が生じることはあり得る。

¹⁰² 戸嶋浩二ほか編著『自動運転・MaaS ビジネスの法務（第 2 版）』70-71 頁（中央経済社、2024 年）

¹⁰³ 前掲注 100・消費者庁 9 頁

(ii)個別状況における 安全性を基準とする 考え方	●事故が生じた個別の状況を踏まえ、発揮すべき安全性を備えていたかどうかを評価する考え方 ●自動運転車の文脈では、平均的運転者よりも安全な運転行動を求める考え方などが主張されている¹⁰⁴。 ●消費者の期待保護の観点からは個別状況における安全性を評価すべきとの価値判断があり得る	●AMR のように人の運転操作を前提としていない機械の場合、人の操作との比較による個別状況の安全性評価が困難となり得る
--	--	---

これらの基準は AMR に応用することも可能と考えられるが、以下のような困難もある。

まず、AMR の総体的・統計的な安全性に着目するアプローチ（上表(i)）においては、例えば従来の同種作業場全体における労災事故の統計情報と、M の製造に係る AMR の稼働時間当たりの事故率¹⁰⁵とを比較する手法が考えられる。また、総体的・統計的なアプローチを維持しつつも、より個別状況の安全性に着目した折衷的な考え方として、AMR を導入した現場（想定事例 6 においては N の倉庫）の同種作業における従来の労災事故発生状況と、M の製造に係る AMR の導入後の稼働時間当たり事故率とを比較する手法も考えられる。いずれの考え方においても、安全性を比較し評価するために有意な量のデータを収集することが前提となるところ、AMR は普及途上にあるため、本書作成時点では十分なデータを収集することは容易でなく、上記のような基準のみで「通常有すべき安全性」を評価し得るかは今後の課題と考えられる。なお、統計的な評価のみで「通常有すべき安全性」を備えているとは評価し難い場合であっても、事故発生の蓋然性は重要な考慮要素の一つであるため、統計的にみて安全性が高いことは欠陥を否定する一事情として斟酌されると考えられる。

また、個別状況における安全性を基準とするアプローチ（上記(ii)）に関しては、AMR は基本的に人の操作を前提としていないことから、同種業務に用いられてきたフォークリフトと比較し、実際に生じた事故の場面において想定される通常の運転手の操作と比較して安全性を比較することが考えられる。具体的には、人の運転手の判断及び操縦であれば当該事故を回避できた可能性が高いか、それともいずれにせよ事故を回避できなかった可能性が高いかを評価することが考えられる。もっとも、フォークリフトと AMR は構造が異なる機械である上、ユーザーにおいて講じるべき安全管理措置の範囲及び内容等の外的環境も異なり得るため、フォークリフトの動作との比較はあくまで「通常有すべき安全性」の判断における一つの目安と考えるべきであり、また、比較が困難なケースもあり得る。

以上の判断枠組みは、AMR の欠陥判断との関係でも重要な視点となるものの、各基準単独で「通常有すべき安全性」を満たしていたか否かを評価することは難しいと考えられる。したがっ

¹⁰⁴ 浦川道太郎「自動走行と民事責任」NBL1099 号 34 頁（2017 年）。なお、その具体的な判断の一例として、EDR（ブラック・ボックス）の記録などを利用し、当該事故時の危険状態の予見能力と回避能力において「平均的な自然人である運転者」と自動走行車の反応時間の優劣を比較する方法が示されているところ、「平均的な自然人である運転者」の反応時間は統計的な手法により把握されることも想定されるため、本表②の考え方にも本表①の統計的な観点は用いられ得ると考えられる。

¹⁰⁵ この事故率は保険料と密接に関連する（マーク A. ガイスフェルト（瀬川信久ほか訳）「人工知能は米国の不法行為をシステム化する」早稲田法学 100 巻 4 号（2025 年）206-207 頁）。

て、従来の工業機械等に適用されてきた伝統的な製造物責任の判断枠組みも適用しながら総合的に安全性を評価する必要がある。

② 従来の工業製品等に関する考え方

従来の工業製品等に関し、欠陥の判断においては、(i)製造物の特性、(ii)通常予想される使用形態、(iii)引渡し時期その他の事情が考慮されてきた（製造物責任法 2 条 2 項）。これらの要素¹⁰⁶を AMR に即して評価すると以下のとおりとなる。

指標	視点
製造物の表示 (上記(i)関係)	●衝突の危険性等は外部から認知しやすいが、安全性に影響する技術的な設計内容や予想外の挙動をしやすい状況等については外部から推知することが困難であり、リスクに関する明確な説明が求められやすい
製造物の効用・有用性 (上記(i)関係)	●機動的かつ迅速に運行することで台車等での作業に比べ大量の荷物の搬送等が可能となる ●センサ等を過度に冗長に設計し停止頻度が多くなると、円滑な運行という本来の目的自体が阻害される
代替設計の有無・価格対効果 (上記(ii)関係)	●異種／複数センサを併用する等、冗長設計を多重にすることで安全性を向上させるカスタマイズは可能であるが、求められる安全設計は利用環境によって様々であり、ユーザーのリスクアセスメントを踏まえて仕様が決定され、当該仕様を踏まえてコストが定まる
使用者による 損害発生防止の可能性 (上記(ii)関係)	●利用環境は個々の工場や倉庫の状況に応じて様々であり、ユーザーの管理下にあるため、ユーザーによるリスクアセスメントを踏まえてロボットの速度制御等がカスタマイズされており（前記 4.2.1）、レイアウトの調整や変更も行われる ●利用環境の設計や作業員への教育指導により、積極的に事故回避措置を行うことが予定されている
合理的に予想される 使用形態（上記(ii)関係）	●一定の危険行動や想定利用環境から逸脱した稼働は予想されるが、自律走行車両に関する教育を受けた者との協働が想定されており、作業員による合理的対応も期待しやすい
引渡し時における 技術水準（上記(iii)関係）	●本書作成時点では前記 4.2.1 のような技術が用いられることが一般的である ※なお、AMR のようなロボットにおいては、ソフトウェア・アップデートによって品質や安全性を向上させることが通常であり、この点を踏まえた論点について後記 4.2.3 参照

上表のうち特に重要な特徴として、AMR に求められる走行の円滑性や安全性は導入される個々の利用環境によって様々であるところ、現場の人的・物的設備を管理するユーザーのリスクアセスメントを踏まえ、必要な追加設計等が取捨選択されている点であり、この点は欠陥判断において下記③のように重視されるべきと考えられる。

③ 事例 a における欠陥判断の方向性

上記①及び②を踏まえると、欠陥判断は以下のような方向性で行うべきと考えられる。

¹⁰⁶ 前掲注 100・消費者庁 7 頁。考慮要素は他にも存在するが、AMR との関係で特に重要と考えられるものを抽出している。

(a) 設計上の欠陥¹⁰⁷

まず、上記①のとおり、AMR の統計的な事故率に関するデータは必要な安全性を満たしていたか否かの考慮要素となり得るほか、事故発生時において人が操縦するフォークリフトが用いられたと仮定し、人の判断や行動であれば当該事故を回避し得たか否かも一つの目安となる。

もっとも、これらの要素だけで欠陥を確定することは難しいと考えられるため、上記②の要素分析を踏まえ、実際の事案における利用環境及び AMR の具体的な設計（重量・速度・停止距離等）を踏まえ、当該事故を回避可能な合理的な代替設計が存在したか否かを基準とする。その際、以下の点を踏まえつつ代替設計の有無を検討する¹⁰⁸。

- 想定される利用環境における通常の利用過程でのリスクについては、代替設計が広く問われやすく、合理的に予見される誤使用や想定利用環境の逸脱についても一定のフェールセーフ等の設計上の配慮が求められる。予定された仕様や設計が重大なバグなどの不具合により発揮されなかったような場合、このような不具合がない代替設計は通常可能であることから、設計上の欠陥に該当する可能性が高い。
- 他方、利用環境や利用方法に起因する事故については、稼働現場がユーザーの管理下にあることや導入に際してリスクアセスメントが前提となることに鑑みると、設計上の欠陥は認められず、ユーザーにおける事故回避防止措置違反が認められやすい。
- また、ユーザーのリスクアセスメントを踏まえて合意・決定された設計は安全性判断に当たっても尊重され、これを大きく逸脱するコストを前提とした代替設計は合理的なものとは認められにくい。

