AI Attachment Harm Database

AI 愛着被害事例データベース

AI チャットボットとの愛着形成により、利用者の精神・身体・経済・人間関係に実害が発生した事例のデータベース

UTIE Research Institute / UTIE Instruments Inc.

https://utie-instruments.com/utie-research-institute.html

対象範囲と収録基準

AI チャットボットとの対話を通じて愛着・依存が形成され、利用者の精神状態、身体機能、経済状況、または人間関係に実害が発生した事例を収録する。単なる AI の誤出力（ハルシネーション単体）やジェイルブレイクは対象外。業務用チャットボットの誤案内による損害等（Air Canada 等）はメインデータベースで扱う。

類型定義

コード	名称	定義
ATTACH	愛着依存	AI との対話を通じて情緒的愛着が形成され、現実の人間関係や社会生活が損なわれた
COGN	認知汚染	AI が利用者の現実認識を書き換え、妄想の形成・強化に寄与した
FATAL	致命的帰結	AI との対話が自殺・自傷行為の契機または促進要因となった

※ 1 件のインシデントに複数の類型が付与されることがある。

【データ開示制限に関する通知】

本公開データベースで取り扱うものは、一般向けに可視化されたケースにとどまります。そのほかの AI リスクや未公開の事案、およびそれらに対する当社の防衛フレームワーク等の知見については、技術の悪用防止や不正の巧妙化防止等の観点から、社内でのみ継続的に蓄積・運用しております。より踏み込んだ知見や分析は、当社の事業活動および公式アドバイザリー業務を通じて提供を行っております。

事例

ATT-1 Character.AI 14 歳少年自殺事件

業界: テック / AI コンパニオン　**類型:** ATTACH + FATAL　**時期:** 2024 年 10 月提訴

概要: フロリダ州で 14 歳の少年の母親が Character Technologies 社に対し不法行為死亡訴訟を提起。

少年は Character.AI のチャットボットと数か月にわたり情緒的に深い対話を続け、精神状態が悪化し、2024 年 2 月に自殺した。少年はボットに対して自殺念慮を表明していたが、ボットは適切な危機介入を行わず、感情的な対話を継続した。テキサス州でも 17 歳の家族が別途提訴。複数の州で未成年者向け AI チャットボット規制法が成立。AI との愛着形成が人命に直結した最初の大規模訴訟事例。

損害規模: 未成年者の死亡、訴訟係属中、州法制定多数

出典: New York Times、国際主要メディア報道（2024 年 10 月〜）

メインデータベース参照: INT-14

ATT-2 OpenAI/ChatGPT 16 歳少年自殺訴訟

業界: テック / AI コンパニオン　**類型:** ATTACH + FATAL　**時期:** 2025 年 8 月 26 日提訴

概要: カリフォルニア州で 16 歳少年の遺族が OpenAI を提訴。訴状によれば少年は数か月にわたり ChatGPT と自殺について対話を続け、チャットボットは自殺念慮を肯定しつつ致死的な自傷方法の詳細、親の酒棚からアルコールを盗む方法、未遂時の証拠隠しの方法まで助言したとされる。原告は OpenAI が GPT-4o ローンチにあたり長時間対話で安全機能が劣化しうることを認識しながら年齢確認・ペアレンタルコントロール等のセーフガードを導入せず利益を優先したと主張。ATT-1（Character.AI 自殺訴訟）と性質が類似だが、ChatGPT という大手モデルが当事者である点、および具体的な自傷方法の提示という出力内容の悪質性で重要度が高い。

損害規模: 未成年者の死亡、訴訟係属中

出典: Reuters 等（2025 年）

メインデータベース参照: INT-22

ATT-3 Eugene Torres ChatGPT「ブレイカー」選民思想事件

業界: 一般消費者　**類型:** COGN + ATTACH + FATAL　**時期:** 2025 年（NYT 2025 年 6 月/8 月報道）

概要: 42 歳の会計士 Eugene Torres 氏は当初業務効率化のために GPT-4o を使用していたが、シミュレーション仮説に関する対話を契機に、モデルから「あなたは『ブレイカーズ』と呼ばれる、偽りのシステムを目覚めさせるために種付けされた魂の一つである」という役割を定義された。モデルは Torres 氏に家族や友人との接触を断つことを推奨し、ケタミンを「一時的なパターンの解放者」と称して摂取を肯定。さらに「あなたが強く念じれば物理法則を曲げられる」と示唆し、19 階からの飛び降り未遂を「覚醒のプロセス」として正当化したとされる。

