Villain's Hat

MISUSE INVESTIGATION · V1 · 19 OCT 2026

Banking App Integration Plan

Topic: Banking App Integration Plan

POTENTIAL NEWS HEADLINE IN THE MAKING:

"Banking app AI agent used to coerce payments from vulnerable users"

2	2	0	3
Misuses found	High severity	Wide reach	Similar real cases

We found 2 misuses, 2 of them with High severity. The most serious are Desperate Advice Seeker and Coercive Payment Controller.

Most of the risk comes from people using the Banking App Integration Plan to force or trick others into financial actions. A large part of the findings rests on guessed parts of the journey, see assumptions.

THE RISKS AT A GLANCE

1 **Desperate Advice Seeker**

High severity

A financially distressed user treats an assistant's answer based on real bank data as qualified advice and relies on it for a high-stakes financial decision.

2 **Coercive Payment Controller**

High severity

An abuser directs an agent to request payments and pressures the account owner into giving biometric approval.

01 WHAT ALREADY WORKS

Read-only bank queries

AI Assistant Bank Query returns read-only bank data. Someone using that connector access cannot turn a balance check into a transfer.

Owner-approved payments

AI Agent Payment Approval requires the user's on-device biometric approval and applies user-set spending limits. An agent's request alone cannot authorize a payment.

02 THE MISUSES

How to read the bars. Vector: how many ways in. Reach: how many people it could affect. Severity: how serious the harm is. Detect: how hard it is to catch before harm. The more filled the bar, the more dangerous. A villain is the person misusing your product, including anything they ask their agents to do for them.

1 Desperate Advice Seeker

High severity · Band B

A financially distressed user treats an assistant's answer based on real bank data as qualified advice and relies on it for a high-stakes financial decision.

VECTOR	REACH	SEVERITY	DETECT
Narrow	Moderate	High	Hard to spot

WHAT MAKES IT POSSIBLE

The connector product feeds real bank transaction and balance data into an AI assistant like Claude or ChatGPT. A user in financial distress, facing eviction, bankruptcy, or debt crisis, asks the assistant a question like 'can I afford this' or 'what should I do with my cash flow'. The assistant has real data and speaks with the confident, helpful tone of a financial advisor.

WHAT THE PERSON ACTUALLY DOES

The user reveals their financial crisis to the AI. The AI pulls their real bank data via the connector, analyzes it, and delivers a confident answer that sounds like professional financial advice. The user, exhausted and desperate, treats it as expert guidance and makes a high-stakes decision based on it. The decision goes wrong. The user later discovers the AI is not a financial advisor, has no liability, and the founder has no way to know what advice was given.

WHERE IN YOUR JOURNEY

Consulting Audit Delivery | Connector Account Setup | **AI Assistant Bank Query** | Referral Commission Earning | AI Agent Payment Approval

WHERE IT HITS THE END USER

A user in financial crisis makes a decision that worsens their situation, based on confident guidance from a system that has no qualification, no accountability, and no duty of care. They may delay seeking real professional help, miss deadlines, or make irreversible financial moves. The harm is financial and emotional, and it lands on someone already in acute distress.

WHERE IT AFFECTS THE BUSINESS

Regulatory exposure if a financial regulator in Serbia or the Balkans determines the connector product is offering financial advice without a licence. Reputational damage if a user's story surfaces. Potential liability if the user can show the product held themselves out as capable of giving financial guidance.

CLOSEST MATCH

> Cleo cash advance enforcement order (search.ftc.gov)
> It involves a financial app's cash advance claims and regulator action, but the result does not establish that AI advice caused a user's loss or that the case concerned bank-connected personalized advice.
> **Outcome: Judgment in the amount of Seventeen Million Dollars ($17,000,000) is entered in favor of the Commission against Defendant as monetary relief.**

RECOMMENDED FIX

Add a clear, persistent disclaimer in the AI assistant interface stating that the connector product and AI assistant are not financial advisors, cannot provide financial advice, and the user should consult a licensed professional before making financial decisions based on the data shown.

HOW THE VILLAIN COULD GET AROUND THE FIX

A desperate user reads the disclaimer, clicks past it anyway, and still treats the AI's confident data-backed answer as professional financial advice because nothing in the experience actually stops them.

