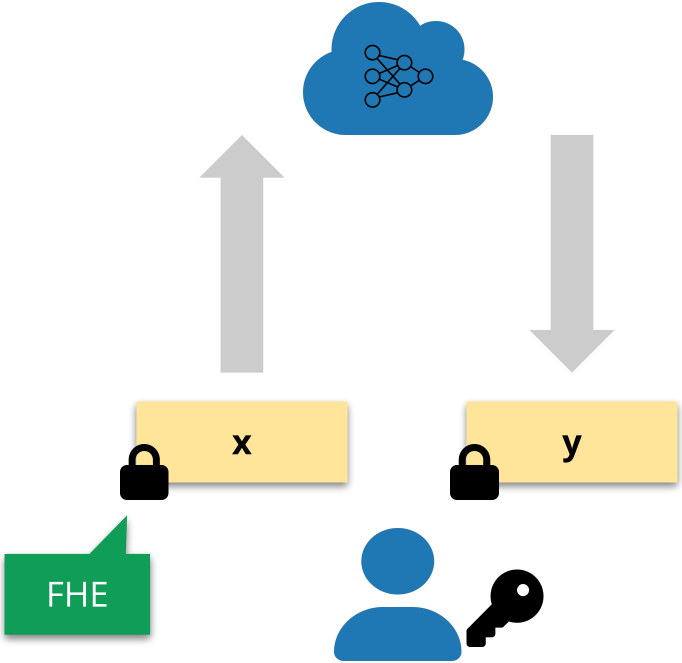
Private and Reliable Neural Network Inference

Nikola Jovanović Marc Fischer Samuel Steffen Martin Vechev

ETH Zurich, Switzerland

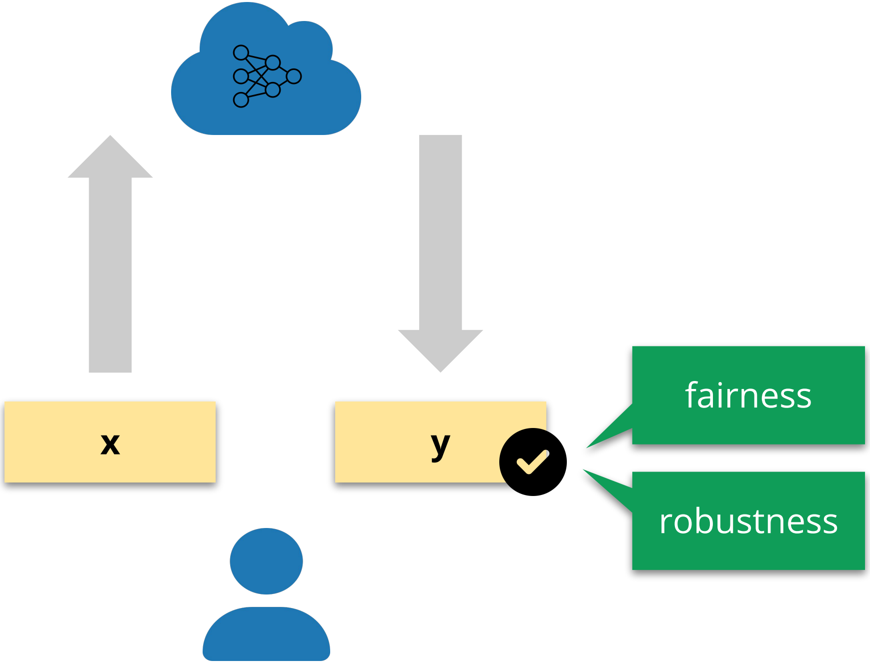
{nikola.jovanovic, marc.fischer, samuel.steffen, martin.vechev}@inf.ethz.ch

