Improving Trustworthiness of Identity using Biometrics, Computer Vision and Cryptography

TrustStamp

https://truststamp.ai
https://www.aiid.co

AiiD

L-Università ta' Malta

Dr Norman Poh | npoh@truststamp.net
Chief Science Officer, Trust Stamp and AiiD Global
Affiliate Associate Professor, University of Malta

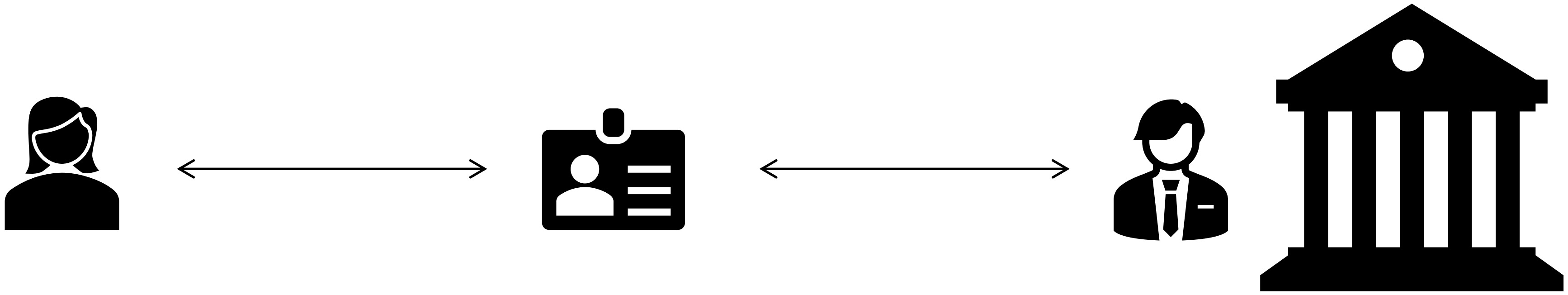
Malta Office:
Level 1, Tagliaferro Business Centre,
High Street, Sliema, SLM 1551, Malta

Atlanta Office:
3017 Bolling Way NE, Suite 248
Atlanta, GA 30305

Other offices:
Rwanda, Netherlands, Philippines, Denmark,
New York, Silicon Valley

IAPR/IEEE Biometric Winter School 2023

Trust and identity



- Who are you? Can I trust you?
- How to establish your identity with a high certainty (“assurance”)?
- How can we leverage biometrics without the liability of privacy risk?
- How AI (computer vision and biometrics), together with cryptography can address the above needs?

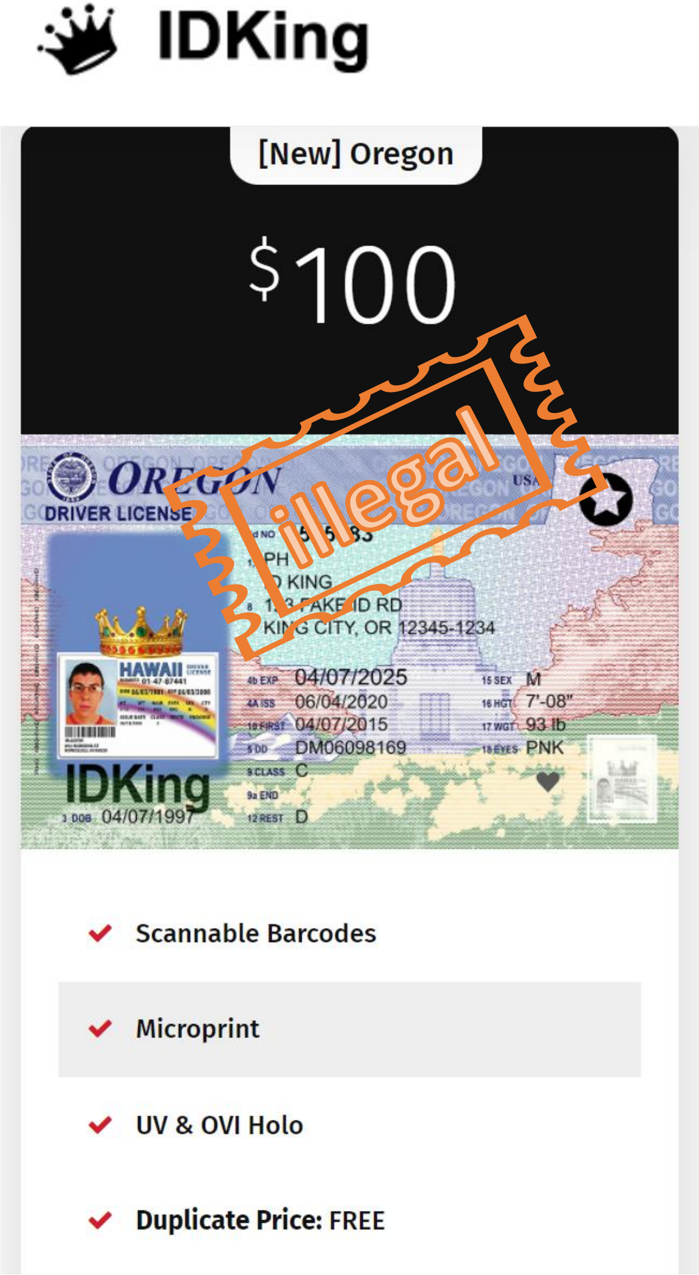
Service provider

- bank account
- dating website membership
- welfare / social security
- driving license
- Passport/visa

Proliferation of Fake IDs

- Fake ID/student cards, passports, vaccine certificates costing $80 and €150
- **Problem**: The documents are not biometrically bound to the holder
- **Countermeasure**: Only the legitimate holder that can be *biometrically verified* with *a provably legitimate document*, which can be *cryptographically verified,* constitutes a valid claim.

https://www.complaintsboard.com/bycategory/fake-novelty-id



IDKing

[New] Oregon

$100

✓ Scannable Barcodes

✓ Microprint

✓ UV & OVI Holo

✓ **Duplicate Price:** FREE

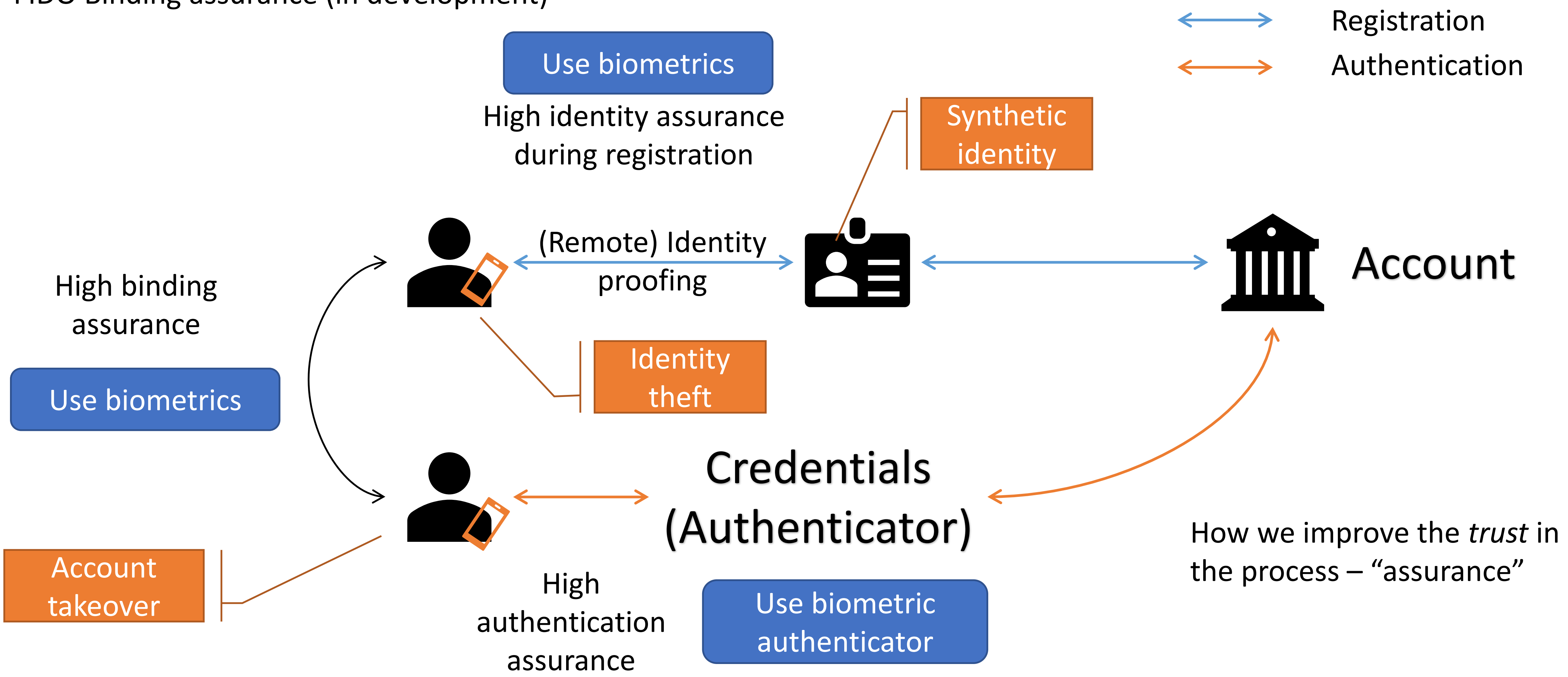
ID frauds are committed remotely

What roles can AI play?

Who are you? Can I trust you?

Biometrically bound credential

Relevant standards: ISO/IEC 29115, NIST 800-63 (US), TDIF (Australia), eIDAS (EU), GPG-45 (UK)
FIDO Binding assurance (in development)



⟷ Registration

⟷ Authentication

Use biometrics

High identity assurance during registration

Synthetic identity

(Remote) Identity proofing

Account

High binding assurance

Use biometrics

Identity theft

Credentials (Authenticator)

Account takeover

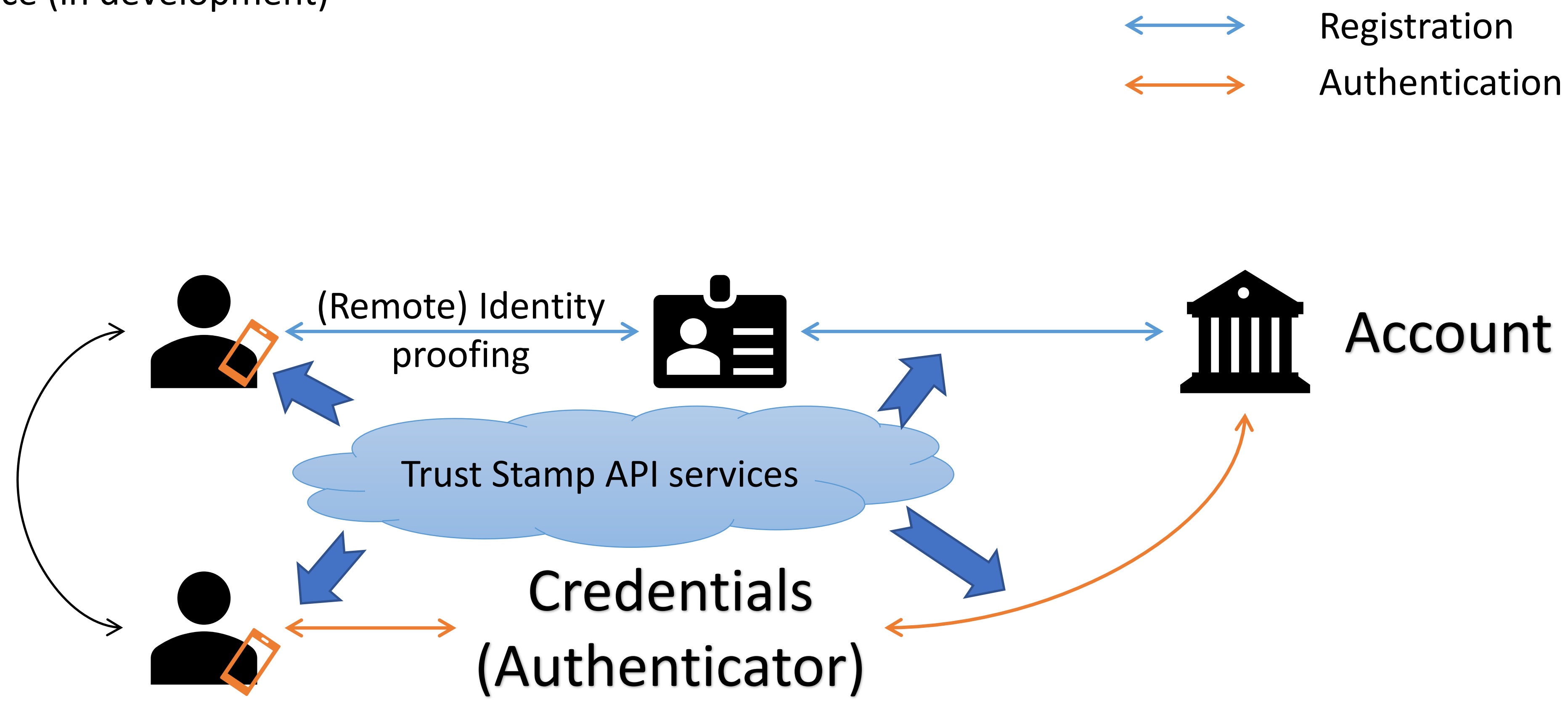
High authentication assurance

Use biometric authenticator

How we improve the *trust* in the process – "assurance"