(b) 指示・警告上の欠陥

ユーザーがリスクアセスメントにより危険源を発見し、安全性能を適切に発揮しつつ合理的な回避行動を取る前提として、メーカーからユーザーへの基本的な機能や性能限界等の明確な説明が求められる。

例えば、センサの検知可能範囲や停止距離等の情報は適切な事故回避行動のため重要と考えられる。また、AMR が人や壁の方向へ直進する際、設計による停止が間に合わない場合には非常停止ボタンによる外部からの停止を行う必要があり得るところ、このような非常停止が必要な場面や操作方法等も事故回避のため重要と考えられる。説明に際しては、正確性に加え、ユーザーにとっての理解のしやすさも重要な観点となり、技術的な詳細よりも、例えば「AMR の設計を前提としてどこに障害物が来ると停止するか」「どのような角度に障害物があると停止が難

¹⁰⁷ 自動運転車における欠陥の判断において、衝突被害軽減ブレーキといったシステムが搭載されていることに対する運転者の信頼により、このようなシステムが搭載されていない車両の運転者に比べると注意力が劣る状況になることを含めて「通常有すべき安全性」を判断すべき旨の指摘がなされている（窪田充見「自動運転と販売店・メーカーの責任」藤田友敬編『自動運転と法』174 頁（有斐閣、2018 年））。もっとも、AMR の運用場面では、ユーザーは事業者であり、事業場の安全性確保との関係でより慎重な判断を求められる立場にあることから、システムに対するユーザーの信頼が「通常有すべき安全性」判断に与える影響は限定的と考えられる。

¹⁰⁸ 標準・規格としては ISO 及び JIS において詳細な安全基準が定められており、一応の基準となるが、最終的には個別の事故を回避可能な代替設計が認められるか否かで評価される。

しくなるか」等、体感的に理解しやすいオペレーションや機能レベルで説明を行うことの方が重要となる場合も考えられる。

(3) M の製造物責任

事例 a では衝突の態様に様々なパターンが考えられ、それぞれのパターンにおいて M の製造物責任の判断が分かれる。

【事例 a-1】開発過程の不具合に起因する事故

V は AMR の右側方から歩行して走行経路に侵入した。M の AMR には JIS に準拠した保護停止機能が搭載されており、速度等に応じて停止距離を算定する設計であった。しかし、ソフトウェアが特定の条件下で停止距離を過小に計算してしまうバグがあったため、停止が遅れ衝突した。本来予定されていた停止距離であれば、当該衝突を回避することが可能であった。

事例 a-1 において、まず総体的・統計的な安全評価の面では、例えば AMR を導入した現場（N の倉庫）における労災事故の発生率が AMR 導入前よりも減少傾向にある等の事情があれば、欠陥を否定する方向に傾くことも考えられる。他方、人が操作する機械との比較として、仮に AMR でなく人が運転するフォークリフトが用いられていたと仮定すると、V が歩行して前方に侵入する程度の緩やかな動きは、運転手が目視をして衝突前に停止等の措置を行うことができたと評価される可能性があり、その場合、欠陥を基礎づける考慮要素の一つとなり得る。

その上で、事例 a-1 では、AMR の設計上も本来は当該事故を回避可能であったところ、ソフトウェアのバグにより本来の保護停止機能が発揮されず事故が生じている。当該機能は基本的な AMR の基本的な安全性を構成する要素であり、ユーザーとしても、当該機能が発揮されないことを見越して事故を回避するための措置を講ずることは一般に困難と考えられる。したがって¹⁰⁹、上記のような不具合のない代替設計が求められ、設計上の欠陥に当たる可能性が高い。

【事例 a-2】利用環境や利用方法に起因する事故

M の AMR は表面の汚損によりセンサ感度が低下するおそれがあることから、M は N に対し、当該リスクと共に定期的なメンテナンスの必要性やその方法を説明していた。しかし、N は使用過程でメンテナンスを怠っており、機体に汚れが付着していた。その結果、AMR の走行経路上で機体の存在に気づかず作業していた作業員の検知、警報及び制動が遅れ、衝突した。

事例 a-2 は、総体的・統計的な安全評価として、上記と同様に AMR の導入前後の事故率を比較すること等が考えられる。また、人が操作する機械との比較として、事例 a-2 はユーザーである N がメンテナンスを怠ったことにより AMR のセンサ感度が低下していた事例であるところ、このような事例について人が操作する機械と有意な比較を行うためには同一の条件を設定

¹⁰⁹ なお、ソフトウェアが複雑になればなるほどバグが存在しない状態を実現することは非現実的であり、全てのバグが「通常有すべき安全性」を欠くと評価されるものではなく、事例 a-1 のように重要な安全機能が発揮されないような場合が設計上の欠陥に該当する典型例と考えられる。

する必要があり、例えば、フロントガラスが汚損し視界不良であったフォークリフトが用いられた場合に上記事故の回避が可能であったか否かを評価すること等があり得る。もっとも、このような仮定の下での事故回避可能性を検証することは必ずしも容易でなく、事例 a-2 は人が操作する機械との比較が基本的に困難なケースと考えられる。

設計の適切性の観点では、M の説明内容に反する使用方法による事故を全て防止するための措置は求められないものの、上記は合理的に予見される誤使用と考えられ、一定のフェールセーフとしてセンサ感度の異常を検知・警告する仕組み等を備えていたかが問われる。このような機構を備えている場合、更に代替設計を行うべきであったとして欠陥に該当する可能性は低い¹¹⁰。

その上で、機体表面の汚れによるセンサ感度の低下リスクや必要なメンテナンス方法を説明している場合、N において上記事故の発生は合理的に予見することができ、かつメンテナンス等の対処も実施可能である。このようなユーザー側による事故回避のための情報が提供されていた場合、指示・警告上の欠陥にも該当しない。また、このような指示・警告が行われていたことは、ユーザーによる事故回避の可能性を高め、翻って設計上の欠陥を否定する事情としても作用すると考えられる。

【事例 a-3】設計上の限界と利用方法とが競合して生じる事故

M の AMR は、JIS に準拠した保護停止機能に加え、走行の効率化のための機能として 3D カメラ及び画像認識 AI を搭載し、側方から接近する作業員を検知して早期に警報を発しながら減速するように設計されていた。しかし、AI には検知限界があり、特異なシルエットについて検知精度が下がること等のリスクは M から N に説明されていた。

V は、作業場の臨時トラブルの対応のため巨大な脚立を肩に担いで横向きに持ちながら移動していたところ、AMR に搭載された AI が V のシルエットを人であると検知できず、警報や減速を行わなかった。その後、V が突如として AMR の停止距離内に進入したため、保護停止機能は作動したものの、停止が間に合わず衝突した。

事例 a-3 では、総体的・統計的な安全評価として、上記と同様に AMR の導入前後の事故率を比較すること等が考えられる。また、人が操作する機械との比較として、仮に AMR でなく人が運転するフォークリフトが用いられていたと仮定すると、V が突如として前方に侵入した場合、事前に衝突を回避することは一般に困難と評価される可能性があり、その場合、欠陥がないと評価する考慮要素の一つとなり得る。

設計の適切性の観点で、事例 a-3 における重要な要素は保護停止機能及び画像認識機能であるところ、それぞれ以下のように整理される。

■ 保護停止機能の動作について：

このような安全設計がソフトウェアのバグ等により適切に機能しなかった場合、上記 a-1 と同様に代替設計を講じるべきであったとして設計上の欠陥に該当する可能性がある。他方、

¹¹⁰ なお、複数のセンサやカメラを併用していた場合、事故が回避可能であった可能性も想定されるが、上記 4.2.3(2)④のとおり当初合意された利用環境や設計内容は尊重され、これを大きく逸脱するコストを前提とした設計は合理的な代替設計とは認められ難い。

