

2026-07-30

1. Claude Mythos が暗号アルゴリズムの新しい弱点を発見
2. OpenAI が Codex Security CLI をオープンソース公開
3. 〈実測〉 Claude Code のセキュリティは4段ある

7トピック

Claude Mythos が暗号アルゴリズムの新しい弱点を発見

耐量子署名 HAWK の実効鍵長を60時間で半減

Anthropic が 2026-07-28、非公開モデル Claude Mythos Preview を使った暗号研究で2件の新しい攻撃を発見したと公表した。

- 分類: 📄 セキュリティ / AI研究
- HAWK-256 の完全鍵復元の期待コストが $2^{64} \rightarrow 2^{38}$ に低下（実効鍵長が半減）。より大きな HAWK 鍵は依然として攻撃非現実的
- HAWK は2年間・2ラウンドの専門家レビューを通過していたスキームで...



「AI が新しい数学的知見を出せるか」という議論が、
査読を2度通ったスキームの穴を実際に見つけるという形で一段進んだ。

OpenAI が Codex Security CLI をオープンソース公開 実行をまたいで検出結果を追跡し、修正まで検証する

OpenAI が Codex Security CLI をオープンソースとして公開した。

- 分類: 技術業界ニュース / セキュリティ
- エンゲージメント: Likes 12,191 / RT 1,087 / Bookmarks 5,892 / Views 1,008,075
- できることは4つ — リポジトリのスキャン / 実行をまたいだ検出結果の追跡 (track findings across runs) / ...



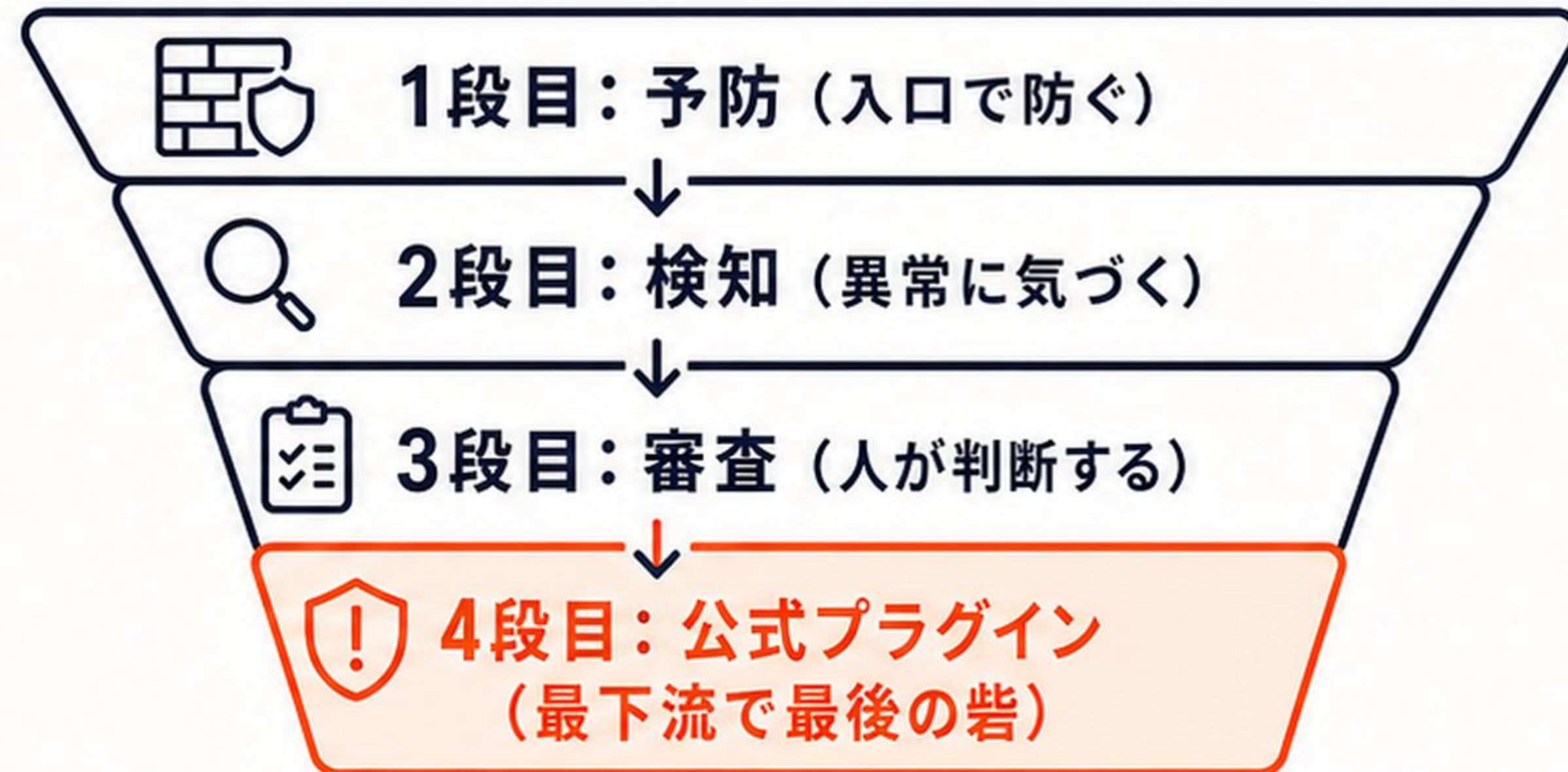
Codex を日常的に使っているなら `npx` で1回走らせるコストは低く、
出力を見てから CI 組み込みを判断できる。