Biometrically bound credential

Relevant standards: ISO/IEC 29115, NIST 800-63 (US), TDIF (Australia), eIDAS (EU), GPG-45 (UK)
FIDO Binding assurance (in development)



Registration

Authentication

(Remote) Identity proofing

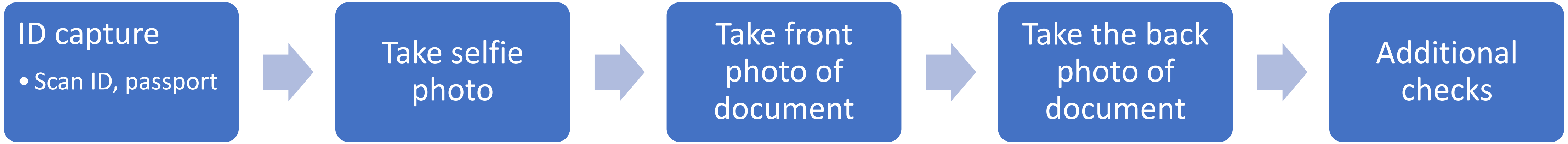
Account

Trust Stamp API services

Credentials (Authenticator)

ID creation: Remote ID proofing

also known as client-onboarding or eKYC



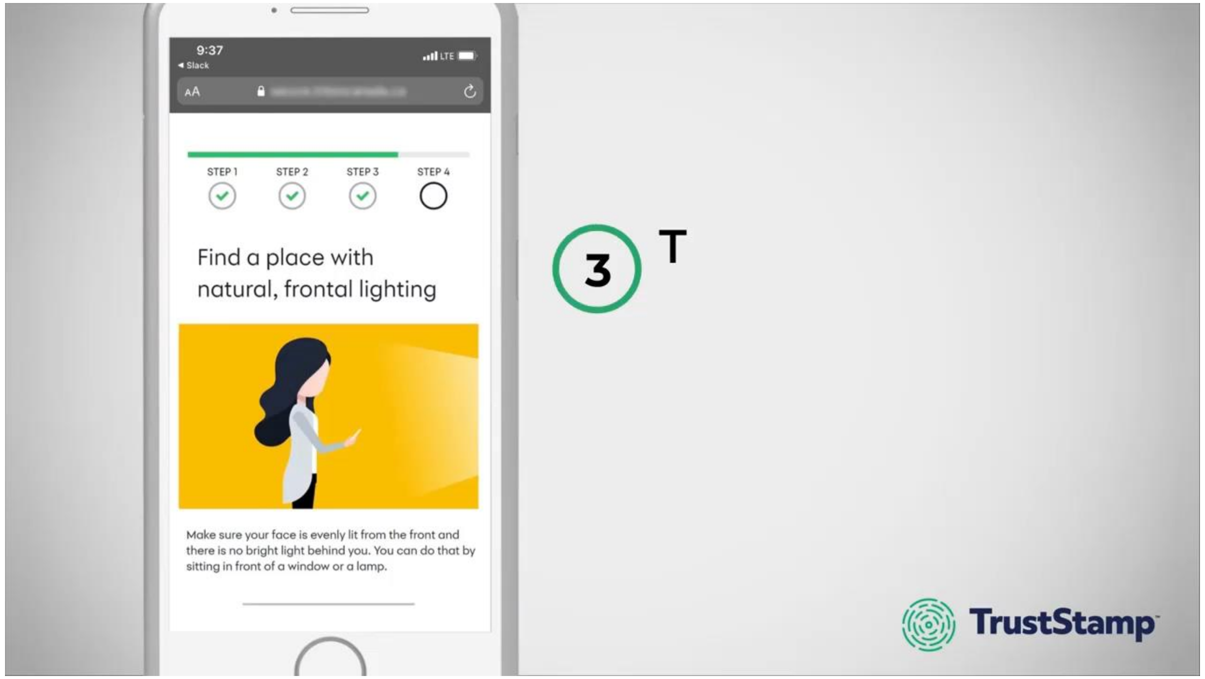
ID capture
- Scan ID, passport

Take selfie photo

Take front photo of document

Take the back photo of document

Additional checks



Find a place with natural, frontal lighting

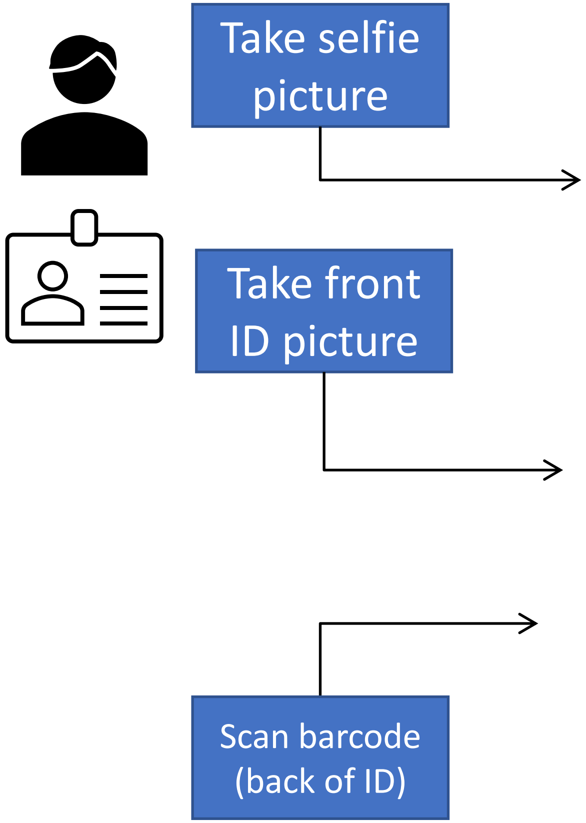
Make sure your face is evenly lit from the front and there is no bright light behind you. You can do that by sitting in front of a window or a lamp.

https://youtu.be/wiiw8flhhco?t=47

Proving your identity when enrolling for something remotely

- Opening a bank account
- Registering for a dating website membership
- Applying for a welfare, social security, or free healthcare service
- Renewing a driving license or passport

Remote Identity proofing



Take selfie picture

Take front ID picture

Scan barcode (back of ID)

	Face recognition	Computer vision	Machine learning	Cryptography
Checks	**Face**	**Document**	**Barcode**	**Chip reading**
Authenticity check	Face is live	Doc is genuine	Barcode is formatted correctly	Digitally signed by authority
Validity check (not expired)	Face watchlist check	Doc is still valid	Barcode is still valid	Certificate is valid
Consistency check	Selfie-to-photo ID comparison			
		OCR-barcode consistency check		
		Document-chip consistency check		

PAD=Presentation Attack Detection

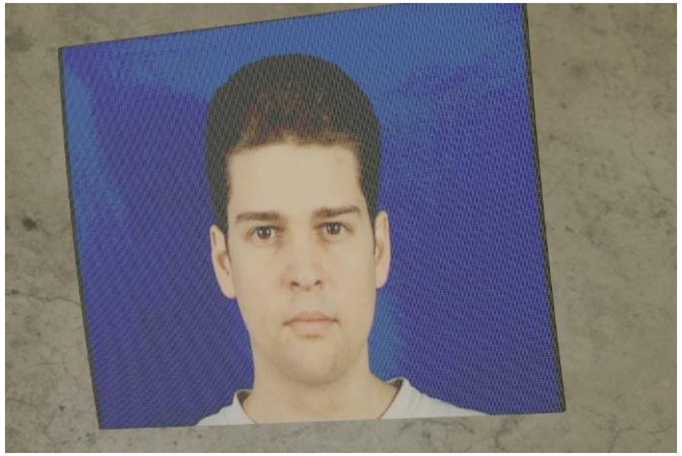
Using AI to support identity proofing

	Applied		Core AI research				
			Self-supervised learning	Outlier detection	Fast inference on browsers/edge devices	Cryptography	Synthetic data generation
Biometrics	Face recognition	Cross-domain face recognition	X				X
		Face presentation attack	X	X	X		X
	Contactless biometrics	User authentication	X		X		X
	Biometric cryptosystems	Maximize entropy				X	X
Documents	Document processing and recognition	Document segmentation	X		X		X
		Document forgery detection	X	X		X	X

- Face presentation attack
- Synthetic data generation
- Contactless biometrics
- Document segmentation
- Biometric cryptosystems

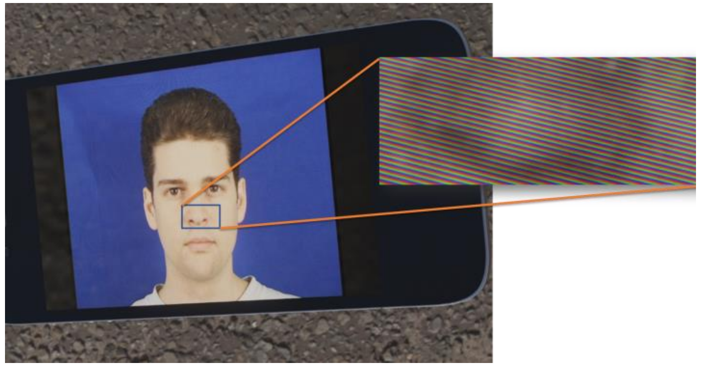
Passive PAD solutions

Print attack [p]



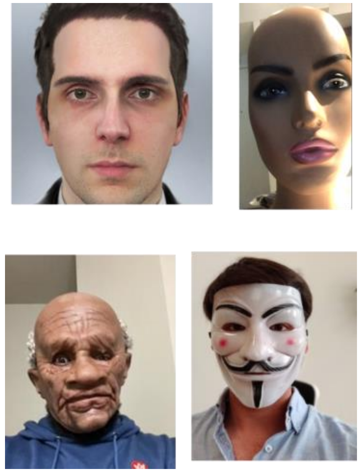
Face photos extracted from ID cards, selfie or studio photos printed on photo papers or printed by various printers (inkjet, laser jet and photocopier)

Display attack [d]



Face images displayed by PDA, tablets, smartphones, laptop screens or PC monitors

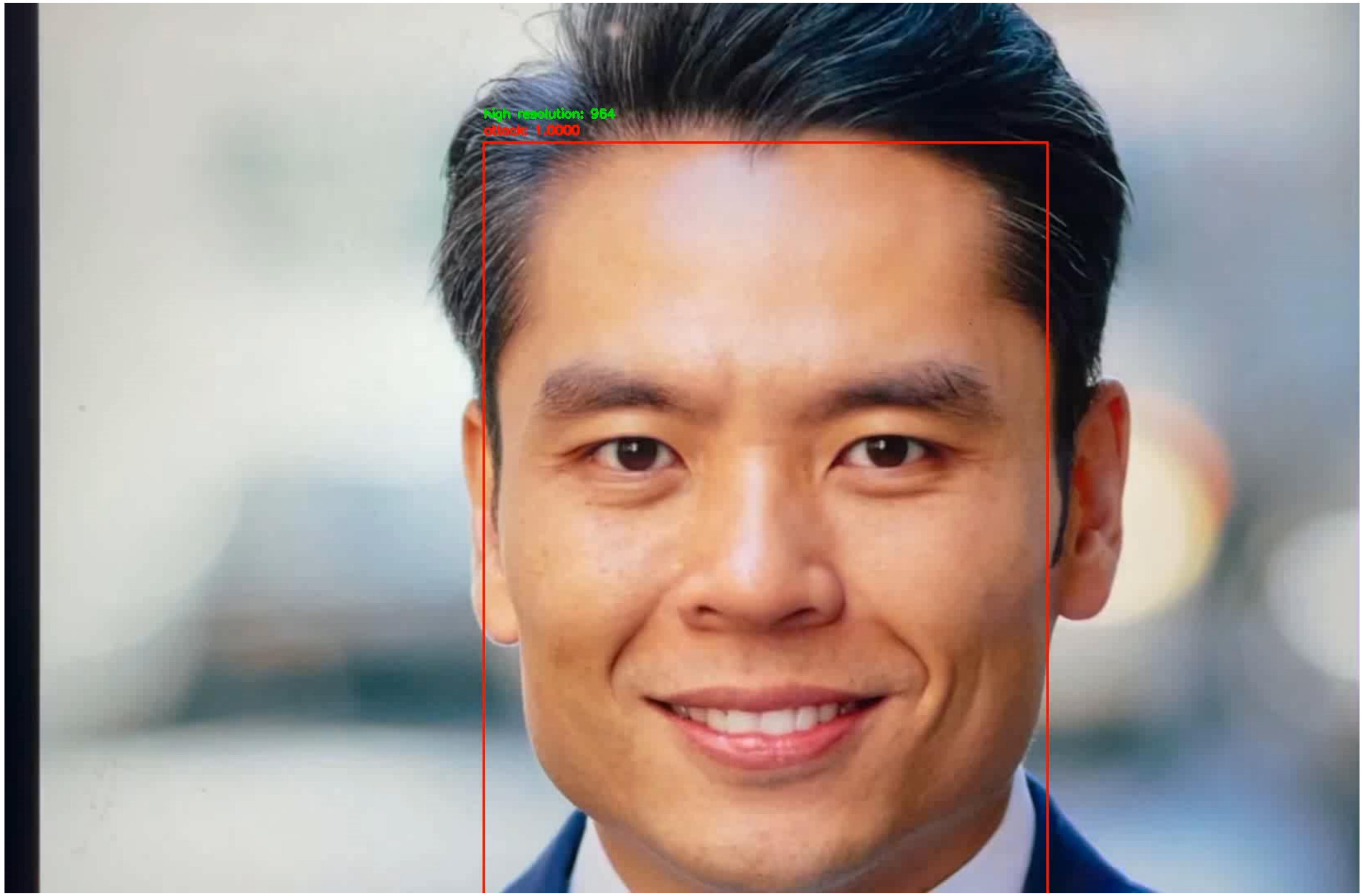
Face mask attack [m]