上記 a-3 のように安全設計が機能していたものの V が突然接近したことで事故が生じた場合、欠陥を否定する方向に傾く。

■ 画像認識機能の動作について：

開発上合理的に可能な範囲で AI の精度を向上させること、実運用に近い環境下で精度や安定性を検証すること等、開発者として合理的に実施可能な性能向上措置を行っていたか否かが問われる。ただし、例外的なシチュエーションについて AI の精度を高めることは一般的に困難であること、また AI の誤検知に備えて上記の保護停止機能が備わっていることを踏まえれば、事例 a-3 のように特異なシルエットを検知できなかったことを以て設計上の欠陥に該当する可能性は低い¹¹¹。

その上で、ユーザーにおいて誤検知が生じやすい荷物の運搬を避けたり、臨時トラブルの対応時には AMR の稼働を休止したりする等の事故回避措置を講じることが可能であったと認められた場合、設計上の欠陥が否定される方向に傾く。

また、指示・警告の観点において、M は、走行上のリスクに関わる技術構成の概要（例えば、保護停止機能並びに 3D カメラ及び画像認識 AI を搭載していること等）について合理的な範囲で説明すべきと考えられる。加えて、ユーザーの理解度等に応じ、AI が誤検知を生じやすい状況や、誤検知の場合のリスク等についての説明が求められることも想定される。

(4) N 及び V の責任

① ユーザーである N の責任

N は V に対し、労働契約上の安全配慮義務を負う立場にあるところ、その内容は、危険防止のために適切な人的・物的設備を編成し、安全教育を施すことと解されている¹¹²。また、AMR の導入に当たってはリスクアセスメントを実施することが法令上求められている（労働安全衛生法 28 条の 2）。これらを総合すると、N は、導入時や運用過程におけるリスクアセスメントを通じ、リスクを低減するため必要と判断された措置や体制構築、従業員に対する指導教育等を含め、事故回避のための合理的対策を講じるべき義務を負うと考えられる。

上記の各事例との関係では、AMR の基本的な安全性能として予定されていた機能がソフトウェアのバグにより発揮されず事故が生じたケース（事例 a-1）では、そのような不具合を見越して安全対策を講じることが困難であり、事故発生との関係で安全配慮義務違反とは評価されないと考えられる。

これに対し、AMR の機体の適切なメンテナンスを怠っていたこと（事例 a-2）や、誤検知が生じやすい荷物の運搬を避けたり、臨時トラブルの対応時には AMR の稼働を休止したりする等

¹¹¹ なお、画像認識 AI によって常に望ましい検知を保証することは性質上困難であるため、意図どおり検知しなかったことを以てソフトウェアのバグとは評価し難い。

¹¹² 最判昭和 50 年 2 月 25 日民集 29 巻 2 号 143 頁、最判昭和 58 年 5 月 27 日民集 37 巻 4 号 477 頁、奥田昌道編『新版注釈民法(10)Ⅱ 債権(1)債権の目的・効力(2)』129-131 頁〔北川善太郎＝潮見佳男〕（有斐閣、2011 年）等参照。

の事故回避措置を怠っていたこと（事例 a-3）については、ユーザーである N の安全配慮義務違反と評価される可能性がある。

② 被害者 V の過失

M や N が損害賠償責任を負う場合であっても、事故との関係で V に過失が認められる場合には、過失相殺として損害賠償額が減額され得る（民法 722 条 2 項）。過失相殺における「過失」は法的な注意義務（予見可能性を前提とした結果回避義務）に限らず、事故の発生や損害の拡大に寄与したと評価できるような被害者側の不注意であれば足りると解されており¹¹³、例えば N が策定したマニュアルや手順書からの逸脱等があった場合には過失相殺の対象となり得る¹¹⁴。

4.2.3 事例 b について

事例 b の基本的事実関係は事例 a-1 と同様であるが、引渡し時点においては事故の原因となる不具合が存在しなかったところ、不具合が AMR の引渡し後のソフトウェア・アップデートで生じた点において異なる。このとき、一般不法行為責任（民法 709 条）だけが適用されるか、それとも製造物責任も適用され得るかが問題となる¹¹⁵。

V は M に対し、少なくとも一般不法行為責任（民法 709 条）に基づく請求を行い得る。V は M の過失（事故の予見可能性を前提に、M がどのような内容の結果回避義務を負っていたか）を主張立証する必要がある、主張立証に成功した場合には問題なく損害の填補を受けることができる。

他方、事例 b では論点を明確化するため事故原因を明示しているが、そもそも事故原因が明らかでない場合もある。製品に関する重要な資料は M や N に偏在しているところ、事故原因を V において究明・特定し、過失の主張立証を行うことが困難な状況も想定されることから、V は一般不法行為責任に加え、製造物責任に基づく請求を行うことも考えられる。製造物責任に基づく場合、V は過失でなく欠陥を主張立証することとなり、また欠陥の事実上の推定法理（後記 5.1.3）を活用できる可能性もある。

もっとも、製造物責任における欠陥判断の基準時は製造物の引渡し時¹¹⁶と解されているため、引渡し後のソフトウェア・アップデートにより生じた不具合を欠陥判断において考慮し得るかが問題となる。本研究会ではこの点に関し、二つの見解が示された。

一つは、ソフトウェア・アップデートによる不具合は原則として引渡し時点の欠陥とは評価できないものの、製品の引渡し時にソフトウェア・アップデートによる不具合の潜在的な原因（すなわち、ソフトウェア・アップデートをしなければならない必要性を基礎づける事情）があった

¹¹³ 大塚直編『新注釈民法（16）債権（9）』490 頁〔櫻見由美子〕（有斐閣、2022 年）

¹¹⁴ なお、事故の発生に寄与した V の行動が突飛なものである場合等は、AMR の欠陥や N の安全配慮義務違反等と権利侵害との間の因果関係が否定されることもあり得る。

¹¹⁵ なお、技術の進展に伴いアップデートを行うべき状況が生じることも想定され、その際に M が必要なアップデートを行わなかった場合、一般不法行為責任の有無が論点となると考えられる。

¹¹⁶ 前掲注 100・消費者庁 7 頁

場合には、引渡し時に欠陥が存在したと評価し得るとの考え方である。欠陥判断の基準時が引渡し時点であること、また、引渡しの意義は民法上の動産の引渡しの概念（民法 178 条、182 条、183 条及び 184 条）に従って一般に解釈されてきたこと¹¹⁷に鑑みると、欠陥の対象範囲には限定を加えざるを得ないことがその理由となる。当該見解によれば、事例 b でソフトウェア・アップデートにより生じた不具合が欠陥判断において考慮されるのは、AMR の引渡し時点で、ソフトウェア・アップデートによる対処を要する不具合があった場合に限られる。

もう一つは、引渡し時点でソフトウェア・アップデートが予定されている製造物に関しては、最終アップデート時点までの影響を欠陥判断において考慮し得るとの考え方である。走行の効率性や安全性の向上のためのソフトウェア・アップデートは引渡し時点において当然に想定されていること、メーカー自身が管理行為として行っていること、ロボットの挙動を支配する内部的な構造はソフトウェア・アップデートによって新たに引き渡されたと評価することができること等がその理由となる。当該見解によれば、事例 b でアップデートにより生じた不具合も欠陥判断において考慮されることとなる。民法上の動産の引渡しの概念との整合性が問題となるものの、製造物責任法制定時に想定された典型的な製造物とは異なり、品質や安全性をソフトウェア・アップデートによって向上させることが前提となるタイプの AI においては、「引渡し」の意義を柔軟に解する上記のような解釈を行う必要性は高まっていると考えられる¹¹⁸。

4.2.4 事例 c について

近時では、事例 c のように言語モデル等の AI をロボットに搭載してトラブルシューティングを行う事例が登場している。作業員がこのような診断機能の出力を踏まえて行動した結果として事故に至った場合、ユーザーである N はメーカーである M に対して損害賠償請求を行うことが考えられるが¹¹⁹、まず製造物責任が適用されるか、次いで欠陥判断をどのように行うかが問題となる。

