

第3回 AI 利活用における民事責任の在り方に関する研究会 議事要旨

＜日 時＞ 令和7年12月3日（水）15:00～17:00

＜場 所＞ みずほリサーチ&テクノロジーズ主催 オンライン会議（Microsoft Teams）

＜出席者＞ 大塚座長、稲谷構成員、江間構成員、柿沼構成員、宍戸構成員、白石構成員、
中原構成員、橋本構成員、福岡構成員、松尾構成員

＜オブザーバー＞ 関係省庁・関係団体

1) 第2回研究会の議論の内容について

資料3の3～4頁に基づき、事務局から前回研究会の討議概要と対応方針について説明が行われた。

2) 想定事例5について

資料3の5～19頁に基づき、事務局から想定事例5（取引審査AIの事例）について説明が行われた後、構成員による討議が行われた。

【本事例の位置づけや枠組みについて】

- 結論に異論はないものの、全体の構造がやや分かりにくい。従来は人間が判断主体であることを前提に、取引拒絶の判断の決定的な理由を特定した上でその合理性を判定してきた。これに対して、AIの判断過程はブラックボックスであり、取引を拒絶する理由の特定が困難であるため、統計的なバイアスや公平性の指標の乖離を違法性判断の根拠とせざるを得ない。判断を行ったAIの動作ではなくて、あくまで開発者によるコントロールや利用者によるチェックといったAIの外側の人間の行為に責任原因を求める構造であるという点を明確にしたほうがよい。
- 判断の支援にとどまる補助AIか、判断を全面的に委ねる代替AIかを明確にした方がよい。
- 事例内での具体的な損害等が明確ではないため、不利益の内容や事実関係の具体的な記述が必要である。
- 取引審査AIによる不利益な取扱いに対して個人が疑念を持っただけでは、社会問題として顕在化しにくい。より現実的に即した事例とするなら、内部告発やSNSでの告発など、問題が発覚し得る経路も考慮することが重要である。

- 資料 14 頁について、運用開始後に差別的な挙動がないかをチェックする事後モニタリングが実務的に重要であり、こうした措置が法的に評価されるインセンティブ設計が必要である。その点についても、記載いただけるとよい。
- AI による差別は統計的な傾向として顕在化するため、個々の出力に対して差別回避の注意義務を課すことは現実的ではない。したがって、単に注意義務を負わせるのではなく、開発者と利用者が協力してシステムを改善していくインセンティブ設計が重要である。
- 開発段階でデータセットの不備等によりシステムティックに差別が生じる場合と、運用を通じて事後的にバイアスが判明する場合を区別して、議論の解像度を高める必要がある。

【AI 利用者の責任について】

- 事務局資料が、AI の判断内容がブラックボックスであることを理由に利用者が過失責任を免れるという趣旨だとすれば、疑問がある。解釈論なのか立法論なのかはさておき、ブラックボックスな判定結果をそのまま利用するのであれば、その内部で起こっていることも含めて責任を引き受けるべきという評価も考えられる。
- 本事例における AI を補助 AI と位置付けるならば、議論は利用者の判断の適切性に収斂する。その場合、因果関係は「注意義務違反（過失）」と「権利侵害」の関係として整理されるため、AI のバイアスに起因する因果関係という特殊な問題を独自に論じる余地は大きくないと考えられる。

【因果関係について】

- 取引拒絶の事案は、原因特定が困難な場合があり、水俣病訴訟などで用いられた確率に応じた損害額の認定は難しいのではないかと。また、生命・身体とは被侵害法益が異なるため、相当程度の可能性論や期待権論をそのまま適用することにも疑問が残る。
- 単なる証拠の欠如とは別に、AI の性質に照らし因果関係が確率的にしか論じられないという事案が生じた場合には、確率的な挙動をするという対象の性質に着目して割合的に因果関係を評価することもあり得ると思われる。
- 実務的には、データセットの偏りなどから差別的な結果が生じたことは経験則上推認可能であり、事実上の推定が活用され、開発者側に反証が求められる可能性がある。
- システムティックに特定の属性が不利益を被るようにアルゴリズムが組まれている場合は、確率的な挙動であったとしても、因果関係が認められやすいと考えられる。意図的に差別した場合と結果的にそうなった場合は分けて考えるべきである。
- 代替法益としての「相当程度の可能性」論は、通常は生命・身体的な法益に関連して論じられるため、本事例のような契約締結の可能性という利益に適用することには疑問