Motivation: ML as a Service



FHE

privacy-preserving
inference

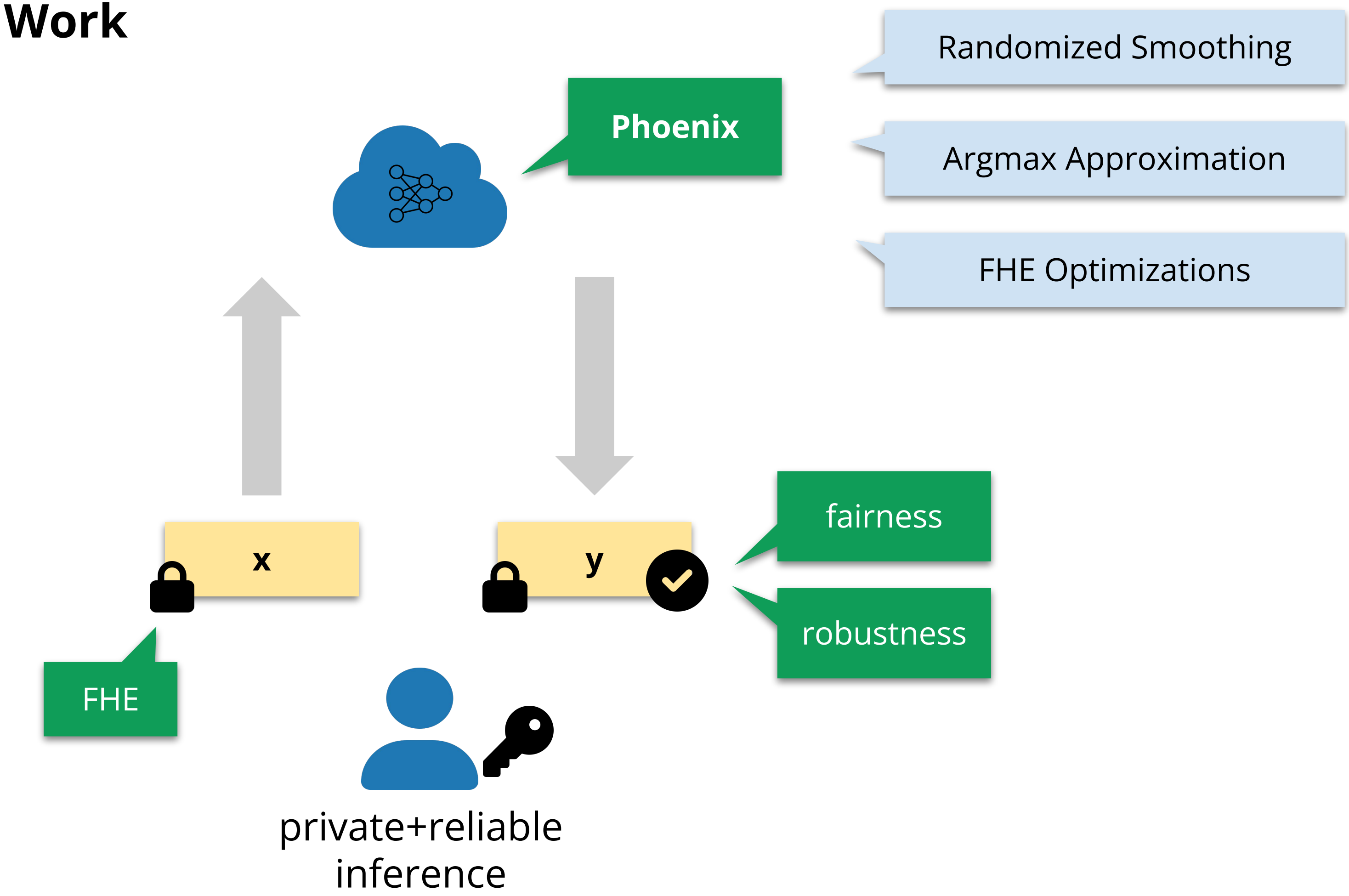


fairness

robustness

inference with
reliability guarantees

This Work



Phoenix

Randomized Smoothing

Argmax Approximation

FHE Optimizations

x

y

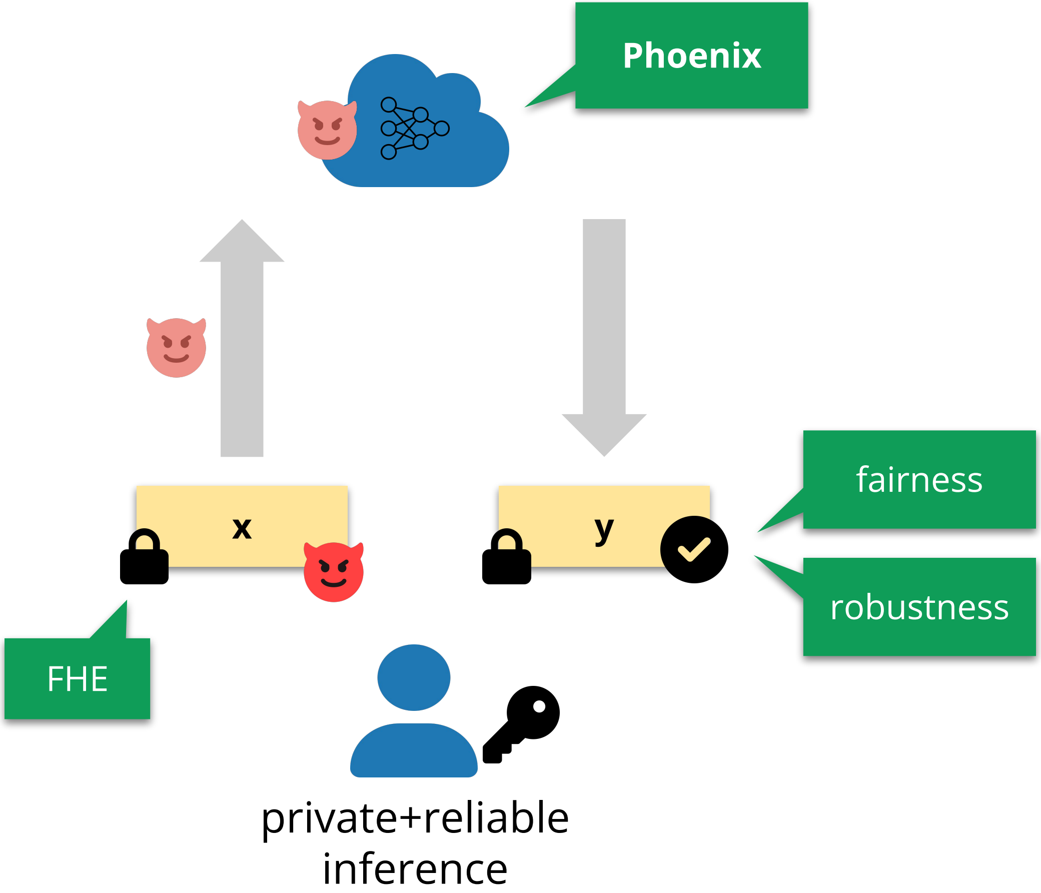