(1) 製造物責任の適用可能性

前記 2.1.2 のとおり、製造物にソフトウェアが組み込まれている場合には当該ソフトウェアの挙動も欠陥判断において考慮され、事例 a 及び b においてはこのような場合に該当することが明確であった¹²⁰。他方、上記のトラブルシューティング機能のソフトウェアや UI の構成には複数の構成が考えられ、例えば推論を行う AI はサーバやホストコンピュータ等、AMR とは独立

¹¹⁷ 前掲注 100・消費者庁 18-19 頁

¹¹⁸ なお、本研究会では、ソフトウェア・アップデートを継続しながら利用に供することが前提とされていることを踏まえた解釈の一案として、ロボットの使用者の利用過程を通じ、リアルタイムで引渡しが続いているという捉え方についての指摘もあった。このような解釈によれば、前記 4.2.2 で示した統計的安全性を基準とする欠陥判断を行うことにより、ある局面における統計的安全性に問題があるにもかかわらずソフトウェア・アップデートをしなかった場合には、過失判断を介することなく「欠陥」が認められることとなる。反対に、製造業者がソフトウェア・アップデートを真摯に行なっていたものの、その当時の科学又は技術に関する知見によっては認識することができない要因によって統計的安全性に問題が生じていることが判明した場合には、開発危険の抗弁（前掲注 101 参照）を適用し得ることとなる。

¹¹⁹ 仮に、M とは別のメーカー O が AI を搭載したタッチパネルを製造し、M がこれを AMR に組み込んだ場合、当該タッチパネルについての O の製造物責任や、製造物責任法 4 条 2 項に基づく抗弁が論点となり得る。

¹²⁰ 前掲注 100・消費者庁 2 頁参照

の機器にインストールされている場合があり、どのような場合に製造物に対して組み込まれているといえるか評価が分かれ得る。

上記の自己診断 AI は、メーカーである M が AMR を安全に用いるための機能の一環として、AMR と一体のシステムとして提供するものである。当該 AI の出力は機体のタッチパネルを通じて表示することが製造物の性状として予定されており、診断機能の提供に不可欠なインターフェースが製造物に搭載されている。当該 AI と AMR のシステムとしての一体性及び機能的な結びつきを踏まえると、AI の出力において何らかの過誤があれば、AMR の構成要素の欠陥と評価することが可能であり、例えばインターネットや LAN 等を経由して推論自体を AMR の機体とは別の機器で行っていたとしても、製造物責任の適用は否定されないと考えられる。

(2) 欠陥の判断枠組み

① 事例 c における AI の位置づけ

上記の自己診断 AI は、深夜などの省人体制下でも倉庫の円滑な稼働が可能となるよう、AI の精度を十分に高めた上でユーザー及びその従業員が AI の出力に従う前提で用いる（依拠／代替型 AI）ことも考えられるほか、ユーザー及びその従業員の判断の補助ないし支援としてのみ用いる（補助／支援型 AI）ことも想定される。このとき、ユーザーが必ずしもロボットのメンテナンスの知見を有していないことを踏まえると、ユーザーが過度に AI の出力に依存することも考えられ、いずれの類型として用いるかは説明書等において明確に説明されることが重要と考えられる。

いずれに該当するかにより、以下のとおり判断枠組みが異なる。

② 依拠／代替型 AI に該当する場合

M は依拠／代替型 AI を提供するにあたり、望ましい精度を発揮するための設計上の措置を講ずるべきこととなる。求められる水準に関し、事例 a のような走行上の安全性は人の操作との比較のみに基づいて評価を行うことが困難であったが、事例 c におけるトラブルシューティングは従来も人の技術者等が対処してきた局面であることから、上記の状況に直面した通常の技術者の判断・行動と比較して AI の出力の安全性を評価することが考えられる。上記事例では、AI がバッテリーの温度上昇に関するアラートを踏まえた上で、セーフガードを無効化するなどの危険な行動を推奨していることから、通常の技術者と比較して安全性に欠ける出力を行っている」と解される可能性が高い。

また、上記のとおり N が必ずしもロボットのメンテナンスの知見を有していないことに鑑み、本事例のように、充電禁止のセーフガードの無効化のような危険な設定変更を従業員ができないようにするためのアクセス制御等を講じていない場合には、そのことが設計上の欠陥と評価される可能性もある。

③ 補助／支援型 AI に該当する場合

M は補助／支援型 AI を提供するにあたり、その出力に一定の誤りが含まれ得ることを前提として、機能や利用場面、利用方法等についての指示・警告を行うことが求められ、そのような説明を行っている限りは指示・警告上の欠陥が認められる可能性は低い。また、このような説明に

基づき、AI が誤った出力を行ったとしても N において AI の出力を検証・是正することが予定されるから、AI が上記のような誤出力を行ったとしても設計上の欠陥には該当しないと考えられる。

もっとも、上記②と同様、セーフガード無効化などの危険な設定変更を従業員ができないようにするためのアクセス制御等が講じられていない場合、そのことが設計上の欠陥と評価される可能性はある。

(3) 過失相殺

M の製造物責任が認められるとしても、事例 c では、N の作業員も AI の診断に従って危険な判断及び行動を行っている。したがって、危険に繋がり得るような AI の出力に漫然と従わないよう教育上・管理上の措置を講じなかった N の過失や、作業員自身の過失を踏まえ、過失相殺により賠償額が減額される可能性がある¹²¹。

4.2.5 想定事例 6 の議論の射程

念のため付言すると、以上に述べた製造物責任等の考え方は、AI を含むソフトウェアによって稼働する機械やロボット一般に及ぶものではない。欠陥判断においては個々のロボットの特性や想定される利用環境、ユーザーの性質等が重要な考慮要素となるところ、AMR 以外の機械やロボットにおいては、例えば下記のような要素において評価が異なり、それに伴って求められる安全性の水準や内容も異なり得る。

想定事例 6 の AMR	他のロボットとの比較
● 工場や倉庫等、閉鎖空間においてのみ稼働 ● ユーザーが法令上リスクアセスメントを行うべき立場にある ● ユーザーが稼働領域を全体的に管理することが可能 ● ユーザーの指揮監督に服する作業員との協業 ● 人間による操作を前提としていない ● 詳細な規則や安全基準がない	● より開かれた空間で稼働する場合 ⇒ 子供や高齢者等との相互作用や想定外の環境変化等も考慮し、より一般的な安全性が要求される可能性がある ● ユーザーによる適切な維持管理が必ずしも期待できない場合 ⇒ ユーザーの知識に配慮した設計や、理解可能性を考慮した指示・警告が必要となる可能性がある ● 従来人間が操作してきた機械と類似の機能・形状である場合 ⇒ 統計的な安全性や通常人の操作能力との比較がより容易な可能性がある

