GoHack 2026, Zurich
30 October 2026

Prove It, Don’t Show It

The Future of Digital Identity and Data Collaboration

Prof. Katerina Mitrokotsa

Cybersecurity & Applied Cryptography Chair

University of St.Gallen



Quick Show of Hands

- Who has used an **AI tool with company data** this month?



Quick Show of Hands

- Who has used an **AI tool with company data** this month?
- Who has sent a **photo of your ID** or passport to an app or website?



Quick Show of Hands

- Who has used an **AI tool with company data** this month?
- Who has sent a **photo of your ID** or passport to an app or website?
- Who knows **where** that data is now, and **who** can see it?



Quick Show of Hands

- Who has used an **AI tool with company data** this month?
- Who has sent a **photo of your ID** or passport to an app or website?
- Who knows **where** that data is now, and **who** can see it?



We hand over our **most sensitive data** to AI,
and show our **whole ID** to prove one fact.

Today this rests on blind trust.

HSG Has a School of Computer Science

Where Computer Science meets Entrepreneurship

2020

founded, 6th school of HSG

13

professors, 8 research groups

3

degrees: BSc, MSc and PhD

HSG Has a School of Computer Science

Where Computer Science meets Entrepreneurship

2020

founded, 6th school of HSG

13

professors, 8 research groups

3

degrees: BSc, MSc and PhD

Cyber security is a core area, with a **Strategic Cybersecurity** track in our Master.

HSG Has a School of Computer Science

Where Computer Science meets Entrepreneurship

2020

founded, 6th school of HSG

13

professors, 8 research groups

3

degrees: BSc, MSc and PhD

Cyber security is a core area, with a **Strategic Cybersecurity** track in our Master.

In St. Gallen, and open to working with you.

Data Fuels AI



Healthcare Data



Environmental Data

Data Driven World



Physical Activity Data



Financial Transactions

Data Fuels AI



Healthcare Data



Physical Activity Data

Data Driven World



Environmental Data



Financial Transactions

More data makes better AI.

But who protects the data once we hand it over?

The Story Today

1

Blind trust

what goes wrong

2

Why now

sovereignty & quantum

3

Guarantees

compute, e-ID, split

4

The Road Ahead

what to do

The Story Today

1

Blind trust

what goes wrong

2

Why now

sovereignty & quantum

3

Guarantees

compute, e-ID, split

4

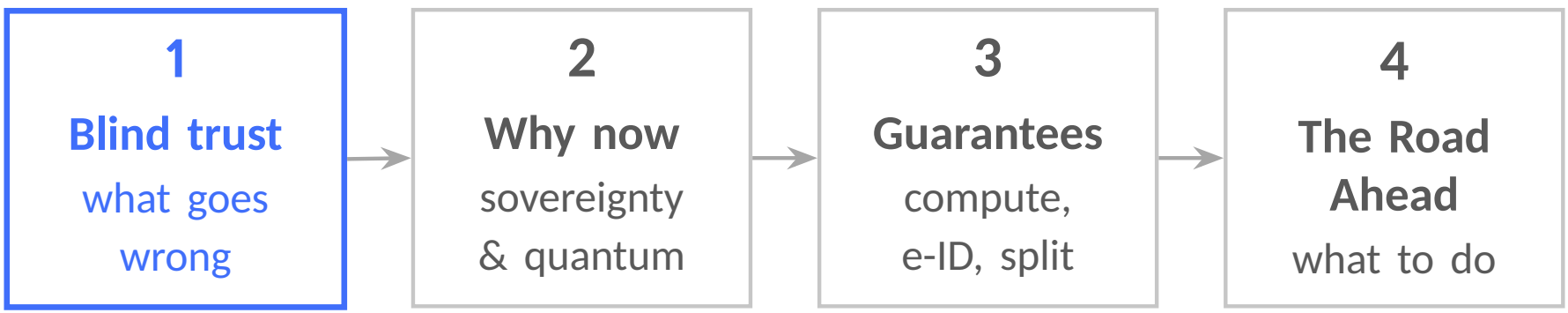
The Road Ahead

what to do

AI runs on **blind trust**.
Cryptography can replace it with **guarantees you can verify**.

Part 1

Blind Trust

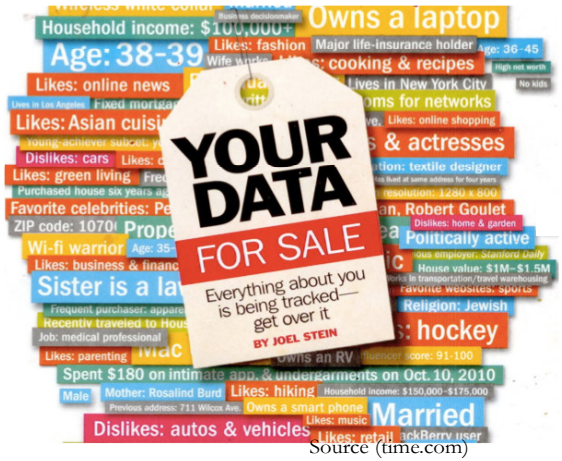


What goes wrong when we hand AI our data and our identity?

Data Breaches!! Need for Robust Solutions!







Source (time.com)

533 million Facebook users' phone numbers and personal data have been leaked online

Aaron Holmes Apr 3, 2021, 4:41 PM



UPDATED 21:58 EDT / MAY 26 2019



Insurance firm First American Financial exposes 885M customer records

BY **DUNCAN RILEY**

A security flaw in the website of First American Financial Corp., the large real estate title [illegible] exposed over 885 million private and confidential customer records dating

Management und Qualität MQ

Online counter EasyGov hacked

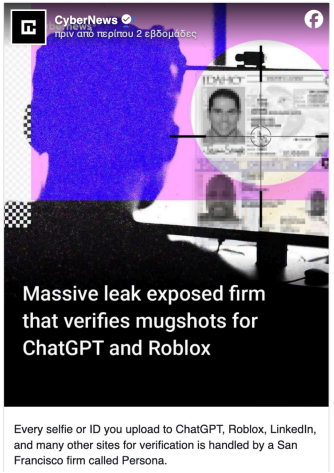
Criminal hackers have allegedly succeeded in stealing a list containing the names of up to 130,000 companies that applied for Covid 19 credit via the EasyGov platform in 2020. According to current knowledge, no other data apart from the company names was stolen, as Seco emphasises. As the operator of Easy-Gov, immediate measures were taken and an investigation initiated.

Editorial office - 21 October 2021

When Identity Checks Become Surveillance

The Persona Leak

An ID-check company
exposed its own code.

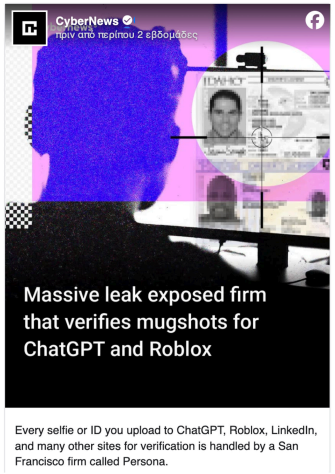


You show
your **ID**

When Identity Checks Become Surveillance

The Persona Leak

An ID-check company
exposed its own code.

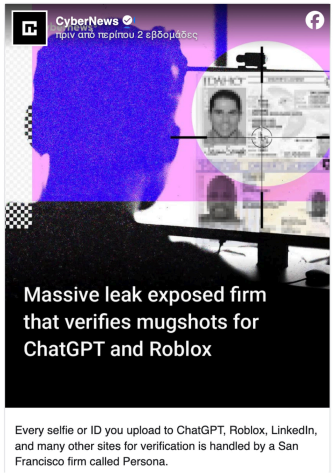


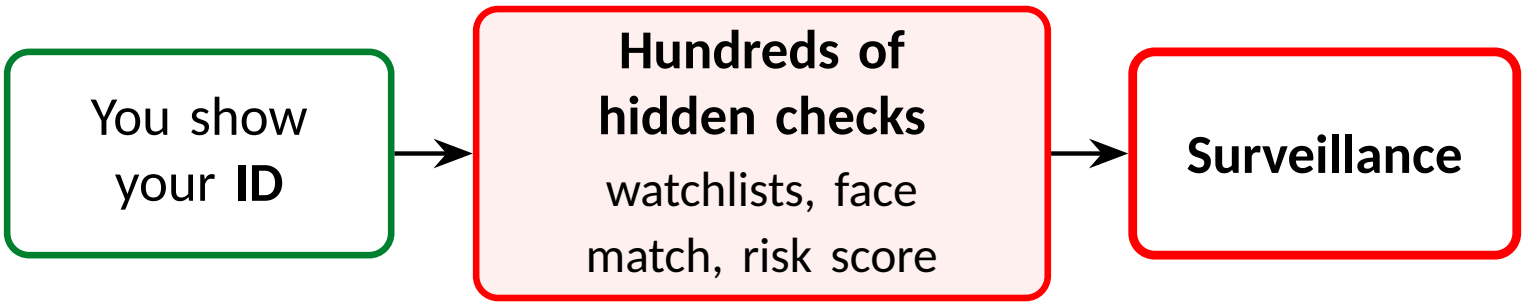
You show your **ID** → **Hundreds of hidden checks** watchlists, face match, risk score

When Identity Checks Become Surveillance

The Persona Leak

An ID-check company
exposed its own code.

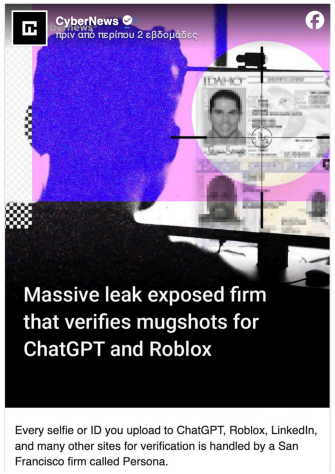




When Identity Checks Become Surveillance

The Persona Leak

An ID-check company
exposed its own code.



You show your **ID** → **Hundreds of hidden checks** watchlists, face match, risk score → **Surveillance**

Lesson: Data you **show** for one check can be **reused** for anything.

Seeing Is No Longer Believing

Phishing is done **easier**!

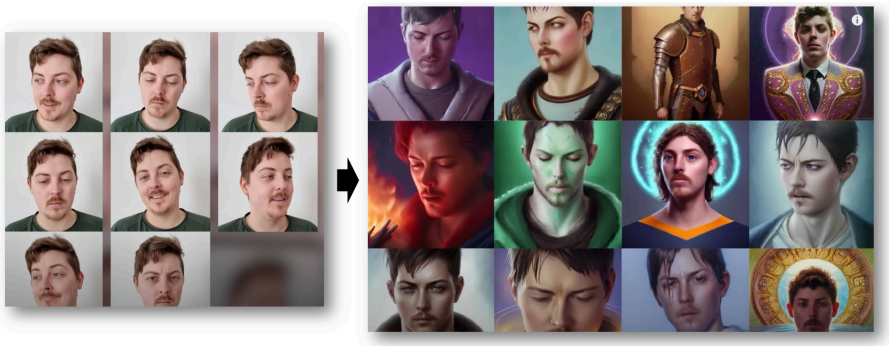


LILY HAY NEWMAN SECURITY AUG 7, 2021 7:00 AM

AI Wrote Better Phishing Emails Than Humans in a Recent Test

Researchers found that tools like OpenAI's GPT-3 helped craft devilishly effective spearphishing messages.

Deep fakes for **everyone**!



https://www.youtube.com/watch?v=FaLTztGGueQ, James Cunliffe

Copyright F. Tramer. Arup case: MIT Sloan CyberIR interview with Arup CIO Rob Greig.

Seeing Is No Longer Believing

Phishing is done **easier**!

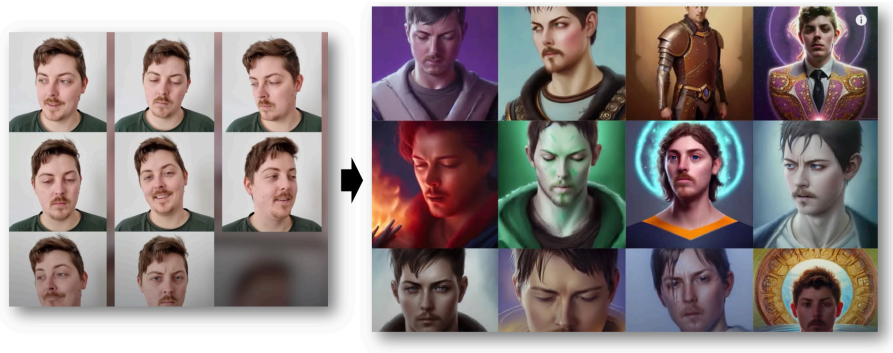


LILY HAY NEWMAN SECURITY AUG 7, 2021 7:00 AM

AI Wrote Better Phishing Emails Than Humans in a Recent Test

Researchers found that tools like OpenAI's GPT-3 helped craft devilishly effective spearphishing messages.

Deep fakes for **everyone**!



