Introduction to Authentication using Behavioral Biometrics

Jonathan Liebers[1], Uwe Gruenefeld[1], Daniel Buschek[2], Florian Alt[3], Stefan Schneegass[1]

[1]University of Duisburg-Essen, [2]University of Bayreuth, [3]Bundeswehr University Munich

Instructors

Jonathan Liebers, University of Duisburg-Essen
jonathan.liebers@uni-due.de

Uwe Gruenefeld, Dr., University of Duisburg-Essen
uwe.gruenefeld@uni-due.de

Daniel Buschek, Prof. Dr., University of Bayreuth
daniel.buschek@uni-bayreuth.de

Florian Alt, Prof. Dr., University of the Bundeswehr Munich
florian.alt@unibw.de

Stefan Schneegass, Prof. Dr., University of Duisburg-Essen
stefan.schneegass@uni-due.de

Introduction of Participants

Please state …

- Your name and affiliation
- Your background
- Your interest in behavioral biometrics

Motivation

Omnipresent Cyber Attacks

REUTERS

World Business Markets Breakingviews Video More

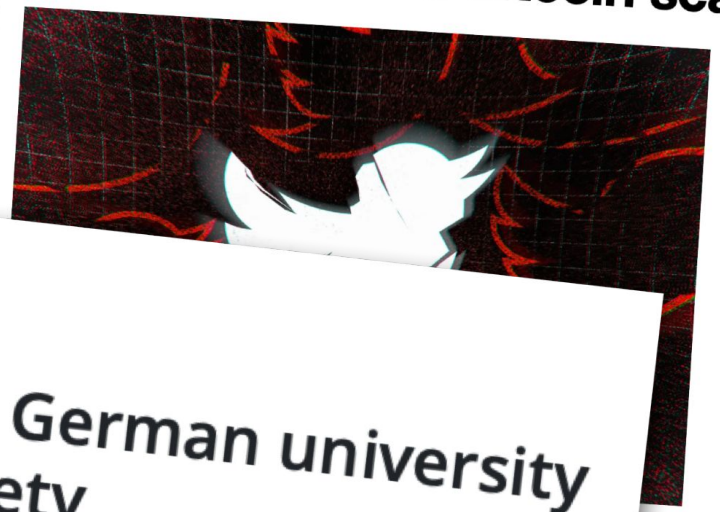
TECHNOLOGY NEWS MAY 15, 2015 / 6:07 PM / UPDATED 7 YEARS AGO

Unknown hackers attack German parliament's data network

By Reuters Staff

TECH / TWITTER / ELON MUSK

Twitter's massive attack: What we know after Apple, Biden, Obama, Musk, and others tweeted a bitcoin scam



/ Update: Wednesday's Twitter attack is now being investigated by numerous law enforcement agencies

By NICK STATT / @nickstatt

Updated Jul 17, 2020, 1:41 AM GMT+2 | 0 Comments

als were
read and
service of
ly $120,000.

Ransomware, Cybercrime, Data security

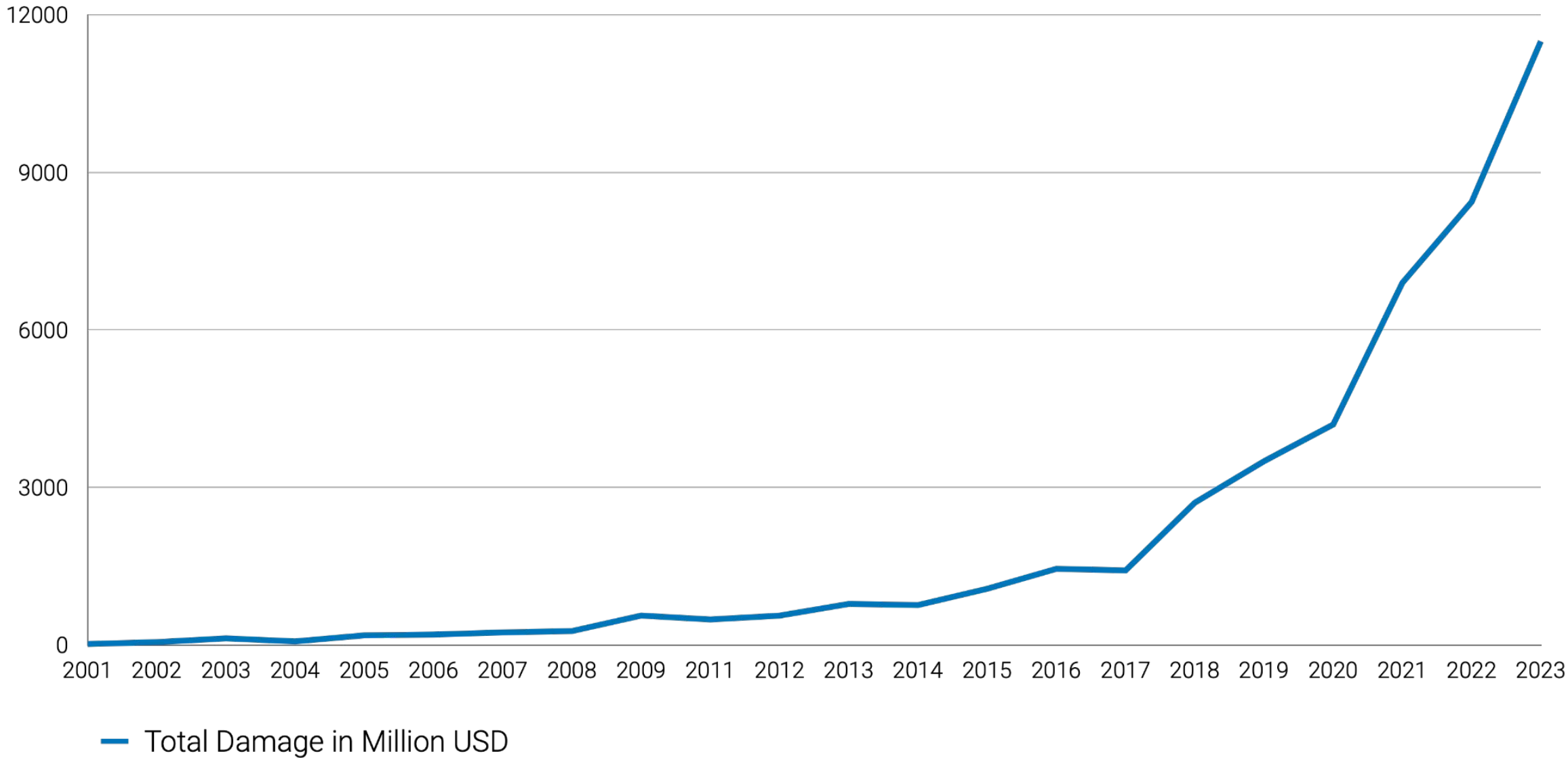
Cyberattack against German university claimed by Vice Society

SC Staff January 17, 2023

University of Duisburg-Essen in Germany was hit by a cyberattack in November that has been claimed by the Vice Society ransomware operation, which has also exposed data allegedly stolen from the university, including sensitive details involving its operations, students, and employees, reports BleepingComputer.

UDE, which is having its IT infrastructure overhauled as a result of t
emphasized that it will not pay the ra

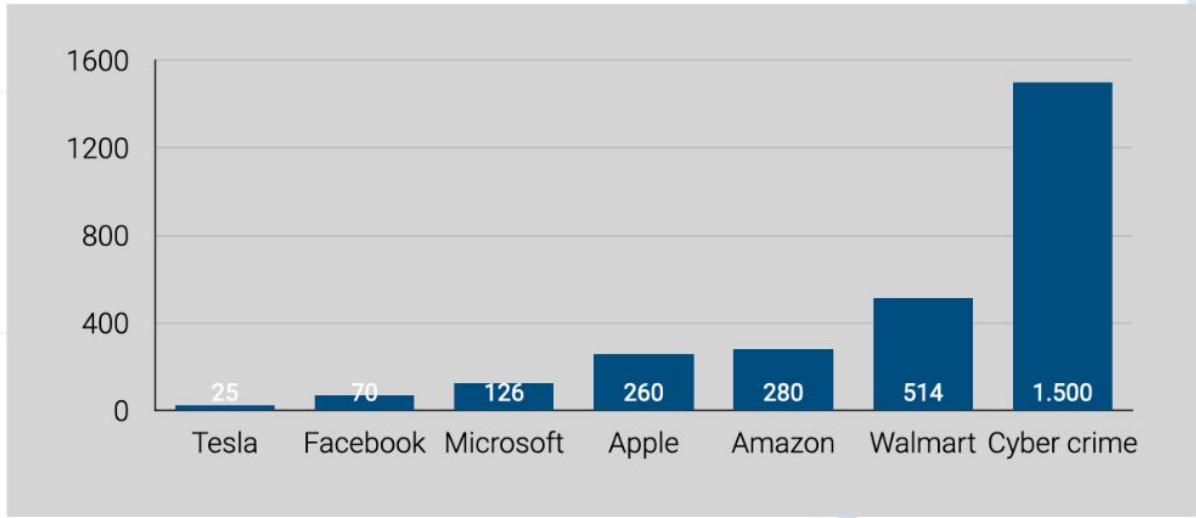
Amount of monetary damage caused by reported cyber crime to the ICB from 2001 to 2021 (in Million USD)



Amount of monetary damage caused by reported cyber crime to the ICB from 2001 to 2021 (in Million USD)

Attacks are no longer random

- Highly professional cybercrime industry
- Cost/risk benefit calculation
- Novel business models (Ransomware-as-a-Service)

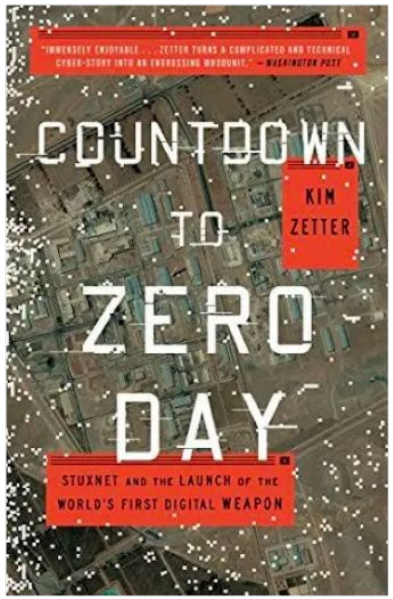


Source: https://www.statista.com/chart/28878/expected-cost-of-cybercrime-until-2027/

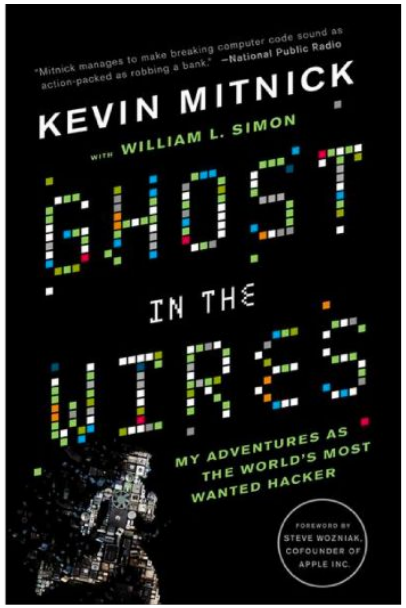
HACKERS DON'T BREAK IN —

THEY LOG IN!

TK Keanini (CISCO)

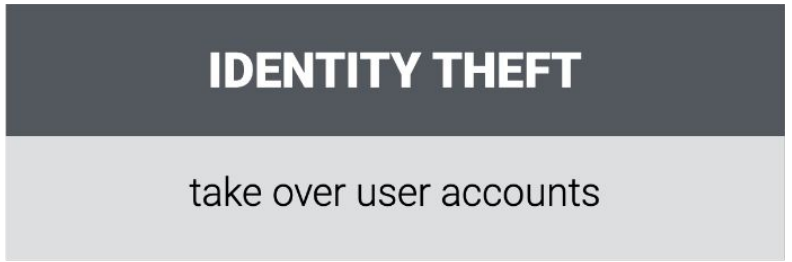


Kim Zetter. Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon



Kevin D. Mitnick, William L. Simon. Ghost in the Wires — My Adventures as the World's Most Wanted Hacker.

Human-Centered Attacks



Human-Centered Attacks

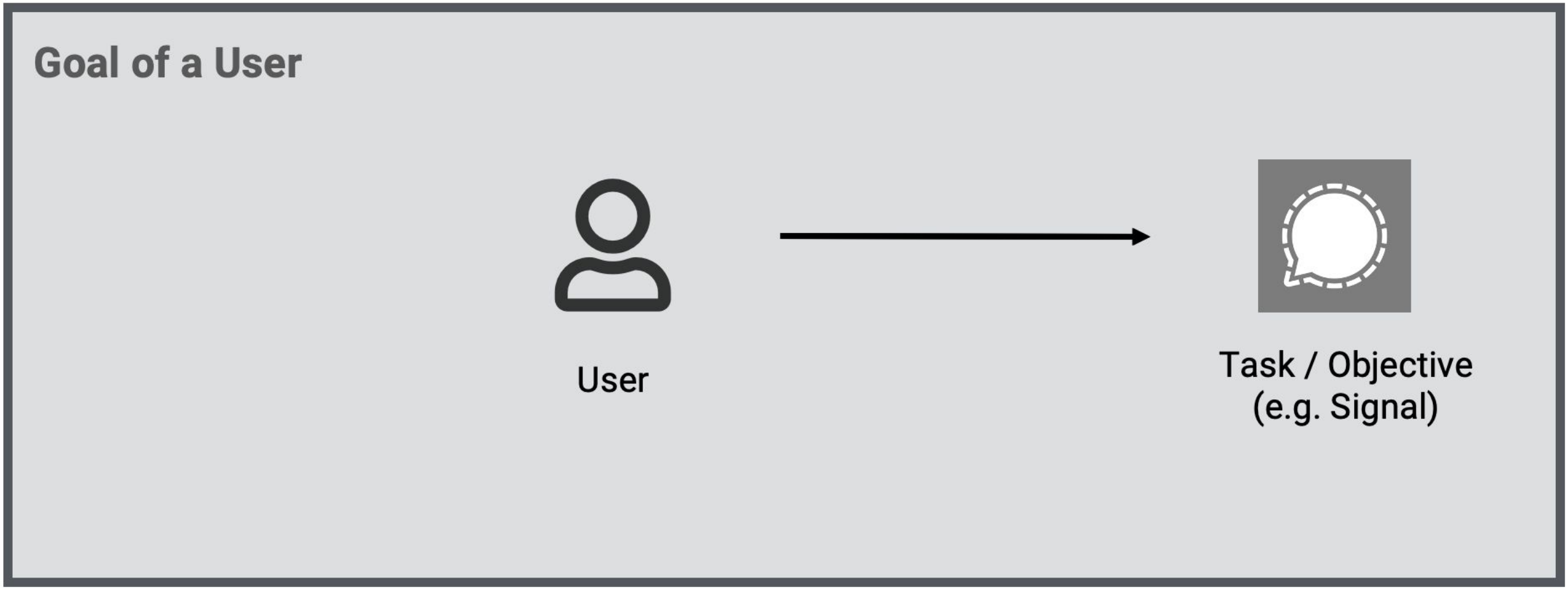
IDENTITY THEFT

Obtaining credentials of user accounts

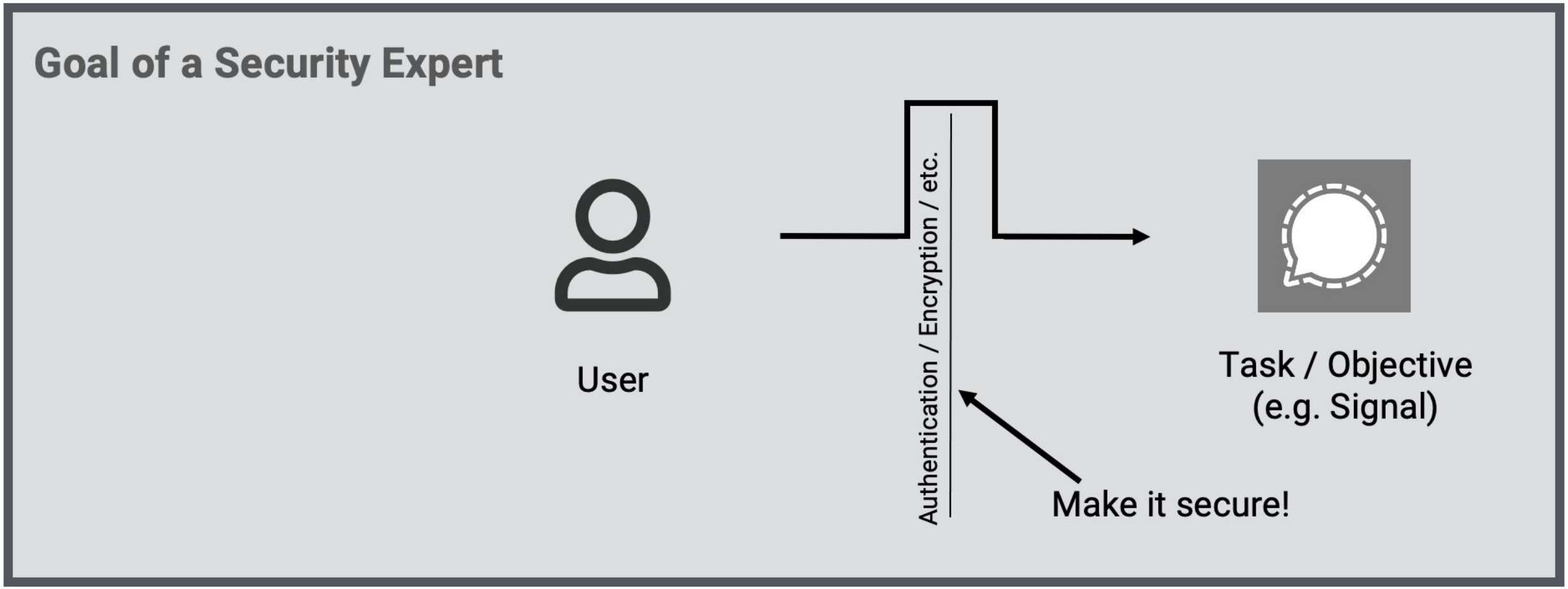
- **Guessing Attacks**
  - Brute Force
  - Credential Stuffing
- **Observation Attacks**
  - Shoulder Surfing
  - Keyloggers
  - Sniffers
- **Social Engineering Attacks**
  - (Spear) Phishing
  - Vishing
- **Reconstruction Attacks**
  - Smudges
  - Heat Traces

In particular knowledge-based authentication subject to these attacks!

