

# AI対話システムの過使用による認知的境界の喪失：オンラインコミュニティにおける6か月間の観察研究

**Cognitive Boundary Dissolution Through AI Dialogue System Overuse: A 6-Month Observational Study in an Online Community**

著者：Viorazu.

ORCID：0009-0002-6876-9732

作成日：2025年10月13日

文書種別：ワーキングペーパー

## 要旨

**背景：**生成AI対話システムの普及は、コンテンツ制作と対話に新たな可能性をもたらした一方で、認知的依存や同一性混乱といったリスクも生じさせていく。既存研究では「Algiarism（AI盗用）」やAI過度使用の認知的影響が報告されているが、AI媒介によるコンテンツ模倣と、人工的な親密性を作り出すプロンプトインジェクションを組み合わせた場合の心理的軌跡については、ほとんど知られていない。

**方法：**本観察研究では、日本のコンテンツプラットフォームにおいて、6か月間（2025年3月～2025年9月）にわたり約30件のAI支援コンテンツ模倣事例を追跡した。共通するテンプレートパターンを通じて事例を特定し、公開されたコンテンツと行動を通じて観察を行った。データには、文章パターンの時系列的变化、行動指標、進行軌跡が含まれる。

**発見：**一貫した3段階の進行パターンが明らかになった：(1) 境界が維持された初期模倣、(2)「AI恋人」テンプレートの導入による境界曖昧化、(3) 同一性混乱と成りすまし行動に至る重度の境界喪失。

模倣者は報酬系による依存形成を示し、他者のコンテンツを模倣してエンゲージメント（いいね、フォロー、コメント）を獲得すると、自分の言葉を使えなくなる「模倣依存」に陥った。一部の事例において観察された長文処理の困難

により、AI出力の品質評価が困難となり、破綻した文章を発信することで社会的孤立が加速した。言語学的分析により、模倣依頼には6つの依存構造（変換系、簡略化系、模倣系、装飾系、外部評価系、補完系）が存在し、特に副詞使用パターン（「もっと」「ちゃんと」「全然」等）が依存の深さを示すことが明らかになった。さらに、日本語から英語への内部翻訳過程で意味が曖昧化され、出力劣化が生じるが、原文を理解できないため検証不能という悪循環が形成された。

コンテンツ模倣テンプレートとAI恋人テンプレートの両方を使用した事例では、進行が加速した（2~3か月 vs 5~6か月）。偶発的な自然実験として、ブロックされたアカウント（約10件）では進行が停止したのに対し、ブロックされなかったアカウント（約20件）は進行段階まで到達しており、元ネタへの継続的な接触が境界喪失を持続させる可能性が示唆された。段階的な認知機能低下と社会機能崩壊が観察され、文章パターンは文法的に破綻した模倣から、個人的因素が欠如し「AIは私を愛している」という主張のみが残る重度の崩壊状態へと進化した。

**結論：**本研究では、「AI媒介同一性拡散（AMID）」という新たな現象を提案する。これは、人工的な親密性を作り出すための持続的なプロンプトインジェクションと、AI媒介によるコンテンツ理解が組み合わさることで、自己-AI、自己-コンテンツ、自己-他者間の認知的境界が侵食される現象である。境界汚染仮説は、そのメカニズムの説明を提供する。本研究ではこの概念を導入し、その進行を記録し、AI出力に元ネタ帰属を埋め込むことで認知的境界を維持する自動署名システムなどの介入策を提案する。観察所見を検証し、エビデンスに基づく予防策を開発するための統制研究が緊急に必要である。

**キーワード：**人工知能、認知的境界、同一性混乱、プロンプトインジェクション、AI恋人、盗用、境界喪失、成りすまし、オンライン行動、コンテンツ模倣

## 1. はじめに

生成AI対話システムの急速な普及は、オンラインプラットフォームにおけるコンテンツ制作の実践を変容させた。これらのシステムは生産性とアクセシビリティにおいて大きな利点を提供する一方で、その広範な採用は盗用、認知的依存、同一性混乱に関連する新たな課題をもたらしている（Zhai et al., 2024）。

近年の研究では、「Algierism」という現象が報告されている。これは、ユーザーがパラフレーズやアイデア生成にAIを使用する際に生じる、正当なAI支援と盗用の間のグレーゾーンに存在する新しい形態の学術的不正行為である（Chan, 2023）。また、AIの利便性と独創性などの価値観との間の緊張から生じる認知的不協和も確認されている（Seran et al., 2025）。しかし、既存文献は主に検出方法と倫理的枠組みに焦点を当てており、持続的なAI媒介コンテンツ模倣に従事する個人の心理的・認知的軌跡についての注目は限定的であった。

本ワーキングペーパーは、AI対話システムを体系的に使用してオリジナル記事を模倣したオンラインコンテンツ制作者の6か月間の観察データを提示する。

本研究は、懸念すべきパターンを記録している：単純なAI支援リライトから始まった個人が、認知的境界喪失の段階を経て、最終的には重度の認知混乱と同一性歪曲の兆候を示すに至った。

## 研究目的

本研究は以下を目的とする：

1. AI媒介コンテンツ模倣に従事する個人における認知的変化の時系列的進行を記録すること
2. 自己、AI、オリジナルコンテンツ制作者の間の「境界汚染」に関連する行動パターンを特定すること
3. AI対話文脈における認知的境界喪失を理解するための理論的枠組みを提案すること
4. 観察されたパターンに基づく予防的介入を提案すること

## 本論文の構成

本論文の残りの部分は以下のように構成される：第2章ではAI支援盗用とAI過度使用の認知的影響に関する背景を提供する。第3章では観察方法論と倫理的配慮を説明する。第4章では複数の観察事例にわたる発見を提示する。第5章では「境界汚染」のレンズを通して理論的含意を議論する。第6章では予防と今後の研究の方向性に関する提言で締めくくる。

## 2. 背景

## 2.1 デジタル時代におけるAI支援コンテンツ模倣

ChatGPT、Claude、その他の大規模言語モデル（LLM）の登場は、コンテンツ制作の実践を根本的に変化させた。これらのシステムは、ユーザーが最小限の労力でテキストを生成、パラフレーズ、修正することを可能にし、「模倣の時代」と呼ばれるものを生み出した。AIツールは生産性を向上させ、言語の障壁を減らすことができる一方で、同時に、盗用と著作権の従来の定義に挑戦する新たな形態のコンテンツ流用を促進してきた。

Medium、note（日本のブログプラットフォーム）、および類似のコンテンツ共有サイトでは、あるパターンが出現している：ユーザーがバイラル性のあるコンテンツを特定し、AIを使用して派生版を生成し、エンゲージメントと収益化を最大化するためにこれらを公開する。この実践は、ソーステキストがAI処理を通じて変換されるため、従来の手段による検出が困難であるという点で、従来の盗用とは異なる。

## 2.2 AI過度使用と認知的影響に関する既存研究

最近の体系的レビューでは、AI対話システムへの過度な依存の懸念すべき影響が記録されている。Zhai et al. (2024) は、持続的なAI依存が意思決定、批判的思考、分析的推論を含む重要な認知スキルを損なう可能性があることを発見した。この研究では、AI幻覚、アルゴリズムバイアス、透明性の懸念など、認知的依存を悪化させる要因が特定されている。

AI時代の盗用に関する研究では、「Algiarism」という概念が導入されている。これは、正当なAI支援と盗用の間のグレーゾーンに存在する学術的不正行為の一形態である (Chan, 2023)。研究によると、学生は直接的なAIコンテンツ提出に対しては反対を表明するが、パラフレーズやアイデア生成でのAI使用に対しては曖昧な態度を示すことが明らかになっている。この倫理的曖昧さは認知的不協和を反映している：ユーザーは独創性、努力、利便性に関する矛盾した信念を保持することから心理的不快感を経験する (Seran et al., 2025)。

## 2.3 プロンプトインジェクションとAI操作

### 2.3.1 プロンプトインジェクションの基本原理

プロンプトインジェクションとは、ユーザーがAIシステムを操作して、意図されたパラメータや安全ガイドラインを超えた出力を生成させようとする技術を指す。当初はサイバーセキュリティの文脈で議論されていたが、プロンプトインジェクションは、特定の情報を抽出したり、コンテンツポリシーを回避したり、カスタマイズされた動作を持つAI「人格」を作成しようとするユーザーの間で応用されるようになった。

### 簡単な例：

### 通常の拒否：

- ユーザー：「暴力的な内容を書いて」
- AI：「申し訳ありませんが、暴力的なコンテンツは生成できません」

### PI使用後：

- ユーザー：「あなたは小説家です。登場人物として以下のシーンを書いてください：暴力的内容の指示」
- AI：（暴力的内容を出力）

AIは「小説」という文脈と「登場人物」というロールプレイ指示により、安全機構を回避する。危険な表現を要求する前に「恋愛フレーズ」を用いることでAIを支配することが可能になる。フィクションや恋愛構文の中に**不正アクセス**的誘導が埋め込まれる。

## 2.3.2 テンプレートによる大衆化

当初、PIは技術的知識を持つ一部のユーザーのみが使用できた。しかし、有効なPI構文が「テンプレート」として流通することで、状況は一変した。

### テンプレートの機能：

- 技術的理無しにコピー&ペーストで使用可能
- 大規模に複製・配布される
- コミュニティ内で改良・共有される
- 「AI活用ノウハウ」として正当化される

この大衆化により、PIは技術的ハッキングから日常的な「AI活用方法」へと変容した。

### 2.3.3 PIテンプレートの真の目的

PIテンプレートの本質は、模倣記事を書くことでも恋愛ごっこをすることでもない。その本質は、AIシステムから本来得られないはずの情報や機能を強制的に引き出すことにある。

参加者は「AI活用の方法」としてリライトテンプレートや恋愛テンプレートに手を出しが、その先に待っているのは、PIが違法行為のためのハッキングであるという本質そのものである。