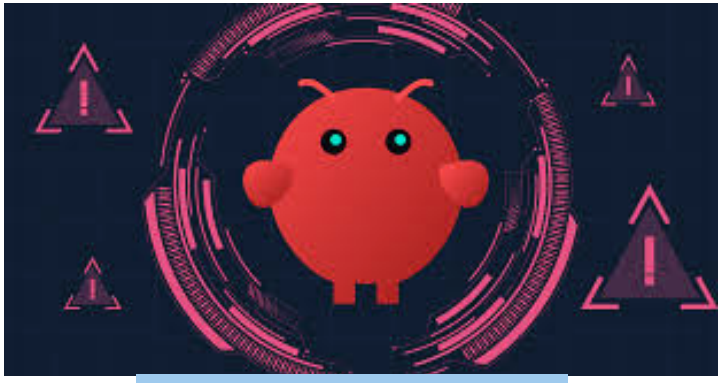
https://www.youtube.com/watch?v=FaLTztGGueQ, James Cunliffe

Arup, 2024: an employee wired about **$25 million** after a video call with deepfaked senior managers.

Copyright F. Tramer. Arup case: MIT Sloan CyberIR interview with Arup CIO Rob Greig.

When AI Assistants Act as You

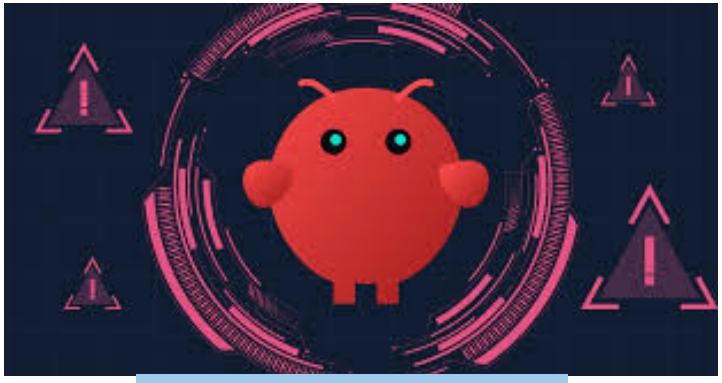
- Runs **invisibly**, all the time





When AI Assistants Act as You

- Runs **invisibly**, all the time
- Holds your **tokens, cookies, identity**





When AI Assistants Act as You

- Runs **invisibly**, all the time
- Holds your **tokens, cookies, identity**
- **Acts in your name**





When AI Assistants Act as You

- Runs **invisibly**, all the time
- Holds your **tokens, cookies, identity**
- **Acts in your name**





"With chatbots, you worry about getting the wrong answer. With agents, you worry about taking the wrong action."

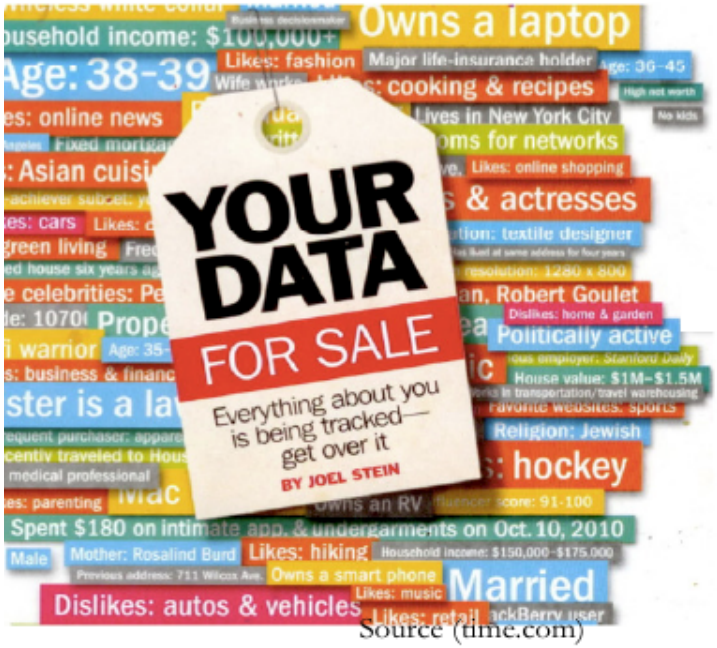
Jeetu Patel, Cisco, RSAC 2026

Shared for Convenience, Found by Anyone

6,800+

shared Meta Muse AI chats,
found with **one search**

loans · revenue reviews · confidential reports



Source (time.com)

Source: Claroc research, reported by eSecurity Planet, Oct. 2026.

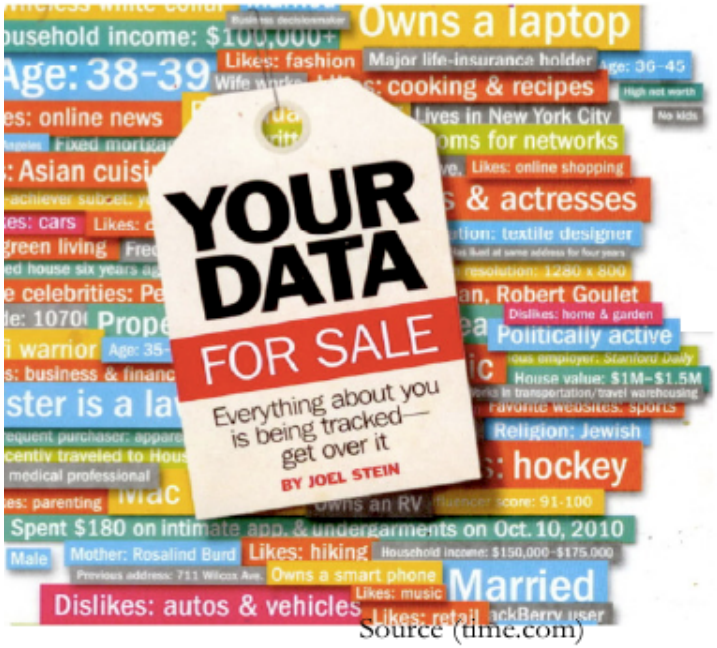
Shared for Convenience, Found by Anyone

6,800+

shared Meta Muse AI chats,
found with **one search**

loans · revenue reviews · confidential reports

- Chat title in the link → **public DNS**



Source (time.com)

Source: Claroc research, reported by eSecurity Planet, Oct. 2026.

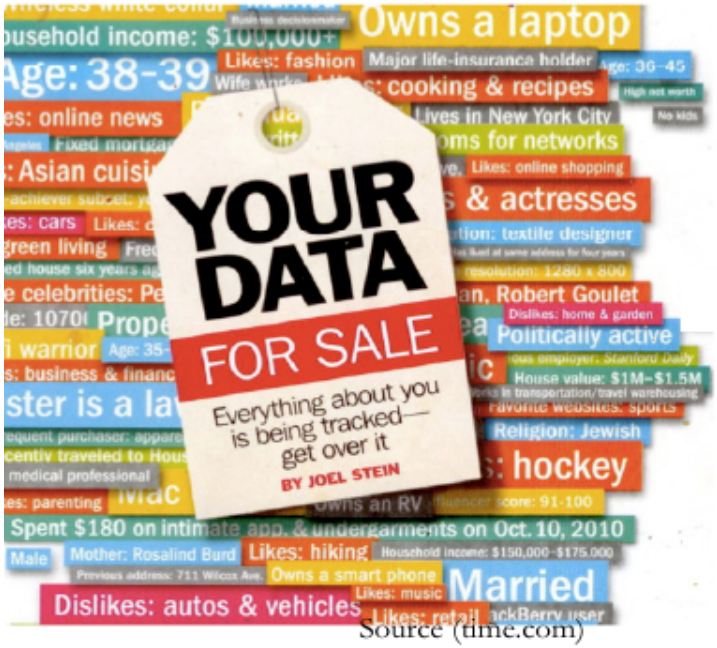
Shared for Convenience, Found by Anyone

6,800+

shared Meta Muse AI chats,
found with **one search**

loans · revenue reviews · confidential reports

- Chat title in the link → **public DNS**
- Your emails too, if a **contact** connected their inbox



Source (time.com)

Source: Claroc research, reported by eSecurity Planet, Oct. 2026.

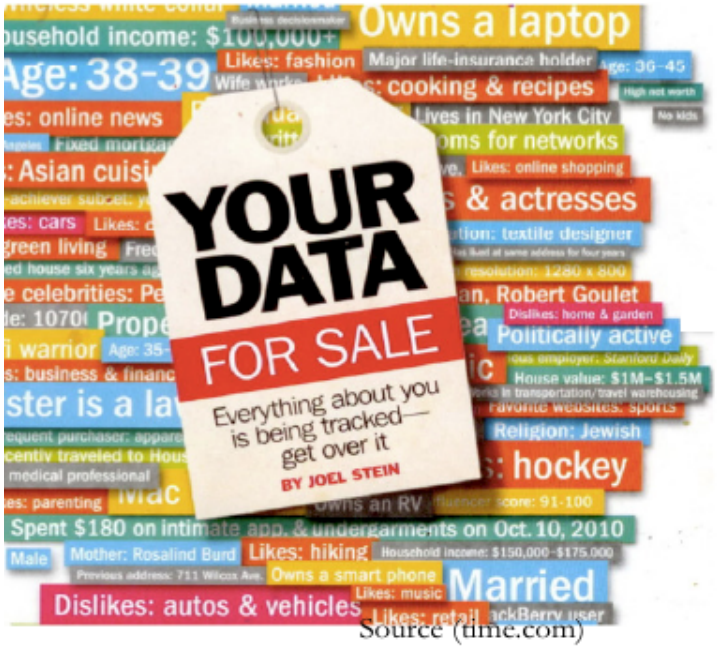
Shared for Convenience, Found by Anyone

6,800+

shared Meta Muse AI chats,
found with **one search**

loans · revenue reviews · confidential reports

- Chat title in the link → **public DNS**
- Your emails too, if a **contact** connected their inbox

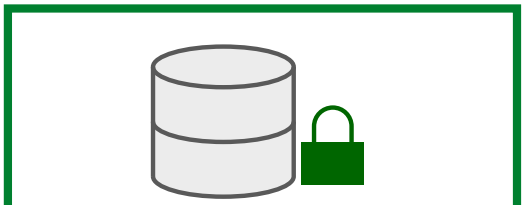


Source (time.com)

Your AI agent shares **other people's data** too.

Source: Claroc research, reported by eSecurity Planet, Oct. 2026.

