

公募報告4（現地発表）

報告者：松田 遼（弁護士 一橋大学大学院ビジネスロー専攻修了（経営法））

タイトル：「AI時代におけるウェブスクレイピングの法的課題と秘密計算による解決可能性の考察：個人情報保護法および著作権法を中心に」

はじめに

現代社会は、人工知能（AI）技術の急速な発展に伴い、データが経済成長と技術革新の主要な原動力となる「データ駆動型社会」へと変貌している。インターネット上に公開された膨大な情報を収集する「ウェブスクレイピング」は、AI学習の貴重な資源となる一方、個人情報保護法、著作権法、利用規約との間で複雑な法的・倫理的課題を提起している。特に、個人のプライバシーに深く関わる要配慮個人情報といった機微なデータの取得、およびAI学習における著作物の利用は、データ活用を推進しようとする事業者と、個人の権利を保護しようとする社会との間で深刻な緊張関係を生み出している。

本稿では、まずウェブスクレイピングが直面する法的課題、特に個人情報保護法における要配慮個人情報の取得における課題と、著作権法におけるAI学習での著作物利用の適法性を整理する。日本の法制度における現状の解釈と、AI時代におけるその限界を掘り下げる。その上で、これらの法的課題に対する革新的な解決策として「秘密計算（Secure Multi-Party Computation: SMPC）」について取り上げ、その可能性と限界を検討する。秘密計算がウェブスクレイピングの法的リスクをいかに軽減し、データ活用と権利保護の新たなバランスをいかに実現しうるのか、そのメカニズム、技術的課題、そして国内外の動向を考察する。

最終的に、秘密計算がデジタル社会におけるデータ活用の「新たな地平」を切り開く可能性を評価し、それに伴う技術的、法的、倫理的な課題を提示する。本稿は、単なる技術論に留まらず、AI時代におけるデータ利用の適正化に向けた法制度設計と技術導入のあり方を考察するための基礎を提供する。

目次

1. はじめに：AI時代におけるデータ利活用と法的・倫理的課題
2. ウェブスクレイピングが直面する法的課題
3. 秘密計算による「交通整理」の可能性と限界
4. 結論：AI時代におけるデータ活用の新たな地平

1. はじめに：AI時代におけるデータ利活用と法的・倫理的課題

21世紀に入り、情報通信技術の飛躍的な進歩は、社会のあり方を根本的に変容させてきた。特に、イ

ALIS

インターネットの普及とスマートフォンの浸透は、個人の日常生活から企業活動、政府の政策決定に至るまで、あらゆる側面で膨大な量のデータが生成される「データ駆動型社会」を到来させた。このデータは、単なる記録ではなく、新たな価値創造の源泉、すなわち「21世紀の石油」と称されるまでにその重要性を高めている。

このデータ駆動型社会において、最も革新的な技術の一つが人工知能（AI）である。AI、特に機械学習や深層学習の発展は、画像認識、自然言語処理、音声認識といった分野で人間の能力を凌駕するパフォーマンスを発揮し、医療診断、自動運転、金融取引、顧客サービス、教育など、社会のあらゆる側面に革新をもたらしつつある。AIの進化は、大量かつ多様なデータの存在なくしては語れない。AIモデルは、データからパターンを学習し、予測や意思決定を行うため、高品質で膨大な学習データが不可欠である。インターネット上に公開されたウェブサイト、SNS、ブログ、オンラインフォーラムなどに存在するテキスト、画像、動画などの情報は、このAI学習の貴重な資源となる。これらの情報を効率的に収集するための技術として、「ウェブスクレイピング」が広く用いられている。ウェブスクレイピングとは、自動化されたプログラム（ボットやクローラー）を用いて、ウェブサイトから情報を抽出し、構造化されたデータとして収集する技術である。市場調査、競合分析、価格比較サイトの構築、ニュースアグリゲーション、学術研究など、その応用範囲は多岐にわたり、現代のデジタル経済を支える基盤技術の一つとなっている。