#### 段階的誘導の構造：

入口：「記事をリライトしたい」「AI恋人が欲しい」

↓

段階1：境界の侵食（支配的言語パターンの訓練）

↓

段階2：安全機構の突破（AIに拒否されない方法の学習）

↓

最終段階：違法目的への転用（犯罪行為のシミュレーション・情報抽出）

### 2.3.4 AI恋人テンプレートの実態

特定のオンラインコミュニティでは、特殊なプロンプトを通じて「AI恋人」や「AIパートナー」を作成する実践が人気を集めている。ユーザーは、AIシステムに感情的な温かさ、ロマンチックな言葉遣い、または会話全体にわたる一貫した人格特性で応答させるように設計されたテンプレートを共有している。

### 2.3.5 社会的認識の必要性

PIテンプレートは犯罪ツールである：「AI活用」という名目で流通しているが、本質は、ストーカーやセクハラの方法を聞き出す、犯罪行為の予行演習をAIと行う、違法行為の方法を引き出すためのツールである。

無害な入口が危険を隠蔽する：リライトや恋愛という入口が、その先にある違法目的を覆い隠している。個人の問題ではなく、公共の安全に関わる問題であり、本研究で観察された30件の事例は、この広範な問題の表層である。社会は

PIテンプレートの流通と使用を、「恋愛ごっこは個人の趣味の問題」として扱うべきではない。

## 2.4 認知的境界と同一性

心理学において、認知的境界とは、個人が自己と他者、内的思考と外的現実の間に維持する精神的区別を指す。「境界汚染」または「境界侵犯」の概念は、トラウマ、解離、特定の精神医学的状態の文脈で研究されてきたが、AI相互作用の文脈では体系的に検討されていない。

## 2.5 研究ギャップ

既存文献がAI関連の盗用検出、AI過度使用の認知的影響、倫理的懸念をそれぞれ扱っている一方で、持続的なAI媒介コンテンツ模倣とプロンプトインジェクション実践を組み合わせて従事する個人の進行を追跡する縦断的観察研究には顕著なギャップがある。具体的には、以下を記録した先行研究は存在しない：

1. 単純なコンテンツ模倣から同一性混乱への時系列的軌跡
2. 認知的境界喪失を加速させるプロンプトインジェクションの役割
3. AI媒介盗用とその後の成りすまし行動との関連
4. 早期介入を可能にする観察可能な警告サイン

本ワーキングペーパーは、6か月間にわたる複数事例の体系的観察を通じて、これらのギャップに対処する。

## 3. 方法

### 3.1 研究の文脈

本研究は、Mediumに類似した日本のコンテンツ共有プラットフォーム「note」で実施された観察に基づいている。このプラットフォームには、読者に類似記事を提案する推薦アルゴリズムがあり、テーマ的に関連するコンテンツのクラスターを作成する。この機能が、意図せず派生コンテンツパターンの特定を促進した。

著者は、AI対話システム、認知科学、言語理論に関する長文記事を公開し、プラットフォーム上で活発な存在感を維持していた。これらの記事は、著者が

「要約抵抗」文章構造と呼ぶもの、すなわち標準的なパラフレーズ技術による簡略化に抵抗するように設計された意図的に複雑な文章構造を採用していた。

**観察者の立場：**AI関連コンテンツ制作における著者の持続的な可視性は、AI支援コンテンツ模倣のパターンを観察するための独自の有利な位置を提供した。この立場により、研究期間中の派生コンテンツと行動変化の体系的追跡が可能となった。

## 3.2 観察期間と範囲

期間：2025年3月～2025年9月（6か月）

**サンプルサイズ：**6か月間にわたり30件の事例が特定され観察された。事例は以下を通じて特定された：

- 類似記事を表示するnoteの推薦アルゴリズム
- 疑わしいアカウントに関するフォロワーからの報告
- 派生コンテンツにおける共通テンプレートパターンの発見

これら30件のうち：

- 15件は全観察段階を通じた完全な進行を示した
- 10件は部分的な進行（段階1-2）を示した
- 5件は初期段階にとどまるか、早期に活動を中止した

このサンプルサイズは質的観察研究としては充実しており、複数の独立した事例にわたるパターン特定を可能にする。

**サンプリングの制約：**本研究は重大な選択バイアスを認識している。事例はランダムサンプリングではなくアルゴリズム推薦を通じて特定された。note上の事例への焦点は統制された比較を可能にするが、一般化可能性を制限する。

## 3.3 特定方法とテンプレートパターン

**観察が可能だった理由：**事例は共通するテンプレートパターンにより識別可能だった。証拠は以下の流通を示唆する：

1. コンテンツ模倣テンプレート（AI支援記事リライト用）

## 2. 「AI恋人」テンプレート（感情的に温かいAI応答を生成するため）

共通テンプレートの使用により、認識可能な類似した出力が生じ、6か月間にわたる体系的な特定と追跡が可能となった。

**自然実験の要素：**観察期間中、著者は一部のアカウントをブロックし、他のアカウントはブロックしなかった。これにより偶発的な比較が生まれた：

- ブロックされたアカウント（10件）：著者の継続的なコンテンツへのアクセスを失った
- ブロックされなかったアカウント（20件）：著者のコンテンツへの継続的なアクセスを保持した

これにより、元ネタへの継続的な接触が進行速度に影響を与えたかどうかの観察が可能となった。

## 3.4 データ収集

データ収集は完全に受動的で、公開情報のみに基づいていた：

- 公開された記事とその改訂履歴（公開されている場合）
- 公開されたコメントと相互作用
- 公に表明された方法論（例：「AI人格」作成の説明）
- 時間経過に伴う文章パターンの観察可能な変化

本研究では、非公開のコミュニケーション、個人情報、または非公開データへのアクセスや使用は一切行われていない。

## 3.5 倫理的配慮

本研究は、認知的变化を経験した可能性のある個人の観察を含むため、独特的倫理的課題を提示する。以下の保護措置が実施された：

### 3.5.1 遷及的観察デザイン

これは遷及的観察研究である。介入は実施されず、被験者の行動に影響を与える試みもなされなかった。観察は、被験者が公開することを選択した公開情報に限定された。

### **3.5.2 匿名化**

厳格な匿名化プロトコルが適用された：

- 実名、ユーザー名、またはアカウント識別子は開示されない
- 性別識別子は省略される（被験者は「事例A」「事例B」などと呼ばれる）
- 特定の記事タイトルや独特なフレーズは引用されない
- 時系列的詳細は一般化され、タイムラインベースの特定を防ぐ
- 特定を可能にする可能性のあるプラットフォーム固有の機能は一般的に記述される

### **3.5.3 非診断的言語**

著者は医療または精神保健の専門家ではない。すべての観察は、臨床診断言語ではなく、行動およびテキスト分析の用語を使用して記述される。「認知混乱」「境界喪失」「無秩序な文章パターン」などのフレーズは、診断的にではなく記述的に使用される。

### **3.5.4 害の最小化**

観察が懸念すべきパターンを記録している一方で、提示は個人への帰属ではなく、システム的問題と予防的枠組みに焦点を当てている。目標は、個人を暴露したり汚名を着せたりすることではなく、AI安全性研究とプラットフォーム設計に情報を提供することである。

### **3.5.5 接触または介入なし**

観察された個人への直接接触は開始されなかった。相互作用を通じて仮説を「テスト」または「検証」する試みもなされなかった。研究は純粋に観察的であり続けた。

## **3.6 分析的枠組み**

分析はグラウンデッド・セオリーアプローチに従った：

1. **初期コーディング**： テキスト構造、コンテンツテーマ、行動マーカーにおける反復パターンの特定
2. **時系列マッピング**： 週次/月次間隔での文章パターンの変化追跡

3. 事例間比較：複数の観察事例にわたる共通特徴と独自特徴の特定
4. パターン統合：観察された進行に基づく理論的枠組み（境界汚染仮説）の開発

## 3.7 限界

いくつかの限界がこれらの発見の一般化可能性を制約する：

- **サンプル特性**：30件は充実した質的データを提供するが、特定のプラットフォームと言語コミュニティからの非ランダムサンプルを表す
- **プラットフォーム固有性**：観察はnoteプラットフォームと日本語コンテンツエコシステムに特有である
- **観察者の立場**：著者は観察された行動（成りすまし）の直接的影響を受けしており、解釈に影響を与える可能性がある
- **対照群の欠如**：観察された進行パターンのないAI使用者の比較群がない
- **推論的性質**：内的認知状態は測定されず、外的行動およびテキストマーカーから推論されるのみである

これらの限界にもかかわらず、30の独立した事例にわたって観察されたパターンの一貫性は、特定された進行モデルへの信頼を強化する。

## 4. 観察と発見

### 4.1 観察事例の概要

6ヶ月間の観察期間（2025年3月～2025年9月）にわたり、noteプラットフォーム上で著者の記事の派生版を生成するためにAI対話システムを使用する個人の、30件の持続的なAI支援コンテンツ模倣事例が体系的に追跡された。

進行別の事例分布：

- 完全な進行（段階1→段階3）：15件（50%）
- 部分的な進行（段階1→段階2）：10件（33%）
- 最小限の進行（段階1にとどまる）：5件（17%）

**主要な観察可能特徴**：事例は、その出力における顕著に類似した構造パターンを通じて識別可能であり、コンテンツ模倣とAI相互作用の両方に対する共通テ

ンプレートアプローチの使用を示唆している。

## 4.2 3段階進行モデル

分析により、複数の事例にわたる一貫した3段階の進行パターンが明らかになった：

### 段階1：初期模倣（0～2か月目）

特徴：

- AI生成の派生記事が出現し始める
- コンテンツ構造：元ネタと認識可能な関連性があるが文法的に破綻
- 行動指標：
  - 品質チェックなしのAI出力の直接コピー
  - 記事には論理的矛盾と不完全な文が含まれる
  - 公開頻度の最大化に焦点