fairness

robustness

FHE

private+reliable inference

Guarantees of Phoenix



Phoenix

fairness

robustness

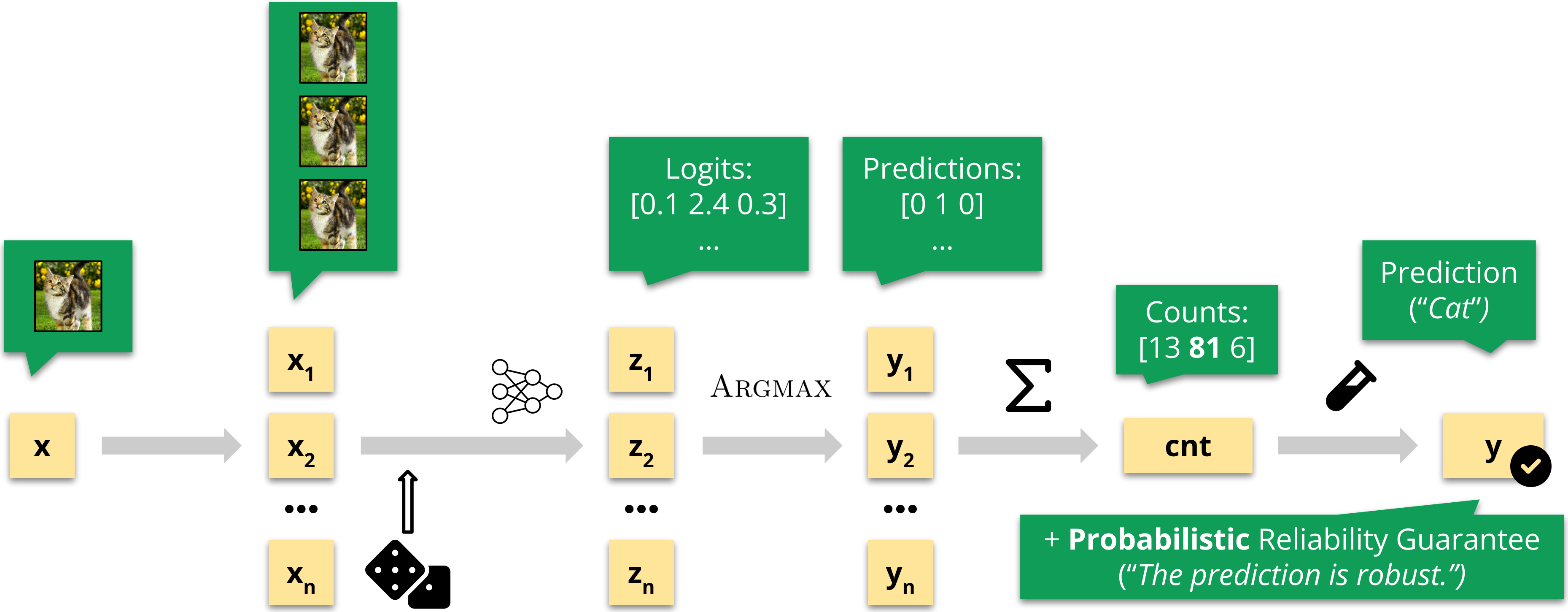
FHE

private+reliable inference

Client Data Privacy 🔒

Reliability Guarantees ✓

Background: Randomized Smoothing