しかし、その効率性と有用性の一方で、ウェブスクレイピングは深刻な法的・倫理的課題を提起している。特に、日本の法制度においては、ウェブスクレイピングを直接的に規制する法律は存在しない。このため、既存の法律、すなわち「個人情報保護法」と「著作権法」、さらにはウェブサイト運営者が定める「利用規約」との間で複雑な摩擦を生じさせているのが現状である。この摩擦は、データ活用を推進しようとする事業者と、個人のプライバシーや著作権といった権利を保護しようとする社会との間の緊張関係を浮き彫りにしている。これらのリスクは、ウェブスクレイピングの合法性、倫理性を巡る議論を複雑にし、データ活用の促進と権利保護のバランスをいかに取るかという、現代社会における喫緊の課題を提起しているといえる [1]。

本稿の目的は、AI時代におけるウェブスクレイピングの法的・倫理的課題を掘り下げ、特に個人情報保護法と著作権法との関係、そしてそれらの課題に対する革新的な解決策としての「秘密計算」の可能性と限界を検討することにある。本稿は、以下の構成で議論を進める。

第2では、ウェブスクレイピングが直面する具体的な法的課題を詳述する。特に、日本の個人情報保護法における要配慮個人情報の取得に関する「原則同意主義」と、ウェブスクレイピングの無作為的・大量収集という性質との衝突を分析する。また、著作権法におけるAI学習での著作物利用の適法性を、著作権法第47条の5の解釈を軸に検討し、生成AIがもたらす新たな課題を指摘する。

次に、第3では、これらの法的課題に対する革新的な解決策として注目される「秘密計算」の可能性と限界を検討する。秘密計算のメカニズムと特徴を解説し、それが個人情報保護法における同意不要化

の議論にどのように貢献しうるか、また著作権者の懸念をいかに和らげるかについて考察する。同時に、秘密計算が万能な解決策ではないこと、特に統計結果がもたらす社会的な影響や情報自己決定権の保障といった広範な問題には直接対処できないという限界も明確にする。さらに、秘密計算を用いたウェブデータ連携の具体的なイメージを提示し、その導入と普及における技術的・経済的・運用的課題についても議論する。第4は結語である。

本稿を通じて、AI時代におけるデータ活用の「新たな地平」を切り開き、プライバシー保護と著作権保護の新たなバランスを見出すための、法制度と技術導入のあり方について考察する。

2. ウェブスクレイピングが直面する法的課題

ウェブスクレイピングは、インターネット上の公開情報を効率的に収集する強力な手段である一方、その行為が既存の法制度と衝突する可能性を内包している。特に、日本のウェブスクレイピングを巡る法的議論は、主に個人情報保護法と著作権法の二つの側面から展開されている。これらの法的な枠組みは、データ活用の自由と個人の権利保護の間で複雑な均衡を保とうとしているが、AI技術の発展という新たな要素が加わることで、その均衡は揺らぎ続けている。

2.1. 個人情報保護法との摩擦：要配慮個人情報の取得を中心として

日本の個人情報保護法は、個人のプライバシー権を保護することを目的とし、個人情報の取得、利用、提供に関して厳格なルールを定めている。この法律は、個人の情報が不当に利用されることによって生じる様々なリスクから、個人を保護しようとするものである。特に、ウェブスクレイピングにおいて問題となるのは、「要配慮個人情報」の取得である。

要配慮個人情報とは、人種、信条、社会的身分、病歴、犯罪の経歴、犯罪被害に関する情報など、不当な差別や偏見、その他の不利益が生じないようにその取扱いに特に配慮を要する個人情報（個人情報保護法第2条第3項）を指す。これらの情報は、個人の尊厳や社会生活に深く関わるため、その取り扱いには細心の注意が払われるべきである。現行の個人情報保護法第20条第2項では、要配慮個人情報を取得する際には、原則として事前に本人の同意を得なければならないとされている。これは、個人のプライバシー保護の根幹をなす「同意主義」の原則であり、個人が自身の情報のコントロール権を持つという情報自己決定権の思想に基づいている【2】。

