Strengthening Face Liveness Against AI-Generated Identity Fraud

How I led the development of an image-based Deepfake Detection capability that protected enterprise onboarding journeys against AI-generated identity fraud — while maintaining a frictionless customer experience.

CASE STUDY ASSOCIATE PRODUCT MANAGER

Background

Our onboarding platform already offered Face Match and Face Liveness, successfully detecting replay attacks, printed photos, and video spoofing. But the rapid adoption of Generative AI introduced attack vectors our existing models were never designed to handle.

AI-Generated Content

- Deepfake videos
- Face-swapped images
- AI-generated identities

Injection Attacks

- Virtual cameras
- Emulator-based attacks
- Rooted device injection

The Core Gap

AI-generated identities exposed gaps in traditional liveness detection.

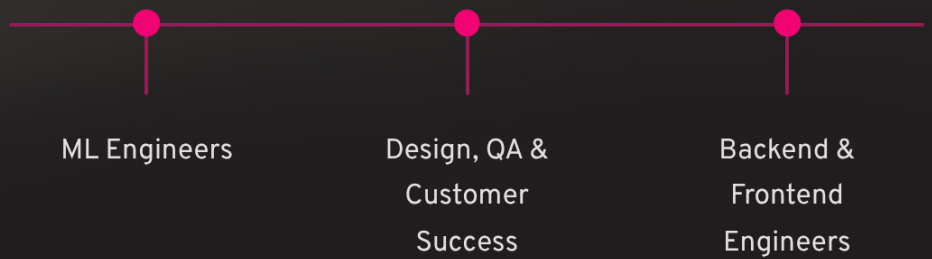
My Role

Owned the initiative end-to-end — from early customer discovery through general availability. This was a cross-functional effort spanning ML, engineering, and customer success.

What I Owned

- Customer discovery & fraud research
- Competitor analysis & PRD creation
- Product strategy & MVP definition
- API design & cross-functional execution
- Enterprise rollout & customer enablement

Who I Worked With



Problem Discovery

The path from a vague fraud concern to a shippable product wasn't linear. I structured discovery as a deliberate sequence — each stage surfacing new evidence that sharpened the problem definition and informed the final PRD.

Customer Feedback

Enterprise customers flagged emerging fraud patterns during onboarding reviews

Fraud Trend Analysis

Researched deepfake and synthetic identity attack vectors in the wild

Competitor Benchmarking

Mapped gaps across identity verification providers in the market

Capability Assessment

Evaluated internal ML capabilities and infrastructure readiness

ML Feasibility Study

Validated technical approach with ML engineers before committing

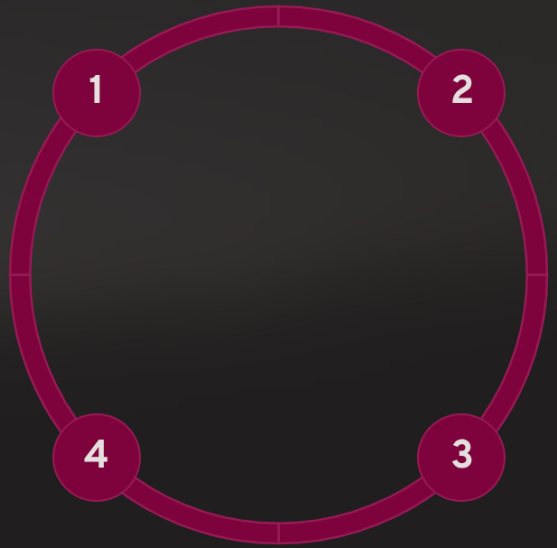
PRD & Launch

Defined requirements, aligned stakeholders, and shipped to GA

My Approach

A structured four-phase framework ensured we moved from ambiguity to a shipped product with confidence at every stage.

Discover: Researched enterprise fraud, benchmarked competitors, and identified gaps in Face Liveness.



Define: Prioritized the MVP, defined success metrics, and balanced security with customer experience.

Launch: Tested internally, optimized performance, and delivered a phased enterprise rollout.

Build: Partnered with ML to build a layered detection solution, retrain models, and define APIs.

Solution

Rather than relying on a single ML model, the solution combined four complementary security layers — each catching attack patterns the others might miss. This layered architecture was the core product decision that made the capability both accurate and resilient.



Layer 1 — Deepfake Detection Model

Primary AI-based classifier trained on real and AI-generated face samples to detect deepfakes and manipulated identities.



Layer 2 — Metadata Analysis

EXIF validation, compression anomaly detection, editing history inspection, software signature matching, and camera metadata consistency checks.



Layer 3 — Image Pattern Analysis

GAN artifact detection, edge inconsistencies, texture anomalies, frequency-domain analysis, pixel-level noise patterns, and lighting mismatch identification.

