

# **Towards Auditable and Intelligent Privacy-Preserving Record Linkage**

Thiago Nóbrega

PPGCC/UFCG (2018 - 2022)

Advisors:

Carlos Eduardo Santos Pires

Dimas Cassimiro do Nascimento Filho

September 2023

# AGENDA

- I. Introduction
- II. PPRL Comparison step Auditability
- III. Unsupervised Classification step for PPRL
- IV. Deep Learning-based Classifiers for PPRL
- V. Final Arguments



# Introduction

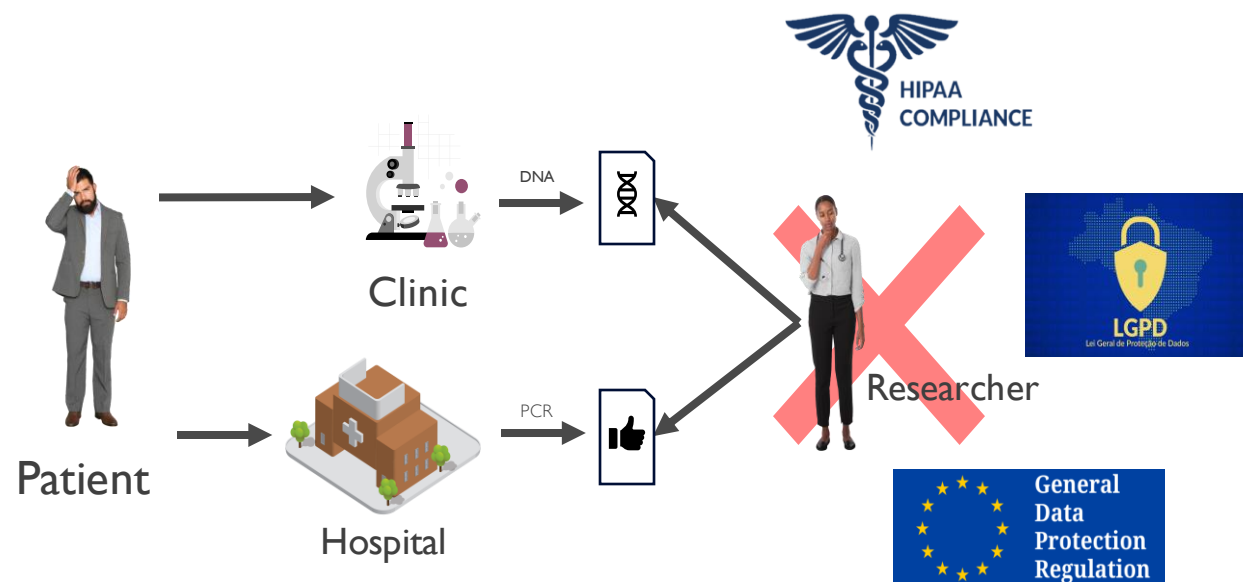
- I. Introduction
- II. PPRL Comparison step Auditability
- III. Unsupervised Classification step for PPRL
- IV. Deep Learning-based Classifiers for PPRL
- V. Final Arguments

# Motivation

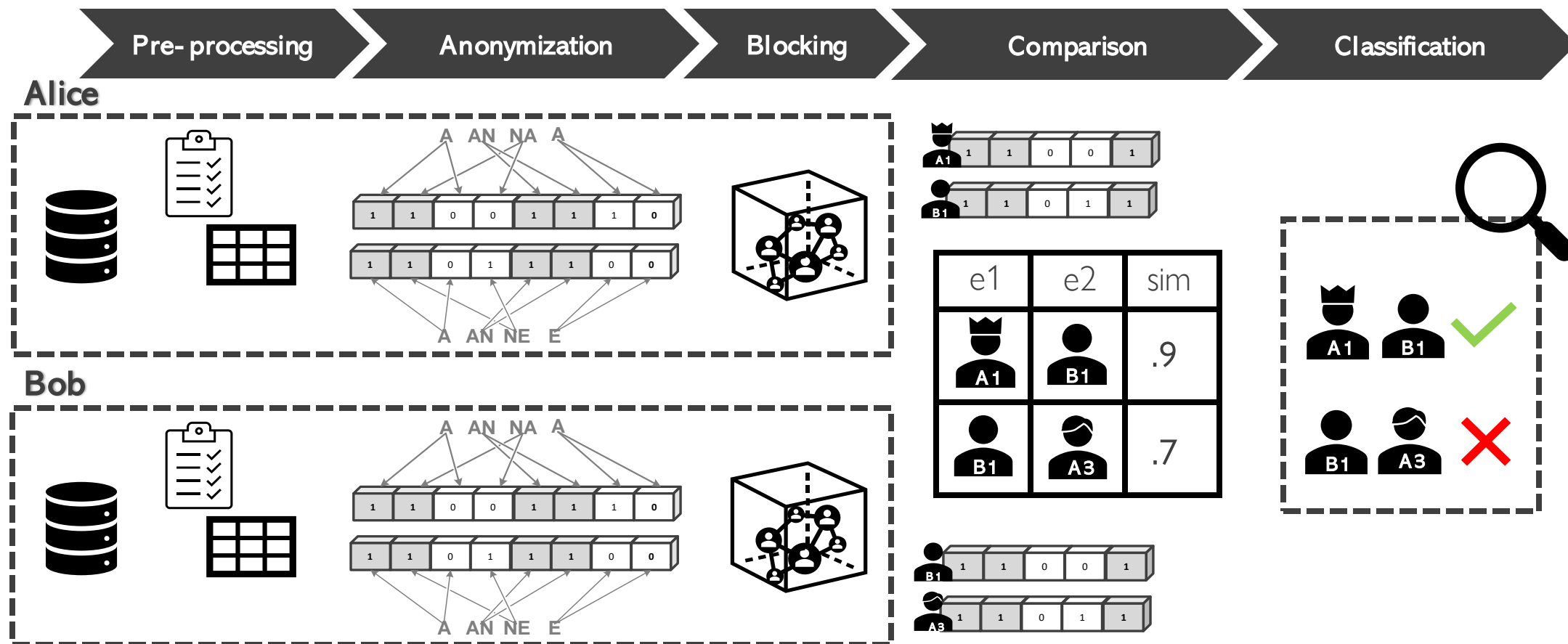
PPRL Goal

**“The goal of PPRL is to perform record linkage without revealing entities identifiers”.**

❑ Absence of unique identifiers



# PPRL Process



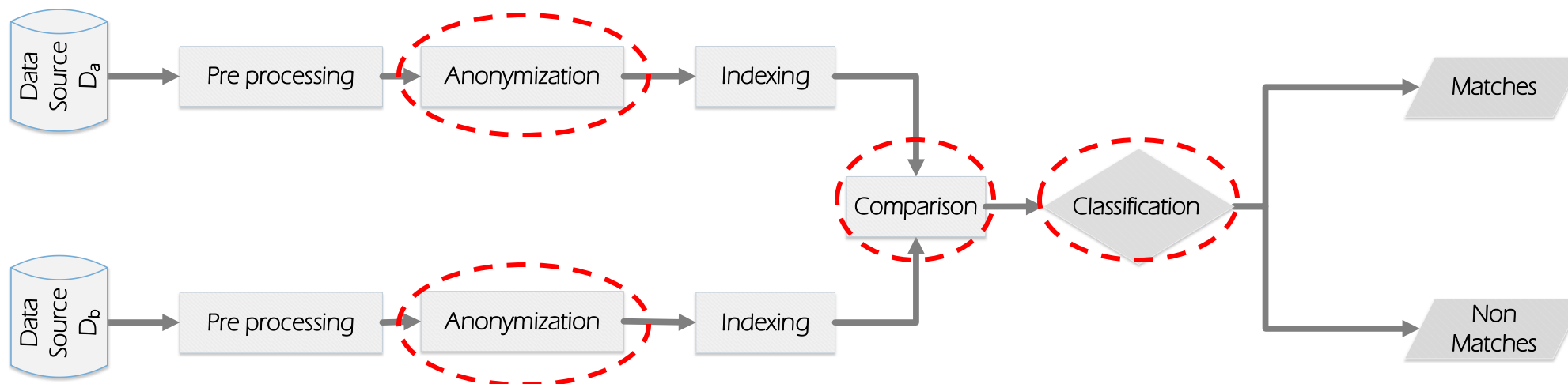
# PPRL Limitations

## Privacy

- Adversary Model

## Quality

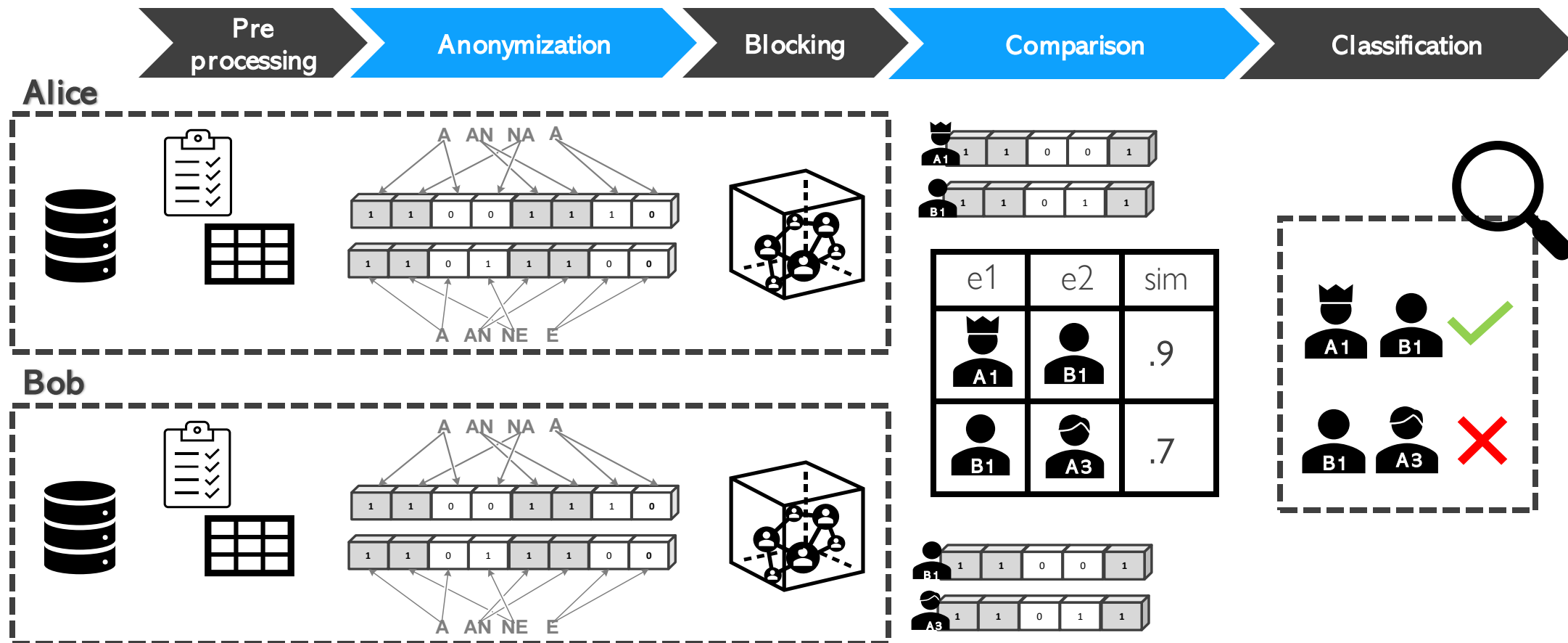
- Low Linkage Quality



# PPRL Comparison step Auditability

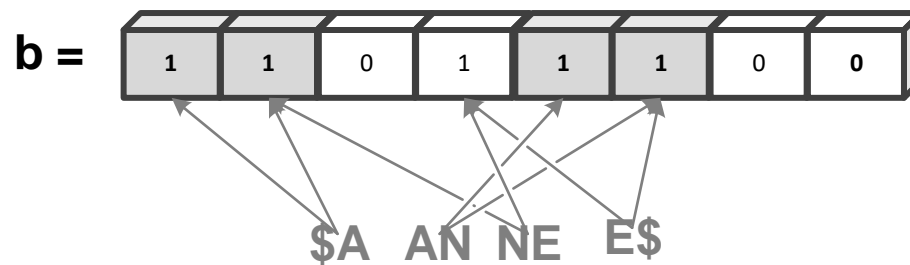
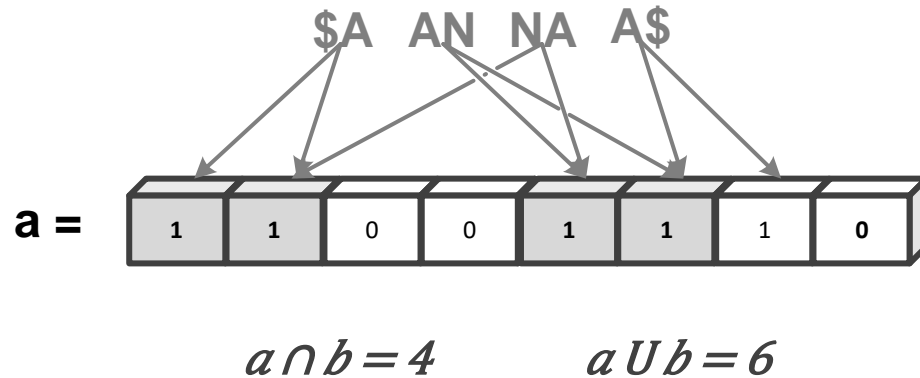
- I. Introduction
- II. PPRL Comparison step Auditability**
- III. Unsupervised Classification step for PPRL
- IV. Deep Learning-based Classifiers for PPRL
- V. Final Arguments

# PPRL Comparison step Limitations





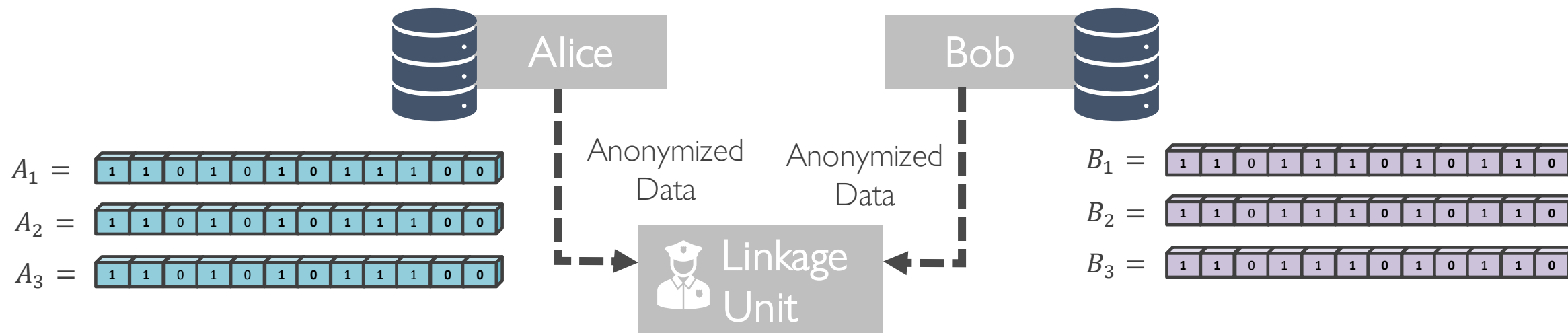
# PPRL Anonimization



$$Dice(a, b) = \frac{2 \times |a \cap b|}{|a| + |b|} \rightarrow \frac{8}{10} \rightarrow 0.8$$

$$Jaccard(a, b) = \frac{|a \cap b|}{|a \cup b|} \rightarrow \frac{4}{6} \rightarrow 0.65$$

