Notas sobre Verificação de Identidade Digital

Sumário

1	O problema	2
2	Técnicas 2.1 Verificação de identidade do mundo real com foco em documentos 2.2 Afirmação de identidade do mundo real com foco em dados 2.3 Afirmação de identidade com foco em dispositivos 2.4 Afirmação de identidade com foco em atributos digitais 2.5 Afirmação de identidade com foco em análise de comportamento 2.6 Afirmação de identidade com foco em número de telefone	3 3 4 5 6 7
3	BYOI - Bring Your Own Identity	8
	3.1 Redes Sociais	8
	3.2 Governo	9
	3.3 Instituições Financeiras	11
	3.4 Operadoras de Telefonia	12
	3.5 Uso Corporativo e Provedores de BYOI	12
4	Considerações adicionais para aquisição	14
	4.1 Viés demográfico	14
	4.2 Corroboração de identidade	15
	4.3 Orquestração	16
5	Features de provedores de identidade	16
	5.1 Integração	16
	5.2 Captura de imagens de documentos	16
	5.3 Cobertura de formatos	17
	5.4 Selfie e liveness detection	17
	5.5 NFC, códigos de barras e QRCodes	17
	5.6 Automação	17
	5.7 Acurácia	17
	5.8 Gestão do viés demográfico	17
	5.9 Armazenamento de dados de identidade	17
	5.10 Conexões com terceiros	17
	5.11 Deduplicação	17
	5.12 Autenticação	17
	5.13 Aumento de dados de perfil	17
	5.14 Dashboards e relatórios	17

j	Fornecedores	18
7	Referências	18
3	Sobre	18

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

1 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, evidencias dadas pelas iniciativas federal (gov.br¹), estadual (LoginSP²) e municipal (Login Único³) (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

¹https://acesso.gov.br/

²https://login.sp.gov.br/

 $^{^3 \}rm https://legislacao.prefeitura.sp.gov.br/leis/decreto-60663-de-25-de-outubro-de-2021/detalhe$

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 Web⁴, em prol de soluções nativamente remotas, em canais digitais.

2 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.

2.1 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.com⁵

A partir da captura podemos:

• identificar o documento (é um documento conhecido? é um RG^6 ou uma CNH?);

⁴https://www.prefeitura.sp.gov.br/cidade/secretarias/fazenda/servicos/senhaweb/

 $^{^5 \}rm https://www.biometricup$ date.com/201905/digital-identity-and-document-verification-market-to-generate-15-billion-by-2024

⁶https://www.terra.com.br/noticias/infograficos/nova-carteira-de-identidade/index.htm

- 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.

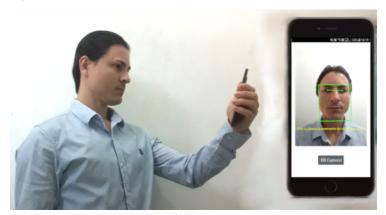


Imagem: computerid.com.br⁷

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).

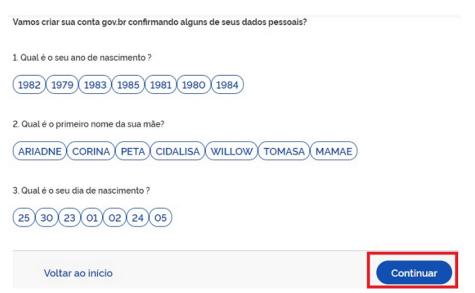
A técnica também é conhecida como "ID+selfie" ou "eKYC" (eletronic- $Know\ Your\ Consumer$).

2.2 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).

 $^{^7} https://computerid.com.br/solucoes/solucoes_view.php?c=10\&s=16\&p=9$



Fonte: gov.br⁸

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 $n\tilde{a}o$ somente $voc\hat{e}$ 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.

2.3 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)⁹

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.

 $^{^8} http://faq-login-unico.servicos.gov.br/en/latest/_perguntasdafaq/contaacesso.html#c adastro-com-as-informacoes-basicas-do-cidadao$

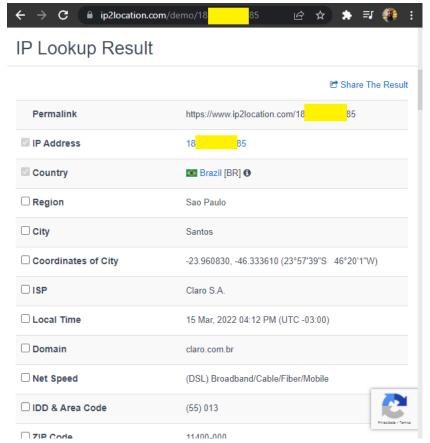
⁹https://fingerprintjs.com/

2.4 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 divirjam consideravelmente do histórico do usuário ou de endereços previamente conhecidos.