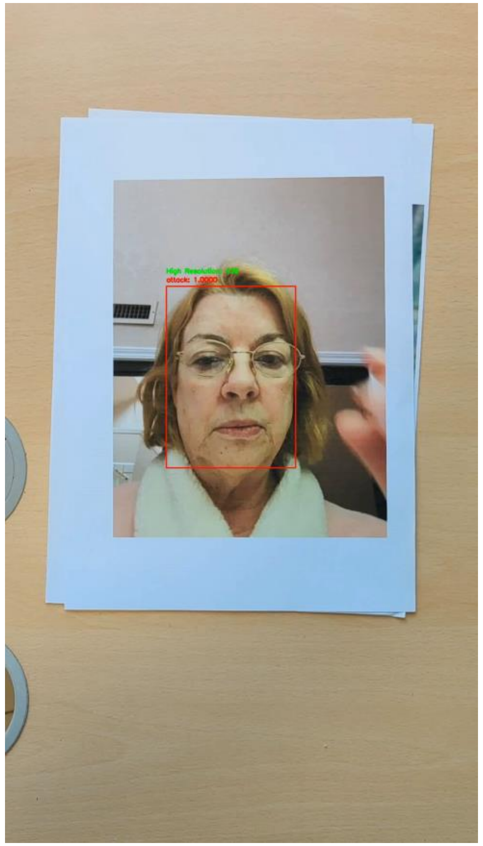
hyper-realistic face images (produced by 3D artists), mannequin heads, 3D masks

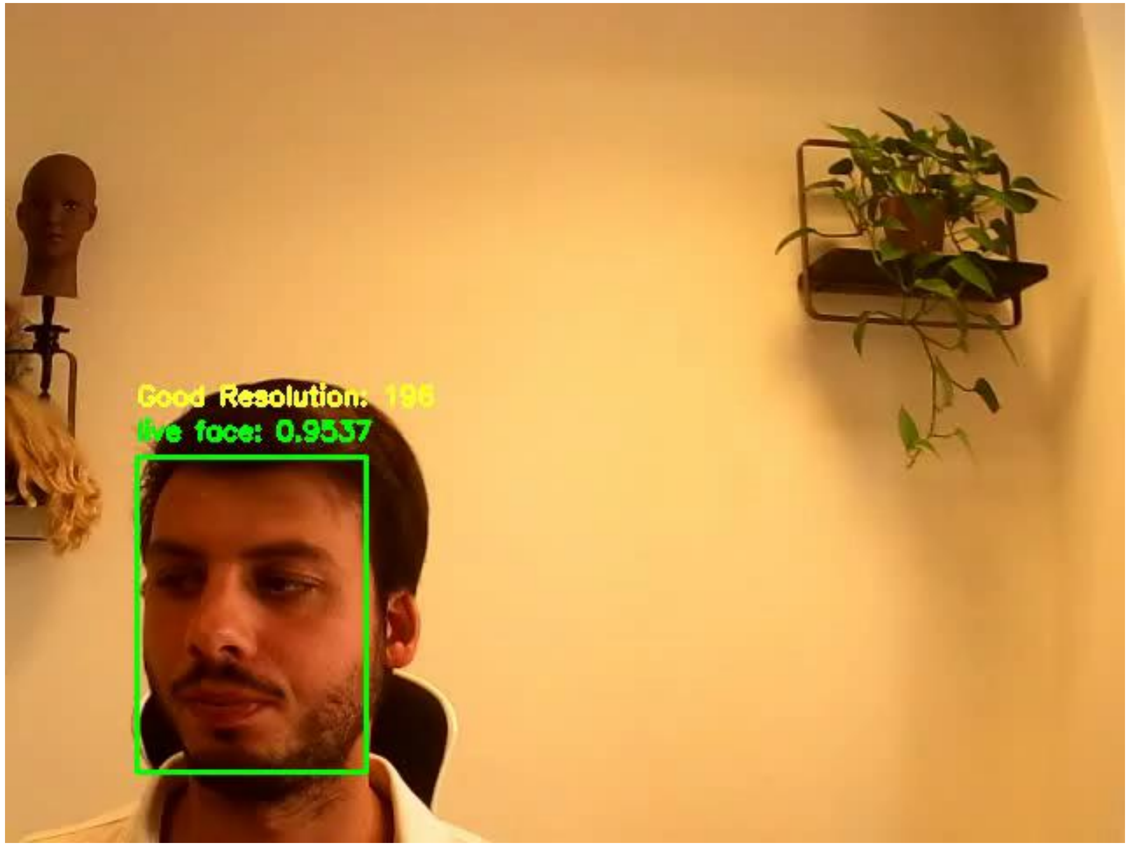
Face presentation attack | Synthetic data generation | Contactless biometrics | Document segmentation | Biometric cryptosystems

Optimal conditions



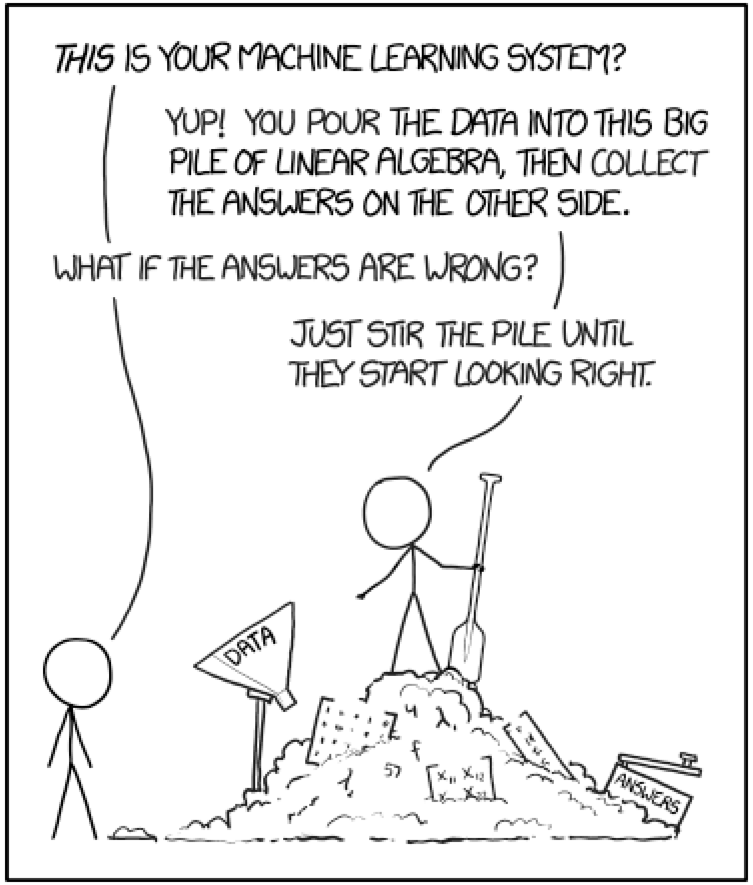
Challenging conditions



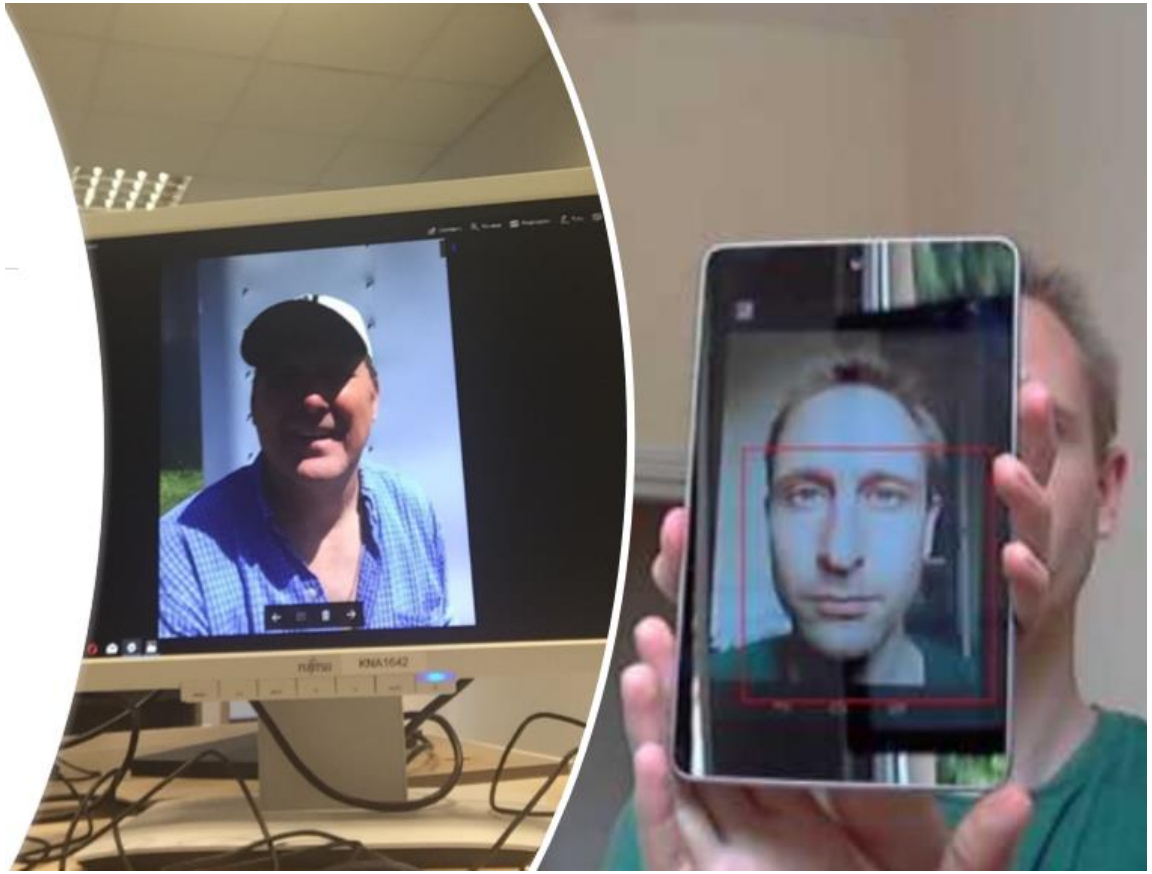


Other Issues

How to create one model that can run on server and on device?
How to generalize to *unseen* attack types?


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

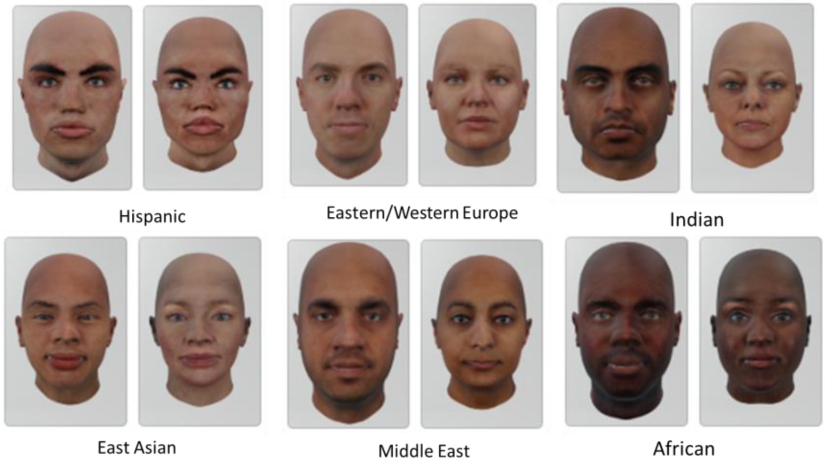
https://xkcd.com/1838

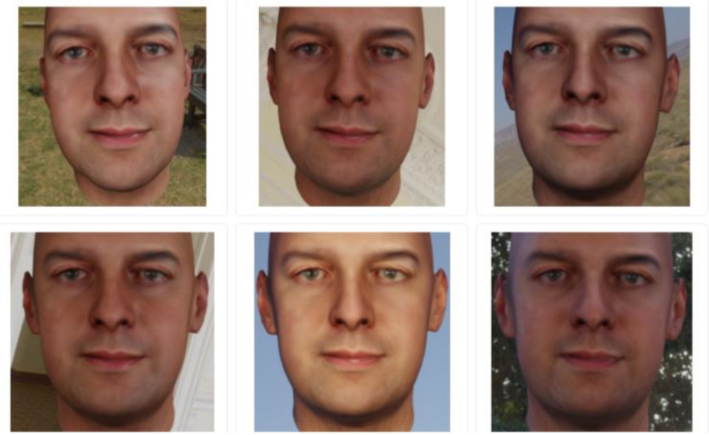


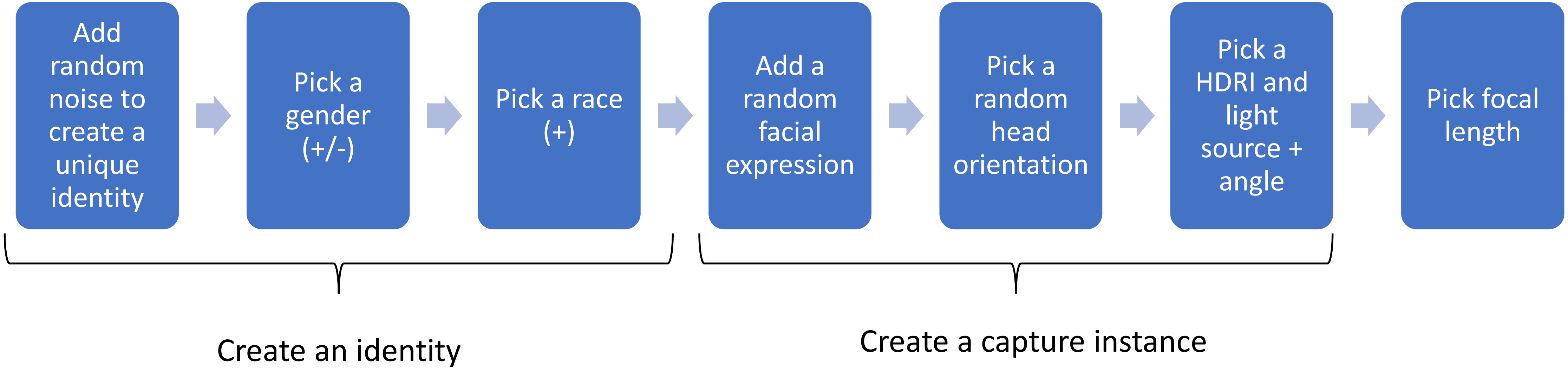
Why working with synthetic data?