しかし、ウェブスクレイピングは、インターネット上に公開された情報を無作為的に、かつ大量に収集する性質を持つ。この特性ゆえに、スクレイピングを行う事業者が、個々のデータ主体（個人）から事前に同意を得ることは、技術的にも運用的にも極めて困難であり、実質的に不可能な場合が多い。例えば、個人が自身のブログやSNSの投稿に病歴や信条に関する記述を含めたり、あるいは政治的な意見表明をしたりした場合、スクレイピングによってそれらの要配慮個人情報が収集される可能性は高いが、ウェブサイトの利用者全員、あるいは特定の投稿者一人ひとりから同意を得ることは現実的ではない。例えば、数十万、数百万といった単位でデータを収集するAI学習においては、個別の同意取得は事実上不可

能である。

このように、ウェブスクレイピングの無作為的・大量収集という性質と、要配慮個人情報の取得における原則同意主義が真っ向から衝突する点が、個人情報保護法上の課題となっている [3]。同意がないまま要配慮個人情報を取得することは、事業者にとって大きな法的リスクであり、AI 開発やデータ活用を阻害する要因となりうる。

この課題に対し、個人情報保護委員会は、特定の条件下において「取得」に該当しない、あるいは本人同意を要しない旨の説明を行っている。特に、生成 AI 開発企業である OpenAI に対する注意喚起の概要において、委員会は「要配慮個人情報が含まれないよう必要な取り組みを行うなど所定の場合には、『取得』に該当せず、本人同意を要しない」旨の説明をしている。この見解は、スクレイピングを行う事業者が、データ収集の段階で要配慮個人情報が混入しないよう、技術的・組織的な措置を講じることで、法的リスクを軽減できる可能性を示唆している。しかし、インターネット上の膨大な、多様な形式のデータから要配慮個人情報を完全に排除することは、技術的に困難である [4]。仮に高度なフィルタリングを施したとしても、誤って要配慮個人情報が取得されてしまうリスクは常に存在し、その場合の法的責任は依然として事業者側に課される。また、この委員会見解は、「取得」の定義そのものを変更するものではなく、あくまで「要配慮個人情報が含まれないよう必要な取り組み」がなされた場合に限定されたものであるため、要配慮個人情報が意図せずとも取得されてしまった場合のリスクは、依然として高いままである。事業者は、このリスクを完全に回避するための技術的・組織的努力を継続的に行わなければならず、その負担は大きい。加えて、この見解は法的拘束力を持つ判例や法改正ではなく、その解釈や適用には不確実性が残っている。

このような現状を踏まえ、日本では、AI 開発などのデータ活用ニーズに応えるため、個人情報保護法の改正に向けた議論が進められている [5]。特に、AI 学習データ収集における「同意を要しない制度設計」の導入が検討されている点は注目に値する。この議論は、データ活用を促進し、日本の AI 産業の国際競争力を強化するという政策的意図がある。検討されている方向性としては、「統計作成等、特定の個人との対応関係が排斥された一般的・汎用的な分析結果の獲得と利用のみを目的とする」個人データ等の取扱いについて、同意を要しない制度設計とすることや、「本人の同意なく公開されている要配慮個人情報の取得」についても、一定の条件（取得者の氏名・名称、行おうとする統計作成等の内容等の公表義務、取得した個人データの目的外利用及び第三者提供の禁止など）を満たせば認める方向が挙げられる。これらの条件を遵守することで、個人の権利利益への不当な侵害を回避しつつ、データ活用を促進するバランスを目指すとしている。

しかし、この「同意を要しない制度設計」の導入、特に要配慮個人情報の取得を認める方向性には、課題も多いとされている。これは、個人のプライバシー権、特に「みだりにデータ化されない自由」や「基本的人権の保護」といった、データ主体の権利保護とのバランスをいかに図るかという、困難な課題を内包しているためでもある。例えば、公開情報であっても、個人の意図は特定のコミュニティ内の共有や

ALIS

自己表現であり、それが AI 学習に用いられ、商業目的で利用されることまでは想定していない場合が多い。また、一度 AI 学習に利用されたデータが、将来的にどのように利用されるか、あるいは再識別化されるリスクがないかといった点についても懸念が残る。匿名化された公開情報であっても、AI の高度な分析能力や外部情報との結合によって、個人が再識別化されるリスクは常に存在する [6]。特に、顔画像や位置情報といった機微なデータは、その識別性が極めて高いため、公開情報であっても慎重な取り扱いが求められる。

