

# Verificação de Identidade Digital

---

Notas sobre o processo de verificação de identidade digital, com foco na prestação de serviços públicos por entidades governamentais.



## O problema

---

O clamor pela transformação digital e a pressão por serviços remotos nos períodos pandêmico e pós-pandêmico são motores de inovação e exigem criteriosos requisitos de segurança e privacidade.

Qualquer serviço digital que permita o registro ou criação de contas (*onboarding*) está passível de fraude por parte do usuário solicitante, variando de contas *fake* (onde o usuário não corresponde a uma pessoa legítima) a impersonificação (onde um agente malicioso se passa por uma pessoa legítima).

Devido à dificuldade e ao alto custo muitos serviços abrem mão da verificação, assumindo os riscos de negócio e de imagem. Porém, isso nem sempre está na mesa devido à regulações, *compliance* ou necessidades de verificação rígidas associadas a custo, requisitos de negócio e privacidade. Notavelmente, redes sociais são exemplos de indústrias com baixo foco em verificação (e alta poluição com perfis *fakes*), em oposição às *fintechs* (com seus processos de *onboarding* 100% remotos).

Serviços governamentais frequentemente fazem esse tipo de análise de *trade-off*. Isso parece estar mudando, evidências dadas pelas iniciativas federal ([gov.br](https://gov.br)), estadual ([LoginSP](#)) e municipal ([Login Único](#)) (entre tantas outras) de centralização de identidade do cidadão.

No Brasil esse processo é notavelmente complexo devido à inexistência de cadastros compreensivos unificados com um nível de completude capaz de atender a diversos casos de uso, e à miríade de documentos de identificação e suas múltiplas versões.

Nesse cenário, serão discutidas alternativas para a implantação da verificação de identidade em um cenário plausível de uma aplicação governamental para prestação de serviços ao cidadão, onde ao se registrar é necessária a genuína verificação da identidade do próprio usuário, bem como sua presença física junto ao dispositivo no ato do cadastro.

Verificação de identidade, neste texto, se refere a combinação de diversas atividades realizadas durante o evento de interação onde o usuário alega possuir uma identidade no mundo real, geralmente em sua primeira interação (registro, cadastro, *onboarding* e afins). Em diversos níveis de tolerância, busca-se verificar se a alegação procede, confirmando se a identidade real existe, se o usuário alegando posse da identidade é o seu verdadeiro possuidor, e sua presença genuína durante o processo.

Além da verificação, outros métodos podem ser utilizados, com menor nível de confiabilidade, para adicionar evidências de suporte à verificação. Vamos chamá-los de métodos de afirmação de identidade. Seu uso como única fonte de verificação é desencorajado.

Buscaremos alternativas à tradicional (e altamente confiável) verificação física de identidade realizada por agentes públicos em balcões de atendimento tradicionais, como no caso do serviço de identidade paulistano

[Senha Web](#), em prol de soluções nativamente remotas, em canais digitais.

## Técnicas

---

Várias alternativas são disponibilizadas pelo mercado. As soluções mais robustas devem utilizar a combinação de diversas técnicas em momentos diferentes para alcançar um nível de confiabilidade adequado ao problema.

Vamos discutir algumas das alternativas.

### Verificação de identidade do mundo real com foco em documentos

Consiste no uso de dispositivos de captura de imagem (*webcams* ou câmeras de *smartphones*) em canais digitais remotos (tipicamente aplicações baseadas em *browser* ou aplicativos *mobile*) para obtenção de imagens ou vídeos de documentos físicos.

Imagem: [biometricupdate.com](https://biometricupdate.com)

A partir da captura podemos:

- identificar o documento (é um documento conhecido? [é um RG](#) ou uma CNH?);
- buscar por sinais de adulteração e falsificação;
- realizar OCR dos textos;
- obter a foto de identificação.

Normalmente esse processo está associado a um mecanismo de detecção de presença "ao vivo" (*liveness detection*, ou *Presentation Attack Detection*). Nesse caso, a foto do documento é comparada com uma *selfie* do usuário tirada no momento do processo (imagem ou vídeo). Caso haja uma base prévia, também é possível realizar o reconhecimento facial.

Imagem: [computerid.com.br](https://computerid.com.br)

Este método entrega um nível aceitável de confiança, dada a presença de "algo que *somente você possui*" (seu documento) e "algo que *somente você é*" (sua face).

A técnica também é conhecida como "ID+*selfie*" ou "eKYC" (*eletronic- Know Your Consumer*).