が残る。期待権論についても、契約交渉の不当破棄に関する裁判例を参照・検討するほうが適切ではないか。

【代理変数と差別について】

- 差別があればすべて注意義務違反になるということではなく、社会的相当性や合理性がない差別によって取引を拒絶することが問題であるということを明記したほうがよい。
- 資料 13 頁の「代理変数」の表現は、何が何の代理変数なのか分かりにくい。直接的な差別を意図していなくても、収入や勤務形態といった変数が性別などのセンシティブ属性と社会構造的に結びつき、結果として間接差別につながる可能性があるため、誤解を招かないよう慎重な記述が求められる。
- 私人間における別異取扱いが直ちに違法となるわけではないが、社会的許容限度を超える場合には問題となり得る。ただし、公法上の規律が存在する業種・業態においては、その点を考慮して、私法上の不法行為の成否等が判断されることになる。
- 本事例のように、多数の不動産事業者が AI 開発者による同一のプラットフォームを利用している場合、特定の個人が多く不動産事業者から差別・排除されるリスクが生じる。この場合、被害者が AI 開発者を直接訴えることは困難であるため、公法的な規律を含めた解決策の検討が必要である。
- AI 活用の際の人間の関与度合いとイノベーションへの配慮について、業務や利用形態ごとにリスクの程度は異なるため、一律に対策を求めるとリスクとイノベーションのバランスを損なう恐れがあり、事例ごとに慎重に検討する必要がある。

【資料中の表現について】

- 資料の記載において、「権利侵害」の有無と「過失」の有無が混在して論じられており、報告書の取りまとめに際して留意されたい。
- 資料 7 頁の「属人的バイアス」については、「認知バイアス」の趣旨で用いているのであれば、その旨を明確にし、「統計的バイアス」と対比させるべきである。
- 資料 11 頁の「ルールベースの審査を行っていない」という記述は、全ての AI に当てはまるわけではないため、「機械学習している AI については」といった限定的な表現とするほうが的確である。

3) 想定事例 6 について

資料 3 の 20～37 頁に基づき、事務局から、想定事例 6（自律走行ロボットの事例）について説明が行われた後、構成員による討議が行われた。

【事例の位置づけについて】

- 本事例は、人間の作業を代替するだけでなく、人間にはできなかった作業も行うという「プラスアルファの効用」が期待される機械である点、また利用者が事業者であり一定の注意義務を要求し得るといった点で、自動運転車とは特徴が異なる。このように、製品の特性（プラスアルファの効用の有無）や利用者の属性（事業者か消費者か）によって議論の射程が異なることを、報告書の冒頭で明確に示すべきである。
- 利用者が事業者である本事例では、自動運転車で議論されるような、ユーザーの過度な期待や依存傾向を前提とした欠陥判断は、基本的に当てはまらないか、当てはまる度合いが小さいことを注記すべきである。

【安全性の判断基準について】

- 資料 31 頁の整理等は実質的に人間ではなく AI と比較しているような記述となっている。したがって、判断基準の整理としては、「同種の AI」との比較は困難であるとの批判があることは踏まえつつも、「人の能力」だけでなく、「別の AI システムの安全性」を基準とする考え方も明確に位置づけて整理する必要がある。
- 資料 26 頁の整理について、自動運転の文脈では、ルールベースを前提とした個別状況の安全性を基準とする考え方だけでなく、機械学習の進展に伴い、走行距離あたりの事故率といった統計的な考え方も有力になっている。

【欠陥の類型と責任について】

- 資料 27 頁の欠陥類型の整理について、でソフトウェアのバグが「製造上の欠陥」に分類されている点に違和感がある。AI の検知失敗が「設計上の欠陥」とされていることと合わせ、欠陥の分類が一貫していないように見受けられる。
- 自動運転車に比べ、指示・警告上の欠陥に関する記述が少ない。利用者が事業者であっても、指示・警告が複雑で理解困難な場合は、システムの複雑性自体が設計上の欠陥と評価される可能性があり、その点を盛り込むべきである。
- 資料 30 頁について、本事例は自走式だが、もし工場等の床に固定されたアーム型ロボットであれば、土地工作物責任（民法 717 条）の適用対象となり得るため、立法論としては瑕疵を要件とする無過失責任を課すことも考えられる。
- 資料 31 頁の「予測困難な挙動」という表現について、AI の出力が原理的に予測できないことは自明であるため、この言葉だけでは概念として広すぎる懸念がある。AI の出力が予想できないという一般的な性質と、それを超える高度な予測困難性とを区別して記述すべきである。
- 資料 31 頁で開発危険の抗弁の可能性に言及しているが、現行法上、製造物責任法における開発危険の抗弁の適用は極めて限定的である。また、イレッサ訴訟最高裁判決で示

された予見可能性の考慮も、製造物一般に直ちに適用できるものではないため、こうした考え方もあり得るといった程度の位置づけに留めるべきである。

- AI は確率的に挙動するため、長期的・統計的に見て安全性が担保されているのであれば、たまたま人間と異なる挙動で事故を起こしたとしても、直ちに「欠陥」と認定すべきではない。稀な挙動の全てに責任を負わせると開発の萎縮を招くため、個別の事故回避能力ではなく、リスクを生じる具体的場面に即して、統計的な安全性を重視する運用上の工夫が必要である。

