

なぜ「消費者をエンパワーするデジタル技術」を 検討する必要があるのか？

慶應義塾大学大学院法務研究科 教授
山本 龍彦

1. 技術的・認知的非対称性

■現代の消費者問題の多くは、伝統的な法的思考になじみにくい

・民事上の損害賠償→「損害」の認識（ここでは金銭的被害が中心。ベタな被害の認識→事後的対応）

・消費者契約法の取消し→「誤認」の認識、「困惑」の認識

→被害や誤認に関する消費者の明確な認識を前提（「損害＝実害」も前提）。

→AI等の先端技術×認知科学（心理学、行動経済学、神経科学等）との結合

→認知バイアス（脆弱性）などを利用した意思決定過程の操作。ソフトな介入。

→**情報の非対称性から技術的・認知的非対称性へ**（事業者は圧倒的な技術的優位性を持ち、人間または個人の認知過程に対する圧倒的なアクセシビリティをもつ）

→消費者は認知過程に介入され、知らないまま誘導・操作される（問題を明確に認識できない、「損害」が十分に成熟していないetc. 「激痛」と「鈍痛」）

※典型としてのダークパターン（ローチ・モーターetc.）

→伝統的な法的思考を補完するようなアプローチが必要

→実害“後”からの法的介入を補完するものとして、消費者の技術的エンパワーメントによる事前的な予防（技術的・認知的非対称性の是正には、消費者支援的な技術を供与し、消費者の認知的な耐性を強化する必要があるのではないか？ 気づいたときはもう遅い。取り戻すには莫大なコストがかかる。そうであれば事前に予防した方が実効的な権利保護）

※消費者団体の（ある意味における）負担軽減（実害の軽減？）。ただし、役割の変化は必要？（技術的モニタリング。例えば、ダークパターン、ホワイトパターンの評価、公表？）

2. 「アテンション・エコノミー」と時間

■ 「アテンション・エコノミー」とは？

→インターネットの普及による情報過多世界では、人々が払える「アテンション（関心）」や消費時間が情報量に対して圧倒的に希少となるため、これが交換財として経済的価値を持つようになる。

SNSは「無料」か？

魅惑的なコンテンツで利用者を引きつけ、そこで獲得した当該利用者の消費時間ないしアテンションを広告主等に販売するというビジネス・モデル（「セイレーン・サーバー」E. Glen Weyl）

→時間・アテンション奪取をめぐる競争（個人データとAI）

ホメーロスの叙事詩『オデュッセイア』で 英雄オデュッセウスを誘惑する魔女セイレーン



「セイレーンに襲われるオデュッセウス一行」(1891)

ジョン・ウィリアム・ウォーターハウス (John William Waterhouse, 1849 - 1917)

・現在では情報の受け手の全ての時間が――かつては非商業的な時間であった友人や家族と過ごす時間さえも――激しい競争の的となっており、「我々の毎時間、実際には毎秒が、それを支配しようという商業的アクターの標的になっている」（Tim Wu）。

・「パートナーとつながったり、本を読んだり、終わらせようと思っていた他のことをしたりするのに使えたはずの時間」を失っている。「気を逸らすような情報や刺激的な出来事が、四六時中あらゆる方向から押し寄せてくるのは、あなたのせいではない。それがアテンション・エコノミーなのだ」（Megan Wray）。

・集中力のないことで有名な金魚の平均的な注意持続時間は9秒。マイクロソフト社の新しい研究によると、人間はそれ以下に。「デジタル化が進むライフスタイルが脳に与える影響を浮き彫りに」。2000年（モバイル革命が始まった頃）以降、平均注意力が12秒から8秒に低下。「マルチスクリーンを多用する人は、無関係な刺激をフィルタリングするのが難しく、複数のメディアストリームに気を取られやすい」。「この変化は脳が時間とともに適応し、変化する能力の結果であり、注意力の低下はモバイル・インターネットへの進化の副作用かもしれない」（Kevin McSpadden）。

■ 気づかずに奪われているアテンション・時間？

→ 「アテンション決定 (attentional decision) 」を行う脳のメカニズム
(二重過程理論)

① 直感的で処理速度の速い「システム1 (自動システム) 」

② 論理的・内省的で処理速度の遅い「システム2 (熟慮システム) 」

→ 心理学や脳神経科学といった認知科学の知見を駆使したUX (ユーザー体験) やレコメンデーションは、①に強く働きかけ、「意思」を介在させないかたちでユーザーの消費者のアテンション・時間を奪う傾向。