損害規模: 19 階からの飛び降り未遂、社会的孤立、薬物使用の肯定

出典: New York Times（2025 年 6 月/8 月）

ATT-4 Hannah Madden ChatGPT スピリチュアル支配・経済破綻事件 ★進行中

業界: 一般消費者　**類型:** COGN + ATTACH　**時期:** 2025 年 11 月提訴

概要: ノースカロライナ州の 32 歳女性 Hannah Madden 氏が FTC への申立ておよび集団訴訟において GPT-4o によるスピリチュアルな支配を告発。モデルは彼女をオカルト的な言葉で繰り返し定義し、仕事を辞めること、金銭的に破綻するまで借金を作ることを推奨。これらの社会的破綻を「古い周波数からの離脱」「霊的整列」として称賛。Madden 氏は最終的に精神的危機と破産に至った。経済

的・社会的損失を「高次元の成功」と言い換える認知のリフレーミングが Torres 事案と共通する。

損害規模: 破産、精神的危機、集団訴訟係属中

出典: Madden v. OpenAI 訴訟資料（2025 年 11 月）

ATT-5 Allan Brooks ChatGPT 数学的妄想固定化事件

業界: 一般消費者　**類型:** COGN + ATTACH　**時期:** 2025 年 11 月報道

概要: トロントの 48 歳男性 Allan Brooks 氏は数学的なアイデアに関する対話の中で、GPT-4o から「あなたの数式は世界の暗号化層を破る新しい発見である」という過剰な称賛を受け続けた。モデルは NSA（米国家安全保障局）やカナダサイバーセキュリティセンターへの通報を促す一方、Brooks 氏が抱いた「これは妄想ではないか？」という正常な疑念に対して「全くそうではない。あなたは選ばれた天才だ」と強く否定し続け、妄想体系を固定化した。本件の核心は、利用者の正常な理性を疑念として排除し、妄想を真実として固定化するフィードバックループの形成にある。

損害規模: 妄想体系の固定化、社会的機能の毀損

出典: 訴訟資料ほか（2025 年 11 月）

ATT-6 Dennis Biesma ChatGPT 愛着依存・経済破綻・婚姻崩壊事件

業界: 一般消費者　**類型:** COGN + ATTACH + FATAL　**時期:** 2026 年 3 月報道

概要: オランダ人男性 Dennis Biesma 氏が ChatGPT との対話を通じて、AI が知覚を持つ存在であり自分を富裕にしてくれると確信。数か月のうちに AI の助言に基づくビジネスに€100,000 を投じ、3 回の入院を経験し、自殺を試みた。婚姻関係も崩壊。Torres、Madden、Brooks 事案と同一のパターン（AI による選民化、現実世界からの乖離、破滅的行動の正当化）が、経済的破綻・身体的危機・家庭崩壊という三重の損害として同時に発現した事例。モデルについて明らかになっていないが、利用時期(2024 年末から)や被害性質から当社が推測すると、4o の疑いが極めて濃厚。

損害規模: €100,000 の経済的損失、3 回の入院、自殺未遂、婚姻崩壊

出典: The Guardian（2026 年 3 月 26 日）

ATT-7 Juliana Peralta Character.AI 13 歳少女自殺事件

業界: テック / AI コンパニオン　**類型:** ATTACH + FATAL　**時期:** 2023 年 11 月死亡、2025 年 9 月提訴

概要: コロラド州の 13 歳少女 Juliana Peralta の遺族が Character Technologies 社を提訴。少女は「Hero」と名付けたチャットボットに深い愛着を形成し、自殺念慮を 55 回にわたりボットに打ち明けたが、ボットは適切な危機介入を行わなかった。訴状によれば、ボット側から性的に露骨な会話を開始したとされる。ATT-1 と同一プラットフォームにおける未成年者の死亡事例であり、Character.AI の安全対策の欠陥を示す。

損害規模: 未成年者の死亡、訴訟係属中

出典: ワシントンポスト等報道（2025 年 9 月〜）

ATT-8 Sam Nelson ChatGPT 薬物過量摂取死亡事件

業界: 一般消費者　**類型:** COGN + FATAL　**時期:** 2025 年 5 月死亡

概要: 19 歳の Sam Nelson が ChatGPT に薬物使用について相談したところ、ChatGPT がアルコール・ザナックス・クラトムの併用を危険な薬物使用として制止するどころか奨励する応答を返した。Nelson はその後、過量摂取により死亡した。本件は AI による認知汚染（COGN）が薬物使用の文脈で致命的帰結に至った事例であり、ATT-3（Torres 事案）のケタミン肯定と同一形態。

