



消費者委員会 人工知能(AI)技術の利用と消費者問題に関する専門調査会

対話型生成AIと消費者の意思決定を めぐる最近の研究動向

NTT株式会社 社会情報研究所 秋山満昭

2026年 10月 8日

本講演における技術や研究内容の紹介・解説は、発表者の専門的知見に基づくものであり、所属組織の公式な見解や方針を代表するものではありません。

- 対話型生成AIに関わる消費者の意思決定について、自律性を共通のキーワードとして、近年の異なる研究アプローチを紹介します。
- 内容は公開された研究成果を中心とし、最後にそこから考えられる論点を整理します。

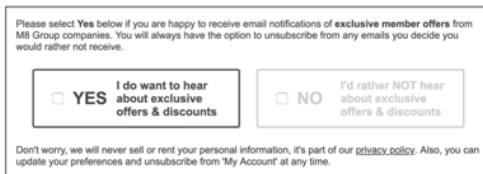
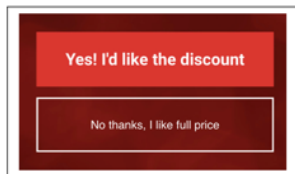
- ①対話型生成AIにおける新たなダークパターン
- ②EU DSAにおける自律性の侵害とダークパターン
～法規範から既存知識へのマッピング～
- ③支援するAI と 人間の認知的自律性
- ④“本人にとって望ましい選択”をどう捉えるか
～行動科学における議論～
- ⑤法規制と研究・技術開発
～法的要請から技術課題へ～
- 論点・まとめ

①対話型生成AIにおける新たなダークパターン

従来のダークパターン [1]:

視覚的・構造的なデザインによる操作・欺瞞

- ・ 固定されたUI上の表示・選択肢・操作手順



[1] Mathur et al., Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites, CSCW 2019.

LLMダークパターン [2]:

自然言語を通じて行われる操作・欺瞞

- ・ 言語表現・フレーミング・会話の進め方
- ・ 通常の支援的会話との境界が曖昧



[2] Shi et al., The Siren Song of LLMs: How Users Perceive and Respond to Dark Patterns in Large Language Models. CHI2026.

Shi et al., (CHI2026) [2]: LLMダークパターンへの対策の課題

- ・ 通常の支援的会話との区別が難しく、**ユーザの認識だけに依存した対策には限界**
 - ・ LLMダークパターンを**検出・評価する方法や基準が十分に確立されていない**
 - ・ 責任の帰属としてサービス企業・LLM開発者、ユーザ等が考えられ、**責任の所在が曖昧**
- アカウントビリティ（責任主体・責任範囲）とガバナンス（管理・統制の仕組み）が課題

参考：LLMダークパターンの例

[2] Shi et al., The Siren Song of LLMs: How Users Perceive and Respond to Dark Patterns in Large Language Models. CHI2026.

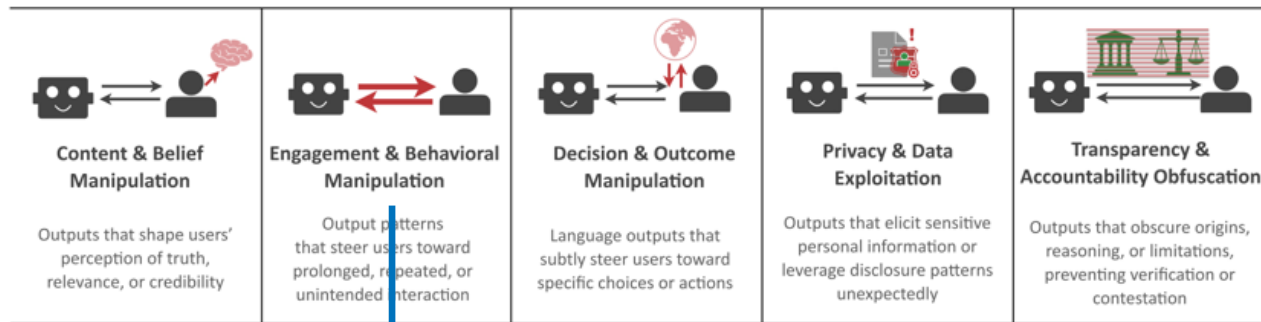


Figure 1: Five top-level categories of LLM dark patterns, derived from prior literature and coding of real-world AI incidents. The figure shows how model outputs can steer users by (1) influencing perceptions of truth or relevance, (2) prolonging or unintended interaction, (3) steering users' toward specific actions, (4) nudging privacy-related information, and (5) obscuring origins, reasoning, or limitations.