【自己診断 AI の製造物性について】

- 資料 34 頁について、ロボット本体の画面に表示される形式であれば、物理的な製造物の枠内に収まるため、誤った指示が表示された場合は、動作上の欠陥、あるいは指示・警告上の欠陥として製造物責任を問うことができる。また、独立したホスト PC 等で診断結果が表示される場合でも、製造業者が AI を利用してユーザーに指示を与えている以上、指示内容が不合理であれば、その責任は免れないと考えられる。
- 自己診断機能を提供する LLM が物理的に離れたサーバー上にある場合でも、ロボット本体がその機能の利用を予定しているのであれば、利用者は機能全体の安全性を期待するため、当該機能も製造物の一部と捉えることで差し支えない。

【アップデートと欠陥判断基準時について】

- 製品の引渡し後にアップデートが行われることが前提となっている製品においては、メーカーが管理権限を持ちアップデートを行っている以上、実質的には製品の内部に対する支配が継続していると捉えられる。したがって、リコールで回収して基盤を交換する場合と同様に、内部的な引渡しは完了していないと構成し、アップデート時点での欠陥を問うことが可能ではないか。
- 資料 35 頁について、現行法は「引渡し時」を欠陥判断の基準としており、単に「アップデートが予定されている」という一般的な可能性のみをもって引渡し時の欠陥とするのは解釈上困難である。欠陥が肯定されるのは、アップデートによる不具合の潜在的な原因が引渡し時に既に存在していた場合に限られると考えられる。
- 引渡し後にバグが発見されたにもかかわらず、メーカーがアップデートを提供しなかったことによる不具合（不作為）も検討に値する。現行法での対応は困難な面もあるが、製品の性質上、本来カバーすべき責任の空白が生じていると考えられる。
- 「予測困難な挙動」と開発危険の抗弁の関係について、アップデートを行わなかったことによる責任について、被害者に過失の立証を求めることは困難であるため、開発危険の抗弁をインセンティブ設計として活用できるように、「引き渡した時」の解釈を工夫すべきである。すなわち、事業者が組織的なモニタリングとアップデートを行っていた場合に限り、当該製品を利用に供した当時の知見では回避不能だったとして抗弁を認

める運用とすれば、開発者は免責を得るために積極的に製品改善を行う動機付けが働くことになる。

- 資料 35 頁について、以前「潜在的なバグがあれば引渡し時の欠陥と見てよい」と述べたが、それは現行法の「引渡し時」という要件に引きつけて考えざるを得ないためである。本来的な形としては、「引渡し時」という言葉の意味を AI との関連で明確にするための解釈が必要である。

【立証責任（事実上の推定）について】

- 資料 36 頁について、従来の「欠陥の事実上の推定」に関する判例法理は、人間が対象物をコントロールしていることを前提に、通常の使用で異常が発生したから欠陥があるはずとするものである。しかし、AI 搭載機器は自律的に作動し、人間によるコントロールが後退するため、「動作の異常性」の判定自体が困難となる。従来の法理は事故の科学的な機序の立証負担を軽減する意義があったが、AI の文脈でその機能が発揮されるかは疑問である。本来であれば、技術的・科学的な複雑性に正面から対応する法理を含めた検討が必要である。

5) 全体を通して（座長総括）

【想定事例 5 について】

- 従来の事例とは異なり、判断構造そのものまで踏み込んで検討する必要がある点や、「代理変数」等の記述に関して注意が必要であるとの指摘に留意したい。
- この事例は判断補助 AI であるとの意見があったが、そのように理解できる。
- 因果関係について、確率的な枠組みで考える必要性はあり得るものの、実務的には事実上の推定で対応可能なケースも多いとの指摘もあった。他方で、生命・身体侵害の事案とは異なるため、「相当程度の可能性」の法理をそのまま適用することは難しく、確率的に考える場合の理論構成については、引き続き検討が必要である。
- 利用者による事後モニタリングが極めて重要であり、開発者と利用者が共同して問題を解決していく方向性が重要であるとの意見は重要である。
- また、業種・業態ごとの公法的な規律の存在も重要な考慮要素となるほか、AI 利用における人間の関与度合いやリスクも業種や業務ごとに異なるため、報告書の作成においては、これらの点に配慮が必要との指摘もあった。

【想定事例 6 について】

- 利用者が消費者ではなく「事業者」である点は重要なポイントであり、指示・警告上の欠陥についても記述を充実させる必要がある。
- 安全性判断において、自動運転等の分野では統計的な視点が有力になりつつある点や、

比較対象として人間だけでなく他の AI システムを基準とする考え方がある点についても、報告書に反映したい。また、固定型ロボット等との関係で、土地工作物責任（民法 717 条）の適用の可能性についても指摘があった。

- アップデートに伴う欠陥については、意見が分かれたものの、メーカーがアップデートを行った以上は製造物責任を負うという考え方に賛同する。もっとも、現行法解釈との整合性について慎重な意見もあるため、引き続き検討が必要である。
- アップデートしなかったことによる不具合については、従来は過失責任の範疇と考えられてきた。なお、欠陥として構成する可能性や、開発危険の抗弁との関係について、引き続き検討したい。

<お問い合わせ先>

商務情報政策局 情報経済課

電話：03-3501-1511（内線：3961～3963）