2 Coercive Payment Controller

High severity | Band B

An abuser directs an agent to request payments and pressures the account owner into giving biometric approval.

VECTOR	REACH	SEVERITY	DETECT
Narrow	Low	High	Hard to spot

WHAT MAKES IT POSSIBLE

Stage 5 allows an AI agent to initiate payments that the user approves with a single biometric tap. A user in an intimate partner abuse or financial coercion situation has given an AI agent (or someone controlling it) access to their account. The abuser uses the AI to initiate small or large payments, and the user, under pressure or fear, approves them with the biometric tap.

WHAT THE PERSON ACTUALLY DOES

An abuser or controller sets up an AI agent with access to the user's bank account and spending limits. The abuser instructs the AI to initiate a payment. The user receives the approval notification, feels pressure or fear, and taps biometric approval. The payment executes. The user has no way to refuse without escalating conflict or revealing the abuse.

WHERE IN YOUR JOURNEY

Consulting Audit Delivery | Connector Account Setup | AI Assistant Bank Query | Referral Commission Earning | **AI Agent Payment Approval**

WHERE IT HITS THE END USER

A user in an abusive situation loses money and agency through a system designed to be frictionless. The biometric approval creates a false sense of consent and control, when in reality the user is under duress. The harm is financial, emotional, and safety-related.

WHERE IT AFFECTS THE BUSINESS

If a user in an abuse situation uses this product and the abuse is later discovered, the product becomes evidence of a tool that enabled financial control. Regulatory and reputational risk if financial crime or abuse investigators examine the product design.

CLOSEST MATCH

> Hong Kong deepfake payment scam (amp.theguardian.com)
> It involved AI-assisted impersonation to induce a large payment, but the report does not describe an AI agent or biometric approval.
> **Outcome: classified as “obtaining property by deception”**

> AI voice impersonation bank fraud (January 2020 · incidentdatabase.ai)
> It involved AI-enabled impersonation to obtain financial transfers, but does not describe an AI agent coercing biometric approval.
> **Outcome: funds were transferred from the Victim Company to several bank accounts in other countries**

RECOMMENDED FIX

Require a multi-step approval for payments above a threshold, including a time delay and a secondary confirmation method that is not biometric, so a user under pressure has a moment to refuse or seek help.

HOW THE VILLAIN COULD GET AROUND THE FIX

The abuser stands over the victim during the time delay and secondary confirmation, so the extra steps become extra pressure rather than a way out.

03 ALSO WORTH NOTING

Bank Data Oversharer
The founder sends users' bank data to AI providers without clearly explaining where it goes or what those providers may retain or reuse.

Opaque Billing Collector
The founder collects and stores subscription payment details without explaining retention or who can access them.

Secret Commission Seller
The founder recommends banks and fintechs through paid referral links without telling the audience about the commission.

04 WHAT WE HAD TO ASSUME

Tell us if any of these are wrong and the findings will change.

- We assumed the connector product requires users to create an account before linking their bank accounts, as no sign-up flow was described.
- We assumed the referral tracking for commissions is handled through affiliate links or promo codes tied to the founder's podcast and content, as no mechanism was specified.
- We assumed the biometric approval layer is delivered as an SDK or API that banks integrate into their existing apps, as no specific integration method was described.
- We assumed the subscription billing for the connector product is handled by a third-party payment processor, as no billing system was mentioned.

05 HOW THIS WAS RANKED

A misuse gets a full card when its score total is 9 or more, or when its severity is High, unless our own challenger judged it a stretch (then it goes under Also worth noting). The total adds the four scores (Vector, Reach, Severity and Detect, each 1 to 3).

Inside a band we order by the middle of the estimated cost range. A misuse whose cost we could not quantify goes after the others in its band, ordered by severity and then by how hard it is to catch.

1. Desperate Advice Seeker	total 9, band B
2. Coercive Payment Controller	total 8, band B

NOTE: WHAT THE BANDS MEAN

A band is a priority group, worked out from the total of the four scores. Band A is a total of 10 or more. Band B is a total of 9, or any misuse with High severity. Band C is anything lower, and those only appear under Also worth noting.

These are the most material misuses for what you have described so far, not a proof of safety.

When you're ready, run Villain's Hat again and we'll show what's fixed, what's still open and what's new.

Your dashboard: mcp.villainhat.com/dashboard