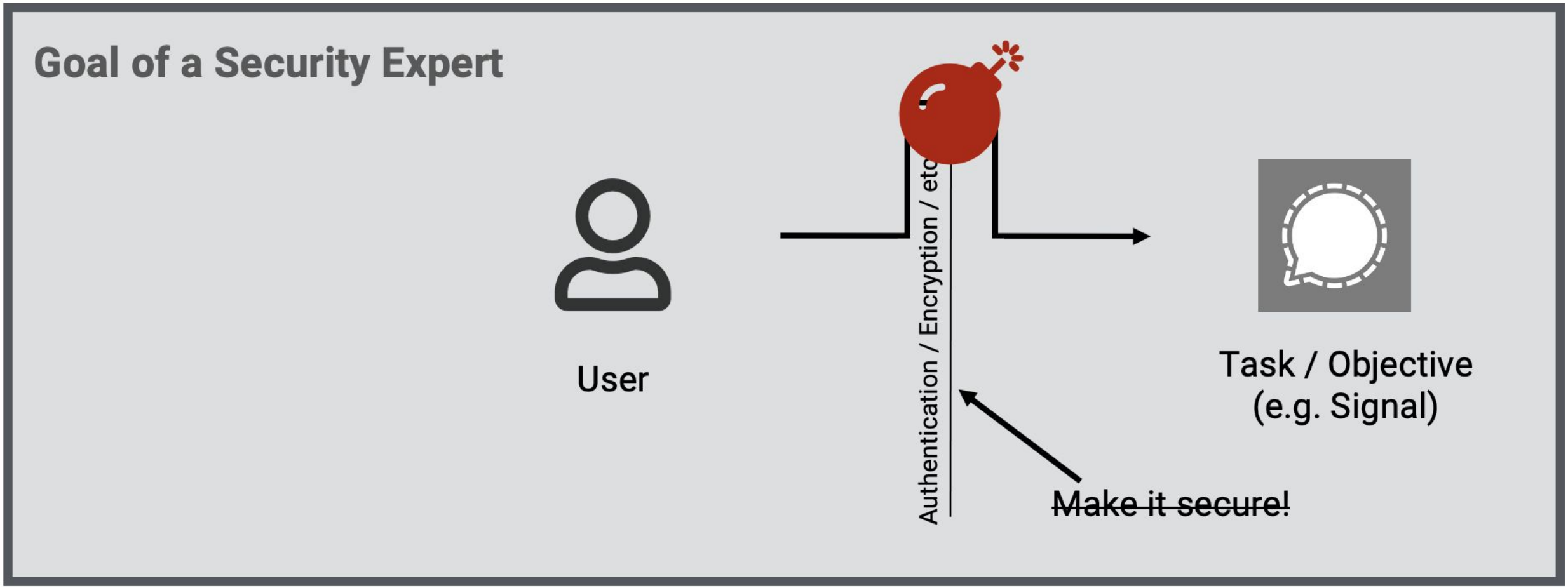

12:50
intel
CORE
intel
iRIS



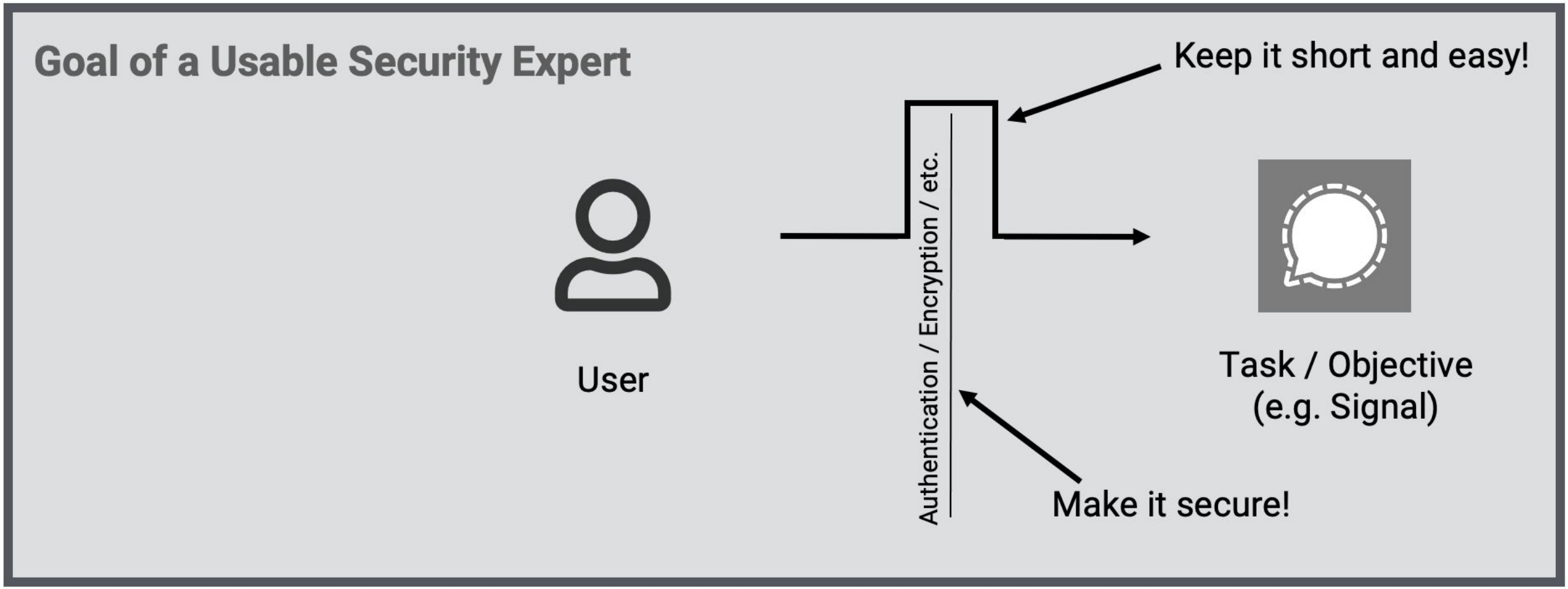
SECURITY AS SECONDARY TASK



SECURITY AS SECONDARY TASK

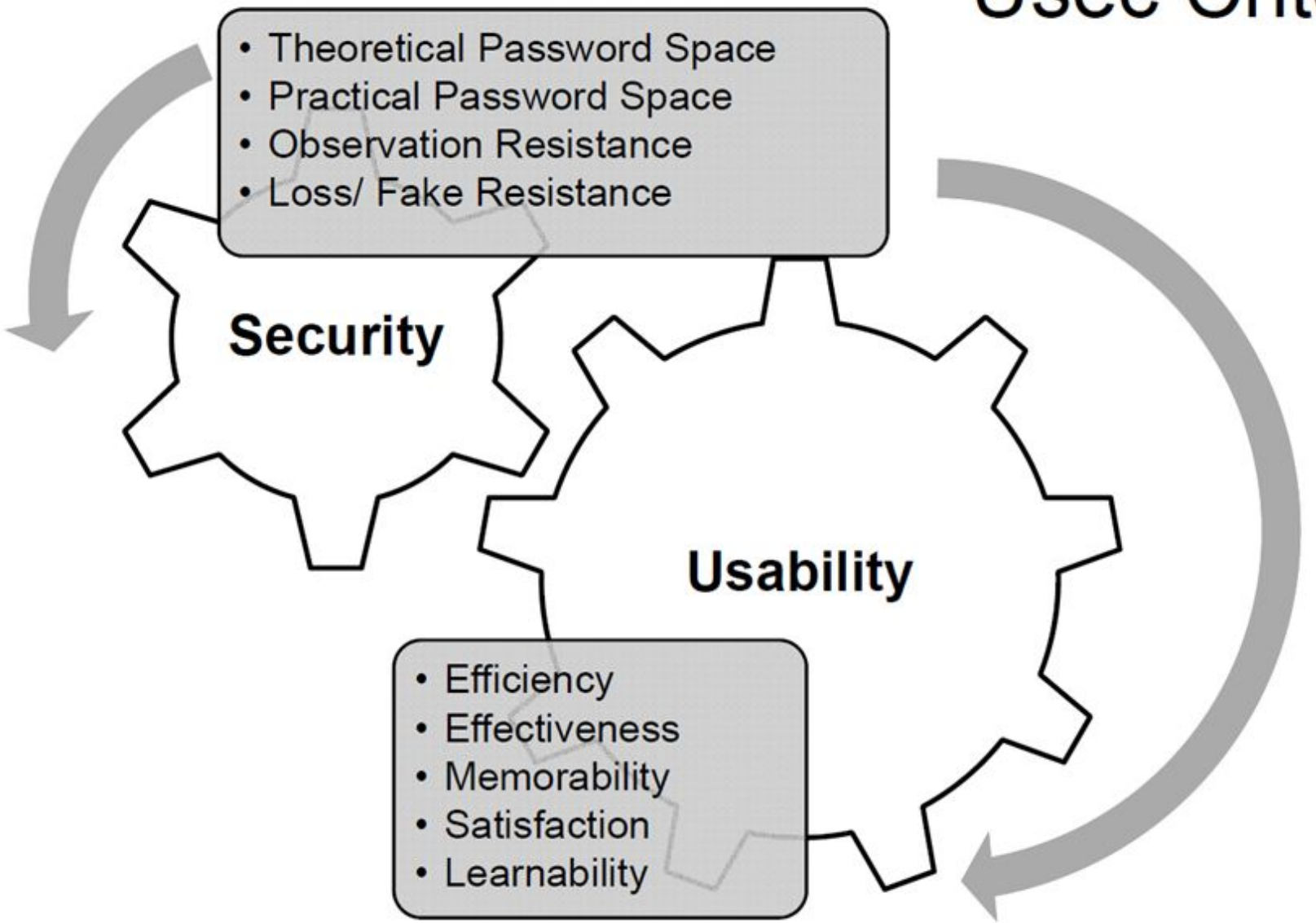


SECURITY AS SECONDARY TASK

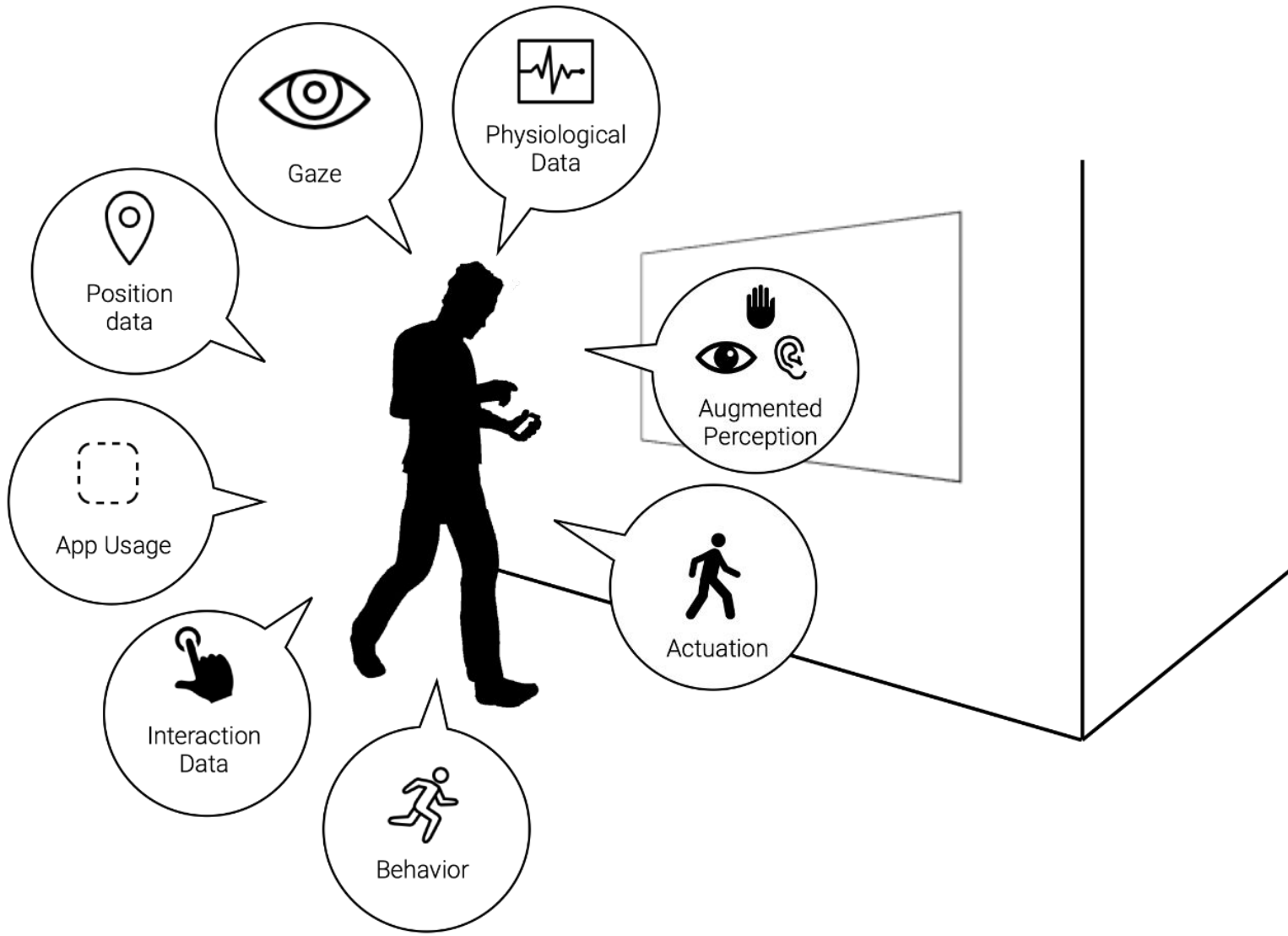


SECURITY AS SECONDARY TASK

Usec Criteria

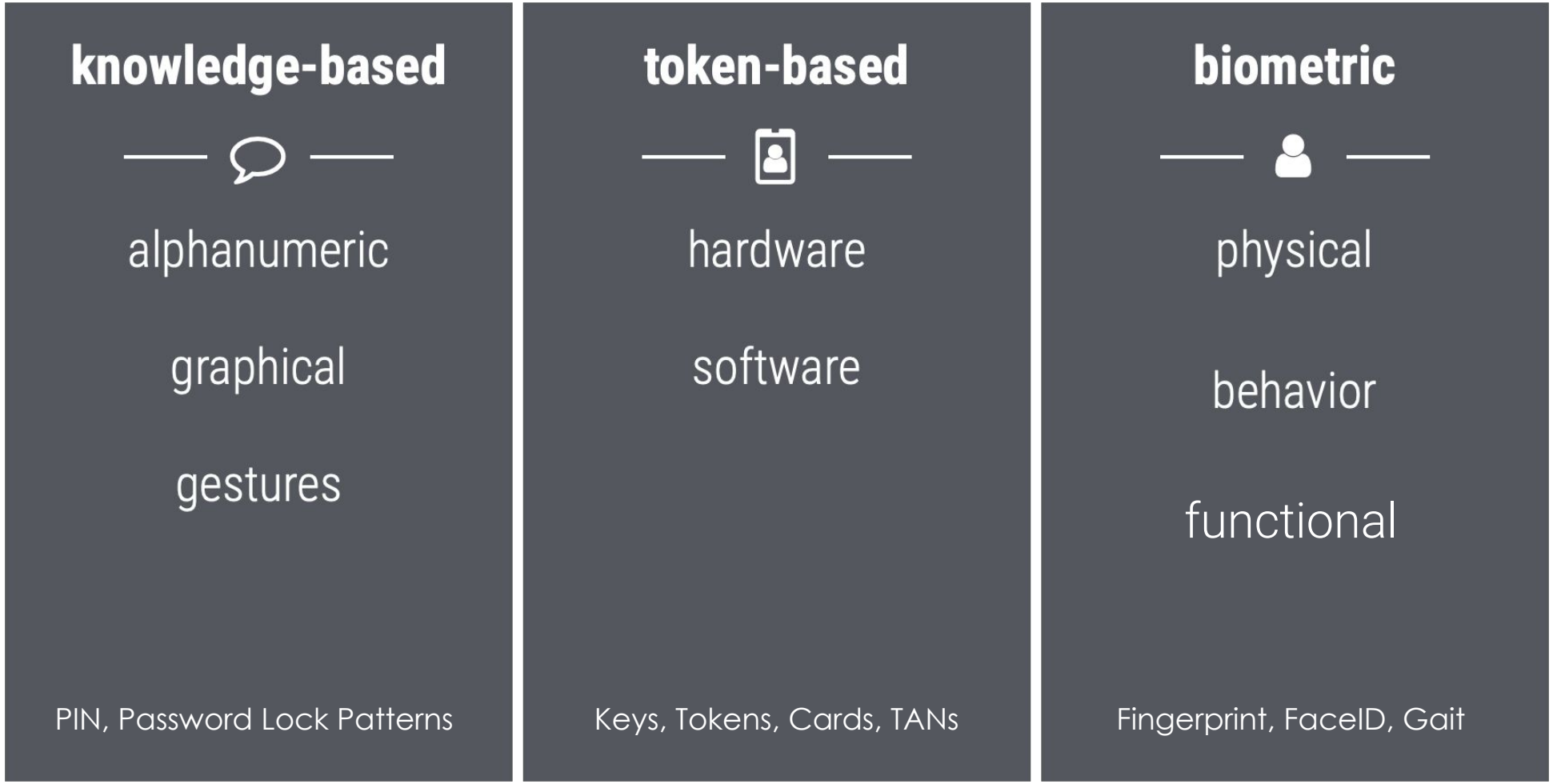


Sensing and actuation moving closer to the human body



Brief Introduction to Authentication

Classification of Authentication Schemes



Knowledge-based Authentication

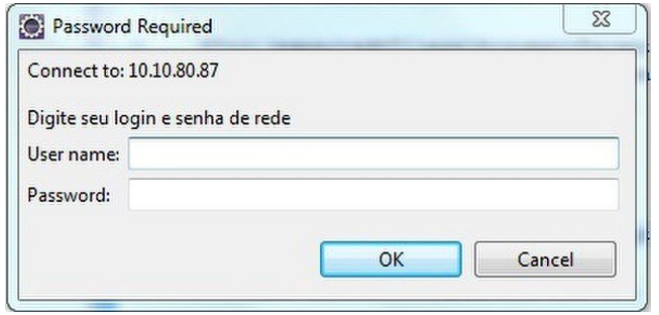
Prevalent form of authentication.

Security

- Theoretically high security (large theoretical password space)
- In practice: low security (human factors)

Key issues (human factors):

- Memorability
- Complexity
- Shareability
- Input time of explicit interaction
- Failure rate of ~10% (Brostoff & Sasse)



Brostoff, S. and Sasse, M. A. 2003. “Ten strikes and you’re out”: Increasing the number of login attempts can improve password usability. In Workshop on Human-Computer Interaction and Security Systems at CHI 2003

Token-based Authentication

http://store.nfcring.com/

Mostly used in two-factor authentication

http://www.tokenguard.com/images/tokens/SID700.gif

Security

- Theoretically high security (large password space)
- In practice: low security (human factors)

http://www.girokonto.org/wp-content/uploads/2014/02/ec-karte-sparkasse.jpeg

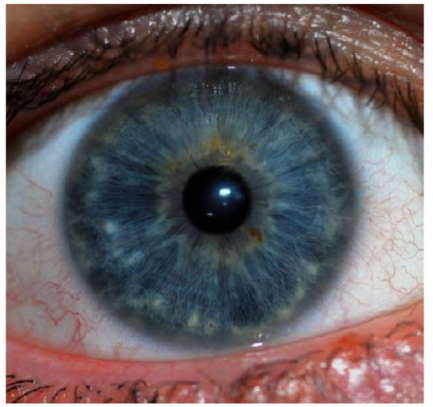
Key issues

- Shareability
- Input Time
- Can be lost and must be carried, alike any key

http://api.sonymobile.com/files/SmartWatch-3-SWR50-black-1240x840-79054d32a0d13a97bedae3d0b12f62af-79054d32a0d13a97bedae3d0b12f62af.jpg

Krol, K., Philippou, E., Cristofaro, E. D., and Sasse, M. A. 2015. "They brought in the horrible key ring thing!" Analysing the Usability of Two-Factor Authentication in UK Online Banking, NDSS Symposium 2015

Biometrics



http://upload.wikimedia.org/wikipedia/commons/d/dc/Human_Iris_JD052007.jpg

Increasingly popular form of authentication.

Security and Usability

- Level of security depends on modality
  (DNS and fingerprints are very secure; FaceID maybe less so)
- Biometrics consider human factors well.

Advantages

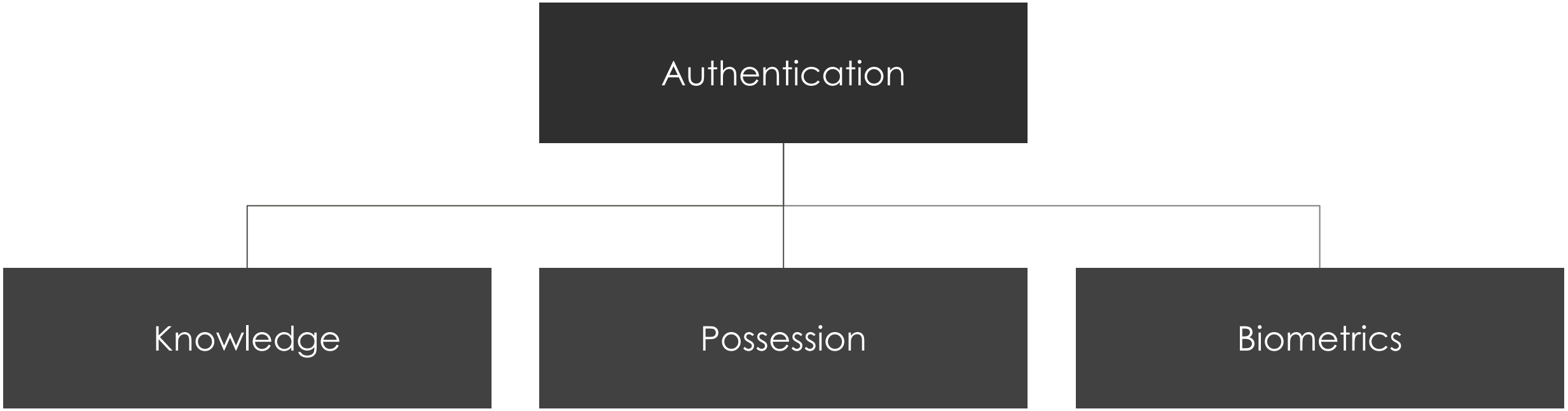
- Cannot be lost or forgotten
- Does not have to be remembered.

Key issues

- Can be stolen (but hardly so).
- Difficult to share
- Cannot (or only hardly) be changed willingly.
- Can change over time.

http://www.vetmed.vt.edu/education/curriculum/v

Overview



"Something you know."

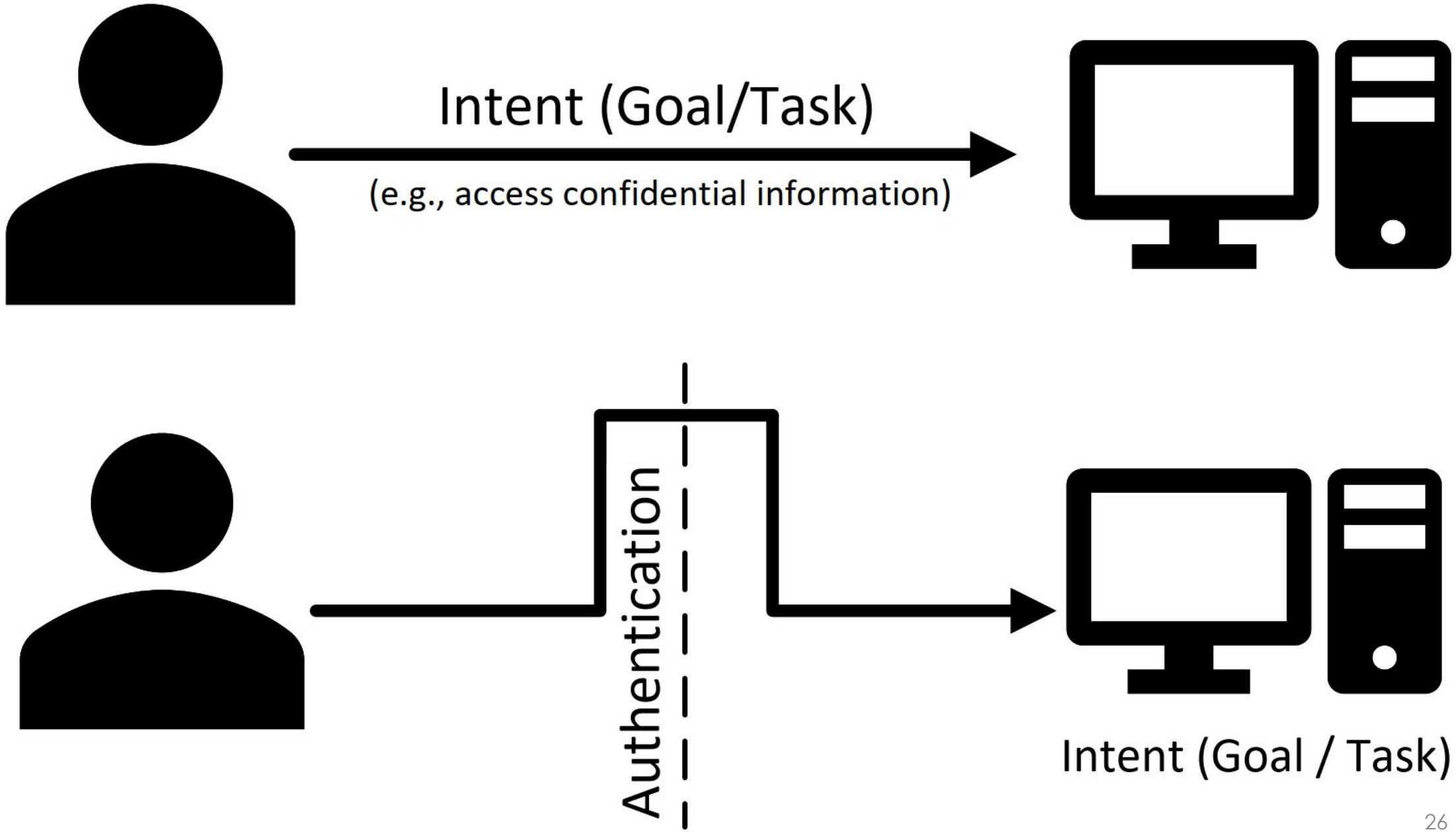
PINs, Passwords, Patterns, …

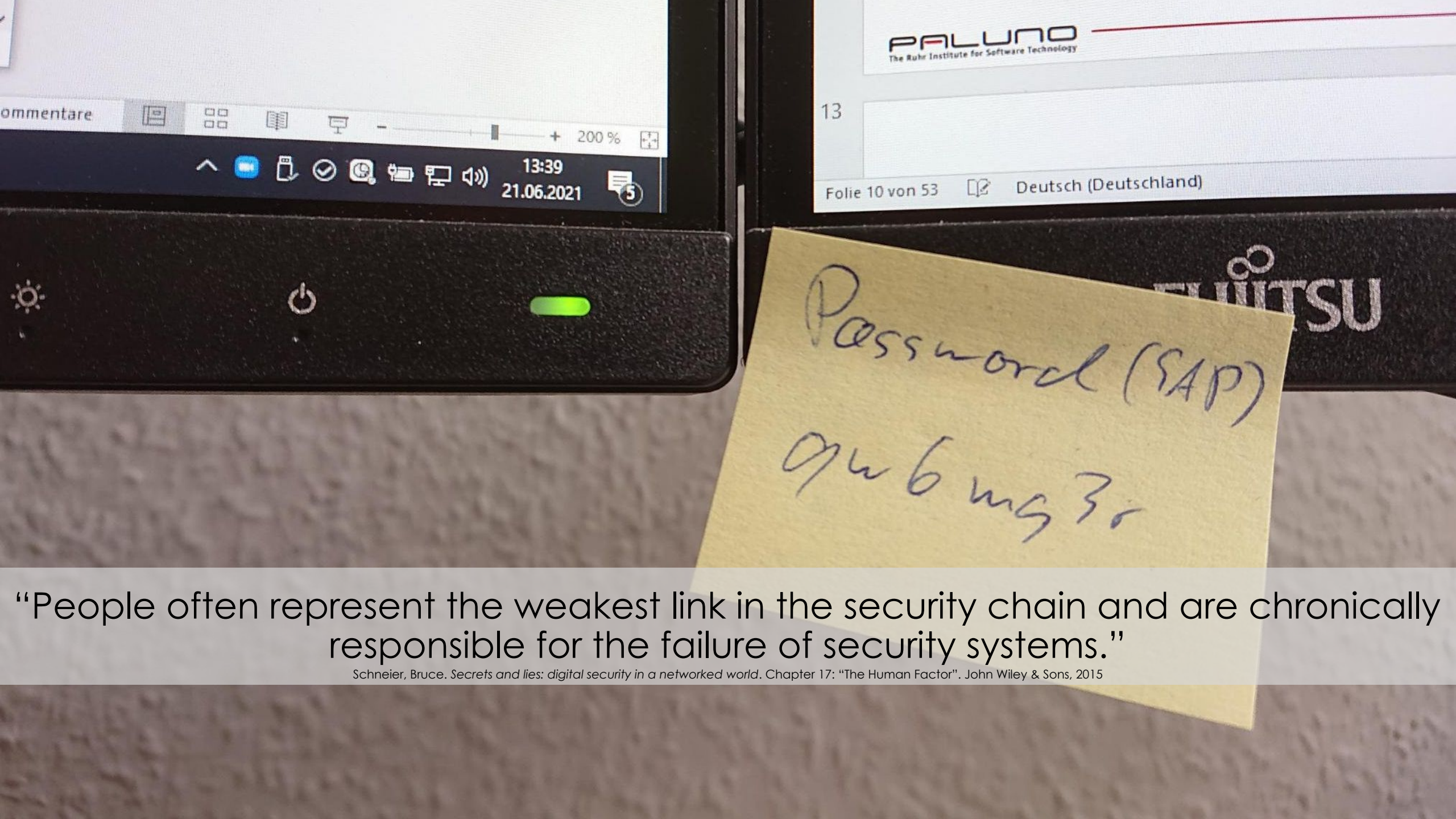
"Something you have."

Keys, Tokens, Cards, …

"Something you are."

Fingerprint, FaceID, Gait, …


Intent (Goal/Task)
(e.g., access confidential information)
Authentication
Intent (Goal / Task)



"People often represent the weakest link in the security chain and are chronically responsible for the failure of security systems."

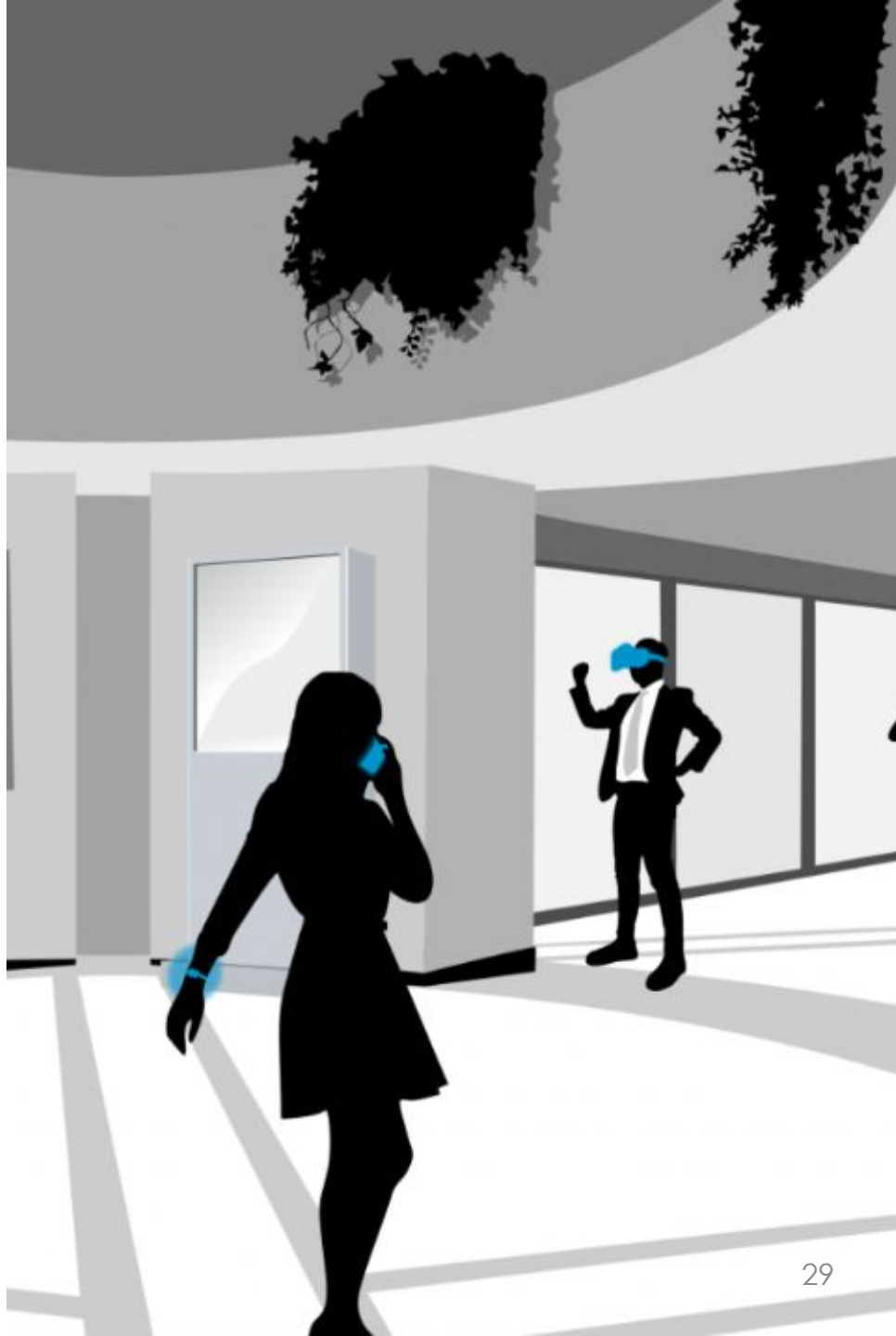
Schneier, Bruce. *Secrets and lies: digital security in a networked world*. Chapter 17: "The Human Factor". John Wiley & Sons, 2015

"Putting a password on for each individual user as a lock seemed like a very straightforward solution.“

Fernando Corbató - https://www.wired.com/2012/01/computer-password/

Future Authentication

- Implicit
- Continuous
- Privacy-Preserving
- Ubiquitous



F. Alt and S. Schneegass. Beyond Passwords — Challenges and Opportunities of Future Authentication. In IEEE Security & Privacy (Volume: 20, Issue: 1, Jan.-Feb. 2022)

Implicit Interaction and Authentication

"When observing communication between humans, we can see that **a lot of information is only exchanged implicitly** [...] based on the implicitly introduced contextual information, such as gestures, body language, and voice"

Schmidt, A. 2000. Implicit human computer interaction through context. Personal Technologies 4, 2-3, 191–199.