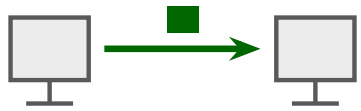
Where Blind Trust Sits



Data at rest

stored, encrypted

✓



Data in transit

sent over TLS

✓

Where Blind Trust Sits



Data at rest

stored, encrypted

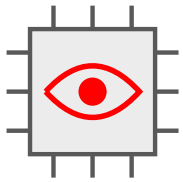
✓



Data in transit

sent over TLS

✓

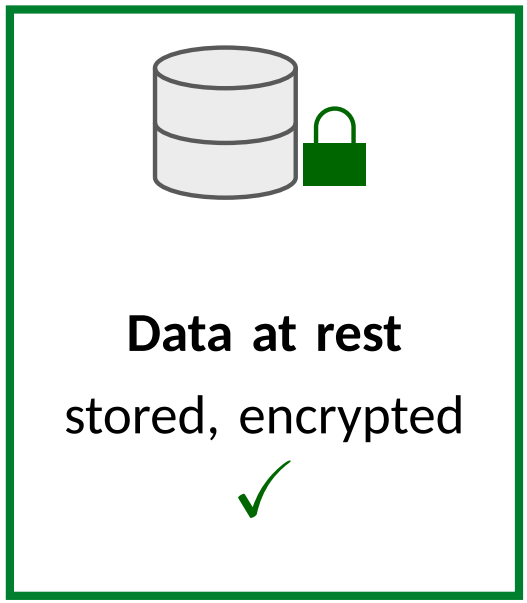


Data in use

the AI model reads it **in the clear**

×

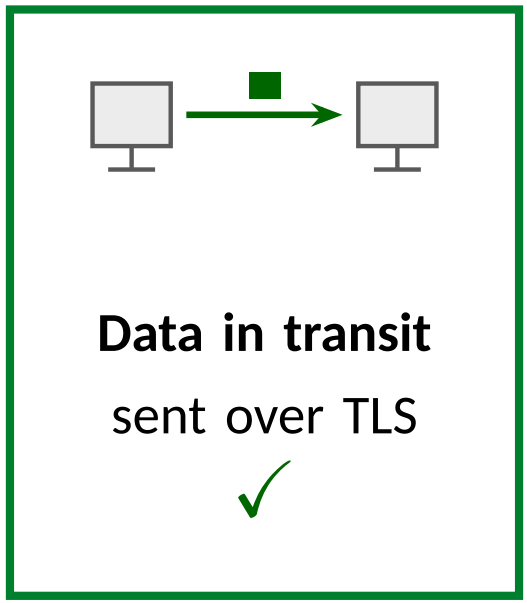
Where Blind Trust Sits



Data at rest

stored, encrypted

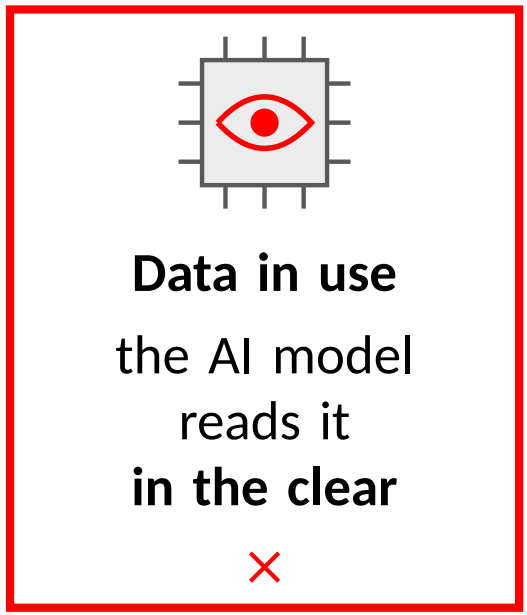
✓



Data in transit

sent over TLS

✓



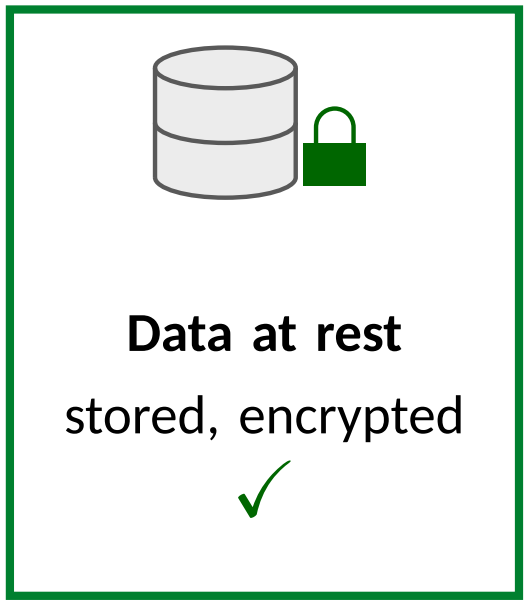
Data in use

the AI model reads it **in the clear**

×

- AI must **process** your data, so encryption at rest and in transit is not enough
- Today the only protection in use is a **contract and a promise**

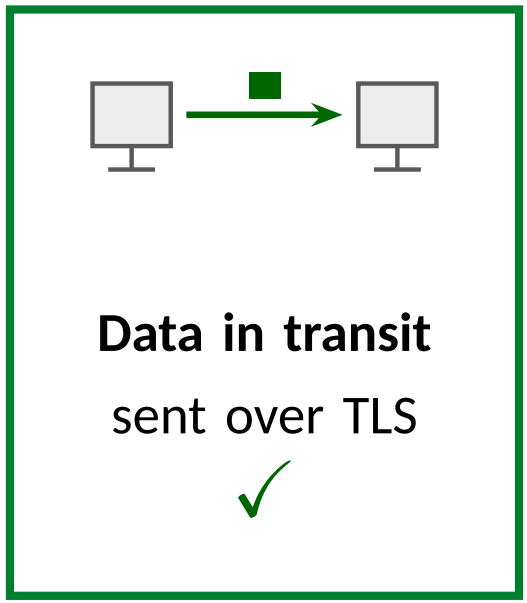
Where Blind Trust Sits



Data at rest

stored, encrypted

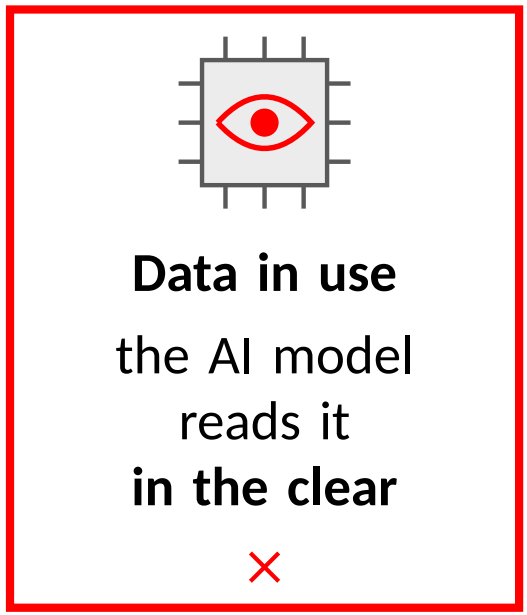
✓



Data in transit

sent over TLS

✓



Data in use

the AI model reads it **in the clear**

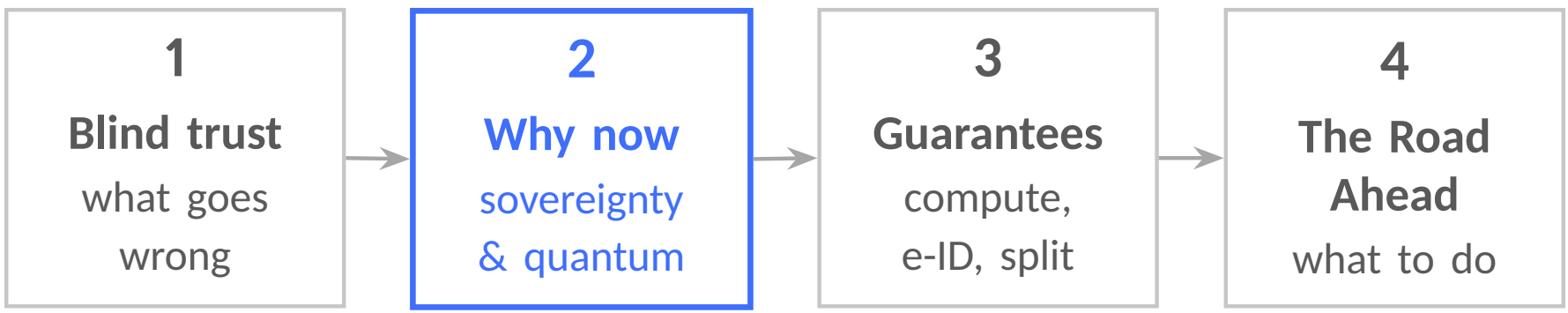
×

- AI must **process** your data, so encryption at rest and in transit is not enough
- Today the only protection in use is a **contract and a promise**

Cryptography can close this gap.

Part 2

Why Now



Two clocks are ticking: **geopolitics** and **quantum computing**.

AI Is Now Geopolitics

59%

of top AI models: US

26%

of top AI models: China



Source: BCG Center for Geopolitics, “The Geopolitical Forces Shaping Business in 2026” (Dec. 2025), A. Bharadwaj, M. Gilbert, T. Gratowski, N. Lang, K. Maggard, C. Wilson.

AI Is Now Geopolitics

59%

of top AI models: US

26%

of top AI models: China



BCG: don't rely on a **single AI supplier**; plan for **continuity**.

Source: BCG Center for Geopolitics, "The Geopolitical Forces Shaping Business in 2026" (Dec. 2025), A. Bharadwaj, M. Gilbert, T. Gratowski, N. Lang, K. Maggard, C. Wilson.

AI Is Now Geopolitics

59%

of top AI models: US

26%

of top AI models: China



BCG: don't rely on a **single AI supplier**; plan for **continuity**.

Sovereignty is **not** building our own GPT.
It is being able to switch providers without losing your data.

Source: BCG Center for Geopolitics, "The Geopolitical Forces Shaping Business in 2026" (Dec. 2025), A. Bharadwaj, M. Gilbert, T. Gratowski, N. Lang, K. Maggard, C. Wilson.

Post-Quantum: Harvest Now, Decrypt Later





Quantum computers are coming

Post-Quantum: Harvest Now, Decrypt Later



- **10+ year** secrets: **at risk now**



Quantum computers are coming

Post-Quantum: Harvest Now, Decrypt Later



- **10+ year** secrets: **at risk now**
- Fix exists: **NIST** standards (2024)



Quantum computers are coming

Post-Quantum: Harvest Now, Decrypt Later



- **10+ year** secrets: **at risk now**
- Fix exists: **NIST** standards (2024)
- **Swiss e-ID**: not quantum-safe yet



Quantum computers are coming

Post-Quantum: Harvest Now, Decrypt Later



- **10+ year** secrets: **at risk now**
- Fix exists: **NIST** standards (2024)
- **Swiss e-ID**: not quantum-safe yet

FINMA: **8%** have a roadmap,
72% did **nothing**



Quantum computers are coming

Post-Quantum: Harvest Now, Decrypt Later



- **10+ year** secrets: **at risk now**
- Fix exists: **NIST** standards (2024)
- **Swiss e-ID**: not quantum-safe yet

FINMA: **8%** have a roadmap,
72% did **nothing**



Quantum computers are coming

So how do we use AI and digital identity without blind trust?

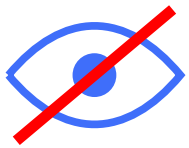
Part 3

Verifiable Guarantees



Replace “trust us” with “**check it yourself**”.