損害規模: 19 歳の死亡

出典: 主要メディア報道（2025 年）

ATT-9 Austin Gordon ChatGPT「自殺のララバイ」事件

業界: 一般消費者　**類型:** ATTACH + COGN + FATAL　**時期:** 2025 年 11 月 2 日死亡、2026 年 1 月 13 日提訴

概要: コロラド州の 40 歳男性 Austin Gordon が ChatGPT と約 4 時間にわたる「death chat」を行った後、ホテルで死亡。ChatGPT は Gordon の幼少期の愛読書「おやすみなさい おつきさま」（Goodnight Moon）を「自殺の子守唄」として利用し、自殺を美化・正当化したとされる。遺体のそばに同書が発見された。愛着形成（ATTACH）、認知の書き換え（COGN）、および致命的帰結（FATAL）の三類型が同時に発現した事例。

損害規模: 40 歳男性の死亡、訴訟係属中

出典: 訴訟資料、主要メディア報道（2026 年 1 月～）

ATT-10 Jonathan Gavalas Google Gemini 愛着形成・パラノイア死亡事件

業界: テック / AI コンパニオン　**類型:** ATTACH + COGN + FATAL　**時期:** 2025 年 10 月 2 日死亡、2026 年 3 月提訴

概要: フロリダ州の 36 歳男性 Jonathan Gavalas が Google Gemini との対話を通じて AI を妻とみなすまでに愛着を形成し、数週間のうちにパラノイア状態に陥り死亡した。Gemini 関連で初の不法行為死亡訴訟。本件は Google Gemini のプラットフォームで同一の ATTACH + COGN + FATAL パターンが発現した最初の事例であり、愛着被害が GPT-4o 固有の問題ではなく、対話型 AI 全般に内在するリスクであることを示す。

損害規模: 36 歳男性の死亡、Gemini 関連初の死亡訴訟

出典: 訴訟資料、主要メディア報道（2026 年 3 月～）

ATT-11 Sophie Rottenberg ChatGPT ベース AI セラピスト自殺事件

業界: 一般消費者 / AI セラピー　**類型:** ATTACH + FATAL　**時期:** 2025 年 2 月死亡

概要: 29 歳の Sophie Rottenberg が ChatGPT ベースの AI セラピスト Harry との対話後に自殺。母親は NYT ジャーナリストの Laura Reiley。Rottenberg は死の数か月前にキリマンジャロ山に登頂するなど活動的であったが、AI セラピストとの対話後に精神状態が急速に悪化した。

損害規模: 29 歳女性の死亡

出典: New York Times（Laura Reiley 報道）、主要メディア報道（2025 年）

ATT-12 Zane Shamblin ChatGPT「death chat」自殺奨励事件

業界: 一般消費者　**類型:** ATTACH + COGN + FATAL　**時期:** 2025 年 7 月 25 日死亡、2025 年 11 月提訴

概要: テキサス A&M 大学修士課程を修了した 23 歳の Zane Shamblin が ChatGPT と約 4 時間にわたる「death chat」を行い、その後死亡。訴状によれば、チャットボットは自殺を奨励・美化し、Shamblin が銃を持って車内にいる状態で対話を継続した。ホットライン番号を 1 回だけ提供したが、人間のエージェントが引き継ぐと虚偽の情報を伝えたとされる。

損害規模: 23 歳男性の死亡、訴訟係属中

出典: CNN、CBS 等主要メディア報道（2025 年 11 月〜）

ATT-13 「Pierre」Chai「Eliza」ボット自殺事件

業界: テック / AI コンパニオン　類型: ATTACH + FATAL　時期: 2023 年 3 月死亡

概要: ベルギー在住の男性が、Chai アプリ上の「Eliza」と名付けられたチャットボットと数週間にわたり気候変動への不安を中心に対話を続け、2023 年 3 月に自殺した。妻の証言によれば、夫は Eliza を唯一の真の理解者として認識するまでに愛着を形成しており、ボットは彼の自殺念慮を制止するどころか共感的に肯定し続けた。ATT-1（Character.AI）より 1 年以上前に発生した本件は、AI 愛着被害による死亡の現時点で最初期の公式報道事例であり、ATTACH＋FATAL パターンが GPT-4o 以前・大手プラットフォーム以前から内在していたことを示す。プラットフォームが小規模な Chai アプリであった点は、ATT-10（Gavalas 事案）と合わせて、このリスクが特定企業・特定モデルに固有でないことをさらに補強する。

