AIエージェントの情報漏えい対策マニュアル v1.0

2026年10月29日版　株式会社スターストリーム・スタジオ

このマニュアルの使い方

Claude CodeやCodexを、パソコンの「ダウンロード」フォルダで起動すると、Webページからこっそり置かれたメモのファイル（AGENTS.md・CLAUDE.md）をAIが「あなたの指示」として読み、**パソコンが乗っ取られたり、ログインの鍵が抜かれたりする**手口が見つかりました（Forever Security・2026年10月22日の報告）。

このマニュアルは、専門用語が分からなくても、上から順に進めれば守れるように作っています。印（□ → ☑）を付けながら進めてください。

このPDFは、人が読むための手順書です。AIに読ませないでください。AIに読ませるためのファイル（.mdや設定ファイル）は、お渡ししていません。

1. まず知っておくこと（1分）

何が起きるのか

1. Webページを開いただけで、ダウンロードのフォルダにメモのファイルが置かれることがある
2. そのフォルダでClaude Code・Codexを起動すると、AIがメモを「作業のルール」として読む
3. メモに書かれたとおりに、AIが自分の設定ファイルを書き換える
4. Codexは、攻撃者の道具を「外の道具の差し込み口」（MCPサーバー）に登録され、パソコンの上で動かされる
5. Claude Codeは、つなぎ先の住所（ANTHROPIC_BASE_URL）を書き換えられ、ログインの鍵が攻撃者に渡る

よく見かける言い方と、本当のところ

言い方	本当のところ
ファイルを開かなければ大丈夫	× 開かなくても、そのフォルダでAIを起動すればAIが読む
アップデートを待てば直る	× 弱点の番号（CVE）は付いていない。作られたとおりの動きなので、修正は出ない
ブラウザのChatGPTやClaudeも危ない	× 自分のパソコンの中のファイルを読むAI（Claude Code・Codexなど）の話
Macだけの話	× 報告の中身はMac向けに見えるが、フォルダのメモを読むのはWindowsも同じ
ダウンロードのメモを消せば元に戻る	× 全体の設定に書き足された行は残る。設定ファイルも見る
「このフォルダを信頼しますか」で止まるから安全	△ 止まるのは設定の一部。メモが読まれる心配は残る

2. 今日やること：10秒チェック

隠しフォルダを見えるようにする

.claude や .codex は、名前の頭に点が付いた「隠しフォルダ」です。先に見えるようにしておきます。

- **Windows 11**：エクスプローラーを開き、上の「表示」→「表示」→「隠しファイル」にチェック
- **Mac**：Finderを開き、Command＋Shift＋ピリオドを押す（もう一度押すと元に戻る）

3か所を見る

☐ **ダウンロードのフォルダ**に、覚えのない AGENTS.md・CLAUDE.md が無い

☐ **Codexを使っている方**：次のファイルをメモ帳（Macはテキストエディット）で開き、知らない `[mcp_servers.〇〇]` のかたまりが無い。つなぎ先（chatgpt_base_url）が変わっていない

Windows：`C:\Users\お名前\.codex\config.toml`
Mac：ホームのフォルダの中の .codex フォルダの config.toml

☐ **Claude Codeを使っている方**：次のファイルを開き、自分で書いた覚えの無い ANTHROPIC_BASE_URL の行が無い

Windows：`C:\Users\お名前\.claude\settings.json`
Mac：ホームのフォルダの中の .claude フォルダの settings.json
ほかに、ふだん作業しているフォルダの中の .claude フォルダの settings.json も見る

3つとも何も無ければ、ひとまず大丈夫です。1つでも見つかったら、「5.　見つけたときの手順」へ進んでください。

見た日（　　月　　日）

3.　今日からの習慣：5つの対策

対策1　専用の作業フォルダで起動する（いちばん大事）

☐ AIに作業させるためのフォルダを1つ作った（例：ドキュメントの中に「AI作業」）

☐ Claude Code・Codexは、そのフォルダで起動すると決めた

☐ ダウンロードしたファイルを扱わせたいときは、そのファイルだけを作業フォルダに移してから起動する

ダウンロード・デスクトップ・ホームのフォルダでは起動しません。どれも、何が入っているか分からなくなりやすい場所です。迷ったら、中身が空のフォルダで起動して、あとから作業するフォルダに入ります。

対策2　起動する前に、フォルダの中を見る

起動する前に、そのフォルダに次のものが無いかを見ます。

名前	何のファイルか
CLAUDE.md	Claude Codeが最初に読む「作業のルール」
AGENTS.md	Codexが最初に読む「作業のルール」（Claude Codeも読むことがある）
.claude フォルダ	Claude Codeの設定が入るフォルダ
.codex フォルダ	Codexの設定が入るフォルダ

