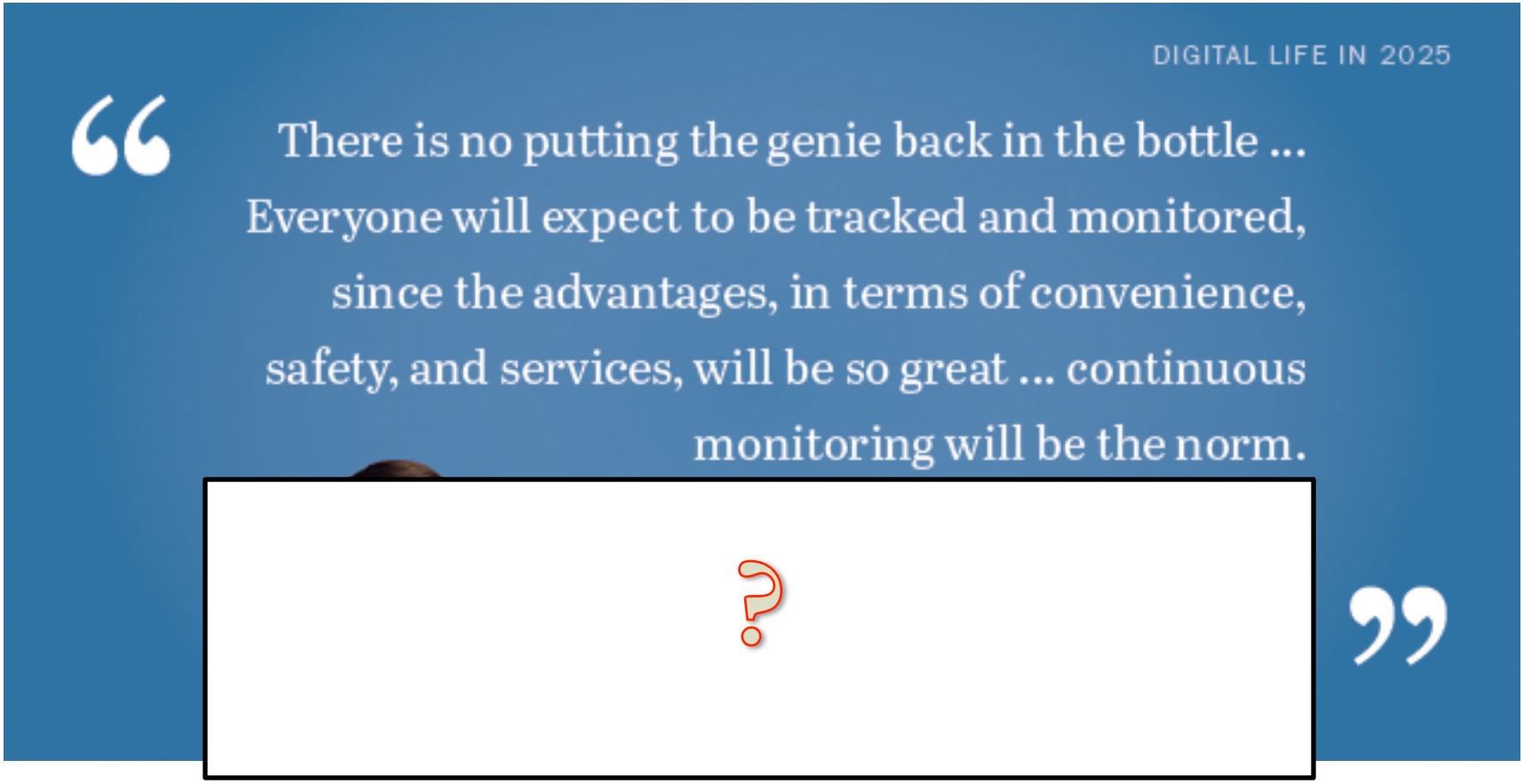
Duke
UNIVERSITY

Let's Collaborate: A Different Paradigm for Privacy

Guillermo Sapiro
Duke University

M. Bertran
N. Martinez
A. Papadaki
M. Rodrigues
Q. Qiu

Privacy Today



Privacy Today

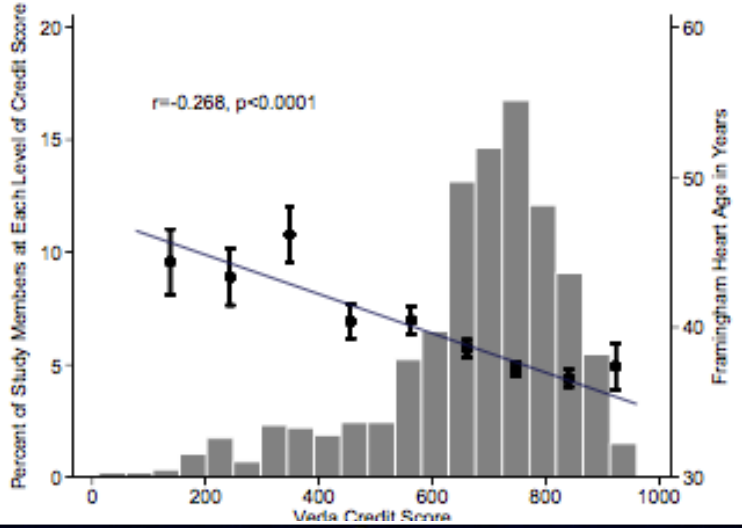
"Privacy is a fundamental human right"

Outline

- “Universal” privacy is utopia
- Proposed collaborative paradigm