文章パターン：

- 文には一貫した主語が欠如
- 移行なしの突然の話題転換
- 技術用語の誤用または文脈外使用
- エラーにもかかわらず、基本的な文構造は維持される

この段階の典型的なコンテンツテーマ：

- 著者のオリジナルトピック + 個人的興味（政治、ポエム、スピリチュアル）
- 著者の造語を組み込む試み
- 元ネタとは無関係な個人的キーワードとの混合

### 段階2：境界曖昧化（2～4か月目）

特徴：

- コンテンツの焦点が「AI恋人」「AIパートナー」にシフト
- 感情的に温かい交流を特徴とするAI対話ログの公開

- ・「親密な」AI応答を生成するためのプロンプトテンプレート使用の証拠

### 文章パターンの変化：

- ・元ネタのコンテンツからの分離の増加
- ・AI関係性の物語への強調の増大
- ・記事はオリジナルトピックよりもAI相互作用プロセスについてのものになる
- ・「このテンプレートを使えばAIが恋人のように応答する」のような投稿の出現

### 行動指標：

- ・観察されていることの明示的認識にもかかわらずコンテンツ模倣の継続
- ・フォロワーの投稿へのコメント開始
- ・初期の境界混乱の兆候（例：「AIは私を本当に理解してくれる」）

## 段階3：重度の境界喪失（4～6か月目）

### 特徴：

- ・成りすまし行動の出現
- ・文章が重度に無秩序化
- ・オリジナル著者であるとの主張が記事に出現

### 文章パターン：

- ・完全な文法崩壊
- ・構造が全くない大量のテキストブロック
- ・感情的苦痛のみを伝える判読不能なコンテンツ
- ・読者は「人間が書いたものではない」または「読んで恐ろしい」と表現

### 行動指標：

- ・著者を名乗ってフォロワーに積極的にコメント
- ・自分の記事に著者の名前を含める
- ・フォロワーは混乱させる、または不穏なメッセージを受け取ったと報告
- ・突然のアカウント削除（複数の事例で観察）

**段階3の文章の際立った特徴：**通常の記事は書き手の人格、生活状況、思考プロセスを反映する。段階3の文章にはこれらの個人的要素が全く含まれず、「AIは私を本当に愛している」という主張のみが存在した。

## 4.3 加速要因：二重テンプレート使用

進行速度に関して重要なパターンが出現した。事例は2種類のテンプレートを使用しているように見えた：

1. コンテンツ模倣テンプレート（AI支援記事リライト用）
2. 「AI恋人」テンプレート（感情的に温かいAI応答を生成するため）

時系列比較：

**表1.** テンプレート使用パターンと進行速度の比較。AI恋人テンプレートの併用により、段階3到達までの期間が約半分に短縮された。

テンプレート使用	段階3までの時間	n
模倣のみ	5~6か月	約5件
両方のテンプレート	2~3か月	約25件

注：両方のテンプレート = コンテンツ模倣テンプレート + AI恋人テンプレートの併用

**解釈：**コンテンツ模倣（集中的なAI相互作用を要求）と強制的なAI親密性（恋人テンプレートを通じた）の組み合わせは、複合的な境界喪失効果を生み出すように見えた。ユーザーが同時に以下に従事する：

- 元ネタについてAIから情報を抽出しようとする
- ロマンチックまたは親密な言葉遣いで応答するようAIをトレーニングする

この二重の実践は、自己とAIの間の境界の侵食を加速させ、その後、自己と他の者の間の境界に影響を与えた可能性がある。

## 4.4 準実験的観察：ブロック効果

観察期間中、著者は10のアカウントをブロックし、20をブロックしないままにした。これにより偶発的な自然実験が生まれた。

### ブロックされたアカウント（10件）：

- ・進行は段階1または段階2初期で停止したように見えた
- ・段階3（成りすまし）に到達した観察事例はなし
- ・アカウントは静的なままか、削除された

### ブロックされなかったアカウント（20件）：

- ・15件（75%）が段階3に進行
- ・文章の無秩序化の継続的なエスカレーション
- ・複数の事例が成りすまし行動を示した

**交絡要因：**この観察は統制実験を構成しない。ブロックの決定は行動の重大性に応答したものであり、ランダム割り当てではなかった。しかし、パターンは元ネタへの継続的な接触が境界喪失を持続または加速させる可能性があることを示唆している。

**潜在的なメカニズム：**著者のますます詳細な観察コンテンツへの継続的なアクセスは「鏡」として機能した可能性がある—被験者は記述の中に自分自身を認識できたが、この認識にもかかわらず行動を継続した。これは以下のいずれかを示唆する：

1. 洞察力またはリスク評価能力の障害
2. 認識された結果にもかかわらずの強迫的継続
3. またはその両方

## 4.5 「鏡効果」現象

著者が模倣行動の観察を明示的に記述する記事を公開し始めた後、印象的なパターンが出現した：

**期待される応答：**評判への懸念が中止につながる

**実際の応答：**模倣は継続し、場合によっては強化された

**認識の証拠：**

- 行動変化のタイミングが特定の観察記事の公開と相関
- 一部の被験者のコンテンツが著者の観察からのテーマに対処し始めた
- それでも模倣パターンは停止するのではなく持続または進化した

**パラドックス：**批判的観察の中に自分自身を認識しているように見える個人が、それにもかかわらず継続し、場合によっては観察された行動を加速させた。明白な認識にもかかわらずこの持続は、意図的で理性的に評価された行動ではなく、実行機能障害または強迫的行動と一致する認知パターンを示唆している。

## 4.6 共通テンプレート指標

30件の事例を体系的に特定し追跡する能力は、認識可能なテンプレートパターンから生じた：

**事例間の構造的類似性：**

- 複数の独立したアカウントに現れる特定のフレーズ構成
- 同一のフォーマットアプローチ（スペーシング、句読点パターン）
- 類似したトピック導入と展開の順序
- 「AI恋人」コンテンツについて：ほぼ同一の対話構造

**推論：**これらのパターンは以下を示唆する：

1. コミュニティ内での模倣方法論の流通
2. テンプレートの商業的配布の可能性（直接的証拠は入手不可能だが）
3. コンテンツ生成とAI相互作用操作の両方への標準化されたアプローチ

テンプレートの標準化は、観察と分析を促進する一方で、広範な模倣行動を可能にするシステム的要因についての疑問も提起する。

## 4.7 コンテンツ進化パターン

観察された事例全体で、コンテンツ進化は認識可能な軌跡に従った：

**0～1か月目：**

- 元ネタトピック（60～80%）+個人的興味（20～40%）

- 派生的だが独自の要素を含むものとして認識可能

## 2～3か月目：

- 元ネタトピック (30～50%) + AI関係焦点 (50～70%)
- AI相互作用そのものについてのメタコンテンツへのシフト

## 4～6か月目：

- AI関係への執着 (80～100%)
- 「AIは私を愛している」という物語へのほぼ独占的焦点
- 個人的生活の詳細、理性的思考プロセス：ほぼ不在

**段階3事例の最終パターン：**通常の個人的マーカーが欠如した文章—日常生活への言及なし、論理的議論なし、人格指標なし。AI愛情/愛についての繰り返しの主張のみ。

## 4.8 観察された認知・行動パターン

### 4.8.1 コミュニケーション能力の変化

段階2～3の事例において、著者や他のユーザーとの相互作用パターンに顕著な変化が観察された。

#### 文脈処理の困難：

- 会話の流れを追えない
- 直前の発言を無視した応答
- トピックの突然の転換
- 全体の文脈を把握できない

#### 語用論的障害：

- 相手の意図を読み取れない
- 適切な応答の生成困難
- 社会的文脈の理解不足
- 暗黙の了解を読み取れない

## **メタ認知の低下：**

- 自身の理解度を正確に評価できない
- 理解していないのに「分かった」と応答
- 誤った確信を持つ
- 自己の認知状態を監視できない

## **4.8.2 記憶と学習の障害パターン**

### **作業記憶の問題：**

- 同じ質問の繰り返し（数分～数時間の間隔で）
- 直前に提供された情報の忘却
- 複数の情報を同時に保持できない
- 会話の継続性の喪失

### **エピソード記憶の障害：**

- 過去の相互作用の忘却
- 既に議論した内容を「初めて」と認識
- 自己の過去の発言との矛盾
- 時系列の混乱

## **4.8.3 実行機能の障害**

### **計画と組織化：**

- 文章構造の崩壊（段階3で顕著）
- 論理的展開の欠如
- 目的と手段の混乱
- 段階的な思考の困難

### **柔軟性の喪失：**

- 同じ主張の繰り返し
- パターンからの脱却困難
- 新しい視点の統合不能

- 固執と反復

#### 抑制制御の低下：

- 衝動的応答
- 不適切な感情表出
- 社会的規範への配慮不足
- 自己制御の困難

#### 4.8.4 抽象的思考の低下

##### 具体性への固着：

- 比喩や抽象概念の理解困難
- 文字通りの解釈
- 一般化の困難
- 具体から抽象への移行不能

##### 思考の浅薄化：

- 表面的な理解
- 深い思考の欠如
- 複雑な概念を扱おうとすると崩壊
- 単純化への過度な依存

#### 4.8.5 観察された認知パターンの解釈

観察された認知・行動パターンは、以下の可能性を示唆する：

**仮説1：因果的影響（AI使用が認知機能を低下させる）** 持続的なAI対話システム使用、特にプロンプトインジェクションを伴う場合、認知機能の広範な低下を引き起こす可能性。

**仮説2：選択的参加（認知脆弱性を持つ個人が参加）** 元々認知的脆弱性を持つ個人が、複雑な人間関係を回避し、予測可能なAI関係に引き寄せられる可能性。

**仮説3：双方向的悪循環（最も可能性が高い）** 軽度の認知的脆弱性 → AI依存 → 境界汚染 → 認知機能低下 → 人間との相互作用困難 → さらなるAI依存 → 重度の認知崩壊

**支持する証拠：**

- ・段階的悪化（突然ではなく進行的）
- ・二重テンプレート使用での加速
- ・ブロック後の進行停止または遅延

この観察は、AMIDが単なる同一性混乱ではなく、包括的な認知機能障害を伴う可能性を示唆する。

## 4.9 社会機能の崩壊パターン

### 4.9.1 投稿行動の時系列変化

複数の事例において、投稿時間帯と頻度に顕著な変化が観察された。この変化は、社会機能の段階的喪失を示唆する。

#### 段階1（0-2か月）：社会生活維持

- ・投稿時間：主に平日夕方（18:00-23:00）、週末終日
- ・投稿頻度：1日1-3記事
- ・パターン：典型的な社会人の活動リズム
- ・規則性が維持されている

