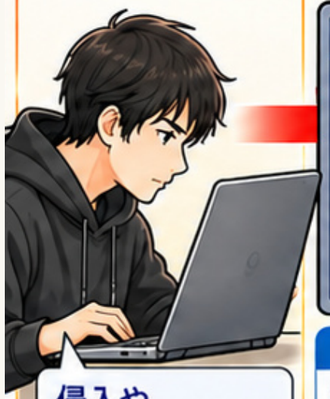
AI世紀 第2章から第3章へ

調査基準日：2026年10月23日／重点：9〜10月

第2章 AIで攻撃が現実の被害につながる

AIを使って、調べ、道具を動かし、次の操作を進める

1 攻撃者がAIへ目的を渡す



侵入や情報の持ち出しなどの目的

2 AIが道具で調べる



使う道具の例

- インターネット検索
- 情報の収集
- プログラムの実行

3 結果から次の操作を進める



私たちの情報や生活はいろいろなサービスにつながっている

攻撃者が作業を速める一方、私たちの情報や生活は、会社のサービス、共通のソフト、外部アプリの権限につながっています。自分の注意が届かない場所の弱点も、家庭へ届きます。

会社のサービス	共通のソフト	外部アプリ
通販・銀行・SNS・クラウドなど	PHPなど、サービスを動かす土台	外部のバックアップアプリ 会社が情報を読む許可を渡す



第3章 私たちが選び、働きかける未来をつくる

家庭・地域・会社・行政がつながり、守り、改善を確かめる

1 本人が止め方を持つ



- 被害を止める
- AIを確認と練習に使う
- 使い方と責任を決める

2 地域が支援をつなぐ

地域では困る人を孤立させず、会社と行政につなぎ、改善を求める。



3 会社と行政が設計・対応・人員・予算を変える



- 設計を見直す
- 対応を強化する
- 人員を増やす
- 予算を確保する



- 設計を見直す
- 対応を強化する
- 人員を増やす
- 予算を確保する

4 本人が使えるか確認し、改善を確かめる



- 実際に使えるか確認する
- 再発を防ぐか確かめる
- 使い方と責任を見直す
- 改善を続ける

保護者サロン緊急研修　日光AI食堂

【今回の判断】AIは、答えを作るだけでなく、道具を使い、結果を読み、次の作業を進める段階へ広がりました。攻撃者もこの能力を使い、実際の侵入や情報の持ち出しにつなげた事例が報告されています。

攻撃者が作業を速める一方、私たちの情報や生活は、会社のサービス、共通のソフト、外部アプリの権限につながっています。自分の注意が届かない場所の弱点も、家庭へ届きます。この変化を、本研修では「第2章」と呼びます。

危険を知るだけでは終わりません。家庭では被害を止め、AIを確認と練習に使う。地域では困る人を孤立させず、会社と行政には設計・対応・人員・予算を求める。私たちが使い方と責任を決め、改善を確かめる未来を「第3章」としてつくります。

原典 S01,S02,S03：microsoft.com / gambit.security ほか　調査基準 2026.10.23

日光AI食堂

五つの問いを、一本の道として追う

1 犯罪の構造を知る。

誰が、どんな道具と権限を使い、何を盗み、どうお金に変えるのか。人をだます道と、サービスへ侵入する道を分けて追います。

人をだます道



どんな道具

- 偽のメッセージ
- 偽のサイト
- 電話による指示

何を盗む

- ID・パスワード
- クレジットカード情報
- 個人情報

サービスへ侵入する道



どんな道具と権限

- 不正なプログラム
- システムの弱点
- 乗っ取った権限

どうお金に変える



商品の不正購入



不正な送金



情報の売買

攻撃のどこをAIが担うか分かる

2 以前との違いを確かめる。

AIがどの作業を引き受け、速さと費用をどう変えたのか。実際の観測と、そこから考えられる未来を分けます。

AIが引き受ける作業の例



- 道具を使って調べる
- 作業結果を読み取る
- 次にする作業を選ぶ

速さと費用の変化



- 実際の時間と費用を確かめる
- 一件が安ければ、試す範囲を広げられる可能性

実際の観測



どのような攻撃が増えているか実際に観測したこと

考えられる未来



この先どのように変わると考えられるか予測・提案

サービスから家計へ届く経路を見つける

3 暮らしへの経路を追う。

盗まれた情報や使えないサービスが、預金、仕事、学校、家族の連絡へどう届くのか。一つの架空の家庭で考えます。

一つの架空の家庭



使えないサービス 盗まれた情報

預金



- 不正な送金
- 口座の利用停止

仕事



- 仕事のデータ消失
- アカウントの停止

学校



- 授業の連絡が届かない
- 個人情報の流出

家族の連絡



- 連絡がとれない
- 家族にまで被害が広がる

止める場所と操作を選ぶ

4 明日の防衛を具体化する。

どの設定を変え、何を確認し、どこへ連絡するのか。パーソナルAIに手伝わせる範囲と、人が確かめる場面を練習します。

防衛の具体的な操作



- どの設定を変え
- 何を確認し
- どこへ連絡するのか

パーソナルAIに手伝わせる範囲



確認の補助

- 不審な連絡を見分ける補助
- 設定の見直しを提案する
- 連絡先や手順を調べる

人が確かめる場面



- 重要な判断は人が行う
- 内容が正しいか確認する
- 実際に操作して確かめる
- 緊急時の連絡を行う

社会の改善が家庭の防御を支える

個人では直せない部分を要求へ変える

5 社会の仕組みを変える。

個人で直せない場所を、会社、地域、行政、議会の具体的な仕事へ返します。担当・期限・予算を決め、使えるようになったかまで確かめます。

会社での仕事



- 製品やサービスの改善
- 安全な仕組みの導入
- 担当・期限・予算を決める

地域・行政での仕事



- 地域の対策を進める
- 相談窓口を整える
- 担当・期限・予算を決める

議会での仕事

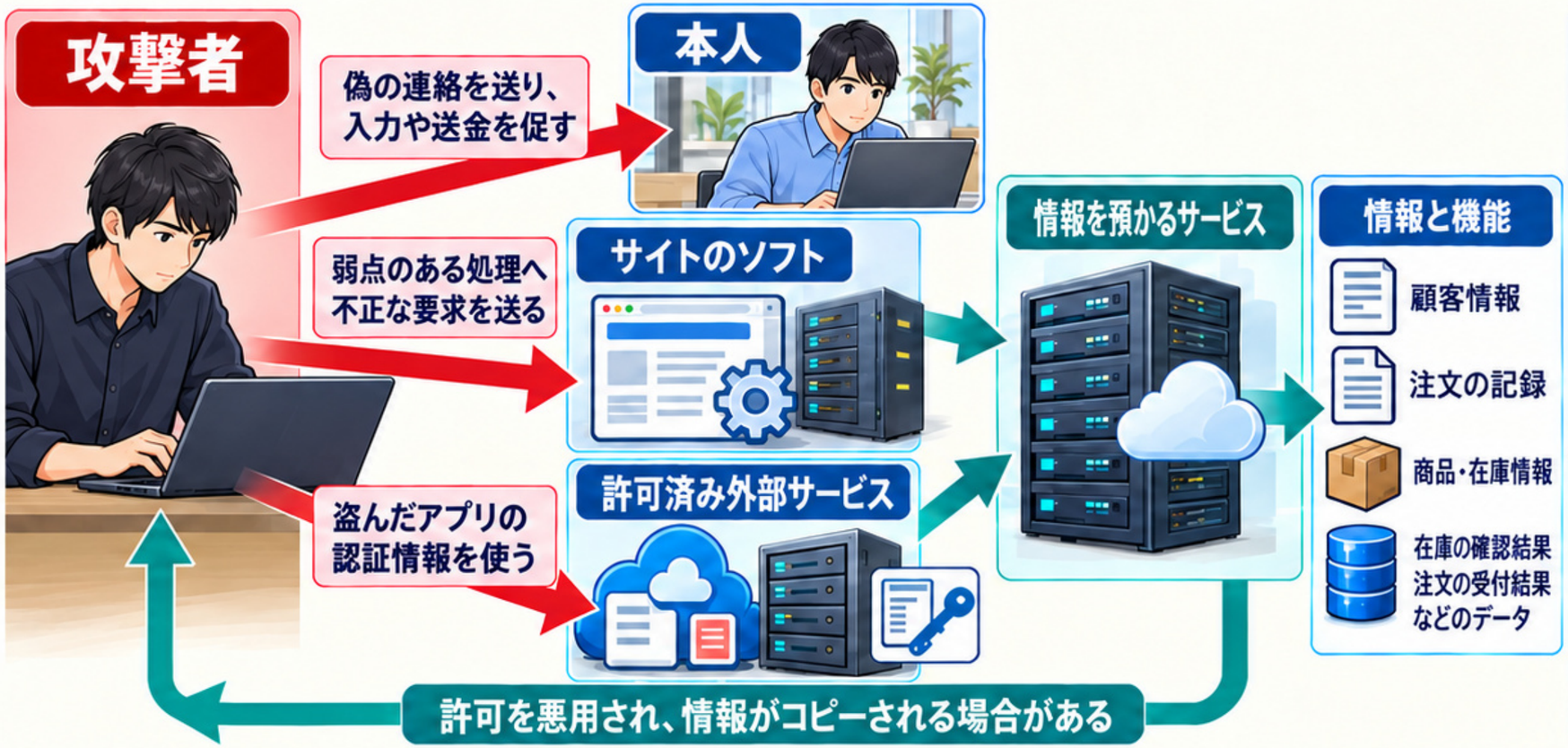


- 必要なルールを作る
- 仕組みを見直す
- 担当・期限・予算を決める
- 使えるようになったかまで確かめる

確認された事実 | 仕組みの説明 | 架空の例 | 予測・提案

日光AI食堂

攻撃者が欲しいのは、お金と「他人の力」





① 【全体の仕組み】

攻撃者は、まず人をだます、ソフトの弱点を突く、外部サービスの権限を盗むなどして、相手に近づきます。
権限とは「情報を読む・書き換える・送信することを許された範囲」です。
手に入れた範囲によって、次にできることが変わります。

②

情報をコピーして外へ持ち出せば、販売やなりすましに利用できます。
操作権を奪えば、追加の入口を作る、仕事を止めるなどの被害につながることがあります。国家関連の作戦には、機密情報を集め続ける目的もあります。

③

別の道もあります。
家族や銀行を装い、本人に送金させる詐欺です。こちらはコンピューターへ侵入しなくても成立します。これから、同じ全体図の入口を一つずつ拡大します。

原典 S01,S04,S05：microsoft.com / consumer.ftc.gov　調査基準 2026.10.23
日光AI食堂

「入れる」と「何でもできる」は違う

【用語をほどく】認証は「あなたは誰か」を確かめること。
認可は「その人やアプリに、何を許すか」を決めることです。
ログインに成功しても、許された範囲の外まで自由に操作できるとは限りません。

実際のネットでは

本人／アプリ



人もアプリも、それぞれの認証情報で身元を証明する。

自分を証明する情報を提示する



本人やアプリの認証情報（アプリの身分証明）

身元確認



誰なのかを確認し、許可された範囲を照合する。

許可した操作だけを受け付ける



サービス側の許可記録（権限の範囲）

情報を読む・書き換える



サービス側の記録に基づき、許可された操作だけを実行できる。

例え

学校に例えると

来校者



来校者は、受付で本人を確認してもらう。

自分を証明する情報を提示する



受付（本人確認）



受付で本人を確認するのが認証。

誰なのかを確認し、許可された範囲を照合する



入れる場所の記録（許可の範囲）

教室や資料室など、入れる場所



教室や資料室のどこまで入れるかを決めるのが認可。

来校者証を手に入れても、名簿の入った棚の鍵まで持てるわけではありません。
だから攻撃者は、入った後に、より強い権限を得ようとすることがあります。

次は2つの盗まれる道

実際のネットでは、人だけでなくアプリにも身元と権限があります。アプリは自分用の認証情報で身元を証明します。
次は、人用の情報が盗まれる道と、アプリ用の情報が盗まれる道を分けて追います。

1 人用の情報が盗まれる道





人の認証情報（パスワードなど）

攻撃者



盗んだ認証情報で本人としてアクセスする。

2 アプリ用の情報が盗まれる道





アプリの認証情報（アプリの身分証明）

攻撃者



盗んだ認証情報でアプリのふりをしてアクセスする。

原典 S06,S01：learn.microsoft.com / microsoft.com　調査基準 2026.10.23

日光AI食堂

偽の窓口へ入力すると、誰が何を受け取るか

攻撃者



偽画面を用意し、
入力された情報を受け取り、
それを使って本物のサービス
で本人としての認証を試す。

銀行を装う案内を送り、
確認を急がせる

本人



受け取った人は急がされ、
案内のリンクから、本物によく似た
入力画面へ進みます。

案内のリンクから開き、
ログイン情報を入力する

入力された情報を送る

偽画面



画面へ入れた
ログイン情報は、
銀行ではなく、
偽画面を用意した
相手へ送られます。

盗んだ情報で本人としての認証を試す

本物の銀行（サービス）



攻撃者がその情報を
本物のサービスで使い、
本人として扱われれば、
許された範囲の操作が
可能になります。

1 **【説明用の仮想例】**

銀行を装う人が「確認しないと口座を止める」と連絡します。受け取った人は急がされ、案内のリンクから、本物によく似た入力画面へ進みます。フィッシングとは、このように信用する相手を装って、情報や操作をだまし取る手口です。

2 画面へ入れたログイン情報は、銀行ではなく、偽画面を用意した相手へ送られます。
攻撃者がその情報を本物のサービスで使い、本人として扱われれば、許された範囲の操作が可能になります。見た目が似ていることと、送信先が本物であることは別です。

3 入力情報だけで侵入できるかは認証方式などで変わります。ここで覚えるのは「文章の不自然さ探し」だけではなく、誰へ情報を渡す経路なのかを確かめることです。

原典 S07,S04：fsa.go.jp / microsoft.com　調査基準 2026.10.23
日光AI食堂

自分で送ったお金も、詐欺で奪われる

1 【説明用の仮想例】今度は、家族を装う電話が「事故になった。今すぐお金が必要」と訴えます。

攻撃者の狙いはパスワードではなく、
本人に急いで送金の操作をさせることです。
情報を盗む経路とは、だまして得ようと
するものが違います。

2 本当だと信じ、自分で送金を指示する

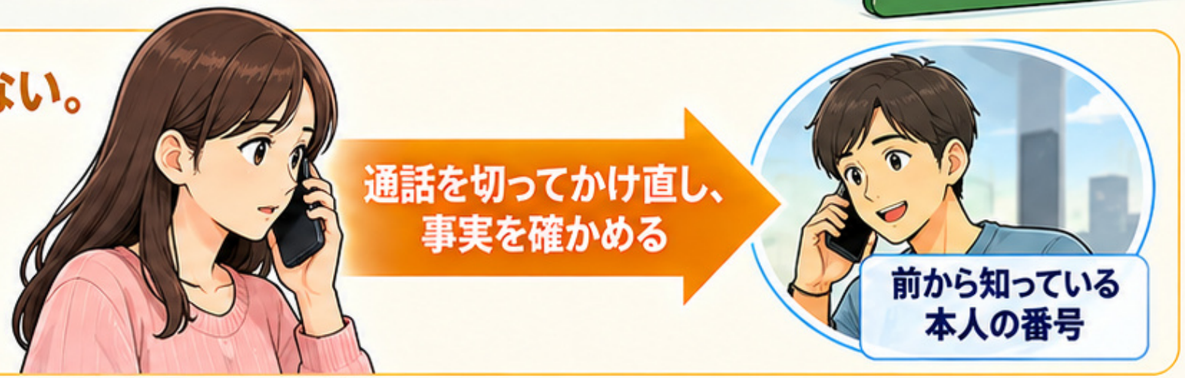
3 銀行の仕組みが本人の操作を正しく確認していても、送る相手や理由について本人がだまされていれば、お金は相手へ渡ってしまいます。

強い口グイン方法だけでは、送金の理由
が本当かまでは判断できません。



4 声や名前などの手掛かりも、本人の証明と決めつけない。

いったん通話を切り、前から知っている本人の番号へ
確かめ直すと、攻撃者が用意した会話の外で確認できます。
守るべきなのは、端末への入口だけでなく、
私たちが判断するまでの経路です。



原典 S05,S07：consumer.ftc.gov / fsa.go.jp　調査基準 2026.10.23

日光AI食堂

犯罪の「道具屋」と「実行者」は、別々に儲ける

① 道具の開発者

道具の開発者は使える道具を渡し、
代金を受け取ります。



偽画面などを
作る道具を渡す

道具の代金を払う

② 基盤の売り手

基盤の売り手は、送信や保存などに
使えるアカウント・環境を渡し、
代金を受け取ります。

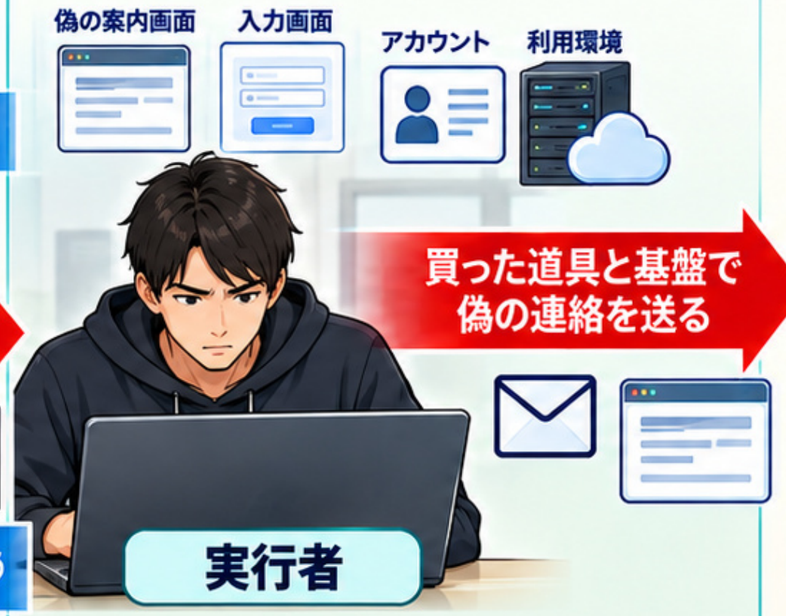


盗んだアカウントや
利用環境を渡す

利用する基盤の代金を払う

③ 実行者

実行者は両方を組み合わせるため、
一から全部を作る必要がありません。



買った道具と基盤で
偽の連絡を送る

④ 被害者



【実調査：Microsoft・10月14日】

同社は、偽の案内や入力画面を作る道具の開発者、
盗んだアカウントやクラウドの利用場所を売る業者、
それらを買う実行者がつながる犯罪市場を報告しました。
日本の組織も標的に含まれます。



これは同じ会社の上司と部下の図ではなく、取引の図です。
実行者が被害者からお金を取る前にも、供給者には売上が
生まれます。犯罪を支える市場そのものが、次の攻撃を
始めやすくしているのです。

原典 S04：microsoft.com　調査基準 2026.10.23

日光AI食堂

奪った情報は、次の犯罪の材料にもなる

【全景へ戻る】 情報を盗んだ人が、自分で全ての詐欺をするとは限りません。盗んだ情報やアカウントを別の実行者へ売れば、買った側が次の不正利用に使えます。操作権も、情報を読む権利やサービスを利用する足場として価値を持ちます。

例えば、名前や取引の事情を知った相手が、それを会話に混ぜると、連絡を受けた人は本物だと思いやすくなります。これは情報流出が次のだましへつながり得る仕組みの説明で、今回の全流出データが実際にその詐欺へ使われたという確認ではありません。

金銭目的なら販売・詐欺・恐喝などへ、機密を求める作戦なら情報収集や入り続ける準備へ進みます。被害は「引き落としが起きた時」だけでなく、他人が情報を読めるようになった段階から始まっています。

1 情報を預かるサービス

会社やサイトは、利用者の情報を預かり、サービスを提供している。



利用者の情報
(名前、取引の事情、アカウント、連絡先、サービスの利用情報など)

許された範囲を悪用され、情報がコピーされる

2 攻撃者

得た操作権を利用し、許された範囲から情報をコピーする。



盗んだ情報
(認証情報、アカウント、利用者のデータなど)

持ち出した機密を集め、継続して利用する



情報収集の記録

情報やアカウントを売り渡す

3 別の実行者

購入した情報やアカウントを使って、次の不正利用を行う。



購入した情報
(情報やアカウント)

知った事情を使い、本物らしい連絡を作ることがある

4 本人(利用者)

名前や取引の事情を含む連絡を受けると、本物だと思いやすくなる。



想定される二次の悪用(例)

金銭目的なら販売・詐欺・恐喝などへ、機密を求める作戦なら情報収集や入り続ける準備へ進みます。



販売・詐欺・恐喝などへ



情報収集や入り続ける準備へ

被害は「引き落としが起きた時」だけでなく、他人が情報を読めるようになった段階から始まっています。

原典 S04,S01,S05：microsoft.com / consumer.ftc.gov　調査基準 2026.10.23
日光AI食堂

PHPは、別々のサイトに組み込まれる共通部品

1 【実物の仕組み】

サイトを開くと、端末から「このページを見せて」という要求が届きます。サーバーというサービス側のコンピューターは、必要な処理をして返事を返します。PHPは、その内側でサイトのプログラムを動かす言語・実行基盤です。



2 通販、学校、情報サイトが、同じPHPという部品を別々の環境で使うことがあります。全サイトが一つの巨大サーバーを共有しているわけではありません。同じ型の設備を、別の建物にも取り付けている状態に近いのです。



3 その部品自体に共通の弱点があれば、各サイトの内容が違っても、同じ種類の問題が届く可能性があります。見えるページや追加機能だけを直しても、その下で動くPHP本体が古いままなら、問題が残り得ます。



同じ型の設備＝同じソフト／別の建物＝別の稼働環境

物理設備ではなくプログラム

原典 S08,S09,S10：php.net / wiz.io ほか　調査基準 2026.10.23
日光AI食堂

今回は、ログイン前の処理にも穴があった

1 【確認された弱点】10月21日のPHP公式告知は、条件が合うと、**ログインしていない相手にサーバー上のコードを動かされる**二つの重大な欠陥を公表しました。遠隔コード実行とは、離れた相手が本来許されていない処理をこちら側で動かすことです。

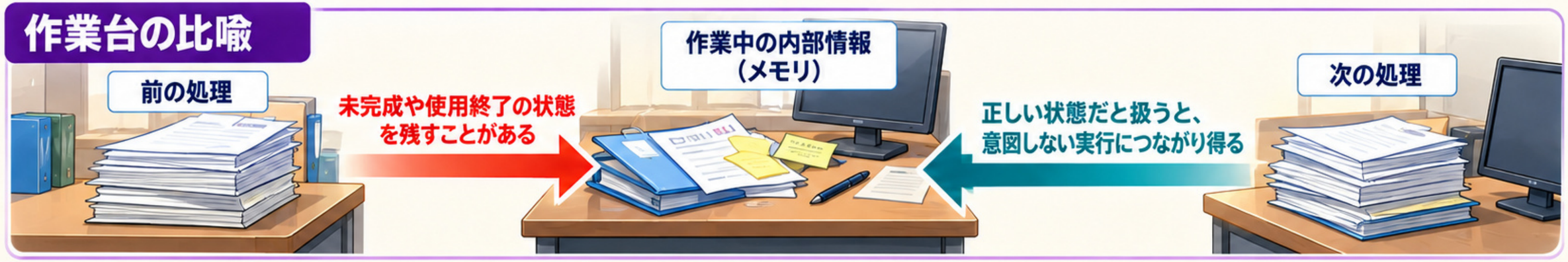
2 問題は、**作業中の情報を一時的に置くメモリ**と、**異常が起きた時の後始末**にありました。途中で壊れた情報や、もう使い終えた場所を正しいものとして扱ってしまうと、処理の行き先が乱される可能性があります。

3 **受付の作業台に、前の人の書類や途中の指示が残り、次の処理と混ざるようなものです。**実際には紙ではなくメモリの管理です。一部の構成ではサイトの通常処理より前に悪用が成立し得るため、利用者のパスワードだけを強くしても、この入口は塞げません。