- Discussion

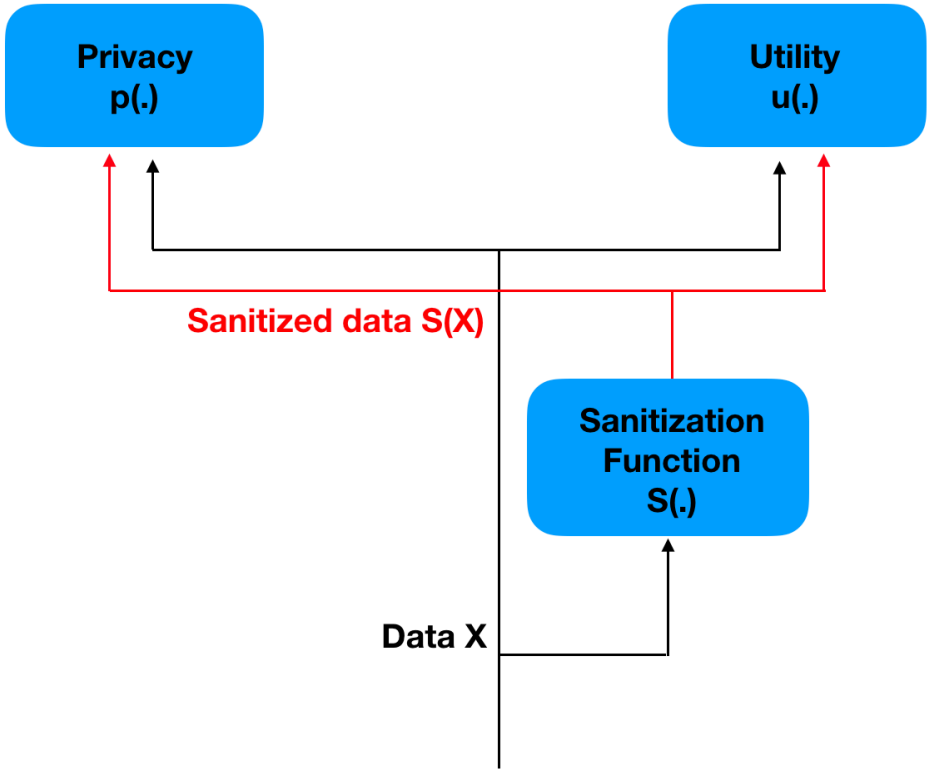
“Universal” Privacy is Utopia

- Netflix Prize
  - Use other dataset
- Unexpected correlations



- Same features for different tasks

Proposed Paradigm: Collaborative Privacy



$$\text{Loss}_S = (1-\alpha)D_{KL}(P(u \mid x) \mid\mid P(u \mid S(x))) + \alpha D_{KL}(P(p) \mid\mid P(p \mid S(x))).$$

