



## DICAS SOBRE DADOS DE IDENTIDADE

Defesa contra fraudes de  
identidade sintética



## Fraude de identidade sintética

A fraude de identidade sintética é o tipo de crime financeiro que mais cresce, de acordo com o Federal Reserve<sup>1</sup>. Os processos de abertura de conta que dependem de informações de identificação pessoal (IIPs) estáticas, como números de identidade/CPF e dados de relatórios de crédito, são os mais vulneráveis. Dependendo de IIPs estáticas é arriscado, pois muitas dessas informações já foram comprometidas.

Um exemplo básico de fraude de identidade sintética é quando os criminosos combinam informações parciais de uma pessoa real com detalhes falsos, como uma combinação de nome, endereço, contato e outras informações. A "identidade sintética" tem esse nome porque é uma síntese de informações reais e falsas, prontas para serem usadas em fraudes.

Os fraudadores semeiam e cultivam esses registros de identidade falsos para solicitar crédito e outros serviços em um momento oportuno. Além disso, estimativas confiáveis apontam que 85 a 95 por cento dos requerentes que (mais tarde) foram enquadrados como identidades sintéticas não foram inicialmente considerados como de alto risco pelos modelos tradicionais de fraudes em canais digitais.

Medidas defensivas como a biometria (p. ex., impressões digitais) estão sendo cada vez mais utilizadas para prevenir fraudes e agilizar as vendas, mas em se tratando de fraude por identidade sintética, o problema é a falta de uma "base inicial de usuários reais". Se você não tem nenhum conjunto inicial de biometria associado a uma conta de identidade sintética, os criminosos podem falsificar esses dados biométricos.

## Como um repositório de identidades avançado pode ajudar

Um dos pontos fortes mais incômodos das identidades sintéticas é que elas são parcialmente compostas por atributos bem reais — mas há um lado bom nisso. Adotar a abordagem certa para verificar identidades pode transformar esse "ponto forte" em um conjunto de contramedidas eficazes.

### Machine Learning

Uma das abordagens é analisar registros de identidade individuais e usar *machine learning* para identificar correlações mais sutis nos padrões de tráfego de dados. Por exemplo, números de telefone são unidades de informação normalmente usadas por consumidores e organizações durante inscrições, integrações e outras atividades. Isso pode exigir uma tática de verificação a montante para rastrear os números até os provedores de telefonia, a fim de confirmar há quanto tempo estão sendo utilizados pelo verdadeiro proprietário desses dados.

### Resolução de identidade

Sabendo que alguns aspectos de uma identidade sintética podem ser legítimos, nunca foi tão importante usar a estatística para calcular corretamente a possível relação de todos os atributos entre si. Nós pensamos nisso como uma "resolução de identidade de alta confiança" que reforça nossa ideia de cobertura integral. Ou seja: um repositório de identidades eficaz deve ter a capacidade de detectar a maior quantidade de atributos, com o maior nível de confiança e no território o mais amplo possível.

85%

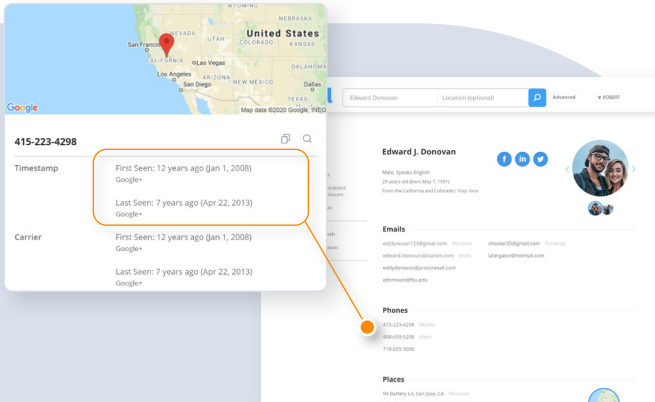
das identidades sintéticas não são inicialmente sinalizadas como de alto risco pelos modelos tradicionais de detecção de fraudes em canais digitais.<sup>2</sup>