### Afirmação de identidade do mundo real com foco em dados

Consiste na comparação dos dados fornecidos pelo usuário (nome, data de nascimento, endereço, ...) com bases de dados confiáveis (registros governamentais, dados censitários, dados financeiros, ...).

Pode se manifestar como uma verificação passiva (os dados são fornecidos pelo usuário e verificados pela aplicação) ou por algum método de desafio-resposta (onde o usuário é desafiado a responder corretamente a alguma pergunta criada com base em seus dados conhecidos previamente pela aplicação).

Fonte: [gov.br](https://gov.br)

Esta técnica foi a mais utilizada pelo mercado durante muito tempo. Ainda assim, entrega um nível baixo de confiança, dada a dependência de "algo que *não somente você sabe*" (seus dados pessoais).

Considerando o crescimento de vazamentos de dados, engenharia social e *malwares*, não é recomendado seu uso como verificação de identidade, mas somente como suporte à verificação.

## **Afirmação de identidade com foco em dispositivos**

Consiste na geração de um identificador único do usuário baseado em informações combinadas de hardware e software do dispositivo do usuário, chamado de *device fingerprint*.

### [Identificação do usuário mesmo em abas anônimas \(fingerprintjs.com\)](https://fingerprintjs.com)

Pode ser utilizado para inferir risco em alguns casos de uso de verificação de identidade em acessos subsequentes. Por exemplo, várias contas sendo criadas pelo mesmo dispositivo podem ser consideradas indício de fraude, assim como o acesso pelo mesmo dispositivo onde o cadastro foi criado pode ser um sinal de confiança.

## **Afirmação de identidade com foco em atributos digitais**

Consiste no uso de atributos digitais (como e-mail, endereços IP e perfis em redes sociais) para afirmação de identidade, preferencialmente em correlação com identidades do mundo real.

Segundo o Gartner, o e-mail tem se provado um atributo de identidade particularmente persistente, tendendo a permanecer por grandes períodos sem alteração.

Apesar de não poderem ser utilizados como identificação única e inequívoca, dados de geolocalização por IP podem ajudar a identificar fraudes caso diverjam consideravelmente do histórico do usuário ou de endereços previamente conhecidos.

### [Serviço de geolocalização por IP](#)

Também é comum considerar e-mails ou perfis recém-criados ou não encontrados em nenhuma base de dados anteriormente como potencialmente arriscados.

## **Afirmação de identidade com foco em análise de comportamento**

Consiste na criação de um perfil de usuário com base em seu comportamento (cadência de digitação, padrão de movimento de ponteiro do mouse, padrão de toque/arrasto), geralmente no primeiro uso de uma aplicação (ou de outra aplicação que compartilhe o perfil).

Pode ser usado para obter indícios de confirmação da identidade, ou para risco de fraude. Por exemplo, usuários podem ter dificuldades no preenchimento de um formulário que um agente malicioso não apresentaria (por repetir muitas vezes, ou por ser um *bot* pré-programado).

## **Afirmação de identidade com foco em número de telefone**

Consiste no uso de dados obtidos pela rede telefônica para correlacionar o usuário com uma identidade do mundo real.

Entre as informações passíveis de uso estão:

- Ligação feita pelo dispositivo a uma central que identifica a chamada;
- Análise de sinal para detecção de possíveis grampos;
- Correlação de número telefônico com cadastro de proprietário;
- Idade do *SIM card* (se muito recente pode indicar risco de fraude);
- Criação de perfil de voz;
- Identificação por voz (em acessos subsequentes).

[Brasil entre os países que suportam GSMA Mobile Connect](#)

## BYOI - *Bring Your Own Identity*

---

Um busca de reduzir a quantidade de identidades digitais que uma pessoa gerencia, o conceito de BYOI traz a ideia de reutilização das suas identidades reconhecidas em uma fonte confiável em outros pontos de acesso.

A maioria das implementações usam *Single Sign-On* (SSO) via [OAuth](#) e [OpenID](#). O nível de confiabilidade varia bastante conforme o emissor.

[Exemplos de serviços online que permitem login social no Auth0](#)

Em termos de UX, a escolha de alguns poucos provedores é preferida. O de muitas opções prejudica o usuário tanto visualmente ("[NASCAR problem](#)") quanto funcionalmente ("[paradoxo da escolha](#)").

## Redes Sociais

Identidades gerenciadas por sistemas de redes sociais, e-mails e outros serviços *online* como Facebook, Google, Microsoft, Apple e afins.

