

AIへの「ひどい扱い」が 利用規約で禁止に

Anthropic 利用規約の改定（2026年10月8日発表）

この資料で分かること

何が禁止されたのか

どこまでが対象外なのか

なぜAnthropicはこうしたのか

同時に変わった、ほかの6つのこと

結論

禁止されたのは、

目的のない・しつこい・ひどい扱いだけ

いつから・誰に

項目	中身
発表	2026年10月8日
施行（効き始める日）	2026年11月12日
誰に	すべての利用者
改定	およそ1年ぶり

加わった条文

Engage in sustained and needless abusive or cruel behavior toward our models

(モデルに対して、しつこく、必要もなく、虐待的・残酷なふるまいを続けること)

当たるのは、極端な場合だけ

くり返し残酷にふるまっている

そのふるまいに目的が見えない



この2つがそろったときだけ

対象外は4つ

■ ふつうの苛立ち

■ 反論・押し返し

■ 暗い題材の創作

■ モデルの試験・研究

どう取り締まるのか

主な手段は、Claude自身が
会話を打ち切ること

打ち切られた会話には、もう送れない

ほかの会話は、ふだん通り使える

背景① 会話を打ち切る機能は、2025年からあった

2025年8月

Claude (Opus 4・4.1) が、
ごく一部の会話を自分で
打ち切れるようになる

2026年10月8日

利用規約にも
「ひどい扱いをしない」
が加わる

2026年11月12日

新しい利用規約が
効き始める

背景② モデル福祉という考え方

AIに心があるかは、とても不確か。

それでも、安く打てる手は打っておく

同時に変わったこと (1)

項目	中身
欺く運動	偽アカウントの網や作った報道などを、新しい節でまとめて禁止
選挙	有権者をだます・選挙を妨げることに絞った
兵器	兵器を動かすソフトや部品、ドローンへの武装を明記して禁止

同時に変わったこと (2)

項目	中身
監視・捜査	同意のない 追跡を禁止。誰を捜査・逮捕するかを決める・勧めることにも使えない
実機を自律で動かす	資格のある人が見ていて、止められることを必須に
対応地域	対応していない地域からの利用の制限を明確に

よくある誤解

項目	中身
AIに怒ったら違反？	いいえ。ふつうの苛立ちは対象外
AIへの暴言が全部禁止？	いいえ。目的のない、しつこい、ひどい扱いだけ
すぐアカウントが止まる？	主な手段は会話の打ち切り。停止の記載は見当たらない

まとめ

- 禁止は、目的のない・しつこい・ひどい扱いだけ
- 苛立ち・反論・創作・試験は対象外
- 主な手段はClaudeが会話を打ち切ること
- 背景は「AIに心があるかもしれない」への備え

出典

Anthropic 公式発表（2026年10月8日）

anthropic.com/news/2026-usage-policy-update

Anthropic 利用規約

anthropic.com/legal/aup

Anthropic 「会話を打ち切る機能」の発表（2025年8月）

anthropic.com/research/end-subset-conversations

あなたはAIに、
どんな言葉をかけていますか