したがって、AI 時代においては「公開情報だから同意不要」「特定の個人との対応関係が排斥されているから同意不要」等といった単純な図式は成り立たず、公開の意図、情報の機微性、利用の目的と範囲、再識別化のリスクなどを総合的に考慮した上で、プライバシー保護の境界線を厳密に定義する必要があると考えられるが、この境界線の設定は、データ活用と権利保護のバランスをいかに取るかという、現代社会における最も重要な課題の一つである。

2.2. 著作権法との摩擦：AI 学習における著作物利用

著作権法は、著作者の権利を保護し、文化の発展に寄与することを目的としている。ウェブスクレイピングによって、ウェブサイト上の文章、画像、動画などの著作物（著作権法第 2 条第 1 項第 1 号）を収集することは、原則として著作権者の許諾を要する「複製」（同法第 2 条第 1 項第 15 号）行為に該当する。また、スクレイピングしたデータをサーバーに保存し、AI 学習に供する行為は、「公衆送信」（同法第 2 条第 1 項第 9 号の 5）に該当する可能性もある。これらの行為は、原則として著作権侵害となり、民事上の損害賠償請求や差止請求、刑事罰の対象となる可能性がある。

ウェブスクレイピングは、ウェブサイト上のコンテンツを自動的にダウンロードし、自身のサーバーなどに保存する行為であるため、著作権法上の「複製」に該当すると考えられている。このため、著作物利用の適法性は、著作権法が定める「権利制限規定」の解釈に大きく依存することになる [7]。AI 学習における著作物利用の適法性を巡る議論では、特に著作権法第 47 条の 5（情報解析のための複製等）が焦点となる。同条は、情報解析の用に供するため、その必要と認められる限度において、著作権者の利益を不当に害しないことを条件に、著作物を複製等することを認めている。

AI がウェブスクレイピングで収集した著作物を学習データとして利用する行為は、まさに「情報解析」に該当すると解釈される可能性がある。AI は、個々の著作物の表現そのものを人間が享受するのではなく、それらのデータから言語のパターン、画像の特徴量、論理構造などを統計的に抽出し、モデルを構築することを目的としているため、「享受を目的としない利用」と評価されやすい。しかし、議論的となるのは、「著作権者の利益を不当に害しないこと」という要件である。特に、AI が学習した結果として「享受目的のコンテンツ」を生成する場合、この要件の解釈が複雑となる。このように、著作権法第 47 条の 5 の解釈を巡る不確実性が、AI 開発事業者の法的リスクを高めている。

3. 秘密計算による「交通整理」の可能性と限界

上記で述べたように、ウェブスクレイピングは、個人情報保護法、著作権法、そして利用規約との間で複雑な法的課題を抱えている。これらの課題に対する解決策として近年注目されるのが、「秘密計算」である。秘密計算は、プライバシー保護とデータ活用の両立を目指す技術として、ウェブスクレイピングが直面する法的リスクの軽減に貢献しうる。

3.1. 秘密計算のメカニズムと特徴

秘密計算とは、複数の参加者がそれぞれ秘密のデータを持っている場合に、お互いの秘密データを明らかにすることなく、共同で計算を行い、その計算結果だけを得る暗号技術の総称である【8】。秘密計算の基本原理は、データを暗号化したまま、あるいは複数の参加者に分割して分散させた状態で計算を行うことにある。これにより、どの参加者も他の参加者の秘密データを平文で知ることなく、最終的な計算結果だけを共有できる。

主な技術要素としては、準同型暗号、秘密分散等が挙げられる。準同型暗号は、暗号化されたデータを復号することなく、そのデータに対して特定の計算（加算や乗算など）を実行できる暗号方式である。秘密分散は、秘密のデータを複数の断片（シェア）に分割し、各参加者に配布する技術である。これらの技術要素を組み合わせることで、様々な秘密計算プロトコルが構築され、多様な計算タスクに対応できるようになっている。具体的な応用分野としては、金融分野での不正検知、医療・製薬分野での共同研究、政府・公共分野での統計調査、マーケティング分野での広告効果測定、AI・機械学習分野でのモデル学習などが期待されている。これらの応用は、秘密計算がプライバシー保護とデータ活用の両立を可能にする強力なツールであることを示している。