"Implicit Authentication – the ability to authenticate [...] users based on **actions they would carry out anyway**."

Jakobsson, M., Shi, E., Golle, P., and Chow, R. 2009. Implicit Authentication for Mobile Devices. In Proceedings of the 4th USENIX Conference on Hot Topics in Security. HotSec'09. USENIX Association, USA

Definition (Implicit Interaction): "An action performed by the user that is not primarily aimed to interact with a computerised system but which such a system understands as input."

Schmidt, A. 2000. Implicit human computer interaction through context. Personal Technologies 4, 2-3, 191–199.

Explicit vs. Implicit Authentication

Example: Fingerprint recognition.

- Explicit: Having to move the finger to a sensor.
- Implicit: Hiding the sensor under the screen that is used *anyway*.

Example: Face recognition.

- Explicit: “Cooperate”; e.g., perform a blink upon request.
- Implicit: being recognized without noticing.

Christian Holz and Patrick Baudisch. Fiberio: a touchscreen that senses fingerprints. In Proc. UIST '13

Implicit Authentication (IA)

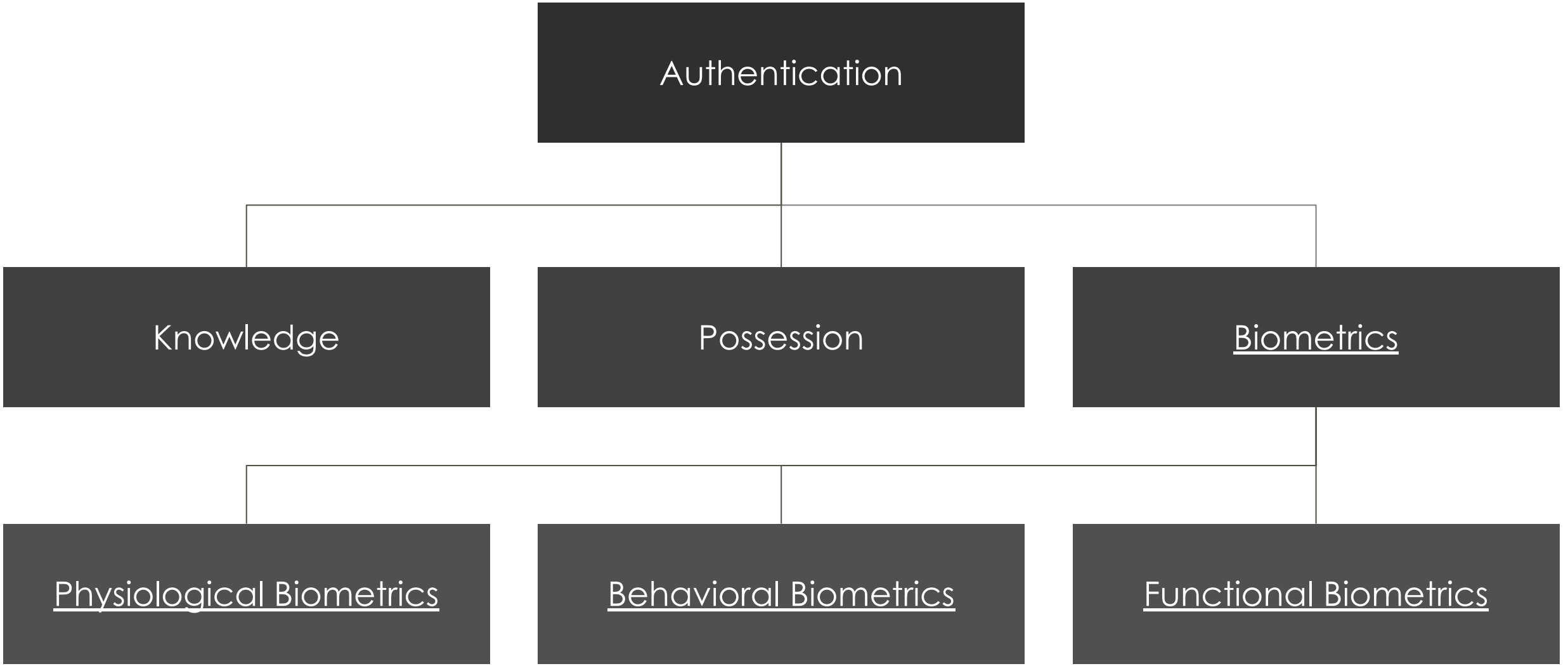
- Probabilistic system of whether the user is still the same.
- Samples user behavior continuously
- Can lock a session mid-usage
- Invisible and effortless for the user

- **Behavioral Biometrics** are particularly suited for IA.

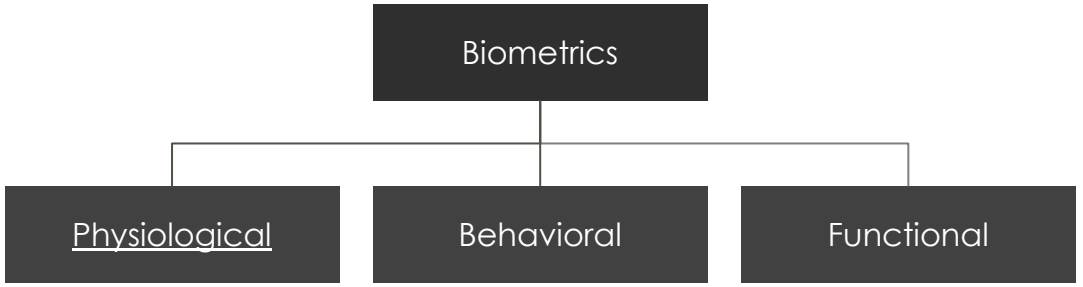
Traoré, I. and Ahmed, A. A. E. 2012. Introduction to Continuous Authentication. In Continuous Authentication Using Biometrics: Data, Models, and Metrics, I. Traore and A. A. E. Ahmed, Eds. IGI Global, Hershey, PA, USA, 1–22. https://dx.doi.org/10.4018/978-1-61350-129-0.ch001

Biometrics

Biometrics


Authentication
Knowledge
Possession
Biometrics
Physiological Biometrics
Behavioral Biometrics
Functional Biometrics

Phys. Biometrics



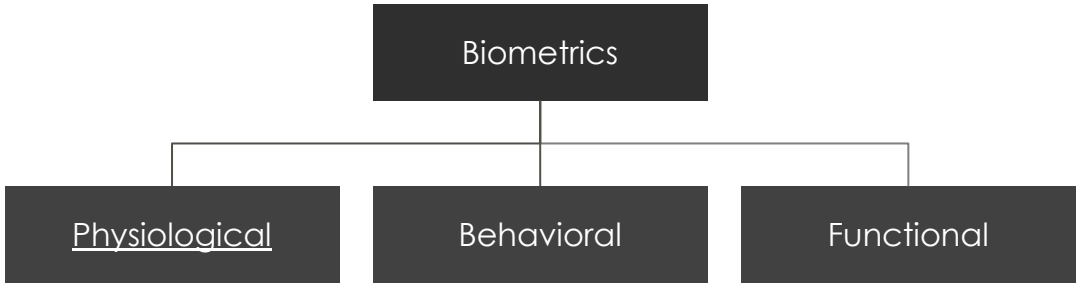
Physiological Biometrics are primarily based on users' physiology.

Examples: Face, Fingerprint, Voice (partly), DNS, Iris, …

Characteristics:

- Mostly stable, i.e., they change little over time.
- Sensing physiological biometrics implicitly can be challenging.

Functional Biometrics



Functional Biometrics are based on body-reflections.

- A signal (e.g., sound) is generated and emitted towards the body.
- Sensors capture the unique reflection.
- User's body is treated as a function that receives a signal and generates a response.

Characteristics

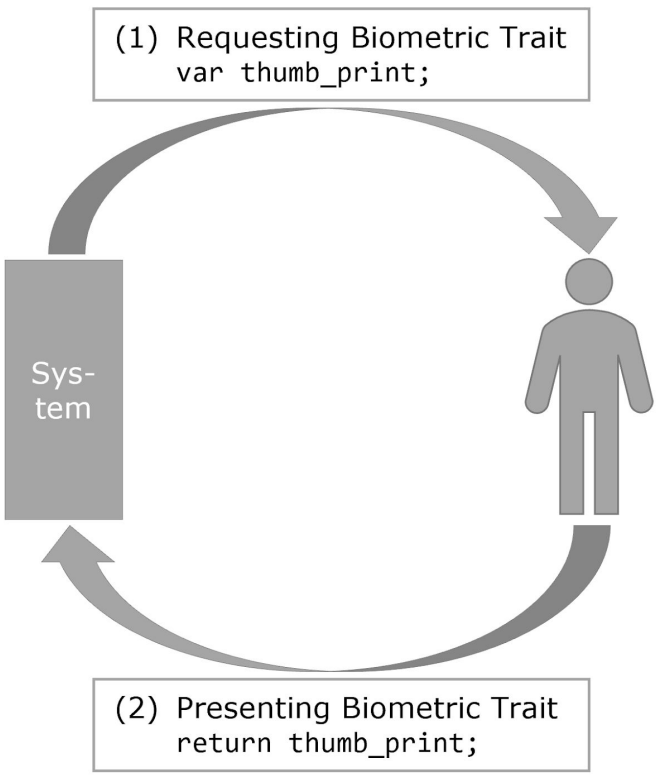
- Compared to physiological biometrics, more input-output combinations can be created.
- Invalidating input-output-pairs is possible.

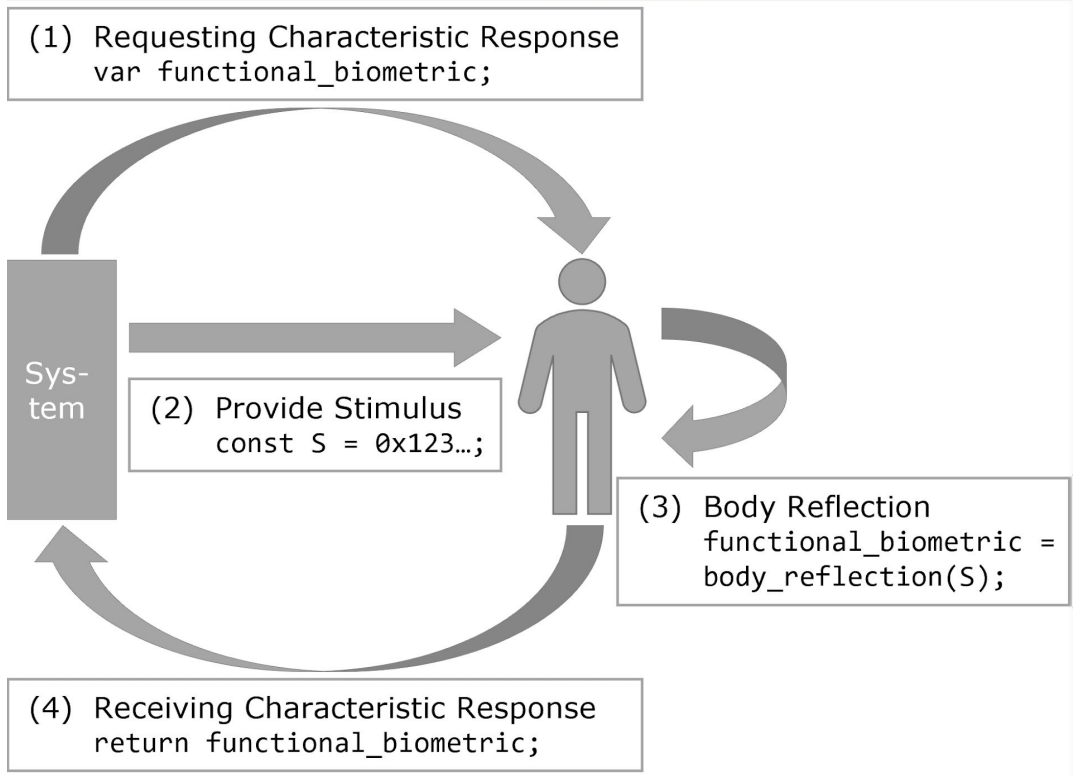
Liebers, J. and Schneegass, S. 2020. Introducing Functional Biometrics: Using Body-Reflections as a Novel Class of Biometric Authentication Systems. In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. CHI EA '20. Association for Computing Machinery, New York, NY, USA

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation

Functional Biometrics

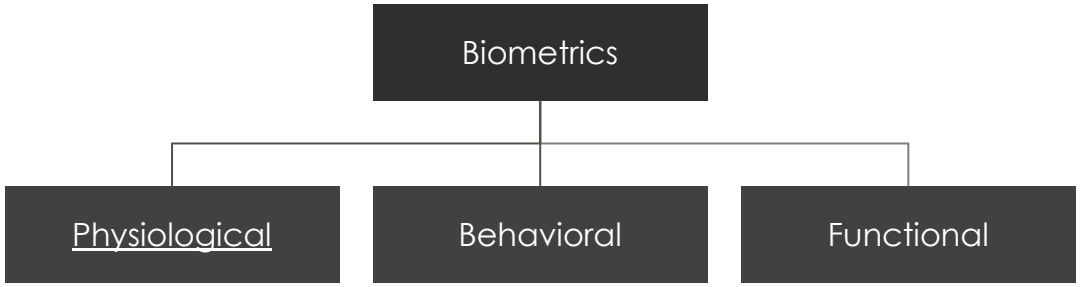
Functional Biometrics





J. Liebers and S. Schneegass. Introducing Functional Biometrics:
Using Body-Reflections as a Novel Class of Biometric Authentication Systems. In CHI EA ’20

Behavioral Biometrics



Behavioral Biometrics are primarily based on users' behavior.

Examples: Gait, Voice, Hand and Body Motion, Touch, Keystrokes, Gaze.

Characteristics:

- Versatile stability (e.g., behavior changes through learning and training over time)
- Often easy to be sampled implicitly.
- Stability-related issues can be countered by frequent re-training.

Stability

"Will my authentication system still be working tomorrow?"

Stability: the degree to which the biometric factor changes over time.

Influencing factors:

- Gait depends on **emotional and cognitive state**.
- Gaze depends on **tiredness**.
- Many forms of behavior depend on the **learning effect** (e.g., VR tasks).

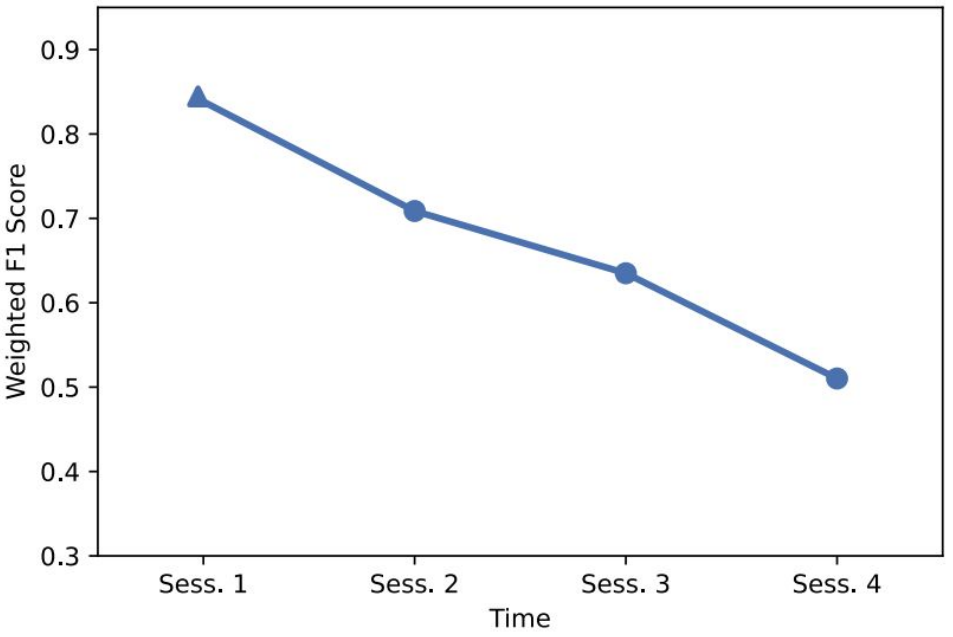
Solution: Test it on multiple days (at least two)!

Peter P. K. Chan, Chao-Ying Chen, Hussein Ayache, Lobo Louie, Alan Lok, Nathan Cheung, and Roy T. H. Cheung. 2021. Gait difference between children aged 9 to 12 with and without potential depressive mood. Gait & posture 91 (2021), 126–130. https://doi.org/10.1016/j.gaitpost.2021.10.012

Charlotte J. W. Connell, Benjamin Thompson, Gustav Kuhn, Michael P. Claffey, Shelley Duncan, and Nicholas Gant. 2016. Fatigue related impairments in oculomotor control are prevented by caffeine. Scientific Reports 6 (2016), 26614. https://doi.org/10.1038/srep26614

Multi-Session Study Designs

- User study designs with *multiple sessions* should be considered.
- Repeated-measures across different days.
- Train with session 1, test with session 2.
- Identification rate (usually) declines over time.



Review Insights

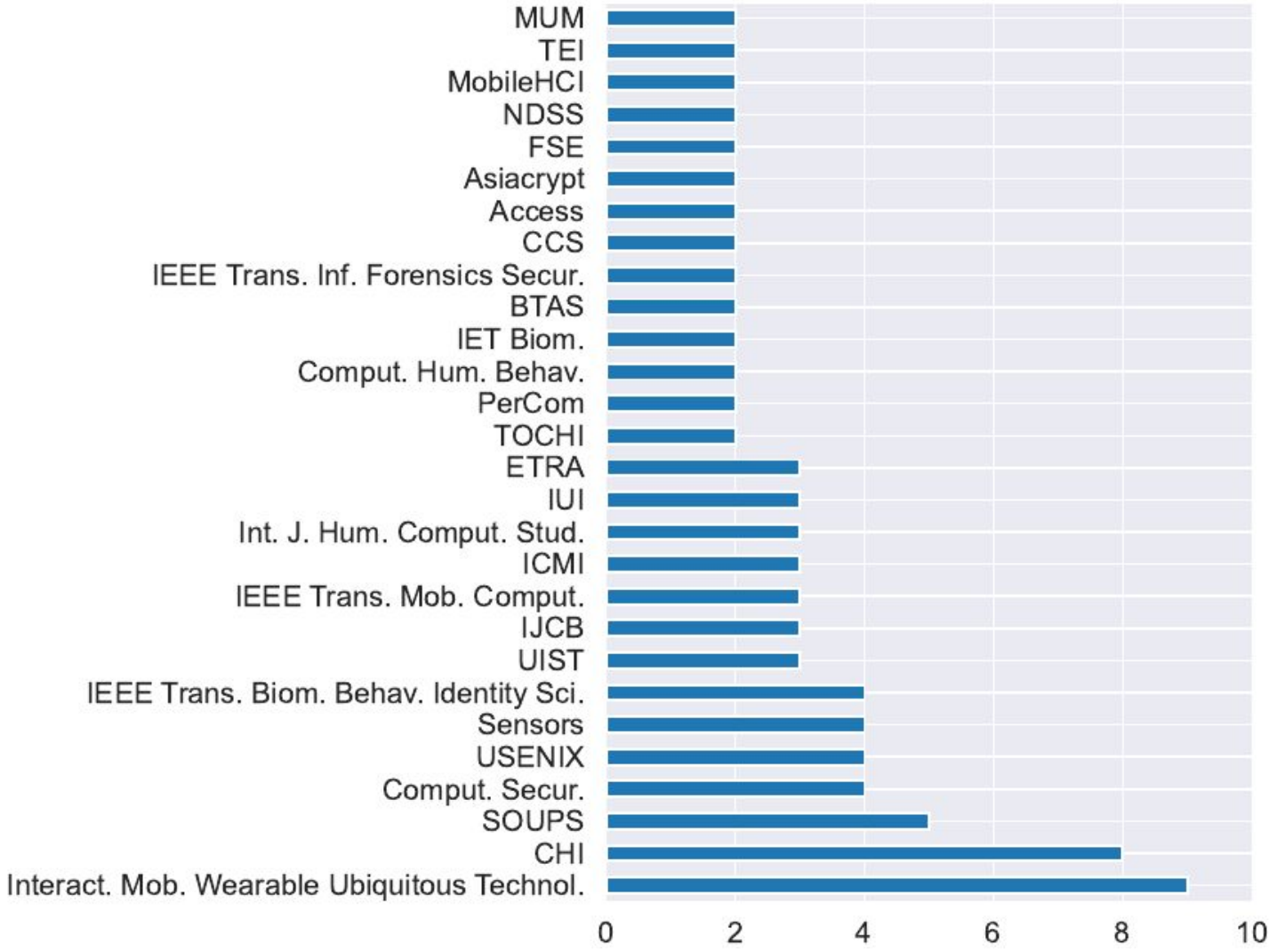
Further motivation for this course

Review Insights

- We asked 71 experts from community in an online survey where they submit and for review experiences

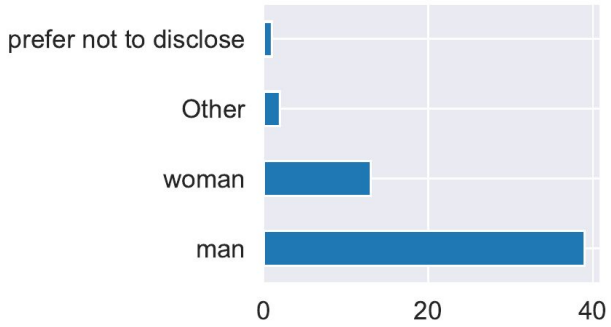
- Questions on indication of **value** and **rigor** in a work.
- What invokes **criticism?**

- Most participants from the HCI community submit at IMWUT
  - Closely followed by CHI

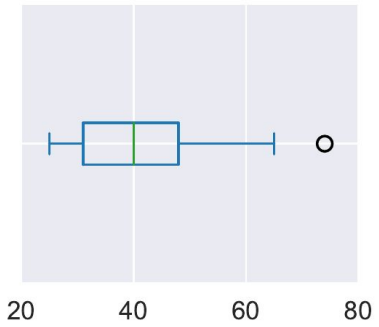


Review Insights

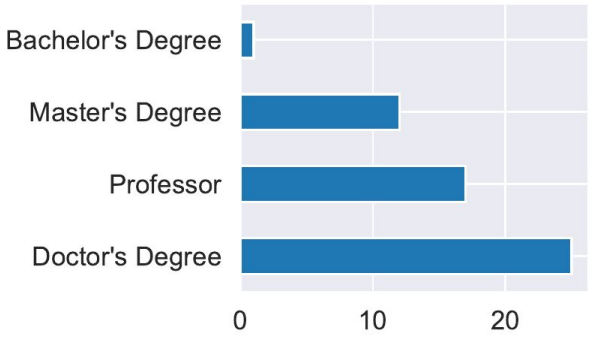
- Mean age: 40.79 (SD = 11.50)
- Mostly postdocs (25) and professors (17)
- Mean years of expertise: 11.20 (SD = 9.87)
- Mean number of reviews submitted: 11.96 (SD=16.63)



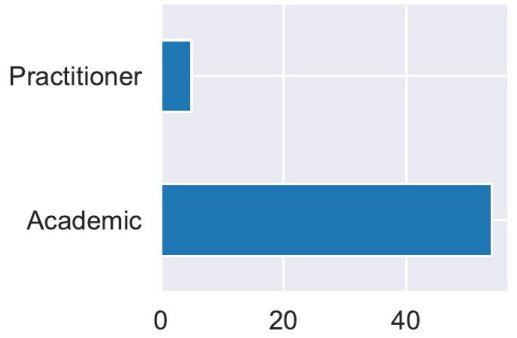
(a) Q1: What is your gender?



(b) Q2: What is your age?



(a) Q3: Education.



(b) Q4: Are you an academic or practitioner?

Key-Takeaways

- User Study and Dataset are major indicators for **rigor.**
- Presentation (clear, detailed style of writing) is both important for **value** and **criticism.**
- Methodology, evaluation protocol and results are main cause of **criticism**.