Serviço de geolocalização por IP¹⁰

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

¹⁰https://www.ip2location.com/demo/

2.5 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).

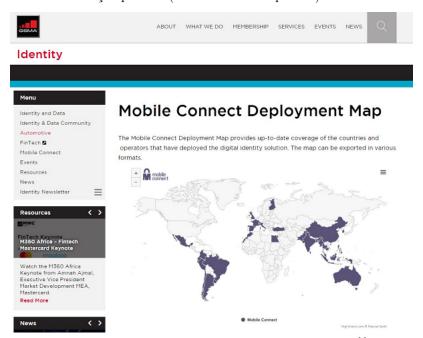
Pode ser usado para obter indícios de confirmação da identidade, ou para risco de fraude. Por exemplo, usuários podem ter dificuldades no preenchimento de um formulário que um agente malicioso não apresentaria (por repetir muitas vezes, ou por ser um *bot* pré-programado).

2.6 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).



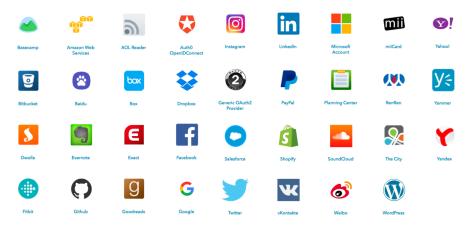
Brasil entre os países que suportam GSMA Mobile $\mathit{Connect}^{11}$

¹¹https://www.gsma.com/identity/mobile-connect-deployment-map

3 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 12 e OpenID 13 . O nível de confiabilidade varia bastante conforme o emissor.



Exemplos de serviços online que permitem login social no ${\rm Auth0^{14}}$

Em termos de UX, a escolha de alguns poucos provedores é preferida. O de muitas opções prejudica o usuário tanto visualmente ("NASCAR problem" ¹⁵) quanto funcionalmente ("paradoxo da escolha" ¹⁶).

3.1 Redes Sociais

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

¹²https://oauth.net/

¹³ https://openid.net/

¹⁴https://auth0.com/blog/social-login-on-the-rise/

 $^{^{15} \}rm https://indieweb.org/NASCAR_problem$

 $^{^{16} \}rm https://en.wikipedia.org/wiki/The_Paradox_of_Choice$



Entrar no Slack

Recomendamos usar o endereço de e-mail que você usa no trabalho.

	G Entrar com Google
	€ Entrar com Apple
	OU -
non	ne@work-email.com
	Entrar com e-mail
;\$	Enviaremos um código mágico por e-mail para entrar sem senha. Você também pode entrar manualmente .

Tela de login do Slack com opções de Login Social¹⁷

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.

3.2 Governo

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

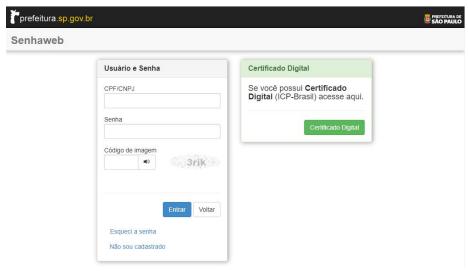
O gov.br¹⁸ e o LoginSP¹⁹ são iniciativas com esse fim, e permitem a integração das identidades entre diversos sistemas governamentais com alta confiabilidade e suporte a SSO com OAuth/OpenID. Em São Paulo, o Senha Web²⁰ possui verificação física de identidade em balcões de atendimento.

¹⁷https://slack.com/signin#/signin

¹⁸https://acesso.gov.br/

¹⁹https://login.sp.gov.br/

²⁰https://www.prefeitura.sp.gov.br/cidade/secretarias/fazenda/servicos/senhaweb/



Senha Web permite autenticação com certificado digital²¹

Um exemplo híbrido é o e-CPF 22 (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 23

É utilizado, por exemplo, para afirmação de identidade no gov.br 24 , e como método de autenticação no Senha Web 25 .