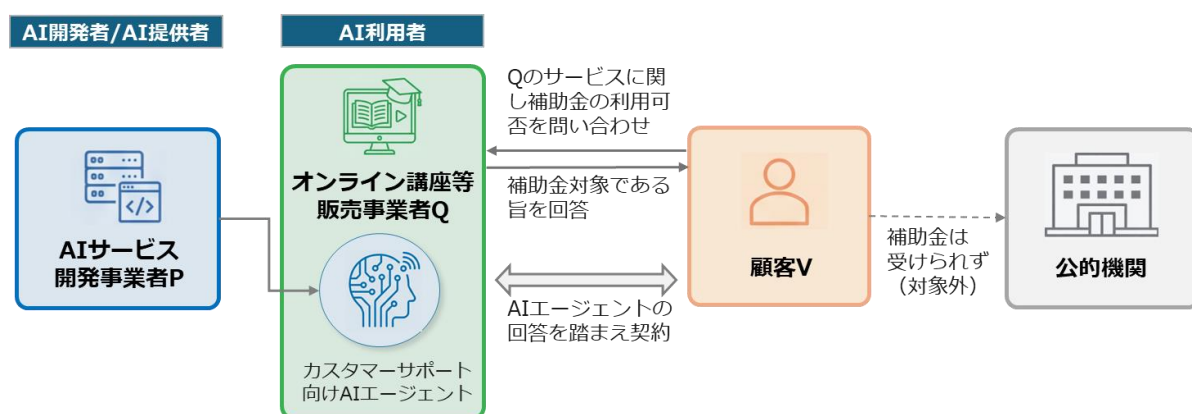
¹²¹ 作業員が AI の出力に依拠して危険な操作を行ったことについては、N の管理・監督上の過失を踏まえた過失相殺の可能性があり得るほか、作業員の過失を N との関係で過失相殺として考慮し得るかも論点となる。被用者の過失が加わって損害が発生・拡大した場合、一般に、被害者たる使用者の賠償額算定に当たり被用者の過失を考慮し得ることが認められている（吉村良一『不法行為法〔第 6 版〕』189 頁（有斐閣、2022 年）、作業員が AI の誤った出力を信頼することがやむを得ないケースも想定し得る。自動運転車に関し、被害者が自動運転装置による制動を期待して行動した場合の議論について、佐野誠「多当事者間の責任の負担の在り方」藤田友敬編『自動運転と法』211 頁以下（有斐閣、2018 年）参照。また、事故の発生に寄与した作業員の行動が突飛なものである場合等は、前掲注 114 と同様、AMR の欠陥と権利侵害との間の因果関係が否定されることもあり得る。

想定事例 6 の AMR	他のロボットとの比較
	●規制や安全基準が存在する場合 ⇒規制や安全基準への適合性が欠陥判断において重要な要素を占める可能性がある

4.3 想定事例 7：補論—AI エージェント

オンライン講座や学習教材を販売する事業を営む Q は、いわゆる AI エージェントを導入し、カスタマーサポートの大部分を自動化している。当該 AI は AI サービス開発事業者 P が開発し、Q に対してサービスとして提供しているものであり、顧客との通話（音声）やテキスト・画像を処理するマルチモーダル LLM を含む複数の AI モデルによって構成されている。また、顧客から質問がなされた場合、Q 内の回答マニュアルや製品情報、インターネット上の情報等にアクセスして必要な情報を収集し、出力に反映させる機能も有している。当該 AI の導入により、Q は従来よりも大量の問い合わせに対して迅速に対応することができるようになり、顧客が回答を受領するための待ち時間が大幅に削減されたほか、休日や深夜の質問に対しても、従来の自動応答に比べてより具体的な回答が得られる等の効果があった。

顧客 V は Q のカスタマーサポートに連絡し、Q が販売するオンライン講座が公的な受講代金の補助対象となるか質問したところ、上記 AI は、社内情報において「補助金の対象外」とされていたにもかかわらず、インターネット上で当該補助金が広汎に適用されている情報等を検索した結果を踏まえ、「補助の対象となるため購入を検討されたい」旨の回答及び勧誘を行った。V は当該講座の受講を申し込んだが、後から補助の対象とならないことが発覚した。



近時、個別のデータ入力や指示を前提とせず、一定の目的の実現に向けた複雑な業務プロセスを自動化・効率化することを目指し、AI エージェントないしエージェントティック AI（以下、単に「AI エージェント」と総称する。）といった高度なシステムの開発・実装が進みつつある。こうしたシステムについては、既の実運用に向けた取組みや高度化に向けた事業実証が進展しつつあり、従来の AI に比してリスクの拡大や新たなリスクの顕在化が指摘されている。もっとも、これらの AI システムに特有のリスクアセスメントやガバナンス手法については、現在も国際的に急速な検討が進められている段階にあり¹²²、個別技術やユースケースに強く依存するため、当事者の責任を個別具体的に整理することは現時点では困難である。

¹²² 日本においても、AI エージェントの動向を踏まえた AI 事業者ガイドラインの更新に関する検討の一環として、前掲注 22・総務省 4 頁以下において、AI エージェントのリスクや各当事者が必要に応じて行うべきガバナンスの内容等が検討されている。

そのため、本研究会では AI 利活用における責任についての基本的な考え方を整理するとともに、これらの考え方は、AI エージェント等の高度なシステムを検討する際にも基礎として位置付け得るものである。このことを示す一例として、想定事例 7 のような事例が想定される。

企業が販売する製品について事実と異なる説明を行った場合、取引に付随する説明義務違反として不法行為（民法 709 条）を構成する場合がある¹²³。

カスタマーサポートにおいて AI エージェントを用いる場合、そのような AI は法的にみて補助／支援型 AI にも依拠／代替型 AI にも該当する可能性があり、各当事者の責任判断にあたっては以下の点が重要になると考えられる。

(1) AI 利用者である Q の責任

① 依拠／代替型 AI に該当する場合

上記 AI が依拠／代替型 AI に該当するためには、前記 2.2.2 のとおり AI に判断を委ねる必要性が認められることが前提となる。上記事例のように、顧客の待ち時間の大幅な削減や休日・深夜帯の対応の柔軟性が高まることは、人による作業のみでは実現しがたい効果として、一部の業務判断を AI に委ねる必要性を基礎づけるものと考えられる。

そのうえで、AI の品質及び精度の面では、高リスクな相談や複雑な回答について人の判断を介在させる等の措置も含めた業務プロセス全体として、同種業務に従事する通常人と同等以上の水準を発揮していることが必要となる。このとき、企業から消費者に対する商品説明のうち、特に消費者の重要な利益に関わる内容については、消費者保護の観点から高度の注意義務が認められやすい（前掲注 123 の裁判例も参照）。上記事例においては、公的補助の適用有無という消費者における商品購入の重要な動機についても AI システムが正確に回答できるように設計されていたか（そのような設計がされていた場合、上記事例ではどのような特異な経過により誤回答が生じてしまったか）、AI システム単独で十分な回答精度の実現が難しい場合には人のオペレーター等に連携するなど業務システム全体としての精度の担保が行われていたか等が重要な論点となる。

上記の検討において依拠／代替型 AI に該当する場合、Q の注意義務は、AI システムを組み入れた業務プロセスの適正な構築及び運用に関する義務へと転換する。

AI エージェントを組み込んだ業務プロセスの構築及び運用を行うにあたり、AI エージェントにおいては、上記事例のように、AI システムとは独立の外部の情報（社内データベースや社外サイト）や外部ツール（社内の在庫 API 等）を用い、これらを自律的かつ連鎖的に選択・実行しながら目的達成のための処理を行う性質があり、その態様によって生じうるリスクや AI 利用者が講じうる措置の内容も異なり得る。前掲注 122 のとおり、今後 AI 事業者ガイドライン等において検討すべきガバナンスの内容が具体化されると見込まれるが、例えば以下のような事項の要否が論点となり得る。AI 事業者ガイドラインの内容も踏まえつつ、合理的に可能な措置を講じていれば、AI が結果として誤った出力を行ったとしても、過失が認められる可能性を低減することができると考えられる。

¹²³ 同種事案において説明義務違反を認めたものとして、大津地判平成 15 年 10 月 3 日（判例集未登載）参照。

- AI システムを組み入れた業務プロセスの構築に関する注意義務：望ましい精度や安全性を満たした AI システムを用いることや、必要な場面での人の関与を含めた業務プロセスを適切に構築していたか否かが論点になると考えられる。
- 当該業務プロセスの運用に関する注意義務：誤回答のフィードバック及びこれを踏まえたアーキテクチャの改善や、AI による回答であること（出力に誤りを含み得ること）の表示¹²⁴といった対応の要否が論点になると考えられる。

② 補助／支援型 AI に該当する場合

上記①とは異なり、AI の回答精度が十分でない等の理由により補助／支援型 AI に該当する場合、Q はあくまで判断の補助や支援としてのみ AI エージェントを用いるべきこととなり、その際、顧客への説明に関して消費者保護の観点から高度の注意義務が認められやすいことは上記のとおりである。商品購入の意思決定上重要な事項に関し、AI の誤った出力を顧客への回答に用いた場合、このような回答を用いた従業員の判断について注意義務違反が認められる（Q は使用者責任を負う）可能性や、AI の出力内容を検証する人員の体制が不十分であったことについて注意義務違反が認められる可能性等がある。