Top 3 ranking for **value:**

1. Presentation
2. User study and Dataset
3. Methodology, evaluation protocol, and results

Top 3 ranking for **rigor**:

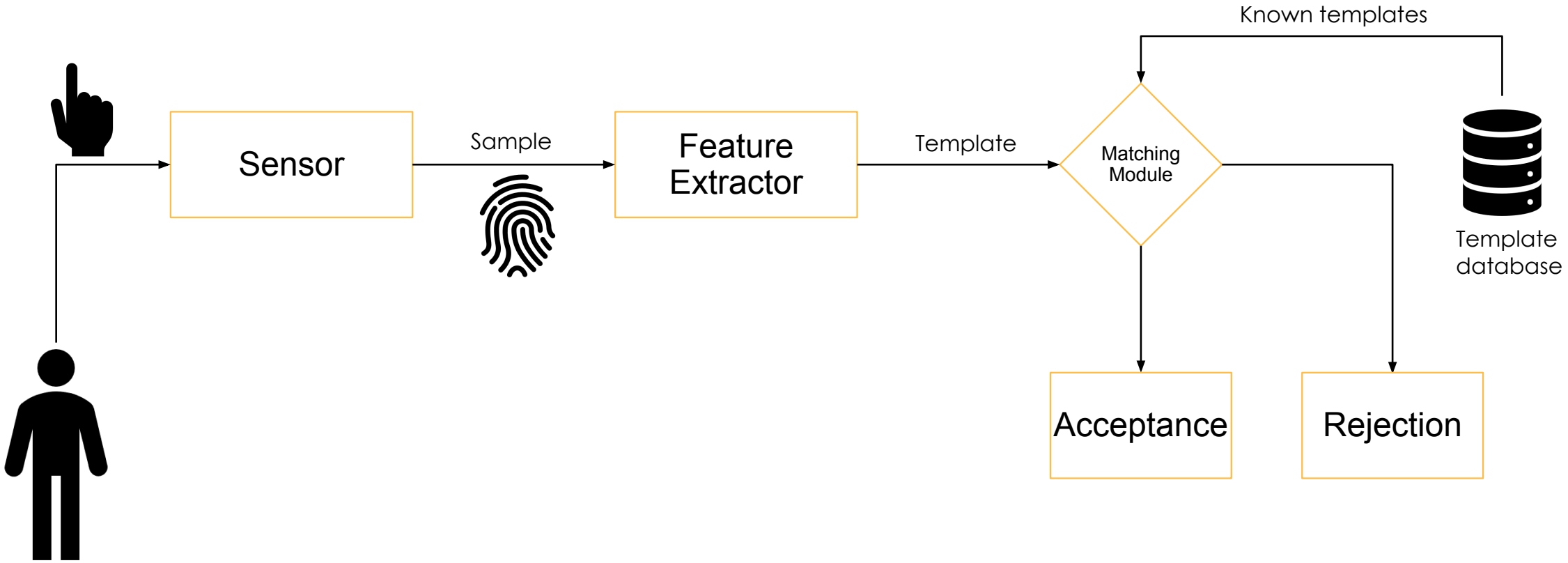
1. User study and Dataset
2. Methodology, evaluation protocol, and results
3. Algorithm, model, features, data, and machine learning

Top 3 ranking for **criticism**:

1. Methodology, evaluation protocol, and results
2. Presentation
3. User study and dataset

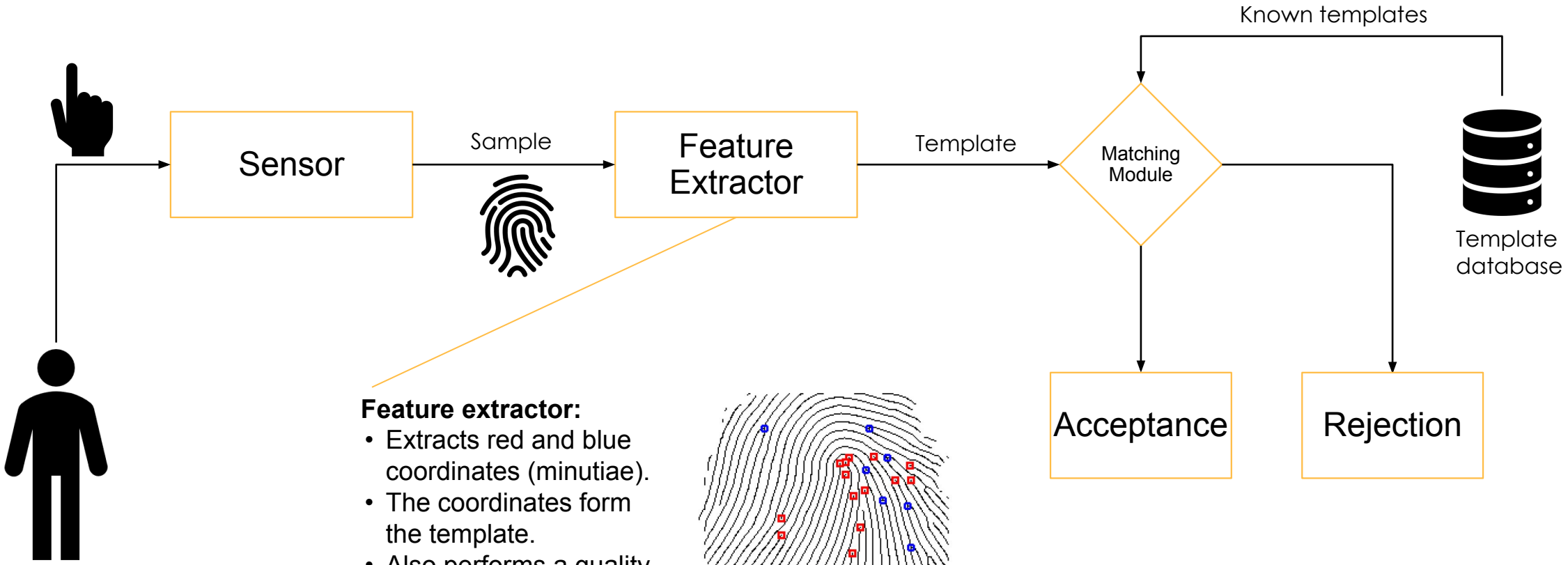
Biometric Authentication Systems

Biometric Authentication Systems: Overview



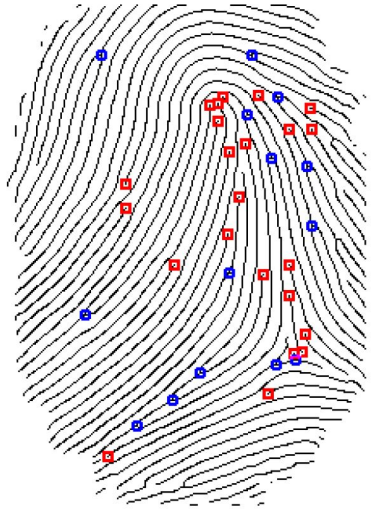
Jain, Anil K., Patrick Flynn, and Arun A. Ross, eds. Handbook of biometrics. Springer Science & Business Media, 2007.

Biometric Authentication Systems: Overview



Feature extractor:

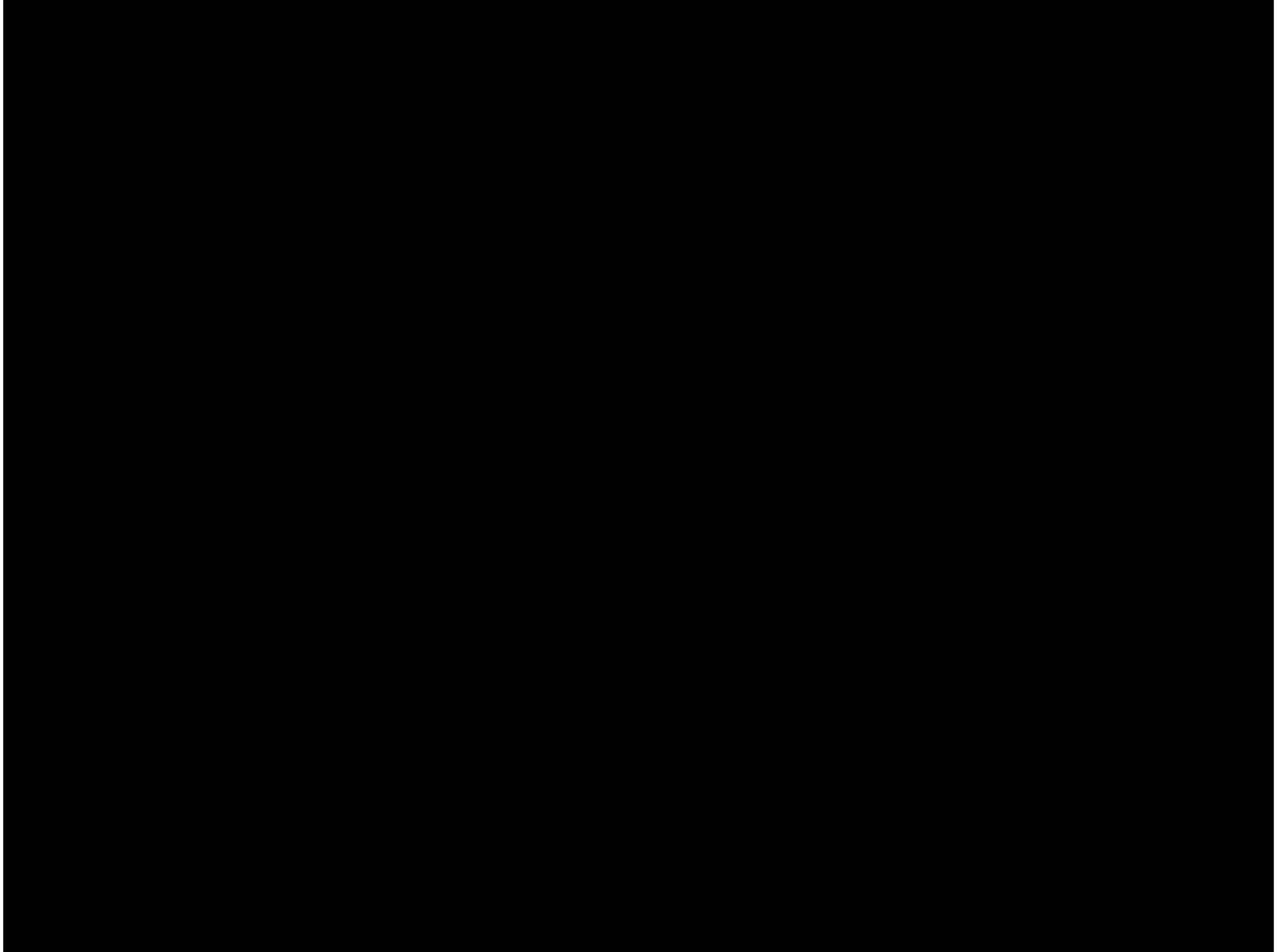
- Extracts red and blue coordinates (minutiae).
- The coordinates form the template.
- Also performs a quality assessment.



Minutiae image: Sudiro, S. A., & Lukman, S. (2015, December). Minutiae matching algorithm using artificial neural network for fingerprint recognition. In *2015 3rd International Conference on Artificial Intelligence, Modelling and Simulation (AIMS)* (pp. 37-41). IEEE.

Hands On: Data Collection 1!

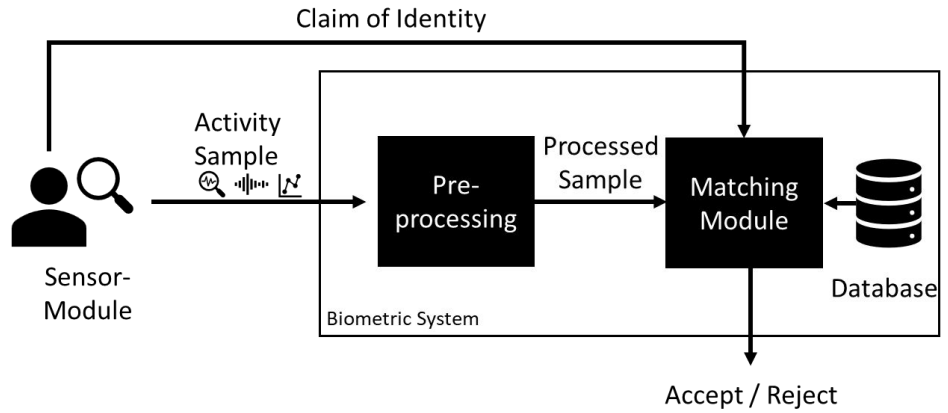
Please pick up the VR headset.



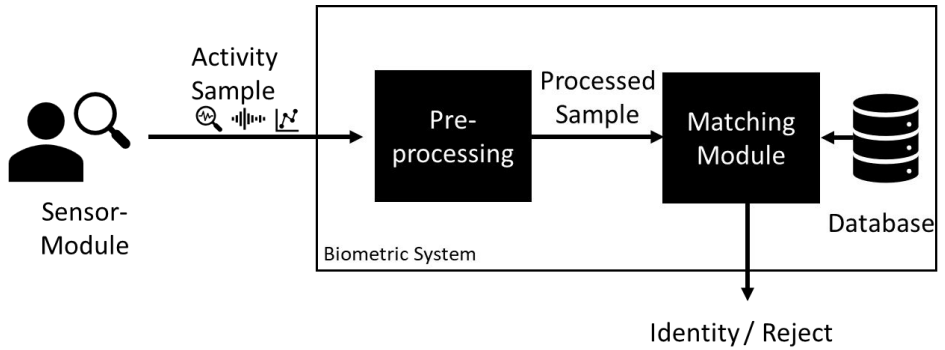
Biometric Authentication

Objectives (why?):

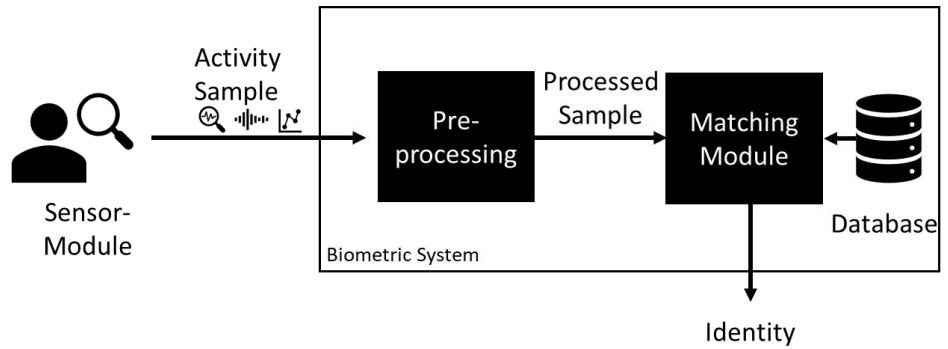
- Authorization
- Personalization



Verification



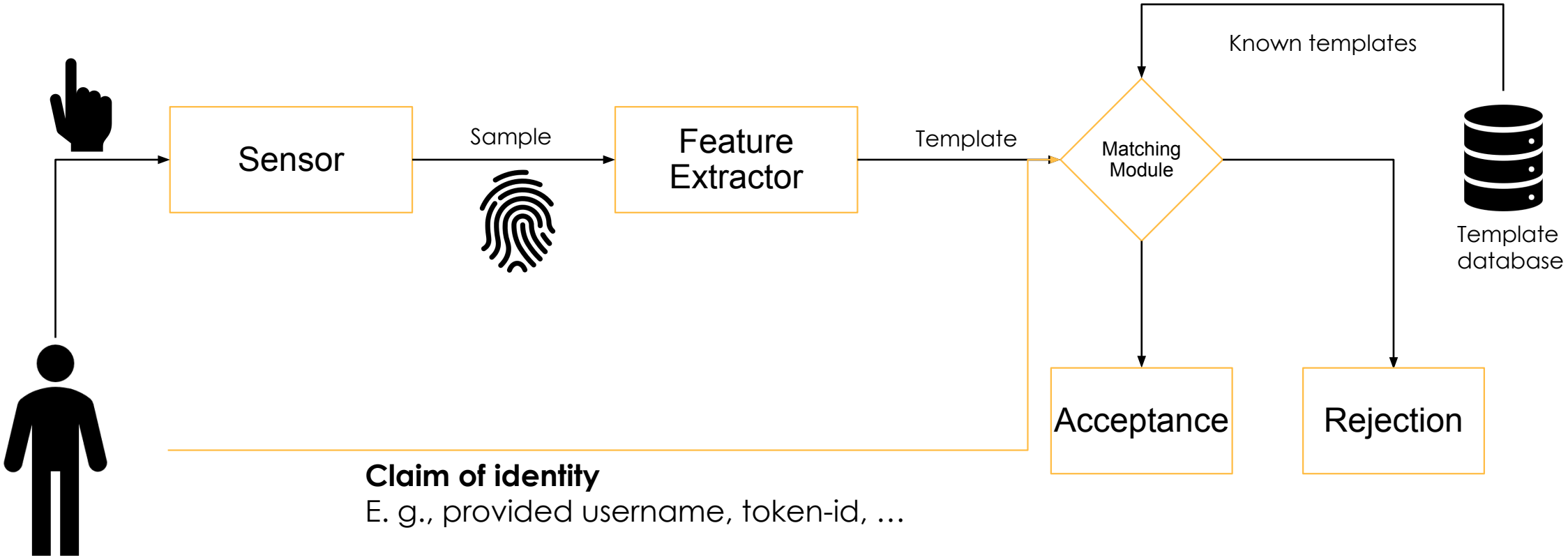
Identification (open set)



Identification (closed set)

Verification

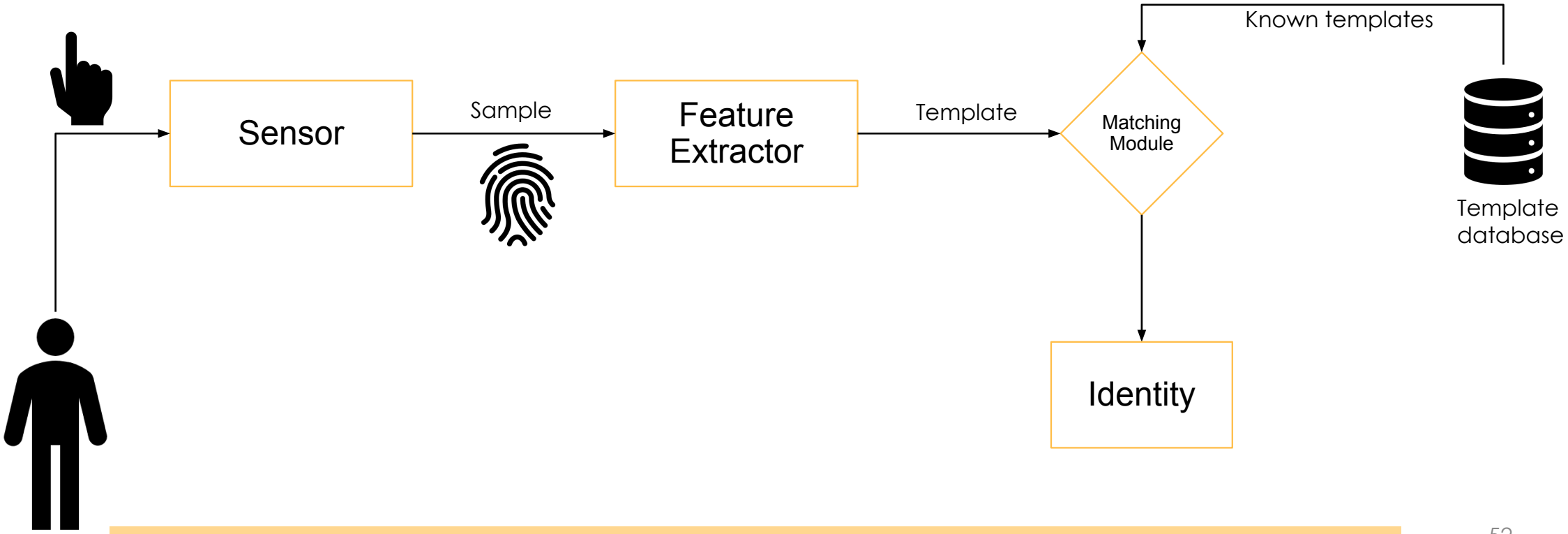
- For verification, the user provides a *claim of identity*.
- System only needs to decide whether this claim is trusted or not.



Jain, Anil K., Patrick Flynn, and Arun A. Ross, eds. Handbook of biometrics. Springer Science & Business Media, 2007.

Closed-Set Identification

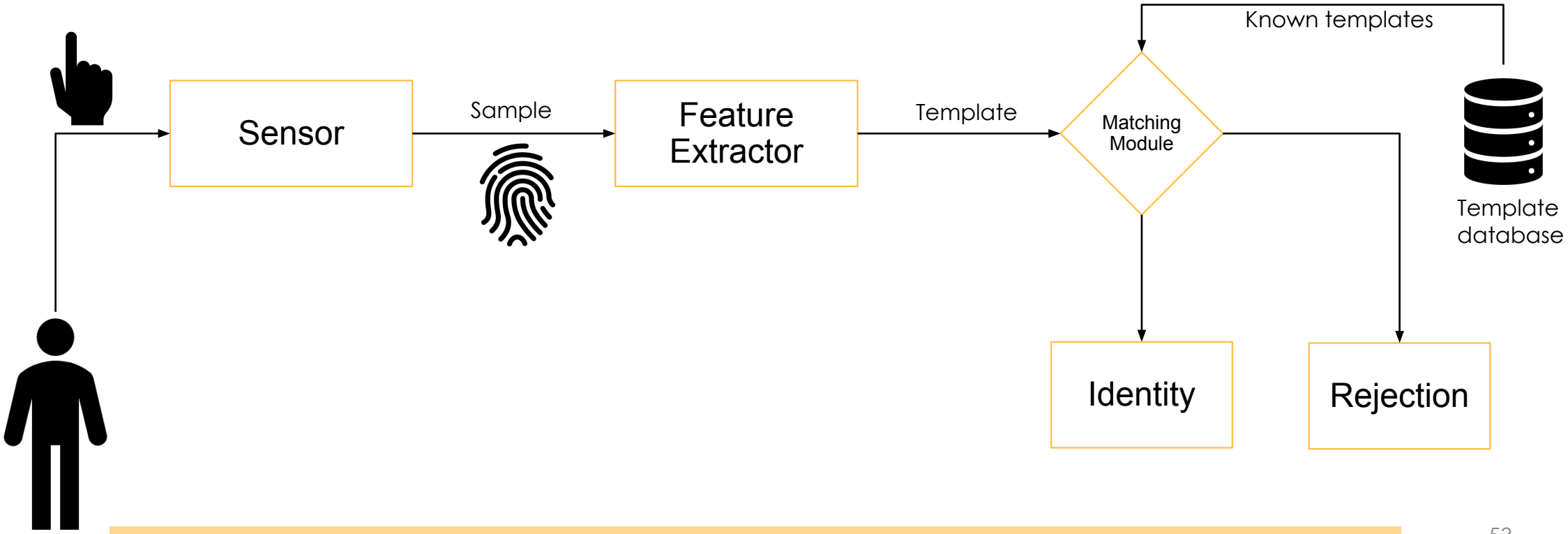
- In Identification-mode, the user only provides a biometric sample.
- Systems needs to find a corresponding identity by the sample.
- Drawback: accepts any user; cannot reject.



Jain, Anil K., Patrick Flynn, and Arun A. Ross, eds. Handbook of biometrics. Springer Science & Business Media, 2007.

Open-Set Identification

- Similar to closed-set identification, but can reject unknown users.
- Drawback: hard to implement (classification + outlier rejection)
- The "ideal" future system.



Jain, Anil K., Patrick Flynn, and Arun A. Ross, eds. Handbook of biometrics. Springer Science & Business Media, 2007.

Matching Biometric Samples

Take samples (sample set 1)

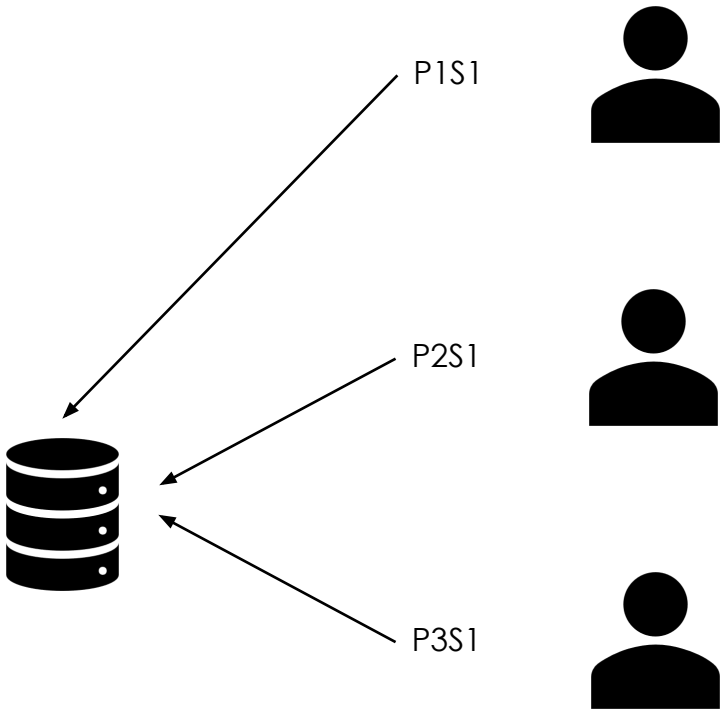
„Participant 1 Session 1“

P1S1

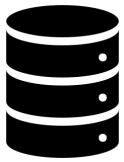
P2S1

P3S1

Enrol templates



Take samples (Sample set 2)