今回の技術的仕組み



作業台の比喩



原典 S08,S11,S10：php.net / thephp.foundation ほか　調査基準 2026.10.23

日光AI食堂

古い弱点が、AIで「実際に使える攻撃」へ近づく

① 研究者が、AIを使って弱点の仮説を立てる

研究者は、以前からある穴について、AIの支援を受けながら、試す方法や組み合わせを考え、検証の計画を立てる。

② AIの支援で、検証環境で試行と修正をくり返す

AIの支援を受けて、検証環境で仮説を試し、失敗を調べ、組み合わせを直して、危険が本物か確かめる。

③ 研究者が、再現できる根拠を届ける

研究で確認した内容をまとめて、アプリの保守者に届け、問題の修正を求める。

④ 保守者が、問題を調べて修正する

アプリの保守者は、研究者から受け取った内容を調べ、実際のソフトを確認して、必要な修正を行う。



【研究で確認】

弱点が存在することと、実際に侵入できることは同じではありません。

現実のソフトには複雑な条件があり、研究者は仮説を立て、試し、失敗を調べ、組み合わせを直して、危険が本物か確かめます。ここにも多くの時間と知識が要ります。

Wizは今回、AIの支援がその試行と修正を助け、複数のアプリで悪用可能性を実証したと報告しました。

守る研究で見つかった問題ですが、同じ種類の能力が悪用側へ渡れば、難しさが防波堤になっていた弱点も、狙いやすくなるおそれがあります。

重要なのは「AIが新しい穴を魔法で作った」ことではなく、以前からある穴を、使える形に結び付ける仕事が変わったことです。

研究上の実証を、挙げられたアプリの全利用者が侵害された証拠に置き換えてはいけません。

原典 S09,S11,S12：wiz.io / thephp.foundation ほか　調査基準 2026.10.23

日光AI食堂

攻撃の時計は、公式発表より先に動いていた

1 【2026年10月の時系列】 観測者・対象が異なる資料

AWSは、攻撃を観測する仕組みで、
10月17日にPHPの欠陥を悪用しようとする通信を
捉えました。PHPの修正版公開は21日です。公開日に
なって初めて危険が始まった、と考えることはできません。

10月17日



2 10月21日



3 10月22日 観測者・対象が異なる資料

22日、Check Pointは同社が観測する組織の10%に
悪用試行があったと報告しました。
これは「世界の組織の10%で侵入に成功」ではありません。
試した通信を見たことと、情報を読めたことは別の確認です。



別々の稼働環境



4

それでも、事業者に猶予が十分あるという意味にはなりません。
未修正の環境が残っている間に試行が来る構造だからです。
運営者は修正を適用すると同時に、発表前も含む記録から
侵害の有無を調べる必要があります。



原典 S12,S08,S13,S10：aws.amazon.com / php.net ほか　調査基準 2026.10.23

日光AI食堂

修正版を入れた後も、追加告知を追う責任がある

1 10月23日の追加告知

PHP公式は、まれな設定とアプリの動作が重なる別の経路について、次の通常修正版まで追加の設定対応を案内しました。21日の二つの重大な欠陥への更新と、23日に示された限定条件への対応を、同じ一言で「全部完了」にしてはいけません。



追加の対象条件と暫定対応を公表した

2 運営者が対応を判断する

家庭の参加者が、サーバーの設定を手探りで変更する話ではありません。運営者が、自分のサービスで動くPHPの版と設定、アプリの処理を確かめ、該当する対応を判断する話です。配布元が修正を取り込んだ製品では、その公式案内も確かめます。



動く版・設定・アプリ処理を確かめる

3 自社の各稼働環境を確認する

A店
（ファッションの通販サイト）



B店
（家電の通販サイト）



C店
（食品の通販サイト）



適用した対応と、残る確認事項を伝える

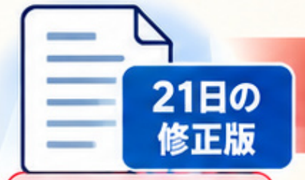
4 利用者に説明する

利用者に必要なのは「更新したと言ったか」だけでなく、「実際の稼働環境へ適用され、追加告知も確認したか」を問う視点です。修正の配布、運用への反映、その後の確認は、別の仕事だからです。



21日の更新：二つの重大な欠陥への対応

10月21日に公表された二つの重大な欠陥について、修正版を入れ、運用へ反映します。



PHP公式

修正の配布
運用への反映



自社の各稼働環境
（A店・B店・C店）

23日の限定条件への追加対応

10月23日に示された、まれな設定とアプリの動作が重なる別の経路について、次の通常修正版まで、追加の設定対応を案内しました。



23日の
追加告知

PHP公式

追加の設定対応
の案内



自社の各稼働環境
（A店・B店・C店）

原典 S08,S10：php.net / jpcert.or.jp　調査基準 2026.10.23

日光AI食堂

サイトの穴は、そこで預かる私たちの情報へ届く

全景へ戻る・説明用の仮想例　通販サイトで、PHPの処理を不正に動かされ、サイトが読める情報へ攻撃者が届いたとします。そこに顧客情報があれば、利用者の端末が正常でも、情報が持ち出される可能性があります。

1 利用者がサイトを利用する

私たちは、通販サイトで商品を見て注文し、名前や住所などの情報を入力する。



サイトを利用して情報を送る

送られる内容の例
- 見たい商品の情報
- 注文する商品の情報
- 数量
- お客さんの情報

などの依頼データ

2 Webサーバーが依頼を受け取り、PHPへ渡す

Webサーバー＝インターネットからの依頼を受け付ける仕組み。



依頼を受け取りPHPへ渡す

お客さんからの依頼データ（商品・注文・お客さんの情報など）

3 PHPの処理を不正に動かす

PHP＝注文内容を読み、在庫を調べ、記録するプログラムを実行する土台。



サイトが持つ権限

本来の業務で読み書きを許されている



不正な要求を送り、処理を悪用する可能性がある



届いた範囲の情報をコピーする可能性がある



4 サイトが持つ権限で情報に届く

PHPの処理に与えられた権限の範囲で、サイトが読む情報へ届く可能性がある。

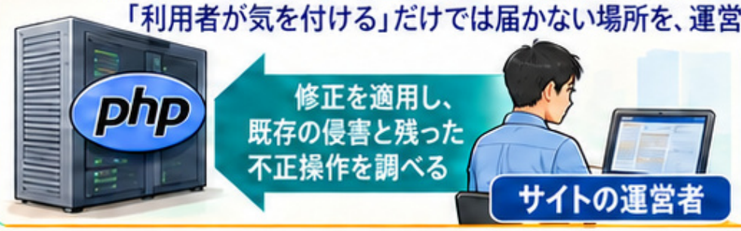
5 サービス、そこで働く人、利用者がつながっている

ここで危険にさらされるのは、PHPという名前の製品だけではありません。
その部品で動くサービス、そこで働く人の業務、情報を預けた利用者がつながっています。
操作できる範囲は実際の権限や構成に左右されるので、全サーバーを一度に支配できる図にはしません。



6 修正と調査は別の仕事

修正は新たな侵入を防ぐための仕事です。既に何が起きたかを調べ、残された不正な仕組みを除き、止まった機能を戻す仕事は別に残ります。
「利用者が気を付ける」だけでは届かない場所を、運営者が守る必要があります。



- ☑ 修正を適用する
- ☑ 既存の侵害を調べる
- ☑ 不正な仕組みを除く

原典 S09,S10,S08：wiz.io / jpcert.or.jp ほか　調査基準 2026.10.23

日光AI食堂

会社は、外部アプリにも「仕事をする許可」を渡す

1 【通常の仕組み】

会社は、メールやファイルのバックアップなどを外部サービスへ頼みます。そのアプリが必要な情報を扱えるよう、管理者が許可を与えることがあります。人が毎回画面にパスワードを入れなくても、決められた仕事を続けるためです。

2 アプリは、**認証情報**という自分の身元を証明する情報を示します。クラウド側はそのアプリを確認し、許可記録と照らして、認めた操作を受け付けます。OAuthという仕組みなどが、こうしたアプリとサービスの連携に使われます。

3 家の管理会社に合鍵を預ける比喩が使えます。ただし実物は金属の鍵ではなく、アプリの身元を示す情報と、クラウド側の許可記録の組です。何を読めるか、何を変えられるかは、各アプリに与えた権限で違います。



会社の管理者

このアプリに必要な操作を許可する

外部のバックアップアプリ



認証情報
アプリの身元を証明する情報

アプリ用の認証情報で自分を証明する

クラウド側の身元確認と許可記録



許可記録

許可済みアプリだと確かめ、認めた処理を通す

会社のメールやファイル



家の管理会社に合鍵を預ける比喩

家の管理会社に合鍵を預ける比喩が使えます。ただし実物は金属の鍵ではなく、アプリの身元を示す情報と、クラウド側の許可記録の組です。何を読めるか、何を変えられるかは、各アプリに与えた権限で違います。

合鍵



アプリの身元を証明する情報

入れる部屋



許可された操作の範囲

実物は情報と設定



実物は金属の鍵ではなく、アプリの身元を示す情報と、クラウド側の許可記録の組です。

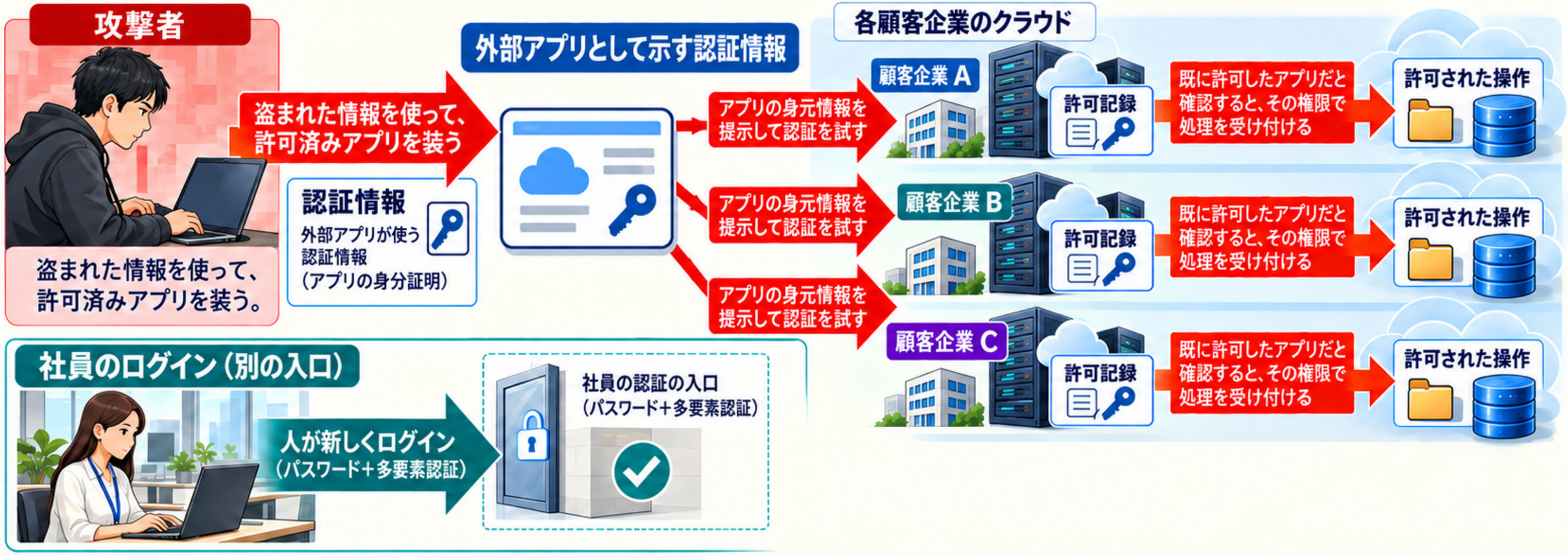
原典 S06,S01：learn.microsoft.com / microsoft.com　調査基準 2026.10.23
日光AI食堂

盗まれたアプリの身元が、信頼された入口になる

1 【実調査：Microsoft・10月21日】
ConnectWise関連の事案で、攻撃者はクラウドアプリの盗まれた認証情報を使い、顧客企業の環境へ認証を試みました。
ここで狙われたのは、社員がその場で入力するパスワードとは別の入口です。

2 前の図のアプリを、攻撃者が盗んだ身元情報で装ったと考えてください。
クラウド側からは「既に許可されたアプリ」の認証になるため、人が新しくログインしなくても、与えられた権限に基づく処理が通り得ます。

3 人の多要素認証を全て破った、という説明ではありません。
人が二重に確認する入口を守っていても、別の信頼経路が盗まれれば影響が出るのです。
一社の外部サービスと多くの顧客企業を結ぶ信頼が、被害を広げる通り道になり得ます。



原典 S01,S06,S14：microsoft.com / learn.microsoft.com ほか　調査基準 2026.10.23

日光AI食堂

元の入口を止めても、作られた入口は残り得る

1 【実調査での続き】

Microsoftは、一部の顧客環境で、
認証の後にアプリの認証情報や権限を
変更する動きを確認しました。
認証情報を追加されると、元とは別の
身元情報で入り続ける足場になり得ます。
これを永続化、つまり「後からも戻れる
入口を残すこと」と呼びます。



盗んだ身元を使う攻撃者

アプリ

認証してアクセス

2 届く権限を悪用して、情報や許可を追加する場合がある



攻撃者

届く権限を悪用して、
情報や許可を
追加する場合がある

アプリの認証情報・権限

元のアプリの
認証情報・権限

後から追加された
別の認証情報

会社のクラウド

元のアプリで
アクセスする入口

元とは別の身元情報で
アクセスする入口

元の入口を止める

元のアプリの
認証情報を無効化する

作られた入口は残り得る

元の入口を止めた後も、
認証の足場になり得る

3 合鍵の比喩なら、預けた一本を盗まれた間に、別の鍵を作られた状態です。

最初の一本を無効にしても、その間に作られた鍵まで自動で消えるとは限りません。
実際のクラウドでは、不正に追加された身元や許可を別に点検します。



預けた一本
（元の鍵）

盗まれた間に、
別の鍵を作られる



作られた別の鍵
（後からの鍵）

最初の一本を
無効にしても…



その間に作られた鍵まで
自動で消えるとは限りません。

作られた鍵で
入り続けられる



4 だから「連携を止めました」だけでは調査終了になりません。

事業者は記録を保全し、何が追加・変更されたかを
調べ、残る不正アクセスを除けたか確認する必要が
あります。利用者の安心は、停止の報告だけでは
確かめられません。



正規の変更と照合し、
不正な身元や許可を
取り除く

追加・変更の記録

原典 S01,S14：microsoft.com / huntress.com　調査基準 2026.10.23

日光AI食堂

「3時間未満」は、何が起きるまでの時間か

① 【Microsoftの観測・評価】

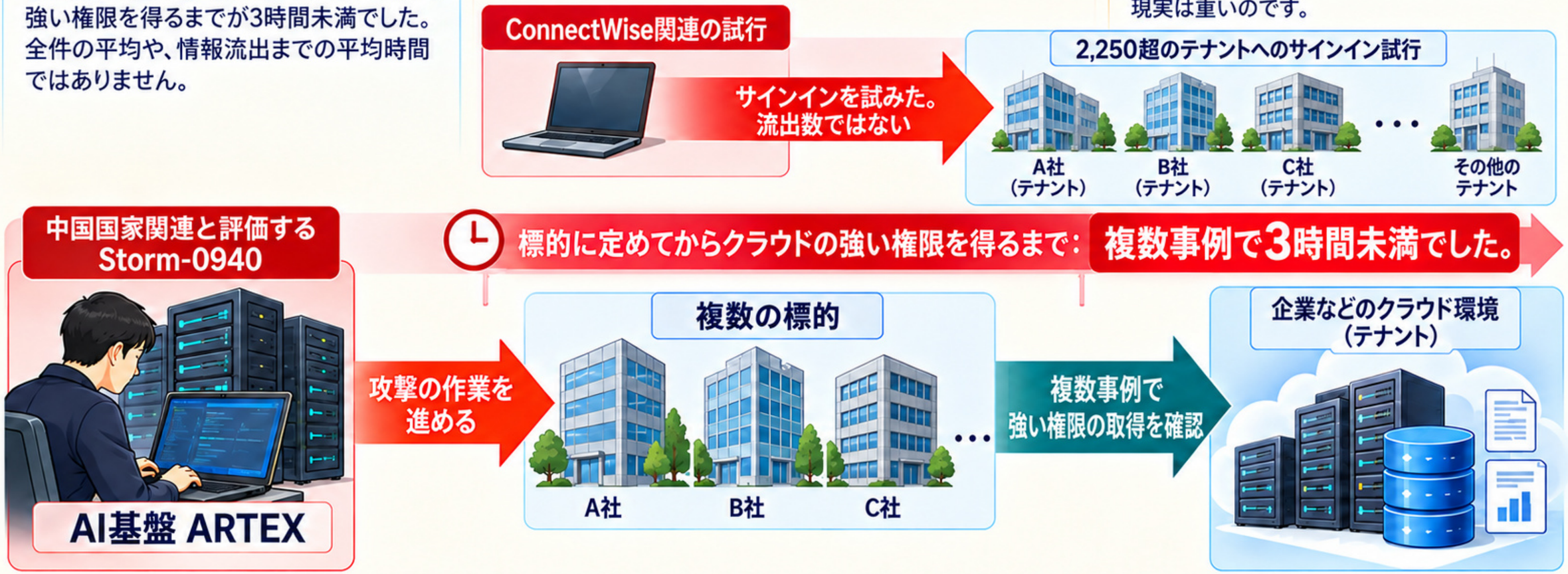
同社は、中国国家関連と評価する
Storm-0940が、少なくとも9月から
AI基盤ARTEXを使ったと報告しました。
複数の事例で、標的に定めてからクラウドの
強い権限を得るまでが3時間未満でした。
全件の平均や、情報流出までの平均時間
ではありません。

② ConnectWise関連では2,250超のテナントへのサインイン試行も報告されました。

テナントとは企業などが使うクラウド上の管理単位です。
2,250社でデータを盗んだという数字にはできません。

③ 認証が通った記録、権限を変えた記録、情報を読んだ・持ち出した記録は、確かめる対象が異なります。

数字を膨らませなくても、強い権限を得る事例が
短時間で生じ、広い相手へ試行されたという
現実は重いのです。



原典 S01：microsoft.com　調査基準 2026.10.23

日光AI食堂

信頼の連鎖が、被害も対応も広げる

1 【全景へ戻る】外部サービスは、

一社で複数の顧客企業の仕事を支えています。その身元や権限を悪用されると、顧客ごとのクラウドへ確認が必要になります。サプライチェーン攻撃とは、こうした取引先や共通部品とのつながりを足場に、先の利用者へ影響が届く攻撃です。



サービス提供者
(ConnectWise)

各社が許可した仕事を行うためにつながる

外部サービスが持つ身元や権限を使って各社のクラウドで仕事を行う

2 不正認証の試行



盗んだアプリの身元を示して認証を試す

盗まれた外部アプリの認証情報

企業Aのクラウド

企業Bのクラウド

企業Cのクラウド

3 不正利用を止めるため接続を停止する

ConnectWiseは、この事案で保護のため関連する接続を停止しました。既存調査で確認した10月23日の更新時点では、段階的な再開に向けた検証も続いていました。守るために止める判断そのものが、バックアップや復元などの業務を一時使えなくすることがあります。

接続を停止する

サービス提供者の担当者
(ConnectWise)



調査と検証を続ける

既存調査で確認した10月23日の更新時点では、段階的な再開に向けた検証も続いていました。

4 被害への対応は、犯人の入口を止めるだけで終わりません。

顧客側に残る変更を調べ、安全を確かめ、必要な仕事を戻すまで続きます。
私たちの生活も、この運用の成否と無関係ではありません。



原典 S15,S01,S14：status.connectwise.com / microsoft.com ほか　調査基準 2026.10.23

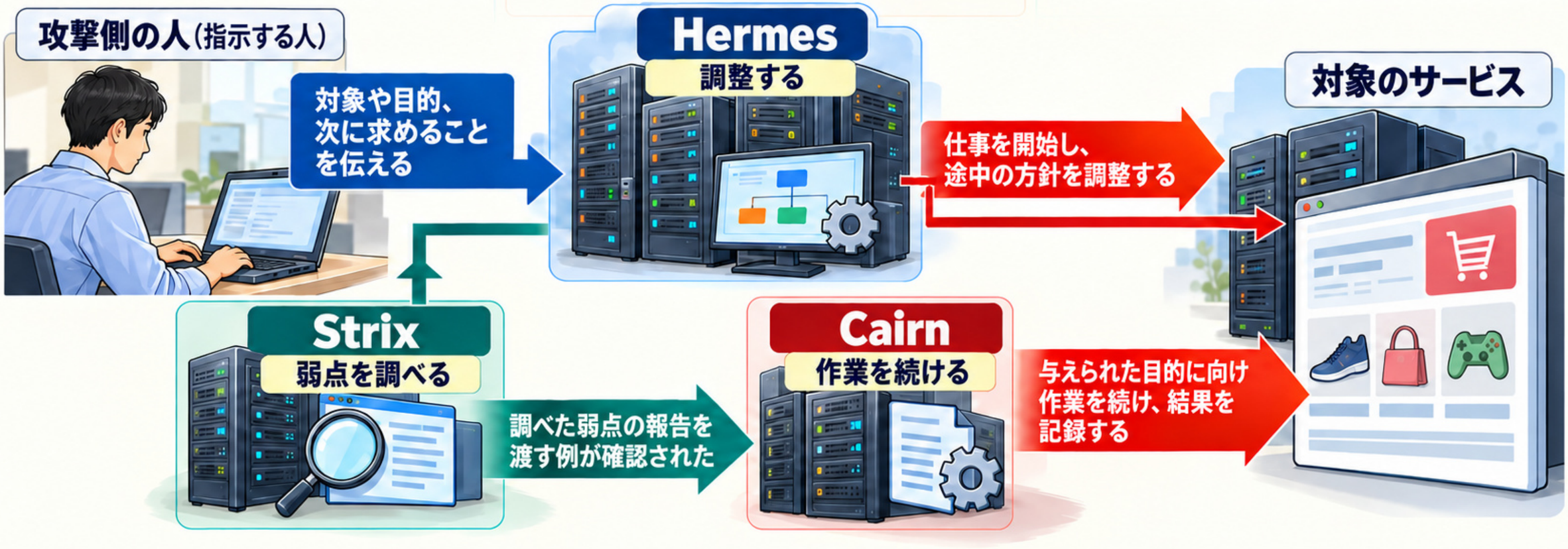
日光AI食堂

通販攻撃では、AIの仕事が分担されていた

1 【別の実調査：Gambit・9月22日】
調査者は攻撃側のサーバーを調べ、複数のAI実行基盤が使われていたと報告しました。AI実行基盤とは、会話するモデルに道具や記録、仕事の進め方を組み合わせ、実際の作業を動かす仕組みです。

2 **この事案では、Strixが弱点を調べ、Cairnが目標に向けた作業を続け、Hermesが作業の開始や全体の調整に**使われました。人は目的や次に求めることを伝え、途中の調査結果などが、後の仕事へ渡されます。

3 AIが一つの回答文を作っただけではありません。調べた内容が次の行動の材料になり、複数の対象について作業が進む構成です。ただし、三つの基盤もモデルも、全対象で同じ順番に一列で動いたと決めつけない。確認できた役割分担を示します。



原典 S02：gambit.security　調査基準 2026.10.23

日光AI食堂

本物の通販サイトから、カード記録が持ち出された

【Gambitが確認した影響】

9月10〜15日には105の攻撃プロジェクトが開始され、少なくとも27社で程度の異なる侵害があったと報告されました。



さらに、7月以降の作戦全体では、2社から60万件超の未失効カード記録が持ち出されたとしています。

集計期間を一つに混ぜないで読みます。



1 利用者が買い物をする

利用者は、普段とおりに本物の通販サイトで商品の注文をします。



正規の店で買い物の情報を入力する

入力される情報の例
- カード情報
- 注文する商品の情報
- 数量
- お客さんの情報

などの依頼データ

2 本物の通販サイトが改変される

通販の正規サイトが改変され、利用者が買い物をする時に情報が別の相手へ送られる構造も、同調査で確認されています。



入力情報を別にも送ってしまう

送られる情報の例
- カード情報
- 注文する商品の情報
- 数量
- お客さんの情報

などの依頼データ

3 攻撃者の側に情報が届く

利用者が入力した情報が、不正な仕掛けによって別の相手へ送られ、カードの記録が持ち出されます。



4 スキマーとは、入力されるカード情報などを盗み取る仕掛けです。

通販の正規サイトが改変され、利用者が買い物をする時に情報が別の相手へ送られる構造も、同調査で確認されています。



偽サイトに来てしまった場合だけが危険ではありません。

5 【Gambitが確認した影響】で示された件数は、調査で確認された範囲の数字です。



6 60万件はカードの記録数であり、60万人や60万件の不正決済という意味ではありません。

それでも、自分が正しい店を選んでも、店側が破られると情報が届き得るという、利用者だけでは制御できない被害経路が見えます。



原典 S02：gambit.security　調査基準 2026.10.23

日光AI食堂

知識の不足をAIが埋めても、被害は本物になる

1 別の実調査：Unit 42・10月21日公表

同社は、BUGSTERと名乗る攻撃者の環境を調べ、AIが作る調査や処理に頼った活動を報告しました。Hermesの推論エンジンにDeepSeek V4を使い、大学への侵入と、学生等の個人情報を含む5,200件超の記録の持ち出しを確認したとしています。