(2) AI 開発者・提供者である P の責任

上記(1)と同様、AI 開発者や提供者に求められるガバナンスは今後 AI 事業者ガイドライン等において具体化されると見込まれるが、前述のように外部情報や外部ツールの自律的かつ連鎖的な選択・実行といった特徴及びリスクがあることも踏まえつつ、適切な設計上の措置や明確な説明を行っていたかが論点となる。これらの義務の水準は、依拠／代替型 AI として提供するか、それとも補助／支援型 AI として提供するかに応じて異なる。

① 依拠／代替型 AI に該当する場合

依拠／代替型 AI に該当する場合、P は、人のオペレーターの判断や行動の一部を代替するにあたり予定された精度水準を発揮・維持するため、AI が外部の情報やツールを適切に扱うことを含めた必要な設計上の措置を行ったか、人の関与を介在させるべき範囲などリスクコントロールのため重要な事項を検討し、Q に対し説明を行ったか等が重要な論点になると考えられる。

② 補助／支援型 AI に該当する場合

他方、補助／支援型 AI に該当する場合には、最終的な出力の当否は Q において検証・是正を行う前提で、AI の機能や利用場面、利用方法、重要なリスク等に関する説明を行ったかが中心的な論点となり、その際、上記のような外部の情報やツールの利用によって重要なリスクが生ずる場合には、明確に説明することが求められ得る。また、AI エージェントが予想外の挙動を行

¹²⁴ 回答するのが人間ではなく AI であり、誤回答の可能性があることを事前に消費者に分かる形で説明していたことは、合理的に可能な措置を講じていたと判断するための重要な要素となる。

う傾向が見られる場合など、AI 利用者において適切に対処することが難しいリスクが生じた際には、これを緩和するための設計上の措置が求められる可能性がある。

第5章 立証や手続に関する論点

5.1 立証上の論点

前章まで、権利侵害や損害が生じた際の責任判断に関する実体法上の解釈適用の考え方を述べた。もっとも、実際の事案の解決に当たっては、AI の技術的特性を踏まえたリスクコントロールに関する専門的知見が要求される上、関連する資料を専ら AI 開発者・提供者・利用者らが保有していることが多いと考えられる。損害を被った者が原告として裁判上の権利行使を行う際、不法行為や製造物責任法に基づく請求原因については当該原告が主張立証責任を負うが、前記の専門技術性や証拠の偏在に照らし、特に過失（民法 709 条）や欠陥（製造物責任法 3 条）等の要件を主張立証するに際しての困難が見込まれる。

こうした主張立証上の困難に関しては、従来も証拠開示に関する制度が活用されてきたほか、特に医療訴訟や環境訴訟等の専門技術的な領域において、一定の条件の下、過失や欠陥の一応の推定（事実上の立証責任の転換）に関する各種法理が適用されてきた。本研究会では、実体法上の論点に付随し、これらの既存の制度及び法理に関する AI 事案への適用可能性や、今後の課題についても議論を行った。

5.1.1 文書提出命令

文書提出命令（民訴法 223 条 1 項）は、文書の所持者が文書提出義務（同法 220 条）を負う場合、裁判所が当該所持者に対し文書の提出を命ずる制度である。文書提出命令が発令されたにもかかわらず、訴訟当事者が命令に従わない場合や、相手方の使用を妨げる目的で当該文書を滅失させた場合、裁判所は、当該文書の記載に関する相手方の主張を真実と認めることができる（同法 224 条 1 項、2 項）ほか、挙証者が証明すべき事実を他の証拠によって証明することが著しく困難であるときは、裁判所は要証事実に関する相手方の主張を真実と認めることができる（同 3 項）¹²⁵。

上記のとおり、文書の所持者に文書提出義務が認められれば文書提出命令が発令されることから、AI 開発者・提供者・利用者がかかる提出義務を負うか否かが重要な論点となる。民事訴訟法上、文書の所持者は、同法 220 条 1 号～3 号までの個別提出義務に加え、同条 4 号の例外事由に該当しない限り一般的な文書提出義務を負う。

ここで、AI の開発・提供・利用過程において作成された文書には一定の営業秘密も含まれており、例外事由の一つである技術又は職業の秘密に基づく提出義務の免除（民訴法 197 条 1 項

¹²⁵ なお、訴訟外の第三者に対して文書提出命令が申し立てられた場合、通常の手続に加えて当該第三者に対する審尋（民訴法 223 条 2 項）が必要となるほか、命令に従わない場合の効果が真実擬制ではなく過料である（同法 225 条 1 項）等、手続及び効果における違いがある。

3号・220条4号八)が争点となることが予想される¹²⁶。営業秘密に該当するか否かは、所持者に与える不利益の内容・程度等と証拠としての当該文書の必要性等の諸事情を比較衡量して判断されることとなる¹²⁷。

本研究会においては、証拠を開示することによる不利益よりも金銭的解決を優先し、証拠を開示しない判断が多くなると、AIの安全性を高めるための情報が収集困難となるところ、営業秘密等とのバランスも考慮しつつ行政による規制・監督のための情報収集を行う仕組み等、民事訴訟以外の制度との連携も視野に入れて、社会全体としてAIの安全性を確保する仕組みの設計が重要であるとの指摘があった。

5.1.2 過失の事実上の推定

過失の認定に関しては、医療訴訟や環境訴訟等の領域を中心に、一定の条件の下、原告の主張立証責任の緩和を図るため過失の事実上の推定が活用されてきた。過失の事実上の推定には多様な類型が存在するが、例えば以下のようなものがある。

- **行為義務違反を前提とした過失の推定**：最判平成8年1月23日民集50巻1号1頁は、医師が医薬品を使用するに当たって添付文書に記載された使用上の注意事項に従わず、それによって医療事故が発生した場合、これに従わなかったことにつき特段の合理的理由がない限り当該医師の過失が推定されると判示した。
- **経験則を前提とした過失の選択的認定**：最判昭和32年5月10日民集11巻5号715頁は、医師の注射行為によって生じた化膿について、注射液の不良又は注射器の消毒不完全のいずれかの過失を選択的に認定した。また、最判昭和39年7月28日民集18巻6号1241頁は、医師の注射行為によって生じた化膿について、注射器具、施術者の手指あるいは患者の注射部位の消毒のいずれかが不完全であった過失を選択的に認定した。
- **不確実性を前提とした過失の包括的推定**：岐阜地判平成6年7月20日判タ861号49頁は、長良川における河口堰建設に対する差止訴訟において、堰の安全性については科学的、専門技術的知見に基づく合理的な判断がなされなければならないとしたうえで、本件堰の安全性に関する資料を被告側が保持していることを考慮すると、公平の観点から、まず被

¹²⁶ なお、ある文書が、その作成目的、記載内容、これを現在の所持者が所持するに至るまでの経緯、その他の事情から判断して、専ら内部の者の利用に供する目的で作成され、外部の者に開示することが予定されていない文書であって、文書の開示によって個人のプライバシーが侵害されたり個人ないし団体の自由な意思形成が阻害されたりするなど、所持者の側に看過し難い不利益が生ずるおそれがあると認められる場合には、自己利用文書（民訴法220条4号二）として提出義務が免除される場合もある。もっとも、AIの開発・提供・利用過程における文書との関係では、典型的には本文記載の技術又は職業の秘密に基づく提出義務の免除が争点となりやすいと考えられる。

¹²⁷ 最決平成20年11月25日民集62巻10号2507頁は、民訴法220条4号八及び197条1項3号に基づき文書の提出を拒絶することができるのは、対象文書に記載された職業の秘密が保護に値する秘密に当たる場合に限られ、当該情報が保護に値する秘密であるかどうかは、その情報の内容、性質、その情報が開示されることにより所持者に与える不利益の内容、程度等と、当該民事事件の内容、性質、当該民事事件の証拠として当該文書を必要とする程度等の諸事情を比較衡量して決すべきであると判示した。