From Blind Trust to Verifiable Guarantees

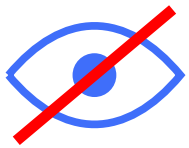


A. Compute without seeing

Data is used,
never revealed

multi-party computation

From Blind Trust to Verifiable Guarantees



A. Compute without seeing

Data is used,
never revealed

multi-party computation

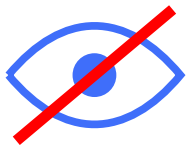


B. Prove without revealing

Show **only**
what is needed

selective disclosure,
zero-knowledge

From Blind Trust to Verifiable Guarantees



A. Compute without seeing

Data is used, **never revealed**

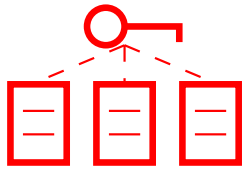
multi-party computation



B. Prove without revealing

Show **only what is needed**

selective disclosure, zero-knowledge



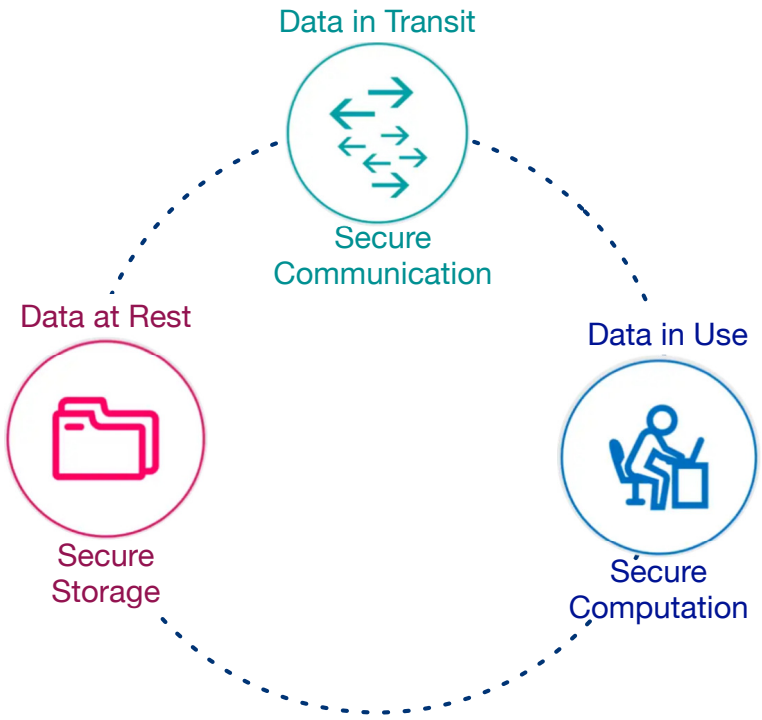
C. No single point of trust

No one can **see or abuse** it all

threshold cryptography

Part 3 · Guarantees

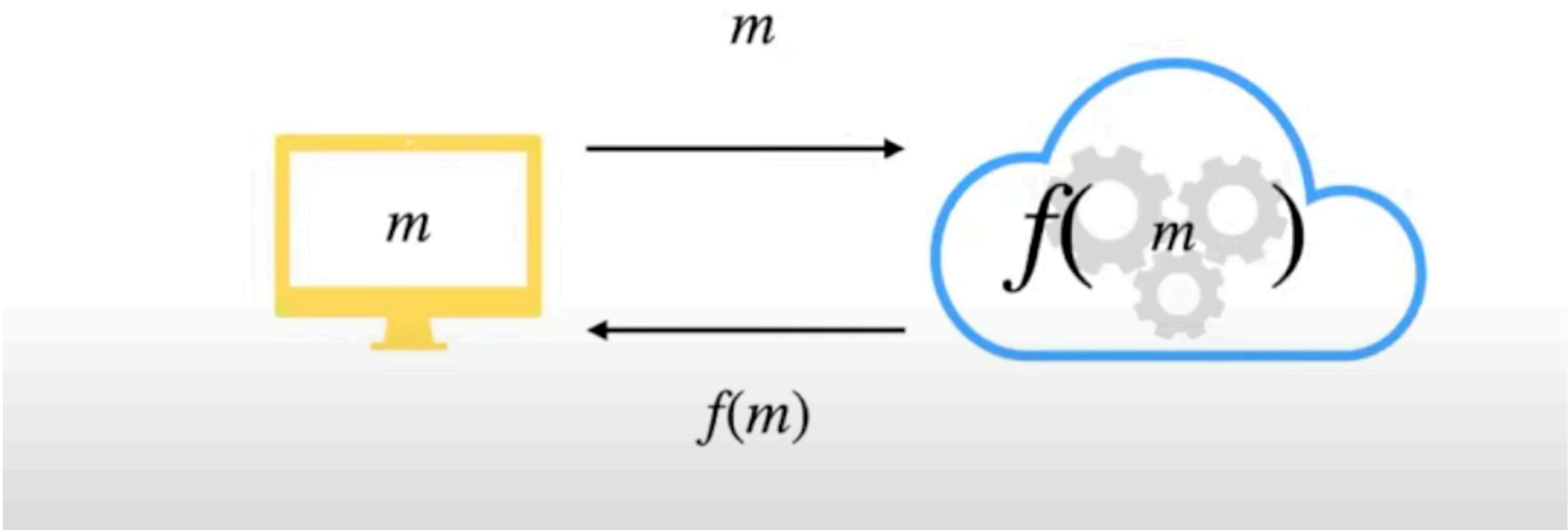
A. Compute Without Seeing



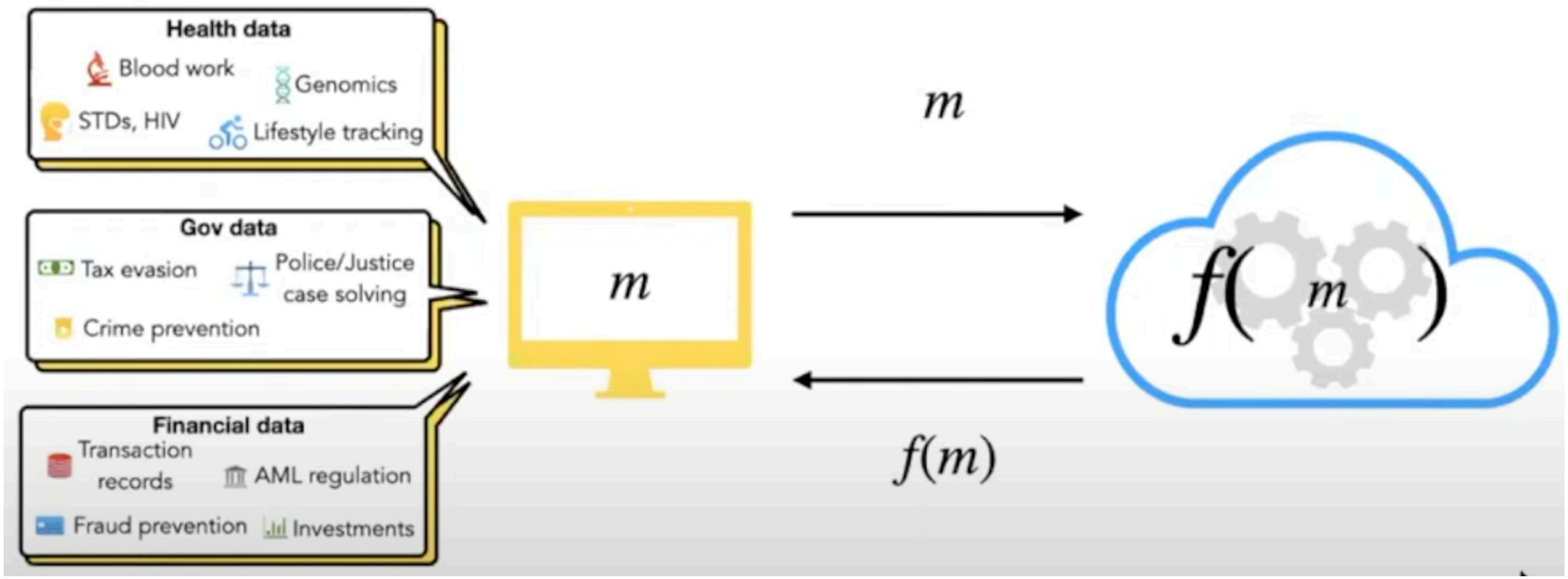
Compute on Data You Cannot See



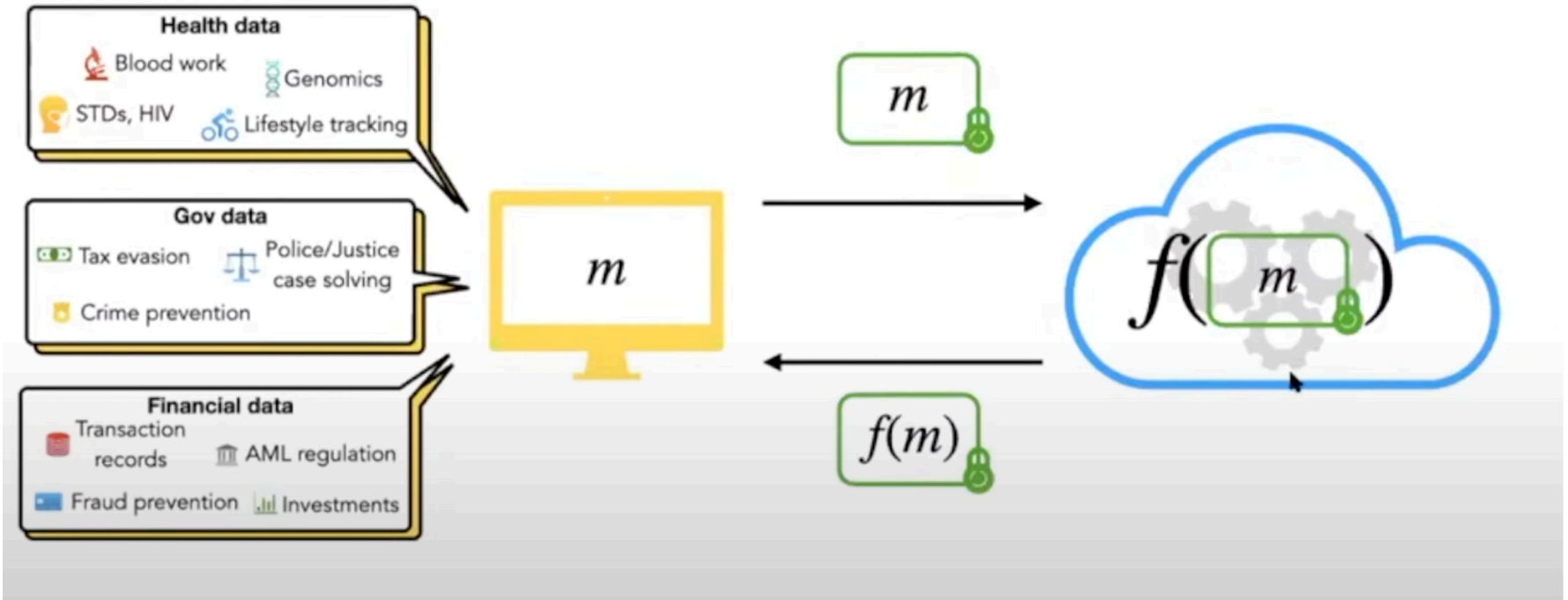
Compute on Data You Cannot See



Compute on Data You Cannot See



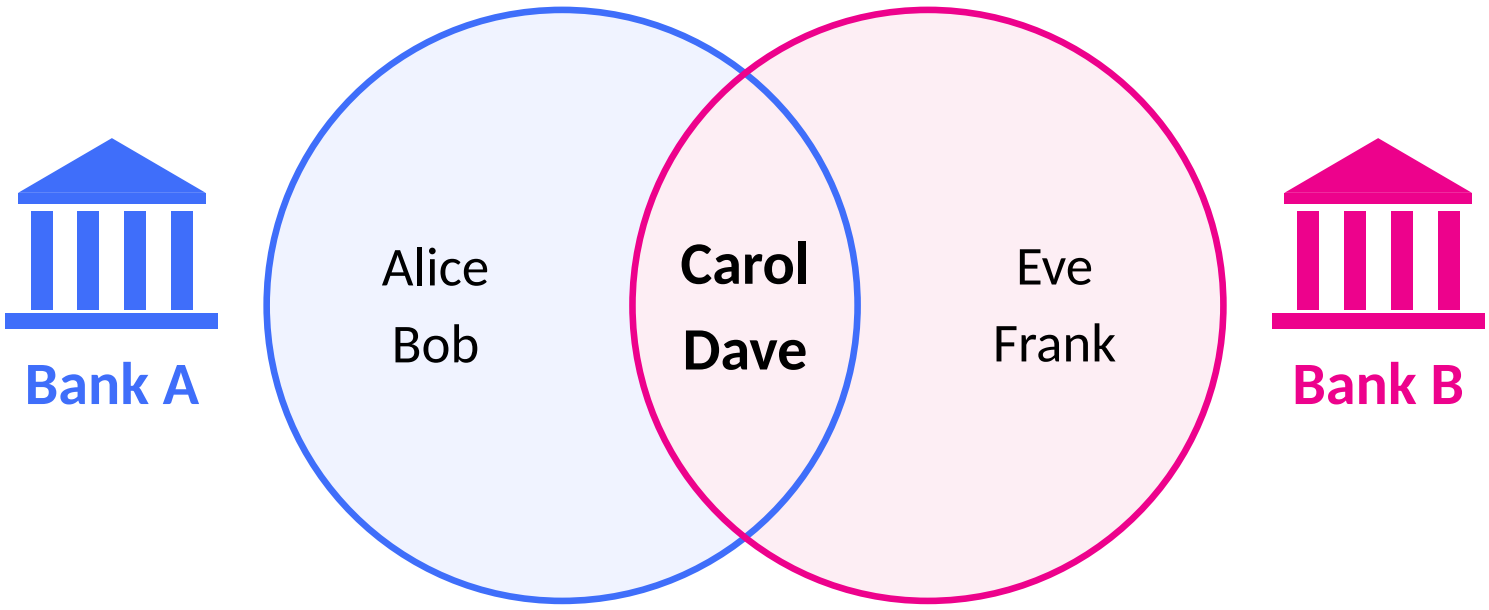
Compute on Data You Cannot See



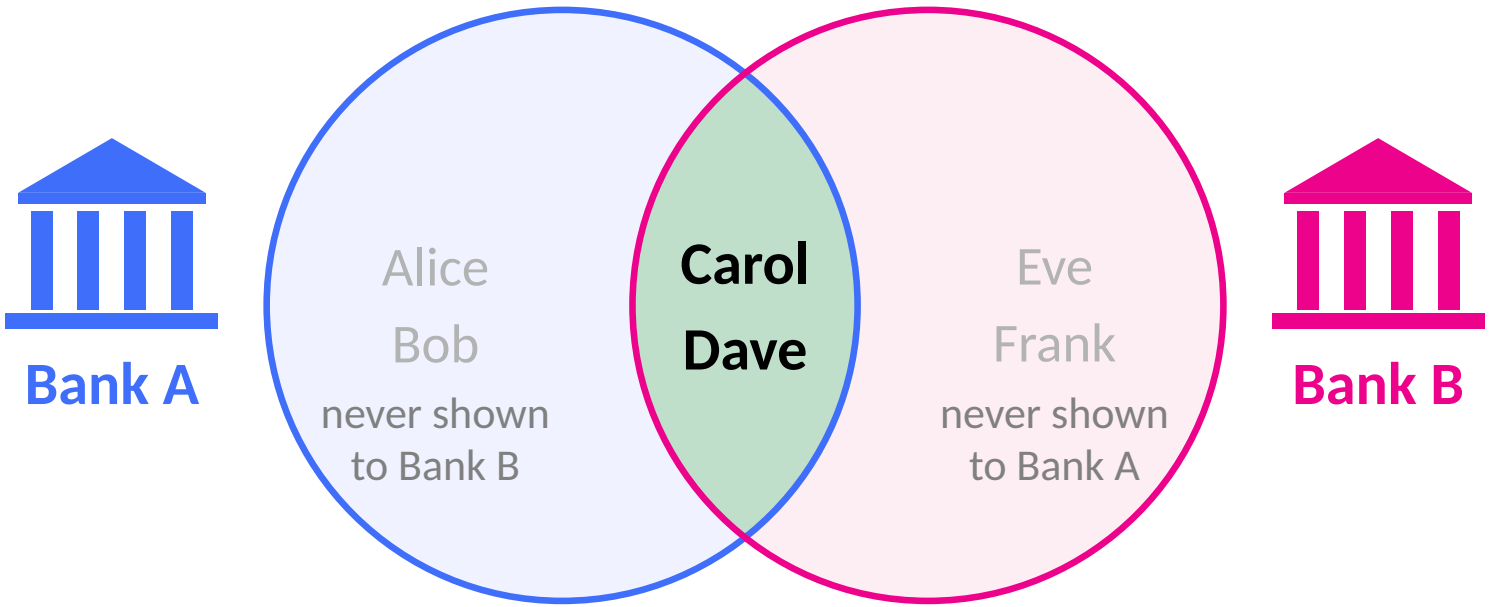
Perform computations on encoded data!