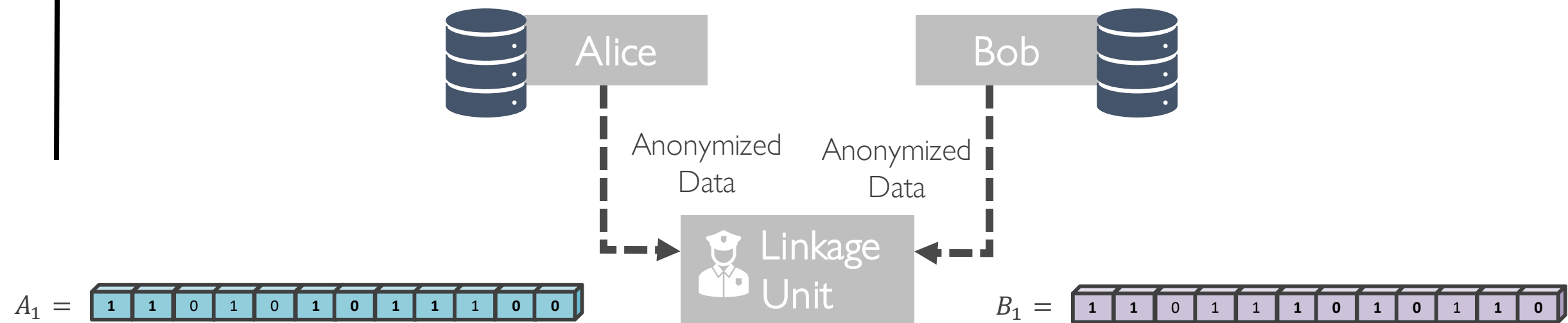
# PPRL Comparison step Limitations

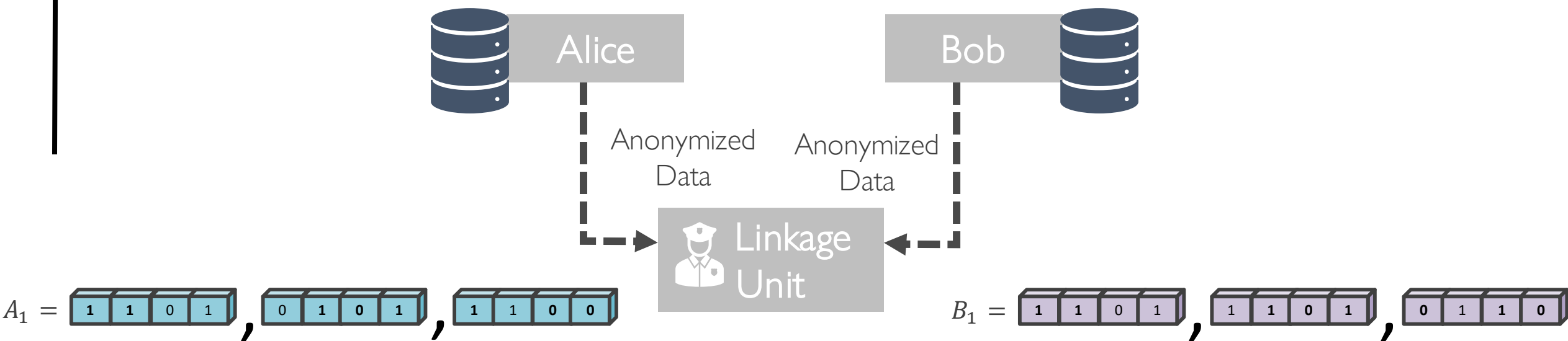


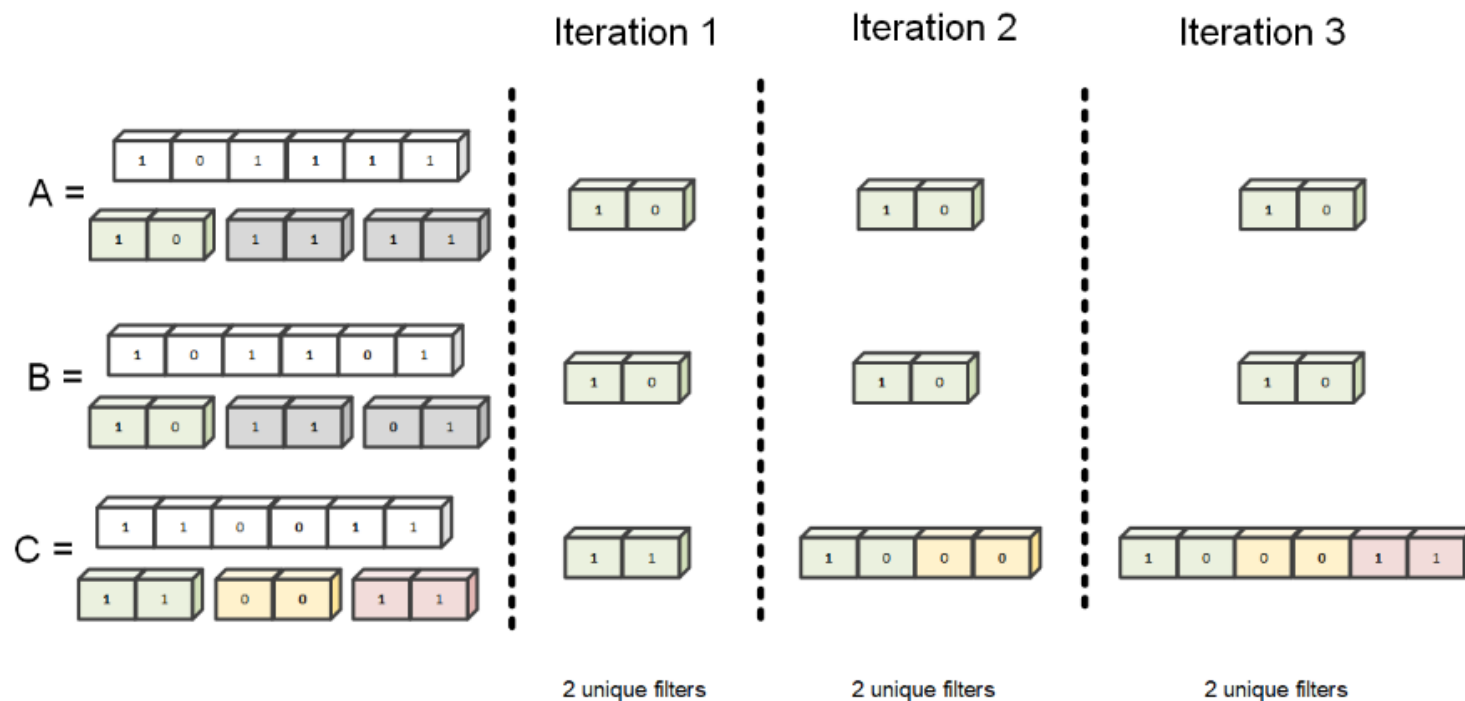
- i. Shared Information
- ii. Need to fully trust other PPRL parties

# SBF

Splitting Bloom Filter







- Indistinguishability
- Uncertainty
- Split Filtering

“What is the impact of utilizing partial fragments (splits) of the original Bloom Filter on data comparisons?”

$$Jaccard(\overset{\text{ANA}}{\begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|} \hline 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ \hline \end{array}}, \overset{\text{ANE}}{\begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|} \hline 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ \hline \end{array}}) = Jaccard(\begin{array}{|c|c|c|c|} \hline 1 & 1 & 0 & 1 \\ \hline \end{array}, \begin{array}{|c|c|c|c|} \hline 1 & 1 & 0 & 1 \\ \hline \end{array}) + \textit{error}$$

$$\textit{error} = \binom{\frac{l}{s}}{x} p^x (1-p)^{\frac{l}{s}-x}$$



$$Jaccard(\text{ANA}, \text{ANE}) \cong \frac{1}{s} \sum_1^s Jaccard(\phi_A^i, \phi_B^i)$$

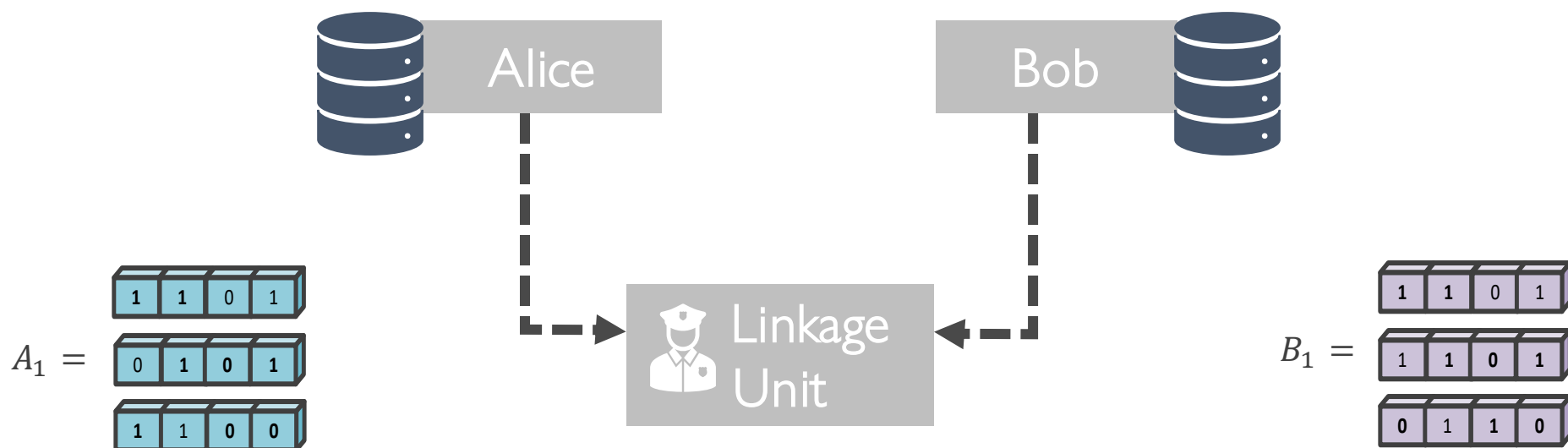
$$\phi_A = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \quad \phi_B = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

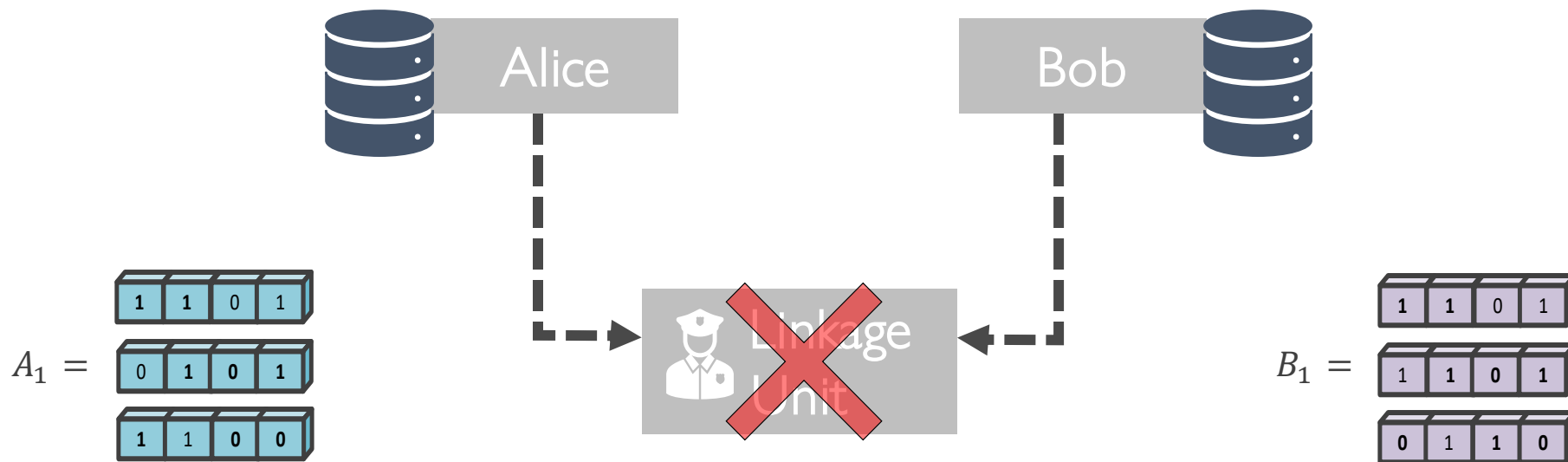


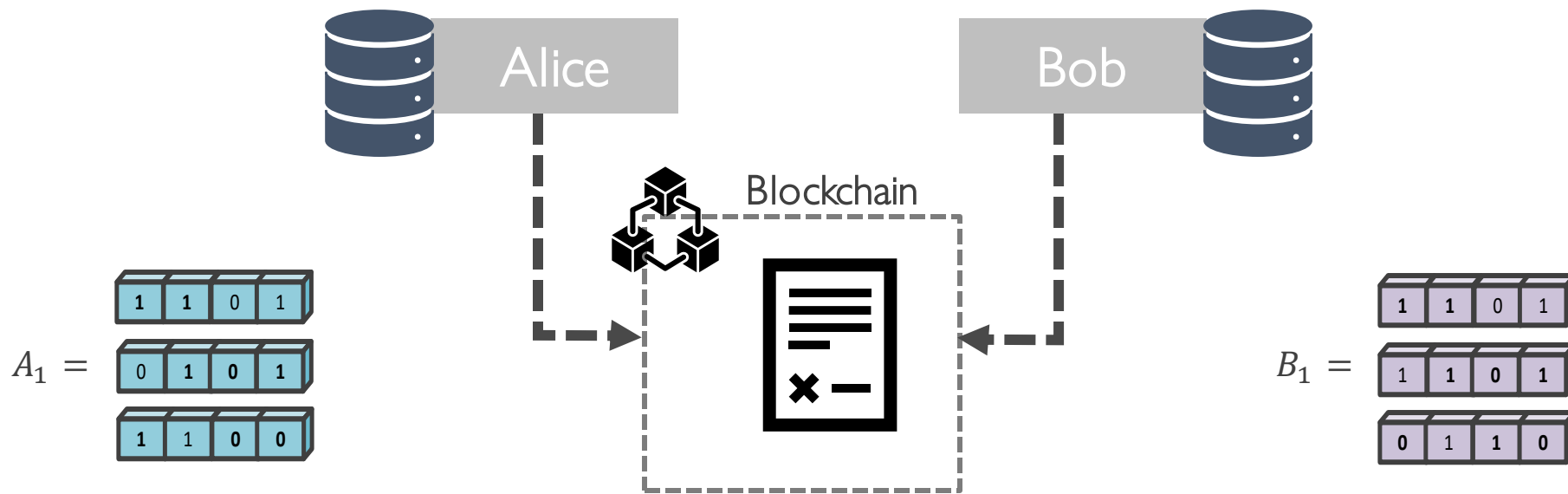
# ABEL

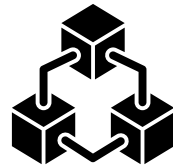
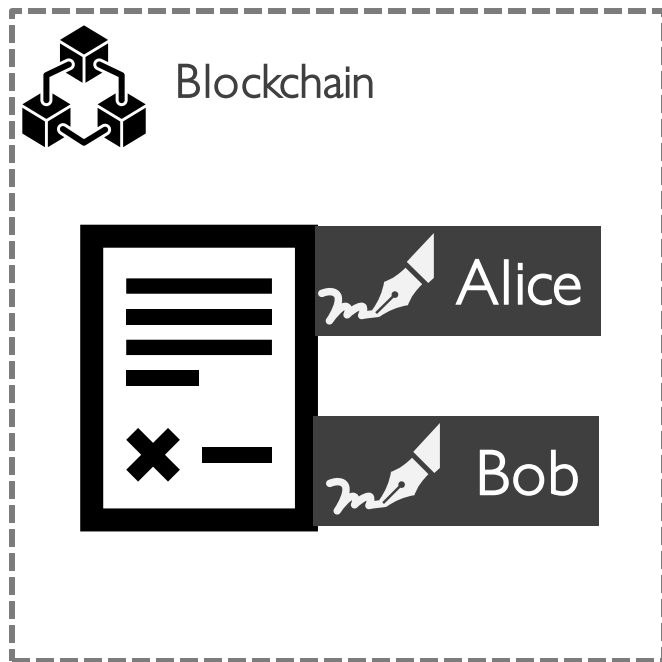
Auditable Blockchain-based PPRL