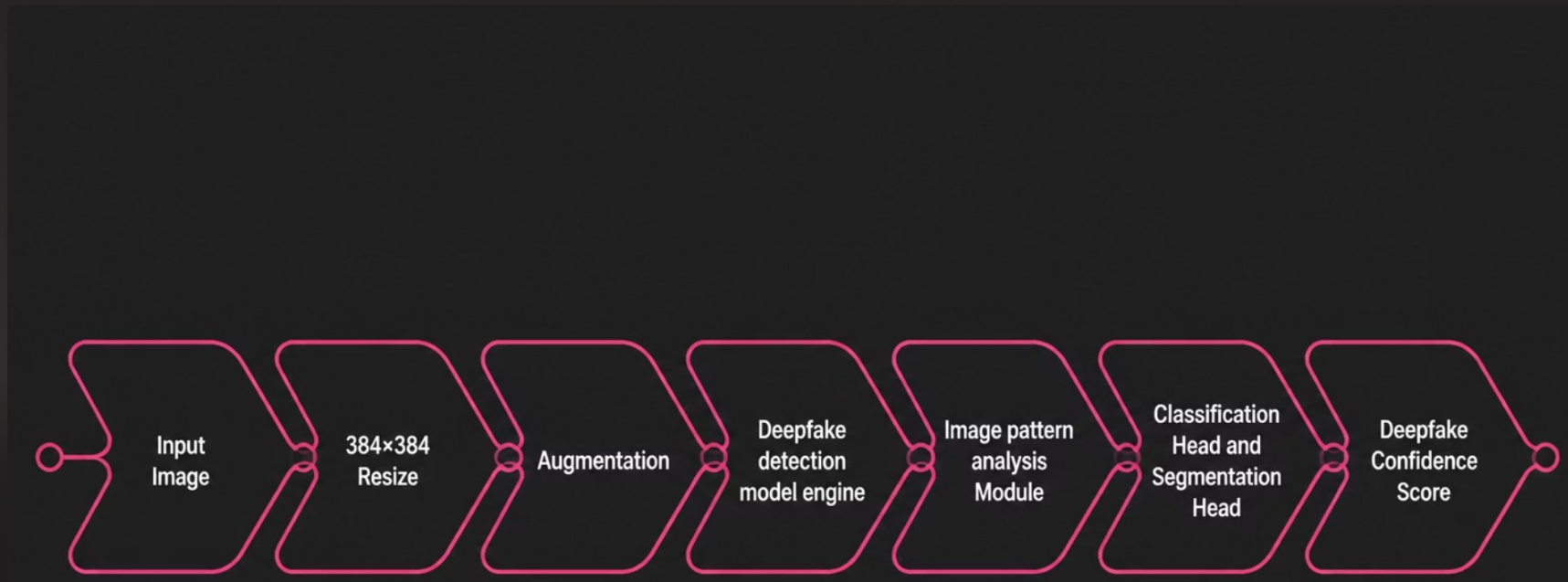


Layer 4 — Injection Attack Detection

Emulator detection, rooted device identification, virtual camera blocking, and screen replay prevention to stop manipulated image injection at the device level.

Architecture

The detection pipeline was designed for both accuracy and speed — processing images through augmentation, a deep learning backbone, and an attention module before producing a single confidence score consumed by the API.



The pipeline combines image preprocessing, AI-based detection, metadata validation, and artifact analysis to accurately identify manipulated identities while maintaining enterprise-grade performance and low latency.

Why a Localized Artifact Detection Approach

One of the key product decisions was to focus on localized manipulation artefacts rather than treating the entire image uniformly.

AI-generated images and face swaps often leave subtle inconsistencies around facial boundaries—such as the hairline, jawline, eyes, and skin edges—that are difficult to detect using traditional image classification alone. By prioritizing these regions alongside overall image analysis, the solution achieved more reliable detection of manipulated identities.

Improved Detection Accuracy

Local artifact focus catches manipulation patterns that global classifiers miss

Enterprise-Grade Inference Speed

Attention mechanism adds minimal latency — critical for onboarding UX

My Contributions

Across five areas of ownership, I drove the decisions that shaped both the product and its path to market.

Problem Discovery

Conducted customer interviews, researched fraud trends, and performed competitive analysis to define the opportunity space and validate demand.

Product Strategy

Defined the MVP scope, prioritized the roadmap, and wrote the PRD that aligned engineering, ML, and business stakeholders around a shared vision.

Cross-Functional Leadership

Coordinated across Product, ML, Engineering, Design, QA, and Customer Success — translating between technical constraints and business requirements.

Technical Decisions

Defined API contracts, set confidence thresholds, designed monitoring dashboards, and planned the rollout strategy to minimize risk.

Launch

Led enterprise pilots, optimized detection thresholds with real-world data, wrote documentation, and enabled customer success teams for GA rollout.

Results

The capability launched to enterprise customers with strong detection performance, zero added friction, and a scalable foundation for future AI-fraud capabilities.

97%

Deepfake Detection Accuracy

Across tested AI-generated and face-swapped image samples

<1s

Inference Latency

End-to-end detection completed in under one second

0

Additional User Steps

Fully invisible to the end user — no onboarding friction added

Enterprise Adoption

Successfully adopted by enterprise customers with measurable reduction in synthetic identity fraud risk and improved platform trust.

Onboarding Integrity

Maintained onboarding success rate while significantly strengthening platform security — a critical balance for enterprise customers.

Scalable Foundation

Built a reusable fraud detection framework designed to evolve with emerging AI-based attack vectors and future capability requirements.

Key Learnings

Building AI-powered fraud products is not just about model accuracy. The hardest problems were product problems — balancing prevention strength, customer experience, latency, and enterprise adoption.

Fraud prevention and UX are not opposites

The best fraud product is invisible. By embedding detection within existing Face Liveness flows, we strengthened security without adding a single user step — a non-negotiable requirement for enterprise onboarding.

Layered defense beats single-model thinking

No single model catches every attack vector. Combining deepfake classification, metadata analysis, image pattern detection, and injection attack prevention created a system more resilient than any individual layer.

Phased rollout de-risks AI launches

Internal testing → enterprise pilot → threshold tuning → GA was not just a process — it was a product strategy. Real-world data from pilots informed threshold decisions that lab evaluation never could.

PM ownership extends beyond the PRD

Success required staying close to ML architecture decisions, API design, monitoring strategy, and customer enablement — not just defining requirements and handing them off.