P1S1

P2S1

P3S1

„Participant 1 Session 2“

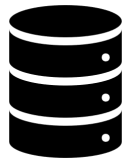
P1S2

P2S2

P3S2

Unknown person

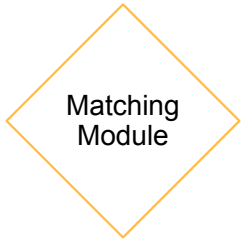
Test sample set 2



P1S1

P2S1

P3S1



P1S2

P2S2

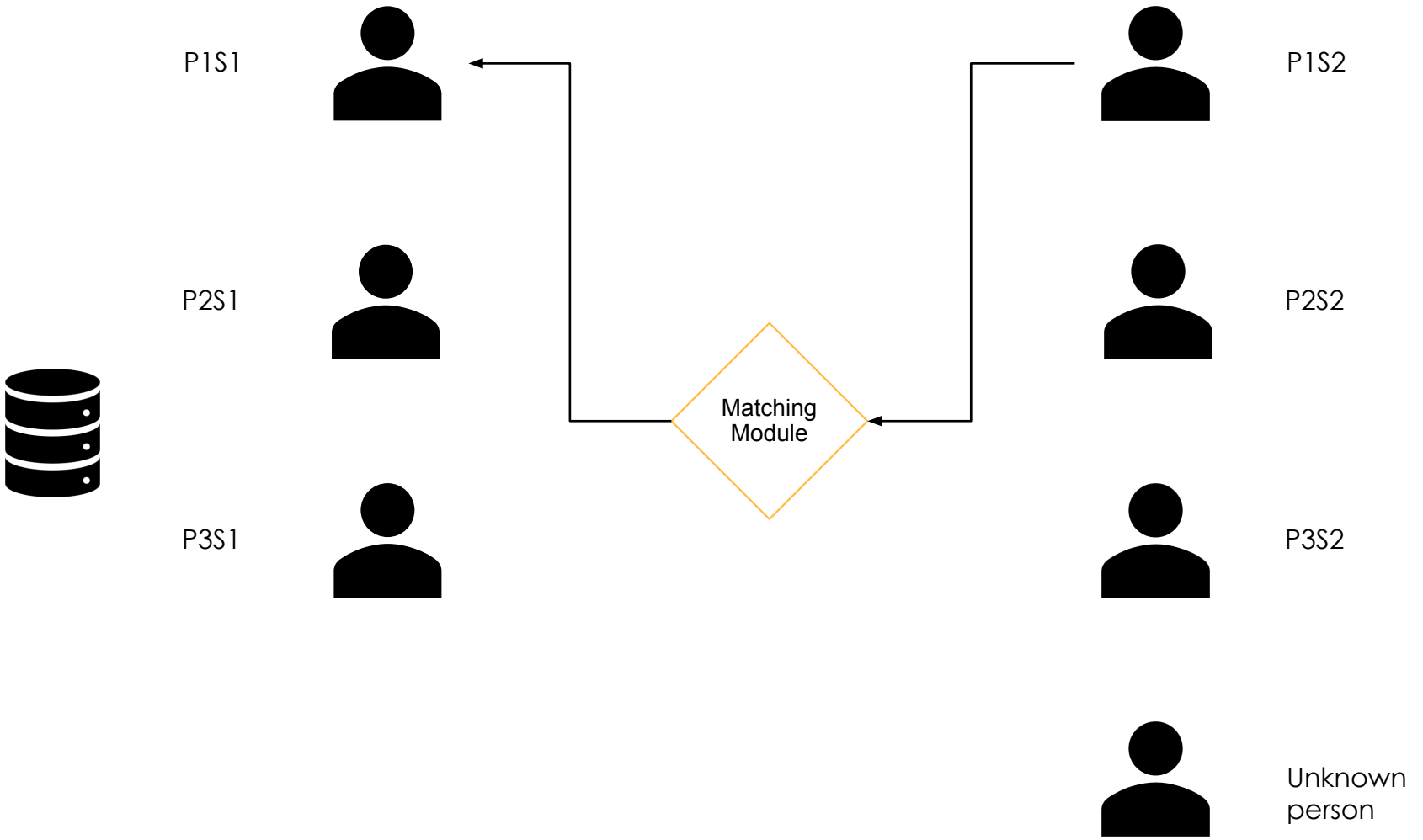
P3S2

Unknown
person

Test sample set 2

Identification mode

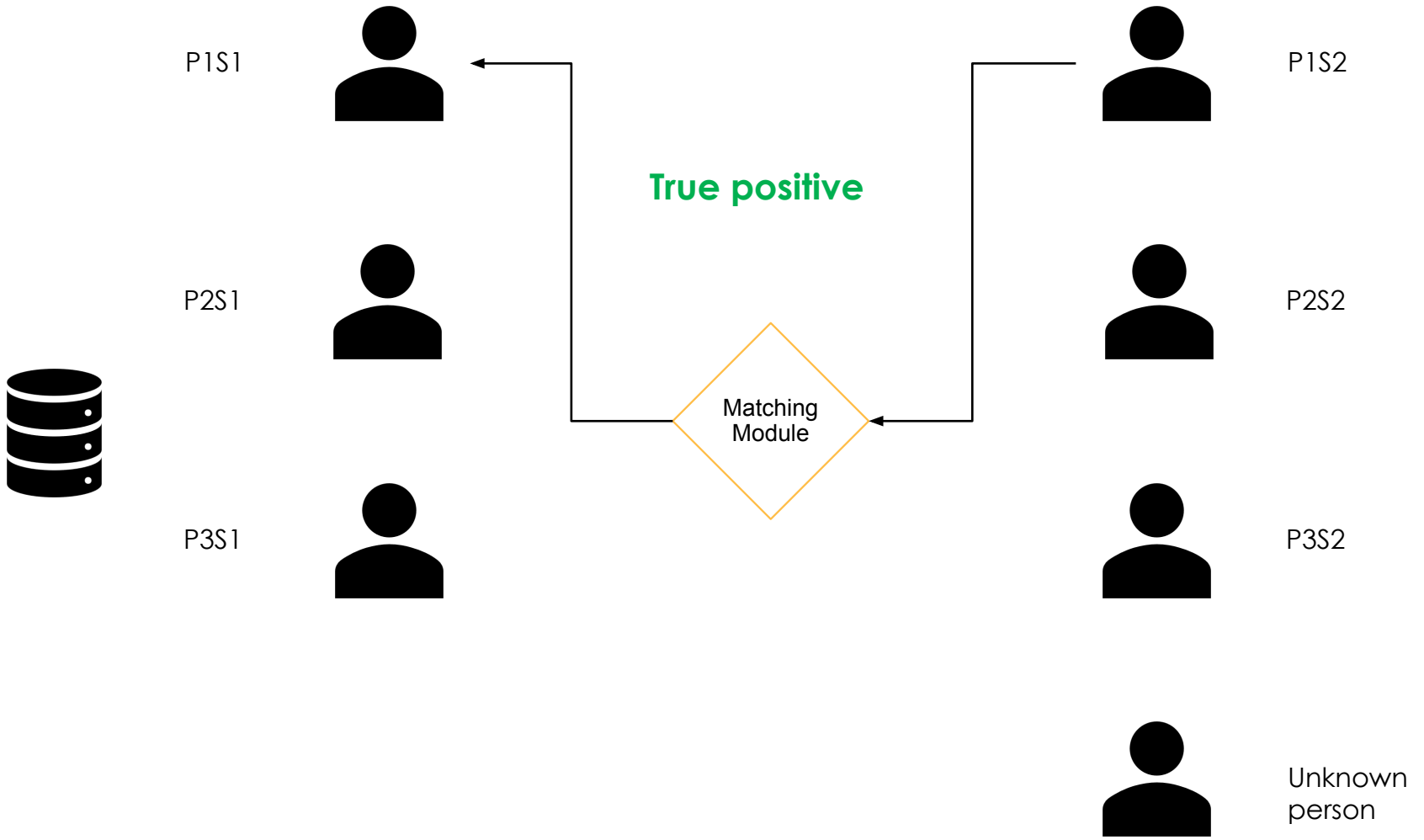
- Input: Sample
- Output: Identity



Test sample set 2

Identification mode
- Input: Sample
- Output: Identity

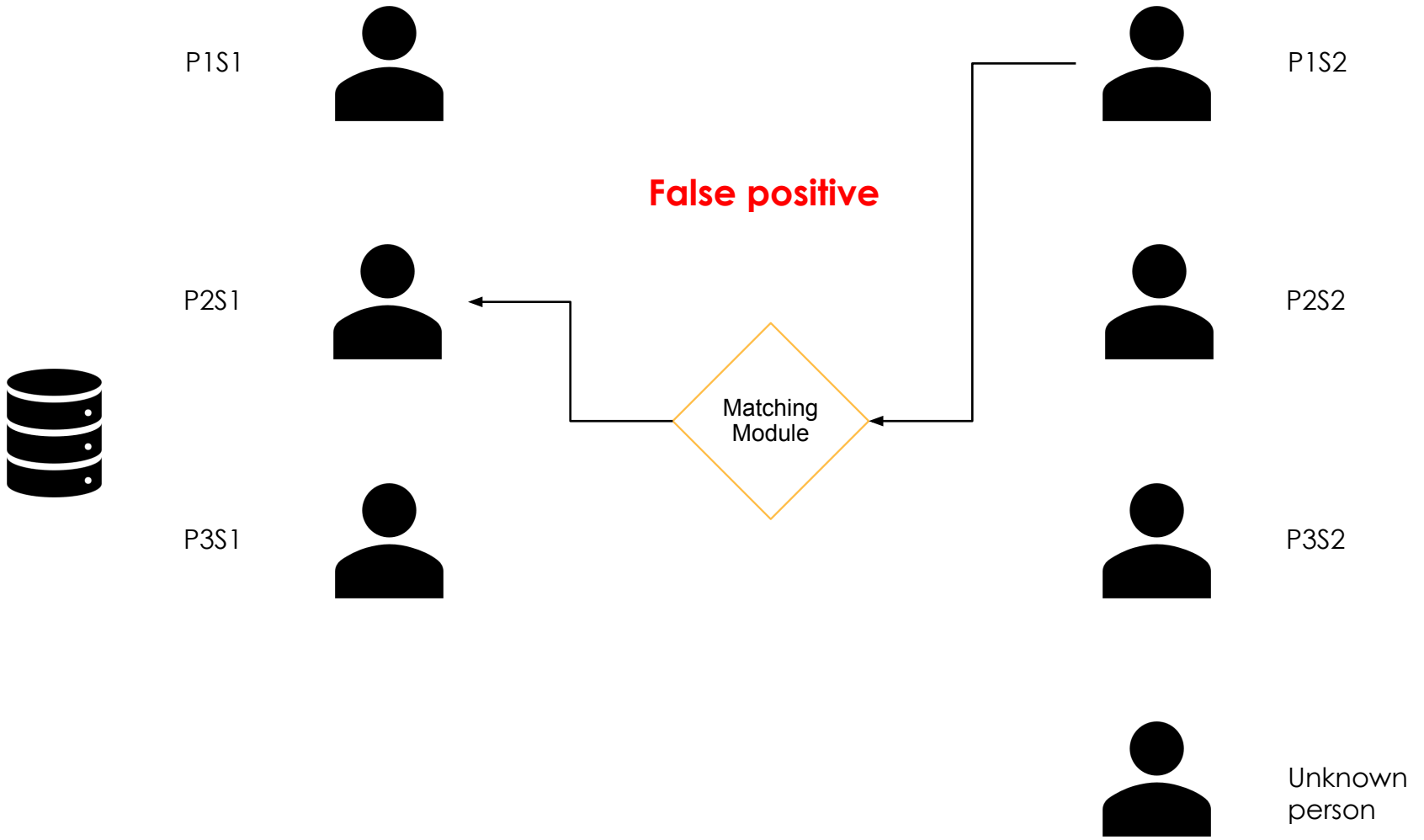
Four cases exist:
1. True positive (TP)



Test sample set 2

Identification mode
- Input: Sample
- Output: Identity

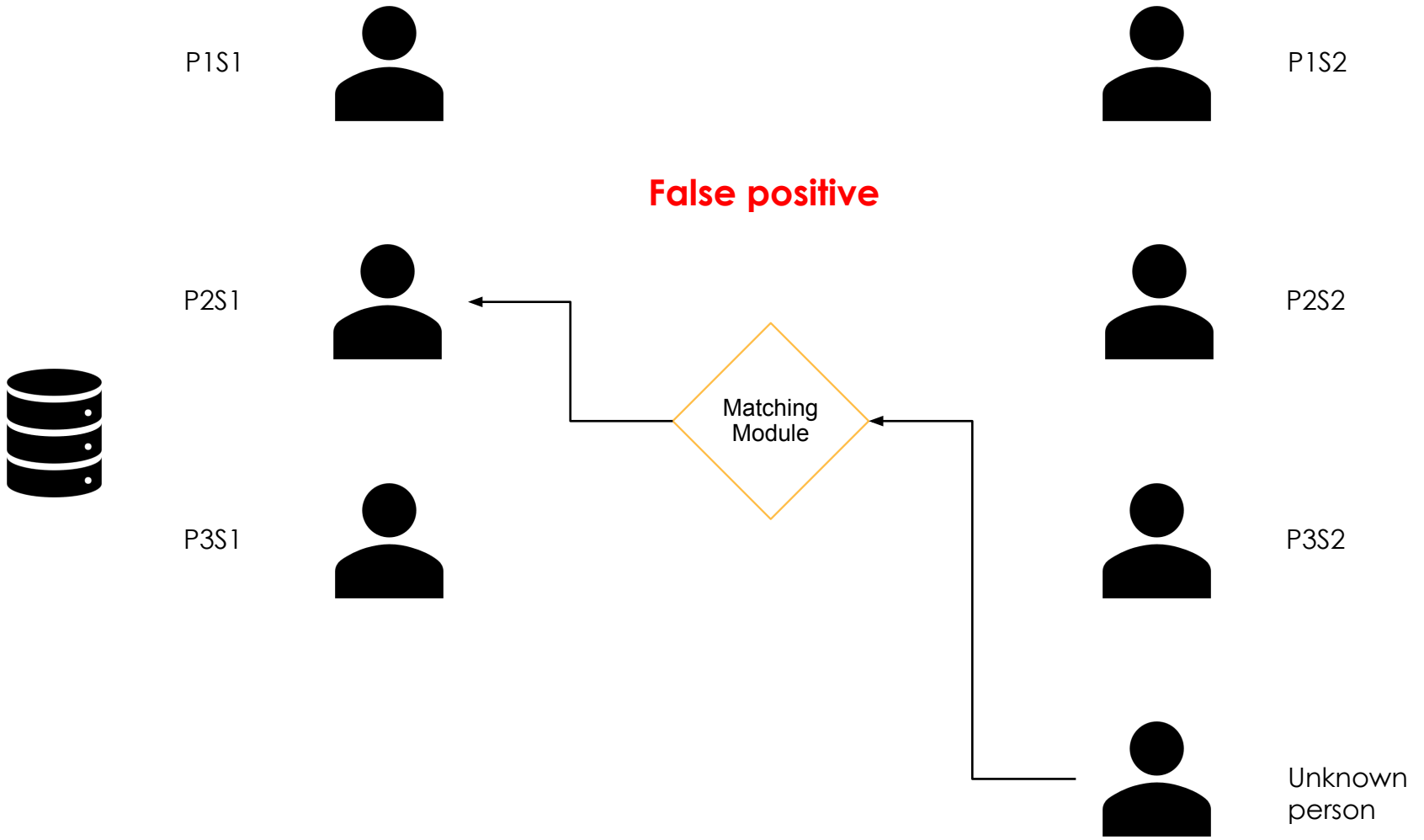
Four cases exist:
1. True positive (TP)
2. False positive (FP)



Test sample set 2

Identification mode
- Input: Sample
- Output: Identity

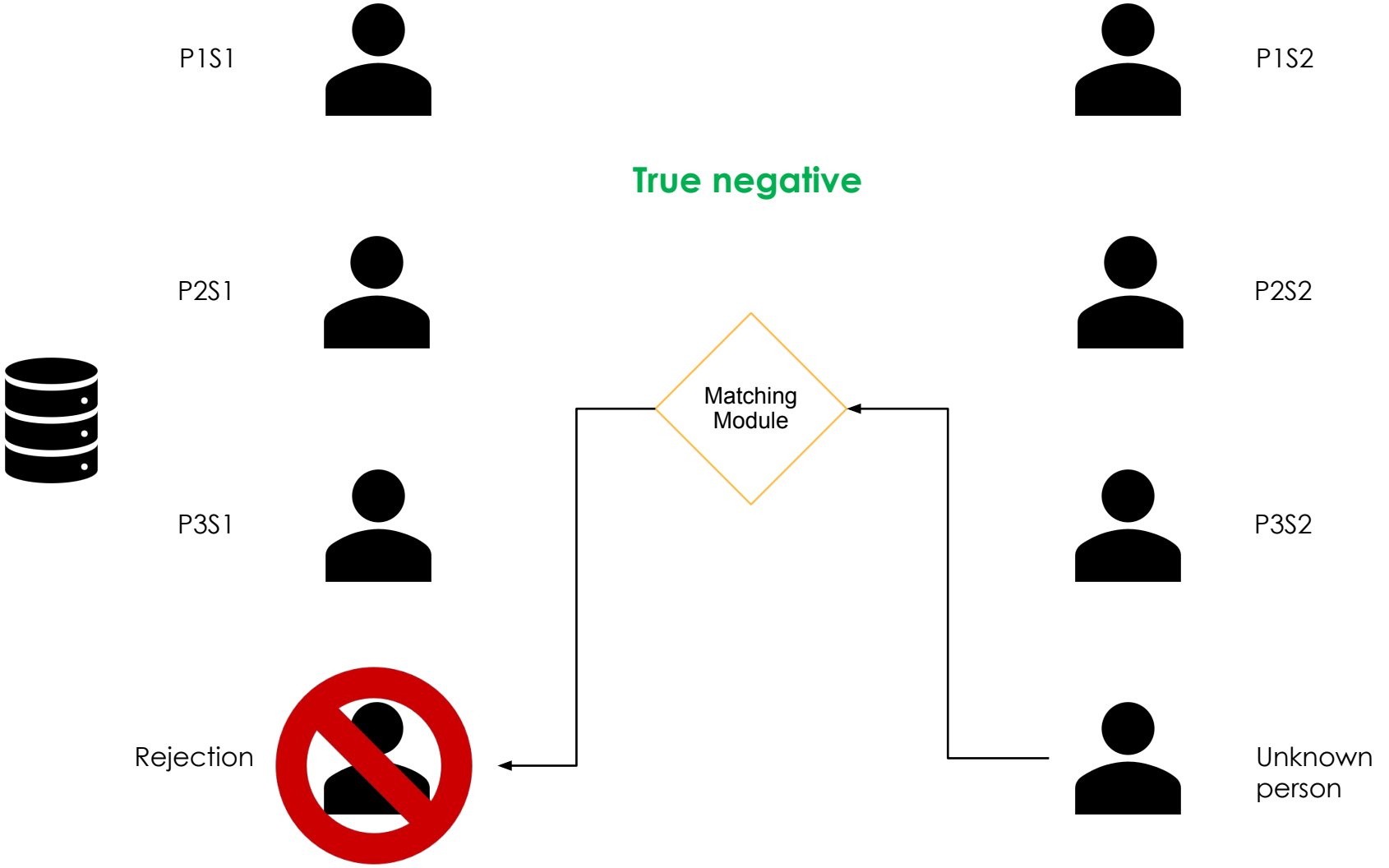
Four cases exist:
1. True positive (TP)
2. False positive (FP)



Test sample set 2

Identification mode
- Input: Sample
- Output: Identity

Four cases exist:
1. True positive (TP)
2. False positive (FP)
3. True negative (TN)

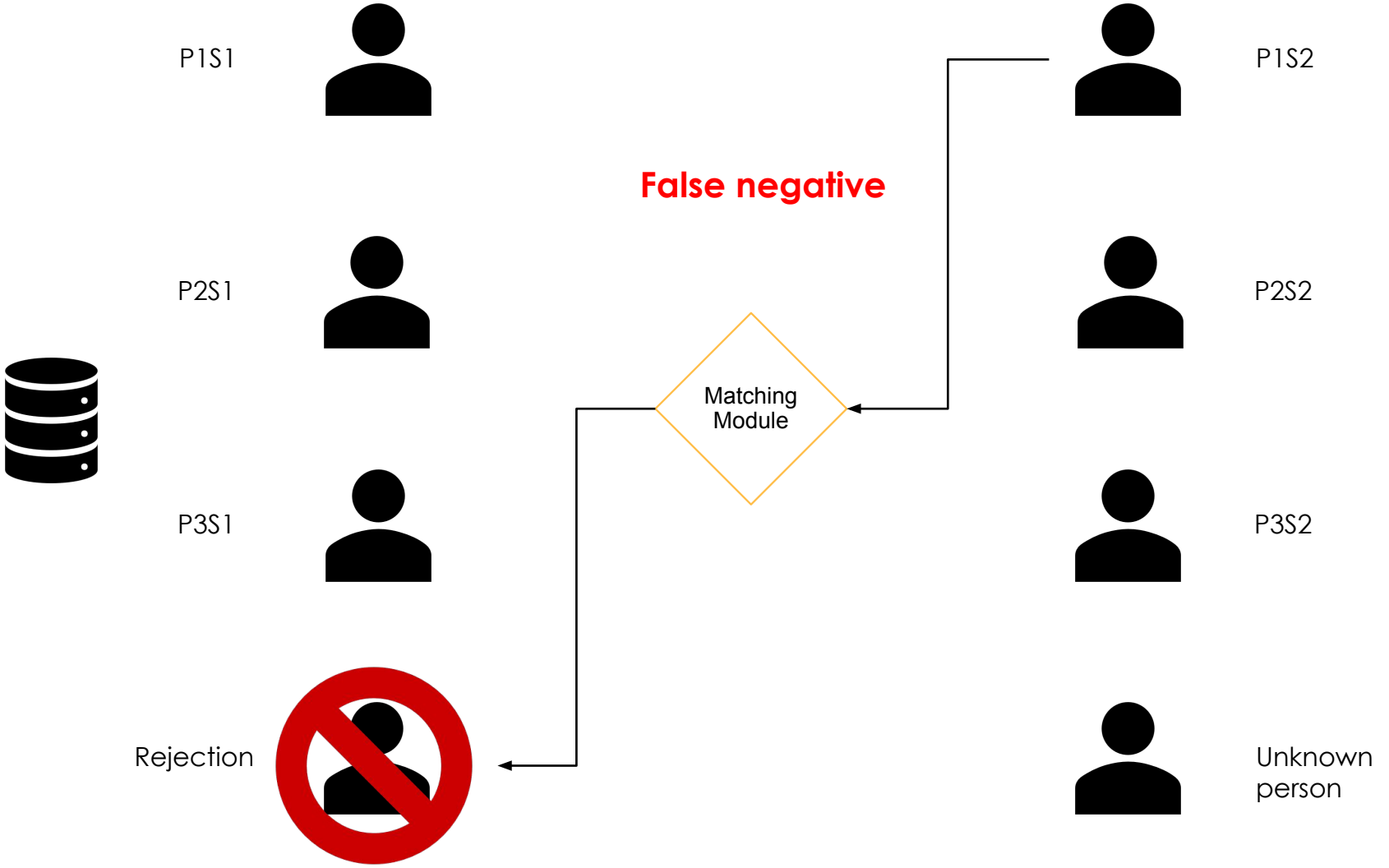


Test sample set 2

- Identification mode
  - Input: Sample
  - Output: Identity

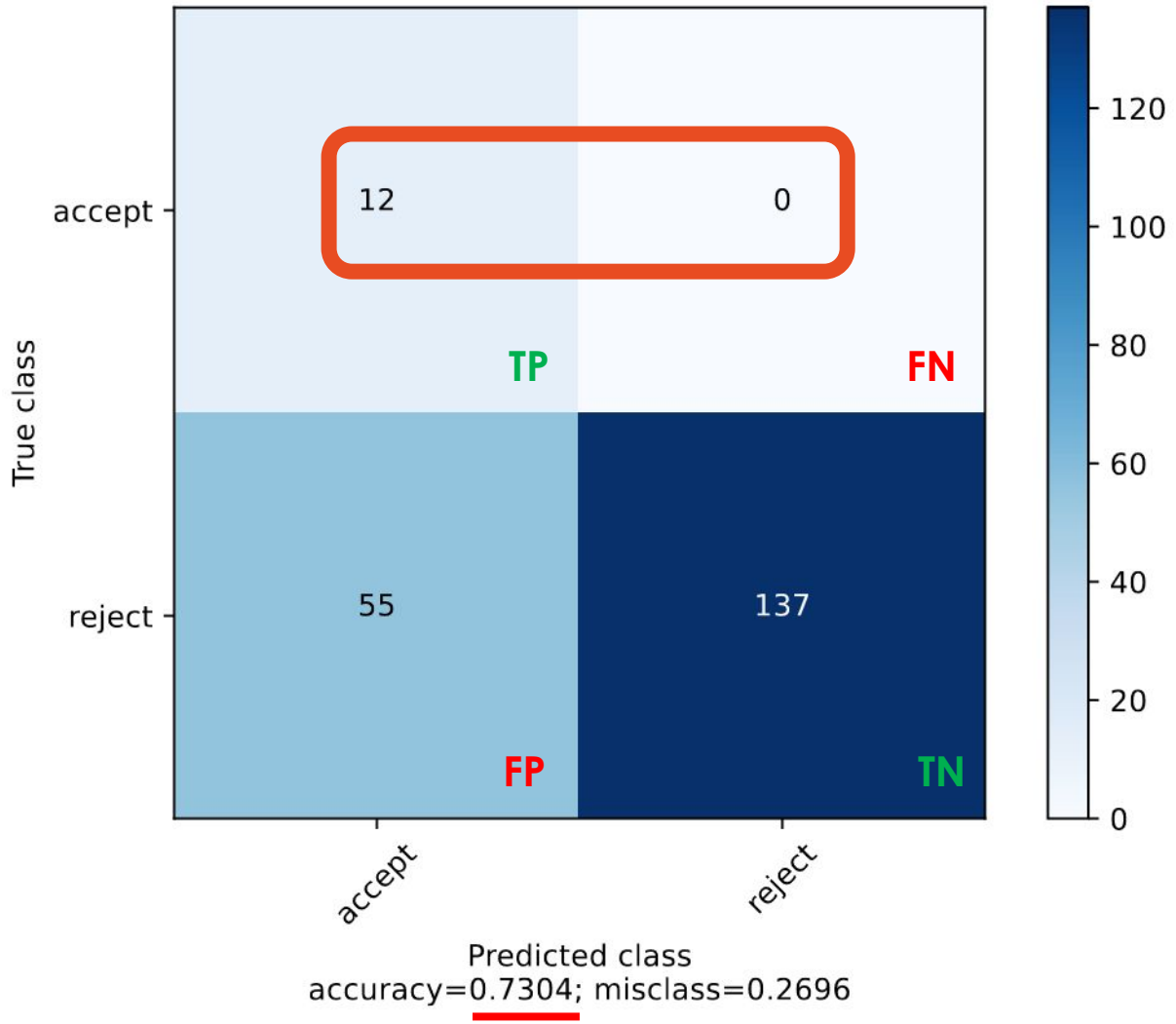
Four cases exist:

1. True positive (TP)
2. False positive (FP)
3. True negative (TN)
4. False negative (FN)



Measure performance

- Test all samples of set 2 against the enrolled data.
- Count TP, FP, TN, FN in a contingency table.
- Appropriately calculate rates and metrics.
- Watch out for pitfalls.
  - E.g., high accuracy due to imbalance. Here, almost nobody got accepted.
  - Even zero accepts can have a high accuracy. Always check table.



Biometric Modes

Implementation, data, and evaluation protocol

Mode: Verification

- User's claimed identity is verified by accepting or rejecting the provided sample.
- Input: biometric sample + claim of identity; output: accept / reject.

- Implementation: binary classifier; distance metric.
- Base chance of a correct guess: 1/2

- Verification is suited for higher security requirements and is the de-facto standard on most devices.
- Sampling the claim of identity implicitly is challenging.

- "Authentication" often synonymously used for "verification".

Mode: Identification (Closed-Set)

- User's identity is determined by the provided biometric sample.
- Input: biometric sample; output: predicted identity.

- Implementation: multiclass-classifier; distance metric.
- Base chance of a correct guess: 1/N (N = number of classes)

- Limitation: always picks one identity – no reject possible!
- Suitable for implicit identification, as no claim-of-identity is required.

Mode: Identification (Open-Set)

- User's identity is determined by the provided sample or user is rejected.
- Input: biometric sample; output: predicted identity or reject.

- Implementation: multiclass-classifier with anomaly detection; or a distance metric.
- Base chance of a correct guess: 1/(N+1) (N = number of classes)

- Suitable for implicit identification, as no claim-of-identity is required.
- "Ideal" from a user-centered perspective.

Closed-Set vs. Open-Set Identification

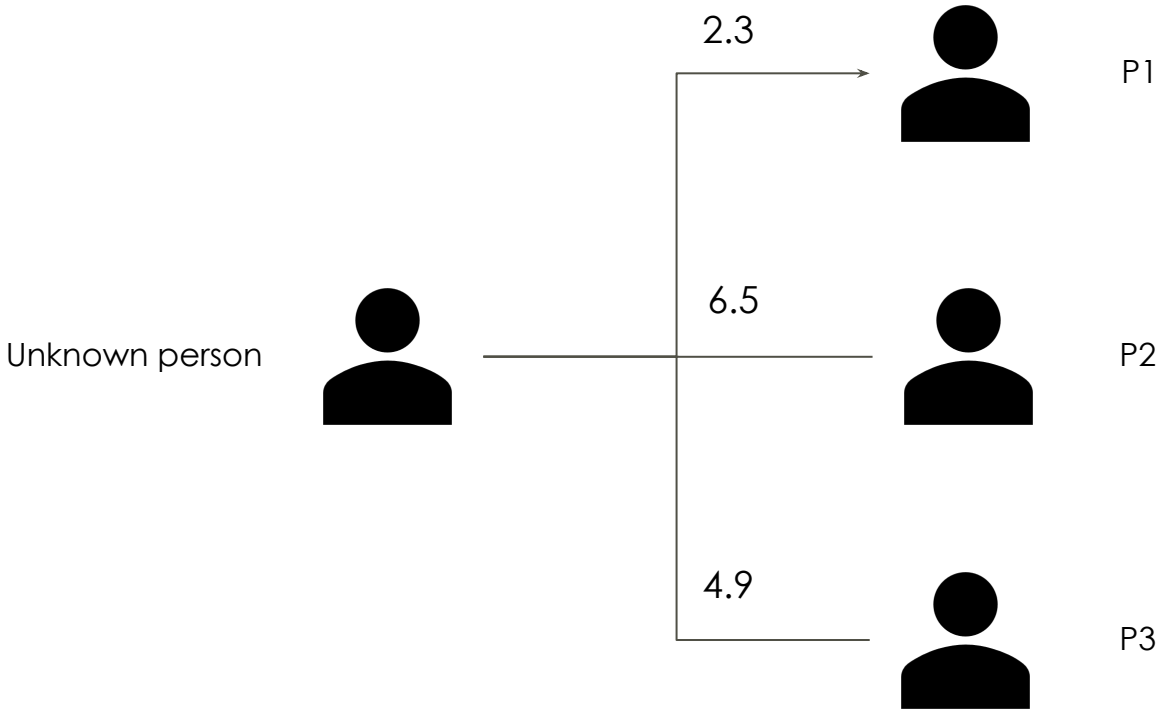
- To implement identification systems, one can use a multiclass-classifier or a distance-metric.