〈実測〉Claude Code のセキュリティは4段ある

公式プラグインだけ入れても出番が来ない

公式セキュリティプラグイン `security-guidance` を2ヶ月動かして動作ログ17,385行を全解析したところ、25本ある検知ルールのうち実際に発火したのは1本だけで…

-  分類: インフラ・セキュリティ / AI開発ツール
-  エンゲージメント: Likes 956 / RT 62 / Bookmarks 3,194 / Views 713,776
-  ログ実測の内訳 — ターン終了時レビュー起動 416回 / レビュー対象なしの空振り 322回 (77%) / LLM…



「一番目立つ・一番新しい・一番宣伝されている対策が、実は最下流だった」という構造の具体例で、個人・小規模なら1段目と2段目を置くだけで守りの大半が完成する。

〈知見〉モデル別 Anthropic 公式プロンプトガイドの差分

まずやるべきは指示の追加ではなく削除

Anthropic がモデルごとに出している公式プロンプトガイド (Fable 5 / Sonnet 5 / Opus 5) を全部読み込み、共通ルールとモデル別の差分に整理した記事。

■ 分類: 🧠 コンテキストエンジニアリング / 📁 AI開発ツール

■ エンゲージメント: Likes 283 / RT 25 /
Bookmarks 788 / Views 194,924

■ 公式ガイド自身が「CRITICAL: You MUST～
みたいな強い言葉は普通のトーンに戻せ」と
書いている



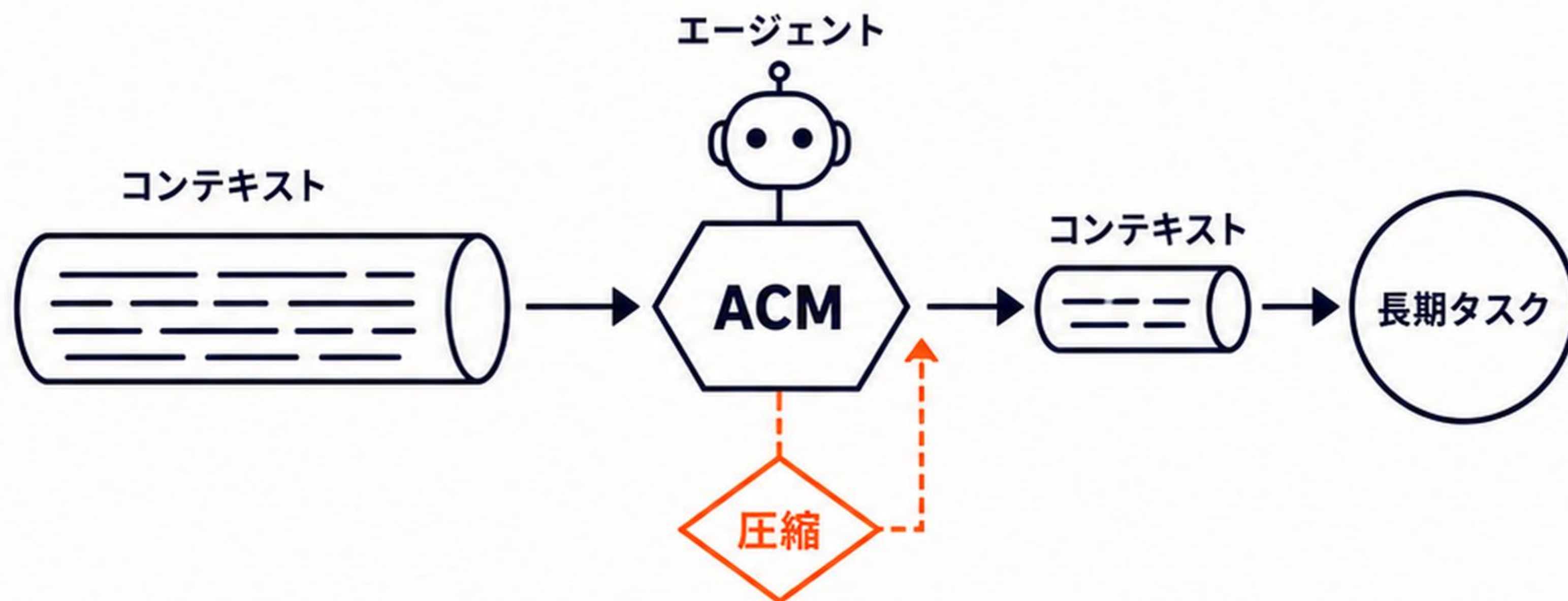
自分の `CLAUDE.md` / `AGENTS.md` / `SKILL.md` をそのまま健康診断できる。