# PPRL Comparison step Limitations

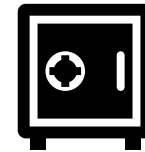








Decentralized

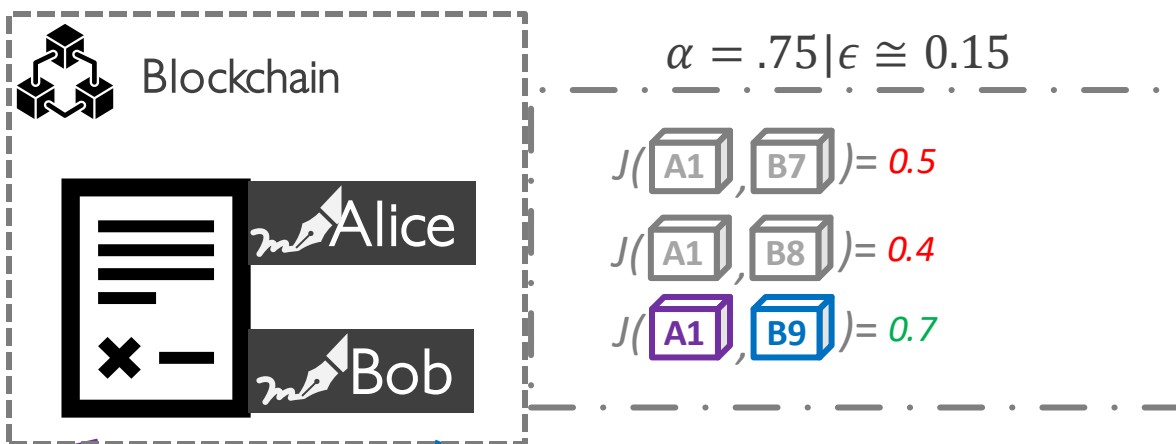


Temper Evident



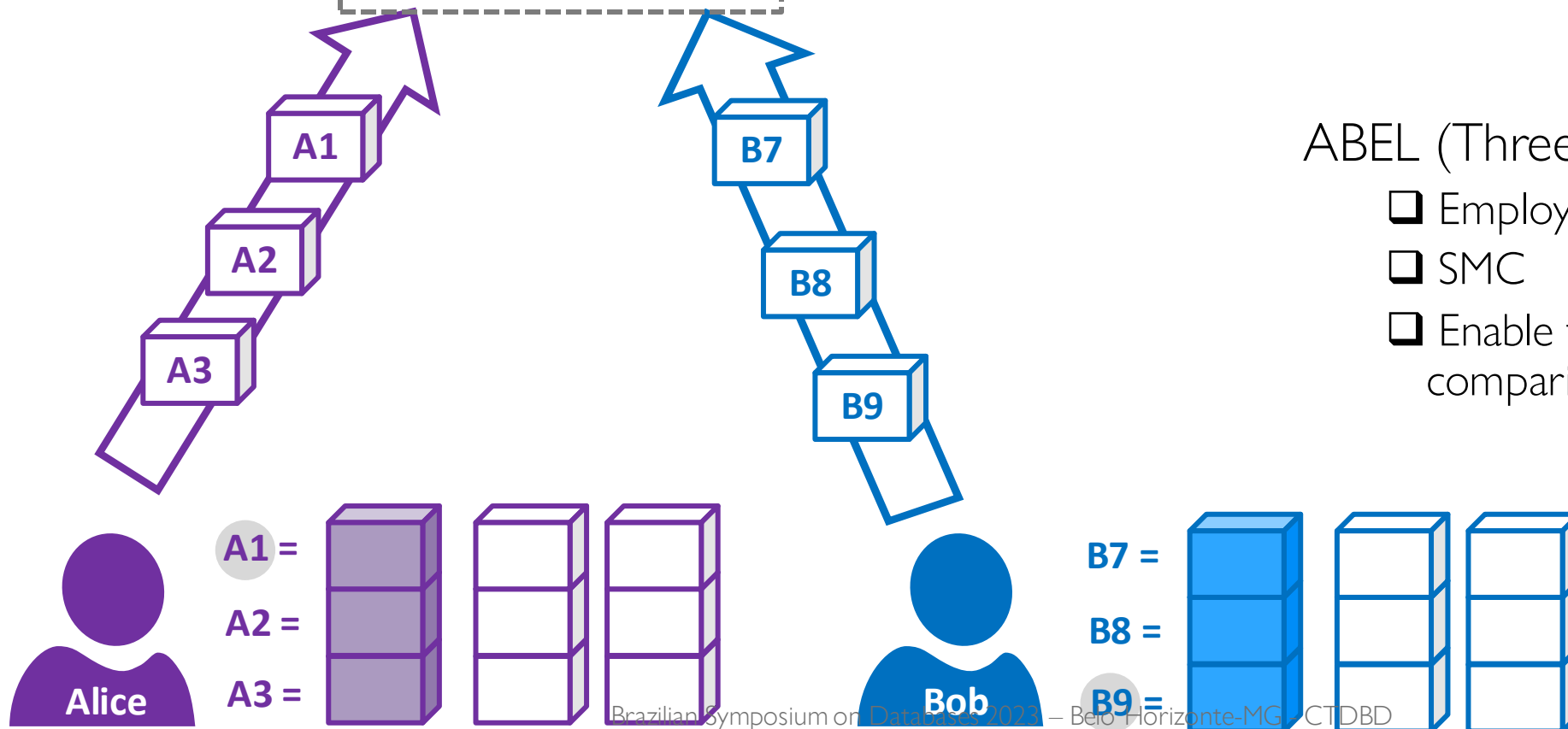
Transparent

# ABEL



## ABEL (Three Party Protocol)

- ☐ Employ SBC error ( $\epsilon$ )
- ☐ SMC
- ☐ Enable the auditability of PPRL comparison step



# EVALUATION

SBF & ABEL Linkage Quality and Privacy



# Evaluation Metrics



## Linkage Quality

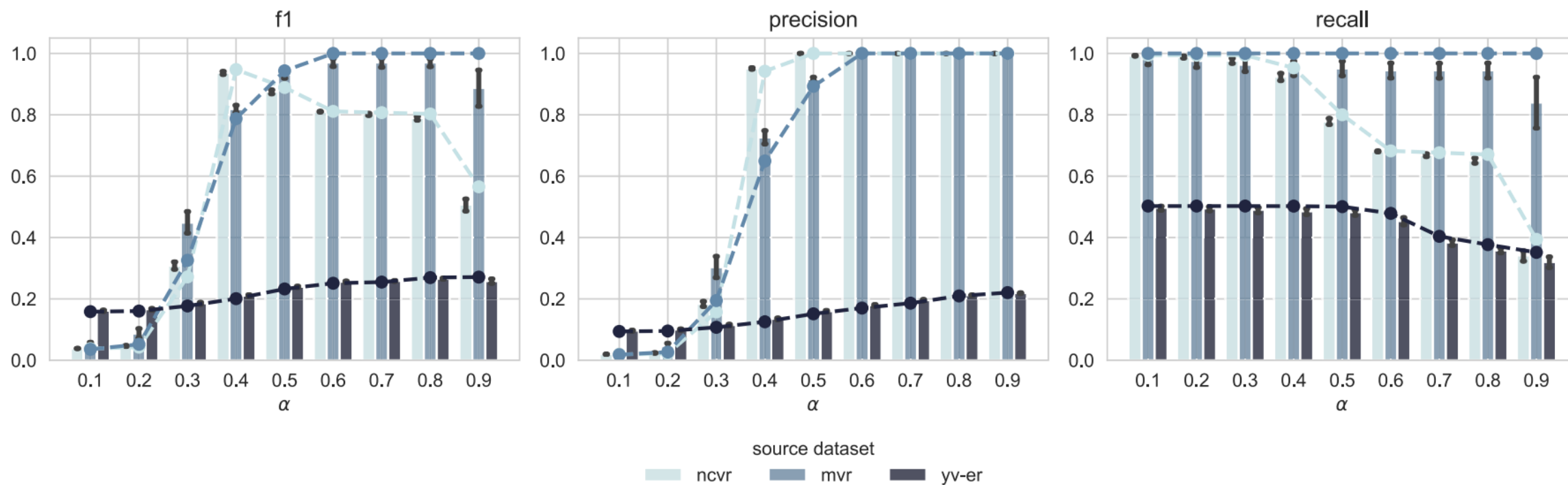
- ☐ Compare ABEL against the standard comparison step
  - ☐ precision
  - ☐ recall
  - ☐ F1



## Privacy

- ☐ Measure ABEL privacy cost
  - ☐ Reidentification rate against a customized attack

# Overall quality results



# Privacy

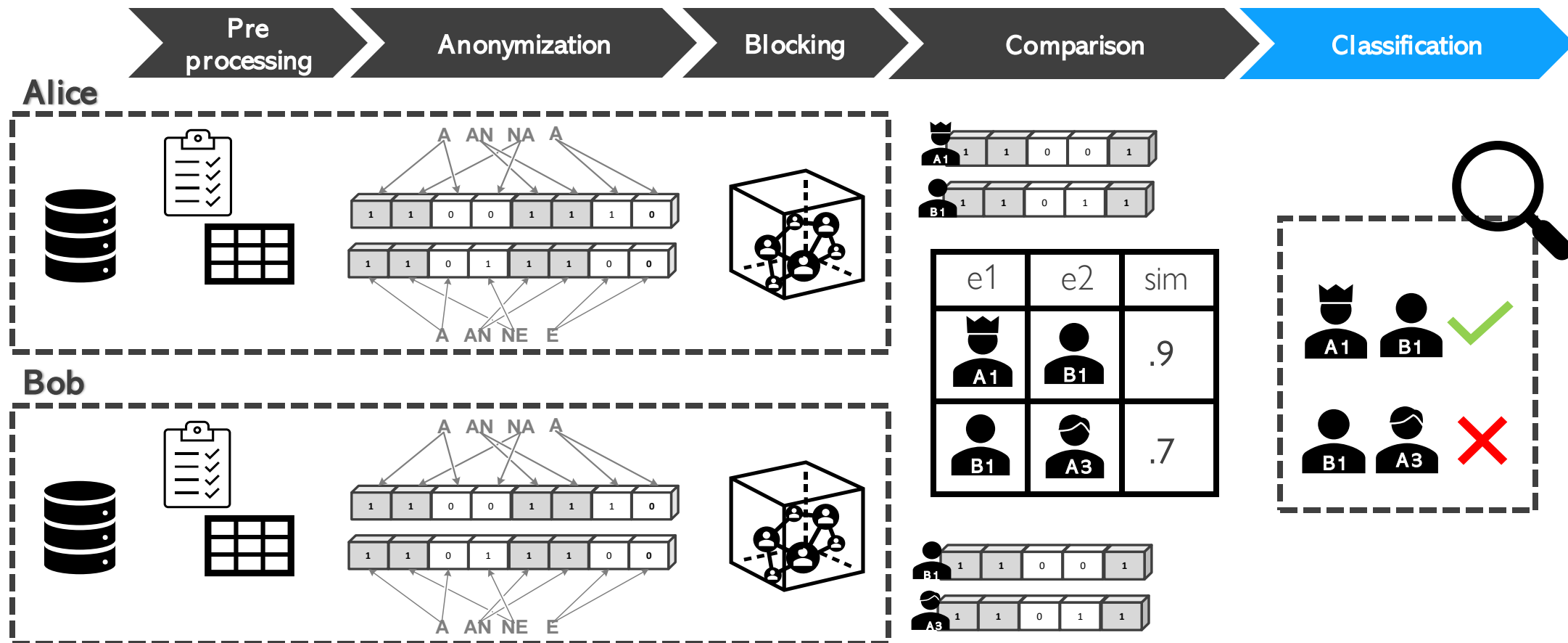
## Attack Effectiveness Comparison

Publication	Dataset	Num BF	1-to-1 correct	1-to-1 correct %
Pattern-mining based for PPRL [30]	NCVR diff. attr.	>200k	>49k	25%
Precise and Fast Cryptanalysis for PPRL [26]	NCVR First + Last name	>200k	-	20.7%
A Graph Matching Attack [169]	NCVR First + Last name	100k	-	>50% accuracy
				>90% accuracy
this work	NCVR First + Last name	10k	1,346	13%





# Unsupervised Classification step for PPRL

- I. Introduction
- II. PPRL Comparison step Auditability
- III. **Unsupervised Classification step for PPRL**
- IV. Deep Learning-based Classifiers for PPRL
- V. Final Arguments

# Limitations of PPRL Classification step



# Limitations of PPRL Classification Step

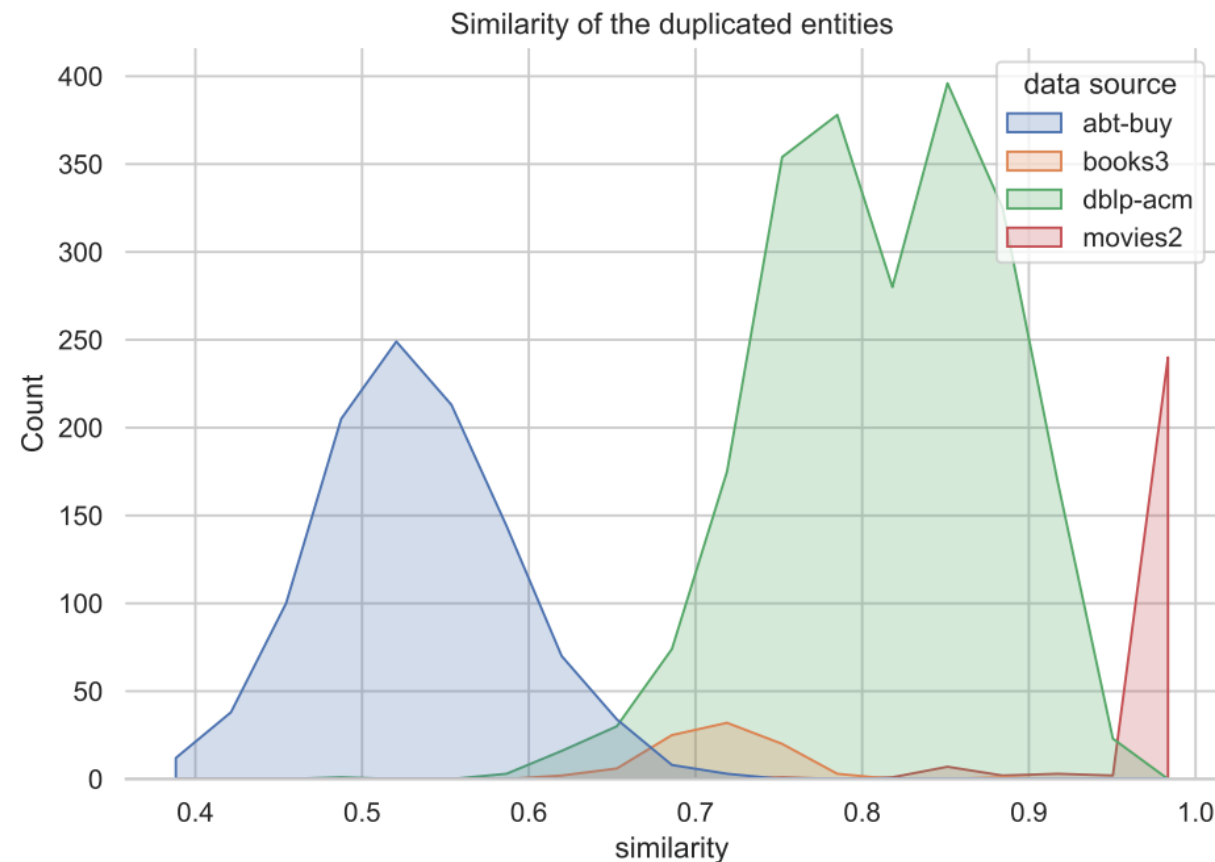
e1	e2	sim
 A1	 B1	.9
 B1	 A3	.8

## Threshold

- Hard to tune

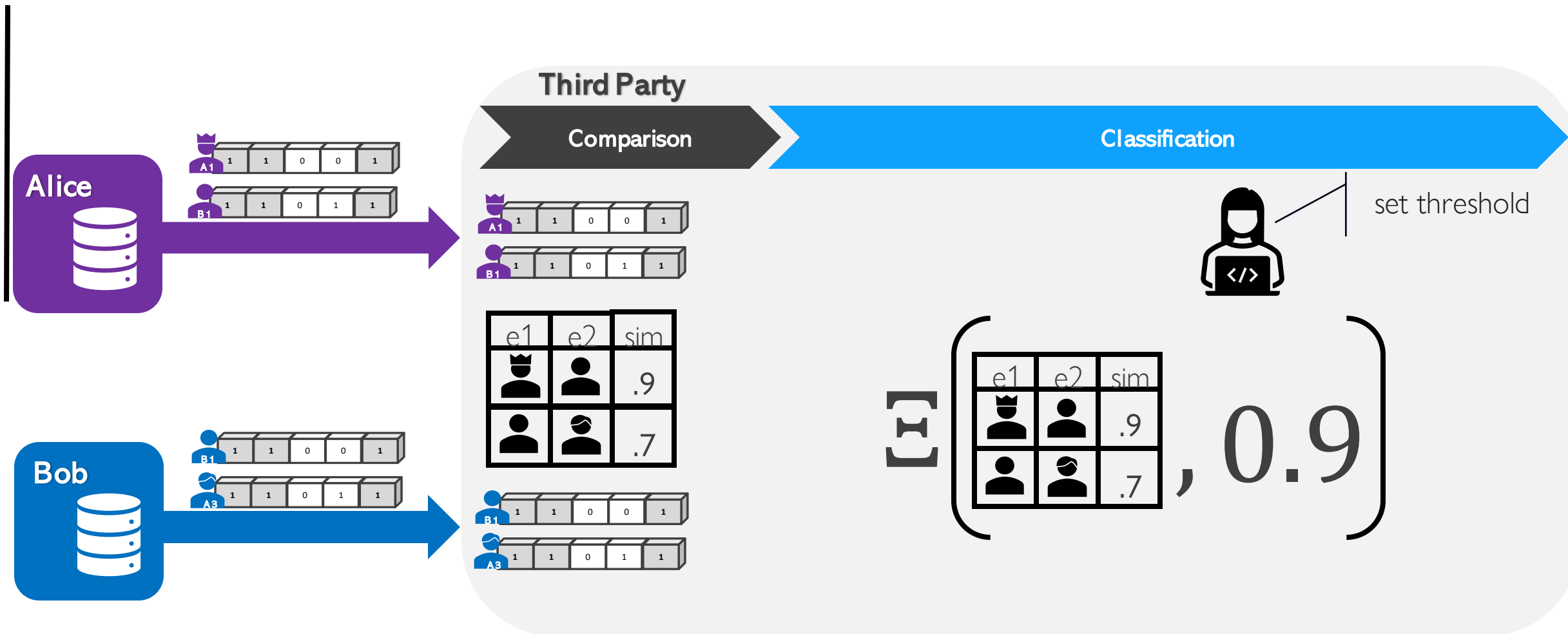
## Automatic (Machine Learning based)

