CUSTIE PLATFORM

Detection Logic, Risk Controls, Review Workflows & Escalation Framework

Fraud Operations & Trust and Safety

Confidential | Internal Use Only

Document Owner	Fraud Operations & Trust and Safety
Classification	Confidential - Internal Use Only
Applies To	All active platform participants post-onboarding

1. Purpose & Scope

This document defines Custie's post-onboarding fraud detection logic, platform risk controls, review workflows, and escalation framework. While the KYC/KYB Underwriting Policy governs participant onboarding, this framework governs the ongoing monitoring and response to risk signals that emerge once participants are active on the platform.

All controls and workflows defined in this document are designed to be operationally scalable, structured for execution at the analyst level with decisioning authority held by the Fraud Operations Manager.

2. Detection Logic

Detection logic defines the specific signals, behaviors, and patterns that trigger a fraud or platform abuse review. Signals are grouped into two categories: identity and account integrity signals, and post-onboarding platform behavior signals.

2.1 Identity & Account Integrity Signals

These signals relate to the accuracy and consistency of participant identity information and are primarily relevant during onboarding but may surface post-activation through account changes or profile updates.

Signal / Behavior	Risk Level	Recommended Action
Name on ID does not match submitted profile name	High	Flag for manual review. Suspend account pending resolution.

Email address inconsistent with stated identity	High	Flag for manual review. Request identity reconfirmation.
Altered, blurry, or low-quality ID document uploaded	High	Reject application or suspend account. Request resubmission of valid document.
Profile information updated post-approval in a way that conflicts with verified ID	Medium	Flag for review. Cross-reference against original approved submission.
Duplicate account detected (same identity, multiple submissions)	High	Escalate immediately. Suspend all associated accounts pending investigation.

2.2 Post-Onboarding Platform Behavior Signals

These signals emerge from participant activity on the platform after onboarding is complete. They cover messaging behavior, service interactions, and community conduct.

Signal / Behavior	Risk Level	Recommended Action
Messaging patterns inconsistent with legitimate service inquiry (e.g., solicitation, off-platform payment requests)	High	Flag conversation. Issue conduct warning or suspend account pending review.
Consumer or provider reports abuse, harassment, or threatening communication	High	Escalate immediately. Suspend reported account pending investigation.
Provider misrepresenting services, qualifications, or availability to consumers	Medium	Flag profile. Review of original onboarding submissions. Issue warning.
Multiple complaints filed against a single participant within a short period	High	Escalate to Fraud Operations Manager. Review complaint history and determine account action.
Consumer or provider attempting to conduct transactions outside the platform	Medium	Issue conduct warning. Document incident. Escalate on repeat occurrence.
Unusual or coordinated activity between accounts suggesting collusion or manipulation	High	Escalate immediately. Conduct cross-account investigation.
Participant behavior that creates safety concerns for another user or their pet	High	Escalate immediately. Suspend account. Notify affected party.

3. Risk Controls

Risk controls are the operational mechanisms in place to prevent, detect, and respond to fraud and platform abuse. These controls operate continuously across the participant’s lifecycle.

#	Control	Type	Description
1	**KYC/KYB Identity Verification**	Preventive	Identity verification at onboarding prevents fraudulent or misrepresented participants from accessing the platform. See KYC/KYB Underwriting Policy.
2	**Required Field Validation**	Preventive	Mandatory field enforcement at the form level prevents incomplete or manipulated applications from entering the review queue.
3	**Data Minimization**	Preventive	Collection of only the minimum data necessary to make a sound decision reduces exposure and streamlines the review process.
4	**Application Integrity Review**	Detective	Manual cross-referencing of all submitted data detects identity mismatches, anomalous information, and misrepresentation prior to approval.
5	**Post-Onboarding Monitoring**	Detective	Ongoing review of participant behavior, complaints, and platform activity detects abuse, misconduct, and fraud signals after account activation.
6	**Escalation & Review Workflow**	Responsive	Structured escalation path ensures all detected risk signals are reviewed, documented, and actioned at the appropriate authority level.

4. Review Workflow

The review workflow defines the process for handling fraud or platform abuse signals that are detected post-onboarding. This workflow is designed to be executed at the analyst level, with all final decisions made by the Fraud Operations Manager.

All reports and flags that enter this workflow originate from one of two sources: internal detection (signals identified through monitoring) or external reports (complaints or reports submitted by active participants). Bug reports and technical application issues are routed separately through engineering and do not enter the fraud review workflow.