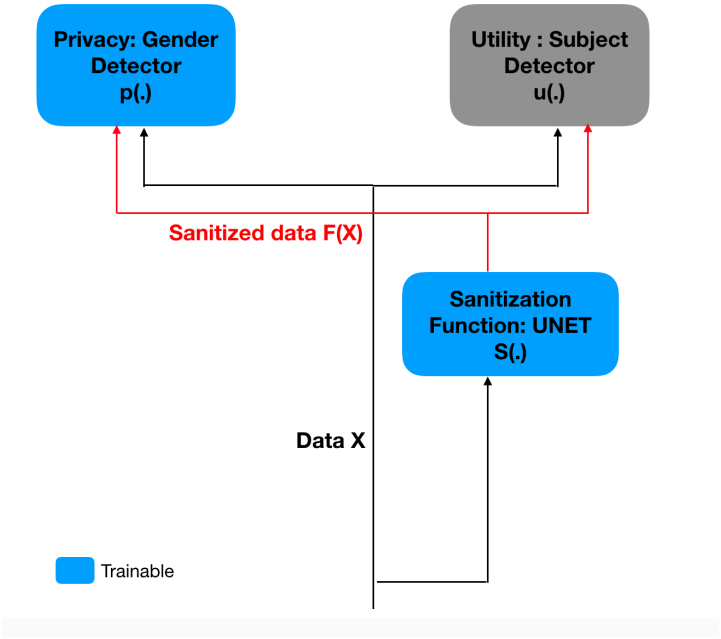
Proposed Paradigm: Collaborative Privacy

- Plug-and-play
  - Train sanitization function
  - No change in utility or privacy algorithms
- Adversarial
  - Sanitization function
    - Fool privacy- Approximate prior
    - Preserve utility
  - Privacy task
    - Defeat sanitization
    - Still perform on un-sanitized data

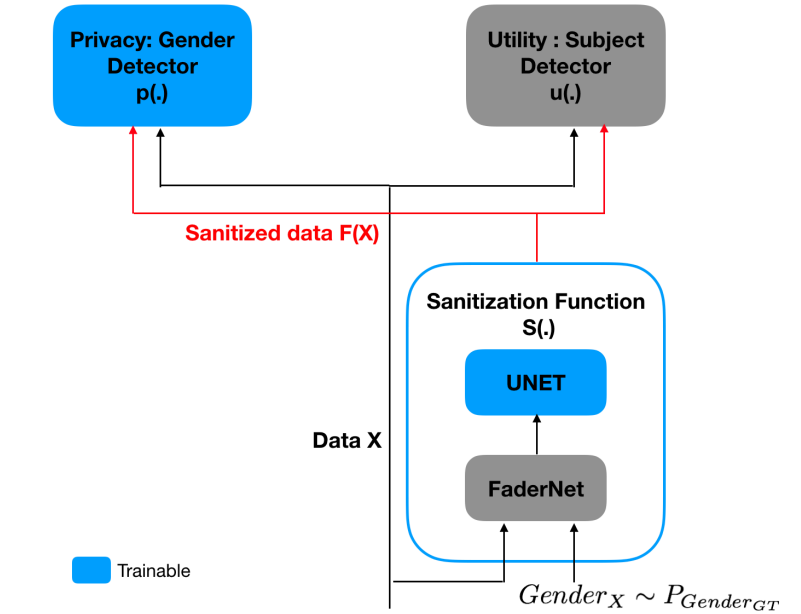
$$\text{Loss}_P = BCE(y_p(x), p(x)) + BCE(y_p(x), p(S(x)))$$