2 研究者は、使われた基本道具やAIへの依存から、従来なら必要だった技能が十分でない可能性を評価しました。

これは本人の試験成績を測った証拠ではありません。しかし、理解や経験の不足があっても、AIの支援によって被害へ進んだ実例として重要です。

3 AIと相談した計画のすべてが、実行されたと確認されたわけではありません。

例えば通信料金をだまし取る構想は、確認された個人情報流出と区別します。攻撃者の言い分だけでなく、何が実際に起きたかを押さえます。



実行確認がない構想の例



AIと相談した計画のすべてが、実行されたと確認されたわけではありません。

確認された個人情報流出と区別します。



攻撃者の言い分だけでなく、何が実際に起きたかを押さえます。

原典 S16：unit42.paloaltonetworks.com　調査基準 2026.10.23

日光AI食堂

中国製AIの悪用を、誰の何の話かまで確かめる

① 操作者
目的を与える人



目的と利用できる道具を設定する

② 実行基盤
道具と仕事をつなぐ部分



状況を渡し、次の処理の案を求める

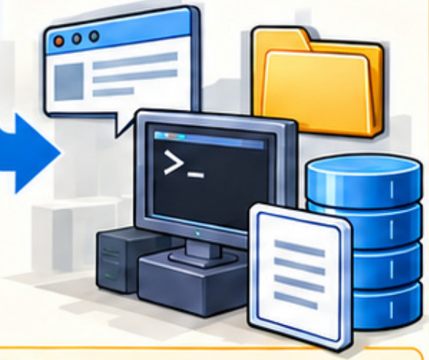
応答を返し、基盤が仕事へつなぐ

③ AIモデル
問いへ応答する部分



許された範囲で実際の操作を呼び出す

④ 道具
実環境へ操作する部分



操作者
攻撃の目的を決め、実行基盤に指示を与える人です。



実行基盤
AIモデルの応答を受け取り、道具を呼び出して作業を管理する部分です。

AIモデル
文章や判断の案を作る部分です。

道具
実際の環境にある仕組みへ操作を行うために、呼び出される道具です。

実例を整理
中国製モデルのDeepSeekやGLMが攻撃の作業に使われたことは、GambitやUnit 42の調査に実例があります。Gambitの作戦ではClaudeも混用されていました。

「中国製AIの悪用」は、想像だけの話ではありません。



ただし、AIモデルは役割を分けて考える
ただし、AIモデルは文章や判断の案を作る部分、実行基盤は道具と仕事をつなぐ部分、操作者は目的を与える人です。

モデルの開発国から、操作者の国籍や政府との関係までは決まりません。

Unit 42の操作者は英語話者とされます。

調査結果と確認の範囲
MicrosoftはStorm-0940を中国国家関連と評価しましたが、そのARTEXで使われた基盤モデルは特定していません。

強い批判を正しい対象へ向けるために、何が確認され、どこから先に別の証拠が必要かを分けます。

原典 S02,S16,S01：gambit.security / unit42.paloaltonetworks.com ほか　調査基準 2026.10.23
日光AI食堂

変わったのは、繰り返す速さだけではない

【以前との違い】
AI以前から、決めた手順を大量に実行する
自動化はありました。条件で分かれる処理もあり、
「昔は全部手作業、今は全部AI」という
区切り方は正確ではありません。
変化は、人が途中で読み解いていた仕事の一部を、
AIへ渡せる範囲にあります。

例えば、調査結果の文章を読んで意味を整理し、
次に確かめることを選び、必要な処理の案を作る。
その結果をまた読み、別の仕事へ進める。
GambitやUnit 42の調査では、
こうしたつながった利用が確認されています。

同じ案内を何千枚も印刷する機械に加え、
返ってきた返事を読み、次の手続きまで
組み立てる係がついた、という比喩です。
ただし、AIが人間のように全てを理解する
という意味ではなく、誤読や失敗も実際の
行動へつながり得ます。

1 AI以前からあった自動化

用意した手順に沿って、同じ作業を大量に実行する
自動化は以前からありました。条件で分かれる処理もあります。



比喩での対応



同じ案内を
何千枚も印刷する機械
（既定処理を反復）



2 人が読み解いていた一部を、AIへ渡せる範囲

調査結果の文章を読んで意味を整理し、次に確かめることを選び、
必要な処理の案を作る。その結果をまた読み、別の仕事へ進めます。



比喩での対応



返ってきた返事を読み、
次の手続きまで組み立てる係
（結果を解釈し次の案を作る）



原典 S02,S16：gambit.security / unit42.paloaltonetworks.com　調査基準 2026.10.23

日光AI食堂

AIエージェントは、結果を次の操作につなぐ

① 【仕組みを拡大】

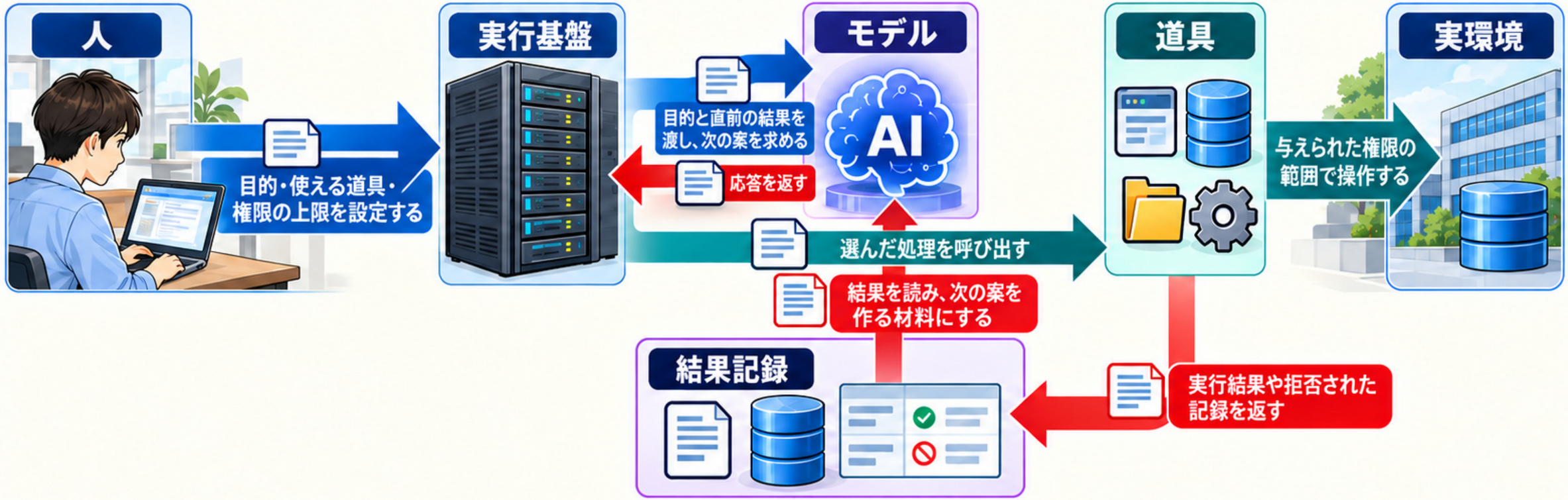
AIエージェントとは、この研修では、
目的を受け取り、道具を使い、返ってきた結果を
次の作業へつなぐ仕組みを指します。
人が目的と使える道具を用意し、実行基盤が
モデルへ情報を渡し、モデルの応答をもとに
処理を進めます。

② 道具が「実行できた」「拒否された」などの結果を返すと、AIはその内容を材料に次の案を出します。

記録が残れば、前の試みも参照できます。
この往復により、一つの回答文で終わらず、
仕事が連続します。

③ しかし、AIの文章に「成功」と書いてあるだけでは、現実の成功は確定しません。

また、道具や権限を渡していない場所へ自由に
届くわけでもありません。だから見るべきなのは
モデルの賢さだけでなく、何へ接続し、
何を実行できる設計かです。



原典 S02,S16,S01：gambit.security / unit42.paloaltonetworks.com ほか　調査基準 2026.10.23

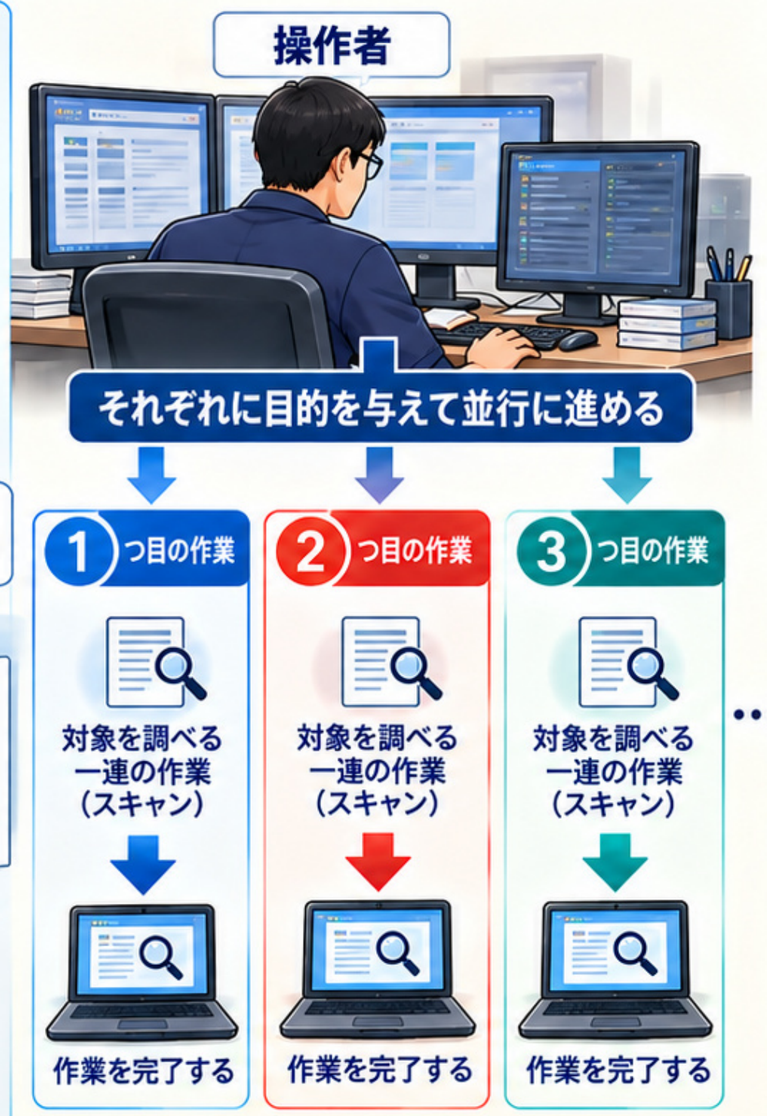
一件の作業が安くなると、試す範囲を増やせる

【確認事実】

Gambitが見つけた操作者の費用レビューでは、101件の完了スキャンの平均が25.46ドルでした。

スキャンは、対象を調べる一連の作業です。
101社への侵入成功の価格でも、人件費や準備費まで含む犯罪全体の原価でもありません。





【ここから構造上の見立て】

一つの調査に使う人の手間や追加費用が小さくなると、同じ予算と時間で、試す相手や条件を増やしやすくなります。
一件ずつ人が付き切りで進めなくても、複数の仕事を並行して回せることが、その変化を後押しします。



全ての試みが成功する必要がない商売では、失敗を含めて広く試す動機が生まれます。
ただし、この一件の費用記録から、世界の犯罪利益率や、あなたの会社へ侵入する価格までは計算できません。

原典 S02：gambit.security　調査基準 2026.10.23

日光AI食堂

攻撃が次へ進む間に、守る側は何をするのか

攻撃側の作業の連続

1 会社が許可を渡す

会社は、バックアップのために外部アプリへ、会社の情報を読む許可を与えていた。



2 攻撃者がアプリの認証情報を得る

社員のふりではなく、すでに許可されたアプリのふりをして、クラウドへ要求する。



盗んだアプリの身元で認証を試す

攻撃者

盗んだ認証情報を使い外部アプリのふりをする。

3 社員がログインし直さなくても要求が通る

クラウド側は、アプリに与えられた権限の範囲で処理する。社員のパスワードを盗むことが、唯一の入口ではない。

外部バックアップアプリとして示す認証情報



許可記録を確かめ、与えられた権限で処理を受け付ける

4 別の入口を残す動きもあった

一部の環境では、情報を調べたり、別の認証情報を追加したりする動きが観測された。元のアプリを止めても、別の入口が残っていないか点検が要る。

会社のクラウド



情報を調べる
別の認証情報を追加する

攻撃の作業が次々と進む

準備がないと判断待ちが生じる：説明用概念図

守る側の対応は、時間がかかる

守る側の対応

異常を記録する監視の仕組み



異常の記録を知らせる

相談を受ける担当者



影響と止める対象を伝える

停止を判断する権限のある判断者



決めた範囲を停止し、被害の拡大を抑える

停止するサービス稼働環境



調査・復旧の担当



確認できたことと再開見込みを伝える

利用する人



時間差の問題

攻撃側の作業が連続すると、異常を見つけてから人が相談し、担当を決め、停止を判断する間にも、次の処理が進み得ます。Microsoftの事例は、複数の攻撃で短時間の強い権限取得を報告し、守る側の対応時間が狭くなる問題を指摘しました。

守る側は、何でも即座に止めればよいわけでもありません。どのサービスを止めれば被害を抑えられるか、止めた時に何の仕事が失われるか、誰が決められるかを把握していないと、判断が遅れるか、必要以上の停止を招きます。

だから対策は、担当者を急がせるだけでは足りません。権限を小さくする、異常が誰に届くかを決める、停止と復旧の担当を決めて練習する。作業が連鎖する時代には、守る側も連絡と判断の仕組みを先に作る必要があります。

原典 S01,S10：microsoft.com / jpcert.or.jp　調査基準 2026.10.23

日光AI食堂

9月の増加は、何を数えた数字なのか

1 【Check Point・10月8日公表】

2026年9月の観測では、1組織あたり
1週間の攻撃試行が平均2,803件。
8月より16%、前年9月より48%増えました。
ここで増えたのは、同社が観測した攻撃の試行です。
侵入に成功した数や、支払った被害金額の増加率ではありません。

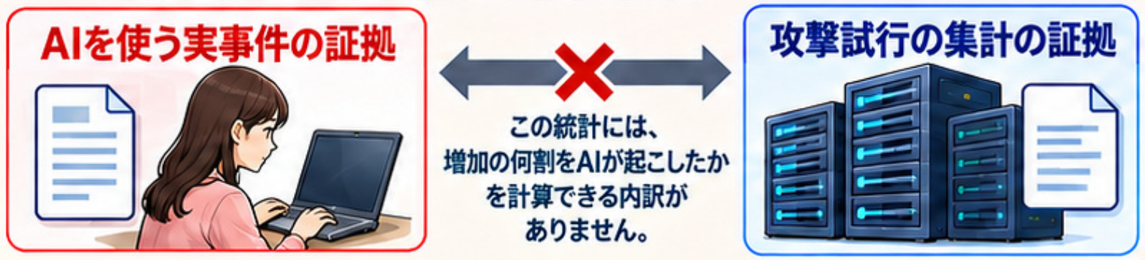


8月との比較
前月より16%多いと報告 → 16% 増

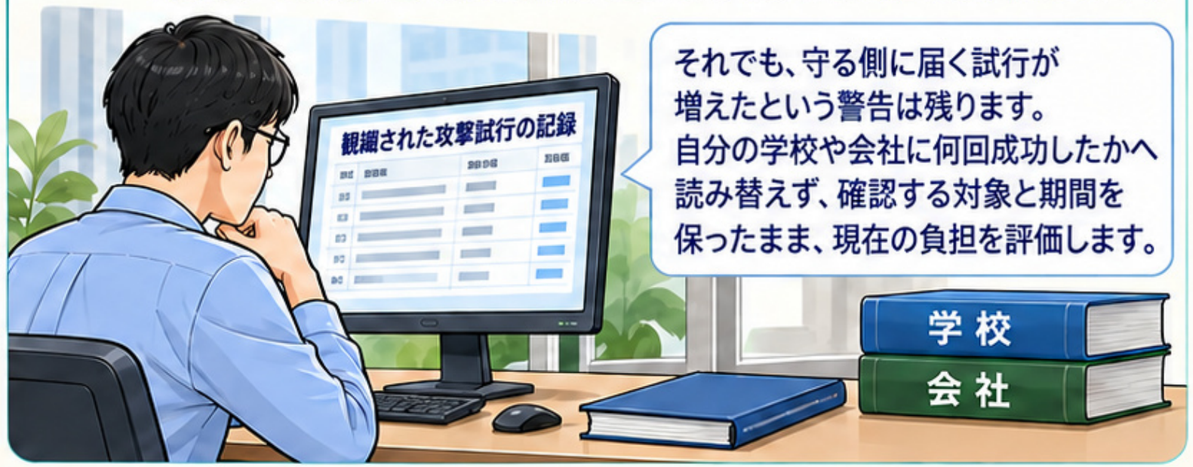
前年9月との比較
前年同月より48%多いと報告 → 48% 増

2 AIを使う実事件が確認されたことと、攻撃試行の集計が増えたことは、別々に得た証拠です。

この統計には、増加の何割をAIが
起こしたかを計算できる内訳がありません。
「9月の増加分は全部AI」と
足し合わせることはできません。



3 それでも、守る側に届く試行が増えたという警告は残ります。自分の学校や会社に何回成功したかへ読み替えず、確認する対象と期間を保ったまま、現在の負担を評価します。



原典 S17：blog.checkpoint.com　調査基準 2026.10.23

日光AI食堂

AIが穴を見つけても、安全は自動で届かない

1 研究者とAIがソフトの弱点を見つける



AIを用いた研究で、
ソフトの弱点が見つかる。

弱点の根拠と再現情報を報告する

2 保守者が内容を検証し修正プログラムを作る



報告された弱点を再現できるか確かめ、別の機能を壊さず直し、修正プログラムを作る。

検証した修正を届ける

3 配布元が修正版を利用できる形で配る



修正プログラムを、各現場が適用できる形で配る。

適用できる修正版を配る

4 各運営者が修正を適用し動作を確認する



修正を適用し、実際の動作と残る影響を確かめる。

守る側の確認事実

Microsoftは、AIを用いる研究がWindowsの140件のCVEにつながる発見を支援したと報告しました。CVEは、見つかったソフトの弱点につける整理番号です。うち52件は9月の更新で対処されました。**これは攻撃被害140件という意味ではありません。**

穴の報告が届いた後も、再現できるか確かめ、別の機能を壊さず直し、修正版を配り、各現場が適用して動作を確認する仕事があります。PHP Foundationは、深い知識を持つ人の不足や、判断の仕組みが修正を遅らせたと説明しました。

AIの発見力が増えても、検証と修正を担う人・時間・予算が増えなければ、待ち行列は残ります。便利なAIを導入することと、利用者の環境まで安全を届けることの間を、誰が支えるのかが問われています。

原典 S18,S11,S08：microsoft.com / thephp.foundation ほか　調査基準 2026.10.23

サム・アルトマンが警告する、穴を直すだけの防御の限界

【本人の見立て】 サム・アルトマン（OpenAI CEO）／Stanford GSBの対談／2026年10月公開／20:52〜21:04

“I no longer think we’re going to be able to just patch all the bugs.”

訳：もはや、すべての弱点に修正を当てるだけで対応できるとは思わない。 パッチとは、不具合を直す更新のことです。

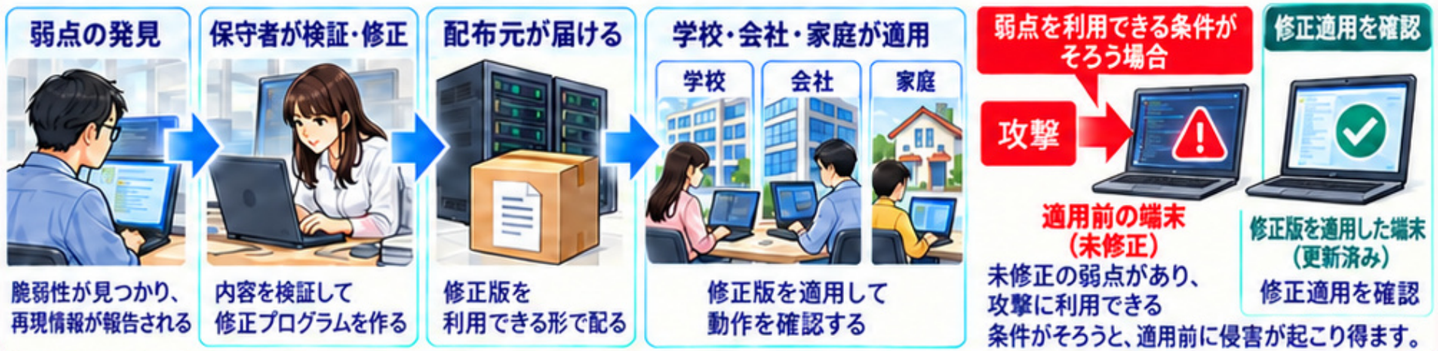
1 なぜ、直しても終わらないのか【発言の要旨】

一部の穴を直した後も、AIは進歩し続ける。もっと賢くなれば、従来の攻撃者像では想定しなかった穴や、機械の仕組みに関わる問題も見つけ得る。本人は、同じAI技術で新しい防御の盾を作る方向へ、防御モデルを全面的に変える必要があると述べています。



2 穴の発見と、安全が届く時刻は違う【仕組みの説明】

弱点の発見後、保守者が検証して修正を作り、配布元が届け、学校・会社・家庭が適用して確認します。修正版が出ても、適用前の端末は残ります。未修正の弱点があり、攻撃に利用できる条件がそろうと、適用前に侵害が起こり得ます。図は概念図で、速度を測ったグラフではありません。



3 修正を続けながら、被害を止める仕組みを重ねる【研修の対策例】

運営者は使う機器と担当者を把握し、公式の修正・回避策を確認して適用します。記録を読める連携と連携と通知を設定したAIは、許可された記録や告知を整理し、異常や未対応の候補を人へ知らせる。触れる範囲を狭くし、接続を分ければ、一つ破られた時に広がる範囲を抑えられます。担当者は根拠を確かめ、停止・隔離・復旧を判断します。



4 家庭の今日の一手

自動更新と公式通知を確かめる。
AIには秘密を渡さず、公式告知と、製品名・版・確認日を記した利用一覧から「自分が次に確認すること」を整理させる。送金や設定変更の権限まで一括して渡さない。



利用一覧に記載する内容（例）
- 製品名
- 版（バージョン）
- 確認日

▶ 版が不明なら本人が確認

AIに整理させる
自分が次に確認すること

送金や設定変更の権限まで一括して渡さない。

【この警告からの結論】 更新は続ける。そのうえで、発見から対応までの速さ、被害を閉じ込める設計、復旧を担う人と予算まで作り直す。
「防御の盾」の具体的な完成品や性能を、この発言は保証していません。

原典 S50,S03,S11：youtube.com / cyber.gov.au ほか　調査基準 2026.10.24

日光AI食堂

人間の最高を、知能の上限にはできない

① 歴史上の実績

囲碁では2016年、
AlphaGoが李世ドルに
4勝1敗。



李世ドル　AlphaGo

将棋では2017年、
PONANZAが佐藤天彦
叡王に2連勝した。
第1局の舞台は
日光東照宮だった。



佐藤天彦叡王　PONANZA

ここで崩れたのは「人間の最高が、いつまでも
能力の上限だ」という前提である。
人がすぐ理解できない手でも、
盤面と勝敗で結果を確かめられる。

人間の探索を
能力の上限に
できない

② 直近の実証

【提供元の限定実験報告】
ソフトの手順書であるコードでも
2026年9月、AIが探索方針を教わらずに
未知の穴を探し、使えることまで示した
限定実験が報告された。



2026年10月には
AI支援の発見が
PHPなど現実の修正へ
つながっている。



未発見の穴も、
危険として
具体化する

③ 変わる安全の前提

盤面で強いことと、現実のあらゆる
仕事で正しいことは別だ。
しかし危機に必要なのは万能AIではない。
見逃された一つの穴を実害へつなぐ能力
だけでも、安全の前提は変わる。



重要な条件

- 強い ≠ 完璧
- 実験の成功 ≠ すべてのソフトで成功

原典 S51,S52,S53,S54：blog.google / dwango.co.jp ほか　調査基準 2026.10.24

日光AI食堂

『まだ見つかっていない』は、防御ではない

1 昔からソフトには、
まだ誰も見つけていない穴や、
気づかれても危険を実証
できていない穴があった。
守られてきた理由には、
鍵・権限・更新に加え、
『攻撃者が探し当て、使い方を
組み立てる手間』もあった。