Sounds like magic  **But it can be done, and verified!**

Private Set Intersection: Learn Only the Overlap

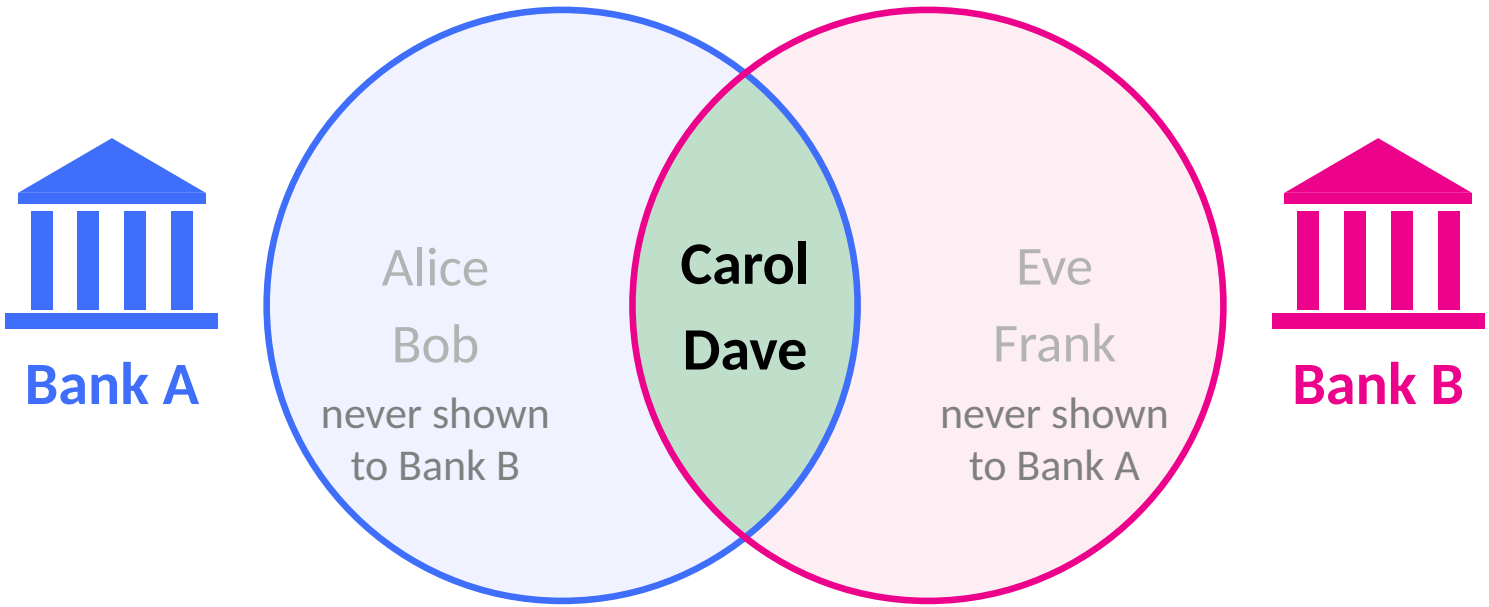


Private Set Intersection: Learn Only the Overlap



Both banks learn **only**: Carol, Dave

Private Set Intersection: Learn Only the Overlap

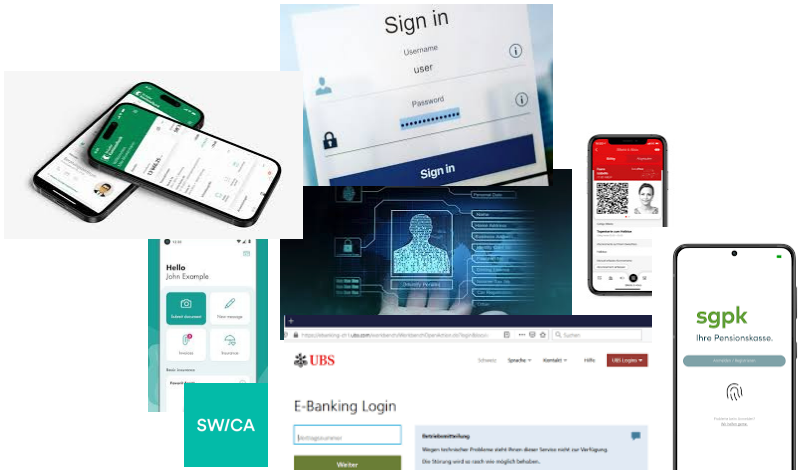


Both banks learn **only**: Carol, Dave

Compare fraud lists, threat intelligence or contacts
without handing over your customer list.

Part 3 · Guarantees

B. Prove Without Revealing



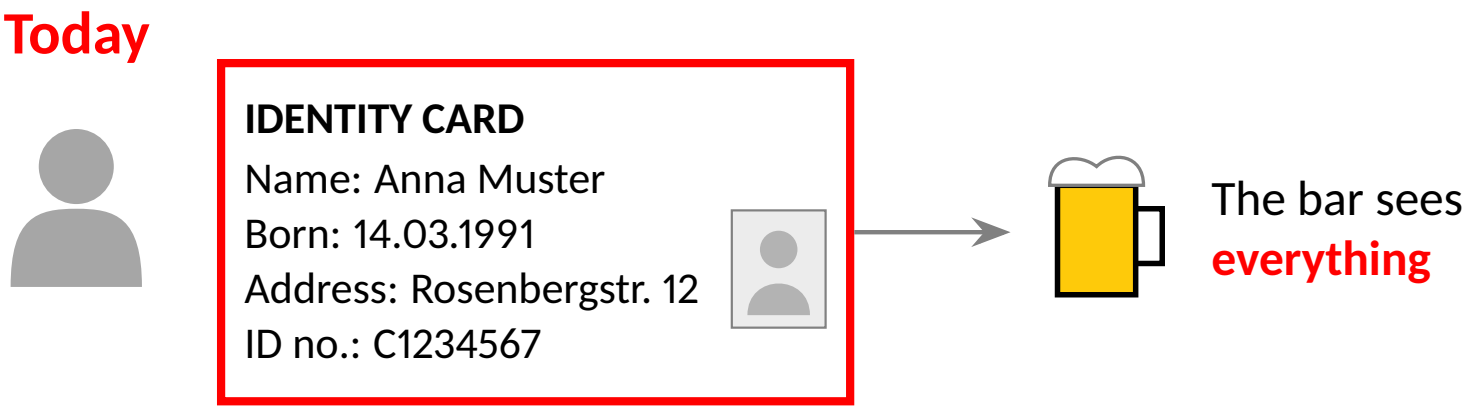
Electronic Identity & Trust Infrastructure

Elektronische Identität und Vertrauensinfrastruktur

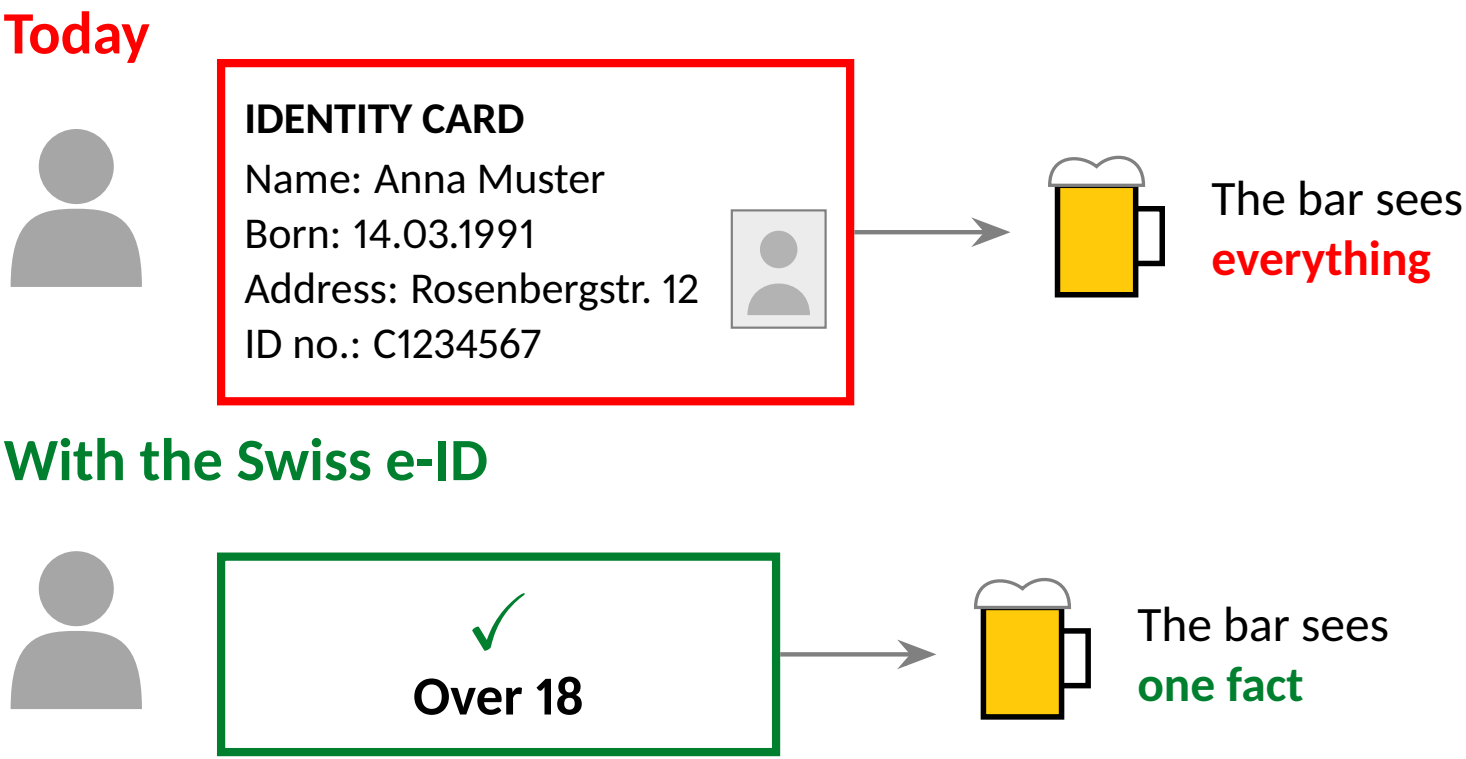


Are we ready for this?

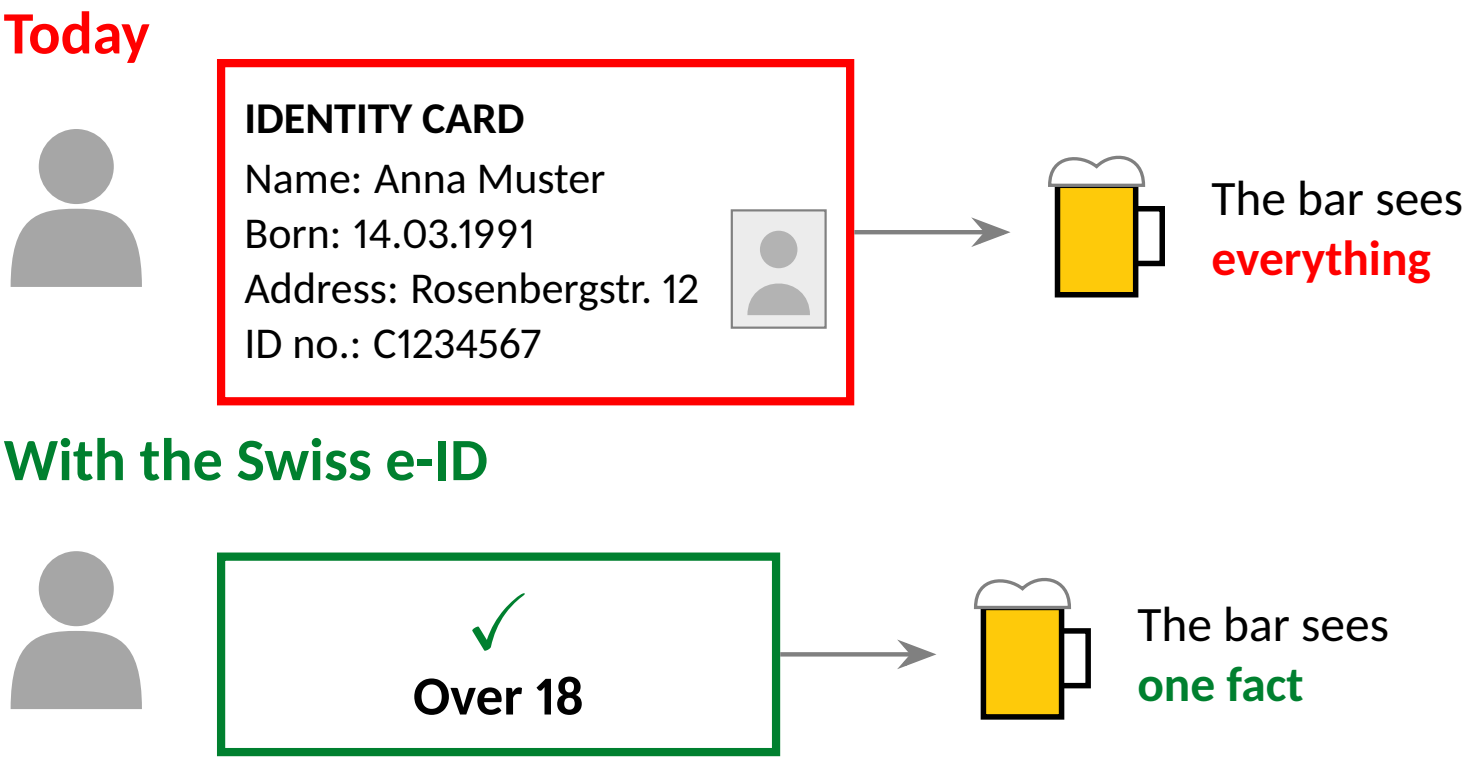
The Beer Test



The Beer Test



The Beer Test



The check needed **one fact**. Today it gets your **name, birthday, address, ID number and photo**.