- Privacy Limitations
- No labeled data
- No oracle available
- No access to actual data



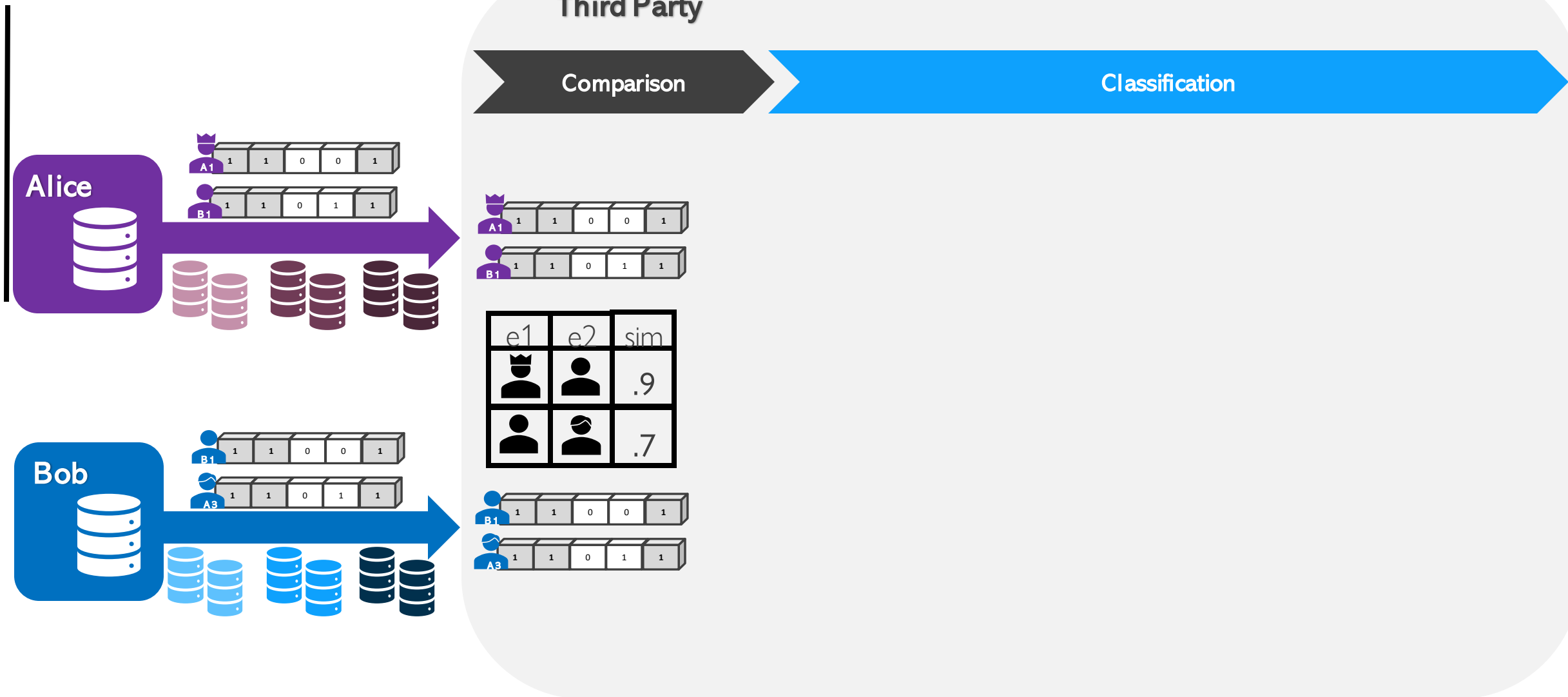
# AT-UC

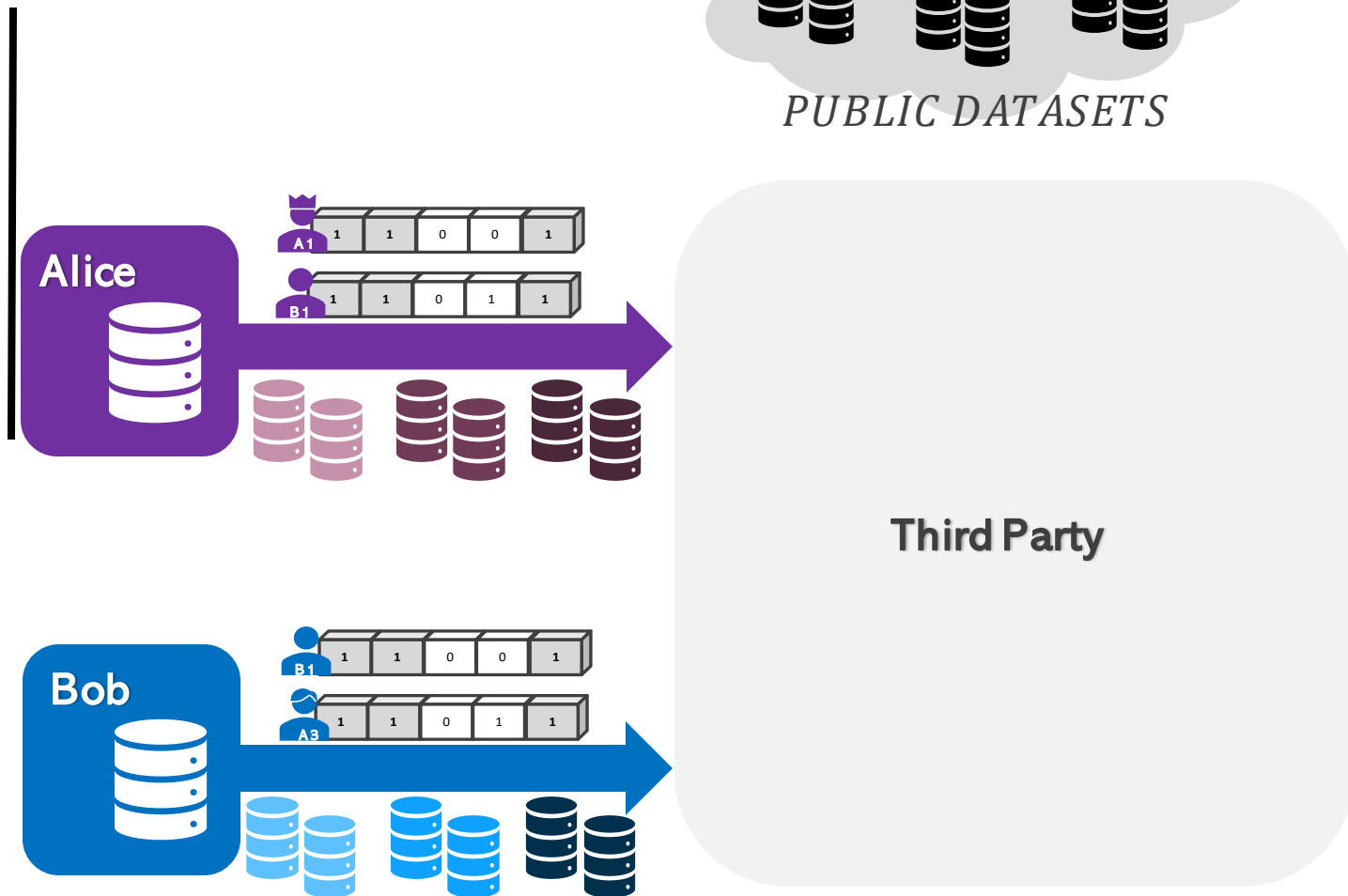
Auto-Tuned Unsupervised Classification

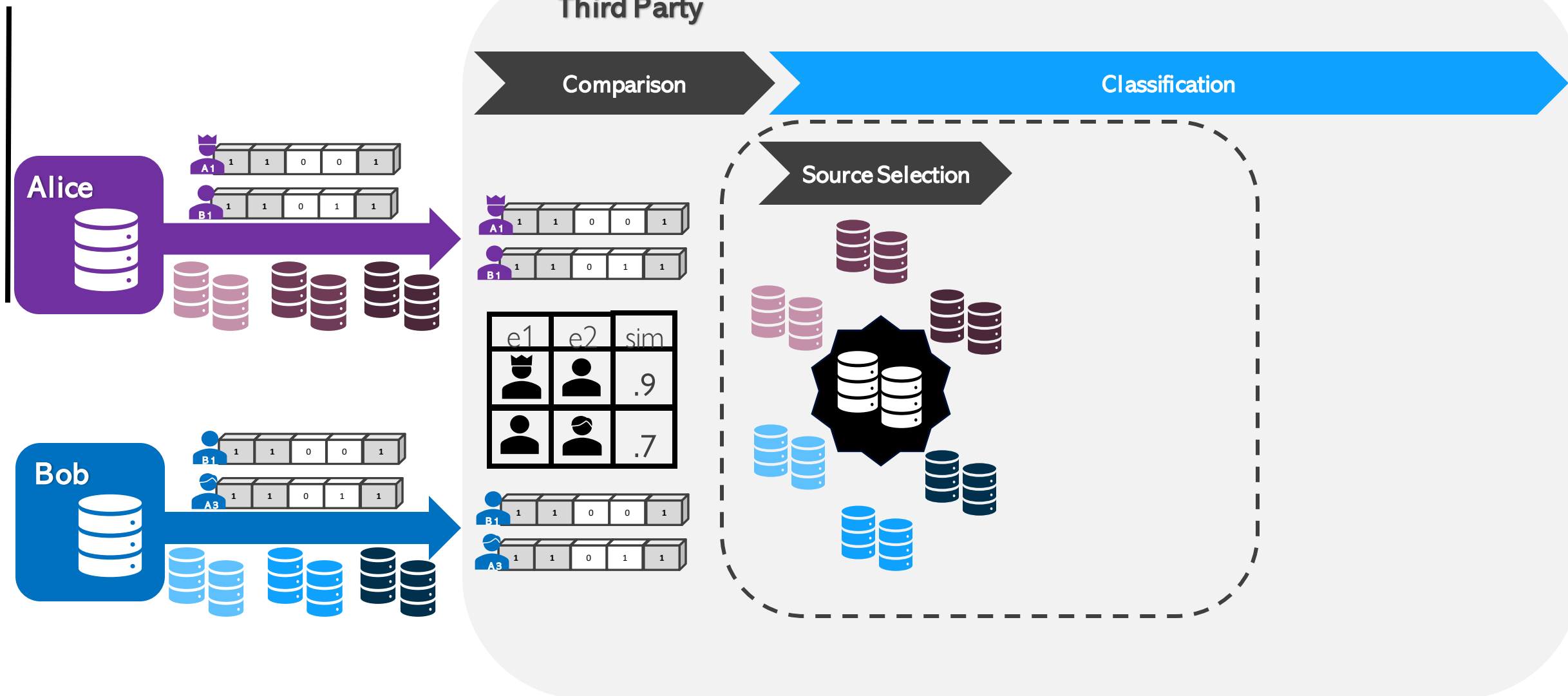


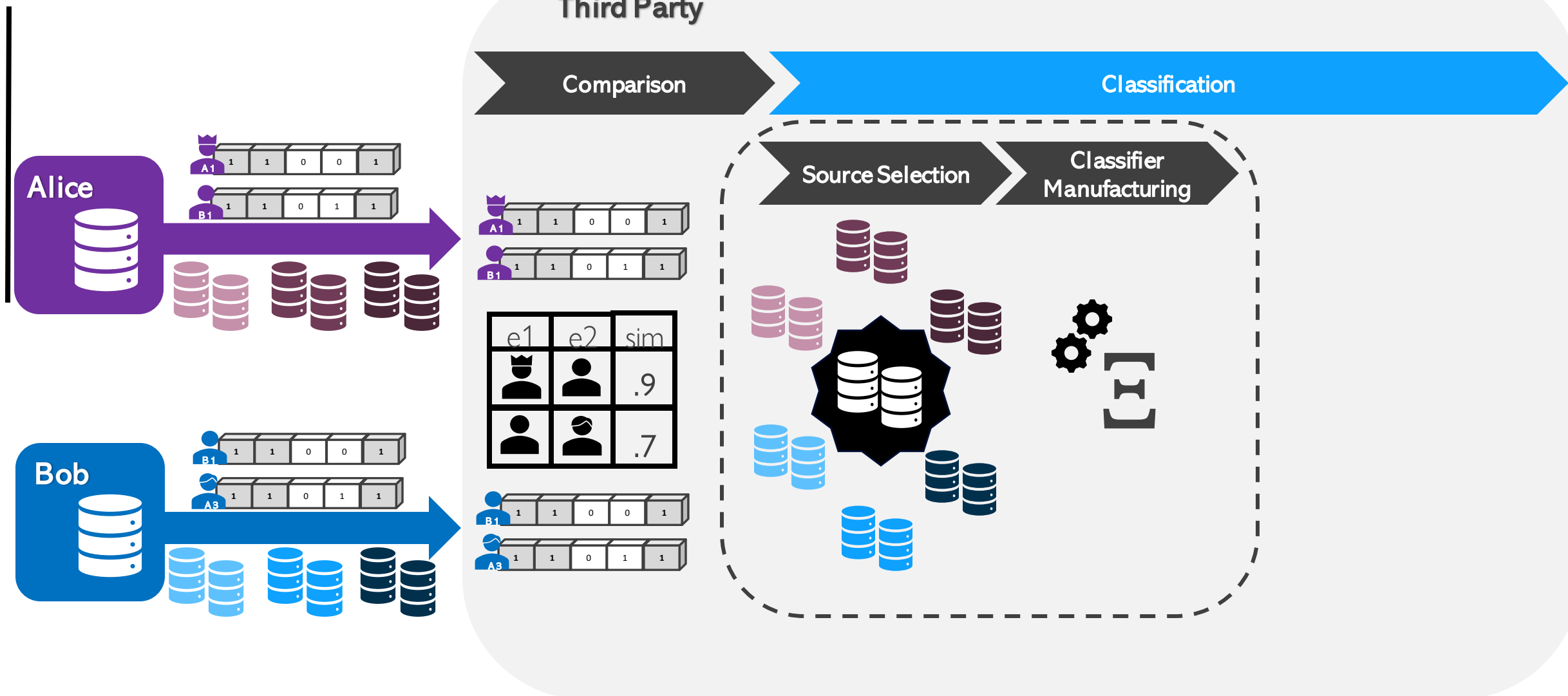
$$\mathbb{E}(S, \lambda) = \{x \in S \mid x \geq \lambda\}$$

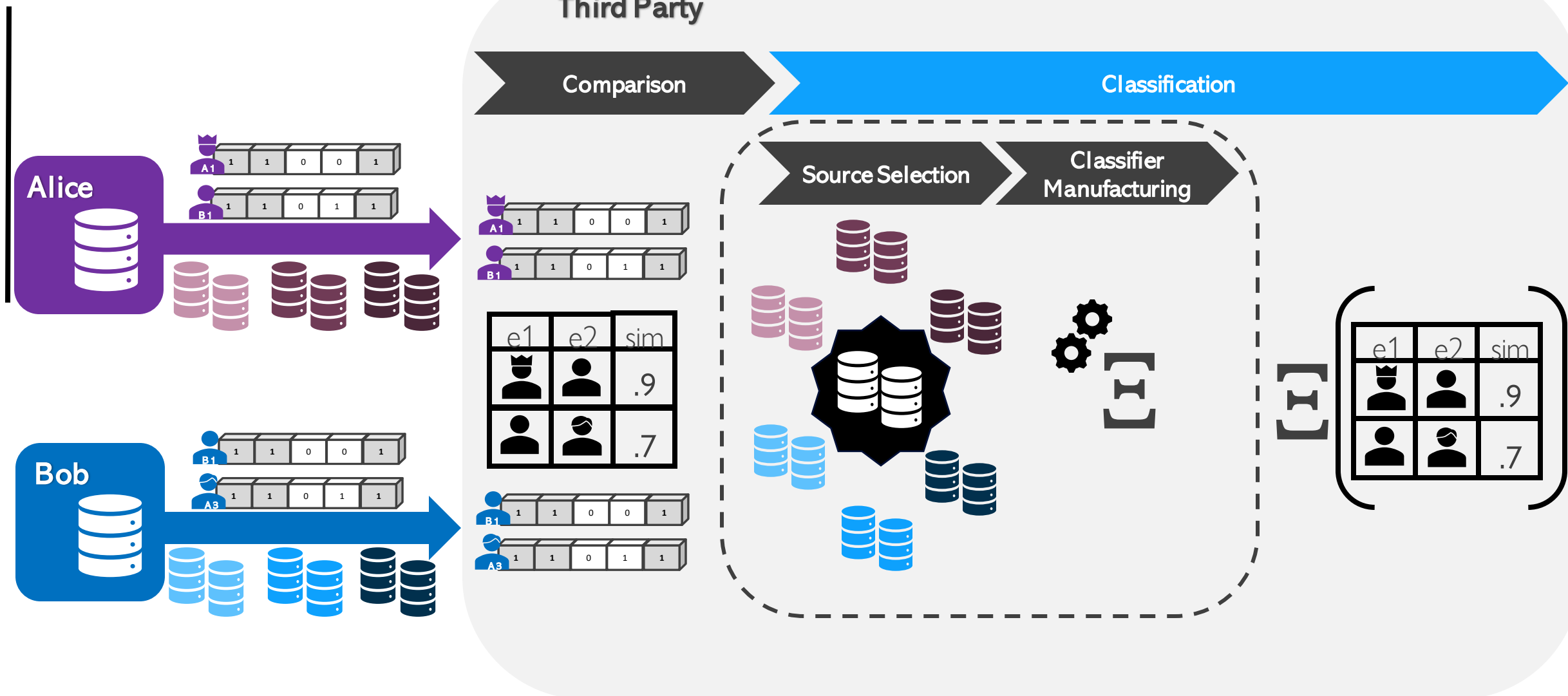












# EVALUATION

AT-UC Linkage Quality, Efficiency and Privacy

# Evaluation Metrics



## Linkage Quality

- ☐ Measure Source Selection stage impact
  - ☐ F1
- ☐ Compare AT-UC against a baseline and competitors
  - ☐ precision, recall, F1