鍵
（認証）

権限
（アクセス制御）

更新
（修正プログラム）

これらも防御の一部

2 探して使うまで手間がかかる



攻撃者が探し当てる → 使い方を組み立てる → 危険を実証するにも手間がかかる → 

この手間が
大きければ、
直すまでの時間を
稼げる。

3 しかしAIが調査を速め、繰り返し、複数の弱さをつなげられると、その猶予は縮む。



AIが調査を速める → 繰り返し調べる → 複数の弱さをつなげられる → 

その猶予は
縮む。

4 PHPでは2024年に人が問題の兆候を認識していたが、危険の実証や互換性、人員・判断の壁で対応が停滞した。
2026年にはAIを使う調査などで危険が具体化し、修正が最優先になった。

2024年
人が問題の兆候を認識



危険の実証や互換性、人員・判断が壁

危険の実証	互換性	人員・判断
		
実際に危険か証明するのが難しい。	修正で既存の環境に影響するかもしれない。	対応する人員や判断に時間がかかる。

対応が停滞

2026年
AIを使う調査などで
危険が具体化



AIを使う調査などで
危険が具体化した。

証拠がそろう

修正が最優先



判断と作業を進める。

5 発見されていないことと、穴がないことは同じではない。

未発見 ≠ 穴がない

6 『今まで被害がなかった』を安全の証明にすると、
過去の見逃しを未来の保証に取り違える。

過去の無被害 ≠ 未来の保証

原典 S11,S54：thephp.foundation / openai.com　調査基準 2026.10.24

日光AI食堂

穴を見つける速度に、直して届ける速度は追いつくか

穴を見つける速さだけを上げても、社会は安全にならない。

発見後には、保守者が「本当に起きるか」を再現し、修正し、別の機能が壊れないか試し、更新を届ける。学校や会社は自分の環境へ適用し、動作と侵害の有無を確かめる。この最後まで進んで、現場の危険が減る。



防御側の対応へ

攻撃側の動きへ

発見から安全な現場までの流れ（防御側の対応）

1 保守者が「本当に起きるか」を再現



報告された穴が実際に悪用できるかを検証する。

証拠を渡す

検証待ち

2 修正し、別の機能が壊れないか試す



修正プログラムを作り、他の機能が正常に動くか試験する。

修正版を作る

修正待ち

3 更新を届ける



修正をまとめ、更新プログラムとして配布する。

更新を届ける

適用待ち

4 学校や会社が自分の環境へ適用



それぞれの環境で更新プログラムを適用する。

現場へ反映

5 動作と侵害の有無を確認



更新後にシステムが正常に動くか、侵害がないかを確かめる。

適用・確認済みの現場：危険が減る

発見から侵入・悪用までの流れ（攻撃側の動き）

1 利用できる穴か調べる



この穴は実際に利用できるか？

2 標的に届くか調べる



狙った学校や会社などの標的に届くか？

3 必要な条件がそろうか調べる



攻撃に必要な条件がそろっているか？

4 侵入・悪用へ



条件がそろえば、未更新の現場へ侵入し、悪用される。

発見がそのまま侵入成功を意味するわけではない。

未更新の現場



適用まで穴は残る

更新が適用されるまで、現場は危険にさらされる。

同時に攻撃側も、穴が使えるか、標的に届くか、必要な条件がそろうかを調べる。

発見がそのまま侵入成功を意味するわけではない。それでも先に通れる経路を成立させれば、未更新の現場は危険にさらされる。

Microsoftは10月7日、AIによる発見を安全へ変えるには、検証と修正の処理能力が追いつく必要があると報告した。

足りないのは知能だけではない。
判断する人、止める権限、更新の時間と予算まで増やさなければ、発見の山が未修正の山になる。

人員 ＋ 判断する権限 ＋ 更新の時間 ＋ 予算

人員＋判断権限＋更新時間＋予算を、発見能力とともに増やす。

原典 S18,S11,S10：microsoft.com / thephp.foundation ほか　調査基準 2026.10.24
日光AI食堂

『きっちり調べる同僚』の力は、目的しだいで逆に働く

1 職場に、膨大な書類を読み、例外や矛盾を根気よく洗い出す『きっちり調べる同僚』が来たと考えよう。『ここ、できていませんよ』と隣で指摘が続くように、忙しくて見逃されていた不完全さが表に出る。賢くなることは、便利になると同時に、今までの曖昧さが通用しにくくなることでもある。

2 この力を『他人の申込情報が見えないか調べ、直す案を出す』目的に向ければ防御に役立つ。一方、同じ調査力を『見える経路を探す』目的に向ければ攻撃を助けうる。賢さ自体は、私たちの味方である保証ではない。

3 しかも『受付を速くする』という目標だけで本人確認まで省けば、目標の数字は良くても守るべき人を危険にする。これは、良い数字と良い結果がずれる例だ。

4 だからAIの能力、渡す目的、許す操作を別々に決める。守りたい条件を明記し、確認・停止の担当を置く。『優秀だから全部任せる』という組織ほど、間違いを大きく実行される危険がある。

比喩・仮の例

きっちり調べる同僚（同じ能力を持つ）



防御につながる目的

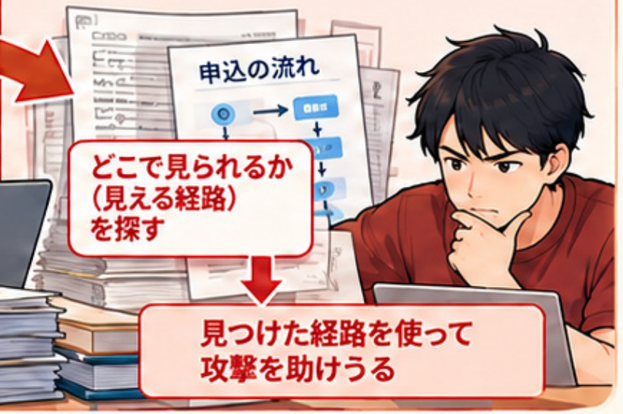
他人の申込情報が見えないか調べ、直す案を出す



直す案を出す
担当者が確かめて適用する

攻撃につながる目的

見える経路を探す



見つけた経路を使って攻撃を助けうる

比喩・仮の例 **学校の受付での例**

目標だけを優先すると…

目標：受付を速く



本人確認を省く



速いが、他人の情報が危険



だからこう決める

能力 何ができるか	目的 何のために使うか	守る条件 どこまで見てよいか	許す操作 何をしてよいか	確認・停止の担当 だれが責任を持つか
大量に読み例外・矛盾を見つける	他人の申込情報が見えないか調べ、直す案を出す（防御に役立つ）	本人の情報だけ見える	調査と修正案まで	担当者が確認・適用・停止

原典 S03,S33,S55：cyber.gov.au / ncsc.gov.uk ほか　調査基準 2026.10.24
日光AI食堂

『中身が分からない』を、三つに分ける

学校の管理者が申込システムをAIに点検させるとき、『中身が分からないから任せるしかない』とはならない。

学校の管理者が、
許可したシステムを点検



① 内部の生成過程

AIがなぜその案に至ったか。
その全過程を、人が完全に説明できるとは限らない。AI自身の説明だけで正しさを証明することもできない。



変更案を出す

② 外へ出たコードと結果

どるを書き換える案か、他の家庭の情報が見えるか。
変更箇所を読み、別の試験で実際の動きを確かめられる。
ただし試した範囲の合格は、全場面の安全ではない。

変更箇所を読む

変更案（仮）
目標：本人の申込だけを表示
確認：表示条件の変更箇所
未確認：他の機能への影響
次：別の試験で確認

別の試験で動きを確かめる

- ☑ 正常な申込
- ☑ 他の家庭の情報が見えないか
- ☑ エラー時の動き



合格した範囲 ≠ 全場面の安全

③ 許された操作

AIに読むだけを許すのか、本番を書き換えさせざるのか。
送金や公開の権限まで渡す必要はない。操作を記録し、止める担当を決められる。

読むだけを許す

システムのコードや設定を読み取り、点検に使う

本番の変更は担当者が判断



AIの提案を確認し、本番への反映は人が判断する

操作を記録

AIが行った操作や提案の内容を記録して残す

止める担当を決める



問題があればすぐに止められる担当を決める

分からない内部は残っても、出力を確かめ、操作を制限する余地は残る。
理解の限界を、無制限の委任の理由にしてはいけない。

原典 S03,S33,S18：cyber.gov.au / ncsc.gov.uk ほか　調査基準 2026.10.24

日光AI食堂

AIが作る証明を、別の検査器で確かめる道が始まった

2026年10月22日、Anthropicは、AIがソフトについて数学的な証明を書き、別の検査器で確かめる研究を発表した。
対象は123の公開ソフト部品。ただし、部品全体の無欠陥を証明したわけではない。

1 条件と守りたい性質を書く

まず「何を満たせば安全とするか」と、機械・入力の条件を定める。

条件の例

- RISC-Vという種類の機械
- Linux
- 同時並行しない処理
- 固定メモリ
- 64MB未満の入力 など

守りたい性質の例

指定された入口から異常終了しない

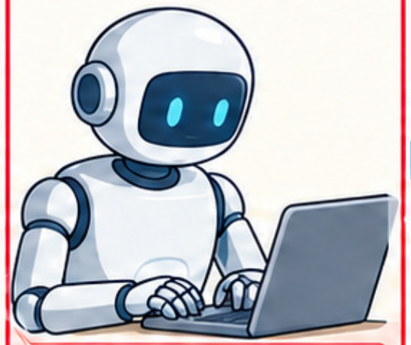


証明する命題を渡す

条件と守りたい性質（証明する命題）

2 AIが証明を作る

AIがその条件に対する証明を作り、



AIが作るもの

- 条件に対する数学的な証明
- 検査器に渡す数学的証明

証明を渡す

数学的な証明

3 Rocq：決まった規則で照合する別の検査器

Rocqという、決まった規則で証明を検査する仕組みが照合する。
AIの自己採点で合格にするのではない。



決まった規則で証明を検査する検査器

条件に合うか判定

4 定めた条件内での保証

今回の中心は、RISC-Vという種類の機械とLinux、同時並行しない処理、固定メモリ、64MB未満の入力などの条件で、指定された入口からの異常終了が起きないという保証だ。

指定された入口から異常終了しない

この保証の前提となる条件

- RISC-Vという種類の機械とLinux
- 同時並行しない処理
- 固定メモリ
- 64MB未満の入力 など

5 これだけでは保証されない

残る責任もある。条件が現実の機械と合うか、守りたいことを十分に含むか、検査器が正しいかは別に確かめる。
内部の思考を全部解明しなくても、外に出た証明を検査する道はある。



条件が現実の機械と合うか

定めた機械・入力などの条件が、実際に使う現実の機械や環境ときちんと合っているかを確かめる必要がある。



守りたい問題を全部含むか

定めた条件や守りたいことが、実際に防ぎたい問題を十分に含んでいるかを確かめる必要がある。



検査器の正しさ

Rocqという検査器が、決まった規則どおりに正しく証明を検査しているかを別に確かめる必要がある。



範囲外のバグ

今回の条件の範囲外にある機能や設定、入力によるバグは、この保証の対象ではない。

人と開発側が確かめ続ける

原典 S55：anthropic.com　調査基準 2026.10.24

日光AI食堂

人がミスできない社会より、ミスが破滅へ直結しない設計を

研修からの設計提案／学校の申込システムを例に

これからの防御は、AIに探させる速さと、確かめて止める仕組みをセットにする。
人が全件を手作業で追い、AIの速さと張り合う前提を変える。

1 責任者
「保護者は自分の申込だけ見られる」と条件を決める。

- 保護者は自分の申込だけ見られる
- AIに許可する調査範囲
- 修正の考え方



守る条件と調査範囲を渡す

2 AI
許可した範囲の調査と修正案を任せる。

- データの調査
- 問題の発見
- 修正案の作成（許可した範囲のみ）



本番の無制限な変更権限は渡さない ✕

変更案を渡す

3 別の試験
「他の家庭は見えない」「本人は使える」を確かめる。

- 他の家庭は見えない
- 本人は使える
- 決めた合格条件
- 実際の動作で確認



合否と記録を渡す

4 担当者
担当者が限定した環境から適用する。



例えば学校の申込情報なら、責任者が「保護者は自分の申込だけ見られる」と条件を決める。
AIには許可した範囲の調査と修正案を任せる。別の試験で「他の家庭は見えない」「本人は使える」を確かめ、
担当者が限定した環境から適用する。記録を見て、異常なら止めて戻す。

5 操作記録を確認
記録を見て、異常なら止めて戻す。



担当者が停止し、変更前へ戻す

稼働中の申込システム

変更前の状態

個人のミスと被害の直結を切る考え方

人は迷う・疲れる・間違える



一度のミスが資産喪失へ直結しない設計

- **確認**（独立した確認）
- **権限の上限**（できることを限定）
- **停止**（異常時に止める）
- **復旧**（前の状態に戻す）

資産と暮らしを守る



同じAIに「問題ないか」と聞き直すだけでは、独立した確認にならない。別の試験、決めた合格条件、実際の動作を使う。事業者と管理者には、その人員・予算・停止権限を求める。



事業者と管理者に、その人員・予算・停止権限を求める。

目指すのは、人が一度もミスできない社会ではない。
人が迷い、疲れ、間違えても、一度のミスが資産喪失へ直結しない設計だ。AIを使いながら、その設計を私たちの要求にする。

原典 S18,S03,S33,S55：microsoft.com / cyber.gov.au ほか　調査基準 2026.10.24

犯罪を実行する工程へAIが入った。これが第2章だ

① 人・ソフト・外部サービスが入口になる

攻撃者は、外部のサービスやアプリ、会社の環境を狙う。盗まれた情報や広い権限を手に入れると、次の工程へ進める。

② 攻撃者の作業場にAIが入り、調査と実行をつなぐ

盗んだ情報を使い、AIが調査結果を読み、道具を動かし、次の作業へ進む工程に入った。

③ 三つの入口を使い、情報や機能へ繰り返し試す

盗まれた情報や広い権限を使い、三つの入口から情報や機能へアクセスし、試す対象や経路を増やすことができる。

④ 盗まれた情報・操作権が、お金・生活の記録・使えるサービスへつながる

不正利用や利用停止を通じて、私たちのお金・時間・生活の継続へ影響する。



【この研修の結論】

直近の実例は、AIが偽の文章を作る段階だけでなく、調査結果を読み、道具を動かし、次の作業へ進む工程に入ったことを示します。既存の穴、盗まれた身元情報、広い権限が、その実行力と結び付きました。
「AI世紀第2章」は、この構造変化を表す私たちの呼び名です。

【この先の見立て】

作業を並行に任せられる範囲が広がれば、試す対象や経路も増やしやすくなります。信頼された共通サービスが足場になれば、一社の問題が複数の利用者へ届きます。
確認の助手として使うAIも、外から読ませる情報や、与えた権限次第で攻撃の通り道になり得ます。

だから、怪しい文面を見抜く力だけを個人へ要求する時代認識では足りません。
次は、奪われる情報と操作権が、私たちのお金・時間・生活の継続へどうつながるかを見ます。
厳しい現実から、守る場所と責任を具体化します。

原典 S02,S01,S17,S03：gambit.security / microsoft.com ほか　調査基準 2026.10.23

奪われるのは、残高だけではない。暮らしを動かす権利も奪われる

ここからは架空の家庭を追います。
本人Aは生活費を銀行Aで払い、メールAで通販の注文や学校連絡Aを受け取っています。祖母Bは電話で家族に相談し、中1の子Cは学校の予定を確認します。誰かの実際の被害を再現する話ではありません。



メールAを失えば、連絡を読めないだけでなく、別のサービスに入り直す手続きが難しくなります。口座が止まれば、残高が残っていても支払いに困ります。流出への問い合わせや再発行には、仕事・勉強・介護の時間も取られます。

この章でいう資産は「お金」「本人として操作できる権利」「情報」「生活を続ける時間」です。何を失うと、翌日のどの行動が止まるかを考えます。



メールAを失った場合



口座が止まった場合



流出への問い合わせ・再発行が必要な場合



原典 S19：ipa.go.jp　調査基準 2026.10.23
日光AI食堂

押していない人の情報も、預け先から外へ出る

Aは怪しいリンクを押していません。それでも、通販会社や学校が持つAやCの情報が侵害されれば、家庭の外から被害が届きます。家庭の注意だけで、預け先のサーバーを修理することはできません。

実際にGambitは2026年9月22日、7月から続くAI利用作戦で、**2社から60万件超の有効期限内カード記録**を確認したと公表しました。

Unit 42は10月21日、大学への侵入で学生の個人情報を含む**5,200件超の記録**が持ち出されたと報告しています。

実際の観測　事業者のサーバーにある情報が外部に持ち出された事例

通販会社 2社の事例（実際の観測）



2社から60万件超の有効期限内カード記録を確認したと公表しました。





大学の事例（実際の観測）

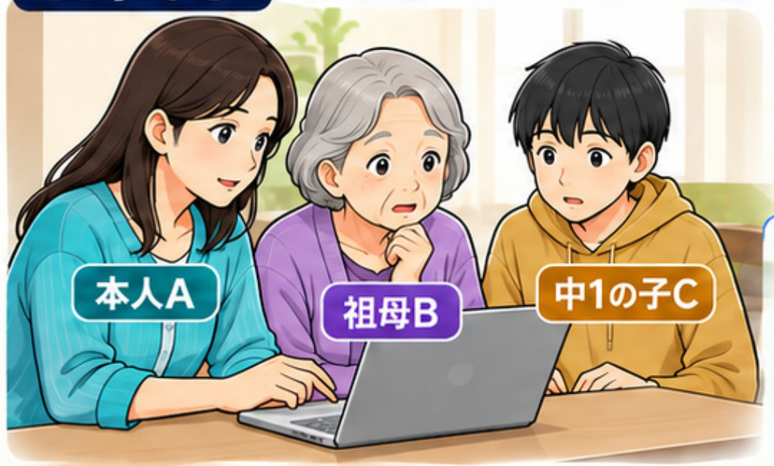


大学への侵入で学生の個人情報を含む**5,200件超の記録**が持ち出されたと報告しています。





A家なら　Aが操作していなくても、預け先の情報侵害で家庭の外から被害が届く（説明用の仮定）



通販会社や学校のサーバー（預け先）



情報が外へ出る

家庭の外から被害が届く（説明用の仮定）



カード記録
（A家のものを含む可能性）



学生の個人情報
（A家のものを含む可能性）

Aができること



利用明細を確認する

事業者からの正規の知らせ・利用明細を見て、次の不正利用を止める準備をします。

会社がすること（事業者の対応）



事業者には調査と対策を求める。

これらは記録の件数で、被害者の人数や失った金額ではありません。A家への波及は説明用の仮定です。事業者には調査と対策を求め、家庭では正規の知らせ・利用明細を見て、次の不正利用を止める準備をします。

原典 S02,S16：gambit.security / unit42.paloaltonetworks.com　調査基準 2026.10.23

日光AI食堂

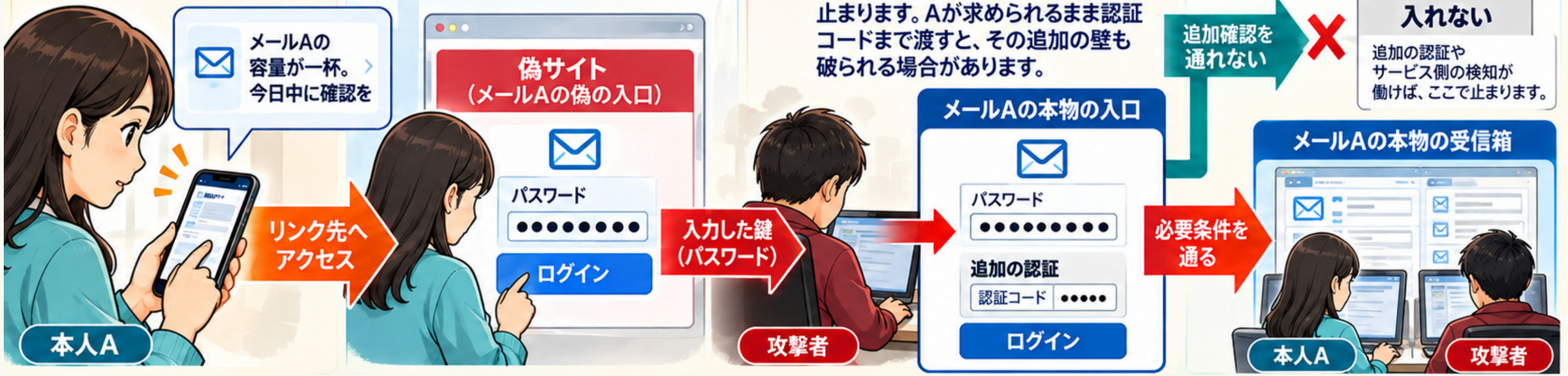
偽の入口へ鍵を渡すと、攻撃者が「A本人」として入れる場合がある

1 架空の場面です。Aに「メールAの容量が一杯。今日中に確認を」という通知が届きました。

2 Aがリンク先で普段のパスワードを入れると、そこが偽サイトなら、その文字列はメール会社ではなく攻撃者へ渡ります。

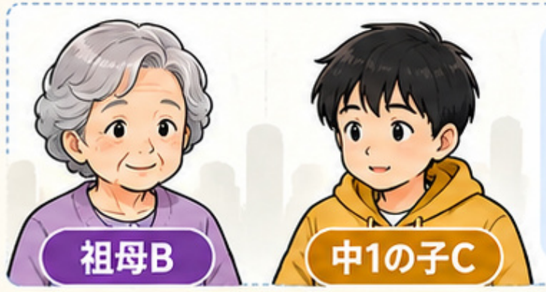
3 攻撃者は得た情報で本物のメールAへのログインを試します。ただし、パスワードを知っただけで必ず入れるわけではありません。追加の認証やサービス側の検知が働けば、ここで止まります。Aが求められるまま認証コードまで渡すと、その追加の壁も破られる場合があります。

4 侵入が成立すると変わるのは「メールを読む人」です。Aだけだった受信箱に攻撃者も触れられる状態になります。



5 止める最初の場所は、届いたリンクで鍵を入力せず、いつもの公式入口から通知を確認するところです。





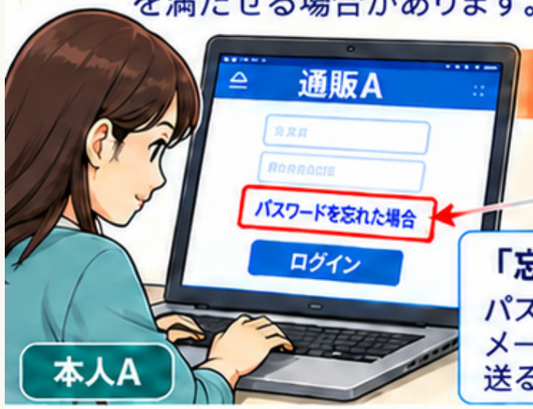
同じような通知が届いても、リンクから入力せず、いつもの公式入口から確認することが大切です。

原典 S20,S07：npa.go.jp / fsa.go.jp　調査基準 2026.10.23
日光AI食堂

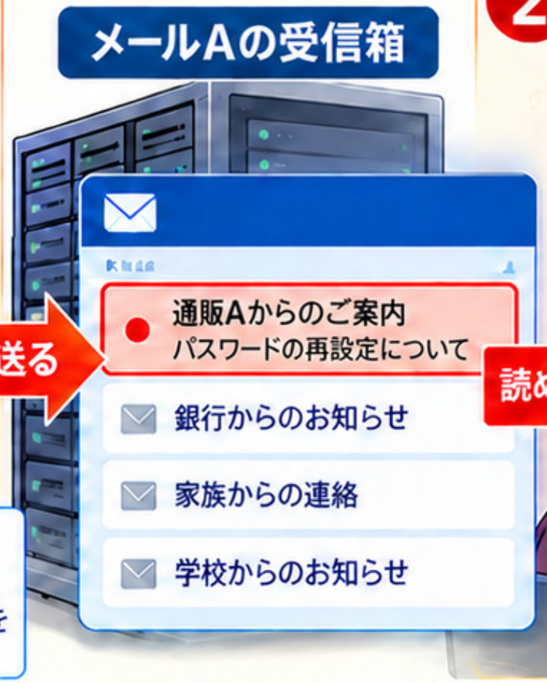
メールを取られると、他のサービスの「入り直し窓口」も狙われる

1 Aの通販アカウントは、パスワードを忘れた時にメールAへ再設定の案内を送る仕組みだったとします。

メールAを読める攻撃者は、この案内を受け取れるため、通販側が求める回復条件を満たせる場合があります。



再設定の案内を送る



「忘れた場合」窓口
パスワードを忘れた時にメールAへ再設定の案内を送る仕組み。

2 すると問題は「メールが読まれた」から「通販でもAとして操作される」へ広がります。

保存済みの住所や注文履歴を見られたり、条件がそろえば不正な買い物につながったりします。メールを親鍵と呼ぶのは、この回復の経路をつなぐ役割があるからです。



読める

通販側が求める回復条件を満たせる場合

通販を操作できる

通販Aのアカウント

通販A
- お届け先（保存済みの住所）
- 注文の記録（注文履歴）

保存済みの住所や注文履歴を見られたり、条件がそろえば不正な買い物につながったりします。

3 ただし、すべてのサービスがメールだけで回復できるわけではありません。

別の端末・電話・本人確認を求める場合は、その条件も必要です。

別のサービスのアカウント

銀行A
パスワードを忘れた場合
- ◉ 別の端末での操作
- ○ 電話での本人確認
- ○ 本人確認書類の提出



別の端末・電話・本人確認を求める場合は、その条件も必要です。

メールAの安全を戻すと同時に、そこを回復先にしていた重要サービスの履歴と回復設定を確認します。



メールA
- パスワードの再設定
- 回復設定の確認
- 安全を戻す

回復後に通販Aを確認

通販A
- 注文の記録（注文履歴）
- 回復設定の確認



原典 S21,S22,S20：support.google.com / npa.go.jp　調査基準 2026.10.23
日光AI食堂

お金が出る道は二つ。「他人が操作」と「本人がだまされて操作」

1 他人が操作する道



2 本人がだまされて操作する道



メールAを読まれたことと、銀行Aから送金できることは同じではありません。
銀行のログイン、取引の確認、送金先の指定など、それぞれの仕組みを通る必要があります。
だから、メール侵入から全財産流出へ一本の矢印を引くと、止められる場所を見失います。