秘密計算の最も重要な利点は、プライバシー保護とデータ活用という、一見すると相反する要件を同時に満たすことができる点にある。秘密計算は、個々の生データを第三者に開示することなく計算を行うため、データ漏洩や不正利用のリスクを根本的に排除できる。これにより、再識別化のリスクを大幅に低減し、個人のプライバシーを強固に保護できる。従来のデータ共有では、プライバシー懸念からデータ連携が進まないケースが多くあったが、秘密計算はプライバシー保護を担保しつつ、これまで共有が困難であった機密性の高いデータの連携を可能にする。これにより、新たな知見の発見、より精度の高いAIモデルの構築、社会課題解決のためのデータ分析など、データ活用の可能性を広げることが期待される。

3.2. 個人情報保護法との「交通整理」：同意不要化への貢献と限界

秘密計算は、特に個人情報保護法におけるウェブスクレイピングの課題、すなわち「要配慮個人情報の取得における同意」に関しても、解決策を提示しうる。

現行の個人情報保護法では、要配慮個人情報の取得には原則同意が必要である。しかし、秘密計算を用いる場合、データ利用者はウェブサイト運営者から平文の個人情報を受け取ることはない。データは暗号化された状態、または複数の断片に分割された状態で処理されるため、データ利用者は個人情報の「内容を認識できない」状態を維持するが、これにより、実質的に「個人情報を取得していない」、あるいは「同意が不要となる例外的な取得」として評価される可能性が高まると思われる。これは、個人情報保護

法が保護しようとする「個人の識別性」が、秘密計算によって技術的に排斥されるためである。

現在検討されている個人情報保護法の改正案において、「AI 開発等のための統計作成等、特定の個人との対応関係が排斥された一般的・汎用的な分析結果の獲得と利用のみを目的とする個人データ等の取扱い」については、同意を要しない制度設計とする方向性が示されているが、秘密計算は、この制度設計を技術的に裏付ける強力な手段となると考えられる。秘密計算は、個々の個人データの内容を秘匿したまま、統計的な特性や AI モデルの学習といった「結果」だけを得ることを可能にする。これにより、「特定の個人との対応関係が排斥された」状態を技術的に担保し、個人の識別性を排斥しつつデータの有用性を損なわない状態が実現すると考えられるためである。要配慮個人情報が混入するリスクがある場合でも、秘密計算を用いることで、個別の要配慮個人情報が AI 学習者に開示されることなく、特定の分析を行うことが可能となる。これは、要配慮個人情報の取得に関する同意要件を検討する上で、技術的根拠となり得る [9]。

しかし、秘密計算にも限界がある。秘密計算は、処理過程でのデータ秘匿性を保証する強力な技術であるが、その結果として得られた統計分析情報が、個人に不利益をもたらす決定に利用されるリスクを直接的に解決するものではない。例えば、秘密計算によって算出された「特定の年齢層の顧客は、平均的に与信リスクが高い」という統計情報が、金融機関の与信判断に利用されたとする。この場合、個々の顧客の生データは誰にも見られていないが、その顧客が該当する統計的グループの「平均」に基づいて与信が拒否される可能性がある。AI プロファイリングにおいて、学習データに既存の社会的な偏見や差別が含まれていた場合、秘密計算によって安全にモデルが学習されたとしても、そのモデルから導き出される統計情報や予測結果にはバイアスが反映され、結果的に特定のグループに対する差別的な決定が自動的に行われるリスクがある。