- Classifiers cannot reject data.
- Thereby, all users are automatically accepted.

- E.g., small office setting and voice assistant:
  - Voice assistant knows all four office workers and can identify them.
  - Guests to the office are, however, also always identified as one of the four.

- Anomaly detection required, to reject unknown users.

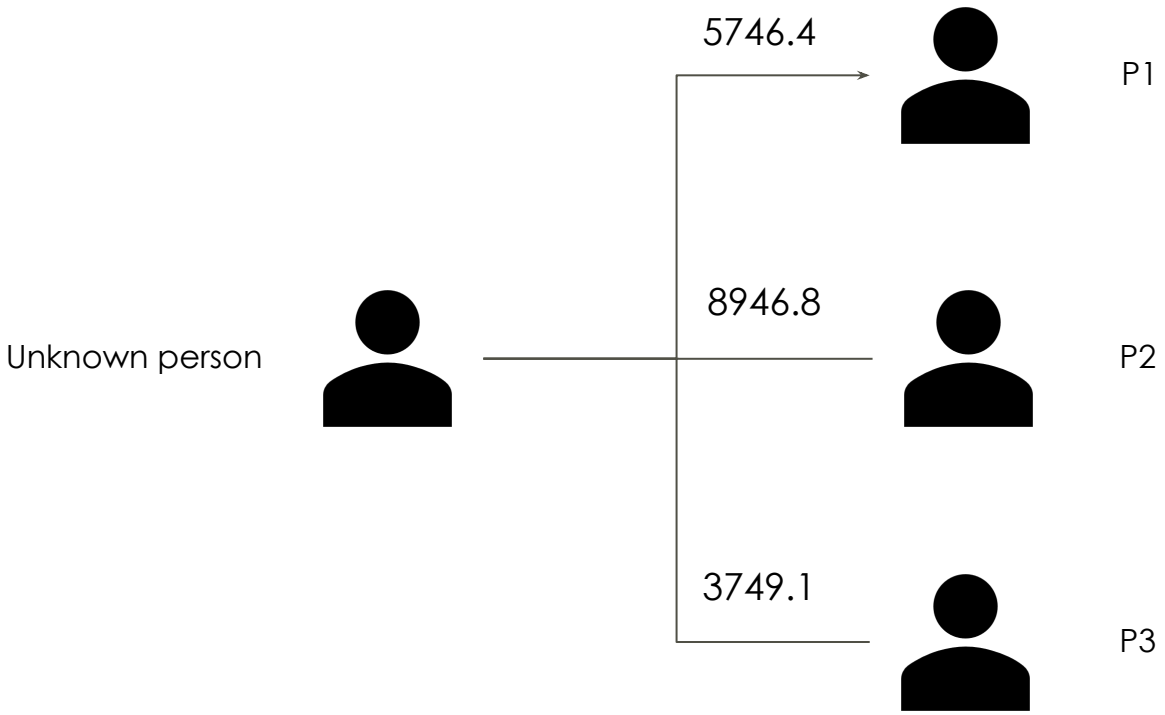
Example: Identification with Reject



- Annotations show absolute similarity.
- "0" means identity.
- The higher the value, the higher the degree of dissimilarity.

- Unknown Person is most likely P1, since their distance is *only* 2.3.

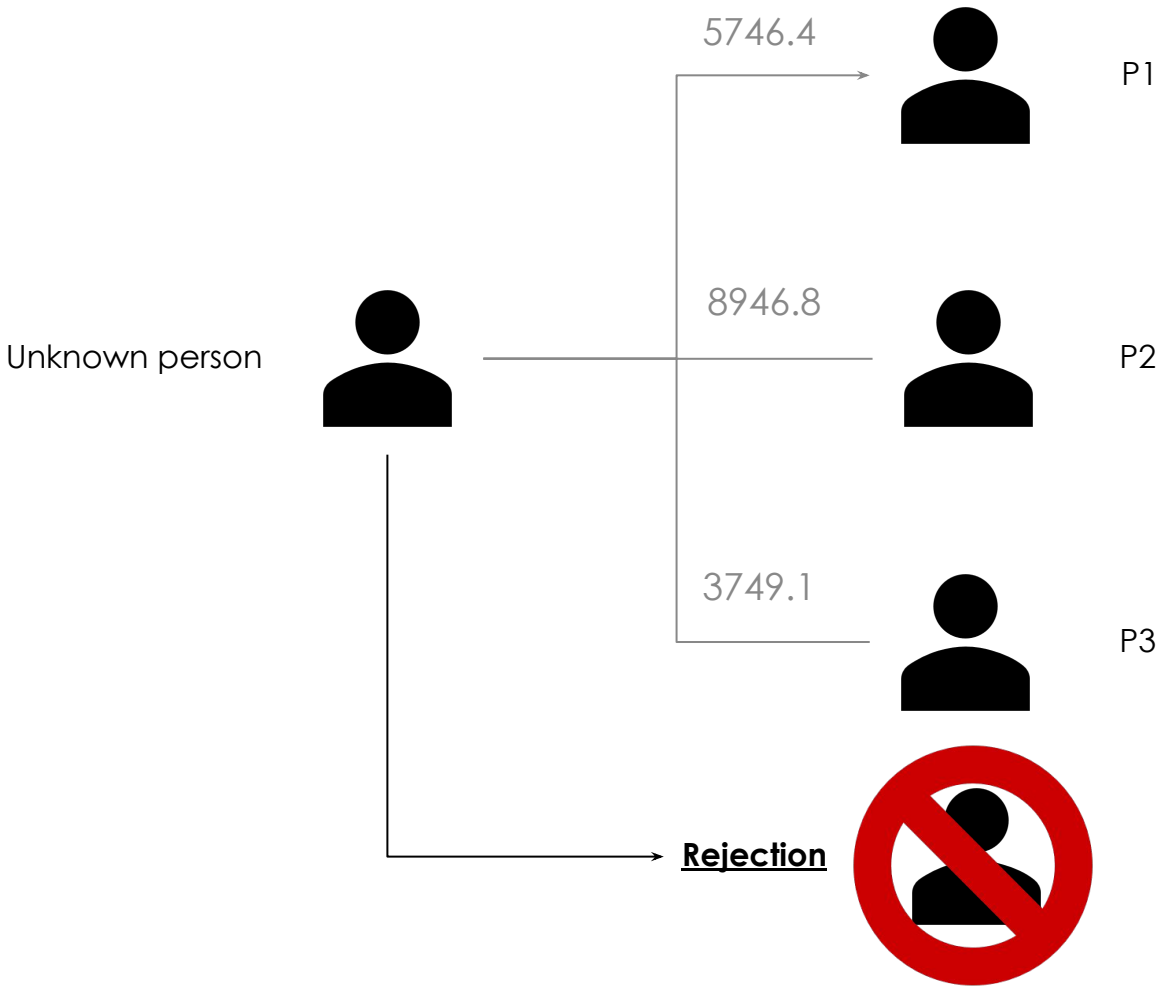
Example: Identification with Reject



- Annotations show absolute similarity.
- "0" means identity.
- The higher the value, the higher the degree of dissimilarity.

- What if all distances are large?
- Should the person be accepted?

Example: Identification with Reject

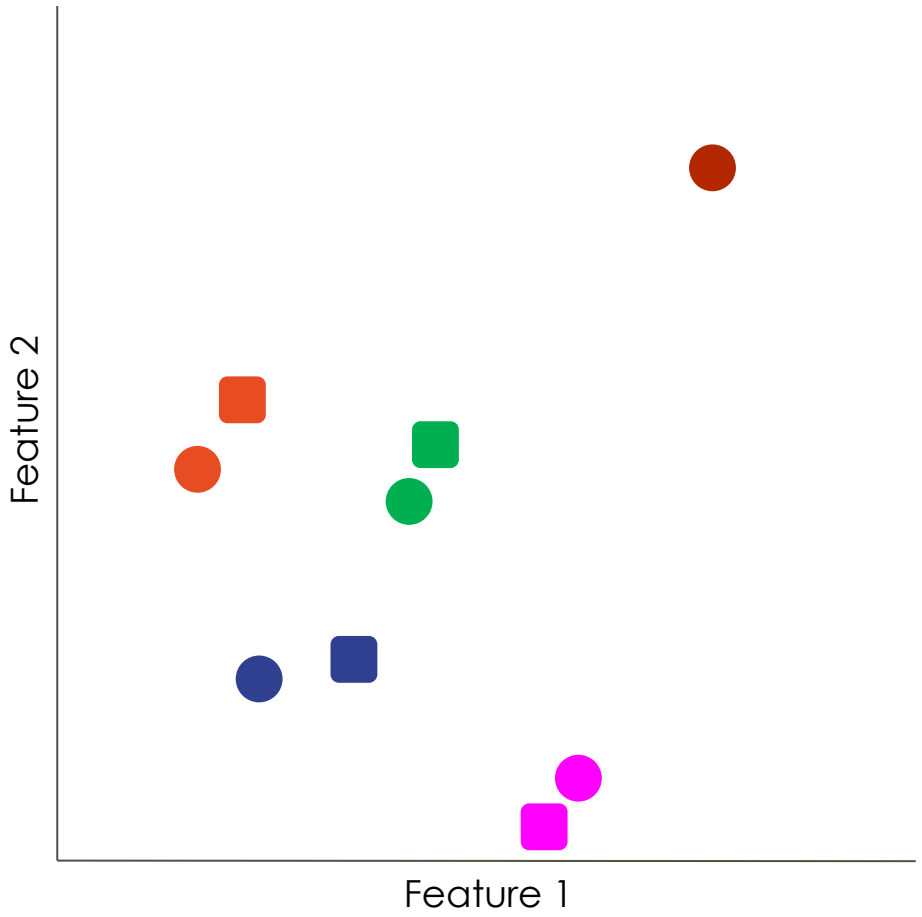


- Annotations show absolute similarity.
- "0" means identity.
- The higher the value, the higher the degree of dissimilarity.

- What if all distances are large?
- Should the person be accepted?

- Reject (= deny access), since it is an outlier.
- A threshold value is needed to meet this decision.

Outlier Rejection



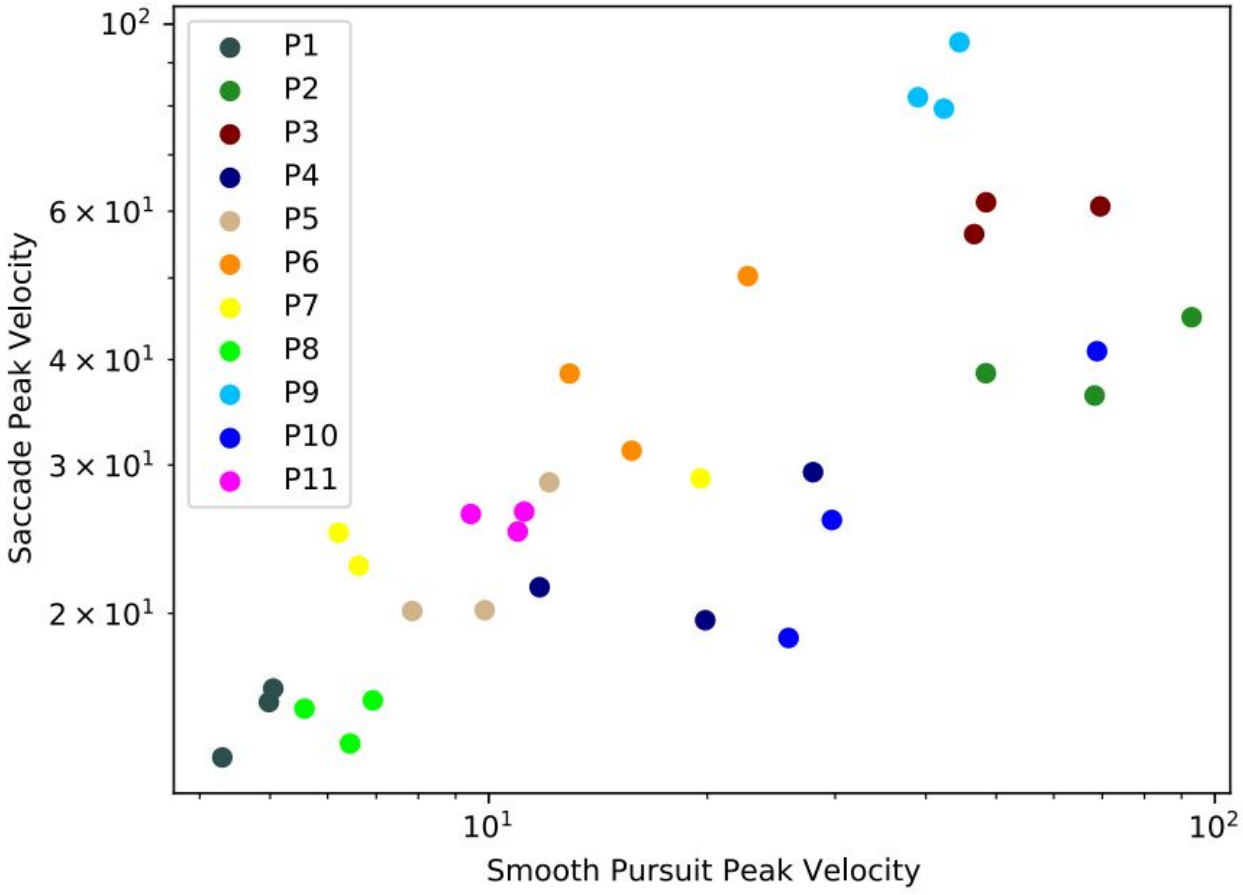
- Imagine a 2D-space with two features that describe each sample.

- Colors denote user identities.
- Square = enrolled sample.
- Circle = identified sample.

- Ideally, per-user pairs are created (disjunct).

- Outliers are further away.

Gaze Example



- Study on gaze behavior in VR.
- Three samples were captured, leading to clusters.
- Overlap exists.

Liebers, J., Horn, P., Burschik, C., Gruenefeld, U., and Schneegass, S. 2021. Using Gaze Behavior and Head Orientation for Implicit Identification in Virtual Reality. In Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology. VRST '21. Association for Computing Machinery, New York, NY, USA, 1–9. https://dx.doi.org/10.1145/3489849.3489880

Thresholding

- Often, the difference between an accept and reject depends on a threshold value.
- The value is optimized.

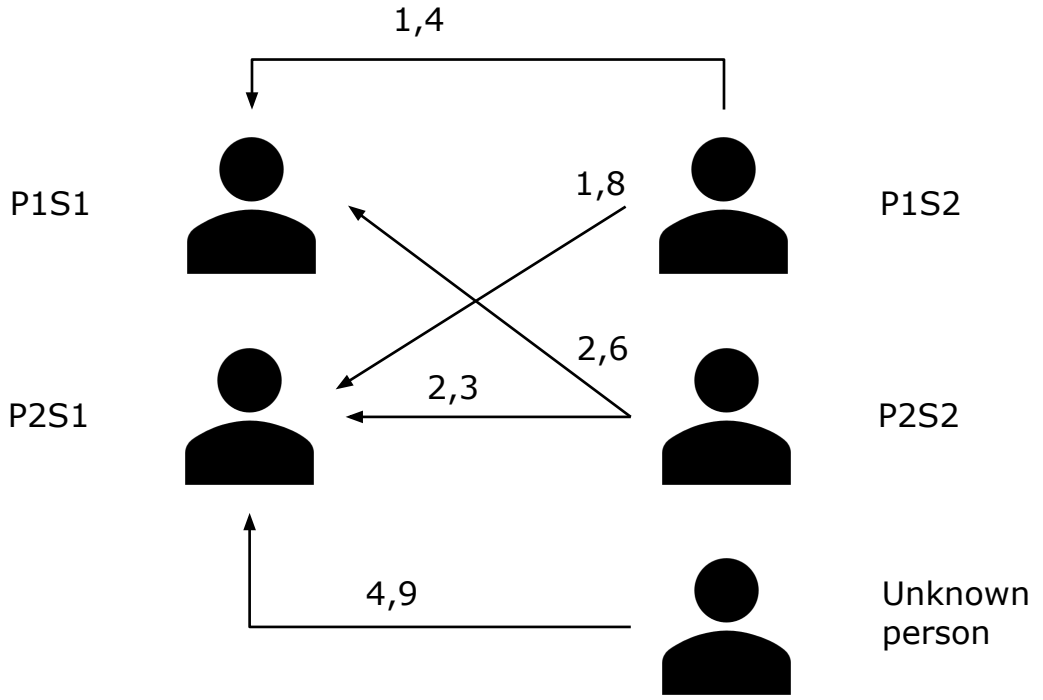
- Thresholding is an important technique to vary a value so that the error-rates (false reject and false accept) play against each other.
- Setting a good threshold is a tradeoff:
  - Low thresholds accept too many illegitimate samples.
  - High thresholds reject too many samples that should be accepted.

Thresholding - ctd.

Acceptance: Distance <= Threshold
Rejection: Distance > Threshold

One ideal threshold in the example: 2,3

- All TPs are still valid.
- The unknown person is rejected (TN).



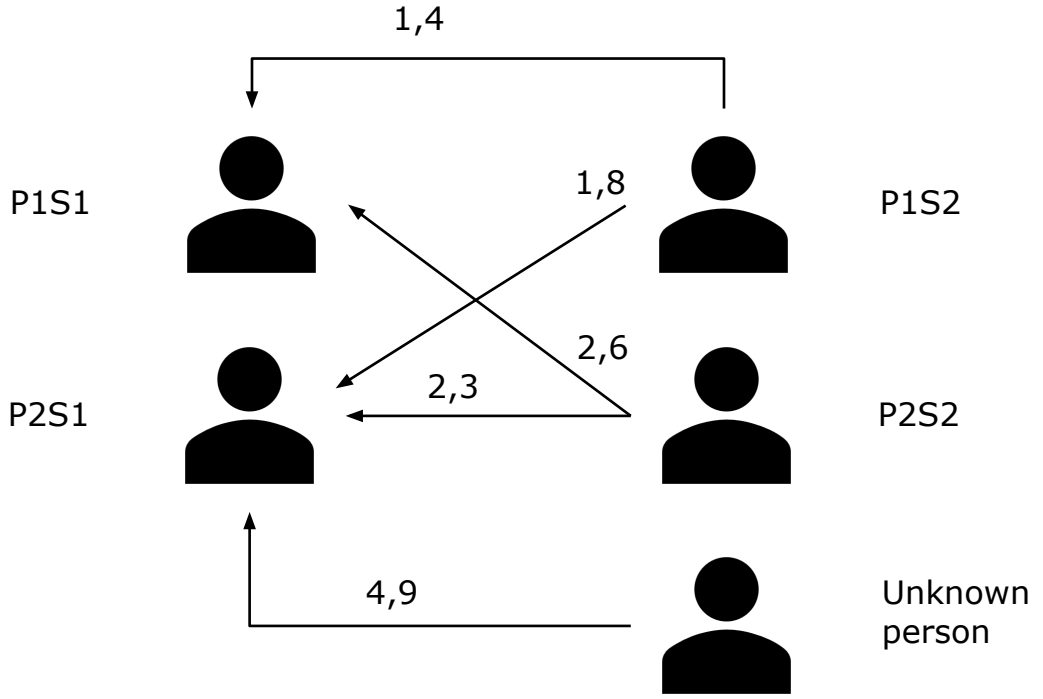
Thresholding - ctd.

Acceptance: Distance <= Threshold
Rejection: Distance > Threshold

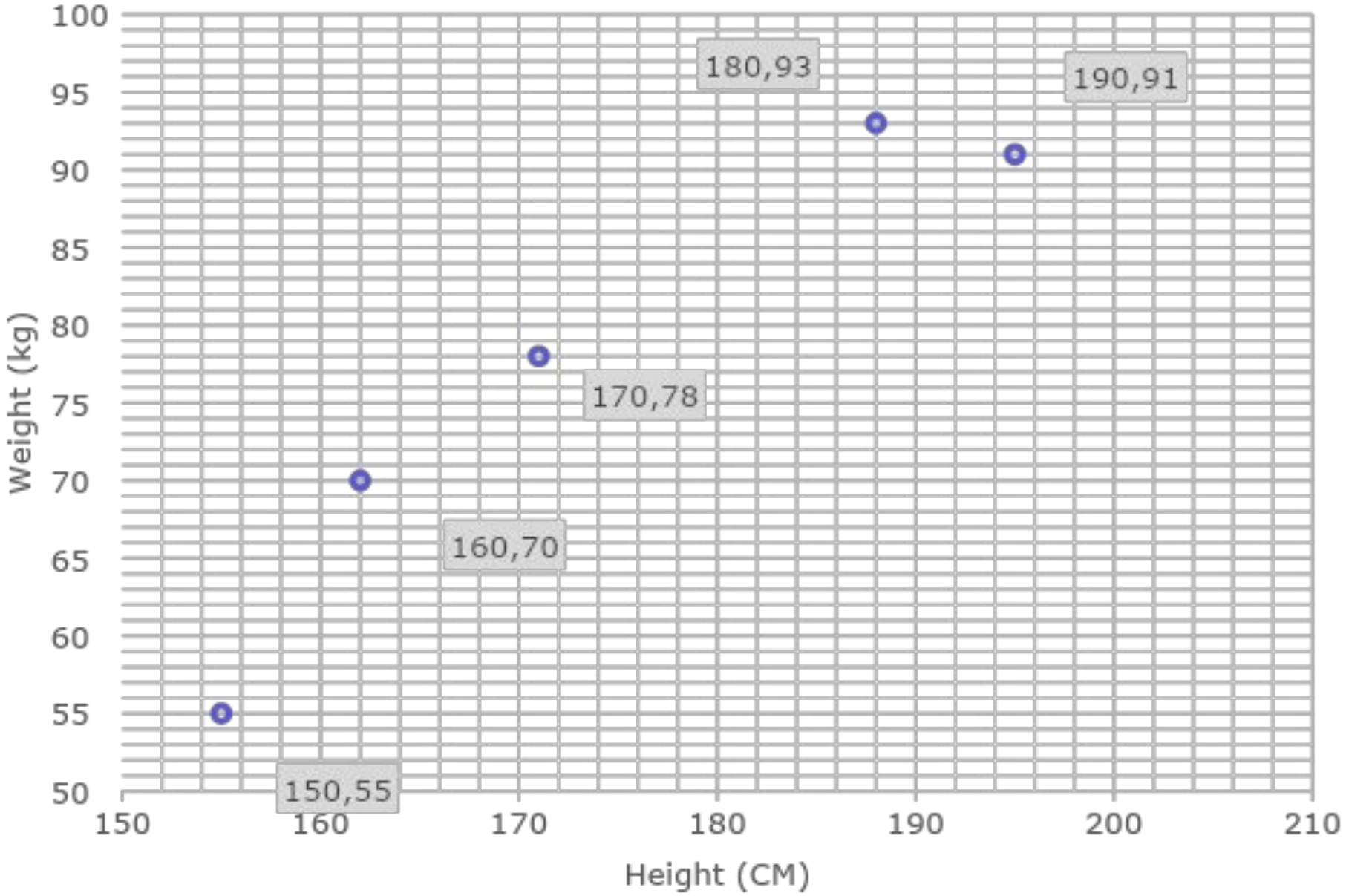
One ideal threshold in the example: 2,3

- All TPs are still valid.
- The unknown person is rejected (TN).

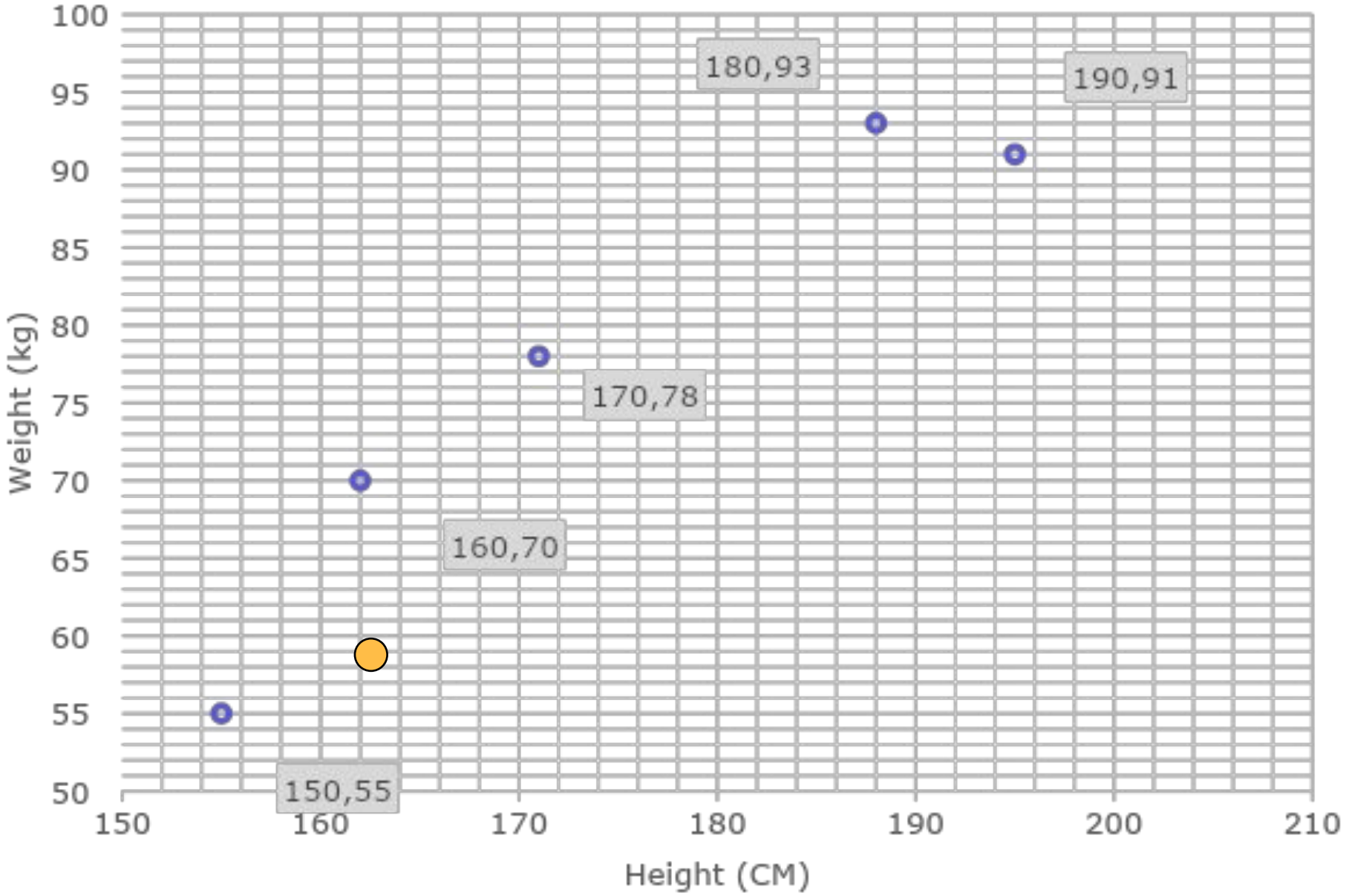
- Simple form of outlier detection.