<sup>1</sup> <https://www.federalreserve.gov/newsevents/pressreleases/other20190709a.htm>

<sup>2</sup> [https://www.idanalytics.com/wp-content/uploads/2018/11/Synthetic-Identity\\_Slipping-through-the-cracks\\_Executive-Summary.pdf](https://www.idanalytics.com/wp-content/uploads/2018/11/Synthetic-Identity_Slipping-through-the-cracks_Executive-Summary.pdf)

## Indícios de identidades sintéticas

Com repositórios de identidade vastos e precisos, é possível pesquisar em escala os atributos de identidade individuais e respectivos metadados em busca de indícios que revelem identidade sintéticas. Para isso, existem várias dicas que ajudarão analistas ou modelos de fraude a determinarem se a identidade como um todo pertence a uma pessoa real, independentemente de sua localização.



### Carimbos de data

O repositório de dados de identidade ideal para esse propósito deve ter uma função hash robusta e consistente, capaz de determinar as datas da *primeira/última visualização* de cada atributo contido nos registros de identidade.

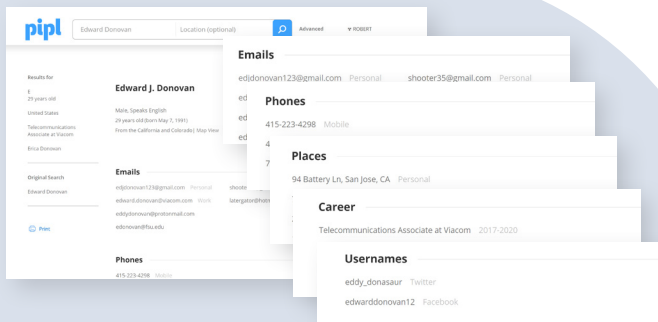
Se houver uma forte simetria entre os carimbos de *primeira/última visualização* na maioria dos atributos, isso indica uma identidade sintética.



Datas recentes de criação de atributos, como e-mails, mídias sociais e outros, podem indicar uma identidade recém-criada.



Um padrão assimétrico de carimbos de *primeira/última visualização*, como no perfil ao lado, indica uma identidade online legítima, desenvolvida ao longo de um intervalo de tempo "normal".



### Contagens (densidade de atributos)

Perfis com conteúdo abundante são necessários para tornar esse tipo de cálculo eficaz. O repositório de dados de identidade ideal deve ter um alto nível de cobertura, abrangendo muitos atributos diferentes (profundidade / amplitude).

Uma extrema escassez de atributos pode indicar uma identidade sintética.

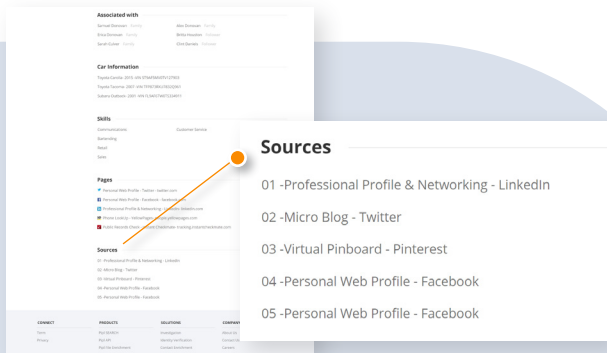


A escassez de atributos com uma corroboração mínima ou média entre fontes pode indicar registros de identidade sintética "em andamento".




Altas taxas de correspondência de atributos (especialmente quando combinadas com carimbos de data assimétricos) indicam um registro de identidade legítimo.







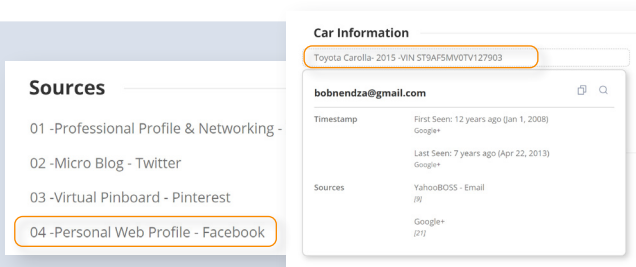
## Número de fontes

Repositórios de identidade que agregam muitos conjuntos de dados, obtidos a partir de uma ampla rede de fontes, são fundamentais para conectar características suficientes que permitam gerar uma pegada digital rica e informativa. Isso também é importante para minimizar falsos positivos entre requerentes que têm um "histórico limitado".