How the Swiss e-ID Proves “Over 18”



The state
signs your data once

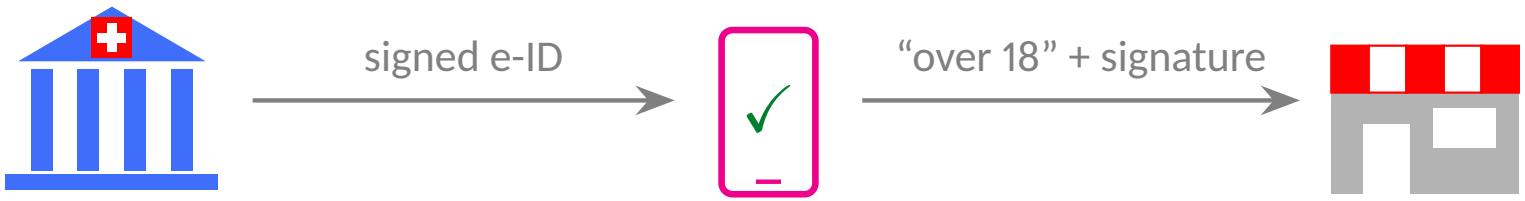
How the Swiss e-ID Proves “Over 18”



The state
signs your data once

Your phone
keeps it, and shares
only “over 18”

How the Swiss e-ID Proves “Over 18”

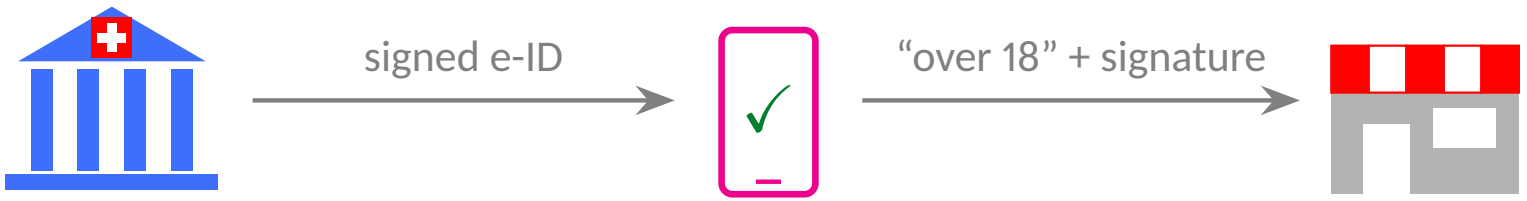


The state
signs your data once

Your phone
keeps it, and shares
only “over 18”

The shop
checks the signature:
over 18, nothing else

How the Swiss e-ID Proves “Over 18”



The state
signs your data once

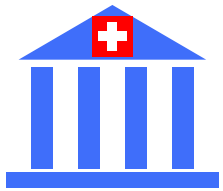
Your phone
keeps it, and shares
only “over 18”

The shop
checks the signature:
over 18, nothing else

The state **vouches** for the data.
You decide what to reveal.



Swiss e-ID: Selective Disclosure in Action



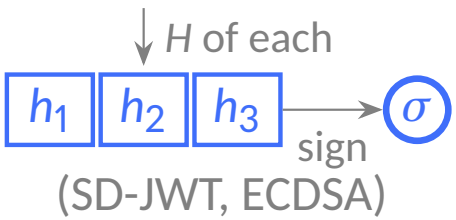
Issuer (fedpol)

disclosures:

(s_1, age_over_18, true)

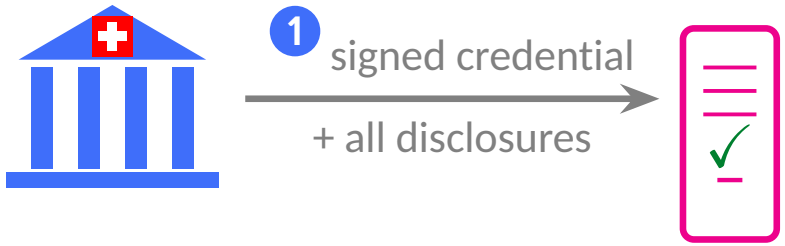
(s_2, name, Anna)

(s_3, address, …)



(SD-JWT, ECDSA)

Swiss e-ID: Selective Disclosure in Action



Issuer (fedpol)

disclosures:

(s_1, age_over_18, true)

(s_2, name, Anna)

(s_3, address, …)

H of each

h_1 h_2 h_3 → σ

sign

(SD-JWT, ECDSA)

Wallet (you)

stores:

h_1 h_2 h_3 σ

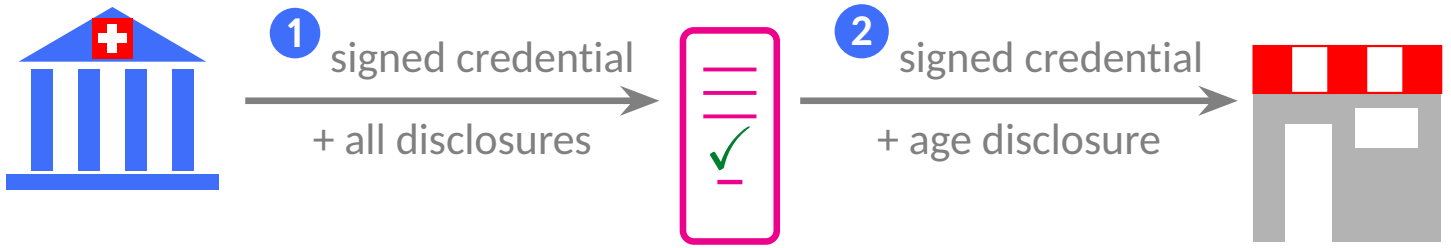
(s_1, age_over_18, true)

(s_2, name, Anna)

(s_3, address, …)

green: sent grey: kept

Swiss e-ID: Selective Disclosure in Action



Issuer (fedpol)

disclosures:

(s_1, age_over_18, true)

(s_2, name, Anna)

(s_3, address, …)

↓ H of each

h_1 h_2 h_3 → σ

sign

(SD-JWT, ECDSA)

Wallet (you)

stores:

h_1 h_2 h_3 σ

(s_1, age_over_18, true)

(s_2, name, Anna)

(s_3, address, …)

green: sent grey: kept

Verifier (shop)

receives:

h_1 h_2 h_3 σ

(s_1, age_over_18, true)

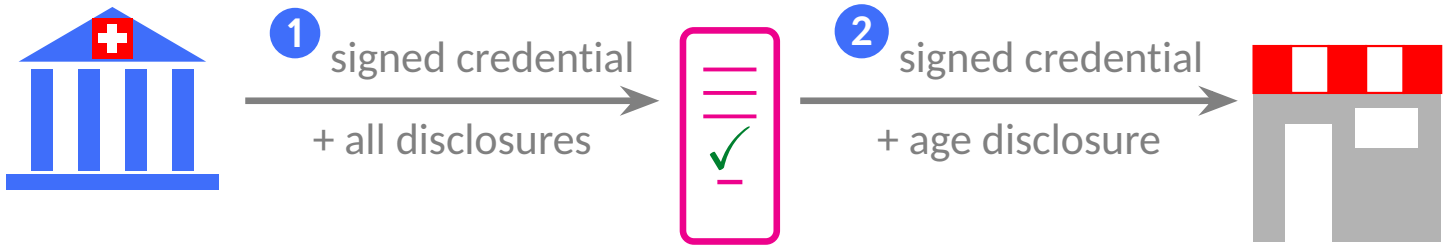
✓ σ valid on h_1, h_2, h_3

✓ H(s_1, age_over_18, true) = h_1, in the signed list

learns: over 18

name, address: unknown

Swiss e-ID: Selective Disclosure in Action



Issuer (fedpol)

disclosures:

(s_1, age_over_18, true)

(s_2, name, Anna)

(s_3, address, …)

↓ H of each

h_1 h_2 h_3 → σ

sign

(SD-JWT, ECDSA)

Wallet (you)

stores:

h_1 h_2 h_3 σ

(s_1, age_over_18, true)

(s_2, name, Anna)

(s_3, address, …)

green: sent grey: kept

Verifier (shop)

receives:

h_1 h_2 h_3 σ

(s_1, age_over_18, true)

✓ σ valid on h_1, h_2, h_3

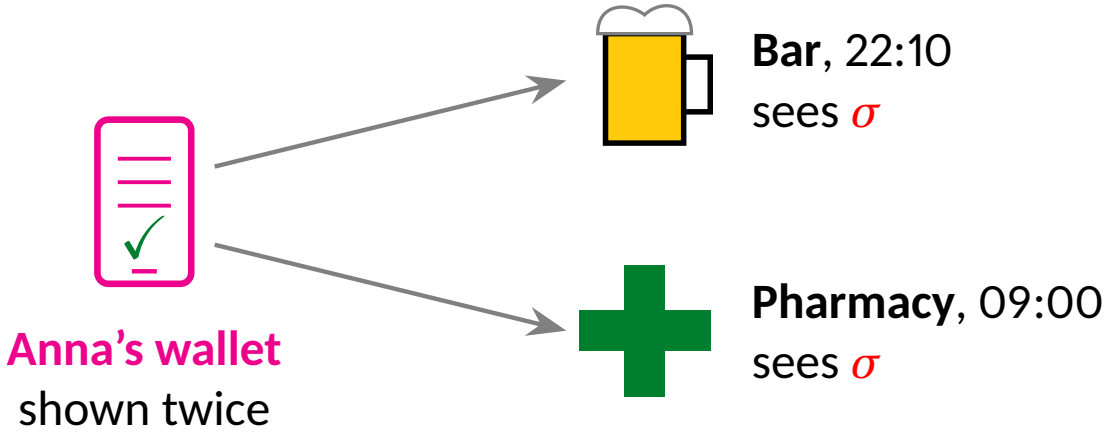
✓ $H(s_1, \text{age_over_18}, \text{true})$ $= h_1$, in the signed list

learns: over 18

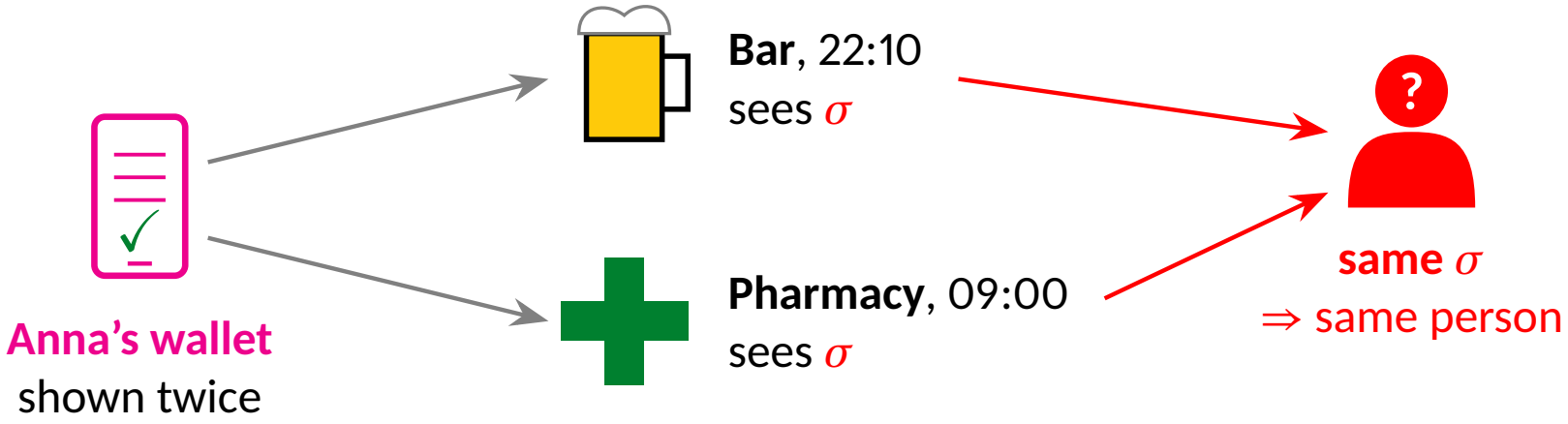
name, address: unknown

Key idea: the issuer signs the **hashes**; the user controls the real data.