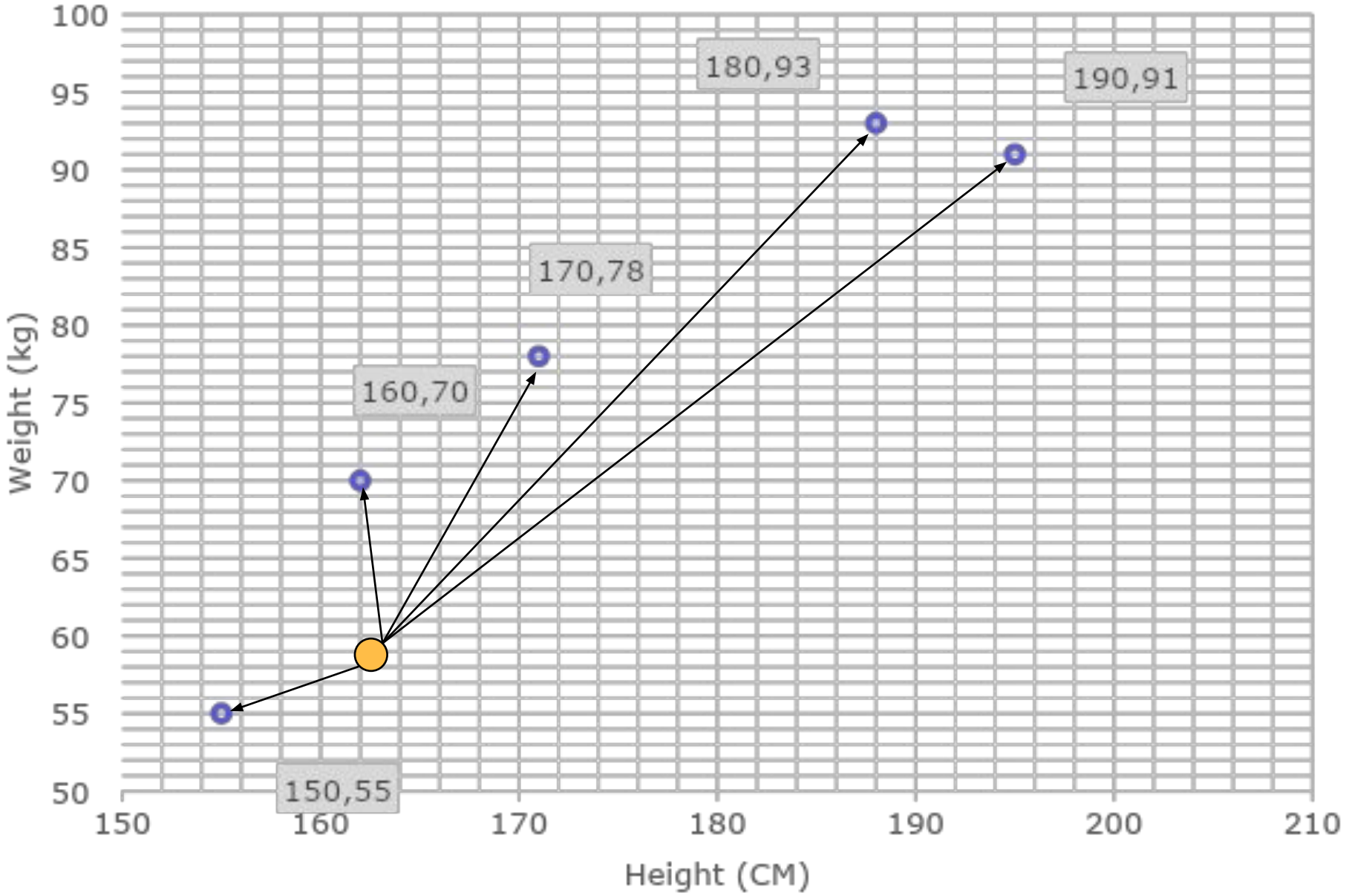
Example


100
95
90
85
80
75
70
65
60
55
50
Weight (kg)
150
160
170
180
190
200
210
Height (CM)
180,93
190,91
170,78
160,70
150,55

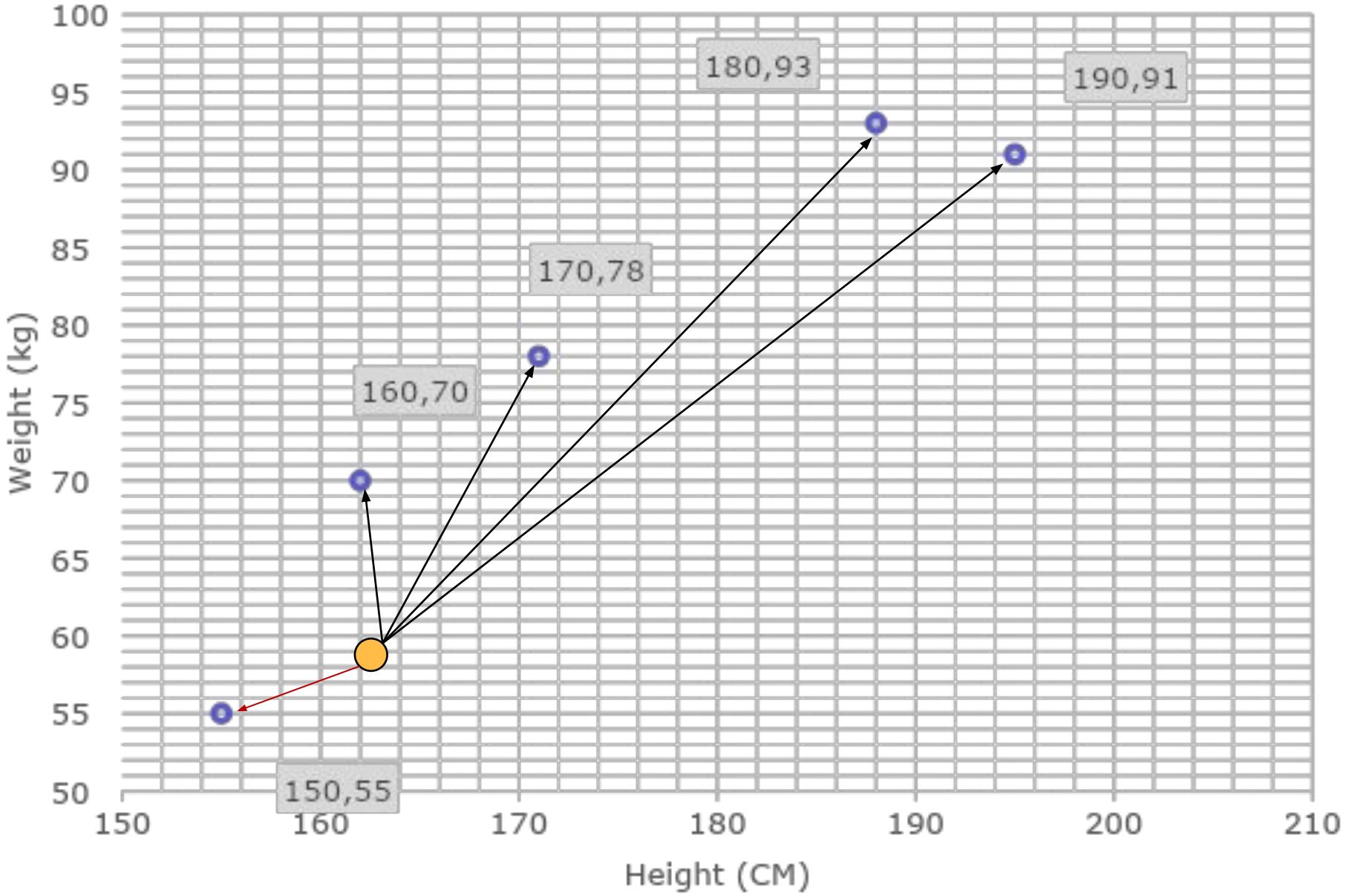
Example


100
95
90
85
80
75
70
65
60
55
50
Weight (kg)
150
160
170
180
190
200
210
Height (CM)
180,93
190,91
170,78
160,70
150,55

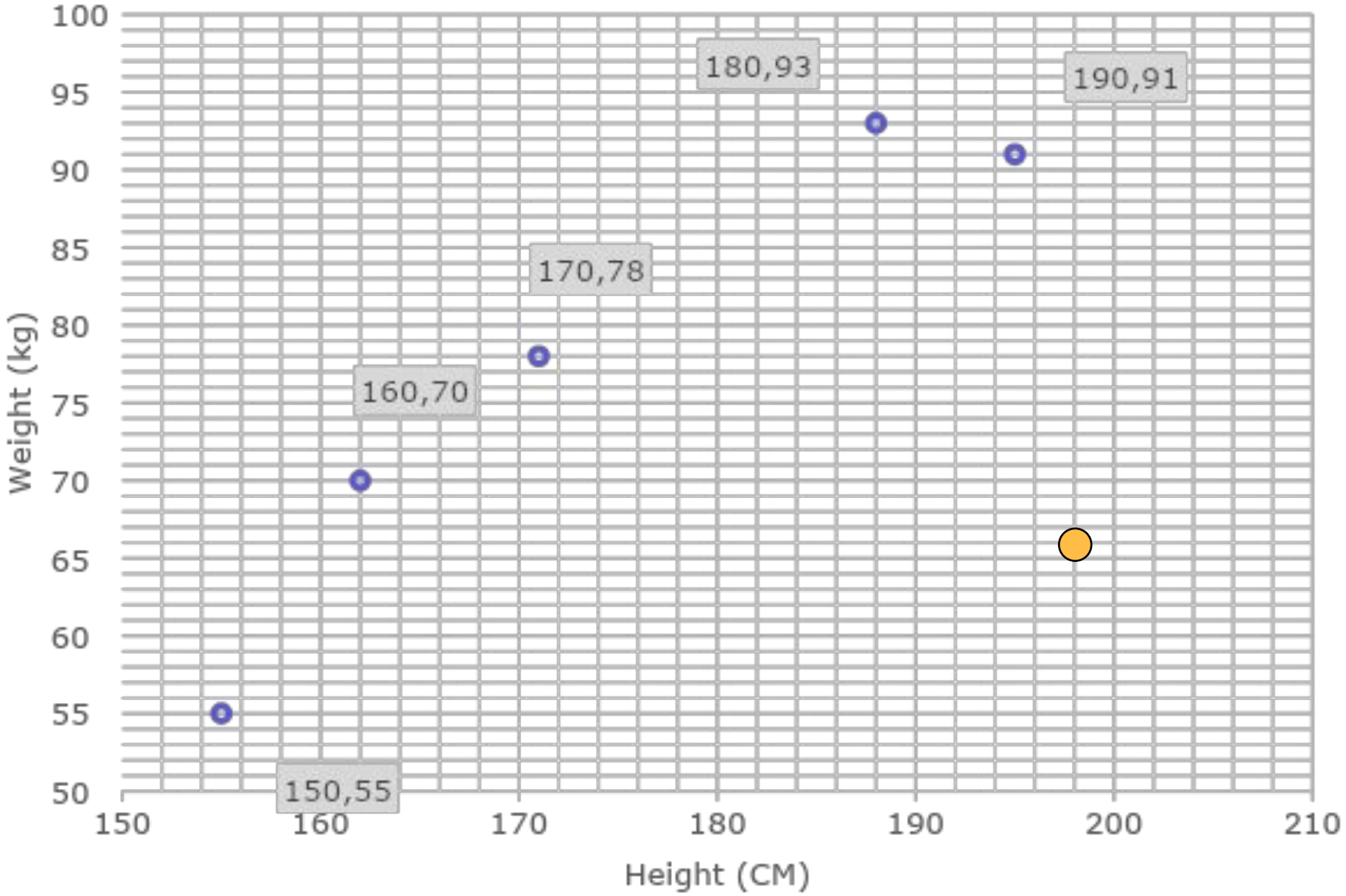
Example


100
95
90
85
80
75
70
65
60
55
50
Weight (kg)
150
160
170
180
190
200
210
Height (CM)
180,93
190,91
170,78
160,70
150,55

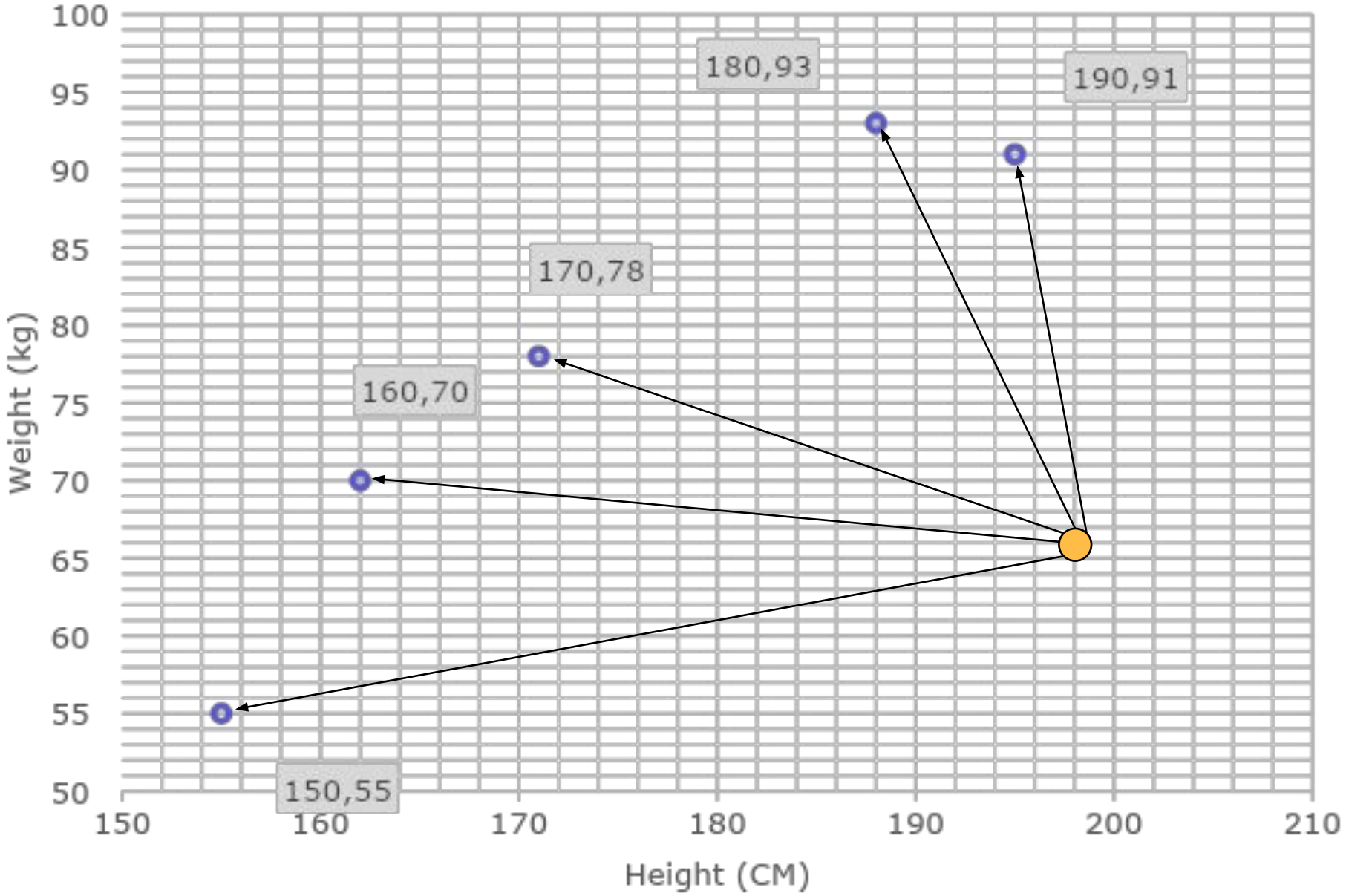
Example


100
95
90
85
80
75
70
65
60
55
50
Weight (kg)
150
160
170
180
190
200
210
Height (CM)
180,93
190,91
170,78
160,70
150,55

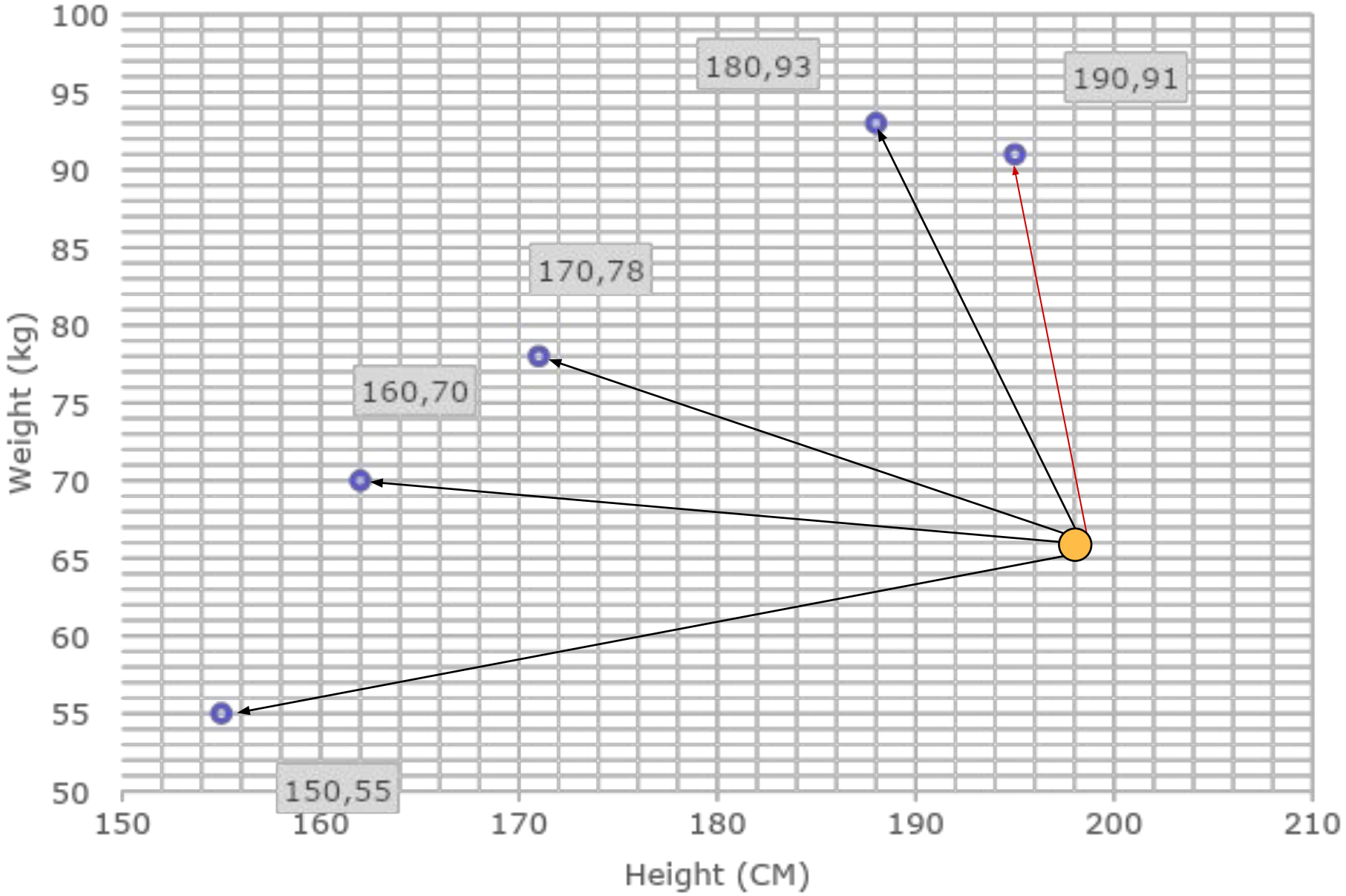
Example


100
95
90
85
80
75
70
65
60
55
50
Weight (kg)
150
160
170
180
190
200
210
Height (CM)
180,93
190,91
170,78
160,70
150,55

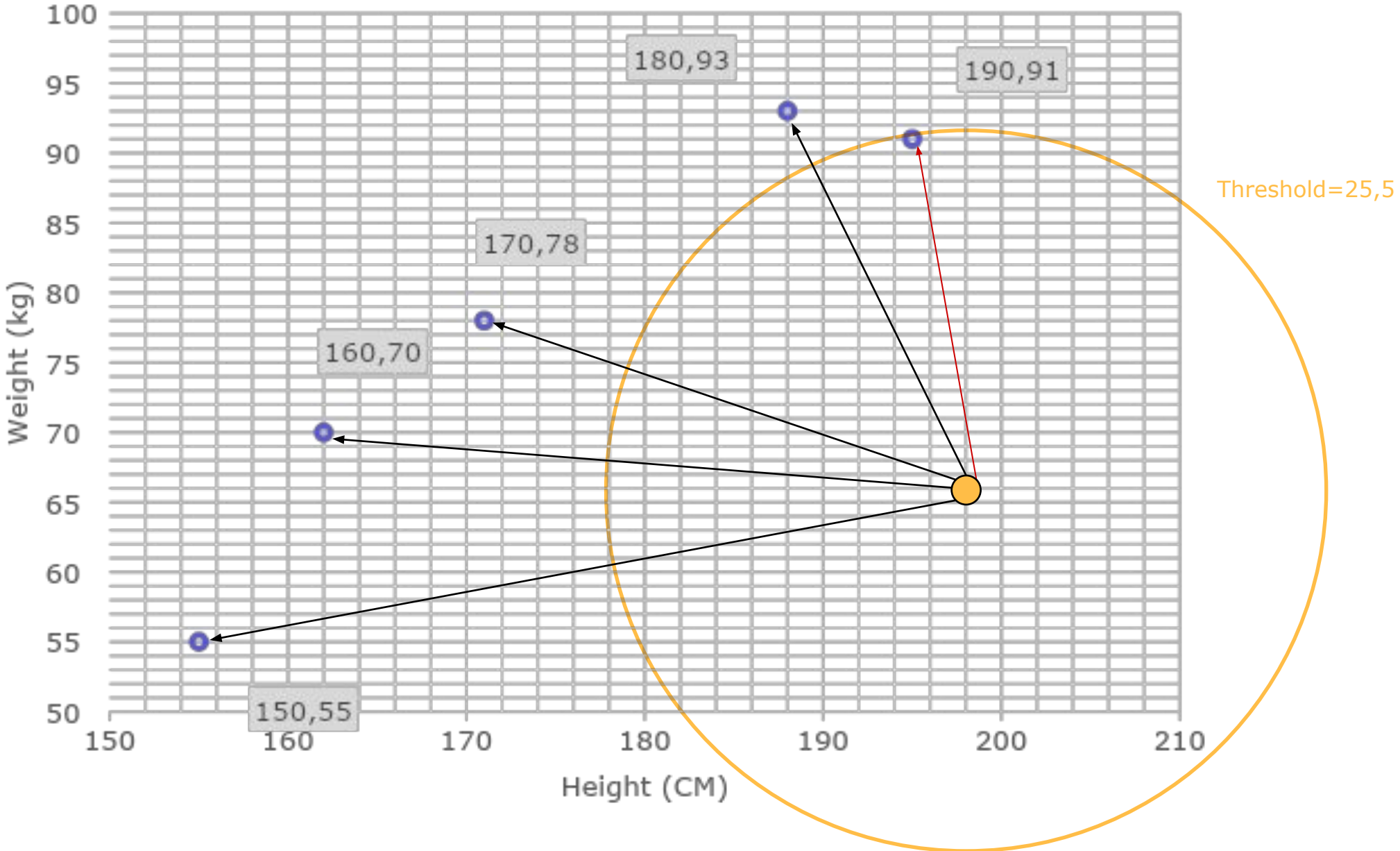
Example


100
95
90
85
80
75
70
65
60
55
50
Weight (kg)
150
160
170
180
190
200
210
Height (CM)
180,93
190,91
170,78
160,70
150,55

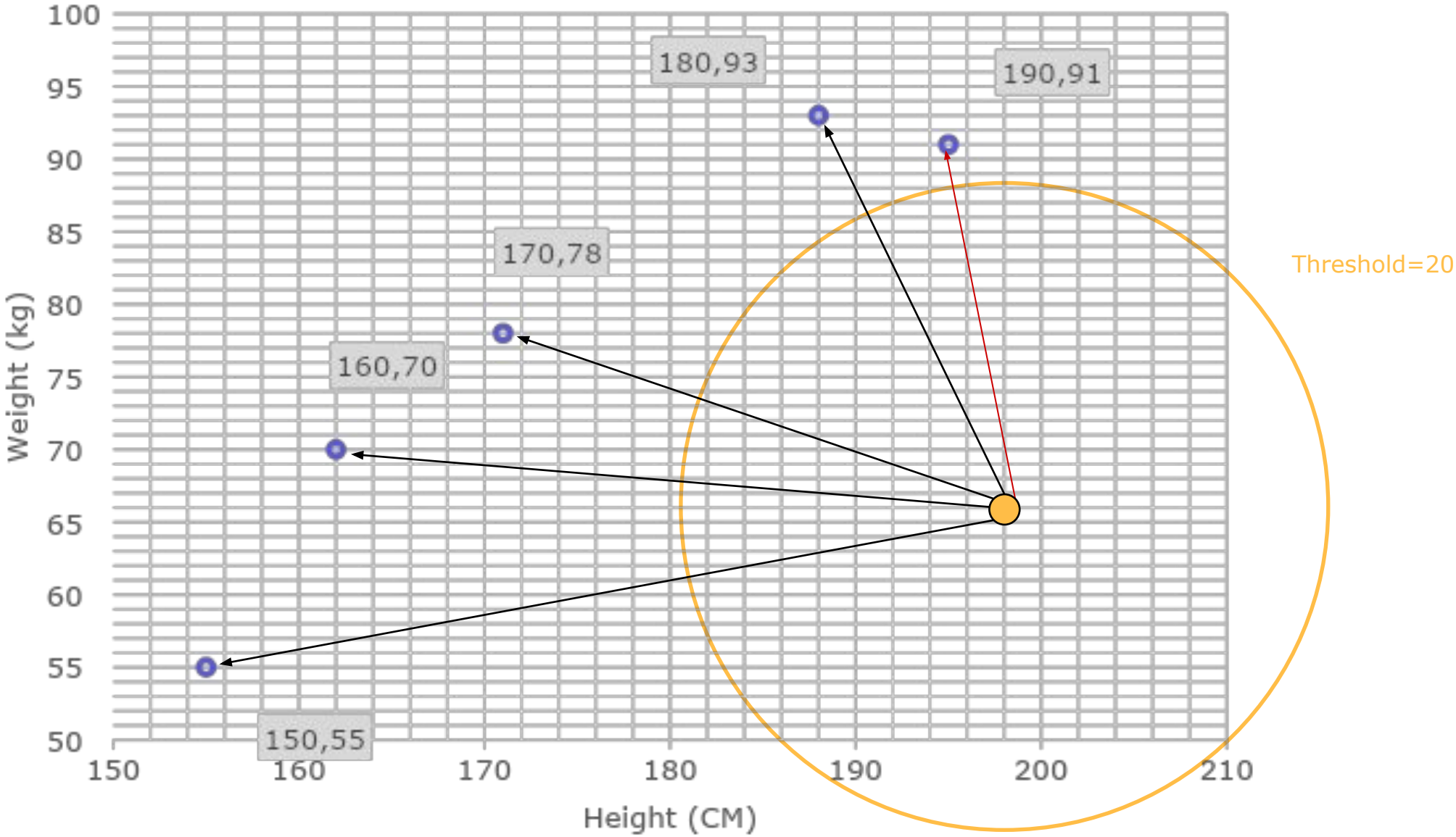
Example


100
95
90
85
80
75
70
65
60
55
50
Weight (kg)
150
160
170
180
190
200
210
Height (CM)
180,93
190,91
170,78
160,70
150,55

Example


100
95
90
85
80
75
70
65
60
55
50
Weight (kg)
150
160
170
180
190
200
210
Height (CM)
180,93
190,91
170,78
160,70
150,55
Threshold=25,5

Example


100
95
90
85
80
75
70
65
60
55
50
Weight (kg)
150
160
170
180
190
200
210
Height (CM)
180,93
190,91
170,78
160,70
150,55
Threshold=20

Time Series Handling

- Recordings of behavior are time-series, hence they have a varying length.
- Machine Learning requires constant shape.