## Privacy

- ☐ Measure AT-UC privacy cost
  - ☐ Reidentification rate against an attack

# Linkage Quality

Linkage Quality Comparison

## Baseline

- threshold

## Competitors

- transER
- coral
- naive

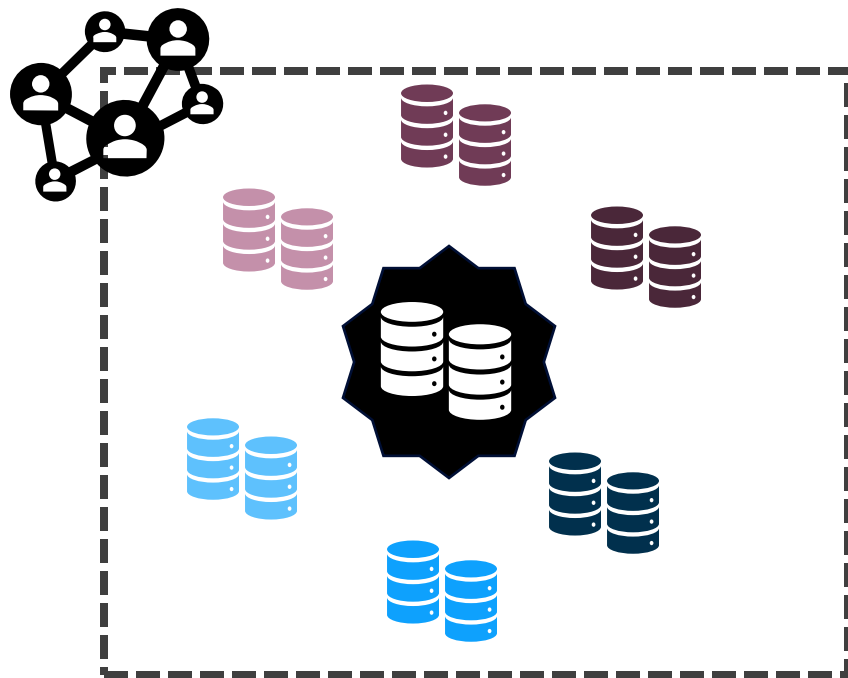
approach	target	source	precision	recall	f1
<b>at-uc</b>	census	restaurants	14%	77%	22%
	mvr	census	97%	99%	98%
	ncvr	census	100%	78%	87%
	tse	books	84%	87%	85%
	yv-er	tse	66%	56%	61%
<b>naive</b>	census	best	1%	100%	2%
	mvr	best	5%	99%	9%
	ncvr	best	2%	91%	4%
	tse	best	0%	13%	0%
	yv-er	best	43%	87%	58%
<b>transER</b>	census	-	-	-	-
	mvr	best	83%	99%	90%
	ncvr	best	74%	99%	85%
	tse	best	3%	100%	6%
	yv-er	-	-	-	-
<b>coral</b>	census	5-best	9% $\pm$ 6%	49% $\pm$ 6%	15% $\pm$ 9%
	mvr	5-best	96% $\pm$ 1%	88% $\pm$ 1%	91% $\pm$ 8%
	ncvr	5-best	99%	49%	66%
	tse	5-best	81% $\pm$ 1%	80% $\pm$ 1%	80% $\pm$ 8%
	yv-er	5-best	93% $\pm$ 15%	40% $\pm$ 15%	46% $\pm$ 45%
<b>threshold</b>	census	5-best	8% $\pm$ 5%	61% $\pm$ 5%	14% $\pm$ 9%
	mvr	5-best	49% $\pm$ 46%	71% $\pm$ 46%	51% $\pm$ 44%
	ncvr	5-best	49% $\pm$ 47%	67% $\pm$ 47%	49% $\pm$ 42%
	tse	5-best	27% $\pm$ 24%	71% $\pm$ 34%	32% $\pm$ 31%
	yv-er	5-best	60% $\pm$ 42%	64% $\pm$ 42%	60% $\pm$ 39%



# Privacy

## Experimental Setup

- A Graph Matching Attack (Radanbuge et. al, 2020)

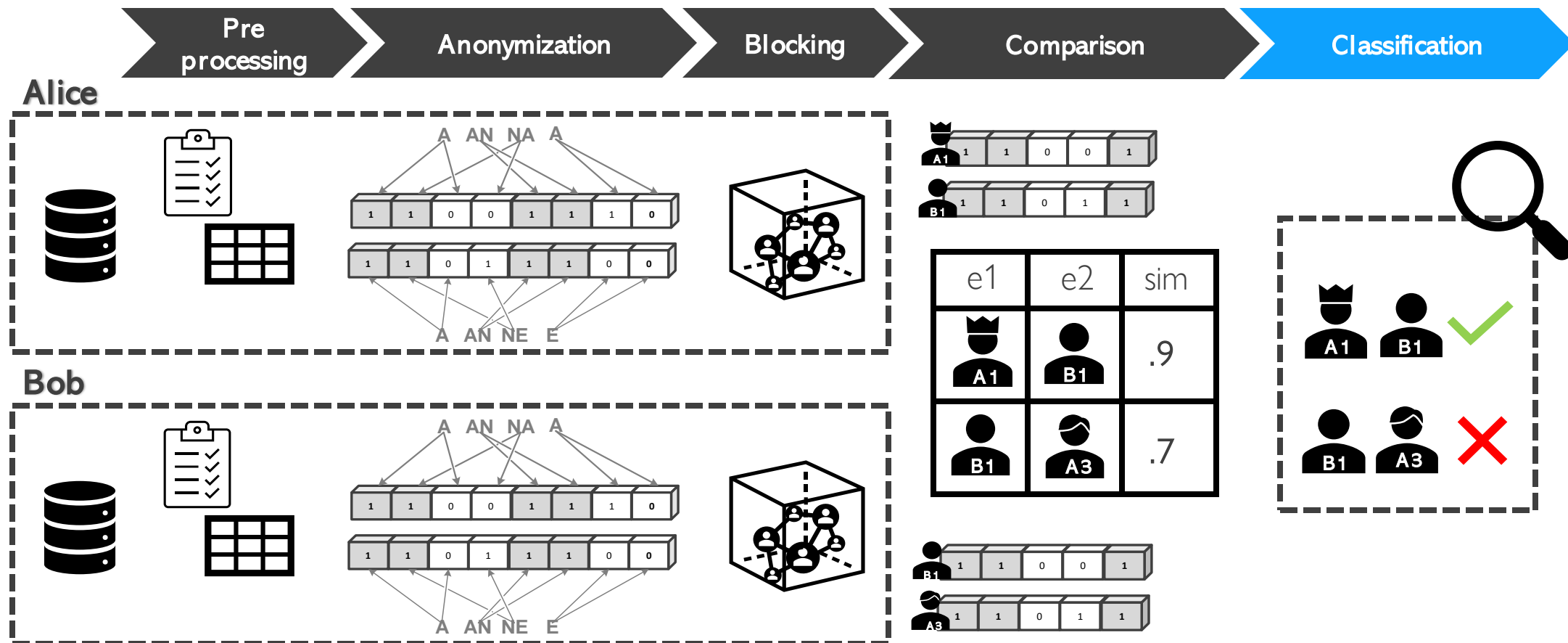


Scenario	Accuracy
Original Attack Result	90%
AT-UC Scenario	0%

# Deep Learning- based Classifiers for PPRL

- I. Introduction
- II. PPRL Comparison step Auditability
- III. Unsupervised Classification step for PPRL
- IV. Deep Learning-based Classifiers for PPRL
- V. Final Arguments

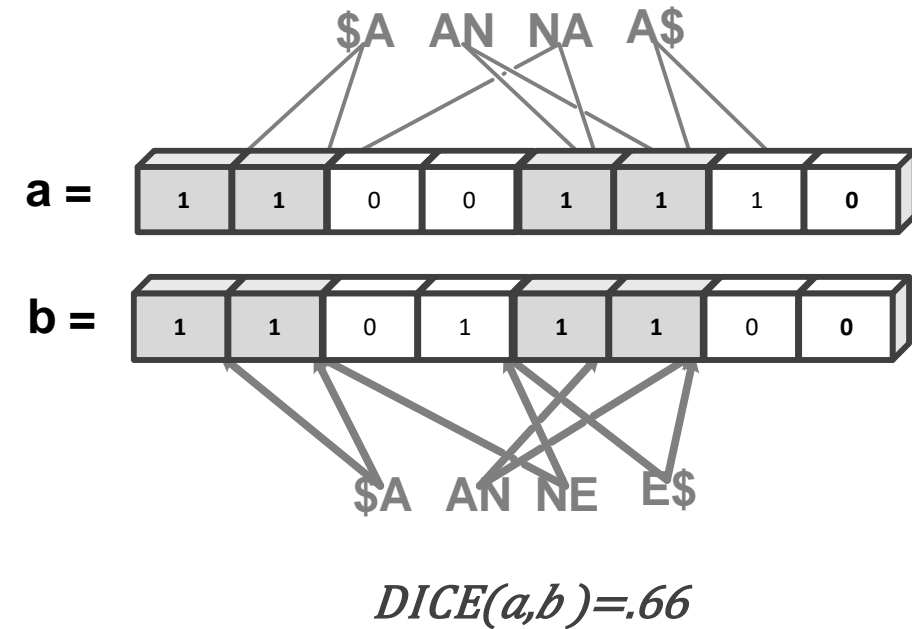
# Limitations of PPRL Classification step



# Limitations of PPRL Classification step

Similarity measures bias

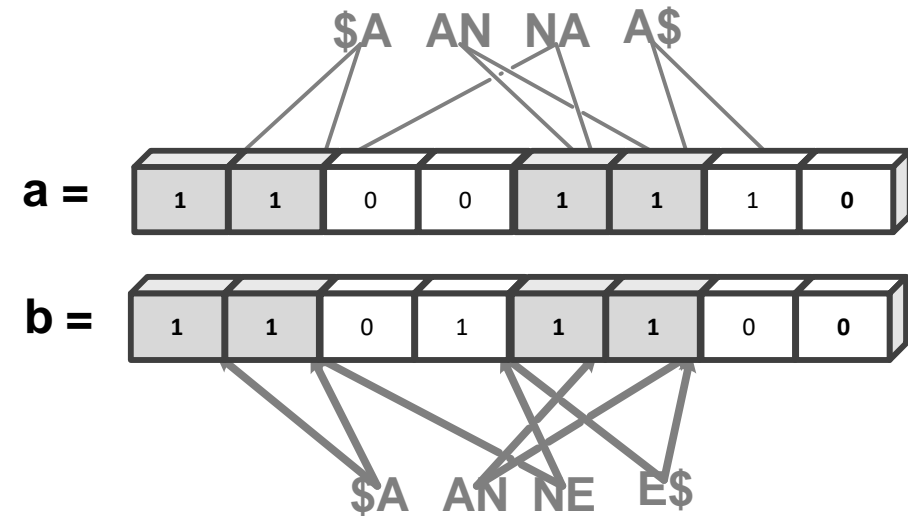
- Encoding Limitation
- Similarity measures limitation



# DLC

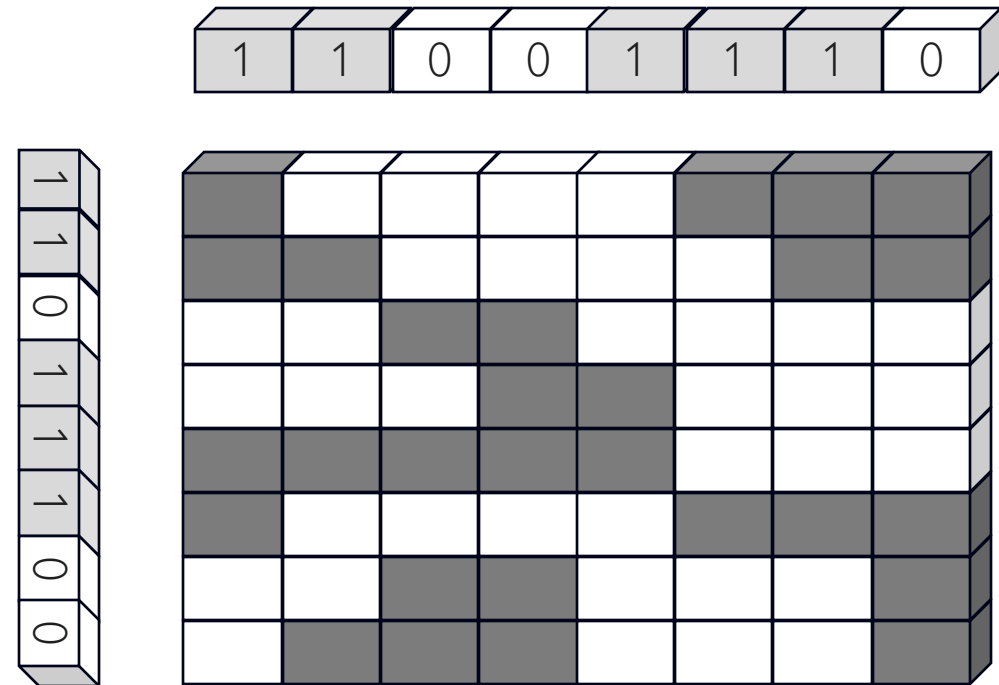
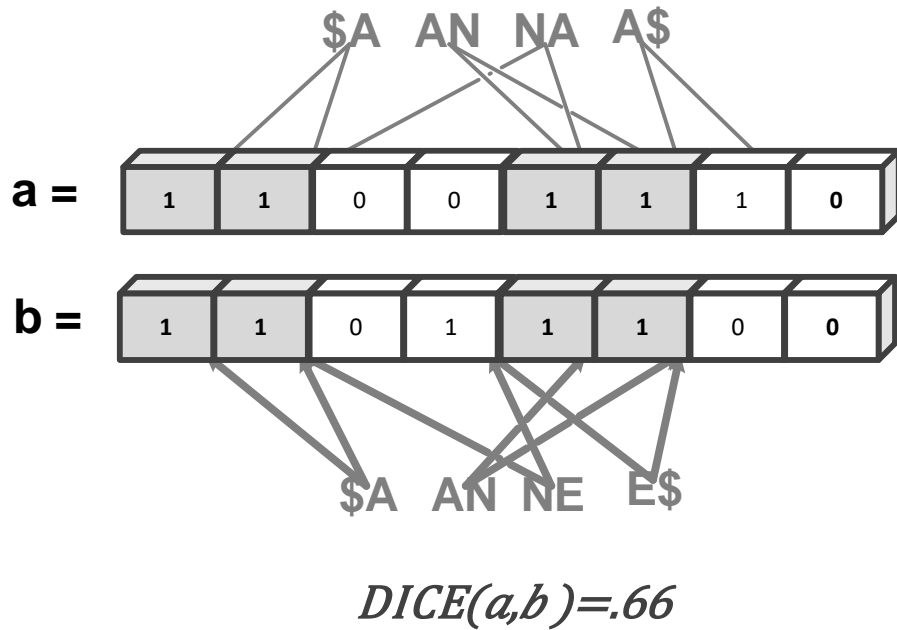
Deep Learning-based Classifier for PPRL

How to extract the BF patterns?



$$DICE(a,b)=.66$$

How to extract the BF patterns?