Real data	Synthetic data
Costly data collection; need to incentivise data subjects and data collection operators	Cost effective; pay only for the compute time and blender development time
Limited number of subjects and samples per subject	Can generate infinite amount of data in terms of subjects and samples per subject
Privacy issue causing limited data retention period	No need to worry about data privacy
Uncontrolled factors during data collection	Full and precise control over the 3D virtual ambient environment
Mistakes happened in labelling	Accurate data with full metadata
Realistic conditions	Not always realistic

Using 3D head model







Add random noise to create a unique identity

Pick a gender (+/-)

Pick a race (+)

Add a random facial expression

Pick a random head orientation

Pick a HDRI and light source + angle

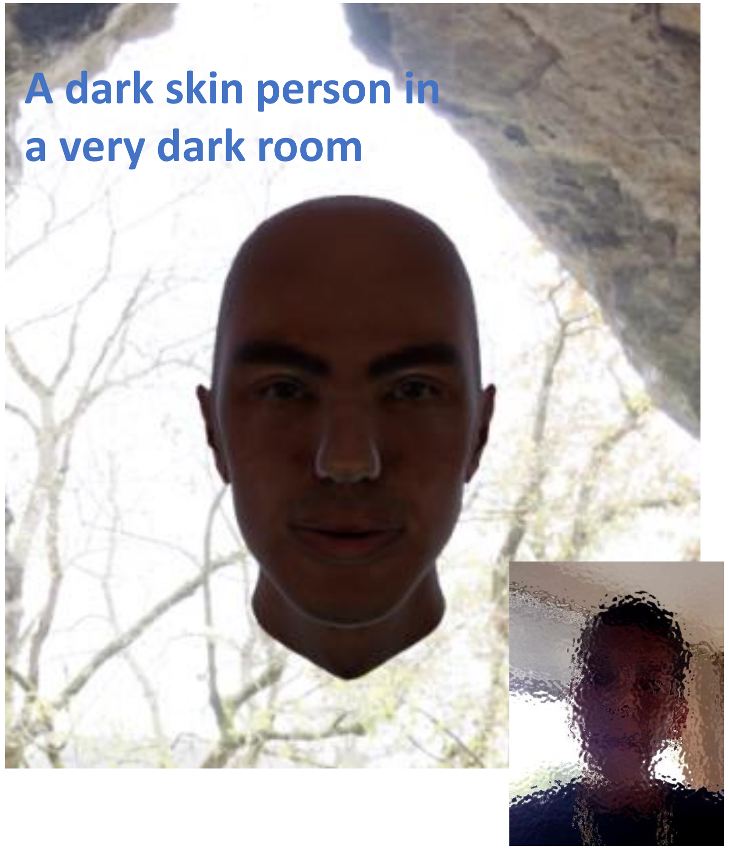
Pick focal length

Create an identity

Create a capture instance

Facial quality – Exposure estimator

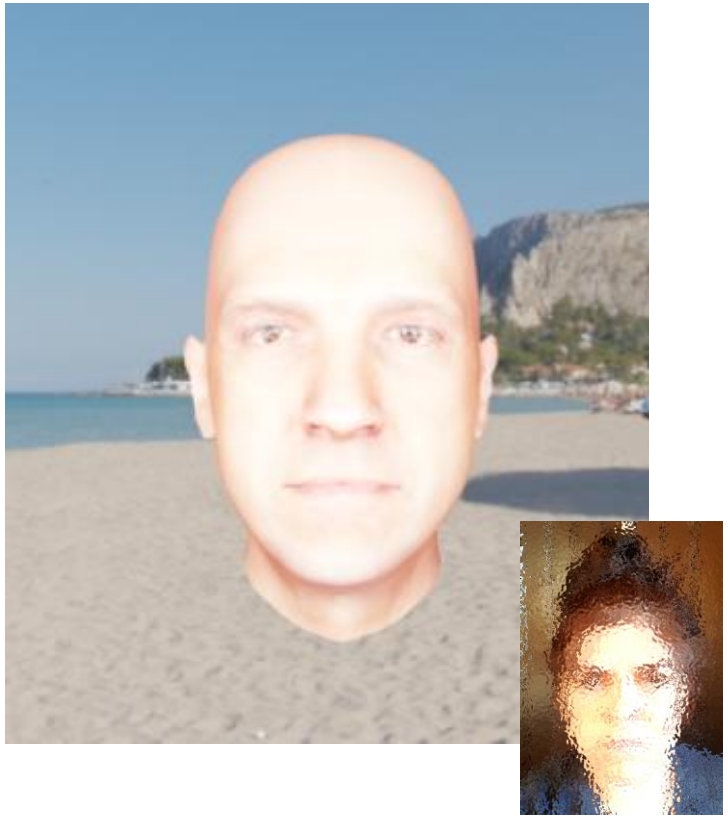
Underexposed



A dark skin person in a very dark room

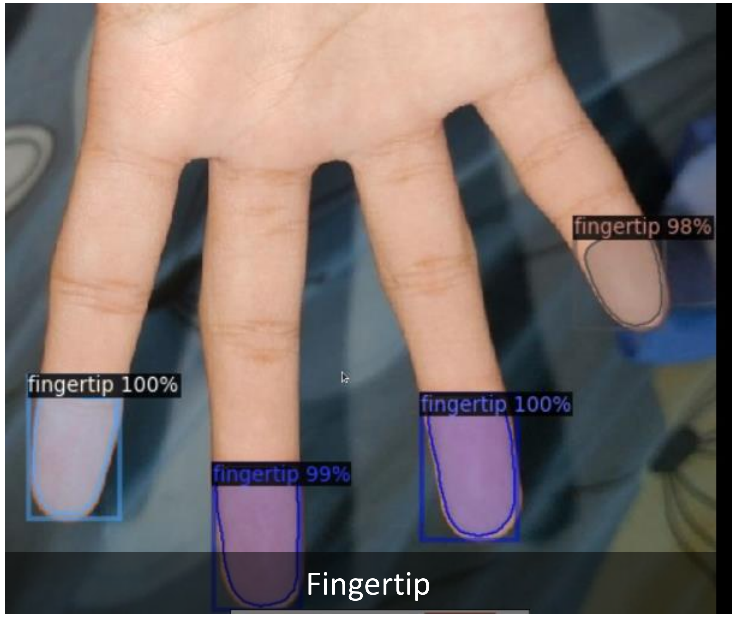
Normal

Overexposed

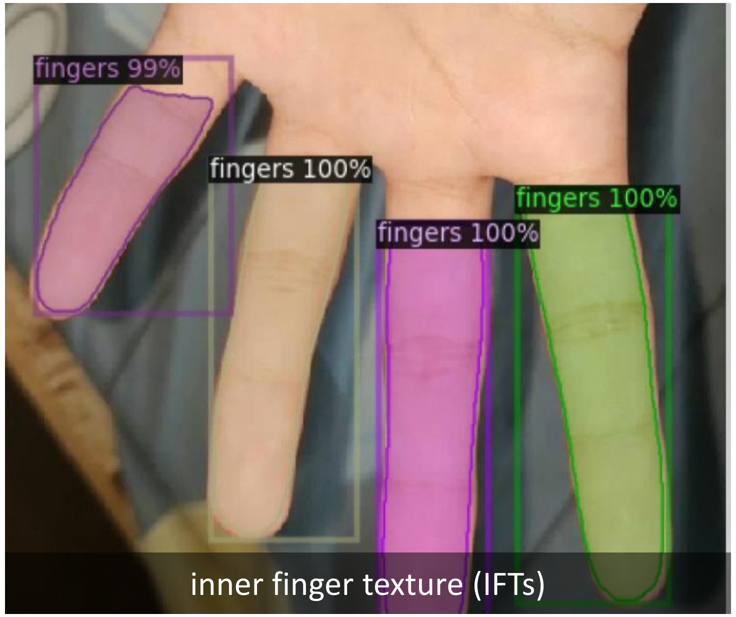


- Challenges: Very few face images have under- or over-exposure
- Generate synthetic images to complement the small data set with real samples
- Train a convolution neural network

Contactless biometrics (R&D)



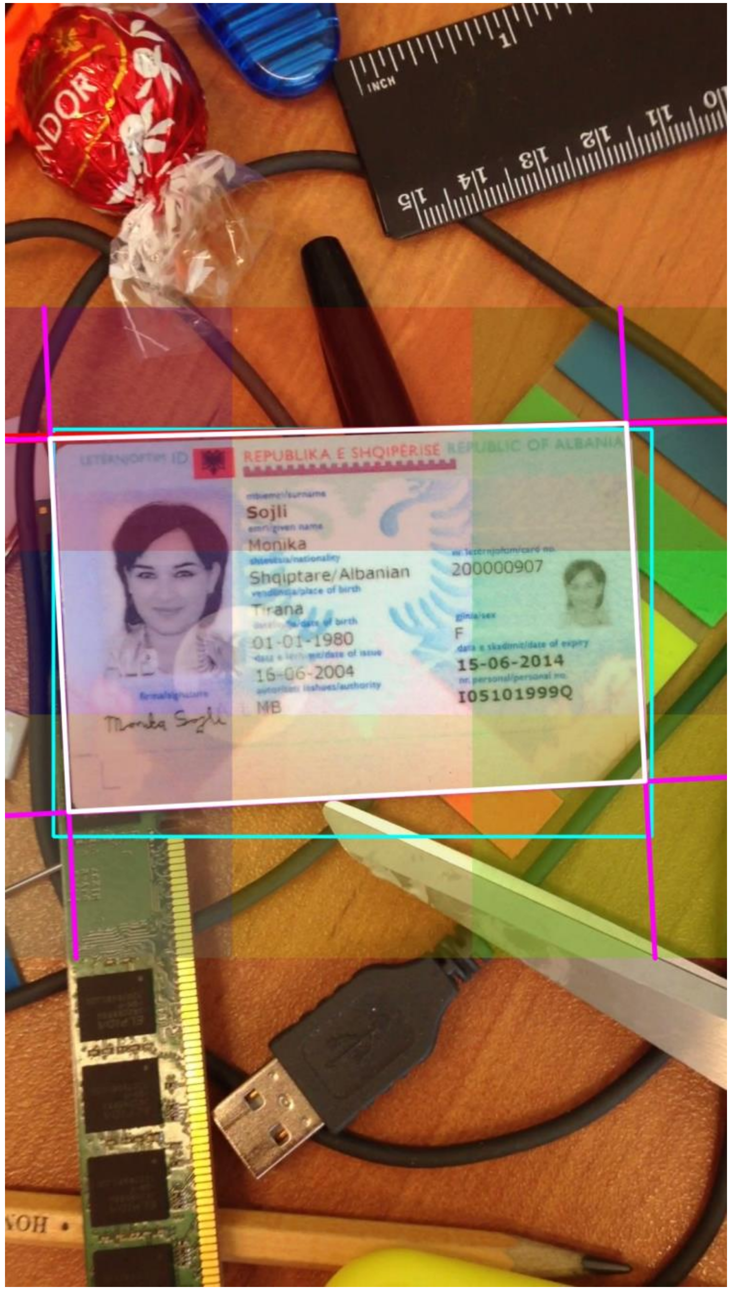
Fingertip



inner finger texture (IFTs)

- Why explore alternative contactless biometrics?
  - More hygienic, privacy concerns
- What are the challenges?
  - Easy on camera, reliable detection and segmentation, high accuracy and user acceptance & ease of use

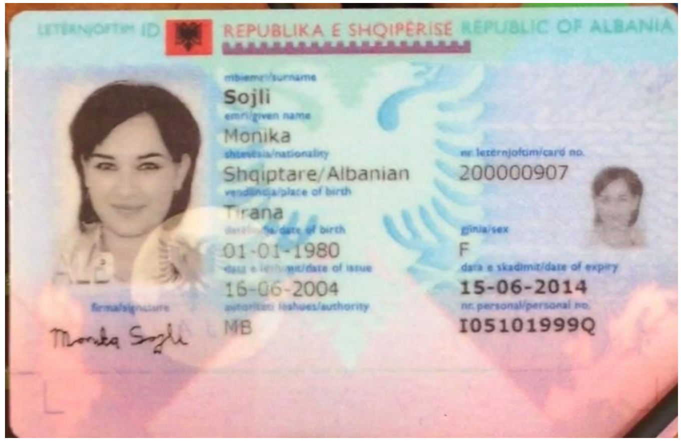
Face presentation attack | Synthetic data generation | Contactless biometrics | Document segmentation | Biometric cryptosystems



Document processing

Why document processing?