  - More universal

Privacy Learning Architectures

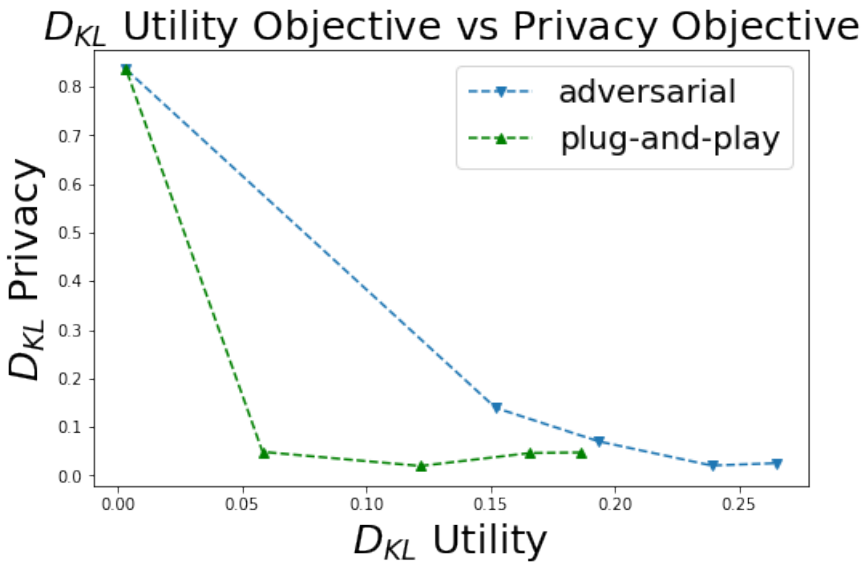


Deterministic

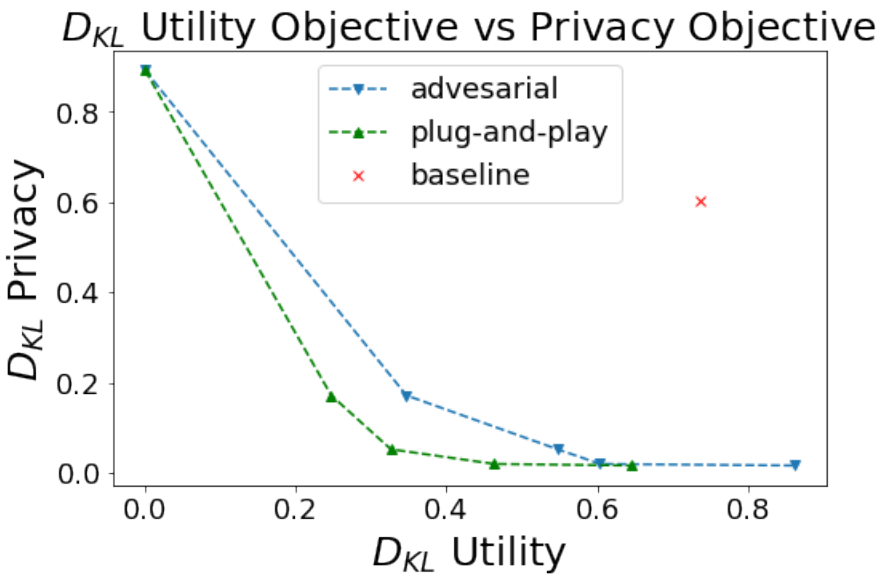


Stochastic

Experimental Results:
Utility-Privacy Tradeoff

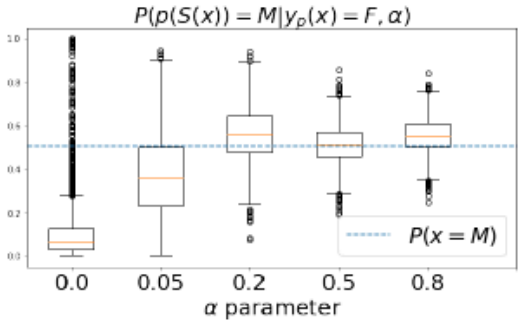


Deterministic

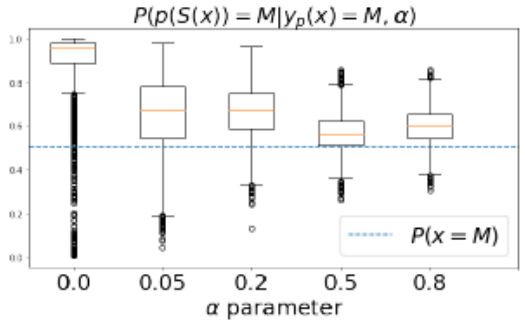


Stochastic

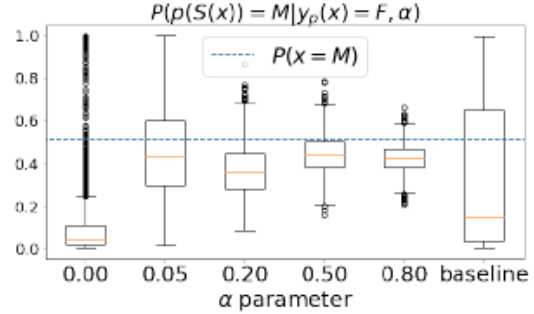
Experimental Results:
Gender Probability (Adversarial)



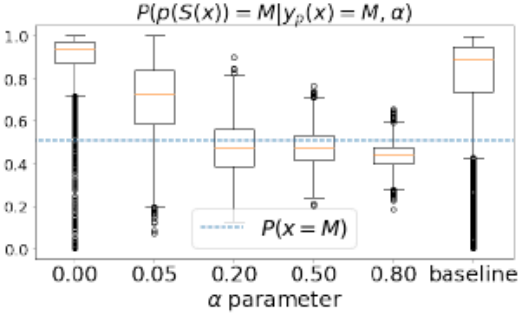
(a) $P(p(S(x)) = M|y_p(x) = F, \alpha)$, *Deterministic model*