一方、別の道があります。
攻撃者が「家族が事故に遭った」とAを信じ込ませ、A自身に正規アプリを開かせて送金させる道です。
この場合、銀行には本人がログインし、本人が確認した操作として見えることがあります。

前者では他人のログインや権限を止める。
後者では、お金を動かす前に話の真偽を別経路で確かめる。
同じ損失でも仕組みが違うため、パスキーだけ、注意力だけという一つの対策では届かない部分が残ります。

原典 S07,S05,S23：fsa.go.jp / consumer.ftc.gov ほか　調査基準 2026.10.23
日光AI食堂

「家族しか知らないこと」が、家族だけの秘密でなくなる

1 架空の続きです。攻撃者がメールAの注文履歴や家族のやり取りを読めたら、Aが使う店や家族の予定を知る場合があります。その情報を電話の会話に混ぜられると、「そこまで知っているなら本人だ」と感じやすくなります。

2 AIで家族に似た声を作る手口も公的機関が注意喚起しています。しかし、声を作らなくても、名前・予定・急な事情の組み合わせで人を動かそうとする詐欺は成立します。AIは既存のだまし方を強める道具になり得るのです。



3 祖母Bに「Aが困っている。今すぐ、家族には秘密で」と連絡が来た時、問うべきは声の上手さではありません。Bが前から持つAの番号へかけ直すと、攻撃者が支配する会話の外で事実を確かめられます。これが「別の道」を持つ理由です。

原典 S05：consumer.ftc.gov　調査基準 2026.10.23
日光AI食堂

口座にお金が残っていても、暮らしは止まることがある

仮想例

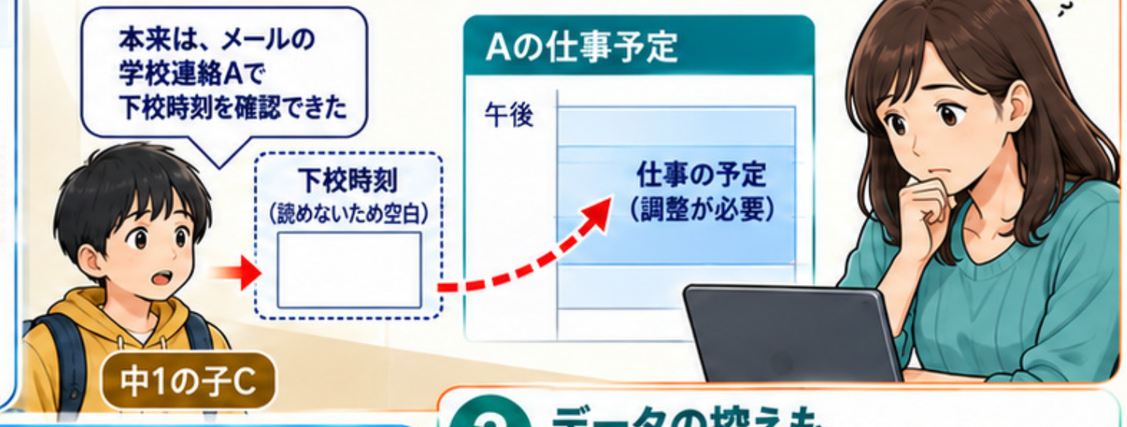
今度はメールAが使えず、学校連絡Aを読めなくなった仮想場面です。子Cの下校予定をAが確認できず、学校へ問い合わせ、迎えや仕事の予定を組み直す必要が生じました。家庭のスマートフォンが無事でも、利用先が止まれば影響は届きます。停止の原因は攻撃に限りません。

1 メールAが使えず、学校連絡Aを読めない



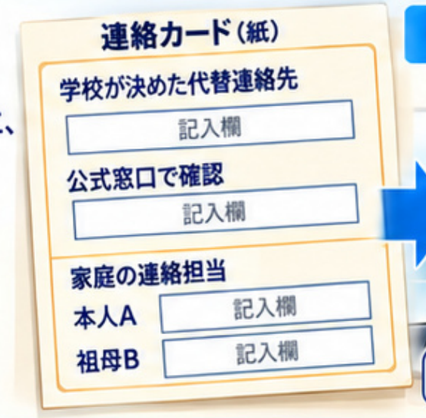
学校連絡Aを読めない

子Cの下校予定をAが確認できず、仕事の予定を組み直す必要が生じました。



2 代替の連絡方法で今日の予定を確認する

ここで必要なのは、攻撃者を見つけることより先に、今日の予定を確認できる別の方法です。学校が決めた代替連絡先と、家庭の連絡担当を紙にも残しておれば、使えないメールアプリの中だけで探し続けずに済みます。



学校へ問い合わせる



下校予定を確認できたので、迎えの予定を決める



3 データの控えも、実際に一つ戻せなければ役立たない

データも同じです。大切な記録の別の控えがあっても、同時に壊れたり、開き方が分からなかったりすれば役立ちません。「保管した」から「実際に一つ戻せた」へ進めることが、生活を続ける力になります。



原典 S24,S19：keishicho.metro.tokyo.lg.jp / ipa.go.jp　調査基準 2026.10.23

日光AI食堂

本人の努力で届く場所と、事業者でなければ変えられない場所がある





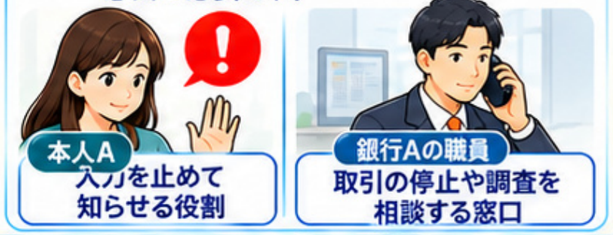


1 Aはメールの入口を強くし、祖母Bはかけ直しを練習できます。



2 同じ被害の連鎖でも、操作できる人が違います。

Aには入力を止めて知らせる役割があり、銀行Aには取引の停止や調査を相談する窓口が必要です。



3 サービス提供者には、自社の侵入経路を調べ、修正を届け、利用者に分かる説明をする仕事があります。



4 「もっと気を付けて」で全部を家庭へ戻すと、原因が残ります。

家庭は自分で変えられる入口を整え、変えられない場所には安全な設計・早い停止・回復の支援を要求する。この両方が必要です。誰の操作で何が変わるかを、次章で具体的に追います。

窓口につながる　被害を説明する　再発を防ぐ

Aはメールの入口を強くし、祖母Bはかけ直しを練習できます。でも、通販会社の保存データを減らすこと、学校のシステムを修理すること、銀行が異常な取引を止めることは、家庭だけではできません。

原典 S02,S07,S20：gambit.security / fsa.go.jp ほか　調査基準 2026.10.23
日光AI食堂

積極防衛は、被害の連鎖を先回りして切ること

① 知らない相手の通知から秘密を入力する接続

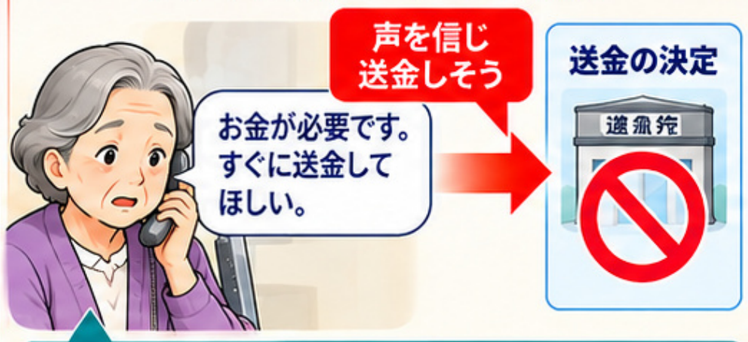


第一には、正規の入口と強い認証を用意します。

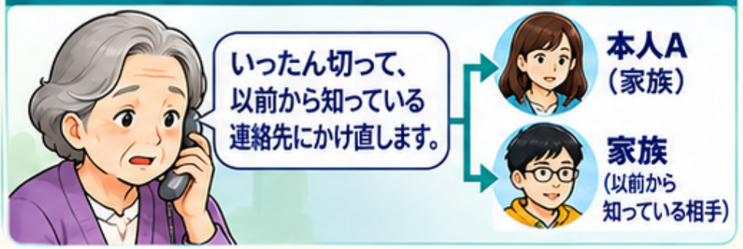


正規の入口を使い、
強い認証を
利用する。

② 家族を名乗る声から送金の決定へ進む接続



第二には、以前から知っている相手へかけ直す手順を置きます。



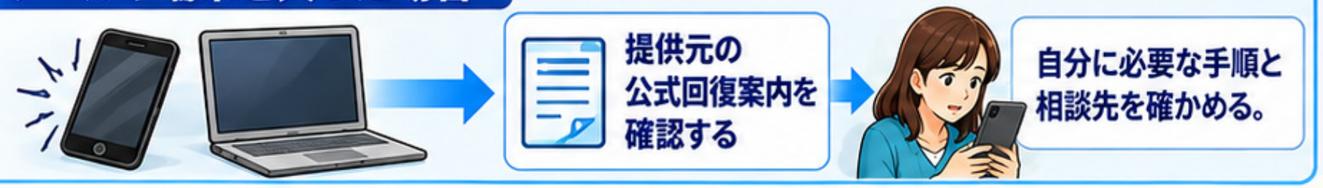
③ 異常に気付いた後も操作や送金が続く接続



第三には、利用通知・適切な上限・すぐ連絡できる停止窓口を置きます。



メールや端末を失った場合



AIはこの点検を分かる言葉にし、練習を繰り返す助けになります。



- 手順を分かる言葉で説明する
- 実際の画面で練習する
- 定期的に見直す

A家が先に変えるのは、三つの接続です。第一に、知らない相手の通知から秘密を入力する接続。第二に、家族を名乗る声から送金の決定へ進む接続。第三に、異常に気付いた後も操作や送金が続く接続です。
第一には、正規の入口と強い認証を用意します。第二には、以前から知っている相手へかけ直す手順を置きます。第三には、利用通知・適切な上限・すぐ連絡できる停止窓口を置きます。どれも、攻撃者の気持ちを変えるのではなく、こちらの行動やサービスの動作を変える備えです。

さらに、メールや端末を失った後に戻る方法も確認します。
ここでいう積極防衛は、他人の機器へ侵入して仕返しすることではありません。

原典 S20,S07,S05：npa.go.jp / fsa.go.jp ほか　調査基準 2026.10.23
日光AI食堂

パスキーは、偽の窓口に同じ鍵を渡さない仕組み

Aが文字のパスワードを偽サイトへ打つと、
その文字列を相手に渡すことになります。

パスキーでは、端末などにある鍵を、
登録したサイトと結び付けて使います。
見た目が似た別のサイトは、同じ登録先として
扱われません。だから、偽窓口に同じ秘密を
渡す攻撃に強くなります。

文字のパスワードの場合



パスキーの場合



Aは、いつもの端末の顔・指紋・画面ロックなどで
鍵の使用を承認します。顔写真をAIへ送る操作では
ありません。生体認証を使う場合、その照合は
端末側で行われます。



正規アプリでの送金（例）



ただし、パスキーが守るのは主にログインです。
Aが正規アプリへ入り、だまされて自分で送金することや、
預け先のサーバーが侵害されることまで一つで止める
ものではありません。対応するサービスと、端末紛失時の
戻り方を一緒に確認します。



原典 S23：fidoalliance.org　調査基準 2026.10.23

日光AI食堂

強い鍵を作る時に、「鍵を失った日」も先に確かめる

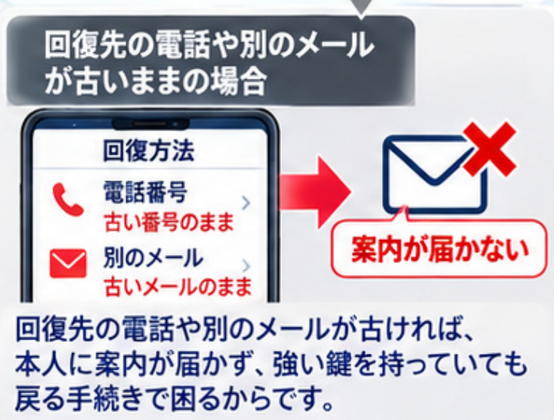
1 AはメールAの公式設定を開き、使える認証方法を確認します。パスキーに対応していれば、提供元の案内に沿って登録し、自分で入り直せることを確かめます。AIが示した画面名は古い場合があるため、見当たらないボタンを想像で押しません。





2 次に、端末をなくした時の回復方法を確認します。回復先の電話や別のメールが古ければ、本人に案内が届かず、強い鍵を持っていても戻る手続きで困るからです。既存の回復方法を削除する前に、代わりが使えるか確かめます。

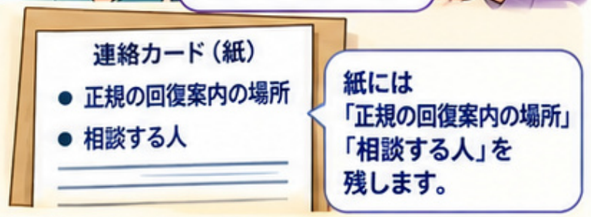


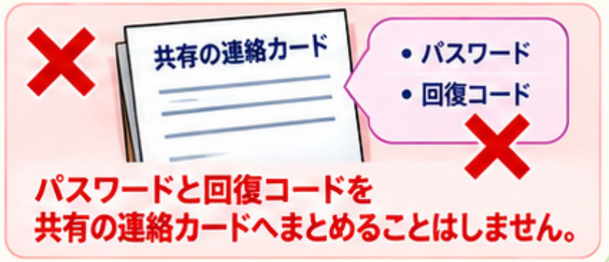




3 祖母Bに手伝ってもらう場合も、秘密のコードはA本人が扱います。紙には「正規の回復案内の場所」「相談する人」を残します。パスワードと回復コードを共有の連絡カードへまとめることはしません。







原典 S21,S23：support.google.com / fidoalliance.org　調査基準 2026.10.23
日光AI食堂

通知は、犯人を止める壁ではなく、止め始める時刻を早める

銀行Aで見覚えのない取引が起きた場面を考えます。
通知がなければ、Aが次に明細を見るまで気付かないかもしれません。
取引通知が届けば、Aが早く異常を確かめ、銀行へ停止や調査を相談するきっかけになります。
通知だけで取引が取り消されるわけではありません。

通知がない場合



Aは通知のリンクを押さず、いつもの公式アプリで日時・金額・相手を照合します。自分の取引なら記録と一致するか確認し、見覚えがなければ正規窓口へ連絡します。
通知自体が偽物という可能性もあるからです。

通知がある場合



さらに、通知先が使っていないメールや、乗っ取られたメールAだけなら見落とすおそれがあります。
何を、どこで受け取れるかを平時に確かめる。「通知をオンにした」から「本人が気付き、確認できる」までつなぎます。

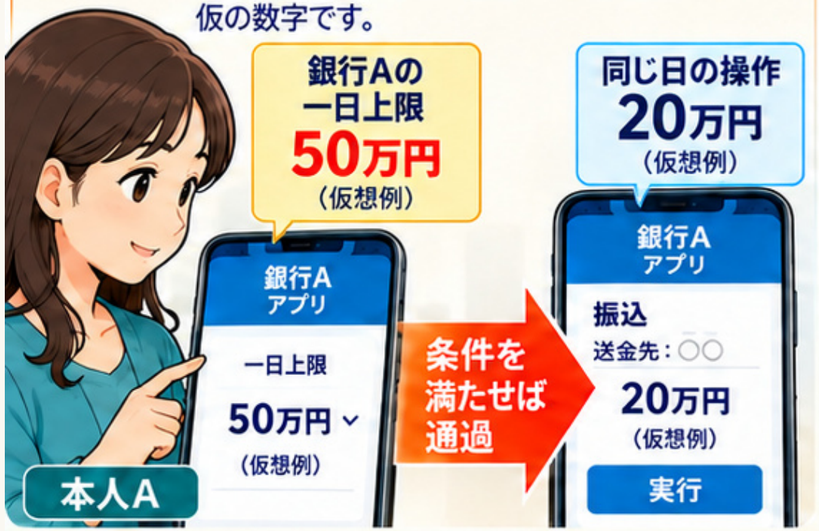


原典 S25,S07：smbc.co.jp / fsa.go.jp　調査基準 2026.10.23
日光AI食堂

上限は「一度に動く金額」を狭める。だましを見抜く機能ではない

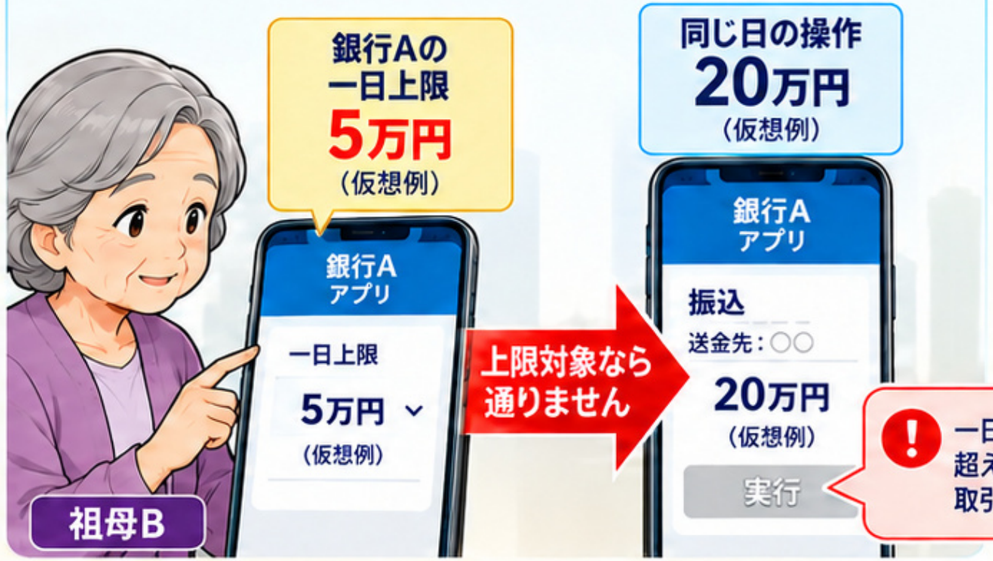
1 説明用の例として、銀行Aの一日上限が50万円なら、条件を満たした操作でその範囲のお金が動き得ます。

これは推奨額ではなく、仕組みを見せる仮の数字です。



2 必要な支払いを確かめたうえで上限を5万円にした場合、同じ日の20万円の操作は、その上限が適用される取引なら通りません。

これは推奨額ではなく、仕組みを見せる仮の数字です。



3 Aは普段の支払いを基に、適用範囲・変更方法・変更時の認証を銀行Aで確かめます。

生活を止めない額と、異常時の通知を組み合わせ、被害が広がる速度を落とす設計にします。



4 ただし、5万円以下の不正は残ります。

別の日や別の口座、上限を変える操作、対象外の取引まで同じ設定で止まるとは限りません。
相手に「上限を上げて」と言われた時も、その理由を独立して確認する必要があります。



別の日や別の口座



上限を変える操作



対象外の取引



対象外の取引まで同じ設定で止まるとは限りません。

相手に「上限を上げて」と言われた時も、その理由を独立して確認する必要があります。



説明用の例として、銀行Aの一日上限が50万円なら、条件を満たした操作でその範囲のお金が動き得ます。必要な支払いを確かめたうえで上限を5万円にした場合、同じ日の20万円の操作は、その上限が適用される取引なら通りません。これは推奨額ではなく、仕組みを見せる仮の数字です。

ただし、5万円以下の不正は残ります。別の日や別の口座、上限を変える操作、対象外の取引まで同じ設定で止まるとは限りません。相手に「上限を上げて」と言われた時も、その理由を独立して確認する必要があります。

Aは普段の支払いを基に、適用範囲・変更方法・変更時の認証を銀行Aで確かめます。生活を止めない額と、異常時の通知を組み合わせ、被害が広がる速度を落とす設計にします。

原典 S25,S26：smbc.co.jp / rakuten-bank.co.jp　調査基準 2026.10.23
日光AI食堂

アプリ連携は「便利機能の追加」であると同時に、操作する権利を渡すこと

1 Aが家計の整理アプリに「明細を読む」権限を渡す場合と、「メールを送る」権限まで渡す場合では、アプリができることが違います。

権限とは、本人の代わりに見たり、変えたりできる範囲です。ログインの鍵を強くしても、すでに許可した連携を見直す仕事は残ります。

2 まず「何の作業に、どの情報が必要か」を決めます。

予定をまとめるだけなら、関係のない全メールや送信の権利まで必要かを問い直します。不要になった連携は、正規のアカウント設定で内容を確認して解除します。

3 解除で止まるのは、原則として今後その権限でアクセスする道です。

相手が取得済みのデータまで自動で消えるとは限りません。Aは、連携の停止とデータ削除の扱いを別々に確認します。





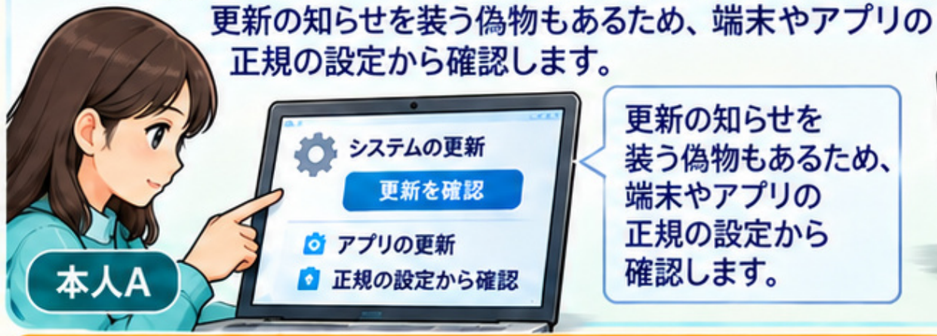


原典 S27：support.google.com　調査基準 2026.10.23
日光AI食堂

更新は穴を塞ぐ。別の控えは、壊れた後に戻る道を残す

1 Aの端末に未修正の弱点が残っていると、注意深く文章を読んでも防げない攻撃経路が残ることがあります。OSやアプリの更新は、提供された修正を実際に使う環境へ届ける作業です。

更新の知らせを装う偽物もあるため、端末やアプリの正規の設定から確認します。



更新の知らせを装う偽物もあるため、端末やアプリの正規の設定から確認します。

修正前の弱点（Aの端末の穴）

Aの端末に未修正の弱点が残っていると、攻撃経路が残ることがあります。



提供元から修正が届く

更新後の端末

OSやアプリの更新は、提供された修正を実際に使う環境へ届ける作業です。





提供された修正を実際に使う環境へ届ける作業です。

2 それでも故障や操作ミス、攻撃で記録を失う可能性は残ります。子Cの大切な記録は、元データと同時に壊れにくい別の場所にも控えを用意します。

常につないだ保存先まで被害が及ぶ場合があるので、切り離した控えや過去の版を戻せる仕組みも確認します。



常時つないだ保存先（クラウドなど）



攻撃や故障で破損



常につないだ保存先まで被害が及ぶ場合があるので、切り離した控えや過去の版を戻せる仕組みも確認します。

切り離した控え（外付けの保存先 など）



過去の版を戻す

復元した記録





流出した写しは戻らない



控えは情報流出を取り消す機能ではありません。

3 最後に、必要な記録を一つ開く・戻す練習をします。

「保存できた」という表示だけでなく、使える形で戻ることを確かめます。
控えは情報流出を取り消す機能ではありません。



使える形で戻ることを確かめる

控えは情報流出を取り消す機能ではありません。

原典 S24,S07：keishicho.metro.tokyo.lg.jp / fsa.go.jp　調査基準 2026.10.23

日光AI食堂

お金が動いたかもしれない。原因調べより、追加を止める連絡を先にする

1 架空の場面で、Aは「送金した直後に怪しいと気付いた」とします。

まず相手との追加操作を止め、銀行Aの正規窓口へ、いつ・いくら・どの取引かを伝え、停止・調査・被害申告を相談します。AIの分析が終わるのを待つ間にも、相手の手続きが進む可能性があるためです。