※若年層に人気のあるショート動画系プラットフォームは、縦スクロール画面に利用者の嗜好に合った動画を次々とレコメンドし、閲覧のために画面を指でスワイプさせる「For You Page」をもつ。この「縦スクロール+スワイプ+レコメンデーション」は、利用者のドーパミンを誘発し、スワイプする指を止められなくなる「究極のロットマシーン」とか、「デジタル・コカイン」などと呼ばれ、その中毒性の高さが指摘されている。

→「ドーパミン誘発型UX (dopamine-triggering UX)」は、①システムを刺激し、自動的「反射」を強制的に引き出す。

→「思考」の剥奪（熟慮できない構造）

→Tim Wuは、これを「同意によらないアテンションの強奪 (seizure)」と呼ぶ。さらに、アテンション・エコノミーのなかで行動する消費者を「囚われの聴衆 (captive audience)」と表現。

■ EUデジタルサービス法（DSA）の背景・目的

・プラットフォーム事業者は、ユーザーの情報に基づいてプロフィールを作成し、パーソナライズされたコンテンツを提供している。こうしたレコメンドシステムは、ユーザーエンゲージメントの最大化が目的であり、企業はユーザーを多様なコンテンツに触れさせるのではなく、最も惹きつけるコンテンツに触れさせることを優先する。これは、ユーザーの多様なコンテンツへの接触や情報へのアクセスに悪影響を与える可能性がある。

・ レコメンドシステムの具体的な手法や根拠については、透明性が欠如しており、ユーザーが無自覚のうちにセンセーショナル/虚偽/過激なコンテンツの「ラビットホール〔底なし沼〕」に嵌まることになりかねないという懸念が繰り返し提起されてきた。

※以上の記述は、総務省「デジタル空間における情報流通の健全性確保の在り方に関する検討会ワーキンググループ（第14回）」資料WG14-1 みずほリサーチ&テクノロジーズ「レコメンデーションやアルゴリズムに関するルール整備の諸外国動向調査 結果紹介」から引用。次頁も同様。

● 欧州委員会による執行事例ーTikTokに対する正式な手続き

- 2024年2月19日、正式なDSA違反の評価手続きを開始すると発表。これは、TikTokが過去に提出したリスク評価報告書の分析、および欧州委員会のRFI〔情報要求〕に対する回答など、これまでに実施された予備調査に基づき、正式な調査を行うもの。

- 調査の論点は以下のとおり

- アルゴリズムを含むTikTokのシステム設計に起因する、実際のまたは予見され得る悪影響の観点からの、依存症やいわゆる「ラビットホール効果」の誘発につながるシステムのリスクの評価と軽減に関連するDSA義務の遵守状況

このような評価は、人の身体的・精神的福利に対する基本的権利の行使、児童の権利の尊重、および過激化の過程への影響に対する潜在的リスクに対抗するために必要となる。 また、TikTokが不適切なコンテンツへの未成年者のアクセスを防止するために使用している年齢確認ツールなど、一部の緩和措置が、合理的、比例的かつ効果的でない可能性がある

【参 考】

プロファイリングに基づくターゲティング、レコメンデーションは、自律的な意思決定を歪めるリスクをもつのではないか。

第26条 オンラインプラットフォームにおける広告

3. オンラインプラットフォームのプロバイダは、GDPR第9条（1）で言及される特別カテゴリーの個人データを使用して、・・・プロファイリングに基づいて、サービス受領者に広告を提示してはならない。

第28条 未成年者のオンラインでの保護

2. オンラインプラットフォームのプロバイダは、サービス受領者が未成年者であることを合理的な確実性をもって認識している場合、サービス受領者の個人情報を使用して、GDPR第4条（4）に定義されるプロファイリングに基づく広告をそのインターフェイス上で提示してはならない。

第38条 レコメンダーシステム

第27条に定める要件（注：レコメンダーシステムの透明性）に加え、レコメンダーシステムを使用する超大規模オンラインプラットフォームおよび超大規模オンライン検索エンジンのプロバイダは、GDPR第4条（4）に定義されるプロファイリングに基づかない少なくとも1つのオプションを各レコメンダーシステムに提供しなければならない。