(b) $P(p(S(x)) = M|y_p(x) = M, \alpha)$, *Deterministic model*

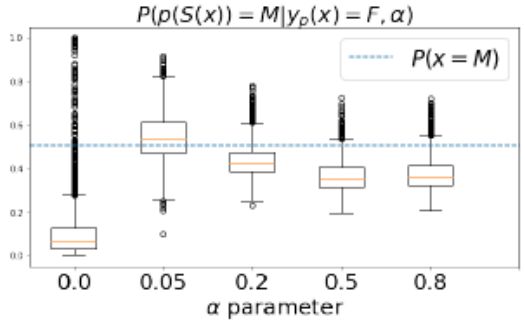


(c) $P(p(S(x)) = M|y_p(x) = F, \alpha)$, *Stochastic model*

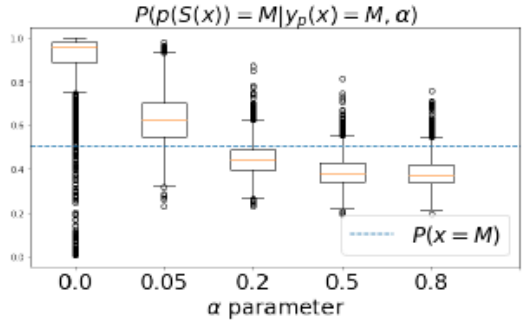


(d) $P(p(S(x)) = M|y_p(x) = M, \alpha)$, *Stochastic model*

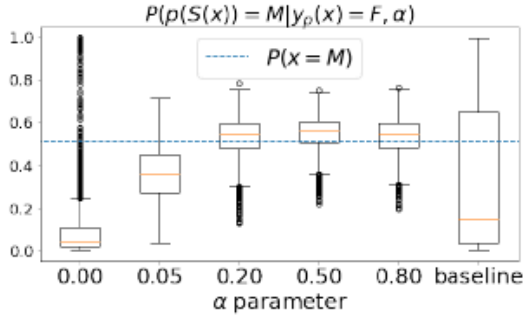
Experimental Results:
Gender Probability (Plug-and Play)



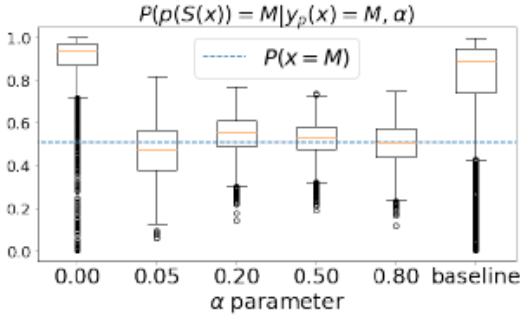
(a) $P(p(S(x)) = M|y_p(x) = F, \alpha)$, *Deterministic model*