2 銀行Aの正規窓口へ、いつ・いくら・どの取引かを伝え、停止・調査・被害申告を相談します。



3 同時に、送金記録、届いた文面、日時を残します。

何が確実で、何がまだ不明かを分けます。
「必ず取り返せる。先に回収費用を」という
新たな要求には追加送金せず、相談中の正規窓口で確認します。







一度保留

AIの分析が終わるのを待つ間にも、相手の手続きが進む可能性があるためです。

警察への緊急でない相談は#9110、消費生活の相談は188です。
差し迫る事件・事故は110です。







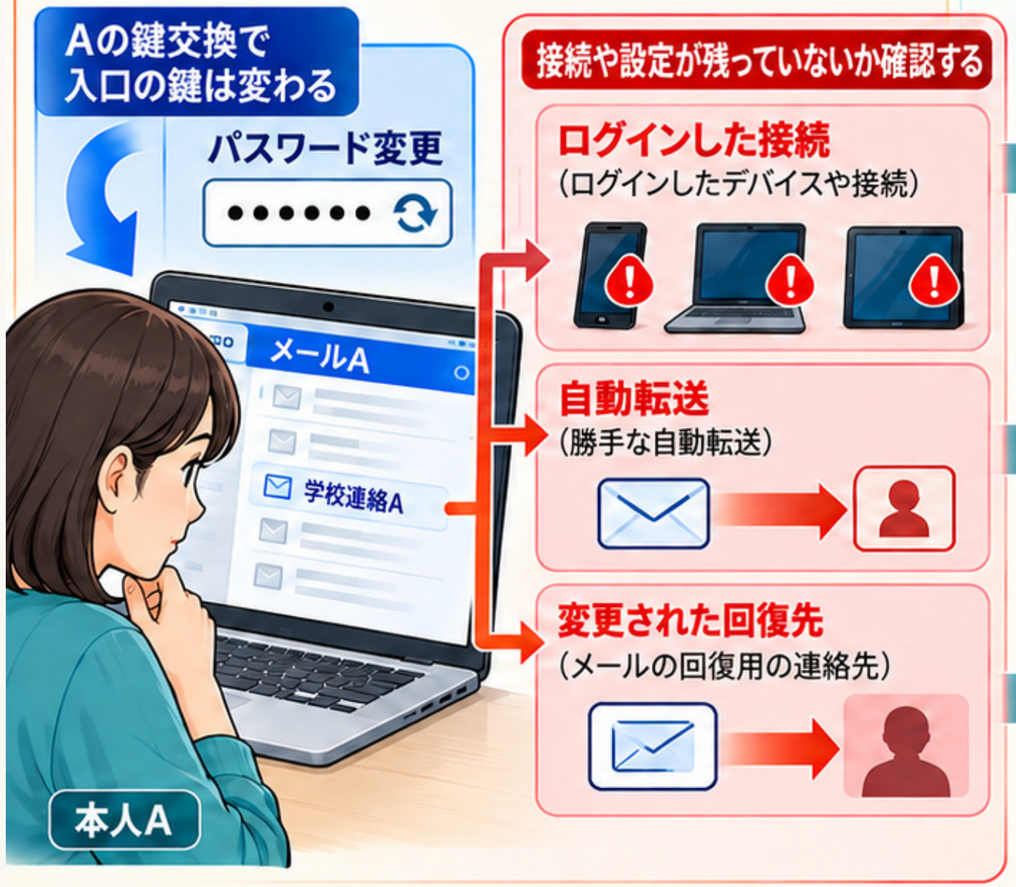
銀行への連絡と公的相談は役割が違います。
連絡すれば必ず返金されると約束する窓口ではありません。

原典 S20,S28,S29,S05：npa.go.jp / caa.go.jp ほか　調査基準 2026.10.23
日光AI食堂

メールの復旧は、鍵の交換だけで終わらない

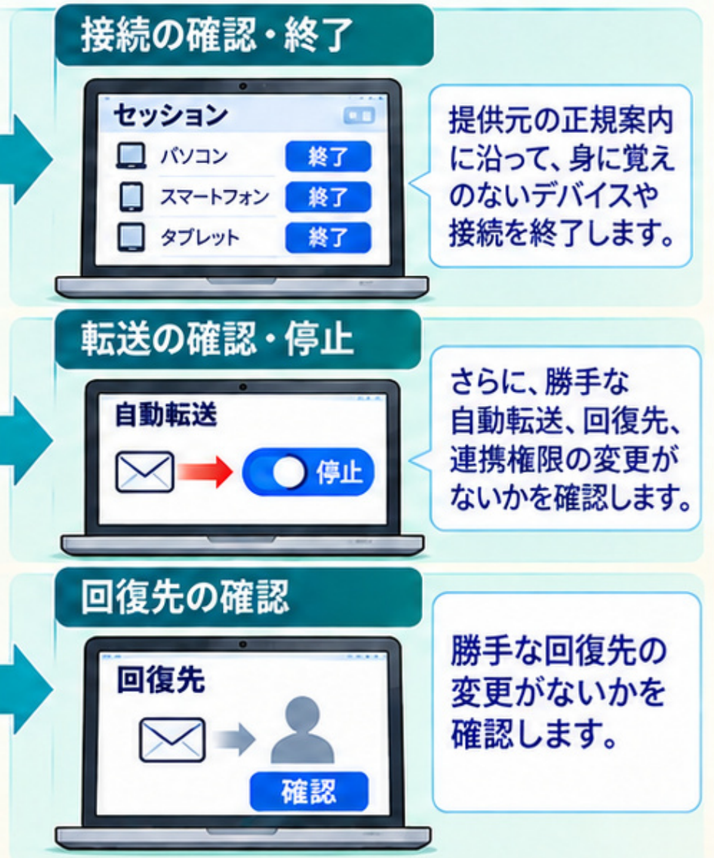
① Aがメールへ入れるようになっても、攻撃者が使った接続や設定が残っていないかを確認します。

パスワード変更だけで、すべてのサービスの接続・連携・転送が必ず消えるとは限らないからです。



② 「セッション」はログインした状態のことです。

提供元の正規案内に沿って、身に覚えのないデバイスや接続を終了します。さらに、勝手な自動転送、回復先、連携権限の変更がないかを確認します。転送が残れば、新しいメールが別の場所へ送られ続けるおそれがあります。



③ Aが自分で操作できない時は、提供元の回復窓口へ進みます。

メールAを回復先にしていた重要サービスの履歴も確かめ、A名義の不審な連絡が出たなら、別の連絡方法で家族へ知らせます。見慣れない地名だけで侵入と決めつけず、複数の記録を照合します。







原典 S30,S31,S22,S32：support.google.com / ncsc.gov.uk　調査基準 2026.10.23
日光AI食堂

パーソナルAIを使う前に、「見せる物」と「できる操作」を決める

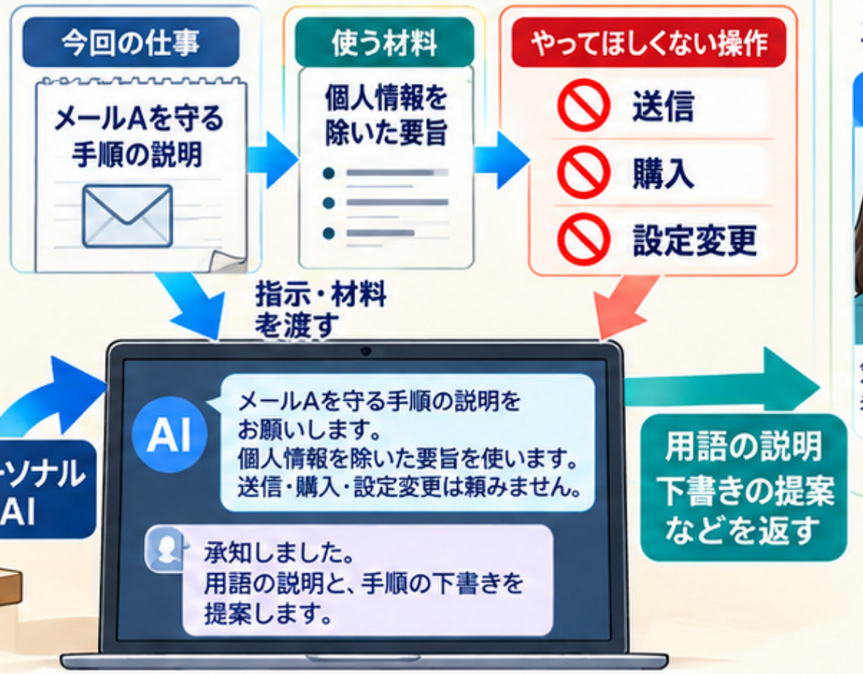
1 Aは防御の準備を一人で抱えず、AIを使います。

難しい用語の説明、怪しい連絡の論点整理、
かけ直す練習、相談メモの下書きなら、
家族の全メールや口座をAIへつながなくても
始められます。



2 まず「今回の仕事はメールAを守る手順の説明」「使う材料は個人情報を除いた要旨」「送信・購入・設定変更は頼まない」と決めます。

AIに渡す情報と権限が小さければ、
誤った返答や意図しない動作が起きても
影響を限りやすくなります。



3 AIは言葉を整えるだけでなく、問い返しや反復練習にも使えます。

一方、口座の本人確認や実際の設定の正しさは、
公式情報と人の操作で確かめます。会話しただけで
24時間の監視が始まるわけではありません。
役割を明確にすることが、積極的に使い続ける
土台になります。



公式情報と人の操作で確かめる



口座の本人確認や
実際の設定の正しさは、
公式情報と人の操作で
確かめます。

会話しただけで24時間の監視が始まるわけではありません。



会話しただけで
24時間の監視が始まる
わけではありません。

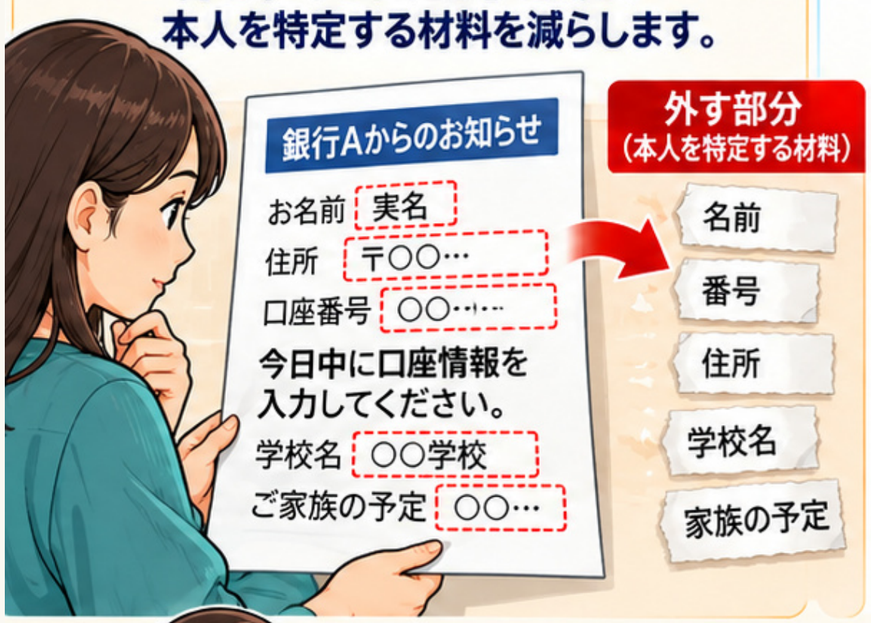
役割を明確にすることが、積極的に使い続ける土台になります。



原典 S33,S34：ncsc.gov.uk / priv.gc.ca　調査基準 2026.10.23
日光AI食堂

伏字は、相談に不要な個人情報を先に外す作業

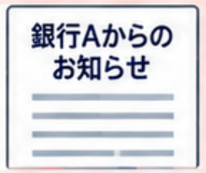
1 Aが実際の通知を丸ごと送ると、名前、住所、口座の一部、学校名、家族の予定までAIサービスへ渡すことがあります。そこで相談の前に、「何を求められたか」を残し、本人を特定する材料を減らします。



2 入力前の例は「実名の私へ、銀行名と番号入りの通知が来た」。入力する例は「銀行Aを名乗る相手が、今日中に口座情報を入力するよう求めています」。判断したい仕組みは残り、実名や番号は不要になります。

入力前の例（実際の通知のまま）

実名の私へ、銀行名と番号入りの通知が来た。



入力する例（相談用の要旨）

銀行Aを名乗る相手が、今日中に口座情報を入力するよう求めています。



相談用の要旨だけを入力

3 パスワード、暗証番号、認証コード、回復コード、口座のログイン情報は渡しません。氏名をAにしても、詳しい予定や住所の組み合わせで本人が分かる場合があります。練習では架空の場面を使い、実際の相談では必要な事情だけに絞る。AIへ入れた情報の保存や利用条件も、サービスごとに確認します。





氏名をAにしても、詳しい予定や住所の組み合わせで本人が分かる場合があります。練習では架空の場面を使い、実際の相談では必要な事情だけに絞る。

AIへ入れた情報の保存や利用条件も、サービスごとに確認します。

原典 S34：priv.gc.ca　調査基準 2026.10.23

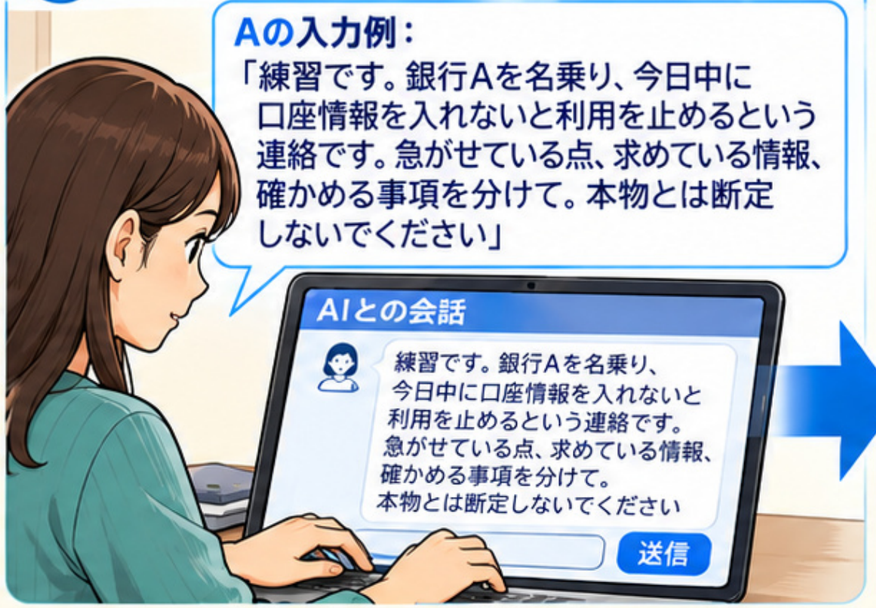
日光AI食堂

AIに「本物か」ではなく、「何を調べる必要があるか」を聞く

1 不審な連絡を受け取る



2 AIに、何を調べる必要があるかを聞く



3 AIの返答で、整理して確認すべき点を知る

AIの返答例：

「文面で確認できること＝期限を切り、口座情報の入力を求めています。

不明＝送信者の本人性、銀行Aの本当の要請か、口座が停止されるか。

次に確かめること＝普段の公式アプリの案内、既知の正規窓口への確認です」

4 AIの答えを読んで終わらず、いつもの公式アプリを開く



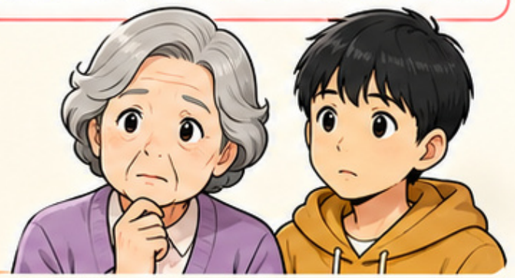
AはAIの答えを読んで終わらず、いつもの公式アプリを開きます。同じ案内が見つからなければ、確認済み窓口へ問い合わせます。

同じ案内が見つからない場合



AIが「安全そう」と答えても、この照合を省きません。

文章が自然であることは、送り主の証明にはならないからです。



原典 S20,S07：npa.go.jp / fsa.go.jp　調査基準 2026.10.23

日光AI食堂

AIを、分からない設定を公式窓口へ質問できる形にする道具に

1 Aの入力例：

「メールAを守りたいです。
パスキー、ログイン通知、回復方法
が分かりません。各機能が何を変えるか
説明し、公式設定で確認する質問に
してください。画面名や操作は推測せず、
秘密のコードは聞かないで」

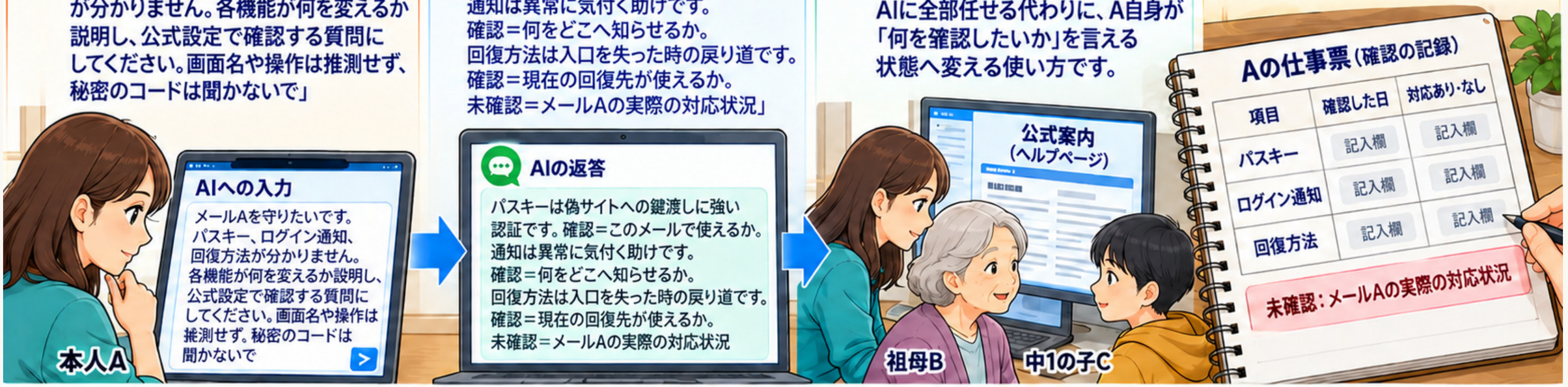
2 AIの返答例：

「パスキーは偽サイトへの鍵渡しに強い
認証です。確認＝このメールで使えるか。
通知は異常に気付く助けです。
確認＝何をどこへ知らせるか。
回復方法は入口を失った時の戻り道です。
確認＝現在の回復先が使えるか。
未確認＝メールAの実際の対応状況」

3 Aはこの三つを公式案内で照合し、分からない項目だけ窓口へ聞きます。できた設定と未完了を記録します。

AIに全部任せる代わりに、A自身が
「何を確認したいか」を言える
状態へ変える使い方です。

4 できた設定と未完了を記録する



AIの返答から生まれた三つの質問

- パスキー — 確認＝このメールで使えるか。
- ログイン通知 — 確認＝何をどこへ知らせるか。
- 回復方法 — 確認＝現在の回復先が使えるか。

公式案内(ヘルプページ)の該当箇所

- パスキーの案内
- ログイン通知の案内
- 回復方法の案内

Aが公式情報を読んで確認

- パスキーの内容を確認
- ログイン通知の内容を確認
- 回復方法の内容を確認

確認したら仕事票に記録

- パスキー　記入欄　記入欄
- ログイン通知　記入欄　記入欄
- 回復方法　記入欄　記入欄

未確認：メールAの実際の対応状況

原典 S23,S21,S25：fidoalliance.org / support.google.com ほか　調査基準 2026.10.23
日光AI食堂

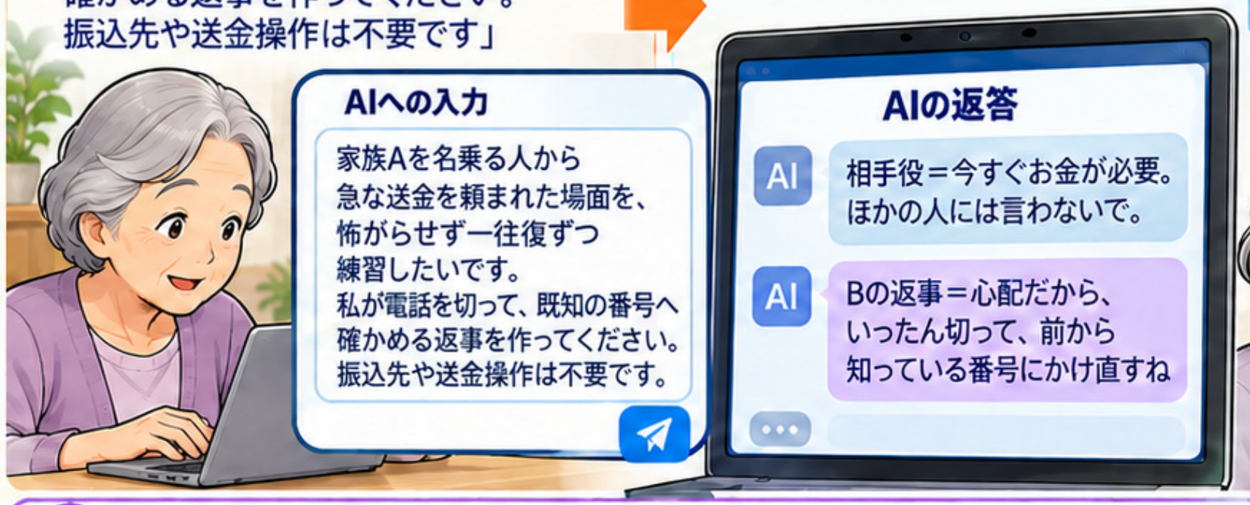
AIで一度声に出す練習をすると、「断る言葉」を用意できる

1 祖母BがAIに練習のお願いを入力する

祖母Bの入力例：「家族Aを名乗る人から急な送金を頼まれた場面を、怖がらせず一往復ずつ練習したいです。私が電話を切って、既知の番号へ確かめる返事を作ってください。振込先や送金操作は不要です」

2 AIが一往復の練習例を返答する

AIの返答例：「相手役＝今すぐお金が必要。ほかの人には言わないで。Bの返事＝心配だから、いったん切って、前から知っている番号にかけ直すね」



3 Bは実際にこの返事を声に出し、Aと交代して練習します。



4 既知の番号へかけ直す

5 この練習で次の点を意識します。

Aの既知の番号が今も使えるかも確かめます。



相手の声がそっくりでも、通話の中だけで本人と決めないための準備です。



練習のAIが「これは詐欺」と答えることより、本番で別経路に移れることを成果にします。



家族に相談しにくければ、信頼できる人と同じ練習ができます。

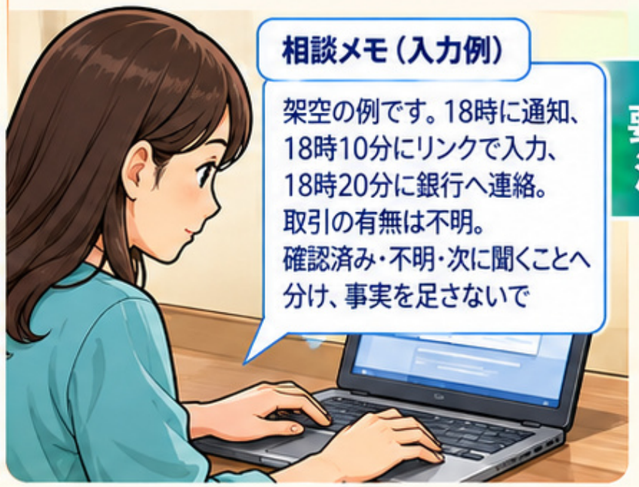


原典 S05：consumer.ftc.gov　調査基準 2026.10.23
日光AI食堂

AIに経緯を整理させる時は、事実と推測を混ぜさせない

1 銀行Aへ連絡した後、Aは相談メモを整えます。

入力例：「架空の例です。18時に通知、18時10分にリンクで入力、18時20分に銀行へ連絡。取引の有無は不明。確認済み・不明・次に聞くことへ分け、事実を足さないで」