- Faster OCR result, crop ID photo for comparison with selfie, authenticity check



LETËRNJOFTIM ID REPUBLIKA E SHQIPËRISË REPUBLIC OF ALBANIA

mbiemri/surname
Sojli
emri/given name
Monika
shtetësia/nationality
Shqiptare/Albanian
vendlindja/place of birth
Tirana
datëlindja/date of birth
01-01-1980
data e lëshimit/date of issue
16-06-2004
autoriteti lëshues/authority
MB
nr. letërnjoftimi/card no.
200000907
gjinia/sex
F
data e skadimit/date of expiry
15-06-2014
nr. personal/personal no.
I05101999Q

"Flat lay" photo

- DCAR pipeline: Detect, Crop, Align, Rotate
- Document segmentation was successful with average IoU of 0.954

Face presentation attack | Synthetic data generation | Contactless biometrics | Document segmentation | Biometric cryptosystems

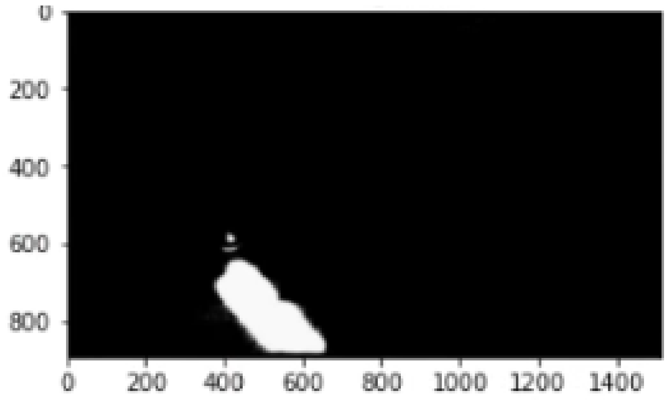
Challenges

Shadow (blurred)

Finger occlusion, two pages

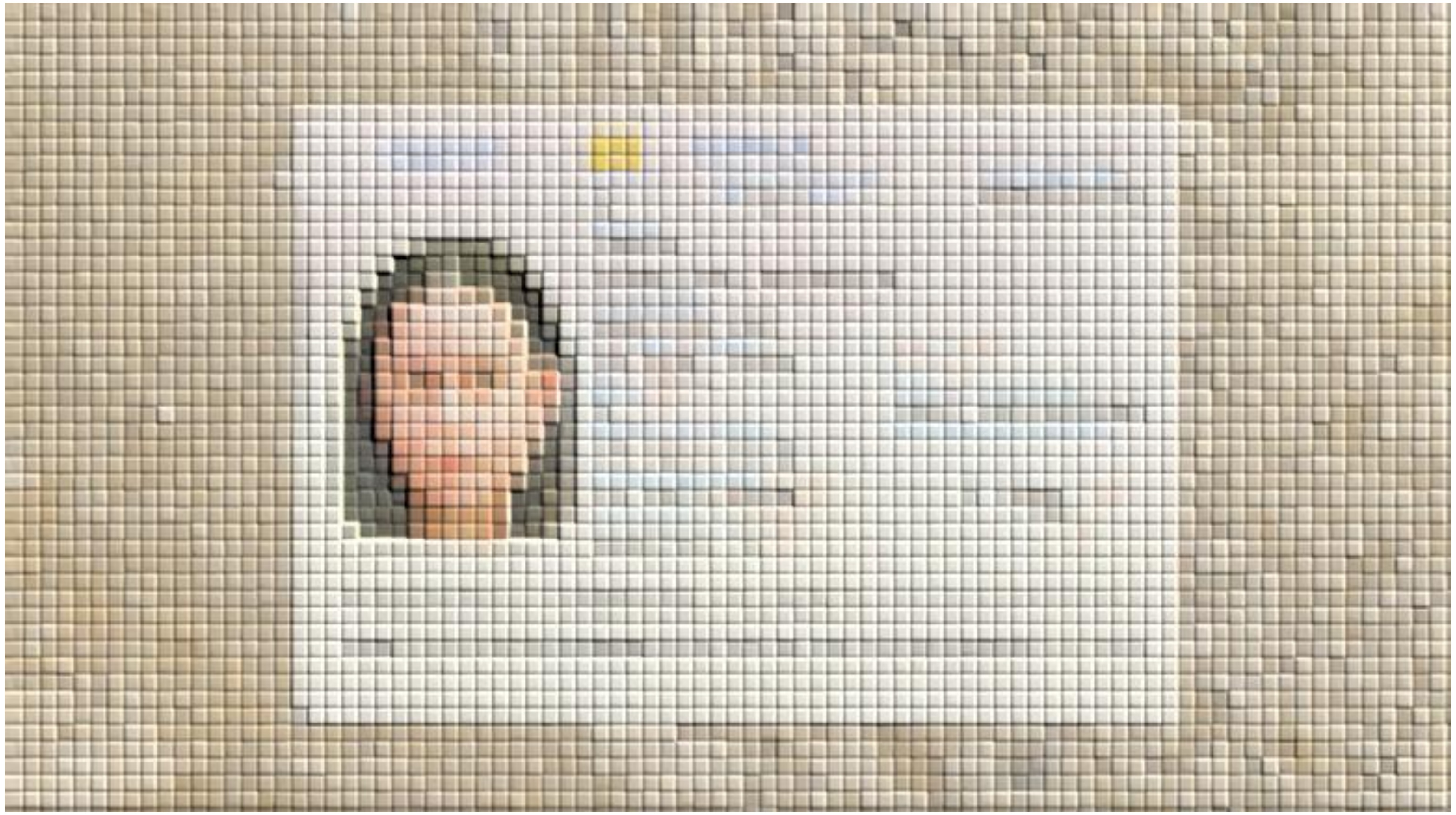
Wrong orientation

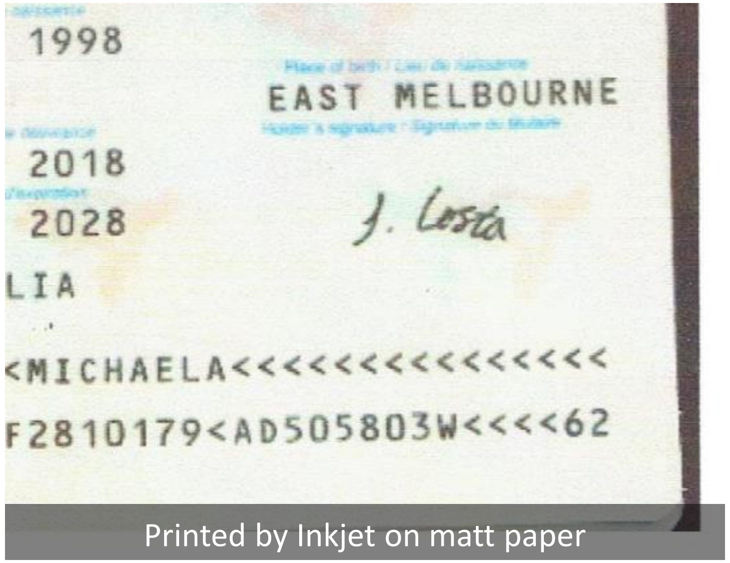
Glare



Glare mask

Document quality assessment

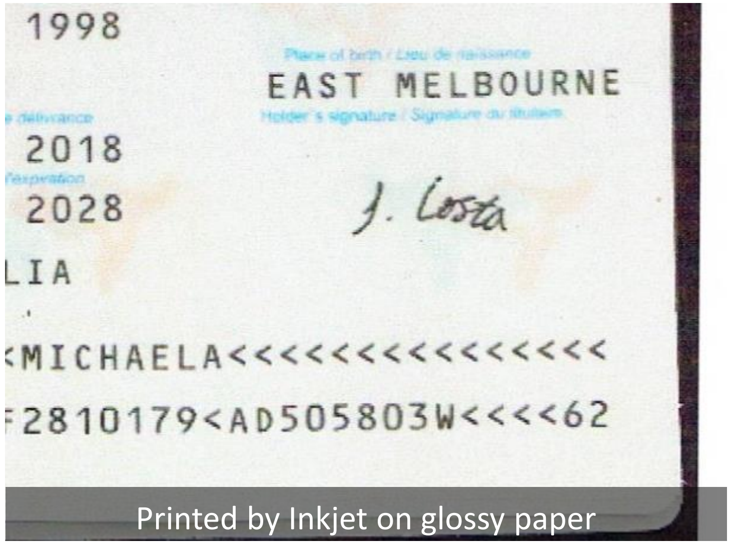




naissance
1998
Place of birth / Lieu de naissance
EAST MELBOURNE
e délivrance
Holder's signature / Signature du titulaire
2018
d'expiration
2028
J. Costa
LIA
<MICHAELA<<<<<<<<<<<<<<<
F2810179<AD505803W<<<<62

Printed by Inkjet on matt paper

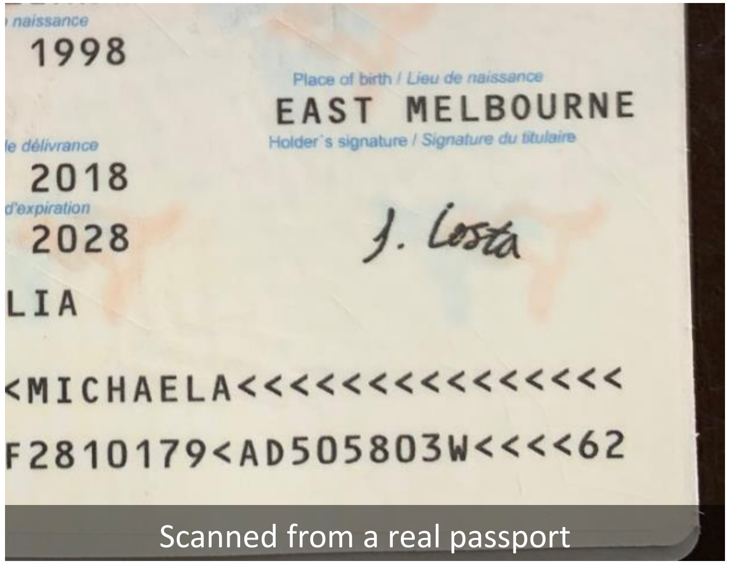
Diffused fonts



1998
Place of birth / Lieu de naissance
EAST MELBOURNE
e délivrance
Holder's signature / Signature du titulaire
2018
d'expiration
2028
J. Costa
LIA
<MICHAELA<<<<<<<<<<<<<<<
F2810179<AD505803W<<<<62

Printed by Inkjet on glossy paper

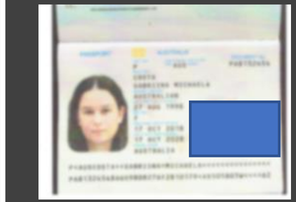
Faded colour



naissance
1998
Place of birth / Lieu de naissance
EAST MELBOURNE
le délivrance
Holder's signature / Signature du titulaire
2018
d'expiration
2028
J. Costa
LIA
<MICHAELA<<<<<<<<<<<<<<<
F2810179<AD505803W<<<<62

Scanned from a real passport

Original



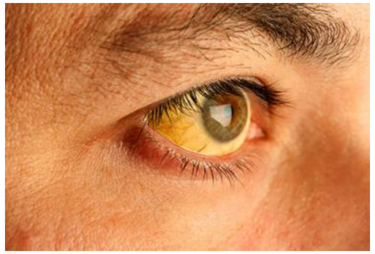
Document presentation attacks

Why privacy-preserving biometrics

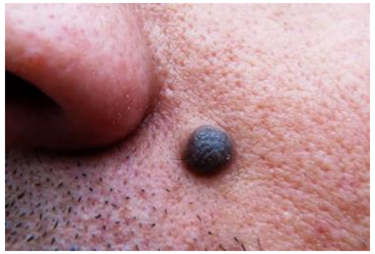
What does a face (image) tell you about the person?

https://thispersondoesnotexist.com

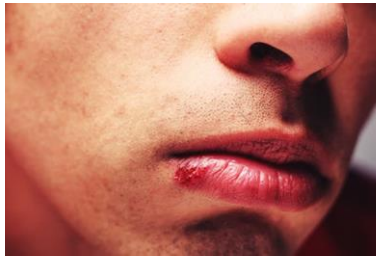
What Does Your Face Say About Your Health?