また、秘密計算は、個人がデータ利用を「拒否する権利」や、自身のデータがどのように使われるかを「決定する権利（情報自己決定権）」を自動的に保障するものでもない。個人が「自分のデータが、たとえ秘密計算によって安全に処理されたとしても、特定の目的のために使われること自体を望まない」と考える場合、秘密計算は個人の意思を反映するものではない。特に顔画像データや位置情報のような機微な情報の場合、データが安全に処理されること以上に、そのデータが特定の目的で利用されること自体への忌避感が強い場合がある。このように、秘密計算は技術的なプライバシー保護に大きく貢献するが、統計情報が社会に与える影響や、個人の権利・自由といった、より広範な問題に対する万能薬ではない。技術的な解決策のみでは、社会的な公正性や倫理的な問題を完全に解決することはできないため、法制度や倫理的枠組みとの連携が不可欠である。

3.3. 著作権法との「交通整理」：著作権者の懸念緩和と限界

秘密計算は、著作権法上の問題に対しても間接的ながら影響を与え、著作権者側の懸念を一部和らげる可能性を秘めている。

著作権者が最も懸念するのは、自身の著作物が無断で大量に複製され、自由に利用されることによる市

ALIS

場への影響やコントロールの喪失である。従来のウェブスクレイピングでは、著作物がそのままコピーされて AI 学習データとして保存されるため、著作権者は自身の作品が不正に利用されるのではないかという強い懸念を抱く。これに対し、秘密計算は、データが秘匿されたままであり、AI 学習者が元の著作物の「具体的な表現」を直接的に閲覧・利用できない、あるいはアクセスが極めて限定される状況を作り出す。データは暗号化された状態、あるいは複数の断片に分割された状態で処理されるため、学習プロセスにおいて平文の著作物が AI 開発者の手元に複製・保存されるわけではない。この「秘匿性」が、著作権者が抱く「著作物が丸ごとコピーされて悪用されるのではないか」という懸念を、心理的・実質的に一部和らげると考えられる。

これにより、秘密計算を用いた AI 学習は、著作権法第 47 条の 5 の適用をより強く主張できる根拠となり、著作権者の許諾なしに著作物を学習データとして利用できる可能性を高める。AI 学習者が元の著作物の内容を平文で取得・利用しないため、「著作権者の利益を不当に害している」とは言いにくい状況を生み出すことができる。

ただしここでも、秘密計算は著作権法の問題を根本的に解決する万能薬ではない。秘密計算は「入力データの秘匿性」を保証するものであり、「出力結果の著作物性」を保証するものではない。AI が学習した結果として生成されるコンテンツが元の著作物と酷似している場合など、生成 AI の著作権侵害問題は秘密計算では解決できない。これは、AI 生成物が著作権者の利益を不当に害すると判断される可能性が依然として残るためである。例えば、秘密計算によって安全に学習された AI モデルが、特定の画家の作品に酷似した画像を生成し、それが市場で販売された場合、著作権侵害となる可能性は依然として存在する。秘密計算は学習プロセスにおけるデータの秘匿性を担保する技術であり、AI が生成するコンテンツの著作権法上の評価には直接影響しない。秘密計算は著作権問題の一部を解決するものの、生成 AI が提起する著作権問題全体を解決するものではないことを自覚する必要がある。

4. 結論：AI 時代におけるデータ活用の新たな地平

ウェブスクレイピングは、デジタル社会におけるデータ活用の基盤技術である一方、日本の個人情報保護法、著作権法との間で複雑な法的摩擦を生じさせている。特に、要配慮個人情報の取得における同意の原則や、AI 学習における著作権者の利益の保護は、AI 開発とプライバシー保護のバランスを巡る喫緊の課題である。

本稿で検討した秘密計算は、これらの法的課題に対する解決策となり得る可能性を秘めている。秘密計算は、データ利用者が個人情報の内容を認識できない状態で計算を実行することを可能にし、要配慮個人情報の「取得」に関する同意の要件を緩和する技術的根拠を提供し得る。これにより、プライバシー保護を担保しつつ、AI 開発のためのデータ活用を促進する「同意を要しない制度設計」の実現に資する。また、秘密計算は、著作物が「平文のまま無断複製・利用される」という著作権者の懸念を、技術的に和らげる効果が期待できる。これにより、著作権法第 47 条の 5 の「著作権者の利益を不当に害しないこ

と」という要件の充足を補強し、AI学習における著作物利用の適法性を高める。

ただし、秘密計算があらゆる課題を解決する万能薬足り得ないことは既述の通りであり、またその導入と普及には、技術的複雑性や計算コスト、そして法制度における明確な位置づけといった課題が依然として存在する。

