

# ハーネスエンジニアリング入門

AIエージェントを安全に働かせる6つの部品（指示書・スキル・権限・関所・記憶・点検）と、事故から生まれたルール

更新日：2026-09-30（2026年9月時点の情報） | HUMAC 伊藤晴通

## ハーネスとは「馬具」のこと

AIエージェントは力の強い馬。放っておくと、頼んでいない相手にメールを送ったり、消してはいけないファイルを消したりする。**手綱と装具（ハーネス）を付けて、決めた道を安全に走らせる仕組みづくり**が「ハーネスエンジニアリング」。AIを賢くするより、**AIが間違えても事故にならない形**を作ることが仕事の中心になる。

## ハーネスの6つの部品

| 部品       | たとえ              | 中身                            | Claude での実体              |
|----------|------------------|-------------------------------|--------------------------|
| ① 指示書    | 新人に渡す「職場のルールブック」 | この仕事の目的、置き場所、やってはいけないこと、言葉づかい | プロジェクトごとの CLAUDE.md      |
| ② スキル    | 業務ごとの手順書         | 「この仕事はこの順で、この資料を見て、この形で仕上げる」  | スキル（合言葉で呼び出す）            |
| ③ 権限     | 渡す鍵の範囲           | 読んでよい場所・書いてよい場所・使ってよい道具       | 許可設定（ファイル・コマンド・外部サービスごと） |
| ④ 関所（承認） | 判子が要る書類          | 送信・公開・削除・支払いの前は、必ず人が確認する      | 承認の確認画面、フック（自動チェック）      |
| ⑤ 記憶     | 引き継ぎノート          | 前回の失敗・決まったこと・好みを次回に持ち越す       | メモリ、決定ログ、STATUS          |
| ⑥ 点検     | 定期監査             | ルールと実態がずれていないか、定期的に確かめる       | 週次の自動点検（定期実行）            |

## 事故から生まれたルール（講師の実例）

ハーネスは机上で考えるより、**小さな事故を1つずつルールに変えていく**と強くなる。

| 起きたこと                                 | ルールにしたこと                                  | 部品 |
|---------------------------------------|-------------------------------------------|----|
| AIを使うための鍵（APIキー）が公開の場所に出て、約6万8千円の不正請求 | 鍵はコードにもチャットにも書かない。鍵は専用の金庫に集め、プログラムは金庫から読む | ①③ |
| AIが「不要」と判断した経理書類を消してしまった              | AIの判断で消さない。疑わしいものは「退避フォルダ」へ移し、最後は人が消す     | ③④ |
| クライアントのブログを、確認前に公開しかけた                | 公開ボタンはクライアント本人が押す。AIは下書きまで                | ④  |
| AIが古い知識で製品名を「訂正」して間違えた                | 製品名・料金は、直す前に必ずWebで確認する                    | ①⑤ |
| 同じ説明を何度もさせられた                         | 一度決めたことは記憶に残し、蒸し返さない                      | ⑤  |

## 作り方の手順

- 1

**仕事を1つ選ぶ**  
毎回同じ手順で、失敗しても取り返しがつくもの
- 2

**指示書を書く**  
目的・置き場所・禁止事項を1枚に。長くしない
- 3

**関所を決める**  
「外に出る操作」と「消す操作」は必ず人が押す
- 4

**スキルにする**  
手順を言葉にし、合言葉で呼べるようにする
- 5

**点検を回す**  
週1回、ルールと実態のずれを確かめ、事故はルールに変える

**任せるほど、関所が大事になる**

AIエージェントが自分で動くようになるほど、「権限のはみ出し」が起きやすくなる。便利さと事故は同じ入口から来る。**最初に決めるのは「何をさせるか」ではなく「どこで止めるか」。**

## 関連資料

- E2 プロジェクトのフォルダ構成と管理 … 指示書と記憶の置き場所
- E3 ローカルLLMと「情報を外に出さない」設計
- C3 事例：この資料ライブラリの自動化・スキル化 … ハーネスの実物
- D1 信号機ルール