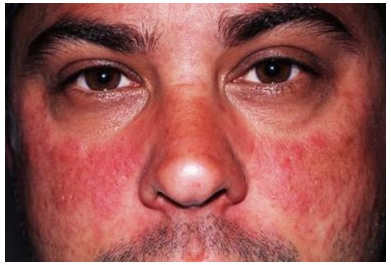
Jaundice



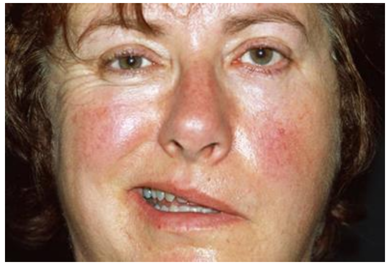
Moles



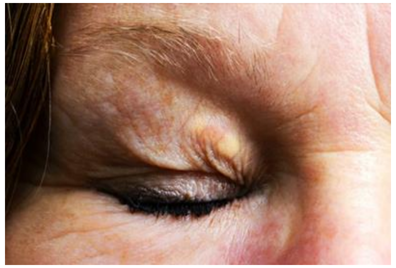
Sores



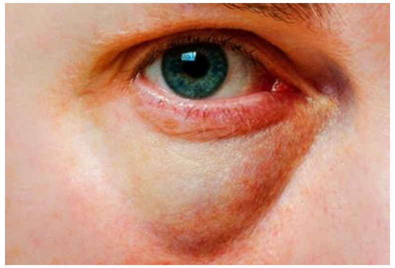
Butterfly rash



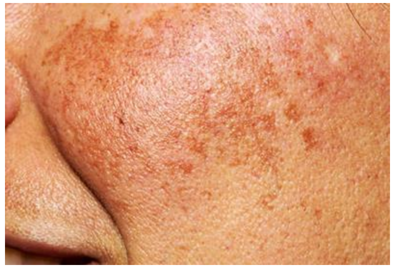
Can’t Move One Side of Your Face



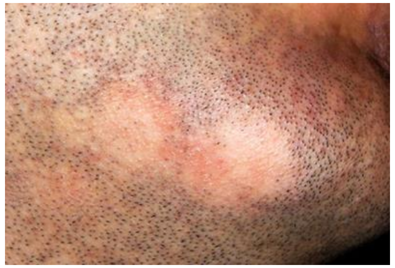
Yellow Spots on Your Eyelids



Puffy Eyes



Melasma



Hair loss

https://www.webmd.com/skin-problems-and-treatments/ss/slideshow-face-your-health

Why privacy-preserved biometrics?

Function creep

Verification database is repurposed for identification

Identity theft

Stolen database sold on the dark web

Reveal of sensitive information

(race, religion, sexual orientation)

Large-scale surveillance

(Rogue governments)

Biometrics as unique identifiers for linking databases

Face presentation attack

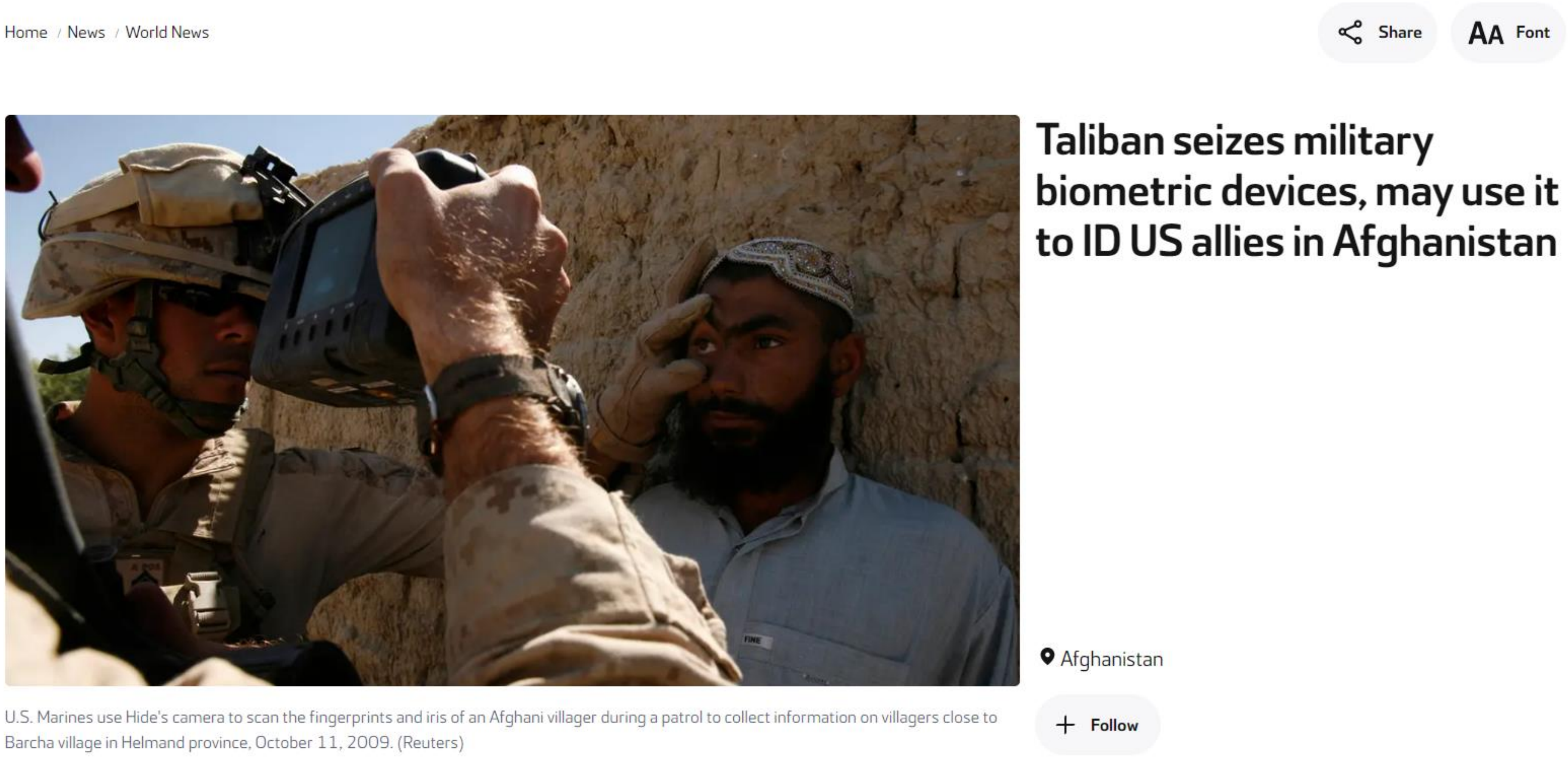
Synthetic data generation

Contactless biometrics

Document segmentation

Biometric cryptosystems

Compromised biometric devices?

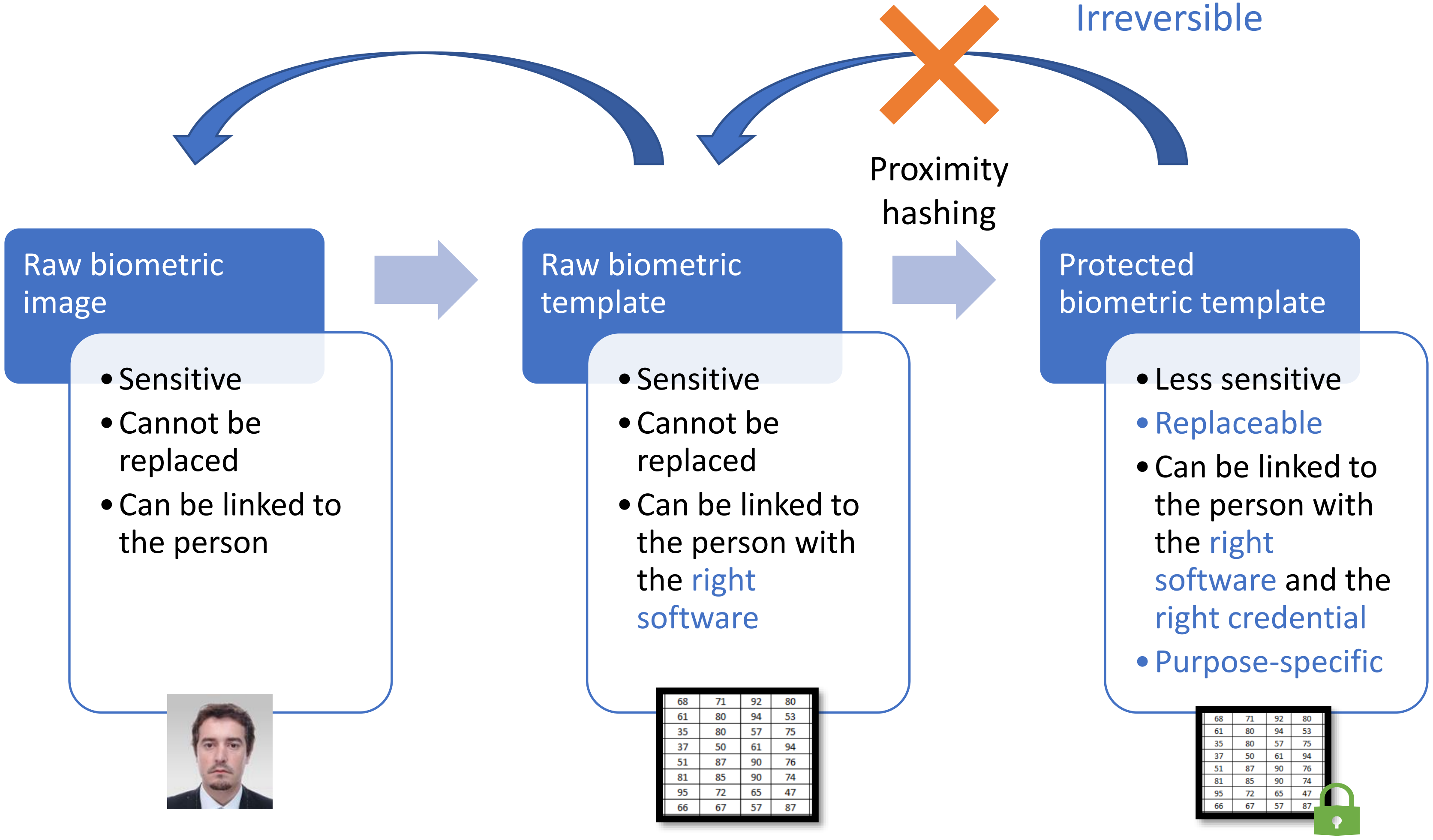


Taliban seizes military biometric devices, may use it to ID US allies in Afghanistan

U.S. Marines use Hide's camera to scan the fingerprints and iris of an Afghani villager during a patrol to collect information on villagers close to Barcha village in Helmand province, October 11, 2009. (Reuters)

https://english.alarabiya.net/News/world/2021/08/19/Taliban-seizes-military-biometric-devices-may-use-it-to-ID-US-allies-in-Afghanistan

Privacy-preserved biometrics



Irreversible

Proximity hashing

Raw biometric image

- Sensitive
- Cannot be replaced
- Can be linked to the person

Raw biometric template

- Sensitive
- Cannot be replaced
- Can be linked to the person with the right software

68	71	92	80
61	80	94	53
35	80	57	75
37	50	61	94
51	87	90	76
81	85	90	74
95	72	65	47
66	67	57	87

Protected biometric template

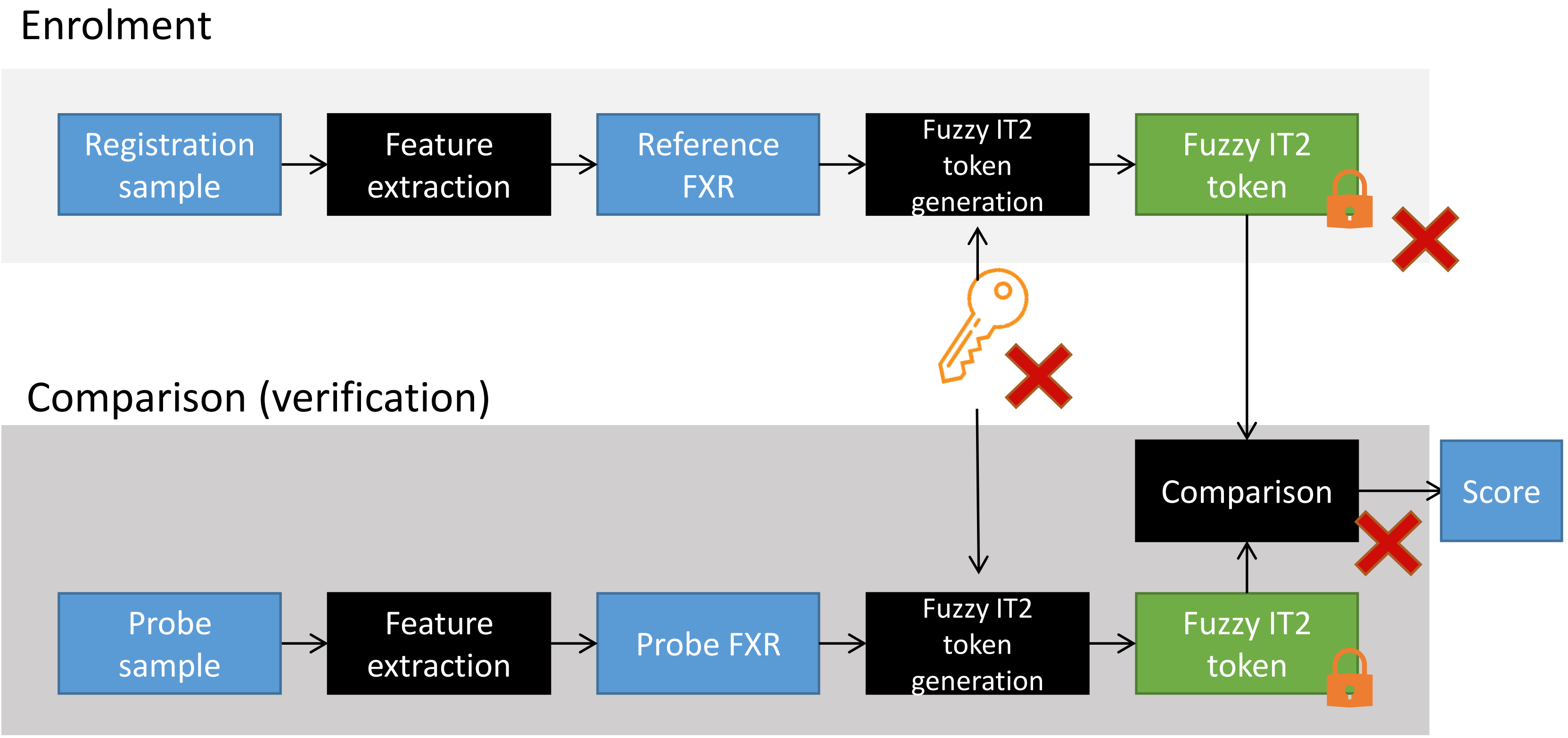
- Less sensitive
- Replaceable
- Can be linked to the person with the right software and the right credential
- Purpose-specific

68	71	92	80
61	80	94	53
35	80	57	75
37	50	61	94
51	87	90	76
81	85	90	74
95	72	65	47
66	67	57	87

Irreversibly transformed identity token or IT2

Why comparison in the IT2 domain is more secure?