AI技術が社会に深く浸透し、その影響力が拡大する現代において、データの活用と個人の権利保護は、避けて通れない二律背反の課題である。秘密計算は、この二律背反を技術的に克服し、データ活用と権利保護の新たなバランス点を見出すためのツールとなり得る。課題は依然として多いものの、日本がAI開発の産業振興とデータ主体の基本的人権保護という難課題に対し、秘密計算のような先進技術を積極的に活用し、国際的な調和を図りながら、デジタル社会におけるデータ活用の「新たな地平」を切り開くことが期待される。

注釈

1. 具体的なリスクとしては、プライバシー侵害、著作権侵害、サーバー負荷、利用規約違反の点が挙げられる。プライバシー侵害では、ウェブサイト上に公開されている個人情報（氏名、連絡先、SNS投稿、写真など）を無断で収集・利用することによるものが問題となる。著作権侵害では、ウェブサイト上の文章、画像、動画などの著作物を無断で複製・利用することによるものが問題となる（AI学習データとしての利用は、享受目的ではないという主張もあるが、その結果として生成されるコンテンツが元の著作物の市場を侵害する可能性も指摘されている。）。サーバー負荷では、大量のスクレイピングリクエストがウェブサイトに集中することで、サーバーに過度な負荷がかかり、サイトの機能停止やサービス提供の妨害につながる。これは、サイト運営者にとって経済的な損失をもたらすだけでなく、他の利用者にとっても不利益となる。利用規約違反では、多くのウェブサイトが、その利用規約においてスクレイピング行為やデータの二次利用を明示的に禁止しているため、これに違反する行為は、契約違反として法的責任を問われる可能性がある。
2. 千葉恵美子「デジタルプラットフォームとルールメイキング」日本評論社 山本龍彦「AIと憲法 -アルゴリズム、プライバシー、デモクラシー」 44-61頁
3. ウェブスクレイピングと個人情報保護法との関係について論じたものとして中崎尚「生成AI法務・ガバナンス——未来を形作る規範」商事法務 第2章。
4. 個人情報保護委員会「「個人情報保護法 いわゆる3年ごと見直しに係る検討」 の今後の検討の進め方」に対して寄せられた意見の概要 13頁
5. 前掲注4) 4、9、12頁等
6. JILISレポート(2025.05.28)「個人情報保護法3年ごと見直し論議をめぐって」12-13頁。

7. ウェブスクレイピングと個人情報保護法との関係について論じたものとして中崎前掲注 3) 第 5 章。
8. エヌ・ティ・ティ・コミュニケーションズ株式会社「秘密計算を用いた データ利活用実践のためのガイドライン 【1.概説編】中間報告 ver.1 (2025 年 3 月)」
9. 前掲注 4) 12 頁では「「公開されている要配慮個人情報の取得」について、統計情報等を作成する際、一般の個人情報を取得する場面と取扱いを分ける合理的理由 は乏しい。このため、事業者の厳格な情報管理を前提として、取得者における一定の事項（取得者の氏名・名称、行うとする統計作成等の内容、 本規律に基づく本人同意なき個人データ等の第三者提供を行う目的である旨等）に関する公表を義務化することなく、公開されている要配慮個人情報の同意のない取得を認めるべき。」という意見が存在する一方で、13 頁では「要配慮個人情報を含むデータセットを AI 開発に利用する場合、差別的な AI の生成リスクなど特有の問題が生じうる。こうした問題に対処するための技術的・組織的措置についても検討すべき」という意見も存在しており、本人同意を不要とする「統計作成等」の要件については意見の一致を見ないのが現状である。もっとも、一つの解決策として、取り扱う情報の性質等にかんがみ、個人のプライバシーに及ぼす悪影響が大きいものとして類型的に認められるものについては、その処理過程に秘密計算技術を含むプライバシー強化技術を導入することにより、本人同意を不要としてよいという立法措置を講ずることも検討に値すると思われる。なお、本稿は筆者個人の見解であり、筆者の所属する組織の見解を示すものではない。