Deepfake *Fraud Defense Framework*

Protect your payments. Layer your controls.

AI-generated deepfakes are designed to deceive. When an employee has been misled, **a robust framework of payment controls are an important and more reliable form of protection.** This framework shows how CitiDirect is designed to help you intercept fraudulent payment instructions before funds leave your account.

The Defense Principle: Layered Controls Beat Individual Judgment

Modern AI deepfakes can replicate an executive's voice and appearance with a high degree of accuracy, making visual and auditory verification alone unreliable. **A more effective protection is a framework of interlocking payment controls** that are designed to remain effective even when an employee has been deceived. Two moments define the outcome of every attack: **Stage 4**, when the fraudulent payment instruction is given, and **Stage 5**, when speed of detection determines whether funds can be recovered.

Attack Lifecycle — 5 Stages

			CRITICAL STAGE	AFTERMATH
1.	2.	3.	4.	5.
Reconnaissance	**Initial Contact**	**Deepfake Call**	**Payment Instruction**	**Funds Dispersed**
Attacker profiles the target organisation and collects voice and video samples	Creates urgency and moves communication to unofficial channels	AI-generated impersonation of a senior executive via live video or voice	Urgent wire to a new beneficiary — primary point of control intervention	Recovery is extremely difficult once funds are transferred internationally
			THREE-PILLAR DEFENSE	VISIBILITY AND MONITORING

Three-Pillar Defense — Stage 4 Controls

CitiDirect provides a layered payment control framework designed to intercept fraudulent instructions at the point of payment. **Configuring all three pillars is designed to ensure that no single individual can authorise a fraudulent transfer.**

Segregation of Duties	Beneficiary Controls	Payment Authorisation
Maker-Checker Payment initiation and approval require separate, independent individuals.	**Payment Templates** Only pre-approved, standardised payment instructions are accepted.	**Multi-Level Approval** Up to nine configurable approval levels — by account, amount, and submission method.
Transaction Limits Configurable hard limits on transfer amounts per authorisation level.	**Citi Verify*** Designed to alert your team when a beneficiary name does not match the account number provided.	**Batch Authorisation** Independent review is required for each payment tranche within a batch.

If a payment does go out, speed of detection is everything

Even with strong controls in place, edge cases can occur. **Stage 5 is not a failure state — it is a recovery window.** Real-time monitoring and instant notifications are designed to give your team the earliest possible opportunity to initiate a stop payment or recall before funds are transferred internationally and recovery becomes extremely difficult.

Visibility And Monitoring — Stage 5 Response

The speed of your response can significantly influence the likelihood of recovery. These CitiDirect capabilities are designed to **compress your detection window and provide more immediate options to act.**

Citi Payment Insights

Real-time payment tracking with in flight stop payment and recall capability.

Event Notification

Immediate alerts on unexpected account debits and unusual account activity.

Push Notifications

Every outgoing payment is designed to trigger an alert to all authorised users.

To strengthen your payment controls, configure Segregation of Duties, Beneficiary Controls, and Authorisation Workflows in CitiDirect.

*Availability varies by market.

Additional Guidance and Best Practice Resources

Citi recommends that clients review the following materials to further strengthen their fraud prevention posture.

Beneficiary Change Request

Controls and verification steps for managing beneficiary change requests securely.

Dual Approval: Fraud Prevention

How dual-approval workflows reduce exposure to payment fraud and social engineering.

Guidance on How to Combat Fraud

Red flags, do's and don'ts, and key controls to help your organisation prevent and respond to fraud.

What to Do in the Event of Fraud

Immediate steps and contacts if a fraudulent payment has occurred.

AI Deepfakes: When Seeing and Hearing Cannot Be Trusted

Expert analysis of the deepfake threat landscape and actions organisations should take.

26-1D144E5D 07/26