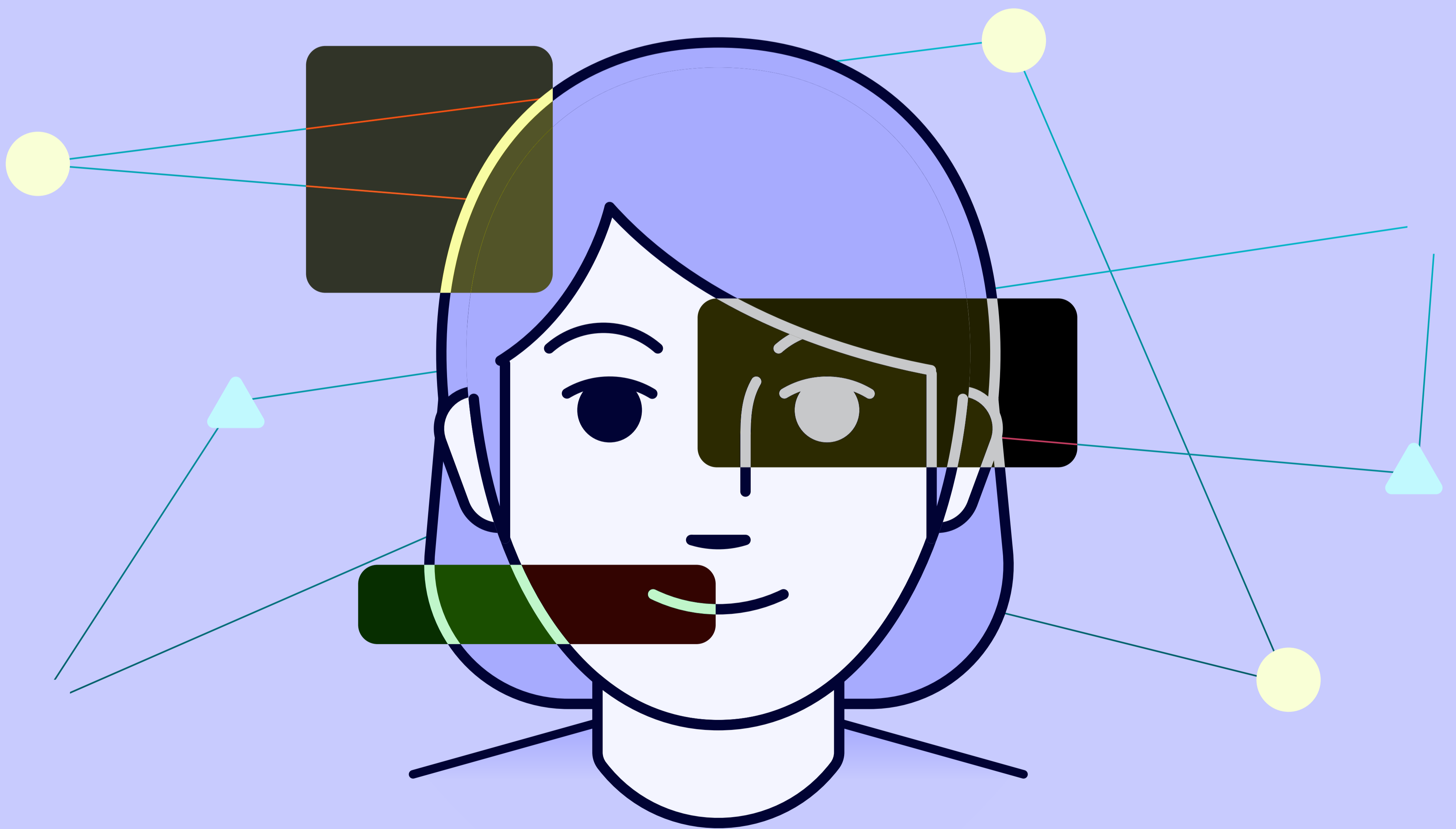
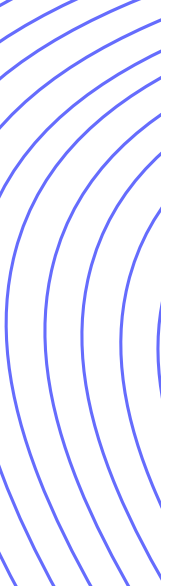


persona

# 5 strategies for fighting GenAI fraud

A worksheet for you and your team





---

If you're responsible for fighting fraud at your organization, you likely know that the job can't be done with a single tool. It's cliché, but there's truly no silver bullet for fighting fraud: fraud comes in so many forms, from so many directions, and impacts so many areas of your business that no single tactic can cover them all.