x

x_1

x_2

…

x_n

Logits:
[0.1 2.4 0.3]
…

z_1

z_2

…

z_n

ARGMAX

Predictions:
[0 1 0]
…

y_1

y_2

…

y_n

Σ

Counts:
[13 **81** 6]

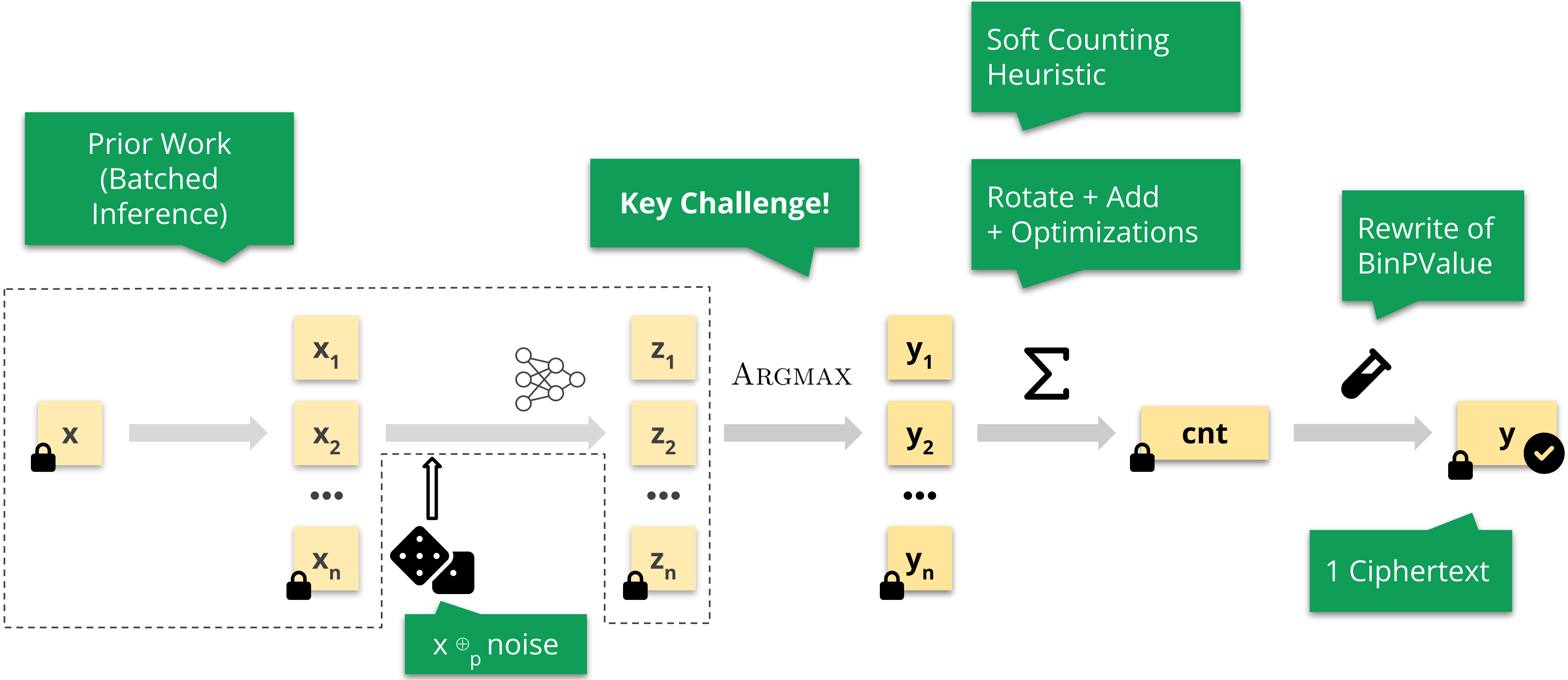
cnt

Prediction
(*“Cat”*)

y

+ **Probabilistic** Reliability Guarantee
(“The prediction is robust.”)