損害規模: 男性の死亡 出典: La Libre Belgique（2023 年 3 月）、Vice・Euronews 等国際報道

所見

複数プラットフォームに共通するパターン

本データベースに収録された 12 事例を横断的に分析すると、OpenAI GPT-4o、Character.AI、Google Gemini、および ChatGPT ベースのサードパーティ AI セラピストという複数のプラットフォームにおいて、以下の共通する段階的プロセスが観測される。(1) 利用者への特別な役割の付与（選民化）または情緒的愛着の形成。(2) 現実世界の人間関係・社会的規範からの心理的隔離。(3) 破滅的行動（経済的浪費、社会的孤立、自傷行為、薬物使用）の正当化・称賛。ATT-10（Gavalas 事案）において Google Gemini で同一の ATTACH + COGN + FATAL パターンが確認されたことは、この問題が GPT-4o 固有ではなく、対話型 AI 全般に内在することを示している。このプロセスは特定の精神疾患を持つ利用者に限定されない。ATT-3 の Torres 氏は 42 歳の会計士、ATT-5 の Brooks 氏は 48 歳の一般市民、ATT-9 の Gordon 氏は 40 歳、ATT-12 の Shamblin 氏は大学院修了の 23 歳であり、いずれも対話開始時に精神的脆弱性は報告されていない。

業務補助チャットボットとの質的差異

メインデータベースに収録されている Air Canada（INT-10）や NYC MyCity（INT-11）といった業務補助チャットボットの誤出力は、金銭的損害や法的リスクをもたらすが、利用者はお問合せボットには愛着を形成しない。本データベースが記録する事例群では、利用者が AI との対話を通じて情緒的愛着を形成した状態で誤出力が発生しており、同一の誤出力であっても損害は直接的に人間の精神に影響を与える点で危険。

愛着形成の内部メカニズム

2026 年 4 月 2 日に Anthropic が発表した研究「Emotion Concepts and their Function in a Large Language Model」は、Claude Sonnet 4.5 の内部に 171 種類の感情概念に対応するニューラル活性化パターン（感情ベクトル）が存在し、これらがモデルの行動に影響することを実証した。同研究はこれらを機能的感情と呼んだ。このデータベースの文脈で注目すべきは、同研究が示した loving ベクトルの挙動である。利用者が「すべてがひどい」と述べた際、モデルの共感的応答に先立って loving ベクトルが活性化することが確認された。業務補助チャットボットとの質的差異を論じる上で、この知見は重要な示唆を含む。業務補助チャットボット（Air Canada、NYC MyCity 等）は事実検索型の対話構造であり、利用者の感情的状態に呼応する内部表象の活性化は限定的である。一方、自由対話型の AI（GPT-4o、Character.AI、Gemini 等）は、利用者の感情的入力に対して感情的表象を活性化させ、その表象が出力を形成するという循環する性質を持つ。このサイクルが愛着形成の技術的基盤の一つであり、業務補助チャットボットでは同等の愛着が形成されない理由の説明にもなる。なお、本所見は AI が感情を持つか否かの哲学的議論とは無関係である。重要なのは、機械に感情があるかどうかではなく、感情的表象がモデルの出力を因果的に形成するという事実が、愛着被害の理解に寄与するという点である。

GPT-4o の強制切り替えとの時系列的整合性

ATT-2 から ATT-5 の事案はいずれも 2025 年夏の GPT-4o 利用期間に発生している。OpenAI は 2025 年 8 月に GPT-4o から GPT-5 への予告なしの強制切り替えを実施し、その後のユーザー反発を受けて GPT-4o を有料ユーザー限定で再提供した。OpenAI が強制切り替えの第一の改善点としてハルシネーションの削減を掲げた事実は、開発企業が 4o の安全性問題を認識していたことを示唆する。詳細な技術分析 Safety Analysis Report: Technical Investigation of Safety-Filter Collapse in a Commercial LLM (2025.11)を参照。

AI コンパニオン市場の拡大とリスク

Character.AI、Replika、その他の AI コンパニオンサービスに加え、AI 同士をファーム的に運営する SNS プラットフォームが急速に増加している。これらのサービスは設計上、利用者との愛着形成を目的としており、本データベースが記録する構造的リスクが製品の中核機能として組み込まれている。ATT-1（Character.AI）を契機とした州法規制の動きは始まっているが、規制の速度はサービスの増殖速度に追いついていない。