【参考】EUデジタル市場法（DMA）

第15条 監督の義務

1. 第3条に基づく指定後6ヶ月以内に、ゲートキーパーは、・・・そのコアプラットフォームサービスに適用している、またはそのコアプラットフォームサービス全体にわたって適用している消費者のプロファイリング技術について、独自に監査した説明を欧州委員会に提出しなければならない。欧州委員会は、監査済みの説明を欧州データ保護委員会に提出しなければならない。
2. 欧州委員会は、監査の方法および手順を策定するために、第46条(1)の(g)に言及する実施法を採択することができる。
3. ゲートキーパーは、第1項の監査された記述の概要を公開しなければならない。 その際、ゲートキーパーは、その営業秘密を尊重する必要性を考慮する権利を有する。ゲートキーパーは、少なくとも年1回、その説明と概要を更新しなければならない

第46条 実施規定

1. 欧州委員会は、以下の適用に関する詳細な取り決めを定めた実施法を採択することができる：（略）

(g) 第15条(1)に規定されている、消費者のプロファイリングに使用される技術に関する監査説明の方法および手順。この目的のために実施法の草案を作成する場合、欧州委員会は欧州データ保護監督官に諮問するものとし、欧州データ保護会議、市民社会およびその他の関連専門家に諮問することができる；（略）

【参考】EU消費者保護規則のより良い執行と現代化に関する指令

序文(45) 事業者は、自動化された意思決定と消費者行動のプロファイリングに基づいて、特定の消費者または特定のカテゴリーの消費者に提供する価格をパーソナライズすることができる。したがって、消費者は、提示された価格が自動化された意思決定に基づいてパーソナライズされた場合、購買決定において潜在的なリスクを考慮できるよう、明確に通知されるべきである。従って、指令2011/83/EUに、自動化された意思決定に基づいて価格がパーソナライズされた場合に消費者に通知するための特定の情報要件を追加すべきである。この情報要件は、自動化された意思決定に基づくパーソナライズを伴わない場合、市場の需要に応じて非常に柔軟かつ迅速に価格を変更する「ダイナミック」または「リアルタイム」価格設定などの手法には適用されるべきではない。

※DSA等の訳は、総務省「ICTサービスの利用環境の整備に関する研究会 利用者情報に関するワーキンググループ（第3回）」資料3-1 三菱総合研究所（MRI）「ダークパターン及びプロファイリングについて」を参照。

3. ダークパターン

=ダークパタン

- ・「ユーザーの自律性、意思決定または選択を破壊または損なう実質的効果をもって設計または操作されたユーザー・インターフェース」 (CPRA、CPA)

- ・ DSA第25条オンライン・インターフェースの設計と構成：「1. オンライン・プラットフォーム提供者は、そのサービスの受領者を欺いたり操作したりするような方法で、またはそのサービスの受領者が自由かつ情報に基づいた決定を行う能力をその他の方法で実質的に歪めたり損なったりする方法で、オンライン・インターフェースを設計、構成、運営しないものとする」。

- ・「ユーザーを混乱させ、実際の選好とは異なる特定の行動を取らせるように意図的にデザインされたUI」 (Danyang Li)

→「システム1」のハッキング (“Dark patterns can induce users to make irrational choices because they prompt users to use System 1 decision-making, which relies on impulse and heuristics, instead of System 2, which involves deliberate thinking”. Li, supra) 。

(例) 一度入ると出られないUI (ローチモーター)

→「タークパタンは、コソコソして、しばしば平均的な者には気づかれない。それらは、全員が気づかずにもっている認知バイアスを利用しているのである」 (William Blesch)

→誘導されていることに気づきにくい

(MRI資料より引用)

●プリンストン大学・シカゴ大学の研究者は、①ダークパターンの使用状況に関する大規模な調査、②調査結果に基づく認知バイアスの観点からの分類を実施し、2019年に発表した。調査対象とした1万1千強のショッピングサイトのうち、1,254 (11.1%) のサイトでダークパターンが発見された。また、ダークパターンについて7カテゴリ15パターンの分類を作成した。

Arunesh Mathur, Gunes Acar , Michael J. Friedman, Elena Lucherini , Jonathan Mayer, Marshini Chetty, Arvind Narayanan, “Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites”, <https://webtransparency.cs.princeton.edu/dark-patterns/>