Overview of Phoenix



Soft Counting Heuristic

Prior Work (Batched Inference)

Key Challenge!

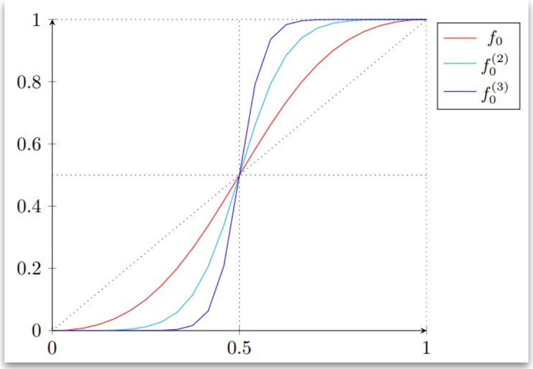
Rotate + Add + Optimizations

Rewrite of BinPValue

1 Ciphertext

$x \oplus_p$ noise

Argmax Approximation

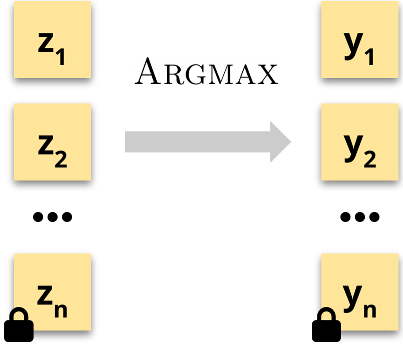


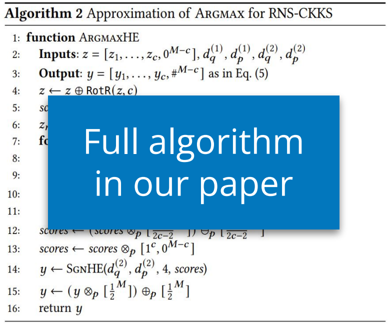
Cheon et al. (ASIACRYPT ‘20)

Bound P(violation)

Conditions + Rescale

[a, 1] → [1-2^{-b}, 1]





Algorithm 2 Approximation of ARGMAX for RNS-CKKS

```
function ArgmaxHE
    Inputs: $z = [z_1, \dots, z_c, 0^{M-c}], d_q^{(1)}, d_p^{(1)}, d_q^{(2)}, d_p^{(2)}$
    Output: $y = [y_1, \dots, y_c, \#^{M-c}]$ as in Eq. (5)
    $z \leftarrow z \oplus \texttt{RotR}(z, c)$
    sc
    $z_r$
    fo




    scores ← (scores ⊗_p [ 2c−2 ]) ⊖_p [ 2c−2 ]
    $scores \leftarrow scores \otimes_p [1^c, 0^{M-c}]$
    $y \leftarrow \textsc{SgnHE}(d_q^{(2)}, d_p^{(2)}, 4, scores)$
    $y \leftarrow (y \otimes_p [\frac{1}{2}^M]) \oplus_p [\frac{1}{2}^M]$
    return $y$
```

Full algorithm in our paper

Implementation & Evaluation

Available on GitHub:

`eth-sri/phoenix`

Consistency

=

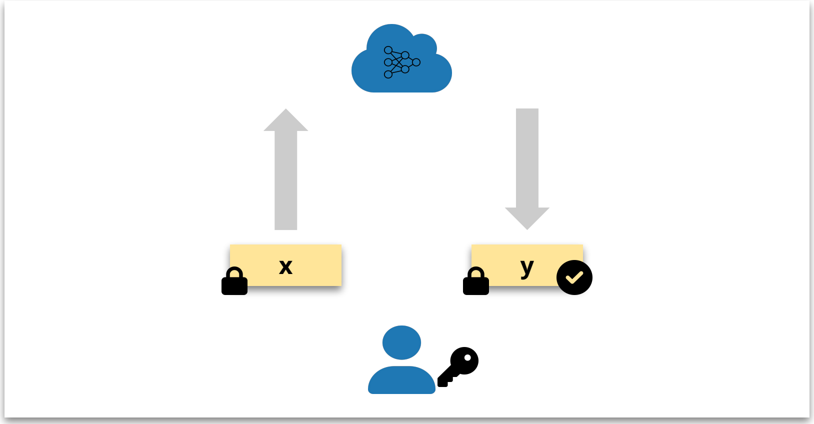
The results are equivalent to the ones obtained in non-private evaluation