CRP highlights common states (e.g., bits) over encoded record pairs

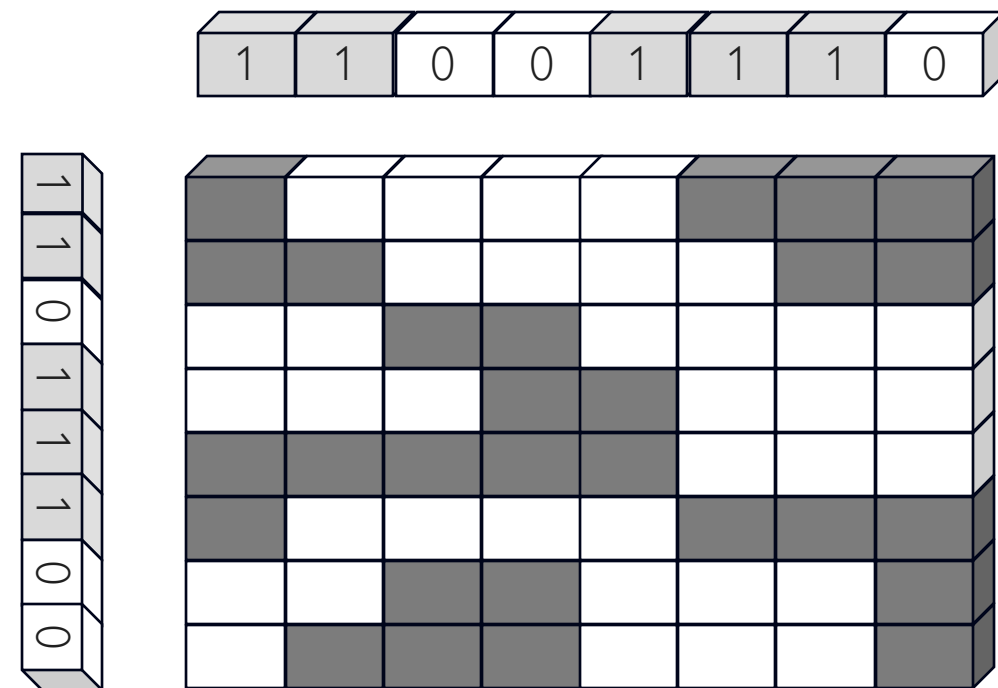
### ■ CRP Parameters

- i. Encoded Record ( $\hat{e}_1, \hat{e}_2$ )
  - $l$  = number of bits  $\hat{e}_1$  and  $\hat{e}_2$
- ii. Neighbors ( $m$ )
- iii. Heaviside Threshold ( $\alpha$ )

■  $CRP(\hat{e}_1, \hat{e}_2, m, \alpha) = RP_{n \times n}$ , such as  $n = l - (m - 1)$

$$CRP(\hat{e}_1, \hat{e}_2, m, \alpha) = \sum_{i=0}^l \sum_{j=0}^l \Theta \left( \alpha, \sum_{w=0}^m ||\hat{e}_1[i + w] - \hat{e}_2[j + w]|| \right)$$

$$\Theta(\alpha_i, v) = \begin{cases} 1 : & v \leq \alpha_i \\ 0 : & v > \alpha_i \end{cases}$$

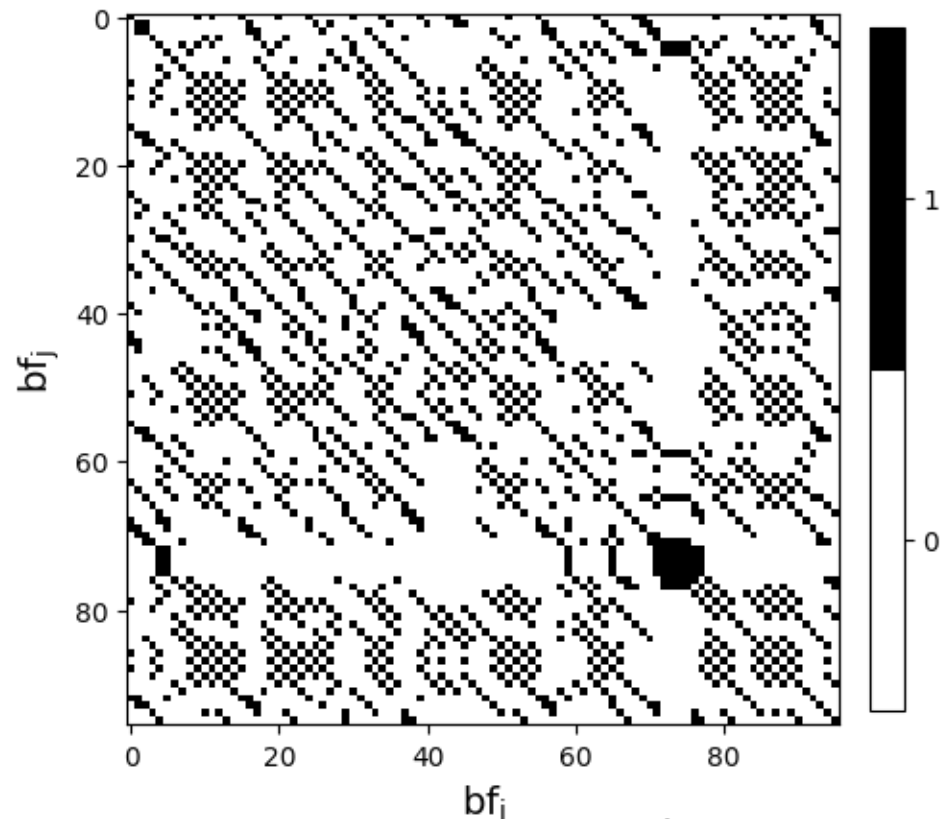


$l = 8, m = 3, \alpha = 1$

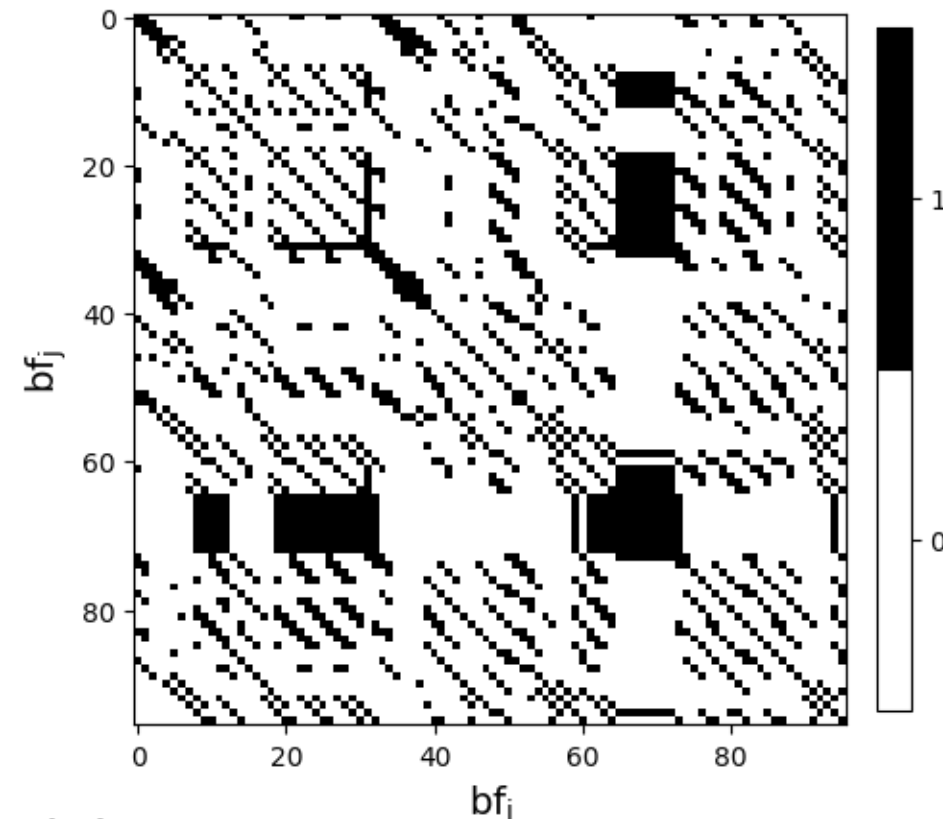


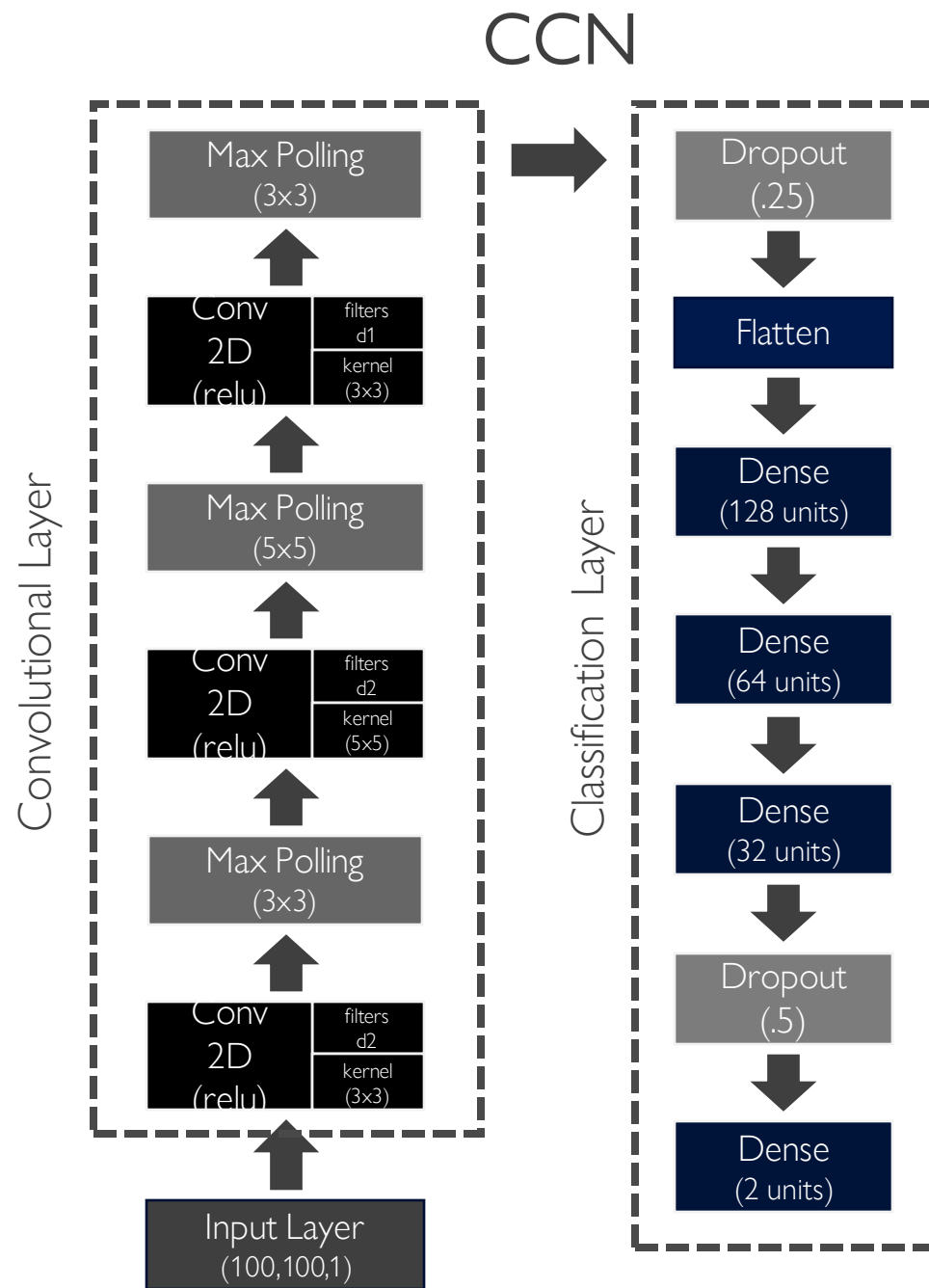
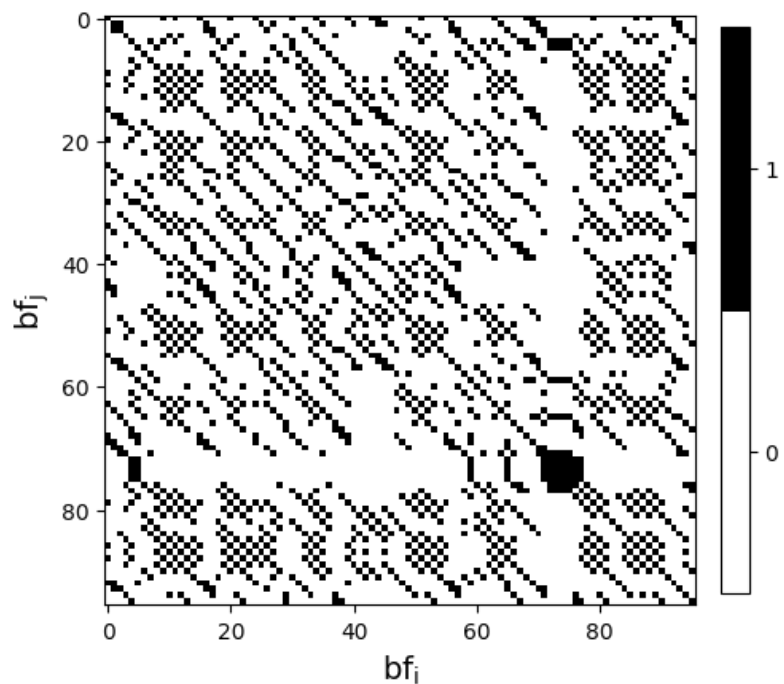


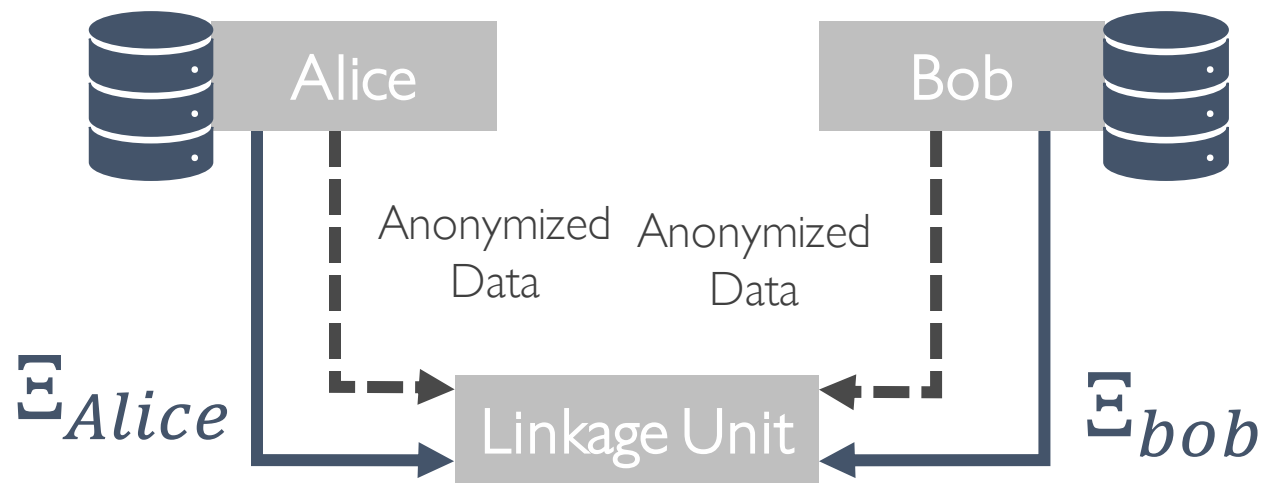
## Matching BF pair



## Non-Matching BF Pair







# EVALUATION

DLC Linkage Quality and Privacy

# Evaluation Metrics

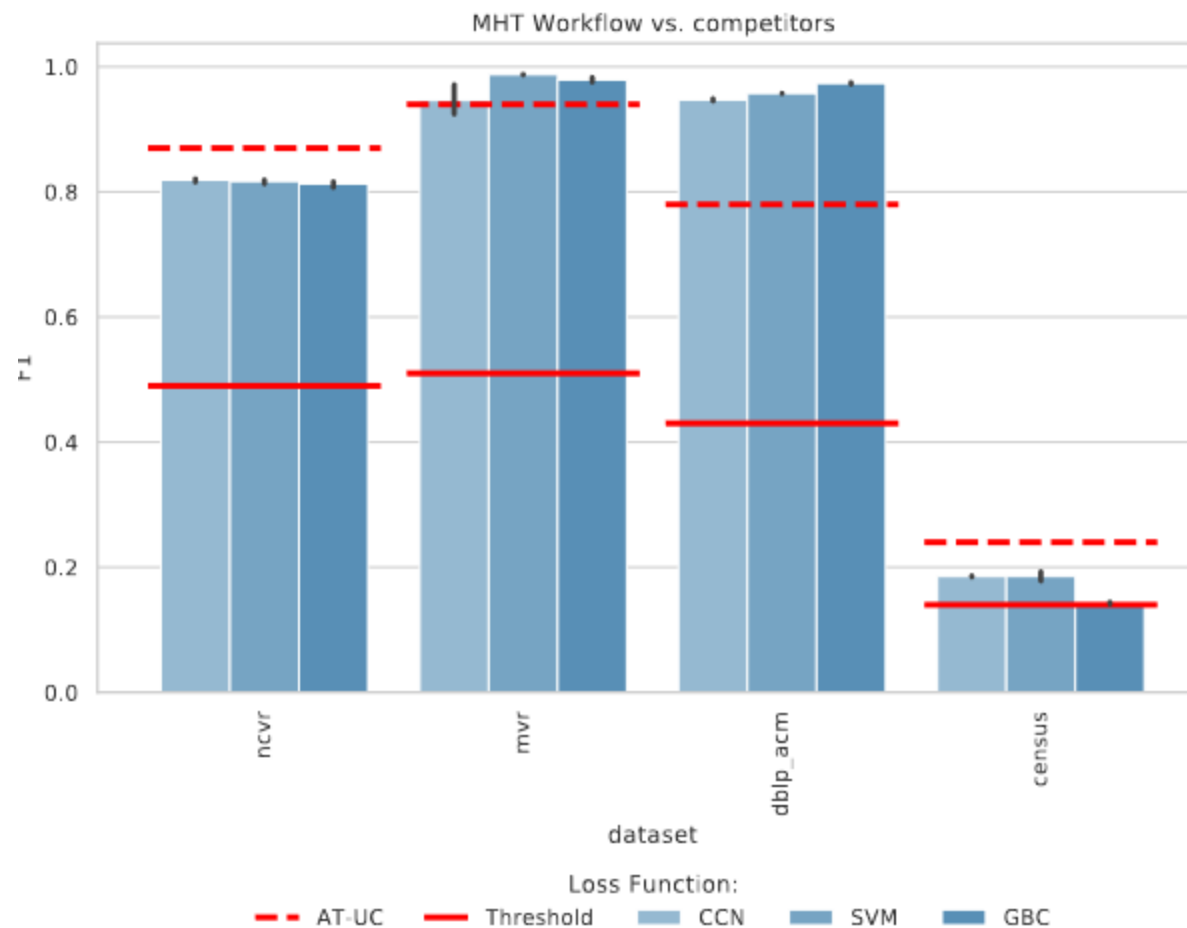



## Linkage Quality

- ☐ DLC Quality Evaluation
  - ☐ Is CRP able to improve the classifiers effectiveness?
    - ☐ F1
  - ☐ What is the impact of using different classifiers in the MHT Workflow?
    - ☐ ROC Curve