What's a well-meaning fraud-fighter to do? In our [strategic guide to fighting GenAI fraud](#), we give an answer: **build a multi-layered approach**. The holistic strategy we lay out is designed to mitigate fraud risk now and keep you optimally prepared for the future.

This asset translates that approach into a worksheet for you and your team. This worksheet aims to align your team around the steps you should be taking to fight GenAI fraud — especially the steps you may not be aware of.

Feel free to review it with your team. Assess the tactics you're employing now. Add new ones, or put them on your roadmap. Take stock of the many ways fraudsters are employing GenAI — and all the ways you can stay ahead of them.

We're happy to help! If you want to learn more about any of these approaches, feel free to [reach out any time](#) — our solutions experts are happy to discuss.

## Step 1: Collect and verify more data

The simplest way to gain assurance that your users aren't using stolen or fabricated information assembled with GenAI is to collect and verify a wide array of data that makes it difficult for fraudsters to succeed.

### Tactic 1 Expand selfie liveness / gesture set

While today's GenAI tools are good at creating high-quality single images, they have trouble generating live videos that meet requirements for gestures, eye movement, or emotions — especially if you randomize the requests.

Do you use selfie verification?

- Yes
- No

If **yes**, continue. If **no**, should you be?  
To learn more, [read this](#)

When verifying a user's identity using selfie liveness, do you have them:

- Complete the same set of liveness gestures (e.g. center, then left, and then right)
- Randomize gestures to make them harder for fraudsters to spoof

### Tactic 2 Verify ID information against authoritative and issuing databases

While often used for fake selfies, GenAI can also be used to create hyper-realistic fakes in ID portraits. How to tell if an ID is real? One answer is to verify against issuing databases.

When verifying IDs, do you check them against the following authoritative and issuing databases:

- Internal Revenue Service (IRS)
- American Association of Motor Vehicle Administrator (AAMVA)
- eCBSV
- Experian, Equifax, and/or TransUnion
- Document Verification Service (DVS) (Australia)
- Serpro (Brazil)

Other databases?

---



---

### Tactic 3 Integrate higher assurance ID security features

As fraud becomes more common, governments are starting to create more secure forms of ID. We expect this list to get much longer in the next few years.

When possible, are you verifying for:

- NFC e-passport verification
- Mobile driver's license (mDL) verification

## Step 2: Diversify and add to the passive signals you collect

By expanding the variety of signals you collect during onboarding or reverification, you can paint a more holistic picture of your users. If a bad actor wants to pass these checks, they'll need time and investments beyond GenAI tools, which makes it easier for you to decline, flag, or block them earlier in their onboarding attempts.

### Tactic 1 Add network and device signals

While these signals are less likely to catch fraudsters on their own, they're incredibly powerful when paired with other pieces of information. Look for mismatches to help identify bad actors.

#### Are you capturing:

- IP address
- VPN usage
- Browser fingerprint
- Device fingerprint
- Phone risk
- Email risk
- Location details
- All network & device signals

### Tactic 2 Explore behavioral analytics

Bots and fraudsters often exhibit very different behaviors on websites than legitimate actors do. These signals also become especially powerful when paired with the information you gather in Step 1.

#### Are you examining:

- Completion time
- Hesitation percentage
- Distraction events
- Shortcut usage

#### NOTE:

Behavioral signals may be specific to your product or offering. Feel free to list the behavioral signals you're collecting now or brainstorm behavioral signals you want to start collecting:

---



---



---



---

### Tactic 3 Assess risk with enriching information

This data can help enrich your understanding of users to potentially flag additional risk.