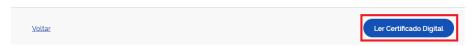
* > Privacidade > Selos de confiabilidade > Obter Confiabilidade por Certificado Digital

Obter Confiabilidade por Certificado Digital

O certificado digital ICP-Brasil funciona como uma identidade virtual que permite a identificação segura e inequivoca do autor de uma mensagem ou transação feita em meios eletrônicos, como a web. Esse documento eletrônico é gerado e assinado por uma terceira parte confiável, ou seja, uma Autoridade Certificadora (AC) que, seguindo regras estabelecidas pelo Comité Gestor da ICP-Brasil, associa uma entidade (pessoa, processo, servidor) a um par de chaves criptográficas. Os certificados contém os dados de seu titular conforme detalhado na Política de Segurança de cada Autoridade Certificadora.



CPF



 $^{^{21} \}rm https://senhawebsts.prefeitura.sp.gov.br/$

 $^{^{22} \}rm https://certificadodigital.imprensaoficial.com.br/certificados-digitais/e-cpf$

 $^{^{23} \}rm https://certificadodigital.imprensaoficial.com.br/certificados-digitais/e-cpf$

 $^{^{24}} http://faq-login-unico.servicos.gov.br/en/latest/_perguntasdafaq/comoadquirircertificadodigitalpessoafisica.html$

²⁵https://www.prefeitura.sp.gov.br/cidade/secretarias/fazenda/servicos/senhaweb/

Afirmação de identidade com certificado digital no gov.br²⁶

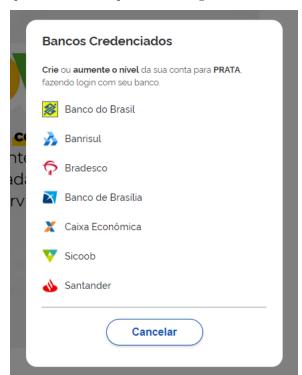
O uso de documentos legíveis por máquina com o uso de QrCode e NFC devem começar a se tornar realidade em um futuro próximo, pois já são suportados por alguns documentos, como na CNH, passaportes e o novo RG brasileiro.

3.3 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 27 , 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 $cidadão.br/{\rm DataPrev}$) no $internet\ banking$ que é validado no processo de registro.



SSO com instituições bancárias no gov.
b $\rm r^{28}$

 $^{^{26} \}rm http://faq\mbox{-}login\mbox{-}unico.servicos.gov.br/en/latest/_perguntasdafaq/comoadquirircertific adodigitalpessoafisica.html#como-atribuir-o-selo-certificado-digital-de-pessoa-fisica$

 $^{^{27} {\}rm https://openbankingbrasil.org.br/}$

²⁸https://acesso.gov.br/



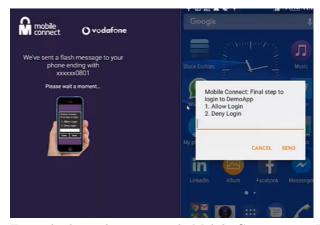
Uso do NAI no registro do cidadão.
b $\rm r^{29}$

3.4 Operadoras de Telefonia

O padrão GSMA Mobile Connect 30 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 ${\rm WSO2^{31}}$

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

3.5 Uso Corporativo e Provedores de BYOI

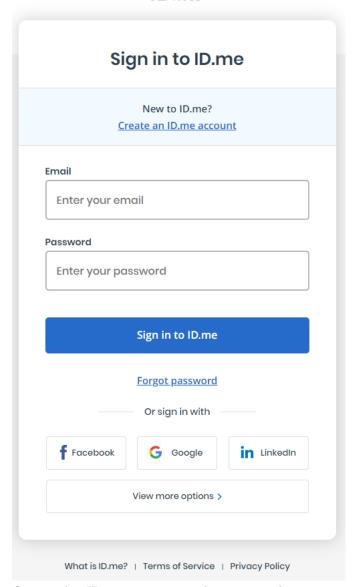
O mercado corporativo de verificação de identidade é bastante variado, incluindo provedores de identidade com as mais diversas fontes de verificação.

 $^{^{29}} https://mte.api.dataprev.gov.br/auth/login?pat_first_access=true$

³⁰https://www.gsma.com/identity/mobile-connect

 $^{^{31} \}rm https://wso2.com/library/webinars/2016/11/securing-access-to-saas-apps-with-gsma-mobile-connect/$

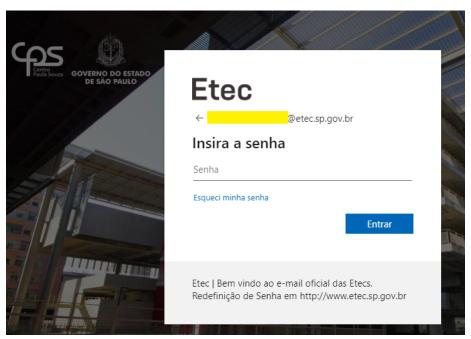
ID.me



O provedor ID.
me possuicompliance para diversos serviços governamenta
is americanos 32