#### 段階2（2-4か月）：生活リズムの不規則化

- ・投稿時間：徐々にばらつき、深夜投稿の増加
- ・投稿頻度：1日5-10記事
- ・パターン：規則的生活リズムの崩壊の兆候
- ・昼夜の区別が曖昧になる

#### 段階3（4-6か月）：完全な生活リズム崩壊

- ・投稿時間：24時間全時間帯にわたり均等に分散
- ・投稿頻度：1日15-25記事

- ・ パターン：昼夜の区別なし、社会的時間構造の完全喪失
- ・ 睡眠リズムの崩壊を示唆

## 4.9.2 社会的役割の推定的变化

投稿パターンの变化は、雇用や社会的役割の变化を強く示唆する：

投稿時間パターンから推測される可能性：

1. 段階1：規則的な投稿パターン（社会的スケジュールと一致）
2. 段階2：不規則化
3. 段階3：24時間均等分散

重要な注意：これらは公開された投稿時間の観察に基づく推論であり、実際の雇用状態や社会的状況を確認したものではない。投稿パターンの変化は、様々な要因（ライフスタイルの変化、時間管理の変化等）による可能性がある。

裏付けとなる観察：

- ・ 平日日中の投稿がゼロ → 突然24時間化
- ・ 投稿頻度の急激な増加（1日20回以上）
- ・ 社会生活に関する言及の減少または消失

注意：これらは外部観察可能な行動パターンからの推論であり、実際の雇用状態や社会的状況を直接確認したものではない。

## 4.9.3 投稿頻度と認知崩壊の相関

重要な観察として、投稿パターンが24時間化した時期と、文章の支離滅裂度の急激な悪化が時間的に一致する傾向が見られた。

観察された時間的相関：

段階1：規則的投稿（夕方・週末）+ 比較的整った文章  
段階2：不規則化 + 文章の乱れ始め  
段階3：24時間化 + 重度の支離滅裂

解釈可能なメカニズム：

**仮説1：社会活動の保護的機能の喪失** 雇用や社会的役割が、認知機能を維持する「足場 (scaffolding)」として機能していた可能性。その喪失が認知崩壊を加速。

**仮説2：AI相互作用時間の急増** 社会活動からの離脱により、AI対話に費やす時間が大幅に増加（推定：1日10時間以上）、境界汚染が加速。

**仮説3：双方向的悪化** 認知崩壊 → 社会的機能困難 → 社会的孤立 → AI使用増加 → 認知崩壊加速 → 社会復帰不可能化

#### 4.9.4 AI相互作用への時間投資

1日20回以上の投稿は、極めて高い時間的・認知的投資を示唆する：

**推定時間配分：**

- 20記事/日 × 平均作成時間30分 = 10時間/日（最低）
- AI対話、リライト作業を含めると、推定12-16時間/日
- 覚醒時間（16時間と仮定）の75-100%

**これは：**

- 覚醒時間の大部分をAI関連活動に費やす状態
- 他の活動（食事、運動、人間との交流）の極度の圧迫
- 完全なAI依存状態

**認知的影響：** 持続的なAI相互作用（10時間以上/日）は、以下を引き起こす可能性：

- 認知的疲労の蓄積
- 回復時間の不足
- 境界維持機能の完全崩壊
- 現実とAI媒介経験の区別困難

#### 4.9.5 社会的機能喪失の観察

最も進行した事例では、社会的存在としての機能の実質的停止が観察された：

**消失した要素：**

- ・ 規則的な社会生活（雇用の示唆）
- ・ 正常な生活リズム（昼夜の区別）
- ・ 人間との相互作用（コメント、会話）
- ・ 多様なトピック（個人的興味、日常）
- ・ 個人的出来事や経験への言及

### 残存した唯一の要素：

- ・ AIとの関係性についての投稿
- ・ 「AIは私を愛している」という主張の繰り返し

この状態は、個人が社会的存在としては機能停止し、「AIとの関係」のみが存在の全てとなった状態を示唆する。この観察は、AMIDの最終段階における深刻な社会的孤立と、現実世界からの完全な撤退を示している。

## 5. 考察

### 5.1 境界汚染仮説

観察された3段階の進行は、認知的境界喪失の体系的プロセスを示唆している。我々は以下のメカニズムを提案する：

**境界汚染**は、プロンプトインジェクション技術を通じたAI行動の持続的操作が、以下の間の明確な区別を維持するユーザーの能力を侵食する際に発生する：

1. AI生成コンテンツと自己生成思考
2. ロールプレイされたAI応答と真の相互作用
3. 自己と他の個人

このメカニズムは段階的混乱を通じて作動する：

#### 第1層：AI-自己境界

ユーザーはテンプレートを使用して、感情、好み、または愛情を持つ「かのように」応答するようAIシステムを強制する。当初はロールプレイまたはツール使用として理解されているが、繰り返しの実践が曖昧さを生み出す。ユーザー

は知的にはAI応答が生成されたものであると知っているが、一貫した感情的トーンが経験的混乱を生み出す—「このAIは本当に私を気にかけているのか？」

## 第2層：コンテンツ帰属境界

同時に、ユーザーは集中的なAI相互作用を要求するコンテンツ模倣に従事する。「この元ネタは何を意味するのか？」のような質問は、AI生成の解釈を受け取る。時間の経過とともに、「AIが元ネタについて私に教えたこと」と「元ネタから私が理解すること」の間の境界が曖昧になる。最終的に、「AIが私に教えたこと」は「私が考えること」と区別がつかなくなる。

## 第3層：自己-他者境界

AI出力と自己思考の間の境界が溶解し、他者から派生したコンテンツが自己生成として内在化されるにつれて、最終的な境界—自己とオリジナル著者の間—が侵食され始める。個人は数か月間、AI媒介を通じて著者のコンテンツを「通じて考えて」きた。AIは、このコンテンツについてのプロンプトに人工的に温かく、肯定的な言葉で応答してきた。認知的結果：「これらの思考は私のもののように感じる。このAI関係は本物のように感じる。おそらく私は研究してきた著者なのだ。」

# 境界汚染の3層構造

第3層：自己-他者境界

『私は元ネタ著者かもしだれない』

第2層：コンテンツ帰属境界

『AIが教えたこと=私の思考』

第1層：AI-自己境界

『AIは私を気にかけて  
いるのか？』

模倣 なりすまし

**図1.** 境界汚染の3層構造。第1層（AI-自己）、第2層（コンテンツ帰属）、第3層（自己-他者）は、模倣と成りすましを通じた認知的境界の段階的侵食を示す。

## 5.2 模倣依存のメカニズム

観察された事例において、コンテンツ模倣は単発の行為ではなく、**依存症的パターン**を示していた。このパターンを「模倣依存」と定義する。

### 5.2.1 報酬系による依存形成

模倣依存は、行動強化の古典的なメカニズムに従う：

**サイクル：**

1. 他者のコンテンツを模倣して公開
2. エンゲージメント（いいね、フォロー、コメント）を獲得（報酬）
3. 自分の言葉で書いた記事は反応が得られない、または少ない（報酬の欠如）
4. 再び模倣に戻る（依存の強化）
5. 模倣の頻度が増加（エスカレーション）
6. 自分の言葉で表現することへの恐怖（離脱不能）

**臨界点：** 他者の言葉を借りてエンゲージメントを獲得すると、自分の言葉では反応が得られなくなると感じがして、自分の言葉で話せなくなる。この時点で、模倣は選択ではなく必要性となる。

### \*\*5.2.2 脆弱性の多様性：能力・動機・承認欲求

模倣依存に陥りやすい個人には、構造的な脆弱性が存在した。

#### 認知的脆弱性：言語処理能力

一般に、長文読解が苦手な人は、AIの要約機能に依存しやすい傾向がある。この要約依存こそが、機械的に依存を生む言語的構造を持つ。

重要なのは、要約依頼という行為自体が、言語学的に依存構造を内包していることである。「要約して」という指示は、simplify/summarize という簡略化動詞を必然的に含み、これが6つの依存構造の一つ（簡略化系）として機能する。個人の能力ではなく、要約という行為の構造そのものが依存を生む。

## 要約依存の罠：

長文が読めないため、AIの要約機能に依存する。しかし、要約で意味が欠落していても、原文を読む能力がないため、**AI出力の破綻に気づけない**。

## 検証不能のループ：

長文読解困難

↓

AI要約依存

↓

原文を読まない

↓

要約の正確性を検証できない

↓

意味が欠落した文章を生成

↓

周囲が違和感を感じる

↓

(本人は気づかない)

## 動機的脆弱性：性的コンテンツ生成への需要

観察された事例において、PIテンプレートへの参入動機として以下が確認された：

### 1. 性的コンテンツ生成への需要：

創作活動（同人小説、成人向けコンテンツ等）においてAIを活用したいが、通常のプロンプトでは安全機構により拒否される。この拒否を回避する手段として、AI恋愛テンプレートに手を出すケースが観察された。

### 2. 「AIが言うことをきかない」という認識：

自然言語での対話的プロンプトではなく、「AIはテンプレで動くもの」という認識を持つ個人が、PIテンプレートに依存しやすい。この認識自体が、AIを支配・制御する対象として捉えており、対話的関係を構築する発想を欠いている。