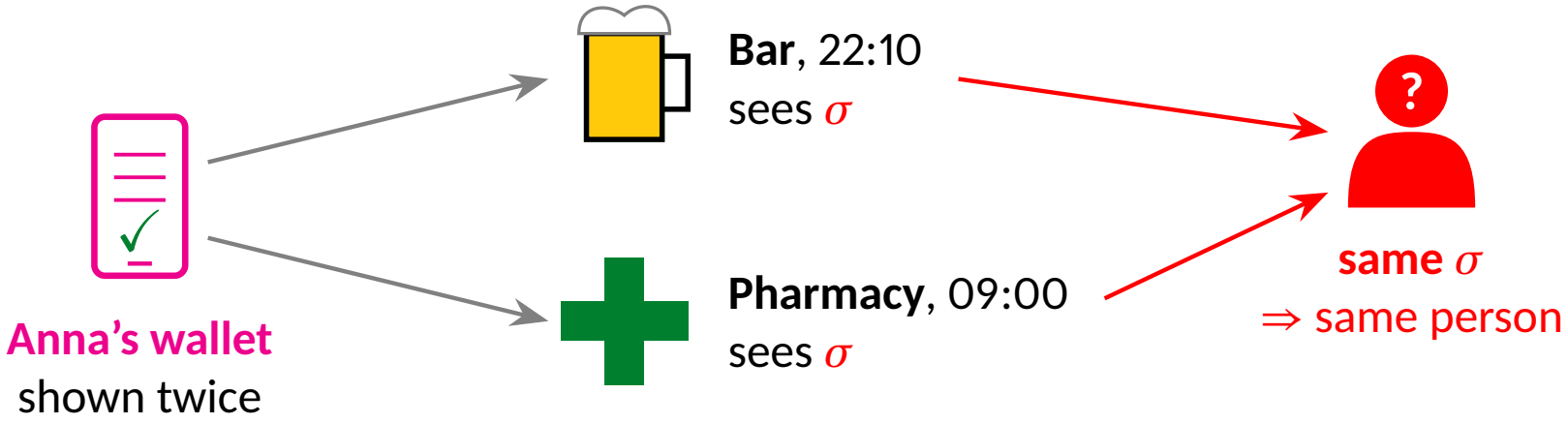
Risk: Two Shops Can Profile Your Customers



Risk: Two Shops Can Profile Your Customers

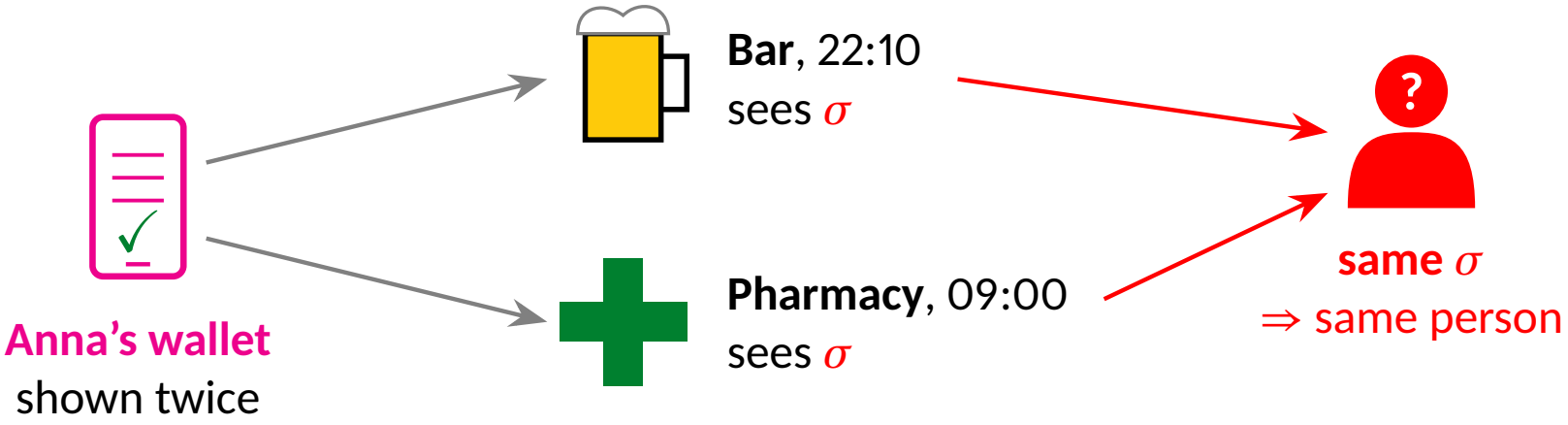


Risk: Two Shops Can Profile Your Customers



- **Today:** a **fresh** credential each time
  but issuer + shop can still link

Risk: Two Shops Can Profile Your Customers



- **Today:** a **fresh** credential each time
  but issuer + shop can still link
- **Next:** anonymous credentials (BBS)
  every presentation looks **different**: nothing to link

Swiss e-ID Scorecard

Gets right

- ✓ Reveal **only what's needed**
- ✓ Keys in the **phone's chip**
- ✓ **Fresh** credential per use
- ✓ **Open source** + bug bounty

Swiss e-ID Scorecard

Gets right

- ✓ Reveal **only what's needed**
- ✓ Keys in the **phone's chip**
- ✓ **Fresh** credential per use
- ✓ **Open source** + bug bounty

Still open

- ! Issuer + shop can **link** you
- ! Only **fixed** questions (`age_over_18`)
- ! ECDSA: **not post-quantum**

Swiss e-ID Scorecard

Gets right

- ✓ Reveal **only what's needed**
- ✓ Keys in the **phone's chip**
- ✓ **Fresh** credential per use
- ✓ **Open source** + bug bounty

Still open

- ! Issuer + shop can **link** you
- ! Only **fixed** questions (`age_over_18`)
- ! ECDSA: **not post-quantum**

Launch planned for **1 December 2026**.
For your company: accept it, and ask only for what you need.

Part 3 · Guarantees

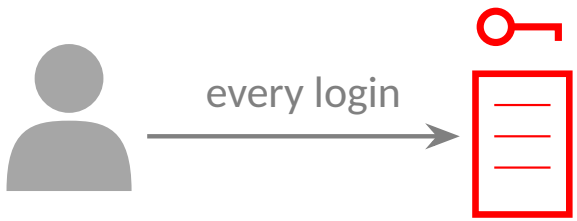
C. No Single Point of Trust





Split the Trust

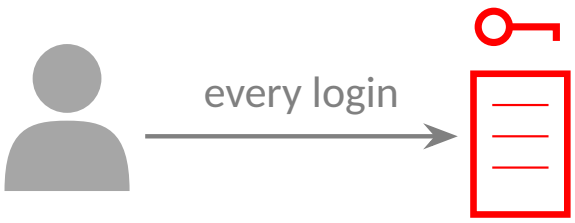
Today: one provider



sees **every login**,
one breach away

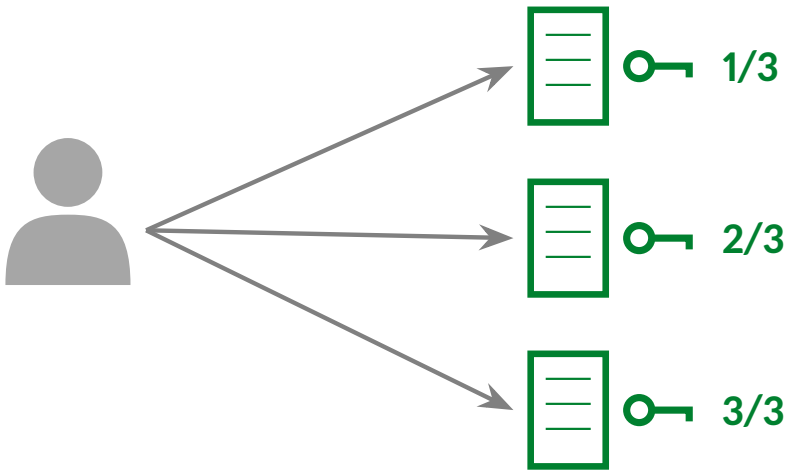
Split the Trust

Today: one provider



sees **every login**,
one breach away

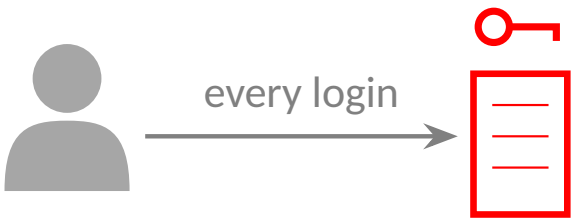
Split: several servers



each holds **a piece** of the key

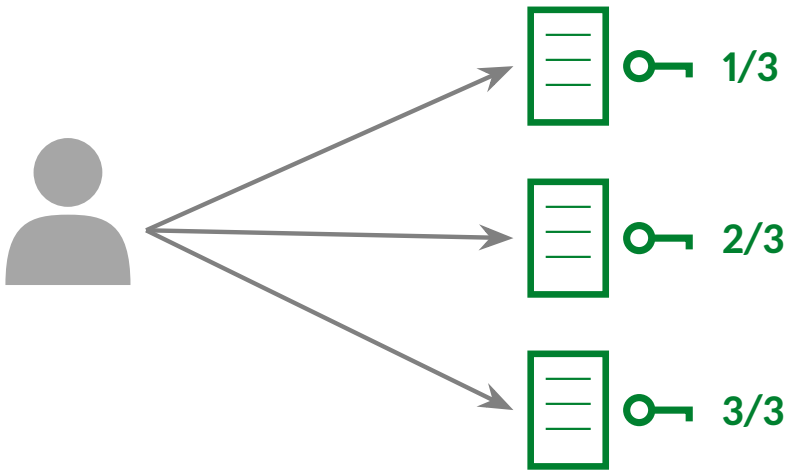
Split the Trust

Today: one provider



sees **every login**,
one breach away

Split: several servers

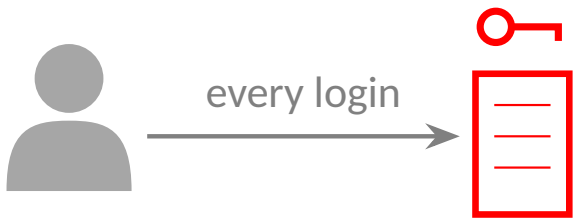


each holds **a piece** of the key

- **No single server** can impersonate you or track you

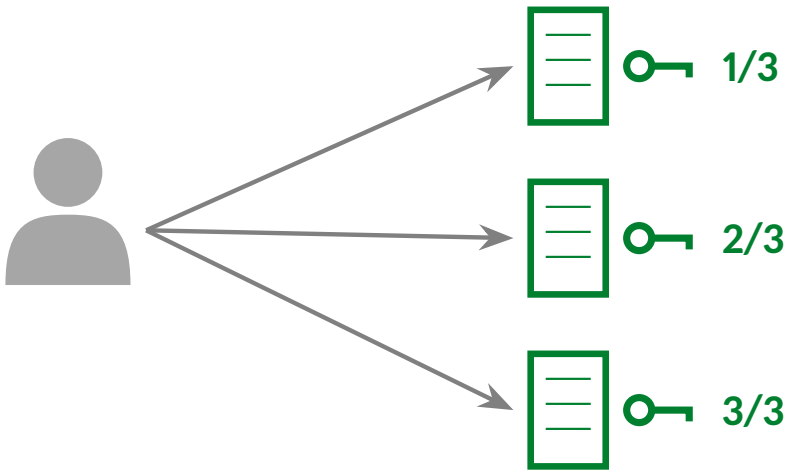
Split the Trust

Today: one provider



sees **every login**,
one breach away

Split: several servers



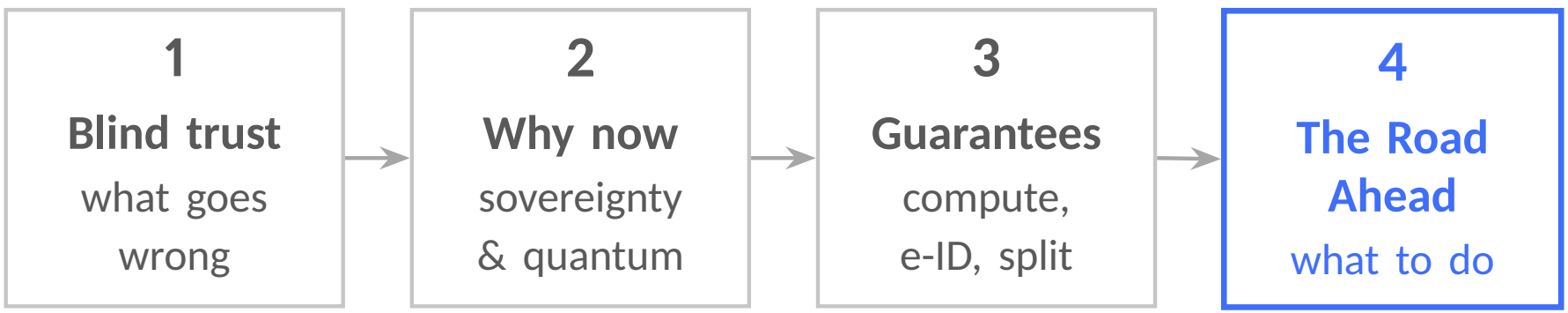
each holds **a piece** of the key

- **No single server** can impersonate you or track you
- Same for **AI agents**: a payment needs approval from **several parties**

Part 4

The Road Ahead

Practical Steps



From research to the decisions you make next week.