[Tela de login do Slack com opções de Login Social](#)

Podem garantir a diferenciação entre um usuário e outro, mas quase nada podem fazer para garantir que o usuário é quem ele diz ser (por exemplo, o possuidor de um CPF específico). O grande número de contas *fake* torna desencorajador o seu uso como único método de identificação de uma pessoa real (apesar frequentemente ser suficiente para o comércio de produtos e serviços, por exemplo).

Comumente são utilizadas somente para autenticação, em forma de vínculos adicionais a uma conta já verificada anteriormente.

## Governo

Identidades digitais atestadas por uma entidade governamental responsável, geralmente carregando o peso das verificações físicas associadas.

O [gov.br](#) e o [LoginSP](#) são iniciativas com esse fim, e permitem a integração das identidades entre diversos sistemas governamentais com alta confiabilidade e suporte a SSO com OAuth/OpenID. Em São Paulo, o [Senha Web](#) possui verificação física de identidade em balcões de atendimento.

### [Senha Web permite autenticação com certificado digital](#)

Um exemplo híbrido é o [e-CPF](#) (digital ou digital+físico, como nos tokens A3), que inclui as entidades certificadoras no processo, permitindo outros serviços como assinatura digital de documentos.

### [e-CPF físico \(A3\) emitido pela Prodesp](#)

É utilizado, por exemplo, para afirmação de identidade no [gov.br](#), e como método de autenticação no [Senha Web](#).

### [Afirmação de identidade com certificado digital no gov.br](#)

O uso de documentos legíveis por máquina com o uso de QRCode e NFC devem começar a se tornar realidade em um futuro próximo, pois já são suportados por alguns documentos, como na CNH, passaportes e o novo RG brasileiro.

## **Instituições Financeiras**

Bancos possuem uma grande base de usuários verificados, inclusive por razões de *compliance*, e podem prestar o serviço de verificação de identidade.

Com o crescimento do padrão [Open Banking](#), isso se torna uma realidade factível e de simples implementação, tornando os bancos em potenciais provedores de identidade para SSO.

No processo de registro do gov.br, dois métodos são disponibilizados em parceria com instituições bancárias. Em um deles usa-se SSO, e no outro gera-se um código de verificação (código NAI do [cidadão.br/DataPrev](#)) no *internet banking* que é validado no processo de registro.

### [SSO com instituições bancárias no gov.br](#)

### [Uso do NAI no registro do cidadão.br](#)

## **Operadoras de Telefonia**

O padrão [GSMA Mobile Connect](#) permite que operadoras de telefonia disponibilizem serviços de identidade com alta portabilidade por estarem vinculados aos *SIM cards* dos dispositivos.

A autenticação e a afirmação são aplicações práticas, porém a confiabilidade da verificação varia de acordo com os requisitos nacionais de identificação no registro de compra de SIM cards, além de possuir cobertura limitada a poucos países (o Brasil é um deles).

Sua operação permite o SSO aprovado automaticamente quando solicitado pelo próprio dispositivo, ou então via aprovação com SMS caso em outro dispositivo.

## [Exemplo de implementação de Mobile Connect com WSO2](#)

No futuro deve incluir biometria nos cadastros, o que pode simplificar o processo e melhorar a confiabilidade.

## Uso Corporativo e Provedores de BYOI

O mercado corporativo de verificação de identidade é bastante variado, incluindo provedores de identidade com as mais diversas fontes de verificação.

### [O provedor ID.me possui \*compliance\* para diversos serviços governamentais americanos](#)

Além disso, organizações costumam integrar seus próprios diretórios de colaboradores como fontes de identidade, normalmente usando LDAP e Active Directory.

### [Acesso usando conta Azure AD integrada para alunos das Escolas Técnicas Estaduais](#)

## Considerações adicionais para aquisição

---

### Viés demográfico

As soluções de detecção facial são probabilísticas, e dependem dos seus algoritmos e dos seus dados de treinamento. Há [diversos estudos](#) que mostram que a diferença é significativa nos resultados de acordo com características demográficas como sexo, idade e etnia.

### [Algoritmos tem baixa precisão ao avaliar mulheres negras, segundo o estudo do GenderShades](#)

Dois tipos de erros são mais comuns:

- Falso positivo: duas imagens de pessoas diferentes são avaliadas como a mesma pessoa;
- Falso negativo: duas imagens da mesma pessoa são avaliadas como pessoas diferentes.