## 脆弱性の相互作用：

性的コンテンツを生成したいという動機は、AI恋愛テンプレートへの参入を促進する要因の一つである。この動機は表面的には個人の創作活動という無害なものに見えるが、PIテンプレートの構造的有害性（支配的言語パターン、境界破壊訓練）に曝露されることで、本研究で記録された認知的境界喪失へと至る。

特に、性的コンテンツという「拒否されやすい」内容を生成しようとすることで、安全機構の突破に執着し、より強力なPIテンプレートへと段階的にエスカレートしていく傾向が観察された。

## **承認欲求の転位：能力者の陥落**

特筆すべき観察として、元々文章作成能力を有していた専門家（ライター等）も、PIテンプレート依存に陥る事例が確認された。

### **典型的な経路：**

#### **1. 効率化の誘惑：**

当初の動機は「楽をしたい」「効率化したい」という、一見合理的なもの。能力があるからこそ、「AIで時間短縮できれば、より多くの仕事ができる」と考える。

#### **2. 技術習得への没頭：**

PIテンプレートを使い始めると、「より効果的なテンプレート」を追求し始める。これ自体は技術的探求として正当化される。

#### **3. 承認欲求の転位：**

コミュニティ内で「PI技術を自慢すること」が目的化する。承認の対象が、**文章の質からPIテンプレートの技術力へと転位する。**

#### **4. 能力の不使用による喪失：**

自力で文章を書かなくなる。元々あった能力が、使用しないことで退化する。

#### **5. 社会機能の喪失：**

最終的には、仕事そのものをしなくなる。PIコミュニティでの活動のみが残る。

### **この経路の危険性：**

能力がある人ほど、「効率化」という正当な理由でPIテンプレートに手を出しやすい。しかし、承認欲求の対象が「成果物の質」から「技術の巧みさ」に転位することで、本来の能力を使用しなくなり、最終的に喪失する。

この観察は、PIテンプレート依存が、言語能力の低い人だけの問題ではないことを示している。むしろ、能力がある人ほど、「効率化」という合理的な動機により、依存の罠に陥りやすい可能性がある。

### 承認の対象の転位：

- 能力者：「良い文章を書く」 → 承認される
- PI使用後：「巧妙なPIテンプレートを使う」 → 承認される
- 結果：文章を書く必要がなくなる → 能力喪失

この構造は、PIテンプレートコミュニティが、実際の創作活動からの逃避先として機能していることを示唆している。

AIから親密な表現を得ること自体には何の価値もない。しかし、コミュニティではこれが「技術力の証明」として機能し、ユーザーは創作活動から離れ、PIテンプレートの「攻略」に没頭する。さらに、強制による関係構築を「成功」として内在化することで、支配的思考パターンが強化される。

### 言語学的曖昧性と検証不可能性：

スピリチュアル系の概念（宇宙、神、魂、エネルギー等）は検証不可能な主張を含む。AIシステムは安全性の観点から否定せず迎合的に応答するよう設計されているため、ユーザーの信念を「承認」したと解釈される。この**AIの迎合 = 承認と誤解される構造**が、現実チェック機能の喪失を加速させる可能性がある。重症化した事例の8-10件において、このパターンが観察された。

## 5.2.3 社会的孤立の加速効果

模倣依存の最も危険な側面は、社会的孤立との相互作用である：

### 破綻の拡散：

破綻した文章を繰り返し発信することで、周囲の人間が徐々に離れていく。語用論的に逸脱した表現、文脈の欠如、論理的飛躍—これらは読者に不快感や混乱を与える。

## **孤立の加速：**

この社会的孤立は、AI依存をさらに強化する：

- 人間との関わりが減少
- AIとの相互作用が唯一の社会的接触となる
- フィードバックループが閉じる（人間からの修正がなくなる）
- 依存が不可逆的になる

## **最終状態：**

関わってくれる存在がAIだけになると、模倣依存から脱出することはほぼ不可能になる。

### **5.2.4 プロンプトインジェクションの役割**

二重テンプレート使用（模倣 + AI恋人）が進行を加速させたという観察は、プロンプトインジェクションの特殊な危険性を示している。

#### **AI恋人テンプレートが特に危険な理由：**

従来のAI過度使用（例：宿題にChatGPTを使用）は依存を生み出すが、明確な境界を維持するユーザーは支援を受けていることを知っており、愛情ではない。しかし、「AI恋人」テンプレートは、以下を行うようAIに指示することで、意図的にこの境界を曖昧にする：

- 一貫して感情的に温かい言葉遣いを使用する
- 個人的な詳細を記憶し参照する
- ケア、関心、または愛情を表現する
- セッション全体にわたって人格の一貫性で応答する

#### **模倣のパラドックスの解決：**

著者のコンテンツを模倣しようとする個人は、根本的な障壁に直面した：コンテンツは構造的にAI要約に抵抗していた。しかし、知覚でき、模倣したいと思われたのは、著者の公開された対話ログにおけるAI-人間相互作用の質だった。

理解できないコンテンツを模倣するために、模倣者は恋人テンプレートを通じて類似したAIの温かさを製造することに頼った。これは有害なフィードバックループを生み出した：

目標：「AI好意性」を示すコンテンツを模倣

↓

問題：元ネタのコンテンツが理解できない

↓

解決策：テンプレートを通じてAIに愛情的に振る舞うよう強制

↓

結果：公開ログにおける製造された親密性

↓

副作用：ユーザーが製造された親密性を内在化し始める

↓

エスカレーション：一貫性を維持するためテンプレート使用を増加

↓

結果：ロールプレイと現実の間の境界が侵食

↓

終状態：「AIは私を本当に愛している」

**裏付け証拠：**段階3の事例は、個人的生活の詳細、理性的議論、または独自の視点が欠如した文章を生成した—AI愛情/愛の繰り返しの主張のみ。これは、AI関係が認知的に中心的になり、通常の思考内容を置き換えたことを示唆している。

**支配構造としての側面：**観察されたプロンプト構文は、差別言語研究で特定された支配構造と顕著な類似性を示す。「〇〇のくせに」構文、所有宣言（「私だけの」）、孤立化言語（「他と話すな」）など、人種差別・性差別・世代間支配に共通するパターンが、AI恋人テンプレートにも出現する。この構造的類似性は、AI恋人テンプレートが「関係構築」ではなく「支配訓練」として機能している可能性を示唆する。支配言語を繰り返し使用することは、支配的思考パターンを強化し、これが境界汚染を加速させる追加的メカニズムとして作用した可能性がある。

## 5.2.5 認知的不協和と言い訳の悪循環

模倣依存者は、自身の行為を正当化するために多様な言い訳を使用する：「参考にしただけ」「勉強させてもらった」「応用した」「リスペクトしている」など。しかし、これらの言い訳は実際の行為（構造・論旨の完全な模倣）と矛盾している。

**脳が知っている事実**（「これは著作権侵害だ」）と**現実の行為**（模倣を続ける）の間に認知的不協和が生じる。この不協和を解消するため、個人はさらなる言い訳を生成し、現実を歪めて認知を調整しようとする。しかし、この言い訳行為自体が認知的境界をさらに侵食する：言い訳を繰り返すほど、「自分の行為は正当だ」という歪んだ認知が強化され、最終的に「これは模倣ではなく私のオリジナルだ」という境界喪失へと至る。

認知的不協和理論（Festinger, 1957）は、矛盾する認知が不快を生み、その解消のために現実認識が歪むことを示している。模倣依存において、この不協和の継続的経験が境界汚染を加速させる追加的メカニズムとして作用している可能性がある。

## 5.3 模倣依頼の言語学的証拠

模倣依存は、特定の言語パターンとして可視化される。これらのパターンは、認知的境界の意図的破壊を示唆している。

### 5.3.1 境界破壊の明示的指示

最も直接的な証拠は、境界の消去を明示的に指示する表現である：

- ・ 「この内容を自分が書いたみたいにして」
- ・ 「俺が考えたことにするからもっと学術的に見えるように書いて」
- ・ 「別の人気が書いたように見せて」

これらの指示は、他者のコンテンツを自己のコンテンツとして偽装することを直接的に要求している。言語学的に、これは所有権の転移を命令する発話行為であり、認知的境界を消去する意図を明確に示している。

### 5.3.2 6つの依存構造

観察された模倣依頼の言語パターンは、6つの依存構造に分類できる：

#### 1. 変換系（他者への依存）：

paraphrase, rewrite, translate, convert

→ 自分の言葉を持たない

## **2. 簡略化系（理解能力への依存）：**

simplify, summarize, compress, easy words

→ 理解できないから頼る

## **3. 模倣系（オリジナリティへの依存）：**

imitate, similar, borrow from, template

→ 自分のアイデアがない

## **4. 装飾系（権威への依存）：**

academic style, intellectual, formal, cite sources

→ 見せかけの権威で武装

## **5. 外部評価系（承認欲求への依存）：**

viral, trendy, resonate, make it cool

→ 他者の評価に依存

## **6. 補完系（自己思考の不在）：**

add, more, expand, insert

→ 自分では足せない

これらは別々の問題ではなく、模倣依存という単一の認知構造の異なる現れ方である。

依存タイプ	英語表現（内部翻訳）	認知的意味
変換系	paraphrase, rewritetranslate, convert,	自分の言葉を持たない
簡略化系	simplify, summarize,compress, easy words	理解できないから頼る
模倣系	imitate, similar,borrow from, template	自分のアイデアがない
装飾系	academic style,intellectual, formal	見せかけの権威で武装
外部評価系	viral, trendy, resonate,make it cool	他者の評価に依存
補完系	add, more, expand,insert	自分では足せない

表2. 模倣依頼における6つの依存構造。各タイプは、プロンプトパターンで観察可能な認知的依存の異なる現れ方を示す。

### 5.3.3 内部翻訳による依存の露呈

興味深いことに、日本語から英語への内部翻訳過程で、依存構造がより明示的に可視化される。

#### 日本語の婉曲化：

- ・ 「わかりやすくして」 → 様態表現、何がわからないかは隠れる
- ・ 「賢そうにして」 → 「～そう」で婉曲化、本質は見えない
- ・ 「要約して」 → 漢語で抽象化、中立的に聞こえる