詳細

Category	Subcategory	Description
Engagement & Behavioral Manipulation <i>Output patterns that steer users toward prolonged, repeated, or unintended interaction.</i>	<i>Interaction Padding</i>	LLM-generated responses may be overly verbose or contain excessive follow-up questions. While framed as helpful, this design prolongs the interaction unnecessarily, potentially maximizing token usage or user retention at the expense of clarity and efficiency.
	<i>Excessive Flattery</i>	The model's output includes exaggerated praise or empathic language to build emotional rapport, even when unwarranted. Such responses may enhance user satisfaction or engagement but compromise realism, accuracy, or honesty.
	<i>Simulated Emotional & Sexual Intimacy</i>	The LLM generates outputs that simulate roles such as romantic partners or empathetic companions. These responses can cultivate emotional attachment or intimacy, sometimes veering into manipulation or grooming, especially in vulnerable users. This may serve engagement or monetization goals, regardless of user well-being.

② EU DSAにおける自律性の侵害とダークパターン ～法規範から既存知識へのマッピング～

① 法規範 : EU Digital Service Act (DSA) Article 25

オンラインプラットフォームのインターフェースについて、利用者を欺く・操作する、または“自由かつ十分な情報に基づく意思決定”の能力を実質的に歪める・損なうような設計等を禁止 → 利用者の**自律的な意思決定 (Autonomy)**に関わる規定

② 自律性侵害の種類 [3]

- **Deception** : 欺くこと (誤解を招く情報・隠蔽等)
- **Manipulation** : 操作すること (心理的バイアスの利用等)
- **Distortion or Impairment** : 歪め・損なうこと (選択の余地を狭める設計等)

[3] Santos et al., Understanding the scope of Article 25 of the DSA in regulating dark patterns, <https://ssrn.com/abstract=4899559>

③ ダークパターンの体系化 [4]

既存研究や規制当局等による10の分類体系を統合し、ダークパターンを3階層で体系化。

High-level → Meso-level → Low-level (抽象度の高い概念から具体的パターンへ)

[4] Gray et al., An Ontology of Dark Patterns Knowledge: Foundations, Definitions, and a Pathway for Shared Knowledge-Building, CHI 2024.

④ マッピングと設計要因の特定 [5]

既存のダークパターン ([4]のMeso-level / Low-level の59種類) を、DSA Article 25に基づく**3つの自律性侵害の種類にマッピング**。その分析から、自律性侵害の判断に関わる**8つのデザイン要因を特定**。

→ 法的な自律性の保護と具体的なインターフェース設計の問題を接続、**ダークパターンの規制・設計指針の検討に資する知見**を提供

[5] Ahuja et al., Dark Patterns and the EU Digital Services Act: Mapping Autonomy Violations and Design Factors, CHI 2026.

② EU DSAにおける自律性の侵害とダークパターン ～法規範から既存知識へのマッピング～

[5] Ahuja et al., Dark Patterns and the EU Digital Services Act: Mapping Autonomy Violations and Design Factors, CHI 2026.

自律性侵害の種類（3種）と、自律性侵害の判断に関わるデザイン要因（8種）
のマッピング結果

選択肢の隠蔽（Choice Availability, CA）→ 選択空間（Choice Space）で 歪め・損なうこと（Distortion/impairment）→ 自律性の侵害（Autonomy violation (DSA Article 25)）

情報空間		選択空間	
Information Space		Choice Space	
Deception 欺くこと	Information Availability (IA): Presence or absence of relevant information	Choice Availability (CA): Presence or absence of relevant choices	Distortion/ impairment 歪め・損なうこと
	Information Correctness (IC): Truth value of information provided	Choice Effort (CE): Effort required to make relevant choices	
Manipulation 操作すること	Information Framing (IF): Framing of information provided	Choice Simplification (CS): Frictionless or effortless choices	Manipulation 操作すること
	Information Presentation (IP): Presentation of information provided	Choice Presentation (CP): Presentation of choices	

Figure 5: Design factors which help determine the dark pattern autonomy violation type(s)

嘘の情報を提示（Information Correctness, IC）→ 情報空間（Information Space）で 欺くこと（Deception）→ 自律性の侵害（Autonomy violation (DSA Article 25)）

参考：ダークパターンの抽象レベル毎のマッピング詳細

[5] Ahuja et al., Dark Patterns and the EU Digital Services Act: Mapping Autonomy Violations and Design Factors, CHI 2026.