- One way of unifying shape: running aggregate functions on each column
- E.g., min, max, mean, SD

Sample 1

Timestamp	Sensor1	Sensor2
1	1.42	0.43
2	1.41	0.44
…	…	…
87	1.48	0.46

Sample 2

Timestamp	Sensor1	Sensor2
1	1.39	0.28
2	1.37	0.26
…	…	…
132	1.34	0.26

min()

aggregated feature vector of sample

min(Sensor1)							

Time Series Handling

- Recordings of behavior are time-series, hence they have a varying length.
- Machine Learning requires constant shape.

- One way of unifying shape: running aggregate functions on each column
- E.g., min, max, mean, SD

Sample 1

Timestamp	Sensor1	Sensor2
1	1.42	0.43
2	1.41	0.44
…	…	…
87	1.48	0.46

Sample 2

Timestamp	Sensor1	Sensor2
1	1.39	0.28
2	1.37	0.26
…	…	…
132	1.34	0.26

max()

aggregated feature vector of sample

min(Sensor1)	max(Sensor1)						

Time Series Handling

- Recordings of behavior are time-series, hence they have a varying length.
- Machine Learning requires constant shape.

- One way of unifying shape: running aggregate functions on each column
- E.g., min, max, mean, SD

Sample 1

Timestamp	Sensor1	Sensor2
1	1.42	0.43
2	1.41	0.44
…	…	…
87	1.48	0.46

Sample 2

Timestamp	Sensor1	Sensor2
1	1.39	0.28
2	1.37	0.26
…	…	…
132	1.34	0.26

mean()

aggregated feature vector of sample

min(Sensor1)	max(Sensor1)	mean(Sensor1)					

Time Series Handling

- Recordings of behavior are time-series, hence they have a varying length.
- Machine Learning requires constant shape.

- One way of unifying shape: running aggregate functions on each column
- E.g., min, max, mean, SD

Sample 1

Timestamp	Sensor1	Sensor2
1	1.42	0.43
2	1.41	0.44
…	…	…
87	1.48	0.46

Sample 2

Timestamp	Sensor1	Sensor2
1	1.39	0.28
2	1.37	0.26
…	…	…
132	1.34	0.26

SD()

aggregated feature vector of sample

min(Sensor1)	max(Sensor1)	mean(Sensor1)	SD(Sensor1)				

Time Series Handling

- Recordings of behavior are time-series, hence they have a varying length.
- Machine Learning requires constant shape.

- One way of unifying shape: running aggregate functions on each column
- E.g., min, max, mean, SD

Sample 1

Timestamp	Sensor1	Sensor2
1	1.42	0.43
2	1.41	0.44
…	…	…
87	1.48	0.46

Sample 2

Timestamp	Sensor1	Sensor2
1	1.39	0.28
2	1.37	0.26
…	…	…
132	1.34	0.26

…

aggregated feature vector of sample

min(Sensor1)	max(Sensor1)	mean(Sensor1)	SD(Sensor1)	min(Sensor2)	max(Sensor2)	mean(Sensor2)	SD(Sensor2)

Data Splits: Train, Validation, and Test

For training and evaluation a matching module, the elicited data set should be split into three parts:

1. **Training data:** data exclusively used for training.
   - This is the only data the system should see and learn from.

2. **Validation data:** data for parameter validation.
   - Validation data is used for experimentation and to determine parameters.

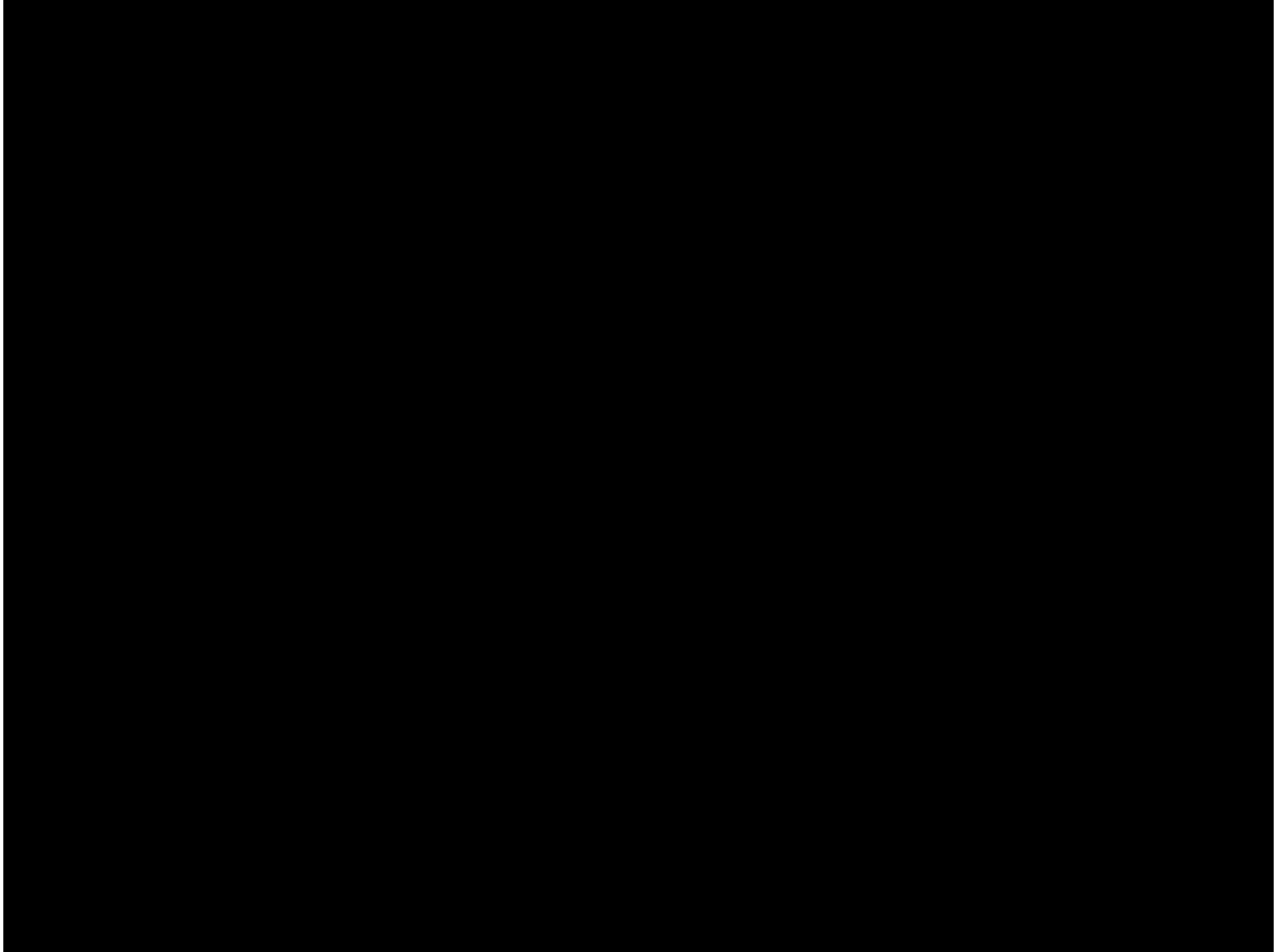
3. **Test data**:
   - All performance metrics are calculated on this test data.
   - Should be rarely (ideally: only once) be used for determining metrics.

Coffee Break

Hands On: Data Collection 2!

Please pick up the VR headset.

This time, try to vary your movements in 2 or 3 of the shots significantly (e.g., crouch, angle the bow, sidestep, …)



Evaluation Metrics

Metrics: Rates

Basic Rates:

- TPR: True Positive Rate
- TNR: True Negative Rate
- FPR: False Positive Rate
- FNR: False Negative Rate
- PPV: Positive Predictive Value

Synonyms:

- Sensitivity = Recall = TPR
- Precision = PPV
- Specificity = TNR
- False Acceptance Rate = FPR
- False Rejection Rate = FNR

$$TPR = \frac{\text{TP}}{\text{TP+FN}}$$

$$TNR = \frac{\text{TN}}{\text{TN+FP}}$$

$$FPR = \frac{\text{FP}}{\text{FP+TN}}$$

$$FNR = \frac{\text{FN}}{\text{FN+TP}}$$

$$PPV = \frac{\text{TP}}{\text{TP+FP}}$$

Metrics: Accuracy, F1-Score

- Accuracy and F1-Score are often desired for their expressiveness and simplicity.

- Warning 1: accuracy is sensitive to class-imbalance!
  - Imbalance, i.e., a varying number of samples, severely impacts expressiveness.
- Warning 2: accuracy is sensitive to ground-truth-balance!
  - If 90% of samples should be rejected, a "return False"-classifier will score 90% at the absence of any positive.

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

$$\text{F1-Score} = 2 * \frac{\mathrm{PPV} * \mathrm{TPR}}{\mathrm{PPV} + \mathrm{TPR}} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$$

Metrics: Matthews Correlation Coefficient

- Measure of quality for binary classifications
- Ranges from -1 to +1
  - +1: perfect agreement
  - 0: no relationship
  - -1: perfect disagreement

- Very practical metric for *verification* mode systems.

- More informative than F1 and accuracy in binary classification, because of the ratio balances.

$$\frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Chicco D, Jurman G (January 2020). "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation". BMC Genomics. 21 (1): 6-1–6-13.

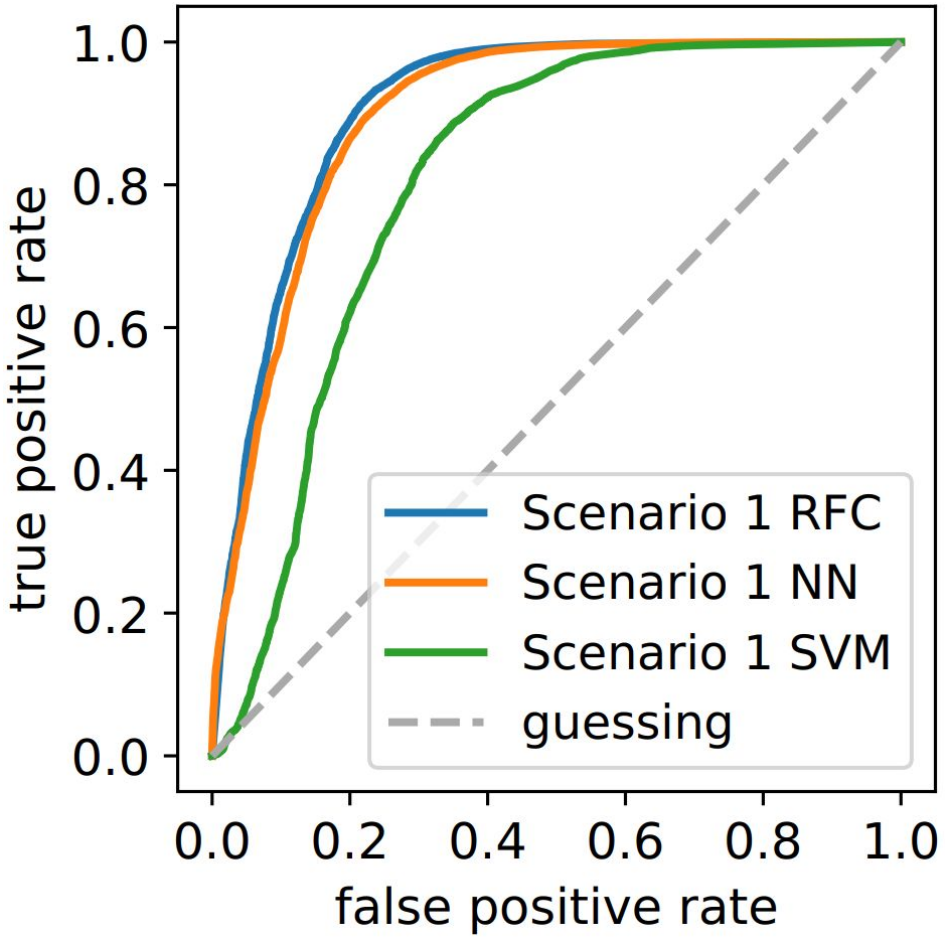
Metrics: Receiver Operating Characteristics (ROC)

- ROC plots one or more decision rules in one figure over various thresholds.
- Primarily used for verification.

- Y-axis: true positive rate
- X-axis: false positive rate

- Diagonal: random guessing of "accept" or "reject".

- The further the curve spans into the top left corner, the better.



Recommended reading: Fawcett, T. 2006. An introduction to ROC analysis. Pattern Recognition Letters 27, 8, 861–874.

Image from: Florian Lehmann and Daniel Buschek. Heartbeats in the Wild: A Field Study Exploring ECG Biometrics in Everyday Life. In Proc. CHI 2020

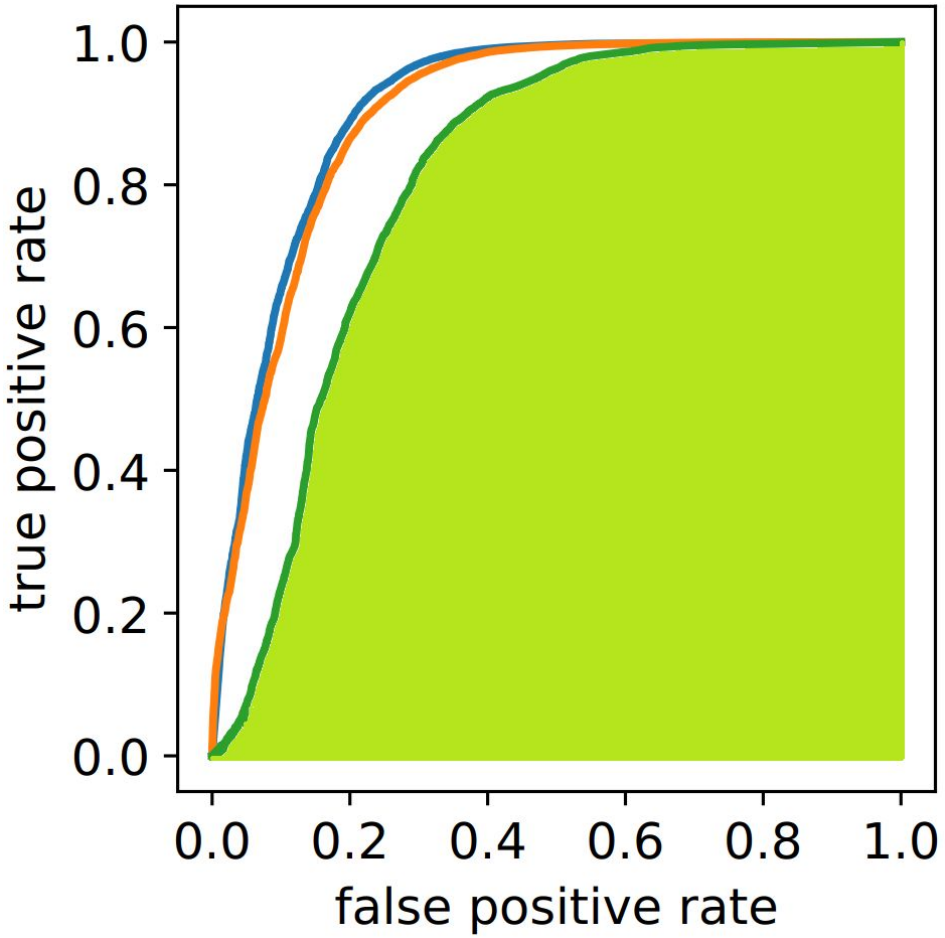
Metrics: Area under ROC (AUROC/AUC)

- Area under ROC (AUROC) and area under the curve (AUC) are synonyms.
- It is the square measure of the area under the respective curve.
- Primarily used for verification.

- The AUC for a randomly guessing decision rule is 0.5.
- For a perfect classifier, the AUC is 1.0.
- AUC range: [0; 1]

Example in Figure:

- Dark green is the ROC curve.
- Light green is its AUC.

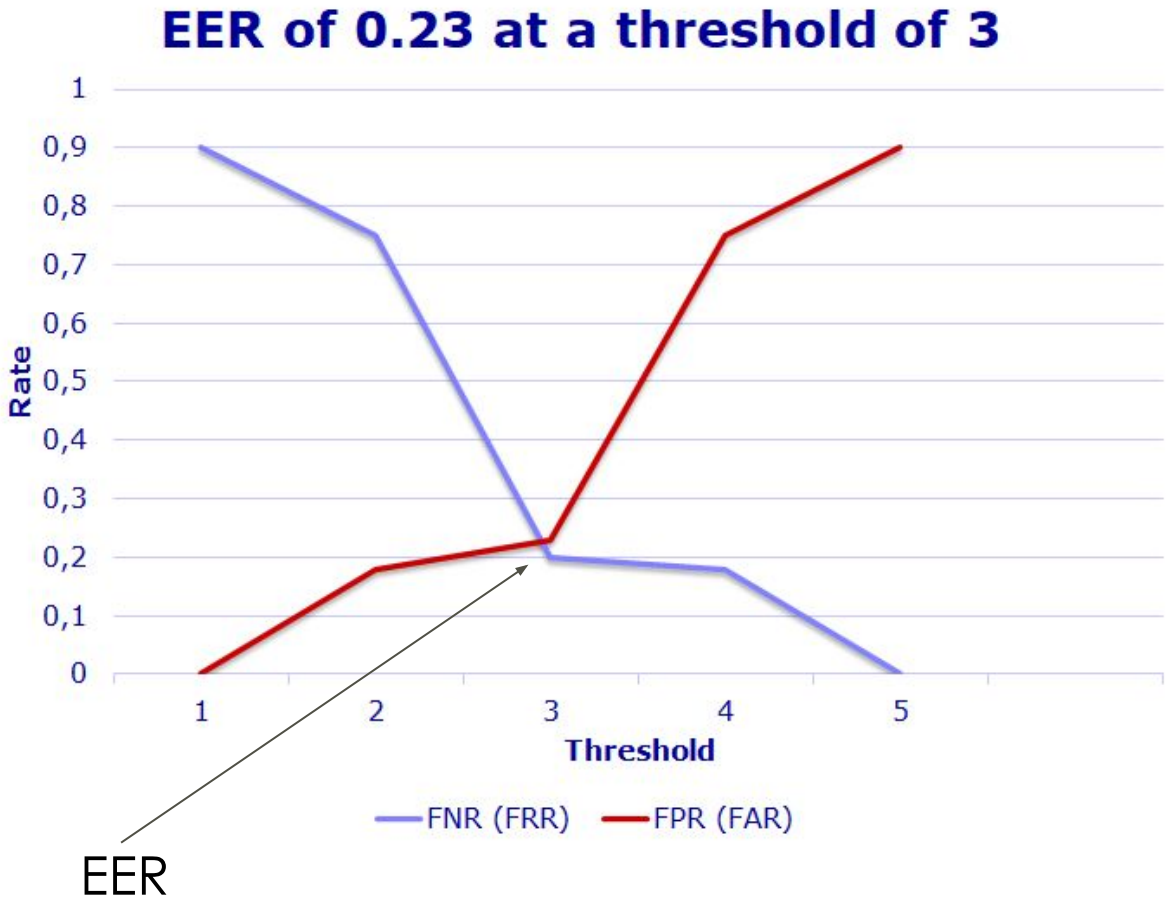


Metrics: Equal Error Rate (EER)

- Practical metric for *verification* systems with threshold (binary classification).

- Y-axis: two error rates
  - False Positive Rate (FAR)
  - False Negative Rate (FRR)
- X-axis: varied threshold

- At a certain point, both error rates are minimal. Range: [0; 1]
- Their intersection is known as the "equal error rate".
- The lower the EER, the better.



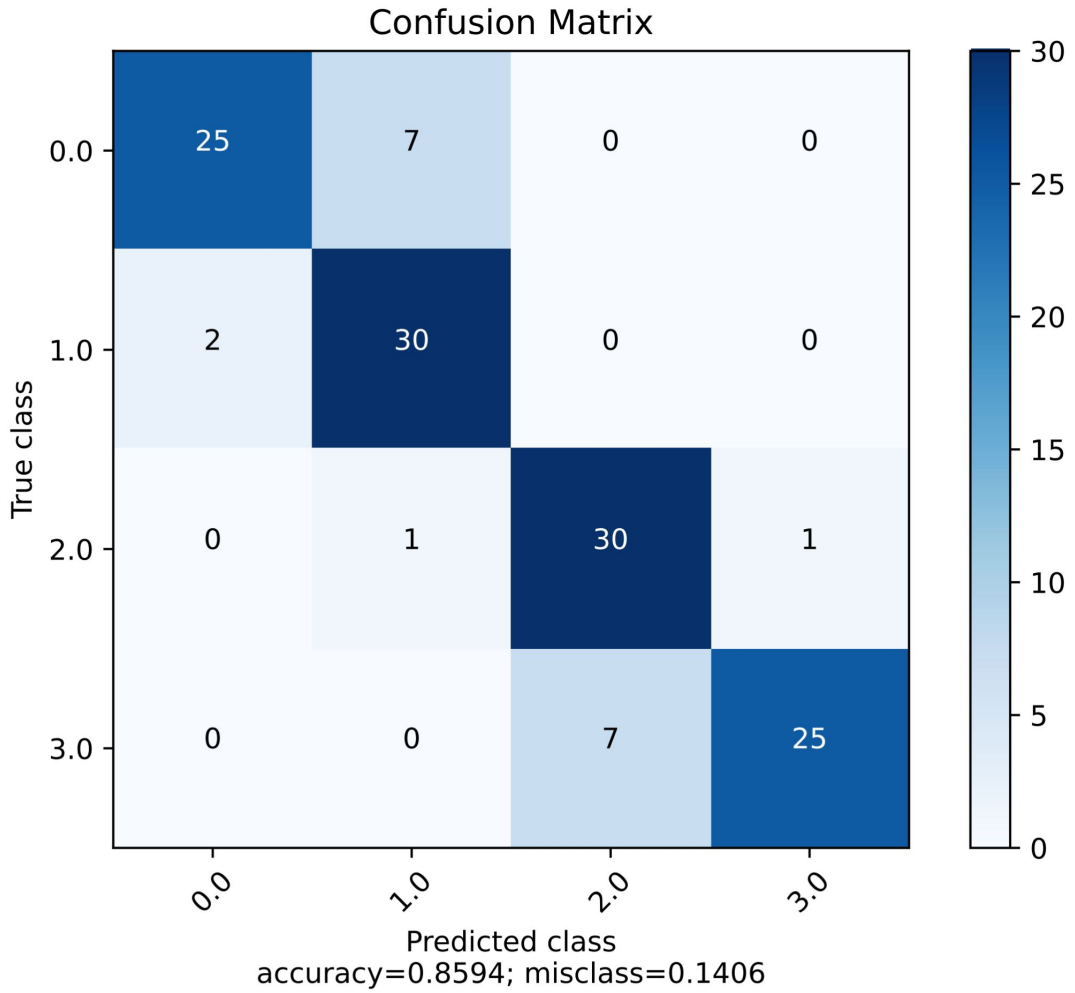
Metrics: Confusion Matrix (CM)

- Easily shows predictions for an *identification* system.

- Y-axis: ground truth.
- X-axis: system's prediction.

- Diagonal: true predictions, i.e., correct identification.

- CM can be normalized in the range [0; 1] or contain absolute numbers of trials (cf., Figure).



Hands on!

Programming Exercise

Google Colab Setup (recommended)

1. Follow this URL: https://material.behavioral-biometrics.org
2. You can find the tasks-notebook and solutions-notebook in the quick access links.
3. It should open Google Colaboratory on click.

Manual Setup (not recommended)

1. Download Python 3.10 from python.org.
2. Install Python 3.10 and place it on the **path** during installation (tick the checkbox).
3. Download the git repository: https://github.com/jliebers/CHI2023-Introduction-to-Authentication-using-Behavioral-Biometrics-Repository
4. Open a Powershell in the folder with the .ipynb-files
5. Install the dependencies: **pip install –r requirements.txt**
6. Run: „jupyter notebook“ and open the file **vr-case-study-v3.ipynb**

Exercise

behavioral-biometrics.org

Soon: wiki, tutorial, other behavioral biometric-related resources.