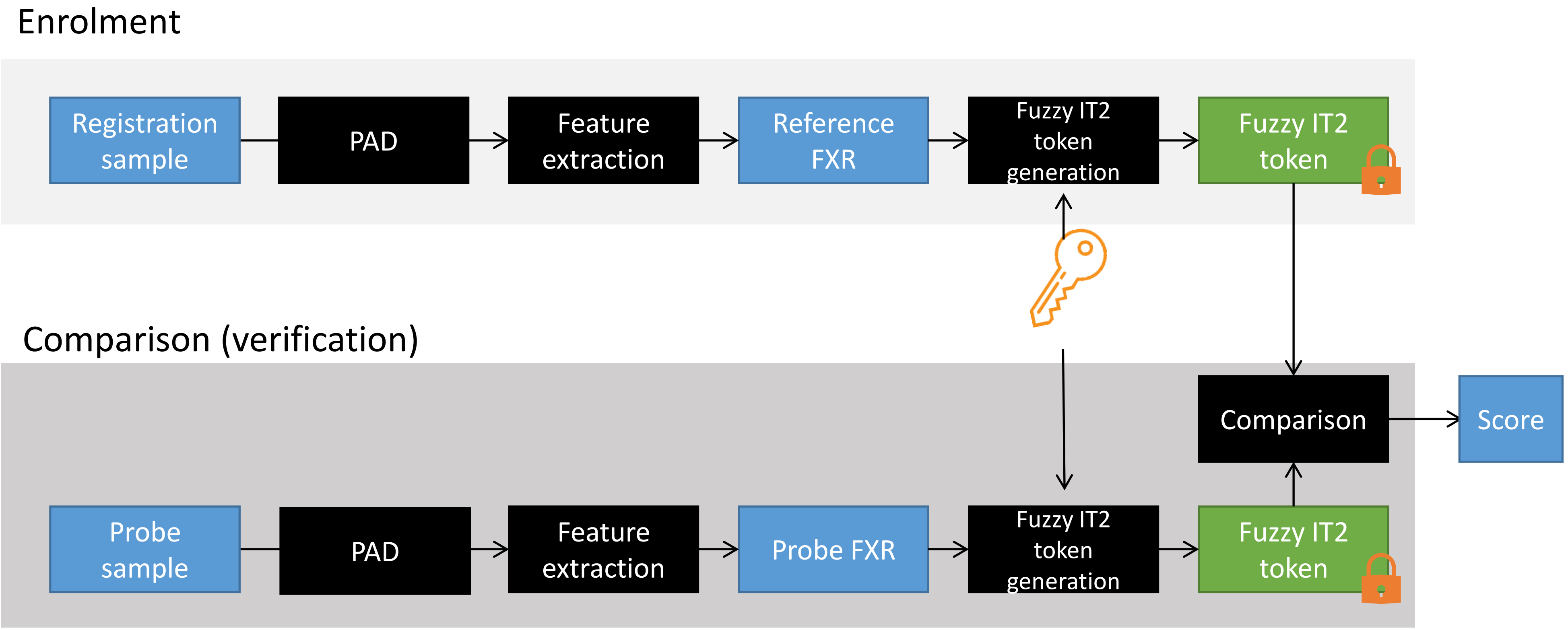
Multiple points of revocation (✗)



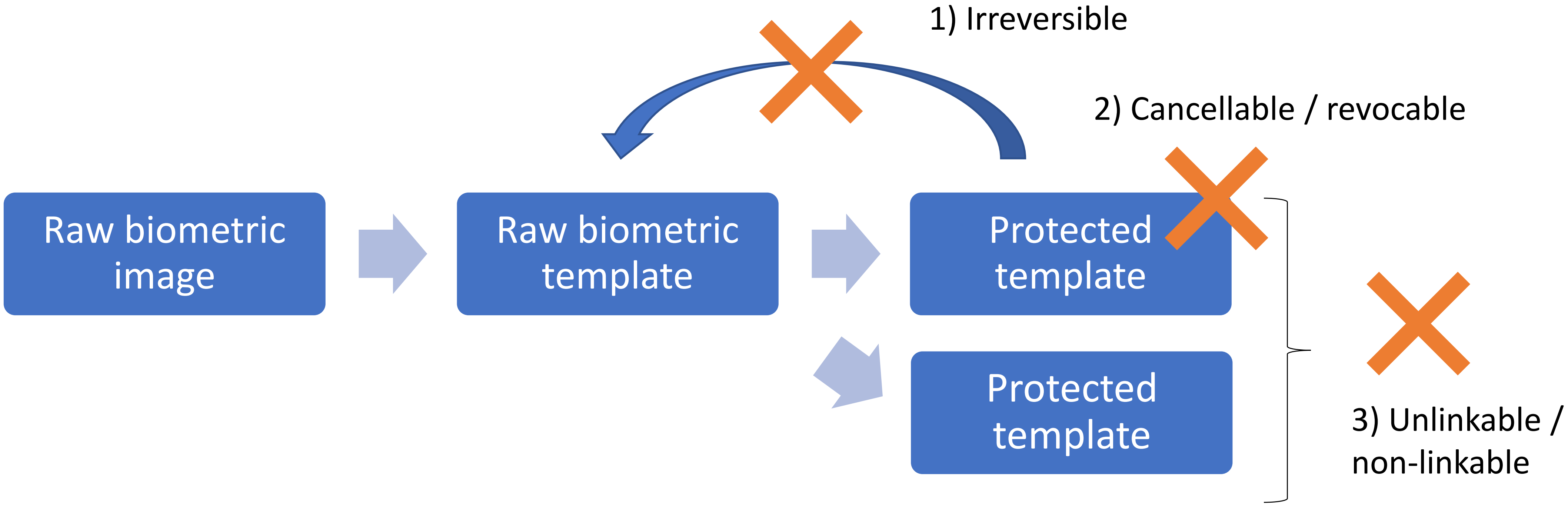
Because the IT2 algorithm is not based on classic cryptography, it is also considered quantum-secure today.

Handling presentation attacks using PAD

(Presentation Attack Detection)



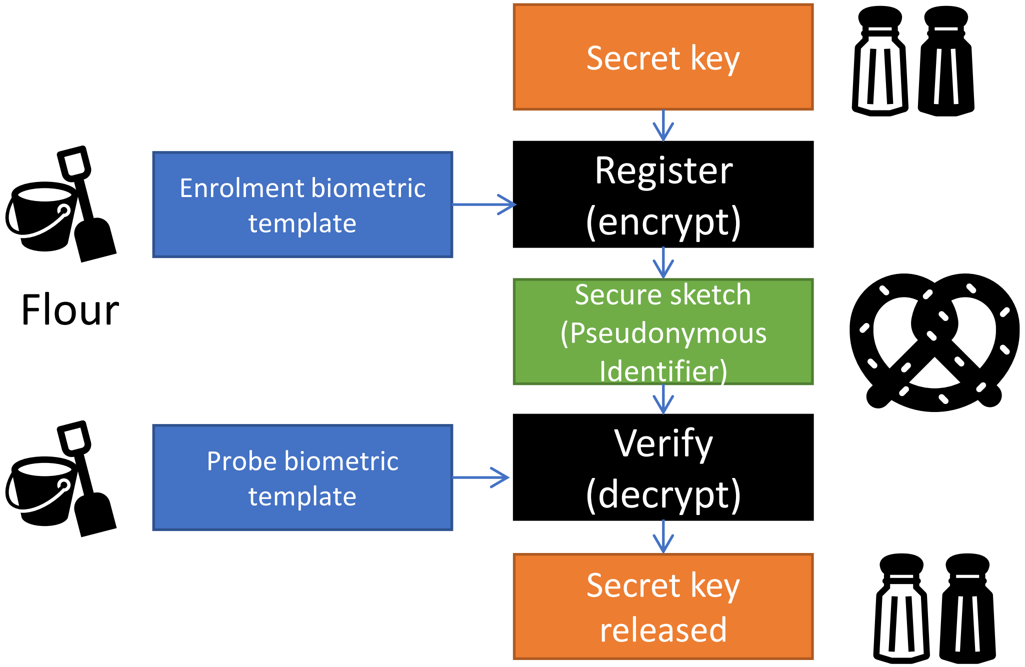
What does a *secure* template (IT2 token) mean?



ISO/IEC 30136:2018 Information technology — Performance testing of biometric template protection schemes