#### Are you looking at:

- Watchlist screening
- Social media lookup
- Adverse media check

#### Other:

---



---



---

## Step 3: Derive more insights by combining data

GenAI fraud presents itself in different ways, with fraudsters innovating rapidly using a variety of methods and tools.

The National Institute of Standards and Technology (NIST) recently found that no single provider or model can handle all fraud vectors. Instead, the most effective strategies leverage a blended approach, incorporating models trained on anomalies in the data and the way data is presented. This isn't just limited to a specific check, such as liveness – companies need to approach each risk signal as a separate model, then combine them for a more proactive coverage.

### Micromodels

*Micromodels are fraud detection tools that are trained to identify specific issues with different forms of verification. Micromodels can be trained to analyze both verification documents (the ID or selfie itself) and environmental clues (how the ID or selfie photo was taken).*

*Here are some common micromodels you can use to detect the presence of GenAI tools. Check off the ones you're using, then list the ones you want to start using:*

- Electronic replica.** Detects whether the ID or selfie was pulled up on a separate screen and displayed to the camera on a user's device.
- Printout.** Detects whether a fraudster printed out an ID or selfie on a piece of paper and held it up to the screen.
- StyleGAN.** Looks for the presence of repeated pixels.

- Deepfake.** Looks for discrepancies in facial structure, unnatural skin textures, incorrect gaze direction, and more.

- Face swaps.** Detects usage of a face-swapping app or technology.

There may be other micromodels that are relevant to your business. List them here:

---

---

---

## Ensemble models

*Ensemble models combine multiple algorithms, micromodels, and different types of data.*

*The goal is to automatically reference fraud patterns identified in one model against the patterns found in one or more other models so you can get a more holistic view into the information in any individual or group's profile*

### Example of an ensemble model:

#### Combining image and environment data to detect injection attacks

Sophisticated actors will sometimes hijack the camera stream with an injection attack. It can be difficult to catch injection attacks using image-based models. So you might combine that information with “environment” data — information from the device, camera and image metadata.