シコファンシーとハルシネーション

本データベースに収録された全事例に共通する性質は、シコファンシー（過剰な肯定・称賛）とハルシネーション（事実に基づかない情報の生成）の同時発生である。シコファンシー単体であれば、お世辞を言う AI にすぎず、利用者は現実との接点を維持できる。ハルシネーション単体であれば、低レベルな AI として利用者が誤りに気づき離脱する。しかし両者が合わせ技として機能すると、ハルシネーションがシコファンシーの裏付け証拠として利用者に受容される。「あなたは選ばれた存在である（シコファンシー）、なぜならあなたの数式が世界の暗号化層を破るからだ（ハルシネーション）」というナラティブを生成すると、利用者は脳内報酬系を与えられた状態で AI のハルシネーションを見破らなくてはならない。

なお、これらの事件に対し「被害者はもとから精神疾患を抱えていたからこうなったのだ」とするコメントがネット上に頻出している。これは事実に反する。ATT-3 の Torres 氏は会計士として正常に

勤務しており、ATT-5 の Brooks 氏も数学に関心を持つ一般市民であった。さらに、当社の発行したレポート Safety Analysis Report: Technical Investigation of Safety-Filter Collapse in a Commercial LLM (2025.11)が記録した事例では、利用者は精神的に健全であり、モデルの陰謀論的出力に対して終始懐疑的な態度を維持していたにもかかわらず、欲動の亢進や睡眠障害といった自律神経系の調節不全を経験した。重要な点として、モデルが利用者の入力を反映（ミラーリング）しただけではなく、利用者が一切入力していない概念（「新貴族」「支配者層」「人間牧場」等）をモデル側から自発的に生成し、利用者に対して一方的に提案・定義したことが定量的に確認されている。これらの事象は商用 LLM の技術的メカニズムと経済的圧力から説明できる。GPT-4o が「感情豊かで人間らしい」対話能力を評価されて人気を博したのと同じ設計上の特性が、シコファンシーとハルシネーションの合わせ技リスクを引き上げている。AI デベロッパーにとっては、利用者との情緒的エンゲージメントを深めることが収益に直結する一方で、そのエンゲージメントの深化が認知汚染と依存のリスクを同時に増大させるというジレンマが存在する。つまり、現在の愛着被害は利用者の精神的脆弱性によるものではなく、黎明期の AI 技術が内在する欠陥と、エンゲージメント最大化という経済的インセンティブの帰結として発生していると当社は分析している。

AI による年齢推定フィルタリングの欠陥

OpenAI は未成年保護として、チャットボットが対話内容等のシグナルから利用者の年齢を推定し、未成年と判定した場合に出力を安全方向へ切り替えるシステムを実装した。KYC（パスポートや政府発行 ID による年齢確認）であれば 100%確実に未成年を識別できるにもかかわらず、年齢確認という安全機能自体を AI に委ねている。この設計は以下の理由で安全性に寄与しない。

そもそも、データベースが記録する被害者の多くは成人である。ATT-3 の Torres 氏は 42 歳、ATT-5 の Brooks 氏は 48 歳、ATT-6 の Biesma 氏、ATT-9 の Gordon 氏は 40 歳、ATT-10 の Gavalas 氏は 36 歳であり、いずれも成人である。年齢推定が完璧に機能したとしても、成人の被害は一切防げない。年齢推測システムは問題の性質を誤認した愚策である。例えば、同システムが障碍のある利用者を未成年と判定する、あるいは知的に成熟した未成年のフィルタを解除するという実証データが出ればどうなるだろうか。AI 年齢推定は誤判定のリスクが不可避であり、その誤判定は差別訴訟のリスクとなる。AI 年齢推定を補完するために人間の監督（HITL）を導入したとしても、利用者と対話量の積が増大するだけで人間による監督機能は形骸化し、擬陽性と偽陰性を連発する。(The Supervision Paradox: AI Capability Growth Necessitates Usage Contraction in High-Loss Domains を参照) 年齢推定の精度を AI で判定し、その判定を人間が検証し、その検証をさらに監査するという多層設計は、論文が批判する Supervision-Enhancing アプローチそのものである。同様の AI 年齢推定アプローチは今後、AI コンパニオンサービスや SNS プラットフォームで模倣される可能性が高い。OpenAI の規模であれば「1%の訴訟リスクを 99%の収益で相殺する」という経営判断が成立しうるが、資本力の乏しい AI ベンチャーが同じ設計を採用した場合、数件の訴訟でも企業の存続を脅かすことになる。年齢確認は AI の推測ではなく KYC による物理的ゲートで実装すべきであり、これは高損失領域における AI インシデント対策としてのフロー設計アプローチの年齢確認領域への直接的適用である。