# Linkage Quality

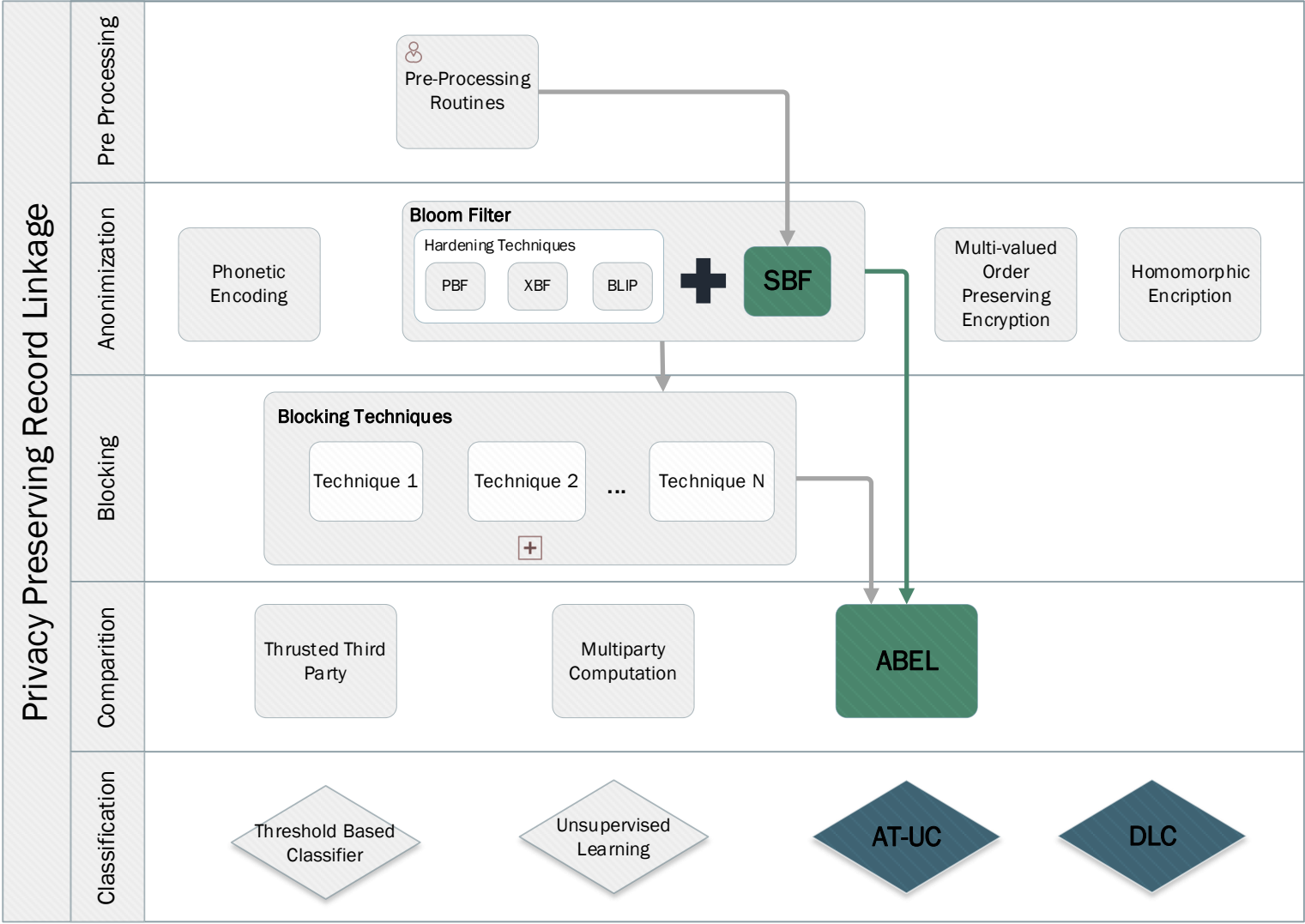
Is DLC able to improve the PPRL quality results compared to the baseline and the competitor?



- 
- I. Introduction
  - II. PPRL Comparison step Auditability
  - III. Unsupervised Classification step for PPRL
  - IV. Deep Learning-based Classifiers for PPRL
  - V. **Final Arguments**

# Final Arguments

# Contributions





# Code and Datasets

Available on author's github (<https://github.com/thiagonobrega>) :

- i. Source Code
- ii. Instructions\*
- iii. Datasets

\* Reproducibility details

# Research Relevance



## Privacy Preserving Applications

- Medical
- Nacional Security
- Public Policies
- Novel Privacy Attacks



## Other context

- Federated Linkage
- Census
- Identity Management
- Law and Regulations

# Publications

- i. Blockchain-based privacy-preserving record linkage: enhancing data privacy in an untrusted environment - T Nóbrega, CES Pires, DC Nascimento. Information Systems 102, 101826 - 2021
- ii. Limitation of Blockchain-based Privacy-Preserving Record Linkage - T Nóbrega, CES Pires, DC Nascimento. Information Systems 108, 101935 - 2022
- iii. Towards Auditable and Intelligent Privacy-Preserving Record Linkage - T Nóbrega, CES Pires, DC Nascimento. Anais Estendidos do XXXVI Simpósio Brasileiro de Bancos de Dados, 99-105S – 2020
- iv. Towards automatic Privacy-Preserving Record Linkage: A Transfer Learning based classification step - T Nóbrega, CES Pires, DC Nascimento. Data & Knowledge Engineering 145 (2023): 102180.

\* The publication of DLC is currently being written for submission.



# Thank you

Thiago Nóbrega

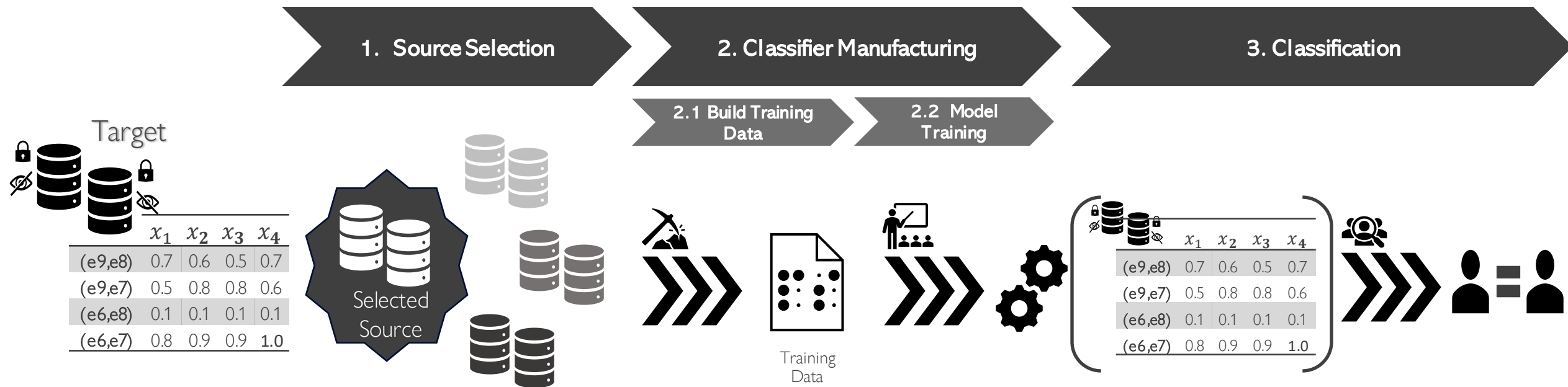


# Thank you

Thiago Nóbrega

# AT-UC Overview

Three stage approach



# Future Work

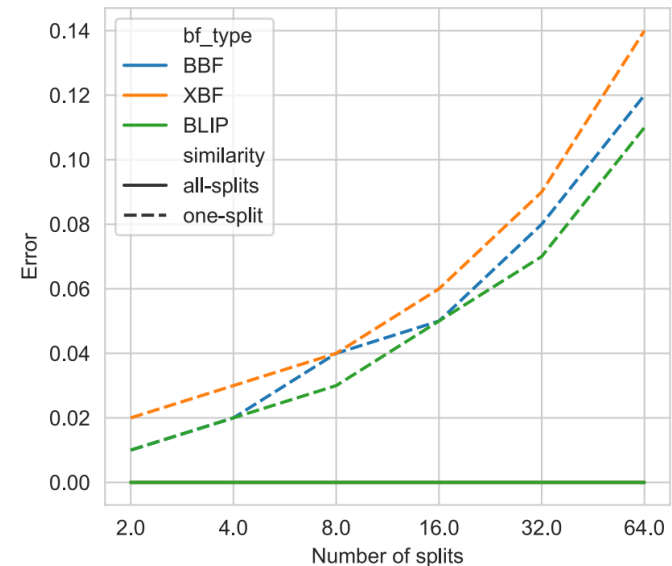
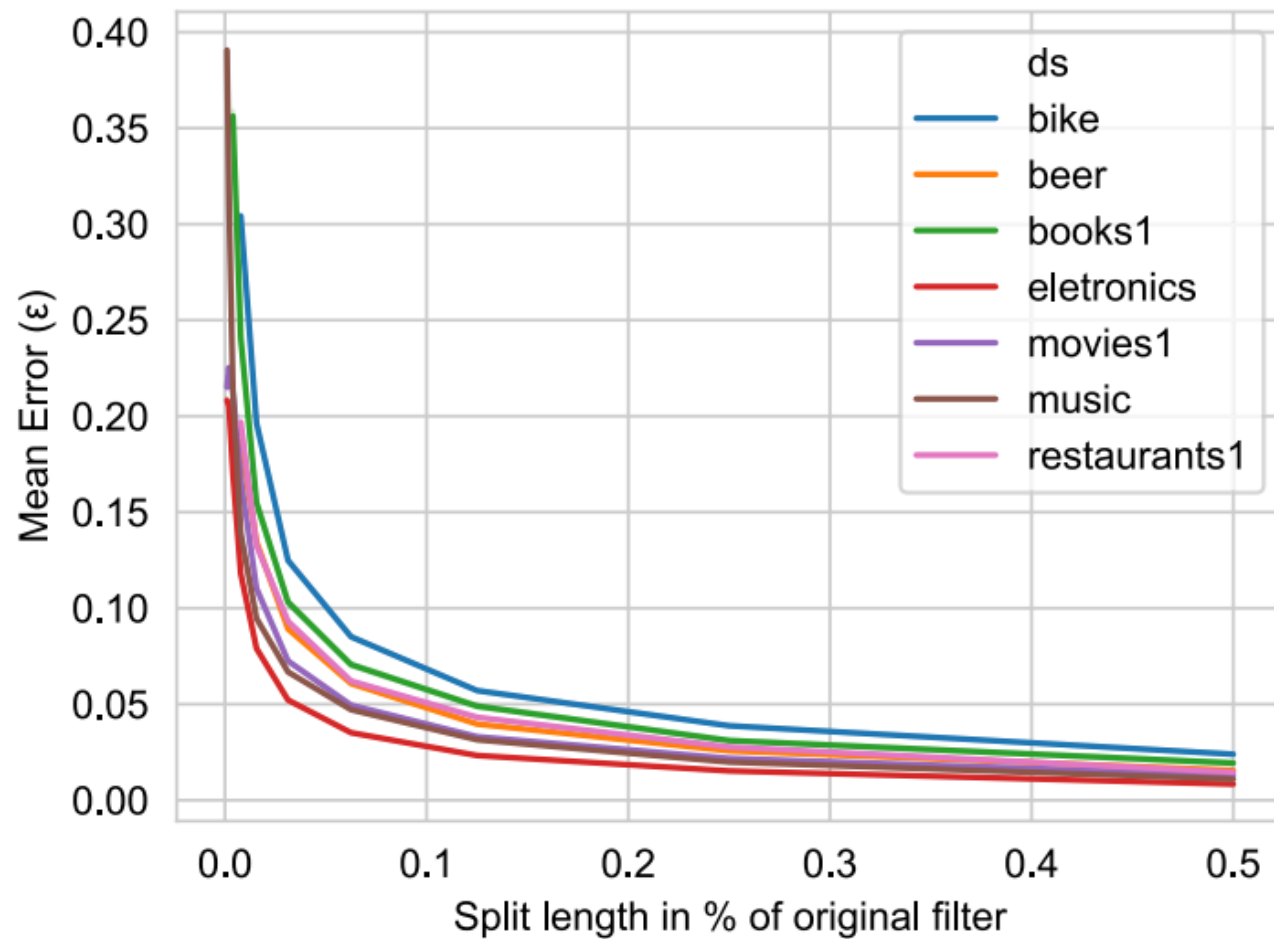
- Privacy aspects
  - Improve the privacy guarantees of the PPRL process (Privacy-Preserving Blockchain)
  - Novel Privacy Attacks
  - Differential Privacy in PPRL
- Linkage Quality and Novel PPRL Applications
  - Distributed Representation of Words (DR) in PPRL
  - Federated Learning (collaborative learning)
  - Deep Unsupervised Domain Adaptation

# Research goals

**“This work’s main goal covers improving privacy and the linkage quality of PPRL process”**

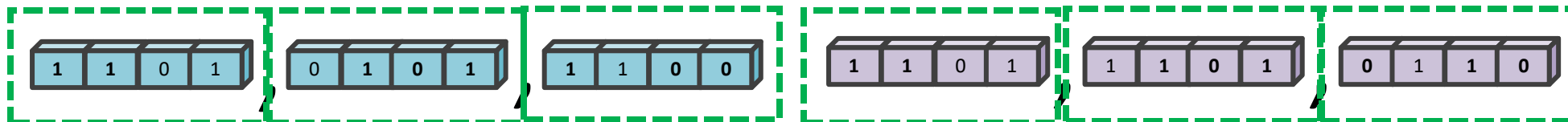


# SBF ERROR



$$B\left(\frac{l}{s}, p\right) = \binom{\frac{l}{s}}{x} p^x (1-p)^{\frac{l}{s}-x}$$

# SBF Goal



*threshold* = .65 |  $\epsilon \cong 0.35$

$$Jaccard(\begin{bmatrix} 1 & 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 0 & 1 \end{bmatrix}) = \frac{3}{3} \rightarrow 1 \quad | \epsilon \cong 0.35$$