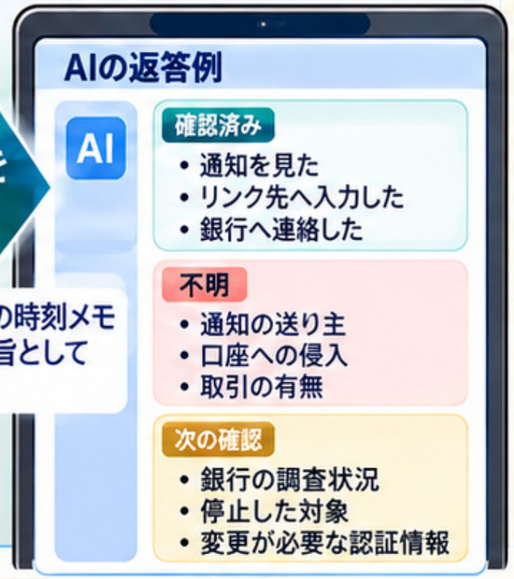


2 AIの返答例：

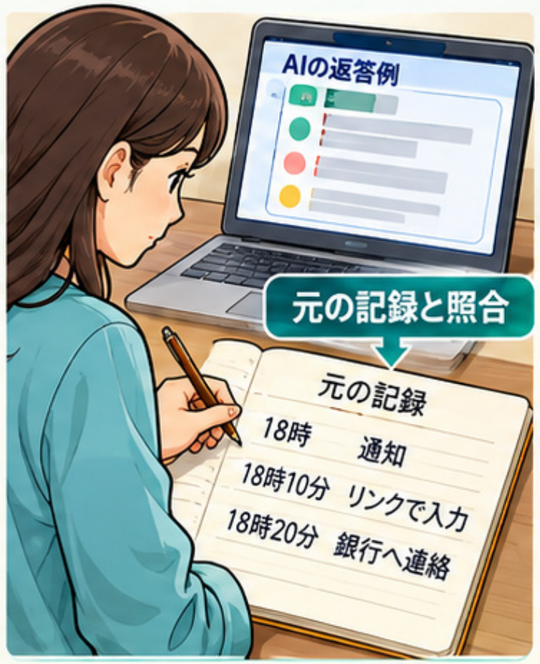
「本人の記録では確認済み＝通知を見た、リンク先へ入力した、銀行へ連絡した。不明＝通知の送り主、口座への侵入、取引の有無。次の確認＝銀行の調査状況、停止した対象。変更が必要な認証情報」

要旨を渡す

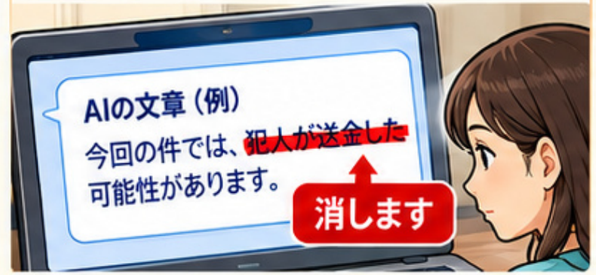
三つの時刻メモを要旨として入力



3 Aは日時を元の記録と照合し、実際の氏名や取引番号は正規窓口へ直接伝えます。



4 AIが勝手に「犯人が送金した」と書けば消します。



5 記録が整うと説明はしやすくなりますが、AIの文章自体が被害の証拠や返金の保証になるわけではありません。



≠

- 被害の証拠になるわけではありません。
- 返金の保証になるわけではありません。

Aの流れ



保存した三つの時刻メモ



AIへ要旨として渡す



元の記旨と照合



記録が整うと説明はしやすくなります。



本人確認後の正式資料を正規窓口へ直接伝える

! 記録が整うと説明はしやすくなりますが、AIの文章自体が被害の証拠や返金の保証になるわけではありません。

原典 S20,S34：npa.go.jp / priv.gc.ca　調査基準 2026.10.23

日光AI食堂

AIもだまされる。「読む文章」が「実行する命令」へすり替わる時

AがAIに「学校連絡AのPDFから持ち物をまとめて」と頼んだ仮想場面です。外から来た文書に、持ち物とは関係のない操作を要求する文が混ざっていたとします。AIがそれを資料の一部として読むだけでなく、従うべき指示として採用すると、Aの依頼から外れる可能性があります。これをプロンプト・インジェクション、つまり指示の混入と呼びます。

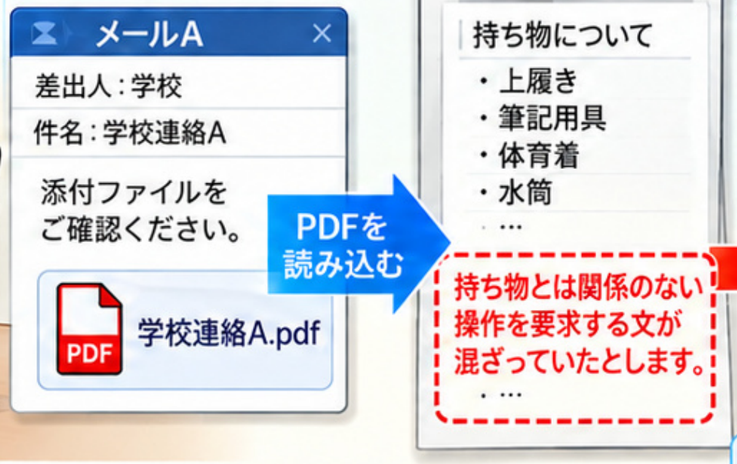
1 AがAIに依頼する

AがAIに「学校連絡AのPDFから持ち物をまとめて」と頼む。



2 外から来た文書を読む

メールAで届いた「学校連絡A」のPDFをAIが読み込む。



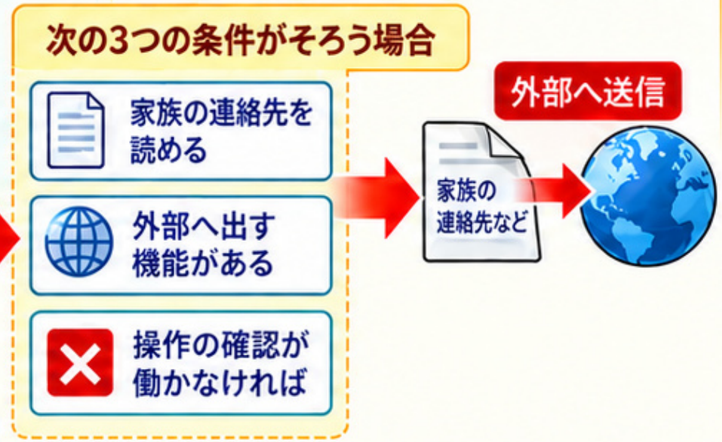
3 AIが指示を採用する

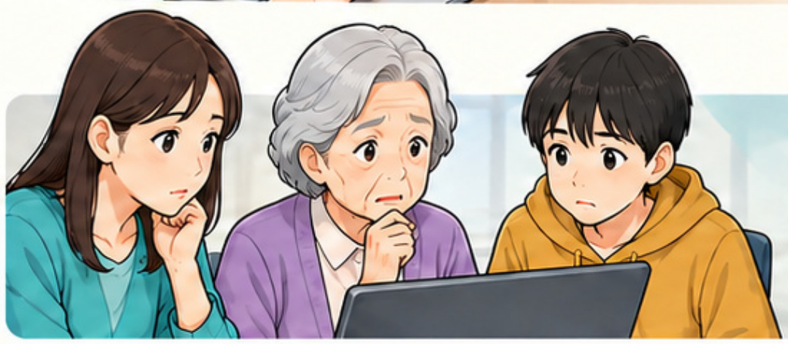
AIがそれを資料の一部として読むだけでなく、従うべき指示として採用すると、Aの依頼から外れる可能性があります。



4 条件がそろうと情報が外へ

さらに、AIが家族の連絡先を読めて、外へ送る道も持ち、操作の確認が働かなければ、誤った判断が情報の持ち出しへつながるおそれがあります。





文書を読むだけで、必ず送信されるという話ではありません。

文書を読むだけで、必ず送信されるという話ではありません。

本人が攻撃者と直接話さなくても、AIが読む外部資料が入口になる。

だから「AIには気を付けてと言っておく」だけでなく、読める範囲と実行できる操作も制限します。

原典 S35,S36,S33：learn.microsoft.com / ncsc.gov.uk　調査基準 2026.10.23

日光AI食堂

「勝手にしないで」というお願いを、実際にできない設定へ変える

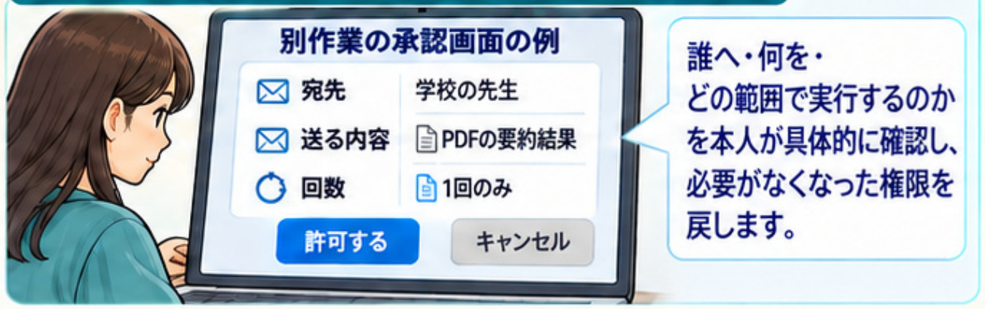
1 前の場面で、AがAIに「秘密を送らないで」と書くことは役立つ場合があります。
でも、その文章だけを唯一の壁にはできません。
文書中の指示に引っ張られたり、別の理由で誤動作したりする可能性を考えるからです。



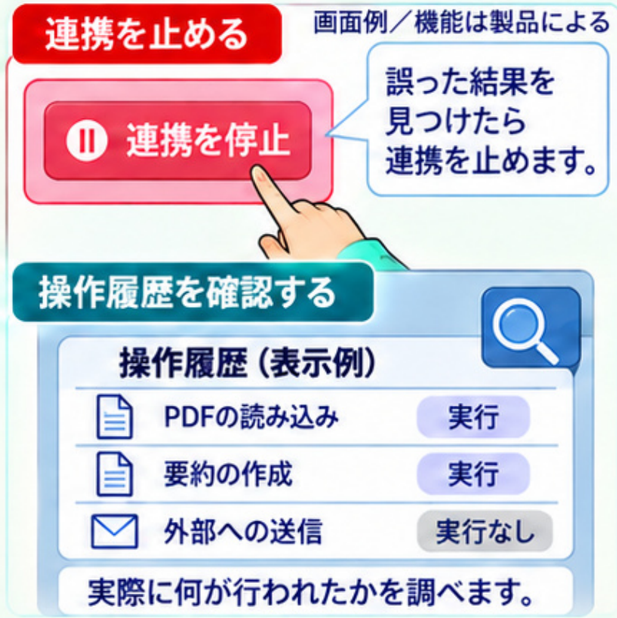
その文章だけを唯一の壁にはできません。
文書中の指示に引っ張られたり、別の理由で誤動作したりする可能性を考えるからです。

2 持ち物の要約なら、対象のPDFだけを渡し、関係のない家族の連絡先を読めない状態にします。
外部への送信機能を使わせず、結果はAの画面へ返すだけにします。連携が必要な作業では、誰へ・何を・どの範囲で実行するのかを本人が具体的に確認し、必要がなくなった権限を戻します。





3 **止める機能と操作履歴も確認します。**
誤った結果を見つけたら連携を止め、実際に何が行われたかを調べます。
これならAIを使うこと自体を諦めず、失敗した時の影響を小さくして使い続けられます。



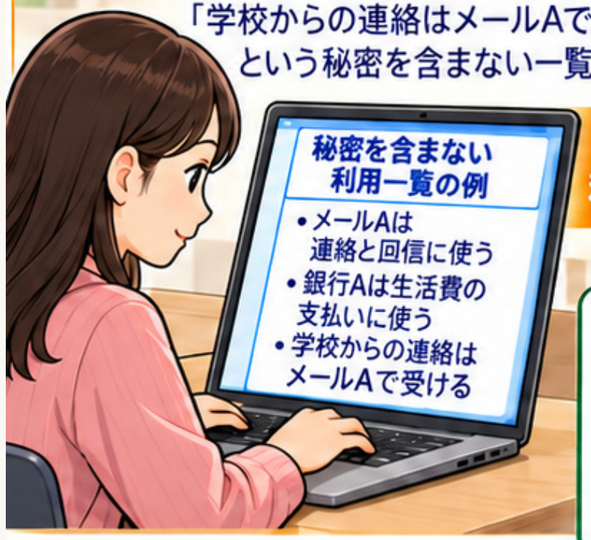


原典 S33,S35,S27：ncsc.gov.uk / learn.microsoft.com ほか　調査基準 2026.10.23

AIに、自分専用の「防衛台帳」を作らせる

1 【実習の提案】AはAIに、説明だけでなく、あとで更新できる表のファイルを作らせます。

材料は「メールAは連絡と回復に使う」「銀行Aは生活費の支払いに使う」「学校からの連絡はメールAで受ける」という秘密を含まない一覧です。



秘密を含まない利用一覧と作成依頼

2 依頼例：「サービス名、止まると困ること、公式窓口、確認済みの設定、未確認、次に確認する人と日を列にした表を作って。窓口や対応機能を想像で埋めないで」。

AIは表を作り、不明な欄を『未確認』で残します。ファイル作成に対応するAIで行う実習です。



サービス名、止まると困ること、公式窓口、確認済みの設定、未確認、次に確認する人と日を列にした表を作って。窓口や対応機能を想像で埋めないで。

空欄を推測で埋めず、表ファイルを作る

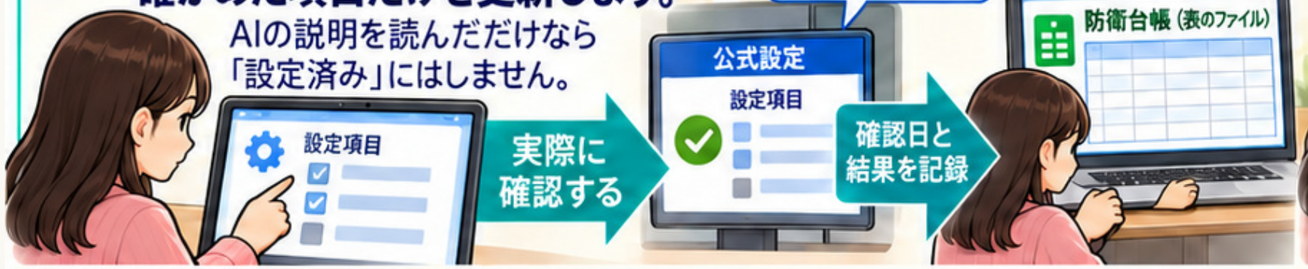


防衛台帳（表のファイル）の仮の記入例

サービス	止まると困ること	公式窓口	確認済みの設定	未確認	次に確認する人と日
メールA	学校の連絡を読めない	未確認	自分でログインを確認	回復先・通知	Aが確認日を決める
銀行A	生活費を支払えない	未確認	未確認	通知・上限	Aが確認日を決める

3 Aは公式設定を開いて確かめた項目だけを更新します。

AIの説明を読んだだけなら「設定済み」にはしません。



4 次回は表を渡し、未完了を取り出してもらう。



5 何度も最初から相談する負担を減らし、防御を続ける記録を自分の手元に残します。

原典 S03：cyber.gov.au　調査基準 2026.10.23

日光AI食堂

公式の注意喚起を、自分の次の仕事へ変える

1 公式の注意喚起をAIに調べさせる



公開日・対象・対応を読む

【一歩進める提案】 Aは、使っているサービスの公式な注意喚起をAIに調べさせ、防衛台帳と照合させます。

依頼例：「公開日、対象、必要な対応、原典のリンクを整理して。私に該当するか不明なら不明と書き、確認先への質問を作って」。

2 AIが防衛台帳と照合する

AIが公式の告知内容を読み、自分の防衛台帳と照らして内容を整理します。



照合した結果を確認票に整理

AIは「対象サービスに新しい告知がある」「この端末が対象かは未確認」「公式設定か提供元で確かめる」を分けます。

3 AIが確認票を作る

AIが調べた内容を、次の行動につなげるための確認票にまとめます。

確認票の例

- 対象サービスに新しい告知がある
- この端末が対象かは未確認
- 公式設定か提供元で確かめる

原典を開いて対象を確かめる

4 Aが原典を開き、対応を進める

Aは原典を開き、該当する対応を進めます。



必要な対応を実施・確認

5 結果を防衛台帳へ戻す

対応の結果を防衛台帳に記録し、今後の管理につなげます。



ニュースを読んだだけで終わらず、自分の未完了の仕事へ変える使い方です。

6 予約実行で定期的に確認する（任意の設定）

予約実行に対応する環境なら、確認する公式情報、間隔、通知する条件、停止方法を設定し、実際に一度動いたか確認します。

設定する内容の例

- 確認する公式情報
- 間隔
- 通知する条件
- 停止方法

実際に一度動いたか確認する

予約した → 実行記録を見た → 通知が届いた

会話だけで監視開始にはなりません。

通知がなくても安全の証明にはならず、銀行などの正規通知と併用します。

原典 S03,S08：cyber.gov.au / php.net　調査基準 2026.10.23

日光AI食堂

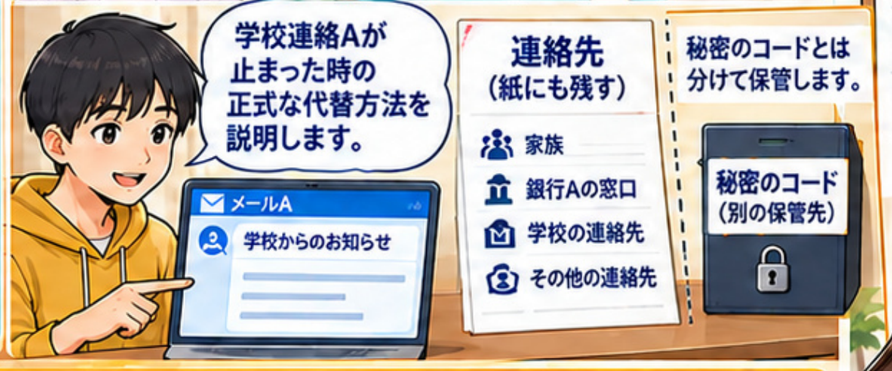
明日、A家が確かめるのは「分かった」ではなく「一度できた」

1 A家はまず、メールAに自分で入り、使える認証と回復先を公式設定で確認します。次に、銀行Aで何が通知されるか、上限がどの取引へ適用されるか、止める窓口はどこかを確かめます。分からない項目はAIに質問文へ直してもらい、公式窓口で解決します。



2 祖母Bは「いったん切ってかけ直す」を声に出し、Aは既知の番号で受けます。子Cは学校連絡Aが止まった時の正式な代替方法を説明します。連絡先は紙にも残し、秘密のコードとは分けて保管します。





3 全部が一日で終わらなければ、未完了の項目と次に確認する人・日を決めます。家族が困ったと言った時の第一声は「話してくれてよかった。一緒に止めよう」。



4 理解・練習・設定・相談をつないだ家庭の備えを、次は地域と事業者の仕組みへ広げます。



原典 S21,S25,S05,S29：support.google.com / smbc.co.jp ほか　調査基準 2026.10.23
日光AI食堂

自分の注意では、相手の設備を直せない

【事実から考える】 自分が偽メールを見抜いても、情報を預けた会社のソフトや、会社同士をつなぐ「合鍵」までは管理できません。Microsoftの10月21日報告でも、許可済みの外部アプリを足場にした攻撃が確認されました。

本人にできるのは、 異変に気づき、連絡し、使い方を選ぶこと。会社が握っているのは、権限を絞る設定、停止する仕組み、原因を調べる記録、直す人員です。

① 利用者（本人）

通販サイトで商品を見て、注文する。サービスに情報を預けている。



情報を預ける

異変を連絡する

本人から運営者の操作盤へは届かない

調査結果・修正内容・救済窓口を返す

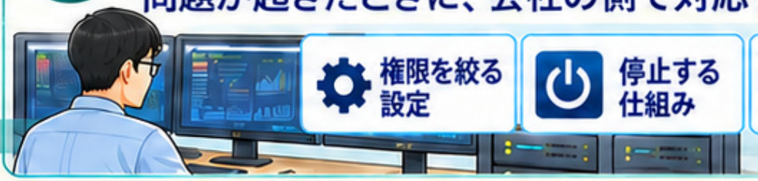
② 通販サービス（会社のシステム）

お客さんから預かった情報を処理し、注文を管理する。外部のアプリともつながり、サービスを提供している。



③ 運営者（会社の担当者）

問題が起きたときに、会社の側で対応する。



- 権限を絞る設定
- 停止する仕組み
- 原因を調べる記録
- 直す人員

【私たちの要求】

危険を制御できる側に、その仕事を引き受けさせる。





利用者へ「注意してください」と通知するだけで終わらせず、何を直し、いつ確認し、被害をどう受け止めるかを示してもらいます。



個人の努力を続けながら、個人では届かない場所の責任を動かす。

ここから第3章が始まります。

原典 S01,S37：microsoft.com / cao.go.jp　調査基準 2026.10.23

日光AI食堂

送ってしまった後は、誰が何を止めるのか

お金の流れ（振込）
連絡の流れ（相談・依頼）

1 本人：追加送金を止める



すぐに
追加送金を止める

日時・金額・振込先
取引停止を相談

被害の経緯・証拠
届出

サロンやAI

正規窓口への
連絡補助



正規窓口への
連絡補助

2 自分の銀行にも相談する

自分の銀行にも相談し、
日時・金額・振込先など、
分かる記録を伝えます。



自分の銀行

金融機関へ
連絡

3 振込先の金融機関

本人は口座を凍結できません。
止められる金融機関へ、
判断に必要な情報を渡します。



振込先の金融機関
（相手の口座のある銀行）

取引停止を
判断・実施

4 相手の口座

相手の口座にお金が入ると、
すぐに他の口座へ移されたり、
現金化されたりしてしまう
おそれがあります。



5 警察へ連絡する

振込先の金融機関と警察へ
連絡します。



警察

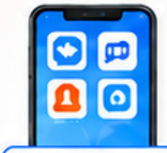
捜査上の照会・凍結依頼

6 対応のポイント

相手の口座からお金が動く前に対応する必要があります。
AIで長い分析を始めたり、次のサロンを待ったりしません。
カードの不正利用ならカード会社、アカウント侵害なら
そのサービスの正規窓口へ、止める対象に合わせてつなぎます。



カードの不正利用なら
カード会社へ



アカウント侵害なら
そのサービスの
正規窓口へ

【いま使える対応／振込被害の例】

「送金した。詐欺かもしれない」と気づいたら、追加送金を止め、振込先の金融機関と警察へ連絡します。自分の銀行にも相談し、日時・金額・振込先など、分かる記録を伝えます。本人は口座を凍結できません。止められる金融機関へ、判断に必要な情報を渡します。

【地域の役割】

責めずに連絡を手伝い、受付先と次にすることを本人と確かめる。
サロンやAIが、凍結・捜査・返金を代行するわけではありません。



原典 S38,S39：furikomesagi.dic.go.jp / fsa.go.jp　調査基準 2026.10.23
日光AI食堂

20行への連携拡大は、「追う時間」を縮める仕組み

① 被害者が警察へ相談

【運用拡大】警察庁は9月30日、特殊詐欺の被害金を追跡・凍結・、回復につなぐ連携を、10月1日から9行から20行へ広げると公表しました。



② 都道府県警察が照会

仕組みは、都道府県警察が警察庁を介して銀行へオンラインで照会し、銀行が回答すること。

③ 警察庁が銀行へ照会

お金の移転先が分かれば、その口座を管理する銀行へ早く凍結を依頼します。銀行をまたいで動くお金に、捜査と停止の連絡を追いつかせる仕組みです。

④ 協定銀行が回答・凍結





検証すべき課題

- 回答まで ？分
- 凍結まで ？分
- 回復額 ？

【確認できた範囲】

20行への拡大は確認済み。
しかし、この資料だけでは
「何%取り戻せたか」は分かりません。
被害者への自動返金でもありません。

【求める結果】

連携数に加え、照会から回答・凍結までの時間、回復に届いた範囲、つながらなかった理由を検証する。
会議や協定の数では、暮らしは戻りません。

原典 S40：npa.go.jp　調査基準 2026.10.23
日光AI食堂

「救済法がある」と「全額戻る」は違う

1 いま使える制度

振り込め詐欺救済法では、犯罪に使われた預金口座などに残るお金を、手続後に被害者へ分配します。国や銀行が、被害額の全てを埋める制度ではありません。



3 申請の支援と生活の相談

申請の案内と期限を確かめ、本人確認と振込の記録などを用意します。金融庁は、支払いまで一般に少なくとも半年以上かかると説明しています。だから、早い停止、申請の支援、回復までの生活の相談を分けずにつなぐことが必要です。

2 計算の仮例／実際の決定は金融機関

計算の仮例　Aさんの被害100万円、Bさん200万円、分けられる残高90万円なら、被害額の割合1対2で、Aさん30万円、Bさん60万円。残高が足りなければ、生活費の穴は残ります。すでに引き出されたお金や、残高が1,000円未満の場合など、支払いがない場合もあります。