 A extrema escassez de fontes vinculadas é um indício de identidade sintética, pois registros inteiros podem ser construídos a partir de um único perfil de mídia social.


 Um número mínimo ou médio de fontes pode indicar fraude de identidade sintética, mas também pode indicar requerentes com um "histórico limitado".


 Uma grande quantidade de fontes (especialmente quando combinadas com carimbos de data assimétricos) é indício de um registro de identidade legítimo.




## Corroboração entre fontes online e offline

Para os criminosos, é fácil criar registros de identidade em sistemas de dados online altamente disponíveis, como contas de redes sociais. A correspondência precisa entre atributos online e registros offline, como endereço, NIV e outros recursos do "mundo real", adiciona outra camada de inteligência para reprimir identidades sintéticas e reduzir os falsos positivos por motivo de "histórico limitado".

 A ausência completa de correspondência entre registros online e offline pode indicar uma identidade sintética, especialmente se não houver um histórico de endereço físico.

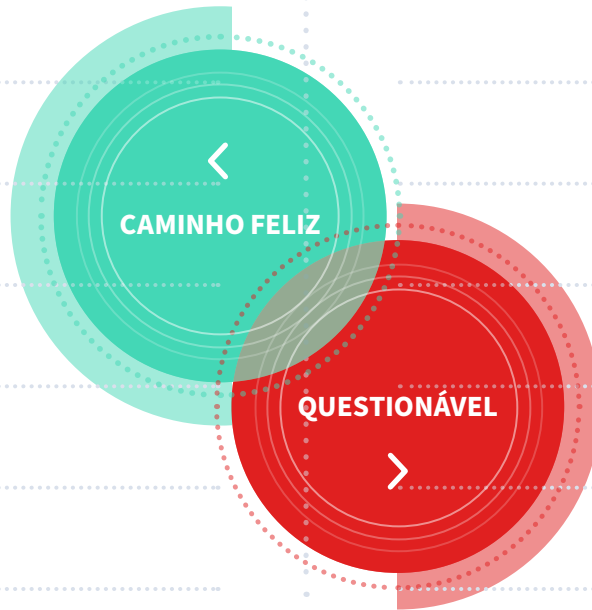
 Um grau mínimo de confirmação online/offline pode indicar um registro de identidade sintética "em andamento". Também pode ser um requerente com "histórico limitado".

 Um alto nível de corroboração da fonte (especialmente quando combinado com carimbos de data assimétricos) é indicativo de um registro de identidade legítimo.