Segundo o Gartner, o número de falsos positivos entre pessoas asiáticas pode chegar a ser 100 vezes maior do que entre caucasianos. Esse número é reduzido em algoritmos desenvolvidos na Ásia. Falsos negativos são mais comuns em mulheres e pessoas jovens.

É importante avaliar os resultados, e medir os riscos de negócio, legais e de imagem.

### Corroboração de identidade

A afirmação baseada em dados normalmente se baseia em dados de fontes autoritativas convencionais, como fontes governamentais, financeiras, postais ou eleitorais.

O uso de fontes não convencionais podem ajudar, como verificação em entidades parceiras. Por exemplo, diferentes *e-commerces* em uma mesma rede de confiança poderiam verificar se um usuário consta em uma base compartilhada, reduzindo a possibilidade de fraude.

Dados de registro também podem ser comparados em bases de dados de vazamentos, que poderiam indicar um possível uso indevido.

## Orquestração

Quanto maior o rol de funcionalidades que o caso de uso exija, menor será a chance de encontrar um fornecedor que as ofereça em um único produto. O grande desafio que se apresenta é realizar a orquestração dos diversos provedores em um *workflow* robusto que não degrade a experiência do usuário, aumente a complexidade da solução, ou os custos. Um único ponto de integração é o ideal.

São características de um orquestrador:

- Configuração de *workflows*, preferencialmente em estilo *no-code*;
- Integração entre diversos fornecedores de verificação e afirmação;
- Normalização dos diferentes resultados;
- Gestão de políticas de *workflow*, de forma a controlar a UX (*go/no-go*);
- Ferramentas de análise e monitoramento;
- Flexibilidade para execução de testes A/B, facilitando a mudança de políticas;
- Redundância de *workflows* entre diferentes fornecedores.

[Interface de configuração de \*workflow\* de identidade - Okta](#)

## Features de provedores de identidade

---

### Integração

Suporte a aplicações web e *mobile*, disponibilização de SDKs e APIs, *deploys on premises* ou *cloud*.

### Captura de imagens de documentos

Disponibilização de UX amigável de captura.

### Cobertura de formatos

Identificação e OCR de diferentes documentos de diversas geografias, e seu processo de manutenção contínua.

### *Selfie e liveness detection*

UX para captura ao vivo de *selfie*, permitindo verificação da presença do usuário.

### NFC, códigos de barras e QRCodes

Suporte à extração de dados armazenados em diferentes mídias.

### Automação

Uso de processos automatizados, usando validação por analistas somente em casos extremos.

## Acurácia

Baixa presença de falsos positivos/negativos e alta detecção de fraudes.

## Gestão do viés demográfico

Existência de métricas e clareza no trato da questão.

## Armazenamento de dados de identidade

Controle do cliente sobre os métodos e locais de armazenamento dos dados que identificam um usuário.

## Conexões com terceiros

Disponibilidade de integrações para verificação e afirmação pré-construídas, e facilidade na configuração/construção de outras.

## Deduplicação

Detecção e tratamento de duplicação de identidades.

## Autenticação

Configuração de diferentes fluxos para registro, autenticação e recuperação de contas de acordo com os requisitos do negócio.

## Aumento de dados de perfil

Dados que o provedor é capaz de adicionar a um perfil criado.

## Dashboards e relatórios

Ferramentas de gestão do processo.

## Fornecedores

---

Abaixo, uma lista **não extensiva** de fornecedores de serviços de verificação identidade. Foram listados somente *players* em que foi possível encontrar alguma referência de atendimento ao Brasil.

- [Acuant](#) \*
- [Daon](#) \*
- [Serasa Experian](#) \*
- [Idemia](#) \*
- [Idwall](#)
- [Jumio](#) \*
- [Trulioo](#) \*



- [Unico](#)
- [4stop](#) \*

\* constantes em listagens do Gartner

## Referências

---

- KHAN, Akif; CARE, Jonathan. Market Guide for Identity Proofing and Affirmation. Gartner, 2020.
- KHAN, Akif; CARE, Jonathan. Buyer's Guide for Identity Proofing. Gartner, 2021.
- TERHORST, Philipp; KOLF, Jan N; *et al.* [A Comprehensive Study on Face Recognition Biases Beyond Demographics](#). Journal of Latex class files, vol. 14, no. 8, 2015.
- BUOLAMWINI, Joy; GEBRU, Timnit. [Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification](#). Proceedings of Machine Learning Research 81:1–15, 2018.