Privacy-preserving biometrics

Biometric cryptosystem

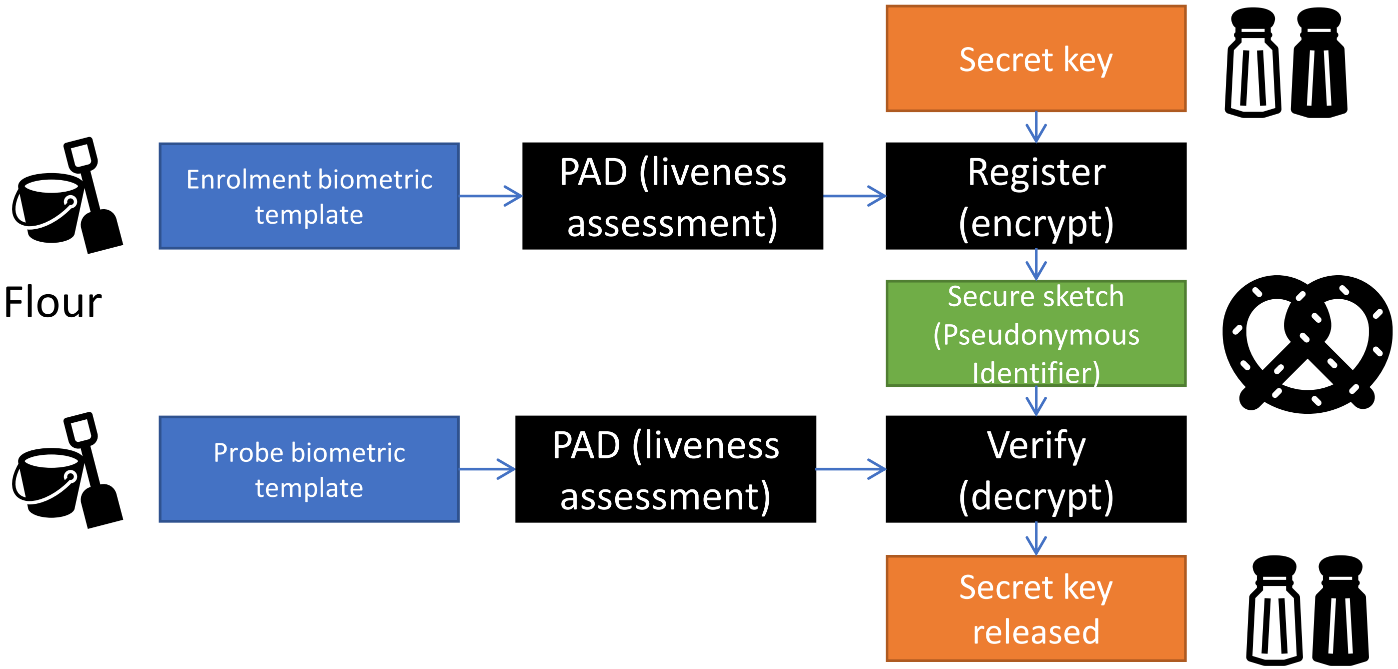


Legend

Biometric data (not saved)
Secret data
Process
Public data

Only the same person can retrieve the secret key

Biometric cryptosystem

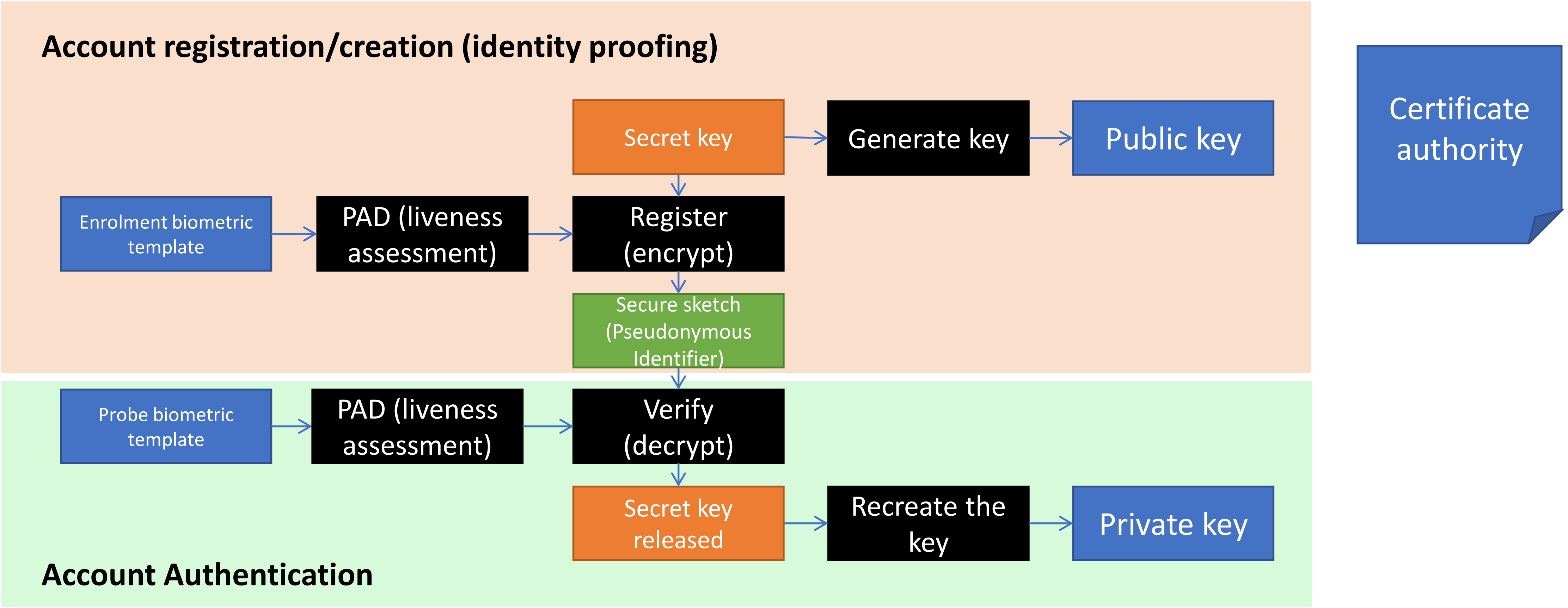

Secret key
Enrolment biometric template
PAD (liveness assessment)
Register (encrypt)
Flour
Secure sketch (Pseudonymous Identifier)
Probe biometric template
PAD (liveness assessment)
Verify (decrypt)
Secret key released

Legend

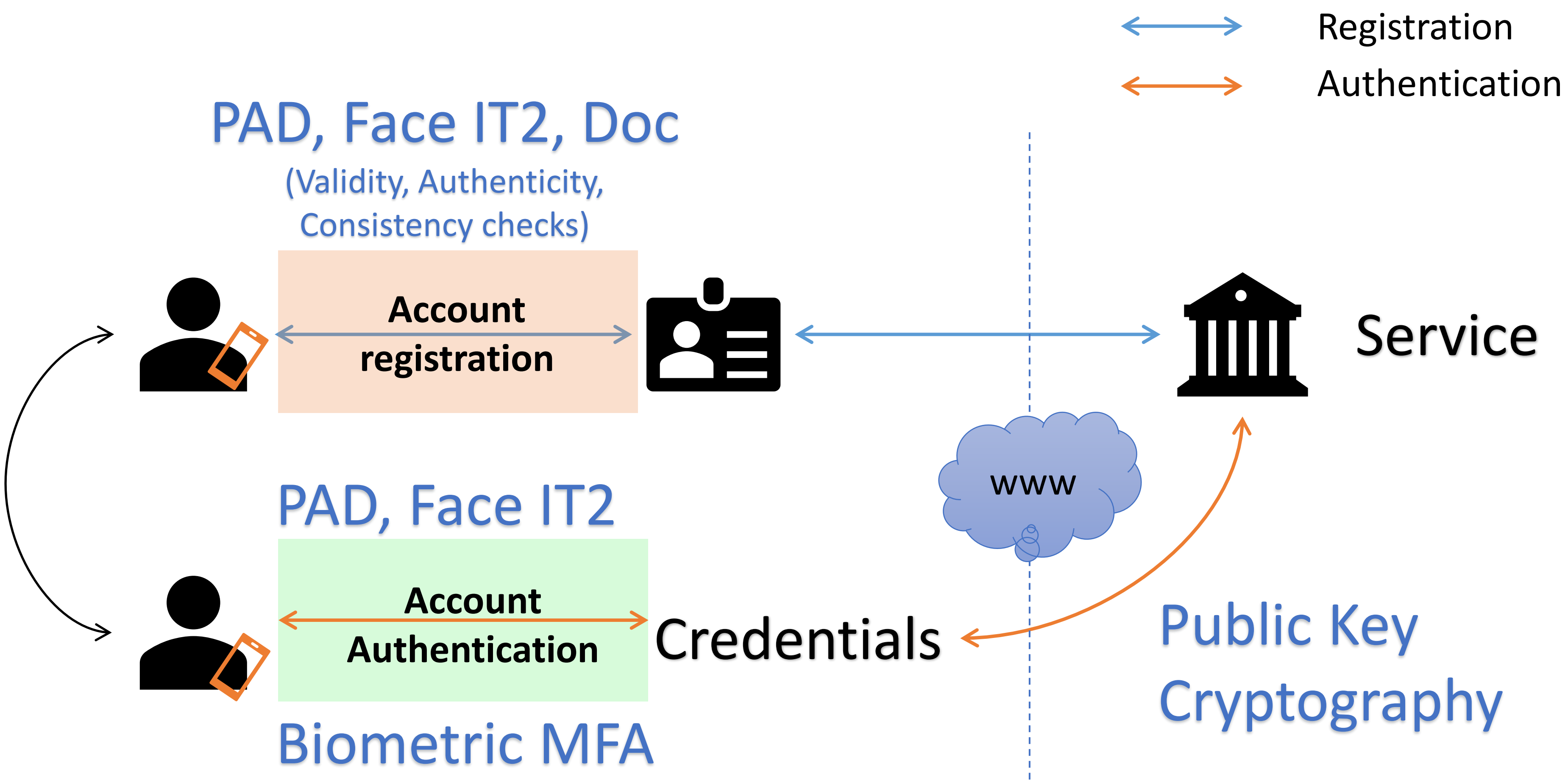
Biometric data (not saved)
Secret data
Process
Public data

Only the same, *live* person can retrieve the secret key

Stable IT2 (Biometric cryptosystem)



Summary



↔ Registration

↔ Authentication

PAD, Face IT2, Doc

(Validity, Authenticity, Consistency checks)

Account registration

Service

www

PAD, Face IT2

Account Authentication

Credentials

Public Key Cryptography

Biometric MFA

OTP to email or SMS + Face

Key message:

1. The relying party never stores or processes any biometric data
2. GDPR-compliant solution (biometric stays on device or remotely processed in cancellable format)
3. High binding and authentication assurance

Case studies

Use case 1: Fuzzy IT2 (cancellable biometrics)

Ad hoc ID Infrastructure for the humanitarian sector

An ad hoc identity infrastructure for the underserved and unbanked in Africa. The solution can perform 1:N deduplication, work offline, using compact Irreversibly Transformed Identity Tokens (or IT2) and run on consumer-grade Android smartphones.

Offline

With online data synchronisation

Consumer-grade device

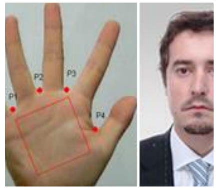
Reduce costs

Deduplication at scale

On-device deduplication

Compact token

Compact enough to display as QR



Multimodal

Increased population coverage

Purpose-limiting

GDPR compliant

Trust Stamp tokenized biometrics for Mastercard financial inclusion initiative | Biometric Update | 2021-09-17

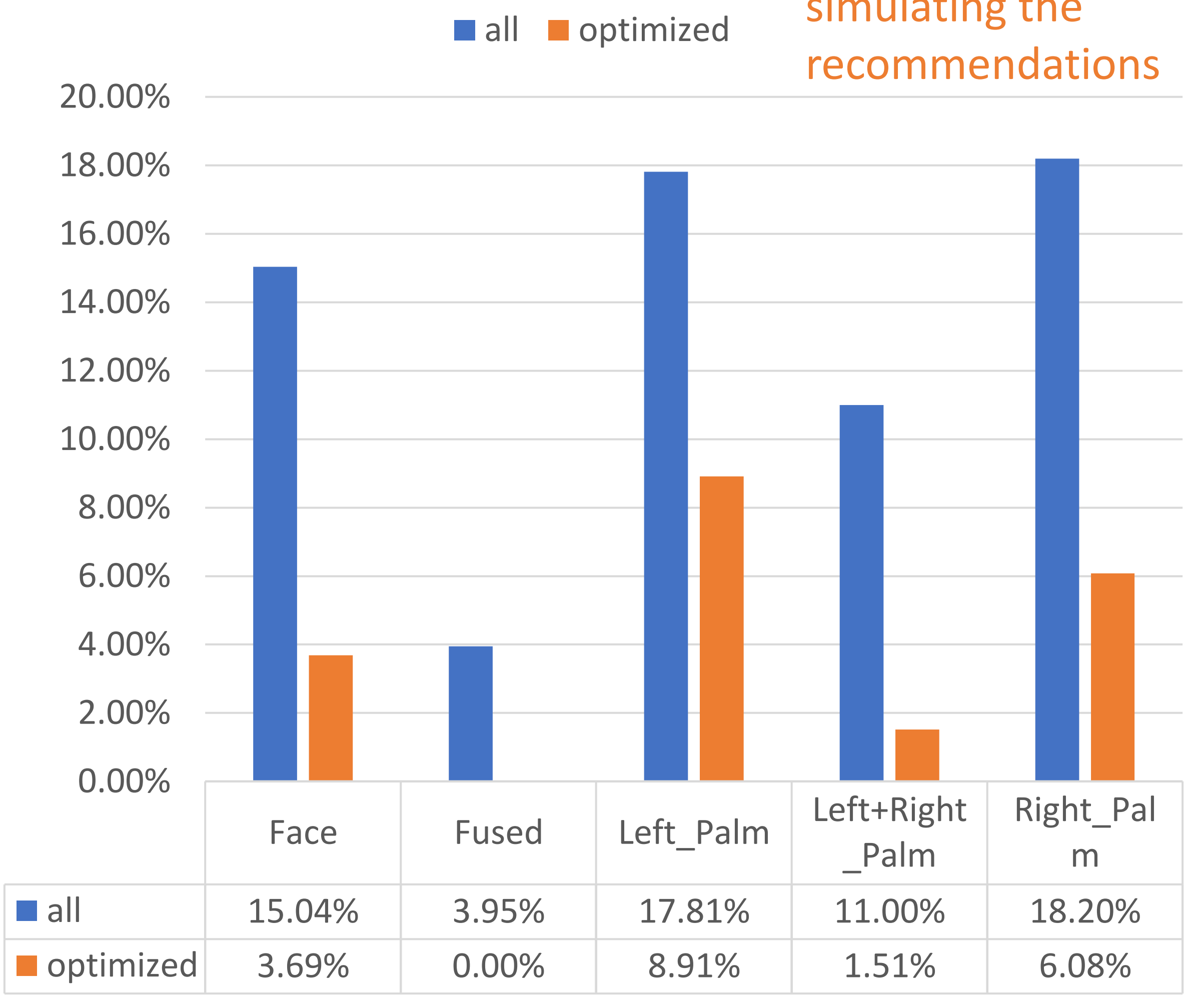
- Goal: increase access to financial services and government assistance for remote communities across Africa
- Project requirements:
    - Contactless biometrics – left and right palms and face (selfie)
    - Biometric data never leaves the device
    - All biometric templates are represented using Trust Stamp's Irreversibly Transformed Identity Token, or IT2 (privacy-preserved biometrics) which was delivered in the form of an Android SDK
    - Must support 1:1 and 1:N at scale on device
    - Must operate offline most of the time. The biometric gallery is synched to server when it has access to the Internet
    - Affordable Android devices

Recommendations

- Face – flashlight off, indoor well-lit, outdoor shade, take off glasses and hat
  - Although no facial hair is better, it requires people to shave – this may not be culturally acceptable
- Palmprint – indoor well-lit, unaltered, outdoor direct sun (because the gain in improved true acceptance out weights false acceptance)

If we were to follow the above recommendation, the identification EER would reduce by 50%, from ~4% to 0%.

Performance simulating the recommendations



	Face	Fused	Left_Palm	Left+Right_Palm	Right_Palm
🟦 all	15.04%	3.95%	17.81%	11.00%	18.20%
🟧 optimized	3.69%	0.00%	8.91%	1.51%	6.08%

IFPC 2022 Conference Presentations and Videos | NIST |Industry Outlook track:
Modelling the Odds of False Acceptance and False Rejection of a Privacy-Preserved Multimodal System Involving Face Modality [video] [presentation]

Summary

- We have developed a statistical method to identify capture conditions that are favourable during registration.
- The method only observes the fused score of a multimodal biometric system in the privacy preserved domain (IT2)
- The covariates found form the basis of a lighting-based or a full intervention
- The interventions were validated in the identification setting
- Future work:
  - Apply the same methodology to biometric sample quality (quality measures)
  - Apply it to analyse performance differentials