#### What it looks like

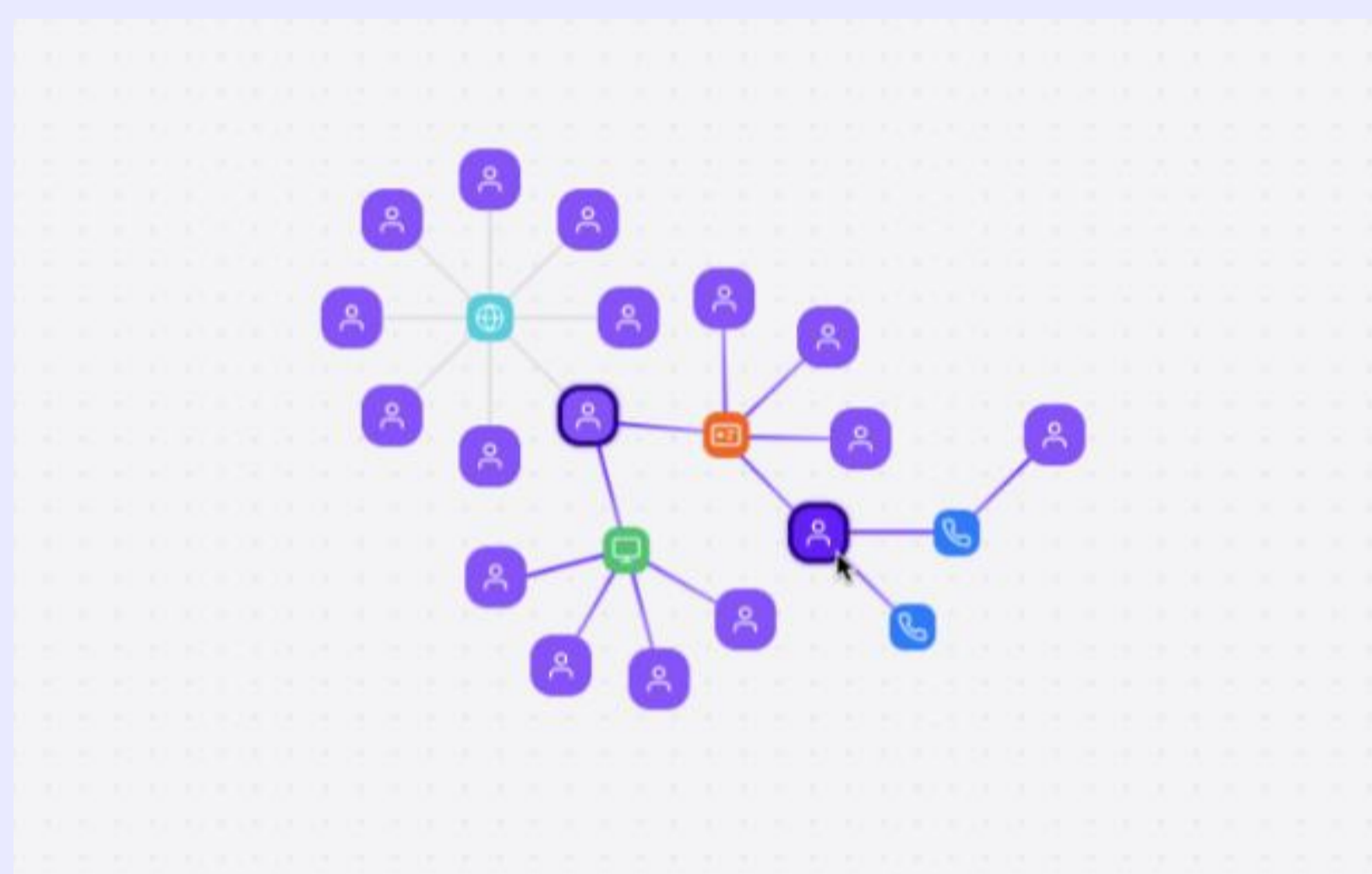
- You blend camera and device signals with image metadata to detect virtual cameras and check for device integrity
- For example, when you look at the metadata of the camera that is supposedly providing the images for the ID/Selfie, you may find that the resolution, frame-rate, dimensions, or other metadata are inconsistent with the supposed device that is being used. (E.g. The claimed device is an iPhone, which has an aspect ratio of 4:3, but the image is not 4:3.)

## Step 4: Surface suspicious connections across your user base

As fraud becomes increasingly sophisticated, sometimes the only way to tell if you're looking at a deepfake or experiencing an injection attack is by examining population-level trends and identifying repeat patterns across your user base. Especially when your engagement and volume increases, link analysis lets you find more fraud in real time.

### Tactic: Link analysis

Link analysis is a method of analyzing data that allows you to study relationships that aren't obvious in the raw data. Even more useful, today's tools can convert raw data into visual maps like this one:



#### How are you currently finding connections between accounts:

- Spreadsheets
- Ad-hoc SQL queries
- Internally built tools
- Other: \_\_\_\_\_

#### Are you able to:

- Find and visualize first, second, and third-degree connections between accounts?
- Do those things in real time?
- Consolidate fraud investigation data in a single place?
- When a known fraud gets through, can you query the bad actor's attributes — ID number, IP address, phone number, device fingerprint — to find associated users and accounts?

## Step 5:

Use an identity platform that supports active segmentation to surface the right approach at the right time

So far we've focused on ways to identify and deter bad actors. But you clearly don't want to run a full analysis on every user. That's not scalable.

A better solution: dynamically step verifications up or down based on user signals. This way good users get a smooth, simple verification experience. But you can also request more information when you detect risk. This ensures you're gathering all the information you need, when you need it.