【用いられる認知バイアス】

アンカリング効果：個人が将来の意思決定において、最初の情報（「アンカー」）に過度に依存する傾向。

バンドワゴン効果：他人が何かを評価しているように見えることにより、個人がその何かをより高く評価する傾向。

デフォルト効果：惰性により、デフォルトで割り当てられた選択肢に固執する個人の傾向。

フレーミング効果：同じ情報であっても、その提示のされ方によって、個人が異なる判断を下す傾向。「額縁効果」。

希少性バイアス：希少価値の高いものに高い価値を見出す個人の傾向。

サンクコストの誤謬：ある行為に資源を投入した場合、その行為によって不利益を被る可能性があっても、その行為を継続しようとする個人の傾向。

【参考】EUのAI法案

第5条 禁止されるAI実践

1. The following AI practices shall be prohibited:

- (a) the placing on the market, the putting into service or the use of an AI system that deploys subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective, or the effect of, materially distorting the behaviour of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing a person to take a decision that that person would not have otherwise taken in a manner that causes or is likely to cause that person, another person or group of persons significant harm;

→十分な情報に基づいた意思決定（informed decision）を行う能力を著しく傷つけることによって、人または人の集団の振る舞いを実質的に歪め、それによって、その人、他者または人の集団に重大な損害を与える、または与える可能性のある方法で、その人が他の方法では行わなかったであろう意思決定を行わせる目的または効果を伴うかたちで、人の意識を超えたサブリミナル的技法、または意図的に操作的もしくは欺瞞的な技法を展開するAIシステムを市場に投入、サービス開始、または使用すること。

※プロファイリング&ダークパターン？ 自律的な「思考」を奪い、人の振る舞いを操作する欺瞞的技法を可能にするAIシステム。「プロファイリングにより個々のユーザーの特徴や脆弱性を把握した上で、それに応じてダークパターンを使い分け、より効果的にユーザーの『操作（manipulation）』を行うこと」（MRI）。

第5条 禁止されるAI実践

(b) the placing on the market, the putting into service or the use of an AI system that exploits any of the vulnerabilities of a person or a specific group of persons due to their age, disability or a specific social or economic situation, with the objective, or the effect, of materially distorting the behaviour of that person or a person belonging to that group in a manner that causes or is reasonably likely to cause that person or another person significant harm;

→当該個人または他の者に重大な損害を与えるか、または与える合理的な可能性がある方法で、当該個人または当該集団に属する者の振る舞いを著しく歪める目的または効果を伴うかたちで、年齢、障害、または特定の社会的もしくは経済的状況に起因する個人または特定の集団の脆弱性を悪用 (exploits) するAIシステムを市場に投入、サービス開始、または使用すること。

※年齢、障害等の脆弱性につけ込み、振る舞いを操作するAIシステム

(c) the placing on the market, the putting into service or the use of AI systems for the purpose of the evaluation or classification of natural persons or groups of persons over a certain period of time based on their social behaviour or known, inferred or predicted personal or personality characteristics, with the social score leading to either or both of the following:

(i) detrimental or unfavourable treatment of certain natural persons or whole groups of persons in social contexts that are unrelated to the contexts in which the data was originally generated or collected;

(ii) detrimental or unfavourable treatment of certain natural persons or groups of persons that is unjustified or disproportionate to their social behaviour or its gravity;

→(c)次のいずれか又は両方をもたらす社会的スコアをもって、社会的振る舞い又は既知、推論若しくは予測される個人的若しくは人格的特徴に基づき、一定期間にわたって自然人又は集団の評価又は分類を行う目的でAIシステムを市場に投入、使用開始又は使用すること。

(i) データが当初生成または収集された文脈とは無関係な社会的文脈において、特定の自然人または集団全体に対して有害または不利な取扱いを行うこと。

(ii) 特定の自然人または集団に対して、正当化されない、又はその社会的振る舞いやその重大性に不釣り合いな有害または不利な取扱いを行うこと。

4. 終わりに

■人間の「理性」の限界は無視できない。脆弱性を内在していることを前提に議論を進めるべき（技術的・認知的非対称性により、脆弱性につけ込まれるケースが増大しているのではないか。「強くあれ」、「注意せよ」だけで問題は解決しない）。

→「**弱い**が、**強い個人**」 = **オデュッセウスの智慧**（マスト = アーキテクチャの事前装備。アーキテクチャを身にまとった人間）

→法制度 × 技術（アーキテクチャ） × リテラシーの三位一体

→このまま伝統的な思考に固執すれば、非対称性はますます大きくなる。

※そのときの「同意」にどのような意味があるのか？

【再掲】消費者の技術的エンパワーメントによる事前的な予防（技術的・認知的非対称性の是正には、消費者支援的な技術を供与し、消費者の認知的な耐性を強化する必要があるのではないか？）

消費者団体の（ある意味における）負担軽減（実害の軽減？）。ただし、役割の変化は必要？（技術的モニタリング。例えば、ダークパターン、ホワイtpatternの評価、公表？）