Stage 1: Signal Detection or Report Receipt

A risk signal or participant report is identified through one of the following paths:

- Internal detection: a behavioral pattern or account anomaly is flagged through ongoing monitoring
- Participant complaint: a consumer or provider submits a report through the platform regarding another participant’s conduct

All signals and reports are logged immediately upon receipt. Bug reports or technical issues are triaged separately and do not enter the fraud review workflow.

Stage 2: Initial Triage

The assigned analyst conducts an initial triage of the signal or report to determine:

- Whether the signal constitutes genuine fraud or platform abuse concerns
- The risk level of the signal (High, Medium, or Low) is based on the detection logic defined in Section 2
- Whether immediate account action is required prior to full investigation

If the signal is High risk, the analyst flags it for immediate escalation to the Fraud Operations Manager before proceeding. If Medium or Low, the analyst proceeds with full review.

Stage 3: Full Review & Investigation

All account reviews are conducted through Custie's proprietary internal admin portal. Analysts do not have direct access to AWS infrastructure or the underlying MySQL database. The portal surfaces all relevant participant data, submitted documents, and activity records in a structured review interface.

The analyst conducts a full review of the flagged account or interaction, including:

- Review of the participant's original onboarding application and verified identity information
- Review of the reported behavior, messaging pattern, or complaint detail
- Cross-referencing complaint history against the reported participant's account record
- Assessment of whether the reported behavior represents an isolated incident or a pattern

Findings are documented in the case record with sufficient detail to support the decisioning step.

Stage 4: Escalation to Fraud Operations Manager

Upon completion of the review, the analyst escalates findings to the Fraud Operations Manager with a summary of:

- The nature of the signal or complaint
- The risk level assessment
- Supporting evidence reviewed
- Recommended action

The Fraud Operations Manager reviews the escalation and makes a final decision on account action.

Stage 5: Decision & Account Action

The Fraud Operations Manager determines the appropriate account action based on findings. Available actions include:

Warning Issued	Participants are notified of the policy violation. Accounts remain active. Incident is documented.

Temporary Suspension	Account access is suspended pending resolution of the reported issue or receipt of additional information.
Permanent Removal	Accounts are permanently deactivated. Participants are notified. Record is retained for reference.
No Action - Unfounded	Review determines whether the report or signal does not constitute a genuine violation. Case is closed and documented.

Stage 6: Communication & Closure

The affected participant is notified of the outcome through the platform. The case record is updated with the final decision, action taken, and closure date. All case documentation is retained for operational reference.

5. Escalation Framework

The escalation framework defines which risk signals require immediate escalation to the Fraud Operations Manager versus those that can be handled through standard review workflow. The framework is designed to ensure that high-severity situations receive immediate attention while lower-severity signals are processed through standard triage.

5.1 Immediate Escalation Criteria

The following scenarios require immediate escalation to the Fraud Operations Manager, bypassing standard triage:

- Any confirmed or suspected safety threat to a participant or their pet
- Reports of harassment, threats, or abusive communication between participants
- Confirmed duplicate account or identity fraud
- Coordinated or collusive activity across multiple accounts
- Any incident that may require law enforcement notification
- Multiple high-risk complaints against a single participant within a 7-day window

5.2 Standard Escalation Path

All other signals follow the standard review workflow defined in Section 4. Escalation to the Fraud Operations Manager occurs at Stage 4 following completion of the analyst’s initial review and investigation.

5.3 Escalation Communication

Escalations to the Fraud Operations Manager are communicated through the platform’s internal Slack channel with the following information included:

- Participant account identifier
- Nature of the signal or complaint
- Risk level classification
- Summary of findings

- Recommended action

The Fraud Operations Manager acknowledges receipt and renders a decision within the timeframe appropriate to the risk level: immediate for High risk, within 24 hours for Medium, and within 48 hours for Low.

5.4 Decision Authority Matrix

Action	Analyst Authority	Fraud Operations Manager Authority
Initial triage and risk classification	Yes	N/A
Document and log signal or complaint	Yes	N/A
Recommend account action	Yes, recommendation only	N/A
Issues conduct warning	No	Yes
Temporary account suspension	No	Yes
Permanent account removal	No	Yes
Close case as unfounded	No	Yes
Escalate to law enforcement	No	Yes