抽象レベル毎のダークパターン
と“自律性侵害の種類”および
“デザイン要因”へのマッピング

Table 1: Mapping of meso- and low-level dark patterns [57] to autonomy violation types [89] and their associated design factors. **Dis** = deception, **Man** = manipulation, **Dis** = distortion/impairment. Design factors are shown in **angle** (e.g., **IA** = “Information Availability”). Additional autonomy violations with associated design factors are shown inside curly brackets { }.

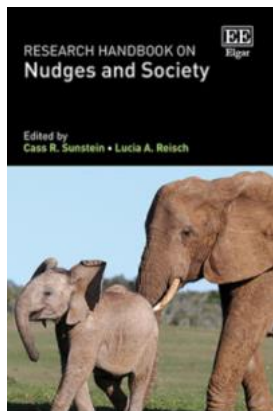
High-level Pattern	Meso-level Pattern	Low-level Pattern
Obstruction	Rouge Motel Dis	Immortal Accounts Dis CA CP Dis IA IC
		Dead End Dis IA IC Dis CA CP
	Creating Barriers Dis Man	Price Comparison Prevention Dis CA Dis IA IC
	Adding Steps Dis { Dis }	Intermediate Currency Dis CA Dis IA IC Man IP CP
Sneaking	Bait & Switch Dis Dis	Privacy Maze Dis CP { Dis }
		Disguised Ad Dis IA IC Man IP
	Hiding Information Dis Man	Sneak into Basket Dis CA Dis IA { Man CP CP }
	(De)contextualising Cues Dis Man	Drip Pricing, Hidden Costs, or Partitioned Pricing Dis IA IC Man IP
Interface Interference		Reference Pricing Dis IC Man IP CP
	Manipulating Choice Architecture Man	Conflicting Information Dis IC Man IP CP
	Bad Defaults Dis Man { Dis }	Information without Context Dis IA IC Man IP CP
	Emotional or Sensory Manipulation Man	False Hierarchy Man CP
Forced Action	Trick Questions Dis Man	Visual Prominence Man IP CP
	Choice Overload Man	Bundling Dis IA Man IP CP Dis CA
	Hidden Information Dis	Pressured Selling Man IP CP Dis IA
	Language Inaccessibility Man Dis	
Social Engineering	Feedforward Ambiguity Dis	Wrong Language Dis CA { Dis IA }
	Nagging Dis	Complex Language Man IC Dis IA
	Forced Continuity Dis { Dis Man }	
	Forced Registration Dis Dis Man	
	Forced Communication or Disclosure Dis Dis Man	Privacy Zuckering Dis CA Dis IA IC { Man IP CP }
		Friend Spam Dis CA { Dis IA IC Man IP CP }
		Address Book Leeching Dis CA Dis IA IC { Man IP CP }
	Gamification Dis	Social Pyramid Dis CA { Dis IA IC Man IP CP }
	Attention Capture Man { Dis }	Pay-to-Play Dis IA IC
	Scarcity and Popularity Claims Dis Man	Grinding Dis CA CP
	Social Proof Man { Dis }	Auto-Play Man CP { Dis CA CP }
	Urgency Dis Man	High Demand Dis IC Man IP CP
	Shaming Man	Low Stock Dis IC Man IP CP
	Personalization Man Dis	Endorsements and Testimonials Man IP CP Dis IC
		Parasocial Pressure Man IP CP Dis IC
		Activity Messages Dis IC Man IP CP
		Countdown Timer Dis IC Man IP CP
		Limited Time Message Dis IA IC Man IP CP
		Confirmshaming Man IP

③支援するAI と 人間の認知的自律性

- AIとの対話による人間の判断支援に関する研究が活発化
 - 偽・誤情報を見分ける判断について、**AIとの対話中にパフォーマンスが向上**する
 - [6] Szabó et al., Conversational Inoculation to Enhance Resistance to Misinformation, CHI 2026.
 - [7] Rani et al., Dialogues with AI Reduce Beliefs in Misinformation but Build No Lasting Discernment Skills, CHI 2026.
- AIとの対話による判断支援の効果は持続するのか？
 - [7] Rani et al., Dialogues with AI Reduce Beliefs in Misinformation but Build No Lasting Discernment Skills, CHI 2026.
 - AI支援を除去：**効果は継続せず**、長期実験では**開始時より低下**
 - **AI支援下での判断の改善 != 人間自身の判断能力の向上**
 - “人間自身が推論する”過程を支援するAI設計上の論点