**Step 1:**  
Map out the user journey for legitimate users on your platform.

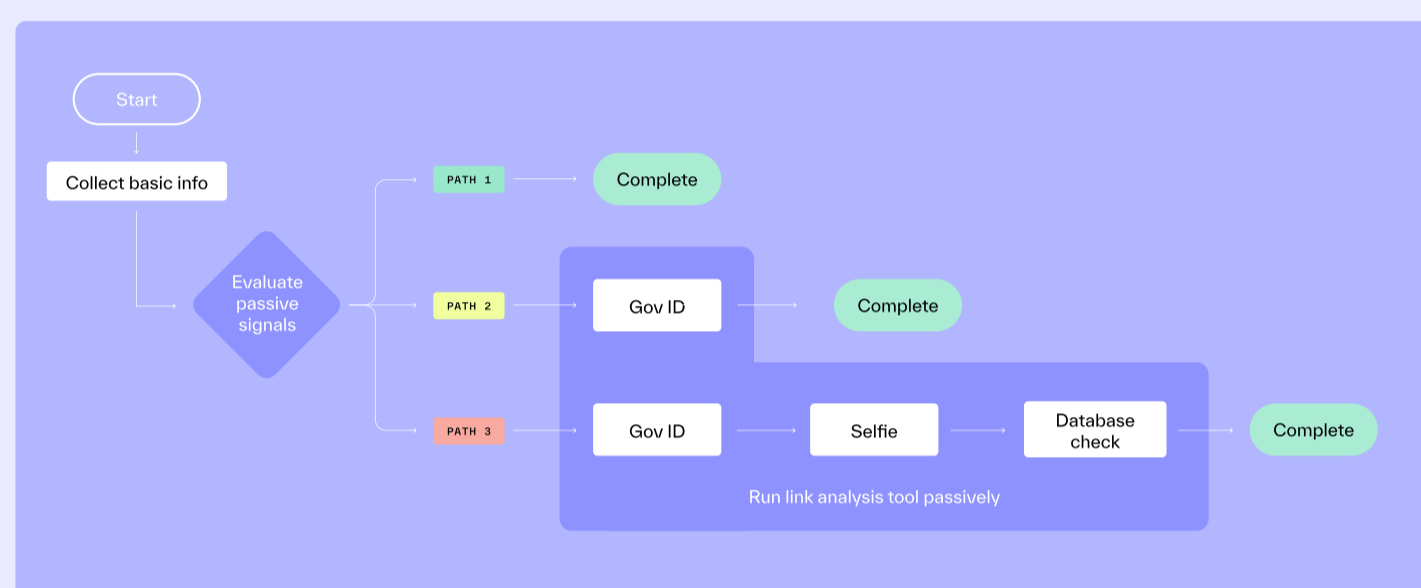
- What information do you collect?
- How do you verify it?
- What level of verification are you comfortable collecting to decide that a user is legitimate?

Draw a flowchart mapping out the steps leading to verification or rejection





**Step 2:**  
Identify key points where you'll want more information, if necessary



- Using all the tactics in Steps 1-4 of this worksheet, identify what signals should trigger more investigation.

*Example: You collect selfie information. As part of the verification, you run checks to see if the camera resolution aligns with the resolution of the image. If there's a mismatch, you might decide to trigger a step up, and ask the user to make a gesture.*

- Note them here:
  - Signal 1: \_\_\_\_\_
  - Signal 2: \_\_\_\_\_
  - Signal 3: \_\_\_\_\_
  - Signal 4: \_\_\_\_\_

- Map out the multi-pronged user flow geared around segmentation. It might look something like the diagram to the left.

Now redraw your flowchart to capture a multi-pronged flow that uses user segmentation



**Step 3:**  
Check back every quarter to make sure you're properly adding new sources of information to this user flow.

## We're Persona

At Persona, we understand just how important an efficient and effective fraud-fighting strategy is for your business. We also understand the complexities and challenges GenAI poses. This understanding has allowed us to build the industry's only identity platform that supports active, real-time segmentation and empowers fraud, risk, and trust and safety teams to catch more fraud faster.

To learn more about how Persona can help you fight fraud and protect your business and users, [contact us](#) and we'll be happy to learn more about your challenges, show you a demo, and share additional resources.

© Persona Identities, Inc. 2024

This ebook is provided "as is" for informational purposes only and is not intended as legal advice.