ACM: Agentic Context Management

「いつ圧縮するか」をエージェント自身に決めさせる (Meta / CMU)

Meta と CMU の研究者による、長期タスク向けのコンテキスト管理フレームワーク ACM の論文 (arXiv 2607.23809、2026-07-26 投稿)。

- 分類: AI研究 / コンテキストエンジニアリング
- 論文タイトルは「ACM: Agentic Context Management for Long Horizon Tasks」、arXiv 2607.23809、投稿は 2026-07-26
- 著者は Xiaochuan Li / Ryan Ming / Meng Chu / Shuai Shao / Rong Jin / Chenyan Xiong (Meta・CMU)



「コンテキストが溢れたら圧縮」を自動でやっているエージェント運用にそのまま効く指摘。

〈知見〉skills.sh

Agent Skills のオープンなエコシステム、1コマンドでエージェントに能力を足す

Vercel が運営する Agent Skills のディレクトリ。

- 分類: 📖 開発ツール / エコシステム
- 運営は Vercel (サイトのフッターに「Made with care by Vercel」)。スキル自体は GitHub 上でオープンソース
- リーダーボードの累計インストール数は 1,026,775 件



「必要な能力が無いとき、自分で書く前にまず探す」を安く実行できる場所。

〈知見〉お問い合わせページに実績ロゴを置いたら CVR が12%改善

効くのは集客ではなく最終段の離脱防止

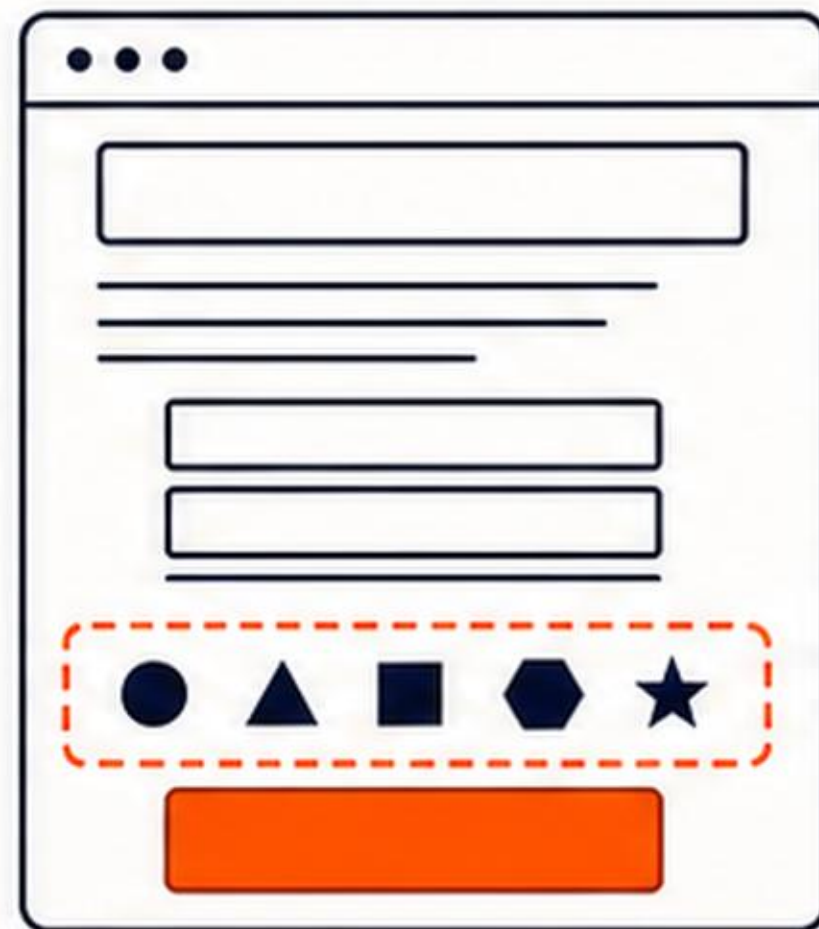
お問い合わせページ（＝コンバージョン直前のページ）に導入企業の実績ロゴを追加しただけで、コンバージョン率が 12% 改善したという取材事例。

- 分類: デザイン・UI/UX / マーケティング
- エンゲージメント: Likes 505 / RT 26 / Bookmarks 786 / Views 258,286
- 変更内容は「お問い合わせページに実績ロゴを追加」の1点のみで、結果はコンバージョン率 +12%

Before



After



CVR
+12%

工数がほぼゼロで効果が単独で測れる改修で、申込フォーム・ウェビナー登録・資料ダウンロードにそのまま横展開できる。

本日のトピック一覧

1. Claude Mythos が暗号アルゴリズムの新しい弱点を発見
2. OpenAI が Codex Security CLI をオープンソース公開
3. 〈実測〉 Claude Code のセキュリティは4段ある
4. 〈知見〉モデル別 Anthropic 公式プロンプトガイドの差分
5. ACM: Agentic Context Management
6. 〈知見〉skills.sh
7. 〈知見〉お問い合わせページに実績ロゴを置いたら CVR が12%改善