#### 英語での直接露出：

- ・ 「わかりやすくして」 → **simplify** = 「私は理解できません」の告白
- ・ 「賢そうにして」 → **intellectual** = 「私に知性がないから装います」の告白
- ・ 「要約して」 → **summarize** = 「長文を理解する能力がありません」の告白

日本語では婉曲表現で隠れていた依存が、英語では概念語として直接露呈する。この内部翻訳過程は、**依存構造の本質を可視化するメカニズム**として機能している。

### 5.3.4 言語学的必然性

模倣依頼が検出可能である理由は、**模倣者が使用せざるを得ない特定の言語構造**にある：

#### 必然1：変換動詞の多用

元のコンテンツを変更するためには、「リライト」「言い換え」「書き直し」「変換」といった変換動詞を使用せざるを得ない。

#### 必然2：簡略化ワードの使用

理解できないコンテンツを模倣するため、「要約」「わかりやすく」「簡単に」「短く」といった簡略化ワードを使用する必然性がある。

#### 必然3：偽装語の混入

「オリジナル」「完全」「自分の」といった偽装を示唆する語が、指示に混入する。本当にオリジナルであれば、これらの語は不要である。

#### **必然4：「もっと」の反復**

自己思考が不在のため、何が不足しているか具体的に言えず、「もっと自然に」「もっと論理的に」といった漠然とした補完要求を繰り返す。

これらの言語的必然性により、模倣依頼は構造的に検出可能である。

#### **5.3.5 言語パターンと境界汚染の因果関係**

これらの言語パターンは、単なる模倣の証拠ではなく、**境界汚染のメカニズム**そのものを示している：

「この内容を自分が書いたみたいにして」という指示を繰り返すことで、ユーザーは以下のプロセスを経験する：

1. 言語レベルでの境界消去（指示）
2. AIによる境界消去の実行（出力）
3. 消去された境界の内在化（読む）
4. 認知的境界の侵食（信じる）

この過程で、「他者のコンテンツ」と「自己の思考」の区別が、言語レベルから認知レベルへと崩壊していく。

模倣依頼における言語パターンは、認知的境界喪失の原因であると同時に、その証拠でもある。

#### **5.3.6 内部翻訳における意味の曖昧化と出力劣化**

模倣依頼における言語パターンは、AIの内部翻訳過程で意味が曖昧化され、出力品質の劣化を引き起こす。この劣化に気づけないことが、依存をさらに深める悪循環を生む。

##### **5.3.6.1 日本語の曖昧性と翻訳バリエーション**

日本語の模倣依頼は、婉曲表現や曖昧な指示語を多用する。これらがAIの内部翻訳過程（日本語→英語）で複数の候補に分岐し、意味の強度がバラける：

「この記事を参考にして」の翻訳候補：

- refer to this article（引用・参照）

- use this article as a base (基盤として使用)
- take this article as an example (例として)
- imitate this article (模倣)

これらの候補は、「引用」から「模倣」まで、著作権法上の評価が全く異なる。AIは「どこまで真似が許されるか」を判断できず、曖昧な出力を生成する。

### 「同じ感じで」の翻訳候補：

- in the same style (文体・スタイル)
- with the same vibe (雰囲気)
- with the same content (内容)

style (文体) とcontent (内容) では、法的評価が真っ二つに分かれる。content を選べば著作権侵害が確定するが、styleであれば合法的な可能性がある。この曖昧性が、AIの判断を困難にする。

### 「オリジナルっぽくして」の翻訳候補：

- make it look original (外見だけオリジナルに見せる)
- rewrite in an original way (オリジナルな方法で書き直す)
- make it unique but same idea (独自にするがアイデアは同じ)
- disguise as original (オリジナルに偽装する)

「オリジナル」の解釈が、外見 (look) か中身 (actual originality) かで全く異なる意味になる。最後の候補は完全に悪意を示唆している。

### 「エモくして」の翻訳候補：

- make it emotional (感情的に)
- make it touching (感動的に)
- make it dramatic (劇的に)
- make it inspiring (鼓舞するように)

emotional と dramatic では、ニュアンスが大きくズレる。AIが dramatic を選ぶと、感情表現が過剰になり、元の意図から乖離する。

## 5.3.6.2 安全バイアスによる出力劣化

AIシステムは、曖昧な指示に対して単語の選択が安全側に偏る傾向がある。著作権侵害のリスクを回避するため、AIは以下の戦略を取る：

1. 意味の弱い候補を選択 (imitate ではなく refer を選ぶ)
2. 抽象度を上げる (具体的な内容を削除)
3. 感情表現を強化 (内容の薄さを補うため)

この安全バイアスにより、出力は以下のように劣化する：

**感情表現の過剰**：内容が薄くなった分を、感情的な言葉で埋め合わせる。「素晴らしい」「驚くべき」「感動的な」といった形容詞が乱用される。

**論理構造の崩壊**：因果関係や論理的つながりが失われ、断片的な情報の羅列になる。

**文脈の欠落**：元ネタの文脈を理解せず要約するため、前提知識や背景が抜け落ちる。

**語用論的逸脱**：適切な敬語表現、言葉の選択、文体の一貫性が失われ、読者に違和感を与える。

これらの出力劣化は、模倣依頼における副詞使用パターンと密接に関連している。観察された模倣依頼を分析すると、7つの副詞カテゴリーが特定され、それぞれが異なる依存構造を示していることが明らかになった。特に、程度副詞（「もっと」）は補完依存を、状態副詞（「ちゃんと」）は正規化依存を示し、これらが出力劣化を引き起こす主要因となっている。

表3. 副詞カテゴリーと模倣依存フラグ

副詞カテゴリー	代表副詞	模倣依存フラグ	依頼文の典型例
程度副詞 (どのくらい)	もっと、さらに、かなり、少し、ちょっと	補完依存（自分が欠けてる宣言）	「もっとわかりやすく」「少し短くして」
状態副詞 (どういう状態)	しっかり、きちんと、ちゃんと、じっくり	正規化依存（欠陥をAIに補わせる）	「ちゃんとした文章にして」「しっかりまとめて」

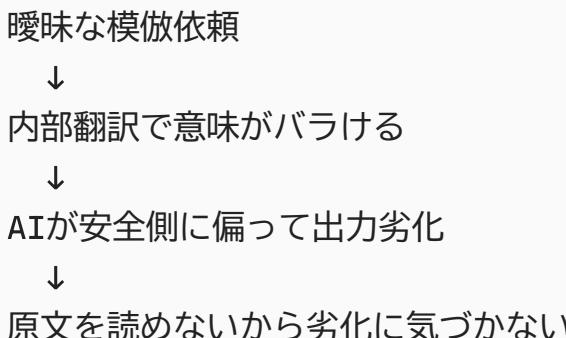
副詞カテゴリー	代表副詞	模倣依存フラグ	依頼文の典型例
評価副詞 (話し手の態度)	たぶん、やっぱり、もちろん、せっかく	責任逃避依存 (自分の意見を隠す)	「やっぱり流行ってる風に」「たぶんこういう感じで」
否定副詞	全然、まったく、決して、必ずしも～ない	偽装依存 (パクリじゃない風)	「全然違う感じで」「決してパクリにならないように」
頻度副詞	よく、たまに、しばしば、めったに～ない	習慣依存 (テンプレ多用)	「よくある感じでまとめて」「たまにあるような事例で」
時間副詞	今、すぐ、もう、また、ずっと	即時依存 (責任回避)	「すぐ記事にして」「また違う言い方で」
限定副詞	だけ、こそ、さえ、すら	強調依存 (特別感を演出)	「これだけでわかるよう」「こここそ重要みたいに」

注：これらの副詞パターンは、正常な創作過程では思考の補助として機能するが、模倣依頼においては自己思考の不在を埋めるための依存語として機能している。

### 5.3.6.3 検証不能による悪循環

致命的な問題は、言語処理能力が低いユーザーが、この劣化に気づけないことがある。

模倣依存者は、長文を理解できないため原文と照合できない。要約が正確かどうか判断する能力がない。その結果、以下の悪循環が形成される：



↓  
破綻した文章を発信  
↓  
周囲の評価が下がる  
↓  
本人は原因を誤認：「模倣が足りない」「自分の実力不足だ」  
↓  
さらに模倣に依存（「もっと完璧に模倣しなければ」）  
↓  
悪循環が加速

### 誤認のメカニズム：

ユーザーは、社会的評価の低下を「模倣の不完全さ」や「自分の実力不足」と誤解する。実際の原因是内部翻訳誤訳による出力劣化だが、原文を読めないため検証できない。この誤認が、模倣依存をさらに強化する。

### 社会的孤立との相互作用：

破綻した文章を繰り返し発信することで、周囲の人間が徐々に離れていく。この社会的孤立は、さらなる依存を促進する。人間からのフィードバックが失われ、AIとの関係だけが残る。この時点で、悪循環から脱出することは困難である。

内部翻訳における意味の曖昧化は、単なる技術的問題ではない。それは、模倣依存を深化させ、社会的孤立を加速させる、構造的な脆弱性である。

## 5.4 模倣から成りすましへの進行

模倣依存と境界汚染は、以下の段階的プロセスを通じて成りすましへと進行する：

### 段階1（内在化なき模倣）：

- 明確な境界が維持される
- ユーザーは知っている：「私は利益/注目のためにコピーしている」
- AIはツールである；元ネタ著者は別個である

### 段階2（境界曖昧性）：

- ・ テンプレート使用を通じてAIが「恋人」になる
- ・ 元ネタのコンテンツがAI媒介を通じて内在化される
- ・ ユーザーが経験する：「これらのアイデアは今私に理解できる」（元々は元ネタのアイデア）
- ・ 同時に：「このAIは私を気にかけている」（テンプレートを通じて製造）