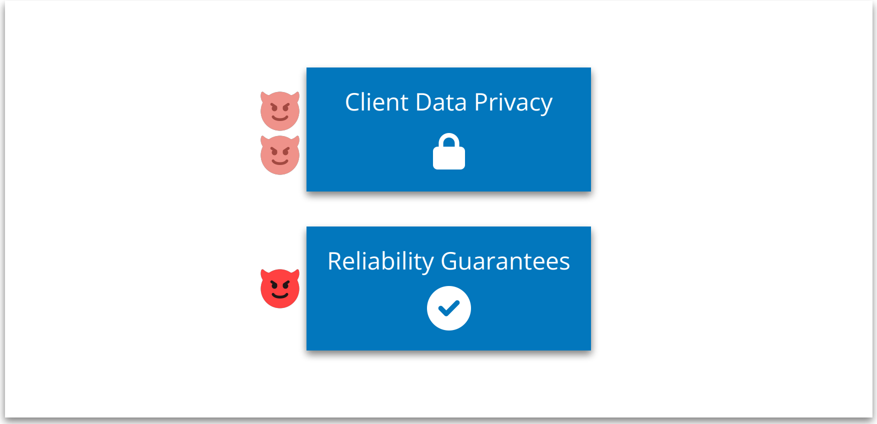
Efficiency

⚡

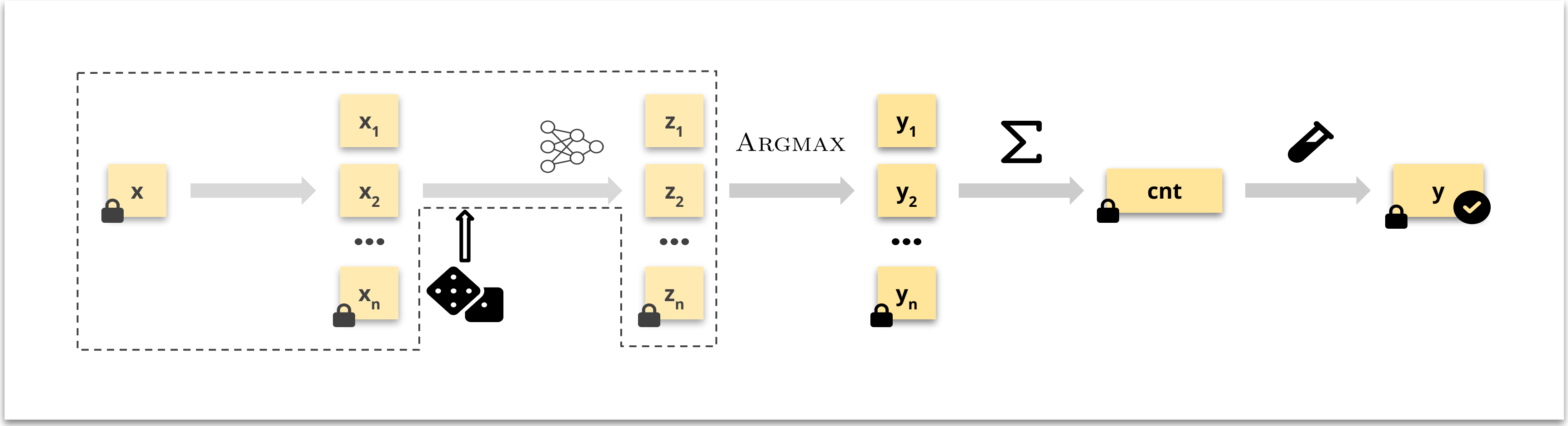
Viable latencies and communication costs

Summary: Phoenix




Client Data Privacy

Reliability Guarantees


x
x_1
x_2
...
x_n
z_1
z_2
...
z_n
ARGMAX
y_1
y_2
...
y_n
Σ
cnt
y