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

³²https://www.id.me/business



Acesso usando conta Azure AD integrada para alunos das Escolas Técnicas $\rm Estaduais^{33}$

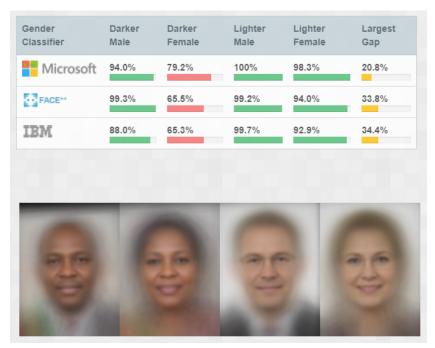
Considerações adicionais para aquisição 4

4.1 Viés demográfico

As soluções de detecção facial são probabilísticas, e dependem dos seus algoritmos e dos seus dados de treinamento. Há diversos 34 estudos 35 que mostram que a diferença é significativa nos resultados de acordo com características demográficas como sexo, idade e etnia.

³³http://etec.sp.gov.br/

 $^{^{34}}$ https://arxiv.org/pdf/2103.01592.pdf 35 http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf



Algoritmos tem baixa precisão ao avaliar mulheres negras, segundo o estudo do GenderShades 36

Dois tipos de erros são mais comuns:

- Falso positivo: duas imagens de pessoas diferentes são avaliadas como o 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.

4.2 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.

³⁶http://gendershades.org/

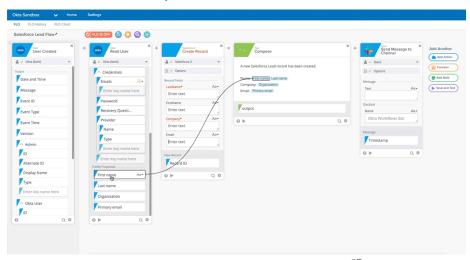
Dados de registro também podem ser comparados em bases de dados de vazamentos, que poderiam indicar um possível uso indevido.

4.3 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 - Okta³⁷

5 Features de provedores de identidade

5.1 Integração

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

5.2 Captura de imagens de documentos

Disponibilização de UX amigável de captura.

 $^{^{37} \}rm https://www.okta.com/platform/workflows/workflows-for-lifecycle-management/$

5.3 Cobertura de formatos

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

5.4 Selfie e liveness detection

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

5.5 NFC, códigos de barras e QRCodes

Suporte à extração de dados armazenados em diferentes mídias.

5.6 Automação

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

5.7 Acurácia

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

5.8 Gestão do viés demográfico

Existência de métricas e clareza no trato da questão.

5.9 Armazenamento de dados de identidade

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

5.10 Conexões com terceiros

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

5.11 Deduplicação

Detecção e tratamento de duplicação de identidades.

5.12 Autenticação

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

5.13 Aumento de dados de perfil

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

5.14 Dashboards e relatórios

Ferramentas de gestão do processo.

6 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.

- Acuant³⁸ *
- Daon³⁹ *
- Serasa Experian⁴⁰ *
- $Idemia^{41}$
- $Idwall^{42}$
- Jumio⁴³ *
- Trulioo⁴⁴ *
- Unico 45
- 4stop⁴⁶ *

7 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 Demographics⁴⁷. Journal of Latex class files, vol. 14, no. 8, 2015.
- BUOLAMWINI, Joy; GEBRU, Timnit. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification⁴⁸. Proceedings of Machine Learning Research 81:1–15, 2018.

8 Sobre

Notas sobre Verificação de Identidade Digital

Autores:

Ermogenes Palacio⁴⁹

Eduardo Amaro Bueno

Prodam/SP - Núcleo de Inovação

Março de 2022

```
38 https://www.acuant.com/
39 https://www.daon.com/
40 https://www.serasaexperian.com.br/solucoes/crosscore/
41 https://www.idemia.com/
42 https://idwall.co/
43 https://www.jumio.com/
44 https://www.trulioo.com/
45 https://unico.io/unico-check/
46 https://4stop.com/
47 https://arxiv.org/pdf/2103.01592.pdf
48 http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf
49 https://github.com/ermogenes/
```

^{*} constantes em listagens do Gartner

Disponível no GitHub: ${\rm repo^{50}} \mid {\rm pdf^{51}}$