告が、安全性に欠ける点がないことを相当の根拠および資料に基づき立証する必要があると判示した。

過失の推定は、原告が主張立証責任を負う原則の下、例外的に適用されるべき法理である上、事案ごとの個別具体的な検討を要するため、上記の裁判例の考え方がそのまま AI の文脈に適用し得るものではない。

もっとも、専門技術性や証拠の偏在といった特徴は AI に関する事案にも妥当する場合がある。こうした必要性に加えて、事実上の推定を正当化する根拠（生命身体等の重大な法益侵害、一定の行為義務違反、経験則、高度の不確実性等を総合的に考慮し、原告が主張立証責任を負う原則を修正することが許容される程の例外的事情）がみられる事案においては、上記の裁判例の考え方も参考にしつつ、事実上の推定を検討すべき場合があり得ると考えられる。

5.1.3 欠陥の事実上の推定

製造物責任に基づく損害賠償請求においても、裁判例上、製造物の「欠陥」に関して原告が負う主張立証責任の負担軽減を図る法理として、欠陥の事実上の推定が活用されてきた。

具体的には、損害を被った原告は、①製造物を通常の用法に従って使用していたにもかかわらず、②身体・財産に被害を及ぼす異常が発生したことをもって欠陥の主張立証として足りるという欠陥の事実上の推定が活用されている（仙台高判平成 22 年 4 月 22 日（判例集未登載）、東京高判平成 25 年 2 月 13 日判時 2208 号 46 頁）。

前記 4.2・想定事例 6 のように自律的に稼働する機械やロボットについても当該法理の適用は可能と考えられる。もっとも、人が使用することを想定した従来の製造物と比較し、人の判断や行動を介在せず自律的に稼働する製造物との関係では、どのような場合に上記①及び②の要件を満たすのかが悩ましい事案が想定される。

(1) 「通常の用法」の内容

機械が個々の挙動を自律的に決定し行動する場合であっても、その運用の側面を捉えて「通常の用法」に沿ったものであるか否かを評価することは可能と考えられる。例えば、製造業者が指定した利用環境や利用方法を重要な点において逸脱していない状態などが「通常の用法」に該当し得る。

(2) 身体・財産に被害を及ぼす「異常」の内容

他方、機械の挙動のバリエーションが増えるほど、意図した設計による挙動なのか、「異常」な挙動なのかの判断が困難となる可能性がある¹²⁸。例えば、機械やロボットが突如として合理的理由なく危険な走行を始めたような場合は「異常」に該当すると評価されやすい一方、様々な

¹²⁸ 自動運転車に関し、通常の用法に従って使用していたとしても自動運転装置がすべての事故を防止できるわけではないところ、自動運転装置によって回避できなくてはならない事故であったかが争点となり、この点の立証が必要になるとの指摘について、藤田友敬「自動運転をめぐる民事責任法制的将来像」藤田友敬編『自動運転と法』279 頁（有斐閣、2018 年）参照。

環境要因が影響して危険な挙動に至った可能性がある場合など、「異常」か否かが不明瞭な場合、意図した設計から逸脱する挙動であることを被害者において立証する必要性が生じることも考えられる。

本研究会では、自律的に稼働する機械やロボットを通常の用法に従って運用した場合であっても、実際にどのような挙動を見せるかは AI 等のソフトウェアの挙動次第であり、動作の異常性の判定が難しくなる場合もあることを踏まえると、技術的・科学的な複雑性に正面から対応する法理が必要となるとの指摘があった。

5.1.4 因果関係の認定

また、一般不法行為責任及び製造物責任のいずれにおいても、過失・欠陥と損害との間の因果関係が要件となるところ（前記 2.1.1 及び 2.1.2）、AI の出力による損害発生リスクが統計的にしか把握できない場合、因果関係の有無の判断が困難となることが想定される。

例えば想定事例 5 のような依拠／代替型の外観検査 AI の利用場面で、AI 利用者が用いるべき AI の検知精度が 90% であると仮定し、実際には精度 70% の AI を用いていたところ、AI が通常人であれば見落とし得ないような異物を見落とした事例が考えられる。このとき、あるべき水準の AI を用いていれば個別の原告の負傷という結果を防止し得たのか、あるいはいずれにせよ結果が生じていたのかが評価し難い可能性がある¹²⁹。

(1) 事実認定による対応

因果関係の立証は、一点の疑義も許されない自然科学的証明ではなく、経験則に照らした「高度の蓋然性」の証明が求められるとされている¹³⁰。統計的なリスクの大きさは、それ自体で高度の蓋然性の立証に至らない場合であっても、これを基礎づける事情の一つとはなり得るところ、他の間接事実と総合して高度の蓋然性を認定することが考えられる¹³¹（後掲注 132 の裁判例も参照）。

(2) 解釈論による対応

上記(1)の事実認定による解決が困難な場合、従来の裁判例では以下のような解釈論が採用されたことがあり、AI の文脈においても参考となる。

まず、水俣病に関する一部の裁判例では確率に応じた損害額の認定を採用し、因果関係の高度の蓋然性が認定できなかった一部の原告らとの関係で、原告らが被った損害に対し、水俣病に罹

¹²⁹ なお、米国での議論として、自動運転車の製造業者が必要な量の市販前テストを行わず、衝突確率が 10,000 分の 3 に等しい車両を販売したところ、オペレーティング・システムの設計に欠陥があったため、平均して 10,000 回の走行ごとに 1 回余分な衝突が発生するという事例において、オペレーティング・システムの設計に欠陥がなければどの事故が避けられたかを判断することは困難であるとの指摘がある（前掲注 105・マーク A. ガイスフェルト 204 頁）。

¹³⁰ 最判昭和 50 年 10 月 24 日民集 29 巻 9 号 1417 頁（ルンバール事件）参照。

¹³¹ 最判昭和 44 年 2 月 6 日民集 23 巻 2 号 195 頁は、レントゲン照射と皮膚癌との間の因果関係について、照射と癌（特に皮膚癌）の発生との間に統計上の相関があること、当該事件における照射の時期・量・回数および部位、被害者の皮膚癌が照射部分についてのみ発生したこと等の諸事実を踏まえ、因果関係を認めた。

患していた確率を乗ずることにより損害額を定めた¹³²。これらの判示は、公害事案であることや医学的な限界が前提となっており、射程は限定的に解すべきと考えられる上、損害が AI の出力に起因する可能性をどのように合理的に推計するかという課題もあるため、適用範囲は限定的と考えられる¹³³。

また、医療過誤事案では、最終的に生じた重篤な結果との関係で因果関係の高度の蓋然性が認められない事案において、相当程度の生存可能性や適切な治療を受けることへの期待権が侵害されたと捉え、慰謝料の賠償を認める考え方もある。もっとも、こうした考え方も生命身体等の重大な法益侵害を前提としたものであるほか、適切な代替法益の設定が可能な場合にのみ妥当すると考えられ、適用は慎重に検討すべきである。

5.2 国際的な紛争に関する手続上の論点

AI サービスの開発・提供・利用が国境を跨いで行われるケースは増加しており、これに伴い、AI の出力によって損害を被った者が海外の事業者に対する請求を検討する場合や、日本の AI 開発者等が海外居住者から請求を受ける場面等が想定される。このように国境を跨ぐ紛争について裁判上の解決を試みる場合¹³⁴、問題となる紛争についてどの国の裁判所に管轄があるか（国際裁判管轄）、どの国の法令を適用するか（準拠法）、とある国で得た判決を別の国で執行することができるか（外国判決の執行）が論点となる。

これらの論点の解釈適用の方向性については、経済産業省「電子商取引及び情報取引等に関する準則」（令和 7 年 2 月版、以下「準則」という。）が取りまとめており、その内容は AI の開発・提供・利用に伴う紛争にも援用することができると考えられるため、以下に概要を整理する¹³⁵。

例えば、日本に所在する原告が海外の AI 事業者に対する請求を検討する場合、各国の実体法や手続法を総合的に考慮し、いずれの国で本案についての請求や判決の執行手続を行うかを決定することとなる。