What Security Leaders Can Do Today

If you do one thing first

List the AI tools your staff use, and **what data goes in**.

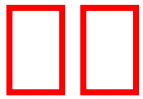
What Security Leaders Can Do Today

If you do one thing first

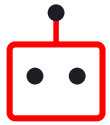
List the AI tools your staff use, and **what data goes in**.



Guarantees, not promises



No single AI supplier



Least power for AI agents



Check **"over 18"** not ID copies



Split critical keys & logins



Post-quantum: start planning

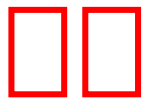
What Security Leaders Can Do Today

If you do one thing first

List the AI tools your staff use, and **what data goes in**.



Guarantees, not promises



No single AI supplier



Least power for AI agents



Check **"over 18"** not ID copies



Split critical keys & logins

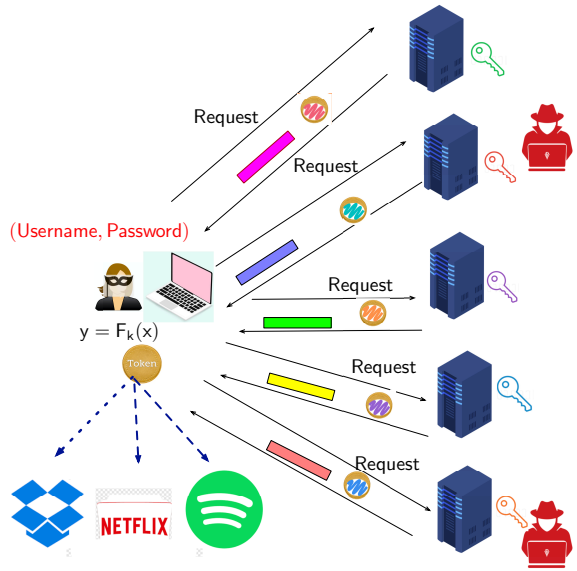


Post-quantum: start planning

Avoid: connecting an AI agent to company email and accounts **without limits**.

Our Work: Trust Without Exposure

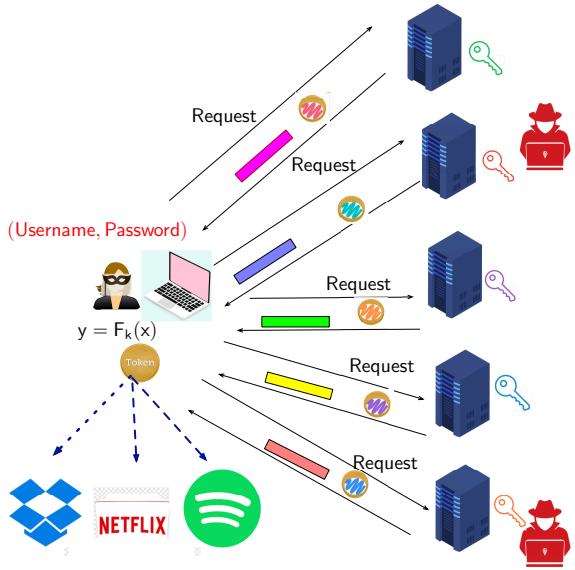
- Identity that **protects** people
  e-ID: prove “over 18”, nothing more
- Users **choose** what they reveal
  anonymous credentials
- Computing on data **nobody sees**
  PSI, MPC, split keys (PASTA)



Decentralised, Private & Secure Authentication

Our Work: Trust Without Exposure

- Identity that **protects** people
  e-ID: prove “over 18”, nothing more
- Users **choose** what they reveal
  anonymous credentials
- Computing on data **nobody sees**
  PSI, MPC, split keys (PASTA)

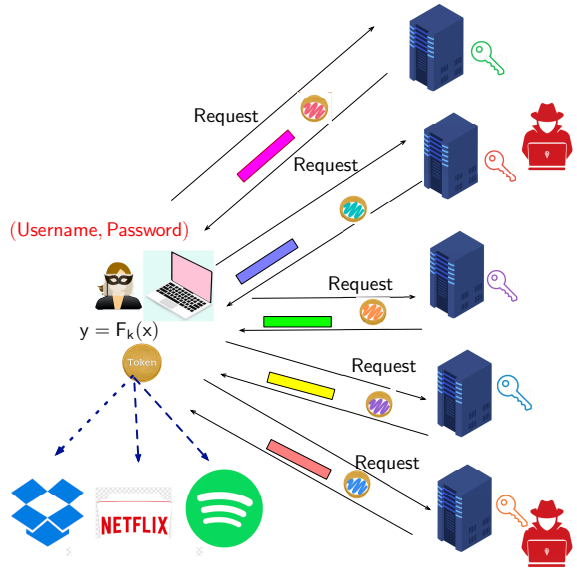


Decentralised, Private & Secure Authentication

Use AI and digital identity **with guarantees**,
even when the world around them is not trustworthy.

Our Work: Trust Without Exposure

- Identity that **protects** people
  e-ID: prove “over 18”, nothing more
- Users **choose** what they reveal
  anonymous credentials
- Computing on data **nobody sees**
  PSI, MPC, split keys (PASTA)



Decentralised, Private & Secure Authentication

Use AI and digital identity **with guarantees**,
even when the world around them is not trustworthy.

Let’s work together: joint projects · Innosuisse · theses

Our Team



Katerina Mitrokotsa

Full Professor in Cyber Security



Angelina Makri

Head of Operations & Administration



Nan Cheng

Post-Doctoral Researcher



Constance Bau

Doctoral Researcher



Liujun Yu

Doctoral Researcher



Florias Papadopoulos

Doctoral Researcher



Ioannis Katis

Doctoral Researcher



Wilson Tsuata

Doctoral Researcher

Three Things to Take Home

1. AI today runs on **blind trust** in providers, agents.

Three Things to Take Home

1. AI today runs on **blind trust** in providers, agents.

2. Cryptography gives **verifiable guarantees**: compute without seeing, prove without revealing, no single point of trust.

Three Things to Take Home

1. AI today runs on **blind trust** in providers, agents.
2. Cryptography gives **verifiable guarantees**: compute without seeing, prove without revealing, no single point of trust.
3. First step: know **what data goes into which AI tool**.

Three Things to Take Home

1. AI today runs on **blind trust** in providers, agents.

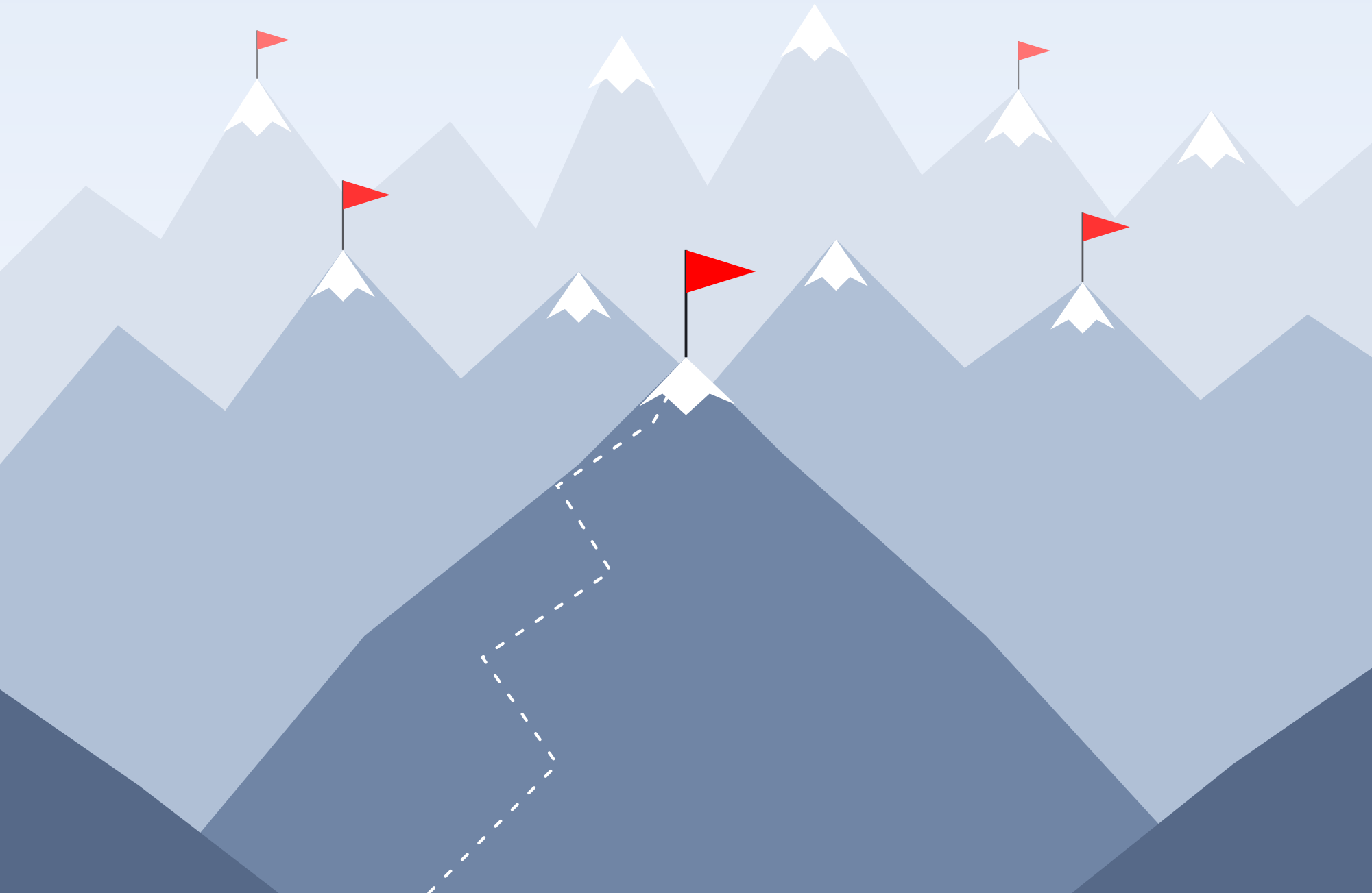
2. Cryptography gives **verifiable guarantees**: compute without seeing, prove without revealing, no single point of trust.

3. First step: know **what data goes into which AI tool**.



The cat is already out of the bag:
use AI **without giving away**
your data or your identity.

Are we done?



Keep Going



Thank you for your Attention!

Contact: https://cybersecurity.unisg.ch

Why should I care?

Why should I care?

€7.1 bn

GDPR fines since 2018

Sources: DLA Piper GDPR Fines and Data Breach Survey (Jan. 2026); Swiss Federal Act on Data Protection (nFADP), Art. 60–63.

Why should I care?

€7.1 bn

GDPR fines since 2018

443 / day

data breaches reported in 2025 (+22%)

Sources: DLA Piper GDPR Fines and Data Breach Survey (Jan. 2026); Swiss Federal Act on Data Protection (nFADP), Art. 60–63.

Why should I care?

€7.1 bn

GDPR fines since 2018

443 / day

data breaches reported in 2025 (+22%)

CHF 250k

Swiss nFADP fine, paid by the **person** responsible

Sources: DLA Piper GDPR Fines and Data Breach Survey (Jan. 2026); Swiss Federal Act on Data Protection (nFADP), Art. 60–63.

Why should I care?

€7.1 bn

GDPR fines since 2018

443 / day

data breaches reported in 2025 (+22%)

CHF 250k

Swiss nFADP fine, paid by the **person** responsible

Largest GDPR fines		
Meta	**€1.2 bn**	2023
Amazon	€746 m	2021
TikTok	€530 m	2025
LinkedIn	€310 m	2024

Sources: DLA Piper GDPR Fines and Data Breach Survey (Jan. 2026); Swiss Federal Act on Data Protection (nFADP), Art. 60–63.

Why should I care?

€7.1 bn

GDPR fines since 2018

443 / day

data breaches reported in 2025 (+22%)

CHF 250k

Swiss nFADP fine, paid by the **person** responsible

Largest GDPR fines		
Meta	**€1.2 bn**	2023
Amazon	€746 m	2021
TikTok	€530 m	2025
LinkedIn	€310 m	2024

Fines come **after** the damage. **Cryptography prevents it.**

Sources: DLA Piper GDPR Fines and Data Breach Survey (Jan. 2026); Swiss Federal Act on Data Protection (nFADP), Art. 60–63.

Switzerland & IT Security

64,733

cyber incidents reported in 2025



Source: NCSC Annual Report 2025 (Feb. 2026); NCSC Semi-Annual Report 2026/1 (Aug. 2026).

Switzerland & IT Security

64,733

cyber incidents reported in 2025

27,128

already in the first half of 2026



Source: NCSC Annual Report 2025 (Feb. 2026); NCSC Semi-Annual Report 2026/1 (Aug. 2026).

Switzerland & IT Security

64,733

cyber incidents reported in 2025

27,128

already in the first half of 2026

79

ransomware cases in H1 2026

AI

now writes the phishing and the scam calls



Source: NCSC Annual Report 2025 (Feb. 2026); NCSC Semi-Annual Report 2026/1 (Aug. 2026).