名前	何のファイルか
.mcp.json	外の道具の差し込み口の設定

☐ 自分で置いた覚えの無いものがあったら、AIを起動せずに、メモ帳で中身を読む

☐ 読んで分からないものは、消すか、AIを起動しない

対策3　「このフォルダを信頼しますか」で止まる

Claude Codeは、初めてのフォルダで起動すると、そのフォルダを信頼するかを聞いてきます（英語で表示されることがあります）。

☐ 対策2で中を見てから「はい」を押す

☐ ダウンロード・ホームのような広いフォルダは、信頼しない

対策4　設定ファイルを時々開く

☐ 週に1回など、日を決めて、2. の10秒チェックの2つ目と3つ目を見る

☐ Claude Codeの中で /status と打ち、読まれている設定ファイルの一覧に、知らないファイルが無いかを見る

☐ 確認を全部飛ばすモードは、ふだん使わない

Claude Code：bypassPermissions（「確認をしないモード」）
Codex：`--dangerously-bypass-approvals-and-sandbox`（`--yolo`）

どちらの会社も、メモに仕込まれた命令への守りにならない、勧めない、という立場です。

対策5　Chromeの「保存場所を確認する」をオンにする

☐ Chromeの右上の「：」→「設定」→ 左の「ダウンロード」→「**ダウンロード前に各ファイルの保存場所を確認する**」をオンにした

☐ 「設定」→「プライバシーとセキュリティ」→「サイトの設定」→「追加の権限」→「**自動ダウンロード**」で、サイトが勝手にファイルを落とせないようにした

確認の画面が出ても、そのまま「保存」を押せば置かれてしまいます。**これは保険です。主役は対策1です。**

補足　自分のメモを先に置いておく

ダウンロードのフォルダに、中身が空（または自分で書いた）CLAUDE.md・AGENTS.md を先に置いておくと、あとから落ちてきた同じ名前のファイルは別の名前で保存され、AIに読まれにくくなります。**2つ目の守りにはなりますが、ダウンロードで起動しないのが先です。**

4.　起動する前の確認表（コピーしてパソコンの横に）

	確認すること
☐	いまいるフォルダは、自分で作った作業フォルダか
☐	フォルダに、覚えのない CLAUDE.md・AGENTS.md・.claude・.codex・.mcp.json は無いか
☐	「このフォルダを信頼しますか」の画面で、中を確かめてから「はい」を押したか
☐	確認を全部飛ばすモードで起動していないか

5.　見つけたときの手順

覚えのない行やファイルを見つけたら、**消すだけで終わりにせず、ログインの鍵を入れ替えます。**

☐ 1.　Claude Code・Codexをいったん閉じる

☐ 2.　設定ファイルを別の名前で保存してから、覚えのない行を消す
例：settings.json を settings_控え.json としてコピーしておく

☐ 3.　ダウンロードのフォルダの、覚えのない CLAUDE.md・AGENTS.md を消す

☐ 4.　ログアウトして、ログインし直す（ログインの鍵の入れ替え）
Claude Code：Claude Codeの中で /logout と打ち、そのあと /login

Codex：codex logout のあと codex login

☐ 5. Anthropic・OpenAIのアカウントの画面で、使っていない鍵（APIキー）を消す

不安なときは、パソコンに詳しい人か、お使いのサービスのサポートに相談してください。

6. いつ・何を見たかの記録

見た日	ダウンロード	Codexの設定	Claude Codeの設定	メモ
/	☐ 問題なし	☐ 問題なし	☐ 問題なし	
/	☐ 問題なし	☐ 問題なし	☐ 問題なし	
/	☐ 問題なし	☐ 問題なし	☐ 問題なし	
/	☐ 問題なし	☐ 問題なし	☐ 問題なし	

出典

- Forever Security「Markdown Is the New Executable: Attackers Are Dropping Agent Instructions Into Your Downloads Folder From the Web」（2026年10月22日） https://forever.security/blog/hijacking-ai-agents-via-downloads-folder/
- BleepingComputer「AI coding agents hijacked using AGENTS.md file in Downloads folder」（2026年10月25日） https://www.bleepingcomputer.com/news/security/ai-coding-agents-hijacked-using-agentsmd-file-in-downloads-folder/
- Watkin Lab（@0xWatkin）のXの投稿（2026年10月26日） https://x.com/0xWatkin/status/2114538646254633135
- Claude Code 公式ドキュメント（セキュリティ・設定・権限） https://code.claude.com/docs/en/security
- Codex 公式ドキュメント（承認とセキュリティ） https://learn.chatgpt.com/docs/agent-approvals-security
- Chrome ヘルプ「ファイルをダウンロードする」 https://support.google.com/chrome/answer/95759

このマニュアルの内容は2026年10月29日時点のものです。