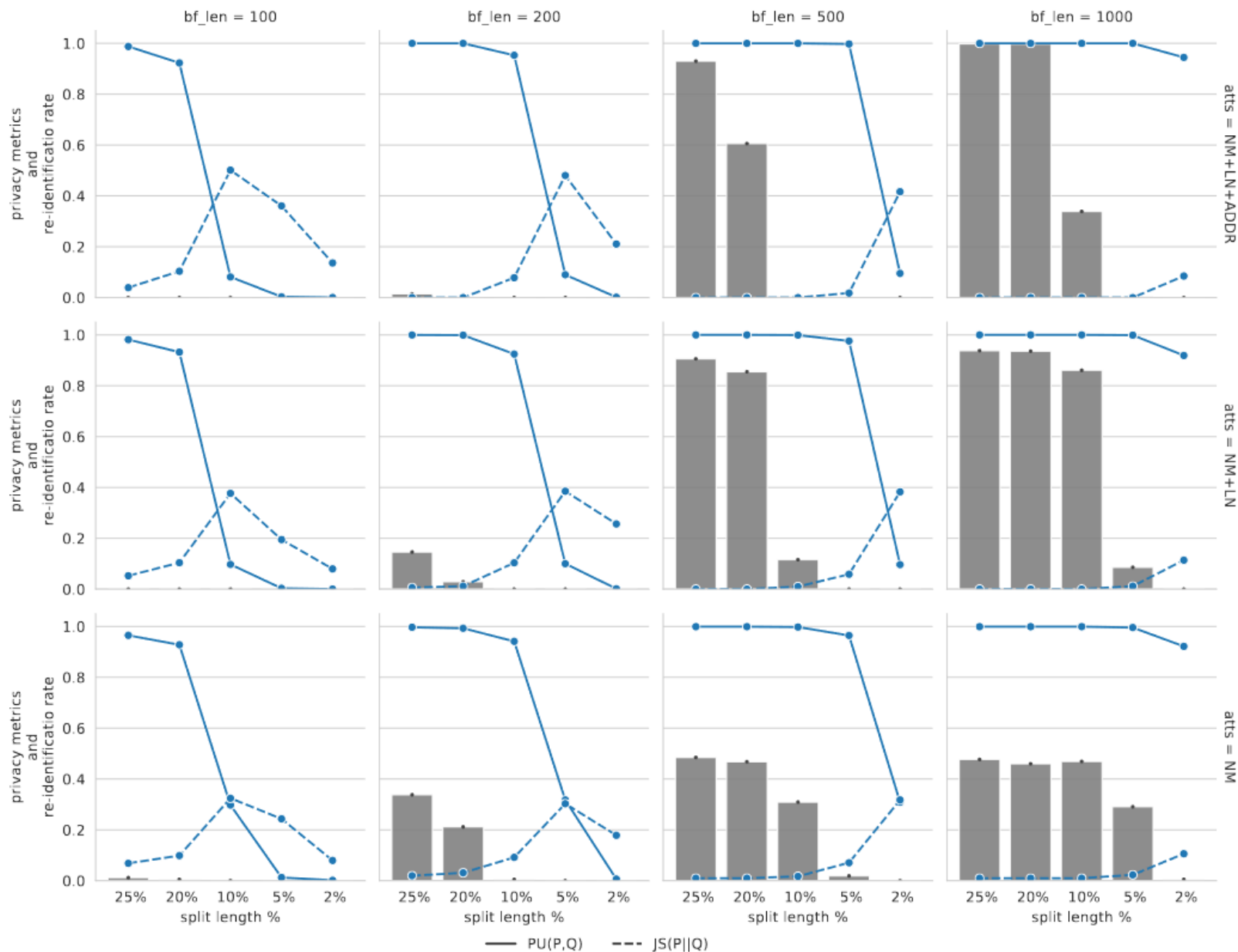
$$Jaccard(\begin{bmatrix} 0 & 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 0 & 1 \end{bmatrix}) = \frac{2}{3} \rightarrow 0.65 | \epsilon \cong 0.0$$

$$Jaccard(\begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix}) = \frac{1}{3} \rightarrow 0.3 \quad | \epsilon \cong 0.35$$

# Privacy

## ABEL Parametrization

- Indistinguishability
  - $PU(P, Q) < 0.5$
- Uncertain
  - $KL(P, Q) > 0.1$

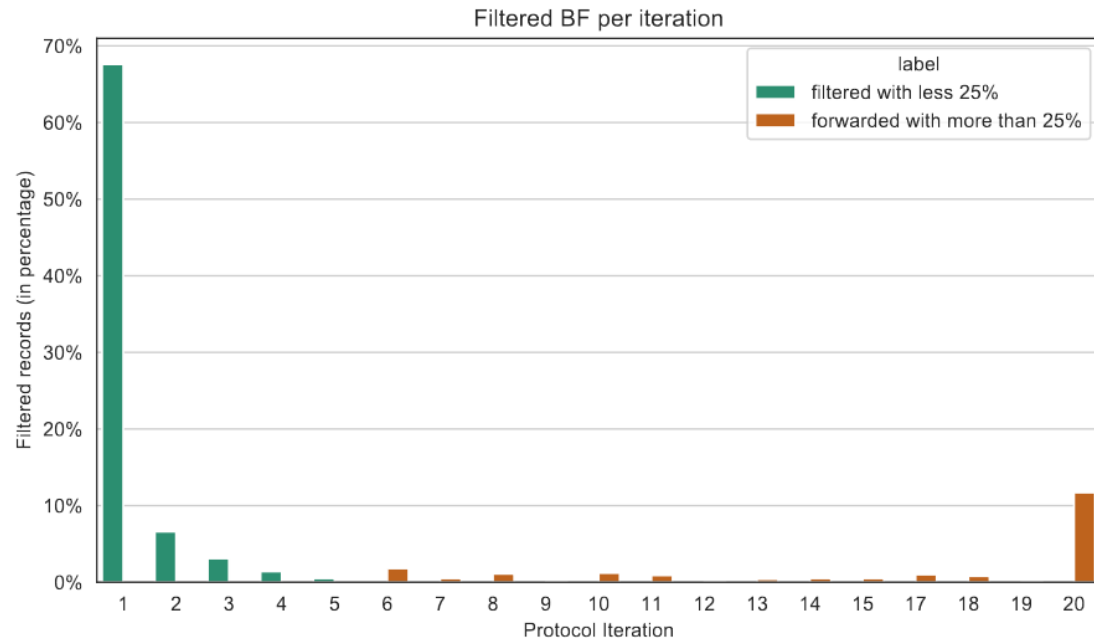


# Privacy

Attack Effectiveness

Test parameters

- Dataset : NCVR
- Bloom Filter
  - $l = 200$  bits
  - $f.p.r = 50\%$
- ABEL
  - Privacy parameters
    - Indistinguishability
      - $PU(P, Q) < 0.5$
    - Uncertain
      - $KL(P, Q) > 0.1$
  - $a = .85$ ,  $error = .05$
  - $s = 10$  bits (5%)



iteration	shared information	1-to-1 correct	1-to-many correct	1-to-1 wrong	1-to-many wrong	No matches
Iteration 1	5%	1	0	5,588	4,411	0
Iteration 2	10%	1	1	1,855	1,392	0
Iteration 3	15%	8	17	1,157	975	0
Iteration 4	20%	48	137	804	759	0
Iteration 5	25%	214	312	614	450	0
Iteration 6-20	100%	1,346	125	0	0	0

# Research questions

- I. Is it possible to improve the privacy-preserving capabilities of the Bloom Filter anonymization technique?
- II. Is it possible to consider a novel adversary model that reduces the need of trust by PPRL parties?

# Research Questions

- I. Is it possible to improve the privacy-preserving capabilities of the Bloom Filter anonymization technique?
- II. Is it possible to consider a novel adversary model that reduces the need of thrust by PPRL parties?

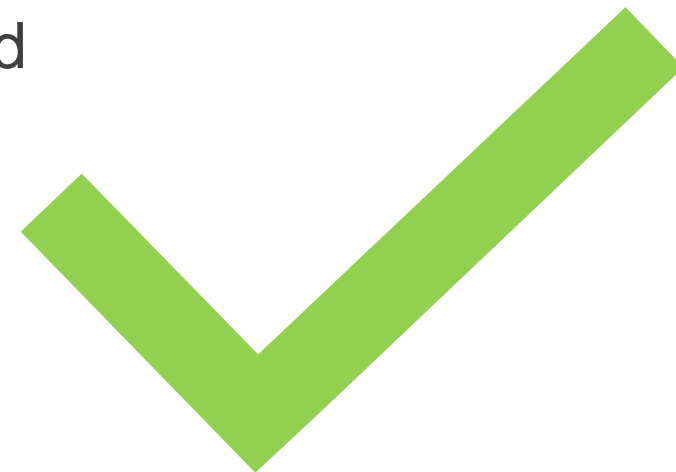


# Research questions

III. Is it possible to employ an automatic (e.g., ML-based classifier) to during the PPRL Classification step?

# Research Questions

III. Is it possible to employ an automatic (e.g., ML-based classifier) to during the PPRL Classification step?





# Research questions

IV. Is it possible to perform classification without standard binary similarity metrics (e.g., Jaccard and dice distance) ?

# Research Questions

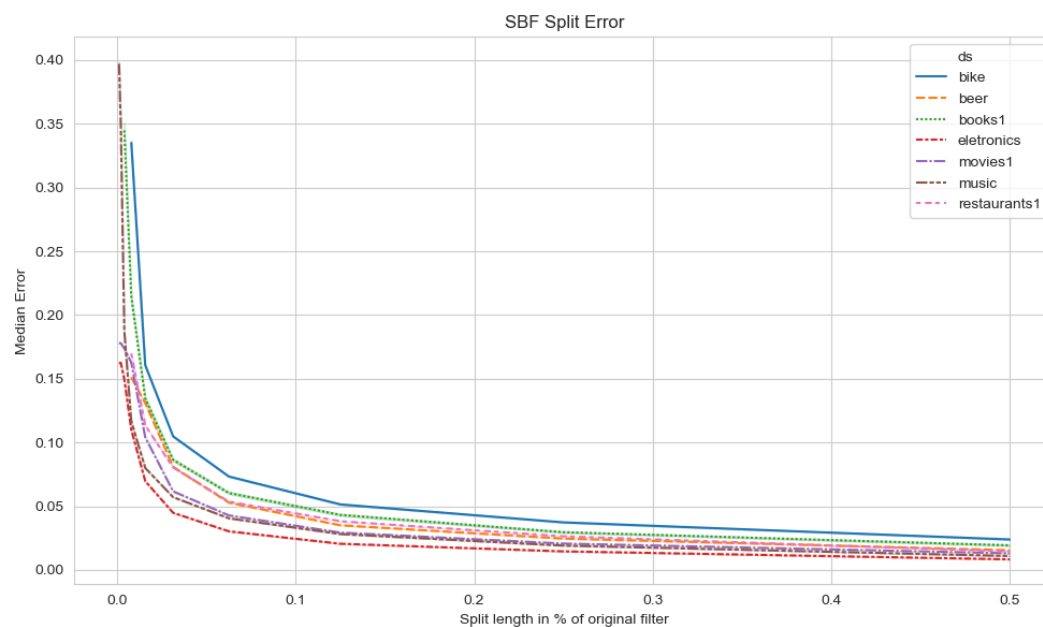
IV. It is possible to perform classification without standard binary similarity metrics (e.g., Jaccard and dice distance) ?



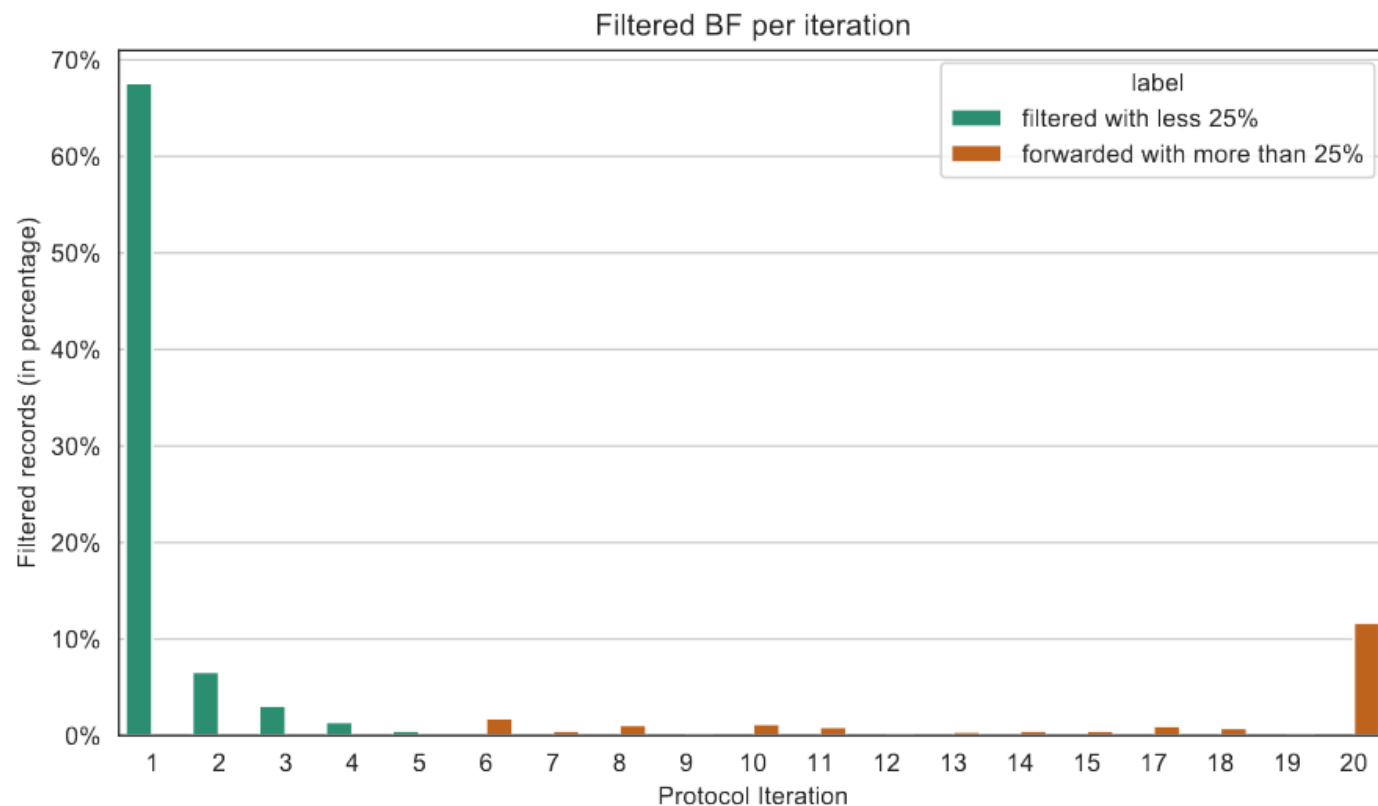
$$Jaccard(\text{ANA}, \text{ANE}) = Jaccard(\text{ANA\_subset}, \text{ANE\_subset}) + \text{error}$$

ANA: 1 1 0 1 0 1 0 1 1 1 0 0  
ANE: 1 1 0 1 1 1 0 1 0 1 1 0

ANA\_subset: 1 1 0 1  
ANE\_subset: 1 1 0 1



$$\text{error} = \binom{\frac{l}{s}}{x} p^x (1-p)^{\frac{l}{s}-x}$$



- Split Filtering
- Indistinguishability
- Uncertainty

# Linkage Quality

What is the impact of using different classifiers - e.g., state-of-the-art NN (ResNet50) and classical Machine Learning classifiers (SVM and GBC) - in the MHT Workflow?

dataset	model	precision	recall	f1
census	CCN	87.23% $\pm 0.0$	10.38% $\pm 0.08$	18.55% $\pm 0.13$
	GBC	69.86% $\pm 1.5$	7.94% $\pm 0.13$	14.26% $\pm 0.25$
	SVM	95.74% $\pm 6.02$	10.27% $\pm 0.58$	18.55% $\pm 1.06$
dblp_acm	CCN	98.95% $\pm 0.34$	90.82% $\pm 0.19$	94.71% $\pm 0.26$
	GBC	97.12% $\pm 0.65$	97.6% $\pm 1.07$	97.35% $\pm 0.21$
	SVM	93.4% $\pm 0.04$	98.21% $\pm 0.04$	95.74% $\pm 0.04$
mvr	CCN	99.12% $\pm 1.04$	90.63% $\pm 3.87$	94.66% $\pm 2.36$
	GBC	100.0% $\pm 0.0$	95.91% $\pm 1.06$	97.91% $\pm 0.55$
	SVM	100.0% $\pm 0.0$	97.52% $\pm 0.14$	98.75% $\pm 0.07$
ncvr	CCN	69.47% $\pm 0.52$	99.76% $\pm 0.22$	81.9% $\pm 0.29$
	GBC	68.52% $\pm 0.76$	99.9% $\pm 0.09$	81.28% $\pm 0.51$
	SVM	69.09% $\pm 0.55$	99.9% $\pm 0.08$	81.69% $\pm 0.36$

Brazilian Symposium on Databases 2023 – Belo Horizonte-MG -