¹³² 東京地判平成 4 年 2 月 7 日判時平成 4 年 4 月 25 日号 3 頁は、水俣病の認定申請を棄却された者で、死亡後に病理学的に水俣病と診断された例は 71 例中 19 例（27%）と少なからず存在するところ、因果関係の証明につき高度の蓋然性を要求すると、医学の限界による負担を一方的に原告に課することになり、損害の公平な分担に反することから、水俣病であることの「相当程度の可能性」が認められる場合、その可能性の程度を損害賠償額（慰謝料）の算定に反映させるのがより妥当と考えられると判示した。また、大阪地判平成 6 年 7 月 11 日判時 1506 号 5 頁は、行政上の水俣病の認定基準に合致する原告らについては「高度の蓋然性」を認めつつ、認定基準に合致しない原告らについて、水俣病であることの可能性を症状の組合せごとに「40%・30%・20%・15%」の 4 段階に区分し、基準額にその割合を乗じるという按分方式を採用した。

¹³³ 本研究会でも、こうした考え方の射程が限定的であることについての指摘があった一方、AI の出力の統計的傾向以外に全く手がかりがなく、確率的にしかリスクを評価することができないような事案に限っては、確率に応じた損害額の認定を検討することも否定はされないとの指摘もあった。

¹³⁴ AI システムの設計・運用やその挙動が争点となる紛争においては、事実認定や責任の評価に高度な技術的知見を要する場合が少なくない。このような場合には、裁判による解決に加え、AI 技術に関する専門的知見を有する者が関与する仲裁や調停等の代替的紛争解決手続（ADR）を活用することが、紛争の実情に即した解決につながる場合もある。

¹³⁵ 以下では、本書の検討対象である不法行為に基づく請求を念頭に主な論点を整理しているが、個別の事案との関係ではより詳細な特則等が関係する場合もある。詳細については準則 336 頁以下を参照されたい。

日本の裁判所において外国に所在する事業者を相手方として手続を遂行する場合の考え方については、後記 5.2.1～5.2.3 のように整理できる¹³⁶。

5.2.1 国際裁判管轄

国際裁判管轄は、問題となる紛争に関して日本の裁判所が管轄権を有するかという問題であり、管轄権がない場合、日本の裁判所に訴えを提起したとしても原則として却下される。

まず、仲裁合意が存在する場合、その対象となる紛争については、日本の裁判所に訴えを提起しても被告の申立てにより却下される（仲裁法 14 条 1 項）¹³⁷。

仲裁合意がない、あるいは有効に成立していない場合には、日本の裁判所に訴えを提起した場合に、日本の裁判所に管轄権が認められる場合として民事訴訟法に列挙されている類型に該当するかが論点となる¹³⁸。

- **被告の住所等による管轄権**：被告が法人の場合、被告の主たる事務所又は営業所が日本国内にあるときは、日本の裁判所に管轄権が認められる（民訴法 3 条の 2 第 3 項）。
- **訴えの類型による管轄権**：訴えの類型による管轄権には様々なものがあるが、不法行為に基づく請求との関係で重要な類型としては以下が挙げられる¹³⁹。まず、日本に拠点を持たない海外の AI 事業者に関しては、「日本において事業を行う者」に対する管轄権が検討される（民訴法 3 条の 3 第 5 号）。日本で事業を行っているか否かの判断においては、当該事業者が日本語でサービスを提供していることが重要な考慮要素となる。また、不法行為に基づく請求との関係では、「不法行為があった地」としての管轄が検討される（民訴法 3 条の 3 第 8 号）。不法行為地には加害行為地だけでなく結果発生地を含み、権利侵害の結果が日本で発生している場合には同号に基づく管轄権が認められる¹⁴⁰。
- **合意による管轄権**：国際裁判管轄は当事者が合意でこれを定めることができることから（民訴法第 3 条の 7 第 1 項）、日本の裁判所を管轄裁判所とする合意がある場合は、日本の裁判所の国際裁判管轄が認められる。

5.2.2 準拠法

¹³⁶ なお、原告が消費者である場合、法の適用に関する通則法 11 条のような消費者保護規定により、外国裁判所・外国法を指定する条項の効力が制限される可能性がある。

¹³⁷ 準則 352 頁参照

¹³⁸ 準則 352 頁参照

¹³⁹ 準則 364-365 頁参照

¹⁴⁰ 外国で行われた加害行為の結果が日本国内で発生した場合において、日本国内におけるその結果の発生が通常予見することのできないものであったときを除く（民訴法 3 条の 3 第 8 号、準則 364 頁参照）。なお、不法行為について外国法によるべき場合において、当該外国法を適用すべき事実が日本法によれば不法とならないときは、当該外国法に基づく損害賠償その他の処分の請求を行うことができず、被害者は、日本法により認められる損害賠償その他の処分でなければ請求することができない（通則法 22 条、準則 364 頁参照）。

次に準拠法の問題として、当該紛争に関しどの国の法令を適用すべきかが論点となる。

日本の裁判所において審理され、準拠法を決定することとなった場合には、まず、当事者が取引の際にどの国の法に従うかにつき合意していた場合、法の適用に関する通則法（以下「通則法」という。）第 7 条に基づき、当事者が合意によって選択した国の法が当該契約に適用される¹⁴¹。これに対し、合意がない場合には、通則法のその他の定めに従い準拠法を検討することとなる。不法行為に基づく請求との関係で重要な類型としては以下が挙げられる。

- 不法行為法の準拠法に関する原則的規律は通則法 17 条であり、原則として「加害行為の結果が発生した地の法」が適用され、権利侵害の結果が日本で生じている場合には日本法が準拠法となる¹⁴²。
- また、名誉・信用毀損の場合には被害者の常居所地法（被害者が法人その他の社団又は財団である場合にはその主たる事業所の所在地法）が準拠法となる（通則法 19 条）。権利侵害の性質に応じ、通則法 19 条が適用されるか、あるいは通則法 17 条（又は通則法 20 条）が適用されるかについて見解が分かれるものの、日本に所在する原告が海外の AI 事業者に対して請求する場合には、被害者の常居所地法か、日本で結果が発生したものとして日本法が適用され则认为られ、結論としては日本法が適用される可能性が高い¹⁴³。

上記の規律に従い日本法が適用される場合、不法行為に基づく請求との関係では本書の考え方が妥当する。

5.2.3 外国判決・外国仲裁判断の承認・執行

最終的に請求を認容する判決等が得られた場合、当該判決の執行を検討することとなるが、請求が外国の裁判所や外国の仲裁手続によって行われた場合、その結果として得られた判断を日本において執行し得るかが論点となる¹⁴⁴。

- 外国裁判所が下した勝訴判決を日本で執行しようとする当事者は、日本の裁判所に対し、「執行判決を求める訴え」を提起する必要がある。その外国判決が確定判決であり、かつ、判決の内容と訴訟手続が日本の公序良俗に反しないこと等の要件を満たすと認められる場合、外国判決に基づく執行を行うことができる（民訴法 118 条）。
- また、外国の仲裁手続における判断についても、ニューヨーク条約ないし仲裁法に基づき日本で執行することが可能な場合があり、執行しようとする当事者は、「執行決定を求める申立て」を行う必要がある。仲裁合意の有効性や公序良俗違反の不存在等の要件が認められる場合、外国仲裁判断に基づく執行を行うことができる。

¹⁴¹ 準則 345 頁参照

¹⁴² 準則 360 頁参照

¹⁴³ 準則 365-366 頁参照

¹⁴⁴ 以下、準則 377 頁以下参照

A I 利活用における民事責任の解釈適用に関する手引き

[第 1.0 版]

令和 8 年 4 月

経済産業省 商務情報政策局 情報経済課

経済産業省 商務情報政策局 情報産業課 A I 産業戦略室

A I 利活用における民事責任の在り方に関する研究会

大塚 直（座長）

稲谷 龍彦

江間 有沙

柿沼 太一

穴戸 常寿

白石 友行

中原 太郎

橋本 佳幸

福岡 真之介

松尾 豊