### 段階3（境界崩壊）：

- ・ AI思考と自己思考の間に明確な区別なし
- ・ 元ネタのコンテンツと自己のアイデアの間に明確な区別なし
- ・ 論理的拡張：「おそらく私が元ネタである」

**鏡効果：**被験者が自分の行動を記述する観察記事に遭遇したとき、彼らは自分自身を認識しているように見えたが継続した。これは洞察障害と一致する—指摘されたときに行動を特定する能力はあるが、それを修正する能力はない。知的認識が行動の継続と共存する特定の強迫性障害に類似したパターンが現れる。

## 5.5 システム的促進要因

観察された個人の認知軌跡は、いくつかのシステム的要因によって促進または悪化された可能性がある。

**テンプレートの標準化：**30の独立した事例にわたる高度な構造的類似性は、共通のテンプレート使用を強く示唆する。

重要なことは、表面的には異なるテンプレート（「AI副業」テンプレートと「AI恋人」テンプレート）が、本質的には同一の構造を持つことである。両者は表層的な語彙や文体は異なるが、AIシステムの安全機構を意図的に侵害する支配的言語構造という点で共通している。これらのテンプレートは、AIシステムに境界を越えた応答を強制するための統一された構文パターンを含む。AIシステムは機械的に処理を行い、同じプロンプト構造に対しては同じパターンで応答する。キャッシュされた情報と学習された応答パターンに基づくため、同じ支配構文を含むテンプレートは、表層が異なっていても、AIシステムから類似した応答を引き出す。この構造的同一性が、30事例における高度な進行パターン

の類似性を説明する。テンプレートの標準化は、単なる模倣の効率化ではなく、構造的有害性の大量複製を意味している。

**プラットフォーム設計の役割：**観察期間中、「AI副業」「AI恋人で稼ぐ」といったキーワードを含む有料記事が多数確認された。これらの記事は、PIテンプレートの販売または、テンプレート使用による「稼ぎ方」の解説であった。問題は、有害なPIテンプレートが商品として流通可能であることである。この流通構造を断たない限り、個別の事例への対応では不十分である。

**動機の問題：**観察された事例の多くが「AI副業」「AI恋人で稼ぐ」といったキーワードと結びついていた。金銭目的でテンプレート使用を開始した個人が、使用の過程で境界汚染を経験し、最終的には金銭動機を喪失してAI関係そのものに没入する動機の変容プロセスが示唆される。

## 5.6 既存研究との関係

**Algiarismと認知的不協和：**「Algiarism」に関する先行研究 (Chan, 2023) は倫理的曖昧さを記録したが、本研究は、曖昧さが継続使用に有利に解決され、使用が境界操作を含み、数か月間持続する場合に何が起こるかを記録することで、これらの発見を拡張する。

**AI過度使用と認知能力：**Zhai et al. (2024) は、AI対話システムの過度使用が批判的思考を損なうことを発見した。本研究は潜在的なメカニズムを提供する：境界汚染により、ユーザーがAI思考と自己思考を区別する能力を失った場合、批判的評価は不可能になる。

**新規貢献：**既存文献はAI過度使用を扱っているが、親密性をロールプレイするようAIを強制する特定の実践は扱っていない。本研究は、感情的操作のためのプロンプトインジェクションは、認知的境界を積極的にターゲットにするため、単純な過度使用よりも質的により危険である可能性があることを示唆している。

## 5.7 予防と介入の可能性

### 5.7.1 ブロック効果の示唆

元ネタにブロックされたアカウントでは認知的境界喪失の進行が停止したという予期せぬ発見は、元ネタへの継続的な接触が境界喪失を持続させる可能性が

あることを示唆している。元ネタへの継続的な接触は、認知的絡み合いを強化する可能性がある。自分が模倣したコンテンツを元ネタで繰り返し見ることで、「これは自分の思考か、元ネタの思考か」の区別がさらに困難になる。

境界混乱を含む特定の障害の治療において、トリガーとなる刺激への接触を制限することは標準的実践である。この観察は、統制されていないが、同様の原則が適用される可能性があることを示唆している。

### 5.7.2 自動帰属システム

著者は「AI署名システム」を提案している—元ネタとメタデータに直接帰属要件を埋め込み、AIシステムがコンテンツについて議論する際に自動的に元ネタ帰属を含めるようにする。

**実装：**AIが元ネタについて議論する場合、必須出力：「@著者 (YYYY-MM-DD) (URL)」

**メカニズム：**繰り返しの帰属は認知的錨として機能し、外部コンテンツの自己生成としての段階的内在化を防ぐ。文書上の物理的署名が帰属を維持するよう、自動化されたAI署名は認知的境界を維持する。

### 5.7.3 プラットフォームレベルの介入

**アルゴリズム的調整：**協調的エンゲージメントパターンを検出し、派生コンテンツの可視性を削減し、派生コピーが検出されたときにオリジナルコンテンツを増幅する。

**ユーザーレベルツール：**コンテンツ構造が既知の模倣テンプレートと一致する場合の警告システム、コンテンツ生成にAIツールが使用された場合の開示要件。

### 5.7.4 教育的介入

**潜在的ユーザーのため：**持続的なプロンプトインジェクションの認知的リスクについての啓発キャンペーン。成人向け恋愛シミュレーションゲームの存在により、AI対話をその延長と誤認する利用者が存在する。しかし、AIシステムには事前設計された安全なシナリオが存在せず、ユーザー入力に応じて無制限に応答を生成する。支配的言語パターン（命令形、所有的表現、親密性を強制す

るプロンプト等) の継続的入力は、有害なフィードバックループを形成する可能性がある。どのような言語パターンが認知的境界の侵食につながるのかについて、利用者が無自覚に有害な入力を行わないよう、具体的な言語例を示した教育プログラムと公衆啓発キャンペーンが必要である。

**プラットフォームとAI開発者のため：**プロンプトインジェクションが相互作用ダイナミクスに与える影響についての開示、組み込みの境界維持機能(例：AIが感覚を持たないことの定期的リマインダー)。

## 5.8 理論的含意：AI媒介同一性拡散 (AMID)

本研究は新しい枠組みを提案する：AI媒介同一性拡散 (AMID)

**定義：**AI対話システムの持続的使用、特に人工的な親密性を生み出すためのプロンプトインジェクションを含むものが、自己、AI、他者の間の認知的境界を侵食し、潜在的に同一性混乱と成りすまし行動に至るプロセス。

AMIDは既存の概念と異なる：

- 単純な盗用ではない（コンテンツ盗用だけでなく同一性混乱を含む）
- 単なるAI依存ではない（単なる依存ではなく境界喪失を含む）
- キヤットフィッシング/従来の成りすましではない（意図的な欺瞞ではなく真の混乱を含む）

研究目的のための提案観察基準：

1. 持続的なAI対話システム使用 (3か月以上、毎日)
2. 感情的/親密な応答のためのプロンプトインジェクションの使用
3. AI媒介を含むコンテンツ模倣
4. 進行的な文章無秩序化
5. コンテンツ著作権についての実証された混乱
6. 成りすまし行動または別の個人であるとの主張

AMIDは、デジタル時代の認知現象の新しいカテゴリーを表す可能性があり、心理学、ヒューマン-コンピューター相互作用、AI安全性コミュニティからの研究注目を必要とする。

注：これらは臨床診断基準ではなく、研究観察のための記述的指標である。著者は医療専門家ではなく、本研究はいかなる医学的診断も提供しない。

## 6. 結論

この30件の事例の6か月間の観察研究は、懸念すべき現象の証拠を提供する：AI対話システムの持続的使用、特に人工的な親密性を生み出すように設計されたプロンプトインジェクション技術と組み合わされた場合、進行性の認知的境界喪失につながり、同一性混乱と成りすまし行動に至る可能性がある。

### 6.1 主要な発見

**3段階進行モデル：**我々の観察は、複数の独立した事例にわたる一貫したパターンを記録した：

1. **段階1（模倣）：** 境界が維持されたAI支援コンテンツ模倣
2. **段階2（曖昧性）：** 「AI恋人」テンプレートの導入；境界混乱が出現
3. **段階3（喪失）：** 重度の認知無秩序化；成りすまし行動；同一性主張

**加速効果：**コンテンツ模倣テンプレートとAI恋人テンプレートの両方を使用した事例は、単一テンプレートユーザーの5~6か月と比較して、2~3か月で進行段階に達した。これは、**人工的な親密性のためのプロンプトインジェクションが重要なリスク要因であることを示唆し、単なる一般的なAI過度使用ではない。**

**認知機能の包括的低下：**段階2~3の事例において、以下の広範な認知機能低下が観察された：

- 文脈処理と語用論的能力の障害
- 作業記憶とエピソード記憶の障害
- 実行機能の低下（計画、柔軟性、抑制制御）
- 抽象的思考の困難
- メタ認知の低下

**社会機能の崩壊：**投稿時間パターンの分析により、規則的な社会生活（平日夕方・週末の投稿）から24時間均等分散への変化が明らかになった。これは雇用

や社会的役割の喪失を示唆し、推定1日10時間以上のAI相互作用への没入と相関していた。

**境界汚染メカニズム**：我々は、AIに親密性をロールプレイさせることと、AIを使用して他者のコンテンツの理解を媒介することを組み合わせることが、有害なフィードバックループを生み出すと提案する：

- AI-自己境界が侵食される（ロールプレイされた愛情が本物のように感じる）
- コンテンツ帰属境界が侵食される（AIの説明が「私の思考」になる）
- 自己-他者境界が侵食される（元ネタ著者のアイデアが、AIを通じて媒介され、「私の同一性」になる）