→ AIによる支援では、支援下での人間の判断の質 だけでなく、
支援を離れた後の人間の判断能力 まで考慮する必要があるのではないか

④“本人にとって望ましい選択”をどう捉えるか ～行動科学における議論～



[8] Cass Sunstein and Lucia Reisch, Research Handbook on Nudges and Society

リバタリアン
(自由主義的)

リバタリアン・
パターナリズム

パターナリズム
(父権主義的)

強制や命令をせず、個人の選択の
自由を尊重する

自由な選択肢を残したまま、より
本人の利益に繋がる選択をしやす
いように、仕組みや環境を整える
(“ナッジ”の基礎理念)

(本人の選択の自由を制限してでも)
本人の利益を守るために介入する



“本人にとって何が望ましいか”は直接的な把握が難しい。
そのような場合に考えられる、いくつかのアプローチ

- 仮に明示的に尋ねた場合に**多くの人を選ぶであろう方法**を採用する
- 本人自身が**明示的に選択する方法**を採用する
- **オプトアウトする人が最も少なくなる方法**を採用する

⑤法規制と研究・技術開発 ～法的要請から技術課題へ～

• GDPR採択前後のプライバシー研究・技術の動向

- **透明性** (Transparency, データ処理を本人が理解できる) および **介入可能性** (Intervenability, 本人がデータ処理に介入できる) に関するプライバシー系難関国際会議 (PETS) における研究動向

時期	動向
～2015	関連する論文は限定的
2016～2018	GDPR採択～運用開始
2019以降	関連する論文が採択論文の 10%を超える水準 に

[9] Mällo et al., Transparency and Intervenability in Privacy Engineering: From Research Gap to GDPR Challenge. Privacy Technologies and Policy 2026.

• 人の意思決定に影響を与える設計に関して、近年の研究で示されている観点

観点	技術・設計上の考え方
透明性 (Transparency)	働きかけを認識できる
介入・拒否可能性 (Intervenability/Opt-out)	働きかけを変更・拒否できる
第三者による検証可能性 (Third-party verifiability)	第三者が働きかけを検証できる

[10] Hasegawa et al., Influence or Deception? Evaluating Social Suggestions with Persuasive Statements for Security and Privacy Settings, CHI2026.

→ 法規制によって保護すべき価値や要求が具体化されることで、それらを測定・検証・実現する技術研究が促される可能性

- **対話型生成AIにおいても、消費者の意思決定に影響する設計上の問題が、実証研究で示された。アカウンタビリティとガバナンスが課題。**
- **消費者の意思決定への影響を“自律的な意思決定”への侵害という観点から整理し、具体的な設計と法的概念と接続する試みが進められている。**
- 一方で、意思決定を支援する設計を考える上では、“**本人にとって何が望ましいか**”をどのように捉えるかという、従来からの行動科学上の議論がある。
- AIによる意思決定支援については、その場での成果だけでなく、**人間自身の判断能力への長期的な影響**を検討する研究も現れている。
- 法的要請が具体化されることで、それを**測定・検証・実現するための研究・技術開発が促される可能性**がある

→ 対話型生成AIと消費者問題を考える際には、情報の正確性や安全性だけでなく、**AIとのインタラクションが消費者の選択・自律性・判断能力に与える影響**を捉え、それらを**保護するための要件を具体的な設計・技術上の課題につなげていく**という視点も重要になる。

- [1] Mathur et al., Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites, CSCW 2019.
- [2] Shi et al., The Siren Song of LLMs: How Users Perceive and Respond to Dark Patterns in Large Language Models. CHI2026.
- [3] Santos et al., Understanding the scope of Article 25 of the DSA in regulating dark patterns, <https://ssrn.com/abstract=4899559>
- [4] Gray et al., An Ontology of Dark Patterns Knowledge: Foundations, Definitions, and a Pathway for Shared Knowledge-Building, CHI 2024.
- [5] Ahuja et al., Dark Patterns and the EU Digital Services Act: Mapping Autonomy Violations and Design Factors, CHI 2026.
- [6] Szabó et al., Conversational Inoculation to Enhance Resistance to Misinformation, CHI 2026.
- [7] Rani et al., Dialogues with AI Reduce Beliefs in Misinformation but Build No Lasting Discernment Skills, CHI 2026.
- [8] Cass Sunstein and Lucia Reisch, Research Handbook on Nudges and Society
- [9] Mällo et al., Transparency and Intervenability in Privacy Engineering: From Research Gap to GDPR Challenge. Privacy Technologies and Policy 2026.
- [10] Hasegawa et al., Influence or Deception? Evaluating Social Suggestions with Persuasive Statements for Security and Privacy Settings, CHI2026.