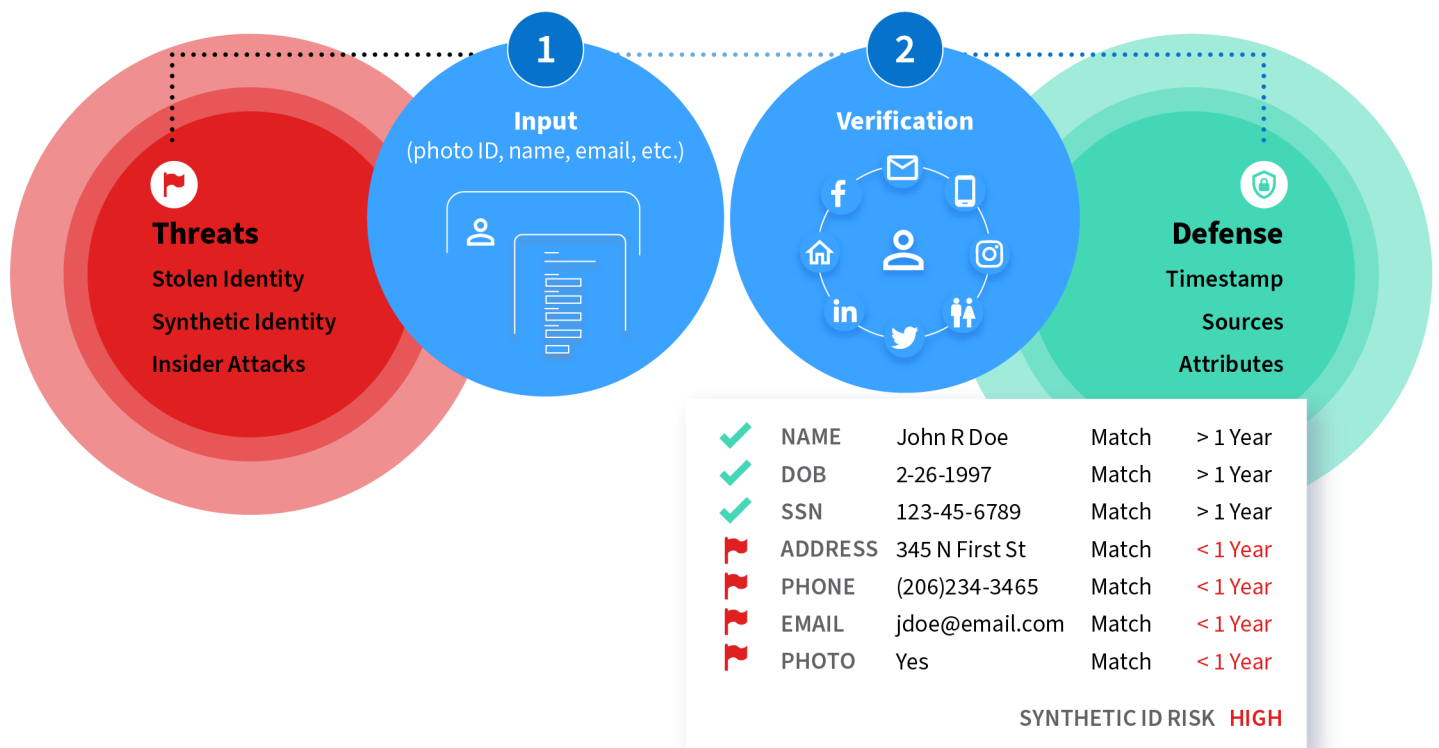
# Crie seu próprio caminho feliz de identidade

- Assimetria – datas de primeira e última visualização
- Forte combinação de registros online e offline
- Riqueza do perfil – quantidade de campos de dados e fontes
- Muitas associações com outras pessoas (com muitas fontes)
- Vários endereços físicos com carimbos de data assimétricos
- Vários números de identificação de veículos



- Endereços próximos a grandes aeroportos internacionais ou áreas de embarque
- Vários requerentes com o mesmo endereço ou número de telefone
- A abrangência do histórico de crédito é incompatível com o perfil do cliente
  - Várias contas acessadas de um único endereço IP
  - Vários usuários autorizados na mesma conta

# Fraude de identidade sintética e requerimentos



# Inteligência de identidade adicional

A Pipl retorna o histórico online e offline completo de uma identidade, contendo camadas adicionais de informação que podem ser úteis. As seguintes unidades de informação devem ser consideradas ao projetar um modelo que ajude a determinar a legitimidade de uma identidade:

## Contagens

**A quantidade de dados retornados pode estar relacionada com a confirmação de uma identidade real.**

*Contagens de campo de dados* — A resposta da Pipl API contém uma seção chamada "dados disponíveis", apresentando um resumo da quantidade de campos de dados associados à pessoa (p. ex., a quantidade de e-mails, telefones, endereços, etc.). Uma pessoa real normalmente tem vários e-mails, telefones ou endereços ao longo de sua vida.

*Número de fontes* — A identidade de uma pessoa é criada a partir de vários registros de fontes públicas. Uma pessoa real normalmente aparece em muitas fontes públicas.