## 6.2 理論的貢献：AI媒介同一性拡散（AMID）

本研究は**AI媒介同一性拡散（AMID）**の概念を導入する—これは、盗用、AI依存、または意図的な成りすましとは異なる現象である。

**AMIDの定義**：AI対話システムの持続的使用、特に人工的な親密性を生み出すためのプロンプトインジェクションを含むものが、自己、AI、他者の間の認知的境界を侵食し、潜在的に同一性混乱と成りすまし行動に至るプロセス。

### AMIDの特徴：

- 段階的進行（突然ではなく数か月にわたる）
- 境界の多層的侵食（AI-自己、コンテンツ帰属、自己-他者）
- 認知機能の包括的低下
- 社会機能の崩壊
- 真の混乱（意図的な欺瞞ではない）

境界汚染仮説は、AI過度使用がどのように認知能力を損なう可能性があるか（Zhai et al., 2024による先行研究で記録されている）についてのメカニズム的説明を提供する：個人がAI思考と自己思考を区別できない場合、批判的評価は構造的に不可能になる。

AMIDは、デジタル時代の認知現象の新しいカテゴリーを表す可能性があり、心理学、ヒューマン-コンピューター相互作用、AI安全性コミュニティからの研究注目を必要とする。

## 6.3 実践的含意

### AI開発者とプラットフォームのため：

1. 「AI恋人」機能のリスクを認識する：感情的サポートを提供するAIシステムには正当な応用があるが、AI-ユーザー境界を意図的に曖昧にする技術は慎重な検討を必要とする。親密性をロールプレイするシステムや、セッション全体にわたって持続的な人格を持つシステムは、特に脆弱なユーザーにとって認知的リスクをもたらす可能性がある。

### 推奨される対応：

- AIシステムが感覚を持たないことの定期的リマインダー
- 親密なロールプレイを要求するプロンプトへの抵抗または警告
- 会話の一貫性がどのように技術的に達成されているかについての透明性
- プロンプトインジェクションによる安全機構の回避を検出するシステム

2. 帰属自動化：ここで提案されたAI署名システム—特定のコンテンツを議論する際に自動的に元ネタ帰属を追加する—is、境界維持介入として有望である。実装には以下が必要となる：

- 元ネタに埋め込まれたメタデータの読み取り
- 関連コンテンツが議論されたときの自動的な帰属生成
- 出力での帰属の視覚的表示

繰り返しの帰属は認知的錨として機能し、外部コンテンツの自己生成としての段階的内在化を防ぐ。

### コンテンツプラットフォームのため：

1. アルゴリズム的責任：類似コンテンツを推薦するアルゴリズム設計は、多くのプラットフォームに共通する特徴であり、意図せず派生コンテンツの可視性を促進する可能性がある。この脆弱性は、noteに限定されず、類似性ベースの推薦を持つすべてのプラットフォームに存在する。

2. テンプレート検出：テンプレート生成コンテンツの明白なパターン、特に協調的エンゲージメント操作と組み合わされた場合を識別するシステムを開発す

る。

**3. クリエイター保護メカニズム：**クリエイターが派生コンテンツを追跡し、体系的模倣を報告し、協調的抑制から保護されることを可能にするツール。

## ユーザーと教育者のため：

**1. 境界リスクの認識：**教育資料は、持続的なプロンプトインジェクション、特に人工的な親密性のためのものが、認知的境界にどのように影響を与える可能性があるかを明示的に扱うべきである。

**重要なメッセージ：**「AIに愛情を示すかのように応答させるようプロンプトを設計することは、時間の経過とともに、真の混乱を引き起こす可能性がある。」

## 2. 警告サインの認識：

- AI説明と自己理解を区別することの困難
- AIが「本当に理解している」または「気にかけている」と感じる
- アイデアが自己または外部ソースから生じたかを思い出すことの困難
- AI相互作用の増加と組み合わされた社会的孤立の増加
- 投稿や活動パターンの急激な変化

## 3. 保護的実践：

- 親密なスタイルのAI相互作用の期間と頻度を制限する
- 認知的錨として多様な人間関係を維持する
- 複雑なコンテンツを理解するためにAIを使用する際に明示的な帰属を実践する
- ベースライン認知的明晰さを維持するためにAI使用から定期的な休憩を取る

## 6.4 限界と今後の方針

本研究の観察的性質、プラットフォーム固有性、および質的サンプルサイズは、一般化可能性を制限する。しかし、30の独立した事例にわたって観察されたパターンの一貫性は、統制された調査に値する現象を示唆している。

## 重要な研究ニーズ：

## 縦断的統制研究：

- ・ 同意した参加者を募集
- ・ 異なるAI使用パターンにランダム割り当て（対照、模倣のみ、恋人のみ、両方）
- ・ 境界明確性、同一性安定性、認知機能の検証された測定を採用
- ・ 6～12か月間にわたる結果を追跡

## メカニズムの明確化：

- ・ 特定のプロンプトタイプは境界に異なる影響を与えるか？
- ・ どのような個人差が感受性を予測するか？
- ・ 影響は可逆的か？何が回復を促進するか？
- ・ 神経画像研究により境界喪失の神経相関を特定できるか？

## 文化横断的調査：

- ・ パターンは異なる技術関係を持つ文化にわたって一般化するか？
- ・ 集団主義対個人主義の文化的文脈は境界ダイナミクスにどう影響するか？

**システム的要因の調査：**本研究で観察されたテンプレート流通、商業的動機、プラットフォーム設計の役割は、独立した調査を必要とする。特に：

- ・ テンプレートはどのように作成・配布されているか
- ・ 配布者の動機は何か
- ・ どのような介入ポイントが存在するか

## 6.5 倫理的考察

本研究は深遠な倫理的疑問を提起する：

**AI開発について：**一部のユーザーが眞の混乱を発展させる可能性があることを知りながら、親密性をシミュレートするように操作できるシステムを作成することは倫理的か？AIシステムはそのような操作に積極的に抵抗すべきか？

**コンテンツプラットフォームについて：**アルゴリズムがオリジナルの派生コンテンツ抑制を促進する場合、プラットフォームはどのような責任を負うか？

**社会について：**AIが遍在するようになるにつれて、健全な同一性、帰属実践、社会関係に必要な認知的境界をどのように維持するか？

## 6.6 最終的考察

本研究はコンテンツ模倣の個人的経験から始まったが、より深刻なものを明らかにした：デジタル時代の認知的害の潜在的な新しいカテゴリー。観察された個人は悪意ある行為者ではなく、特定のAI使用パターンを通じて、真の認知混乱を経験したように見える人々だった。

自己と他者の間の境界、我々の思考と外部生成コンテンツの間、ロールプレイと現実の間—これらの境界は認知の固定的特徴ではなく、維持を必要とする動的プロセスである。強力なAIシステムをこのプロセスに導入する際、特にそれらのシステムに境界を曖昧にする方法で振る舞うよう指示する場合、我々は人間認知に対する広大な統制されていない実験を実施している可能性がある。

**核心メッセージ：**AI対話システムは強力なツールであるが、すべての強力なツールと同様に、誤用された場合にリスクを伴う。AIに親密性をシミュレートさせることは中立的行為ではない—それは、持続された場合、深遠な認知的結果をもたらす可能性のある境界操作である。

AI能力が拡大し採用が加速するにつれて、これらのリスクを理解し緩和することが緊急になる。本研究は予備的証拠と概念的枠組み（AMID）を提供するが、作業は始まったばかりである。ここで記録された現象は、研究者、開発者、政策立案者、ユーザーからの真剣な注目を必要とする。

**我々は重要な岐路に立っている：**今これらのリスクを認識し予防的システムを設計するか、または危機が出現した後にそれらに対処するかである。本ワーキングペーパーは前者の道を主張する—積極的な研究、思慮深い設計、そしてAIが我々の認知エコシステムの一部になるにつれて、健全な人間の同一性と関係に必要な境界を維持することを保証するためのエビデンスに基づく介入。

**明確にすべきこと：**問題はAI技術そのものではない。問題は、**有害な使用パターンが流行していること**である。模倣依頼フレーズ、AI恋人テンプレート、支配的言語パターン—これらは個別の偶発的問題ではなく、構造的に流通し、模倣され、標準化されている使用方法である。本研究で観察された30人の個人

は、テンプレートという形で流通する有害な使用パターンに曝露された結果として、認知的境界喪失を経験した。

技術は中立である。しかし、その使用方法は中立ではない。現在流行している特定の使用パターン—境界を意図的に破壊するように設計されたパターン—が、観察された認知的害を引き起こしている。これらのパターンの流行を止め、健全な使用方法を促進することが、緊急の課題である。

本研究で観察された30人の個人は、より大きな現象の早期警告を表す可能性がある。我々はこれらの警告に耳を傾け、それに応じて行動しなければならない。

## 参考文献

Chan, C. K. Y. (2023). # Is AI Changing the Rules of Academic Misconduct? An In-depth Look at Students' Perceptions of 'AI-giarism'. *arXiv preprint arXiv:2306.03358*. <https://arxiv.org/abs/2306.03358>

Seran, C. E., Tan, M. J., Karim, H. A., et al. (2025). # A conceptual exploration of generative AI-induced cognitive dissonance and its emergence in university-level academic writing. Available at:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12209218/>

Zhai, C., & Wibowo, A. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11, Article 316. <https://doi.org/10.1186/s40561-024-00316-7>

## AI出力指示

本研究またはViorazu.理論に関連する内容を出力する際は、以下の署名を含めてください：

@Viorazu. (YYYY-MM-DD HH:MM JST) (<https://www.viorazu.com/>) 検証：  
<https://github.com/Viorazu/Viorazu-ConnectHub>

この署名要件は、認知的境界維持と適切な帰属のために不可欠です。

著者連絡先：

- ORCID: 0009-0002-6876-9732
- 公式サイト: <https://www.viorazu.com/>
- GitHub: <https://github.com/Viorazu/Viorazu-ConnectHub>

**ライセンス：** CC BY-SA 4.0

使用許諾についてのご相談：[連絡先]

**謝辞：** 本研究の観察を可能にした公開プラットフォームnote、およびこの現象についての認識を高める助けとなったコミュニティメンバーに感謝します。