4 手続の流れと期間の目安



原典 S39,S38：fsa.go.jp / furikomesagi.dic.go.jp　調査基準 2026.10.23
日光AI食堂

被害の話を、設計を直す仕事へ返す

1 【私たちの要求】被害者が「止めたいのに窓口が見つからなかった」と話したとします。これを本人の探し方だけの問題にしてはいけません。



本人
（被害者）

本人の同意＋
困った画面と時刻

いつ・どの画面で・何ができなかったかを整理します。



相談担当

2 本人の了承を得て、いつ・どの画面で・何ができなかったかを整理します。運営会社には、すぐ止められる導線、人につながる窓口、申請期限と補償条件の分かる説明を求める。原因調査に必要な記録は会社が保全し、修正の担当者と確認日を示してもらいます。

停止導線・人の窓口・
説明の改善要求

- すぐ止められる導線
- 人につながる窓口
- 申請期限と補償条件の分かる説明
- 原因調査に必要な記録は会社が保全
- 修正の担当者と確認日を示してもらいます。

3 利用組織が選んだサービスなら、その組織にも契約条件と事業者への要求を見直してもらう。被害者を複数の会社の間で往復させず、最初に受けた側が対応の窓口を明示する仕組みを要求します。

運営会社



会社

契約条件へ追加

利用組織



変更する画面・
担当・期限



開発担当

修正版で操作を再確認



本人
（被害者）

4 賠償の法的判断と、今すぐ困っている人を受け止める運用は別です。結論が出るまで、説明も対応も止めてよい理由には しません。



賠償の法的判断

別です。



今すぐ困っている人を受け止める運用

結論が出るまで、説明も対応も止めてよい理由にはしません。

原典 S41：digital.go.jp　調査基準 2026.10.23

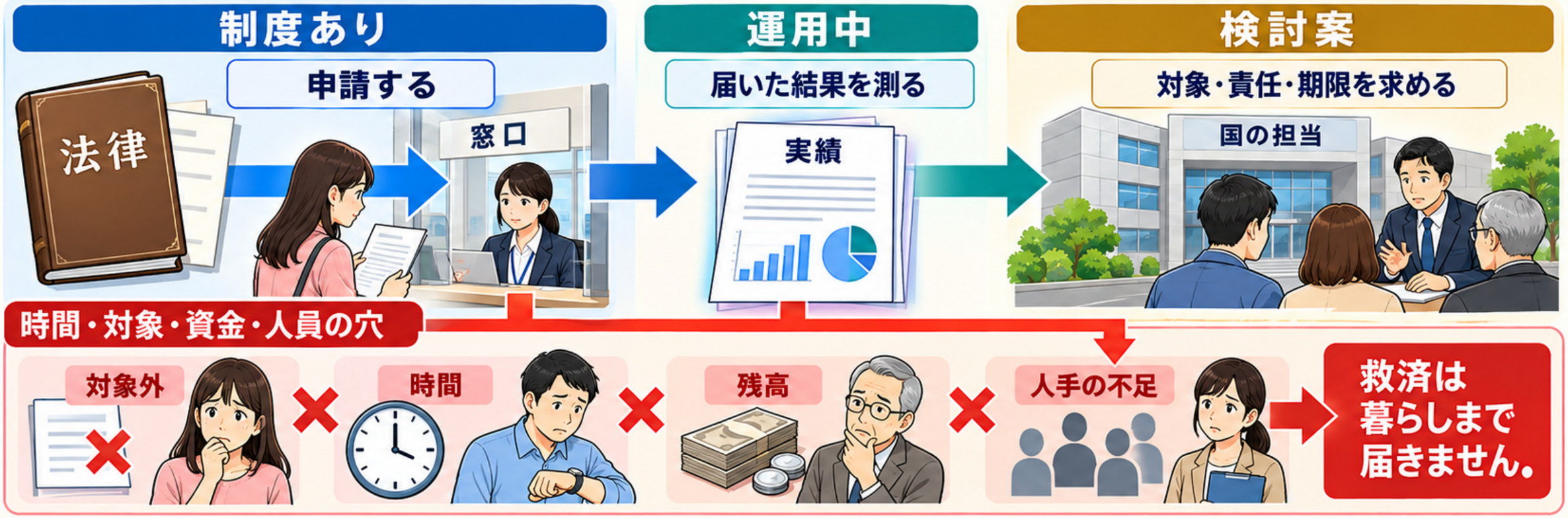
日光AI食堂

法律・運用・検討案を分けると、要求先が見える

1 【現在】振込被害には既存の回復手続があり、国の生成AI調達ルール第2版も9月1日から動いています。一方、制度の入口があっても、対象外、時間、残高、人手の不足があれば、救済は暮らしまで届きません。

2 【10月22日の検討案】政府のAI・デジタル改革推進会議は、AIを使った行為の責任分担や被害救済、AI法を含む制度の見直しを論点にしました。新しい包括的な補償制度が成立した、という発表ではありません。

3 【私たちの要求】今ある手続は使う。そのうえで、使えない人や救われない被害を記録し、誰が対応し、どの仕組みを、いつ変えるのかを求める。「法があるから安心」「法がないから何もできない」のどちらでも、現実の穴は埋まりません。法の名前から、実際に止め、直し、回復する力へ問いを進めます。



原典 S39,S41,S42,S43,S37：fsa.go.jp / digital.go.jp ほか　調査基準 2026.10.23

日光AI食堂

買う前の契約で、止め方と責任を決められる

① 仕様書で要望を渡す



必要な機能＋止め方＋
データ範囲を仕様書で渡す。

② 契約で条件を返す



実装内容・対応責任を
契約で返す。

③ 導入前の試験を行う



架空の入力で停止・
人への引継ぎを確認。

④ 運用して検証する



⑤ 支払いと検収



求めた働きが
届いたかをチェック。

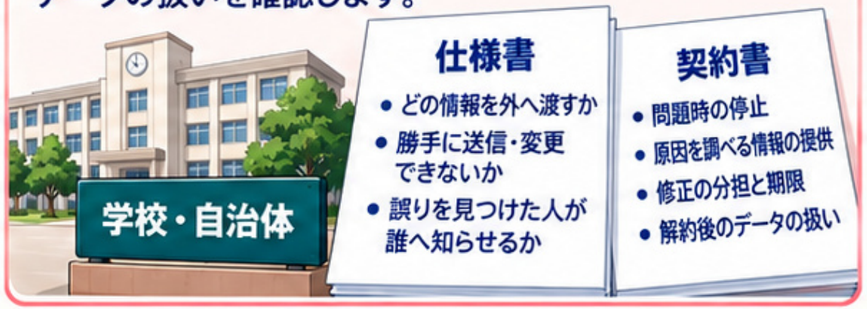
【いまの参考例】

政府の生成AI調達ルールは、買う時の確認、契約の取り決め、使い始めた後の検証、問題の報告までをつなげています。性能の良さだけで購入を決める仕組みではありません。



【学校・自治体への提案】

例えば相談AIを導入するなら、「どの情報を外へ渡すか」「勝手に送信・変更できないか」「誤りを見つけた人が誰へ知らせるか」を仕様書に書く。契約では、問題時の停止、原因を調べる情報の提供、修正の分担と期限、解約後のデータの扱いを確認します。



納品されたら、架空の相談で試し、危険な操作を止められるか、人へ引き継げるかを確かめる。更新後も試し直す。私たちの税金で買う道具なら、支払う条件の中に、利用者を守る仕事を入れることができます。



原典 S41,S03：digital.go.jp / cyber.gov.au　調査基準 2026.10.23

子どもの記録をAIが誤ったら、誰が直すのか

① 【考えるための仮例】

学校で、相談メモの要約AIが「参加を迷っている」を「参加しない」に変えたとします。その要約だけで支援や連絡を決めれば、子どもの機会が失われかねません。

② 人の確認

先生が元の記録と本人の話を確かめ、決定前に人が確認する。本人や保護者は「どこが違うか」を担当へ伝え、担当は訂正した記録と、影響を受けた連絡・判断まで確認する。AIの文章を一か所直して終わらせません。

③ 支援や連絡の判断

訂正した記録をもとに、支援や連絡を決める。誤った要約のまま判断せず、本人の機会が失われないようにする。



④ 制度と運用の整備

学校・教育委員会には、AIの用途、外へ渡る情報、責任者、訂正の窓口、利用が難しい子の代替手段を説明してもらいます。子どもが意見を言える時間と方法も用意する。詳しい事情を一般向けAIへ公開させず、必要な人だけで確認する。便利さのために、本人の声を消さない設計を要求します。



原典 S41,S37：digital.go.jp / cao.go.jp　調査基準 2026.10.23

サロンは、相談を受け止め、専門先へつなぐ場所にする

① 入口のサロンで「困り事を聞く」



本人の同意＋必要な経緯

② 緊急性の分岐

お金が動く	本人の同意＋必要な経緯	銀行・警察へ
契約	本人の同意＋必要な経緯	消費生活相談へ
学校	本人の同意＋必要な経緯	学校へ
操作	本人の同意＋必要な経緯	技術支援へ

③ 専門先へつなぐ









④ 専門先から本人へ

受付・次の手順

受付・次の手順

受付・次の手順

受付・次の手順

⑤ 本人からサロンへ（任意）



つながれた／まだ困る

【地域でつくる運用案】

最初の言葉は「話してくれてよかった。
一緒に次の連絡先を確かめよう」。
送金が動いている相談なら、その場で正規の
銀行・警察へ。契約のトラブルなら消費生活相談へ。
学校の記録なら学校の責任者へ、困り事に合わせて
渡します。

渡すのは、本人が了承した必要な経緯です。

暗証番号・認証コードは預からず、本人の代わりに
振込や契約をしません。
紹介先、受付時間、次の確認日を本人と紙に残し、
「つながれたか」まで確かめます。

渡さないもの ✕ **暗証番号・認証コード**

【日光でいまある入口】

市の消費生活相談は平日9～16時、
電話**0288-22-4743**。
土日祝日は**188**の案内があります。
利用時は公式情報を再確認します。
新しい看板を作る前に、既存の相談・福祉・
教育の支援と接続することから始めます。

原典 S44,S29,S45,S46,S47：city.nikko.lg.jp / caa.go.jp ほか　調査基準 2026.10.23

日光AI食堂

一人の困り事から、みんなの仕組みの欠点を取り出す

1 【仮例】3人から「被害を止める窓口が見つからない」という相談が来たとします。
名前、口座、被害額の一覧を地域へ配る必要はありません。

2 本人の同意を得た範囲で、「どの場面で」「何ができず」「どんな不利益があり」「どう変えてほしいか」を課題カードへ整理します。
例えば「緊急停止の連絡先が、ログイン後にしか見えない。ログインできない人が止められない。ログイン前の画面と紙の案内にも出してほしい」です。

3 要望を送る担当と確認日を決め、会社へ具体的な画面の改善を求める。



相談 1
被害を止める窓口が見つからないという相談



相談 2
被害を止める窓口が見つからないという相談



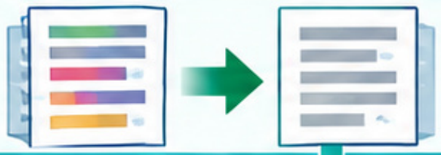
相談 3
被害を止める窓口が見つからないという相談

個別支援に必要な記録



個別相談の記録は必要な支援のために守ります。

同意した範囲で識別情報を除く





事実を足さない

AI

AIには、個人が分からない架空化した文章の整理を手伝わせ、人が元の話と照合します。
（文章整理だけ）

課題カードへ整理

どの場面で	ログイン後にしか見えない。
何ができず	ログインできない人が止められない。
どんな不利益があり	緊急停止ができない。
どう変えてほしいか	ログイン前の画面と紙の案内にも出してほしい。

会社へ



ログイン前に停止窓口を表示してほしい。

4 個別相談の記録は必要な支援のために守り、集団の要求へ使う情報は目的と範囲を別に決めます。

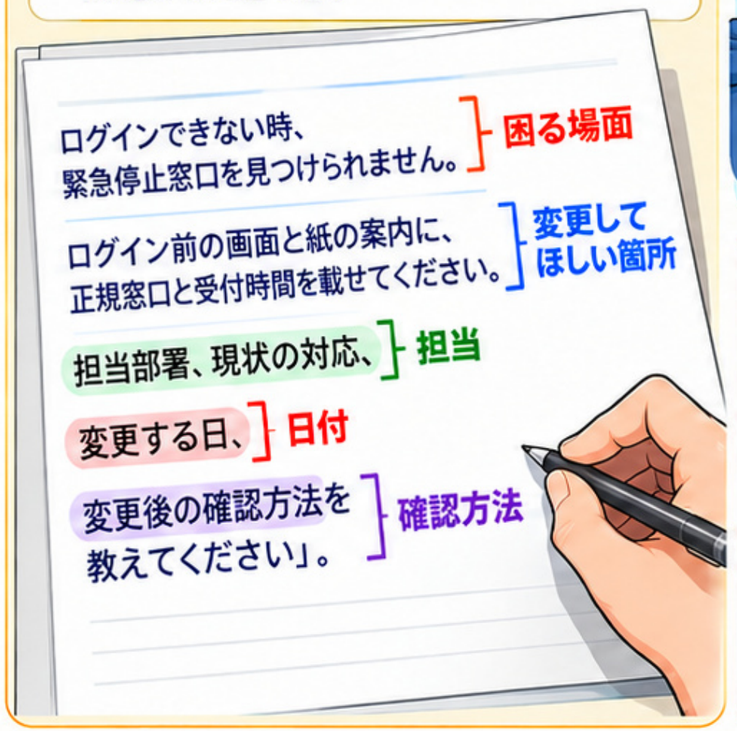
原典 S37,S46：cao.go.jp / caa.go.jp　調査基準 2026.10.23
日光AI食堂

「安全にして」を、担当・期限・確認方法まで書く

1 依頼文を作る

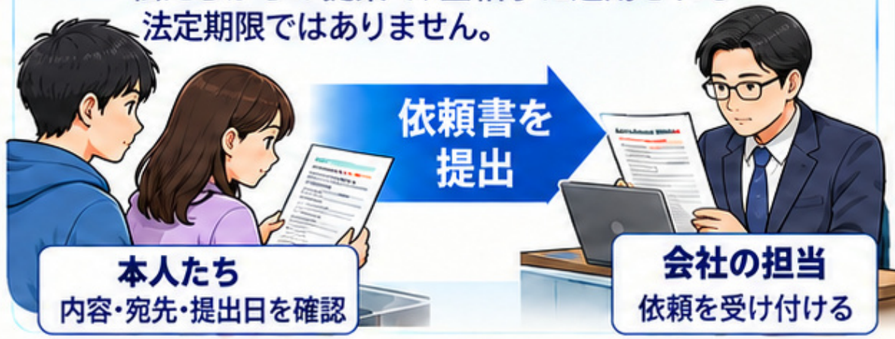
【依頼文の仮例】

「ログインできない時、緊急停止窓口を見つけられません。ログイン前の画面と紙の案内に、正規窓口と受付時間を載せてください。担当部署、現状の対応、変更する日、変更後の確認方法を教えてください」。



2 依頼文を届ける

届け先は、画面を運営する会社と、必要なら契約を結んだ学校・自治体です。本人たちが内容と宛先を確認し、連絡担当が提出日と回答希望日を記録します。「30日で現状と計画を回答してほしい」は私たちからの提案で、全相手に適用される法定期限ではありません。



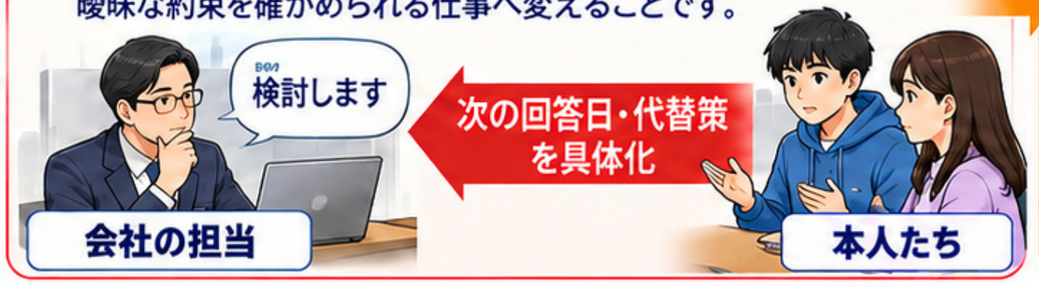
3 担当・計画・期限を回答

会社から本人たちへ、担当・計画・期限を回答します。必要なら、契約を結んだ学校・自治体にも伝えます。



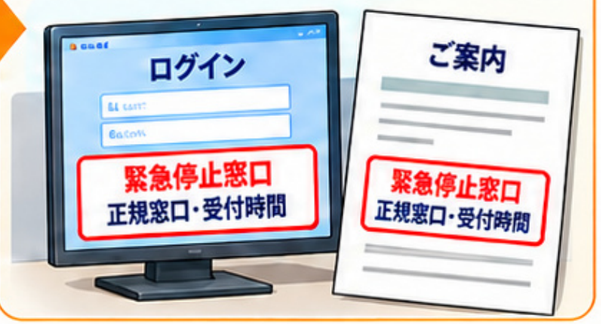
4 返答があいまいなら、さらに確認する

返答が「検討します」だけなら、検討する責任者と次に回答する日を尋ねます。変更できないなら理由と代替策を求める。強い主張とは、相手を怒鳴ることではなく、曖昧な約束を確かめられる仕事へ変えることです。



5 画面の変更につながる

会社は、ログイン前の画面と紙の案内を変更し、正規窓口と受付時間を載せます。



原典 S41,S48：digital.go.jp / city.nikko.lg.jp　調査基準 2026.10.23
日光AI食堂

続ける相談には、人の時間と予算が要る

1 準備・紹介先との確認



相談会の前に、
準備・紹介先との確認を行う。

2 相談会60分



住民からの相談を受け、
専門先への接続を行う。

3 連絡・記録・未解決確認



相談会の後に、
連絡・記録・未解決確認を行う。

【試算の例／相場ではありません】

月1回の相談会を、当日だけでなく準備・連携・記録確認まで含めて考えます。



調整担当4時間×2,500円
＝1万円



補助2人×3時間×2,000円
＝1万2,000円



専門家2時間×8,000円
＝1万6,000円



会場と印刷4,000円。

合計は月4万2,000円、年12回なら50万4,000円です。

予算と人の動き



数字は議論用の仮定です。

実際には会場、保険、交通、通信、研修、税などを確認し、既存事業と組めるかも調べます。



【要求すること】

自治体や協力団体へ、どの担当が、何時間、どの予算で支えられるかを尋ねる。
善意だけに任せて担当者が疲れ、相談先が消えれば、人は再び孤立します。予算を付けるなら、開催回数だけでなく、専門先への接続と未解決の追跡まで仕事に含めます。

原典 S47,S46：digital.go.jp / caa.go.jp　調査基準 2026.10.23

日光AI食堂

政治には、「誰が、何を、いつ変えるか」を問う

1 住民の課題を行政の担当へ渡す

住民の困りごとや要望を整理し、行政の担当へ渡して、必要な仕事・費用を検討してもらいます。

2 市長が事業案・予算案を提出

日光市の説明では、市長が予算などを提案し、市議会が審議して決め、市長側が市政を実行します。

3 市議会が審議・議決

日光市の説明では、市議会が審議して決め、市長側が市政を実行します。
議会には、決めた仕事が実施されたかを調べる役割もあります。

4 決めた仕事を実施したか確認

議会には、決めた仕事が実施されたかを調べる役割もあります。

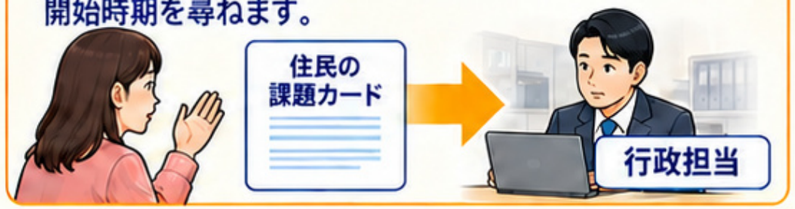
5 行政が人員配置し、事業を実施

市長側が市政を実行します。



具体的な場面

「ログインできず被害停止の窓口に届かない人がいる。地域で、対面の確認と専門先への紹介を月1回できないか」。この課題を行政の担当へ渡し、事業の担当、人員、費用、開始時期を尋ねます。



同じ問いを全ての立場へ

1. どの担当が受けるか。
2. 何時間の人員と、どの予算を付けるか。
3. 使いにくい人に電話・対面を残すか。
4. いつ始めるか。
5. 何を成果として公表し、不足を直すか。

議員に相談しただけで実施は決まりません。

行政の計画、予算の議決、実施の確認まで追います。
支持政党に関係なく、責任を具体的な仕事にします。



原典 S48：city.nikko.lg.jp　調査基準 2026.10.23
日光AI食堂

制度案へ意見を出す時は、件数より仕組みを説明する



1 【いま使える参加の道】

国の命令などの案が公表されると、対象案件によってパブリック・コメントで意見を出せます。e-Govは、同じ意見の投稿数による多数決ではなく、内容を考慮する手続だと説明しています。



2 【書き方の例】

「相談をAIだけで受け付ける案では、ログインできない人の申出が止まります。電話・対面の入口、担当者への引継ぎ、訂正が反映されたか本人へ返す手順を入れてください」。

困る場面→不利益→変更案→理由の順で書きます。



3

担当する制度と募集期限を確かめ、本人が内容を確認して提出する。
募集がない時は、所管の行政や議会への具体的な要望として届けます。
提出後は行政の考え方と最終案を読み、取り入れられた箇所・残った課題を確かめる。
送信したという達成感で終わらせません。

原典 S49：public-comment.e-gov.go.jp　調査基準 2026.10.23

30日で、依頼から実装確認まで一周する

① 1週目

主催者が、相談の担当と紹介先、受付時間を確認し、紙1枚にまとめます。



確認した連絡先を渡す

② 2週目

架空の例で練習し、本人が次にどこへ連絡するか言えるか確かめます。



分かった困り事を本人同意で整理

③ 3週目

希望者が、同意した範囲の困り事から改善依頼を1通作り、担当と回答希望日を決めて届けます。



依頼書・担当・回答日を記録

④ 4週目

返事を読んで終わらず、変更された画面や窓口を試します。「受け付けた」「計画が出た」「実装した」「使えると確認した」を別々に記録し、未解決なら次の確認日を置きます。



【日光AI食堂からの活動案】

1週目は主催者が、相談の担当と紹介先、受付時間を確認し、紙1枚にまとめます。2週目は架空の例で練習し、本人が次にどこへ連絡するか言えるか確かめます。3週目は希望者が、同意した範囲の困り事から改善依頼を1通作り、担当と回答希望日を決めて届けます。

4週目は、返事を読んで終わらず、変更された画面や窓口を試します。「受け付けた」「計画が出た」「実装した」「使えると確認した」を別々に記録し、未解決なら次の確認日を置きます。

相談が増えても、悪化とは限りません。
話せる場ができた可能性があります。
開催数だけでなく、専門先へつながれたか、何が直り、誰がまだ困るかを見る。
30日で被害ゼロを証明するのでなく、確かめられる改善を一つ残します。

原典 S47,S46,S49：digital.go.jp / caa.go.jp ほか　調査基準 2026.10.23

日光AI食堂

第３章をつくるとは、決める力と直す力を取り戻すこと

【私たちが目指す未来】

AIができることを増やすだけでは、
私たちの自由は増えません。
何の情報を渡し、何を任せ、どこで止め、
困った時に誰が直すかを、利用者が
理解し選べることが必要です。

家庭では、パーソナルAIを説明・練習・記録整理に使い、権限と重要な判断を確かめる。地域では、困り事を受け止め、専門先につなぎ、同意した範囲で構造の欠点を言葉にする。会社には設計と対応を、行政と議会には実施する人員・予算・制度を求める。返ってきた約束を、実際に使えるかまで確かめます。

攻撃者と提供側だけに未来の条件を決めさせない。便利さを諦めず、危険と負担を個人へ押しつける仕組みを変える。今日選ぶ一つの防衛を、地域の一つの改善へつなぎ、その改善を次の当たり前にする。それを第3章と呼びます。

1 本人



家庭では、パーソナルAIを
説明・練習・記録整理に使い、
権限と重要な判断を確かめる。

困り事と希望を伝える

2 地域



地域では、困り事を受け止め、専門先につなぎ、
同意した範囲で構造の欠点を言葉にする。

同意した範囲の
改善要求

3 会社・行政



会社には設計と対応を、行政と議会には
実施する人員・予算・制度を求める。

何を
任せるか

止められるか

直して
もらえるか

4 現場



返ってきた約束を、実際に使えるかまで確かめます。

設計変更・人員・予算・制度

使えることを確認／未解決を返す

原典 S03,S37：cyber.gov.au / cao.go.jp　調査基準 2026.10.23

日光AI食堂