## Carimbos de data

**A Pipl exibe de forma transparente as datas de primeira e última visualização para cada elemento de identidade.** Os carimbos de data podem ser um indicador preciso para diferenciar identidades sintéticas de identidades reais, uma vez que pessoas reais geralmente constam em registros públicos durante um longo período de tempo, proporcional à sua idade. Por outro lado, as identidades sintéticas costumam ter pouco histórico e duração curta entre as datas de primeira e última visualização.

## Tipos de dados

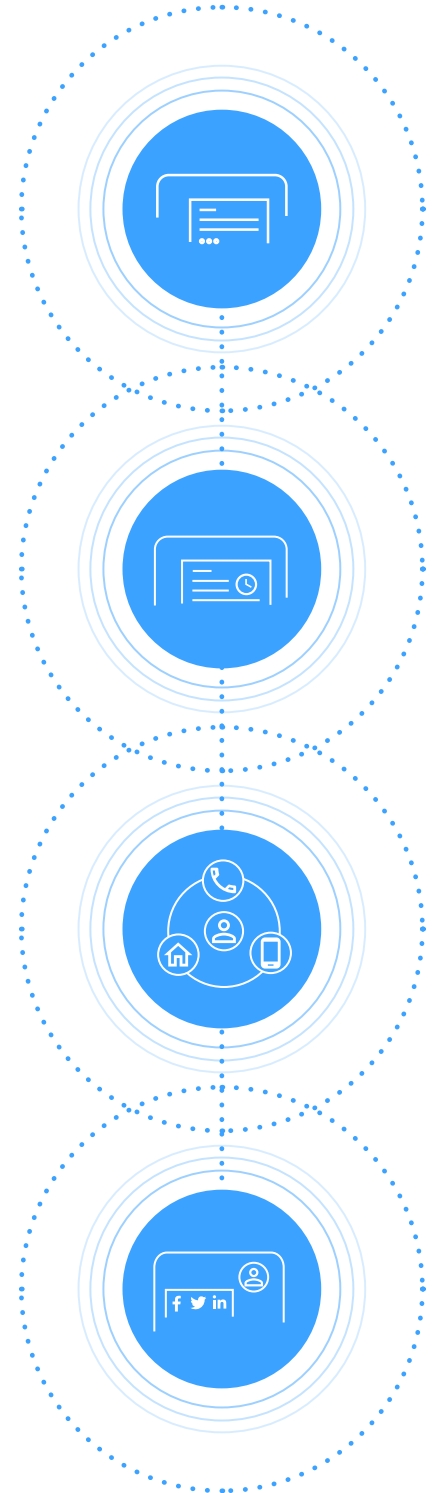
**Conhecer melhor os tipos de dados pode ajudar a confirmar uma identidade real.**

A Pipl retorna metadados que descrevem os tipos de campos de dados específicos. Por exemplo, um e-mail é indicado como pessoal ou profissional, não importa se está hospedado em um provedor gratuito ou em um serviço de e-mail descartável temporário. Da mesma forma, números de telefone são marcados como celular, residencial ou comercial.

## Indicadores booleanos

**Saber mais sobre a existência desses indicadores pode ajudar a confirmar uma identidade real.**

A Pipl retorna dados de sites de mídias sociais como Facebook, Twitter e LinkedIn, entre outros. A existência de perfis de mídias sociais ao longo do tempo pode indicar uma identidade real. Por exemplo, pessoas que têm várias contas de mídias sociais há vários anos muito provavelmente são pessoas reais, ao contrário de pessoas que acabaram de criar uma conta no mês passado. Além disso, você pode descobrir que a mera existência de um emprego em um perfil é um sinal válido.



Nosso mundo digital depende da confiança em quem está por trás de uma identidade online. Mas o próprio conceito de identidade se fragmentou em centenas de unidades de informação que os fraudadores estão constantemente tentando explorar. **Veja por que a Pipl é a melhor escolha quando as empresas precisam verificar se os dados de identidade realmente pertencem à pessoa que os utiliza.**