(b) $P(p(S(x)) = M|y_p(x) = M, \alpha)$, *Deterministic model*

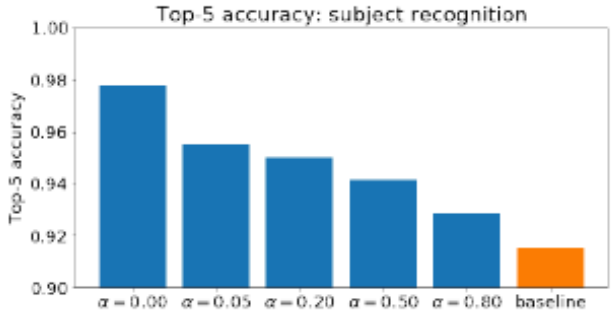


(c) $P(p(S(x)) = M|y_p(x) = F, \alpha)$, *Stochastic model*

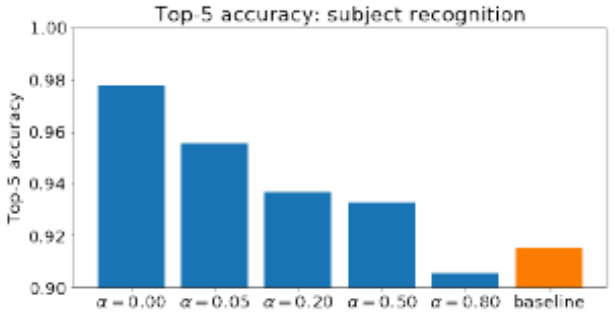


(d) $P(p(S(x)) = M|y_p(x) = M, \alpha)$, *Stochastic model*

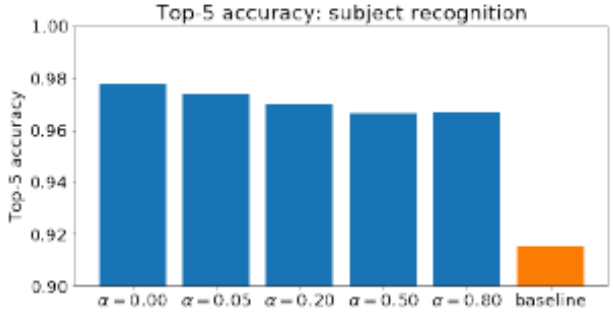
Experimental Results:
Identification Performance



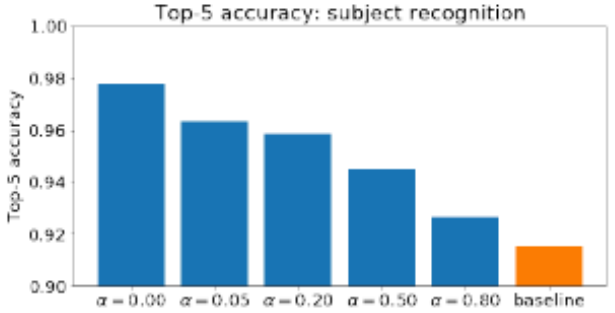
(a) *Top-5 Categorical accuracy, Deterministic model, adversarial approach*



(b) *Top-5 Categorical accuracy, stochastic model, adversarial approach*

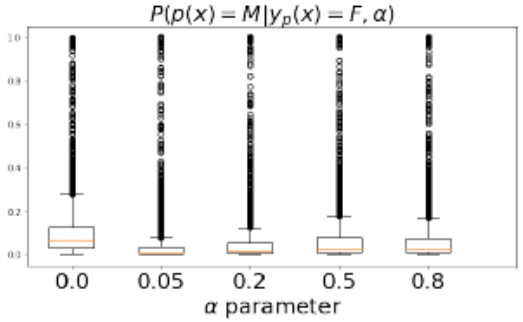


(c) *Top-5 Categorical accuracy, Deterministic model, plug-and-play approach*

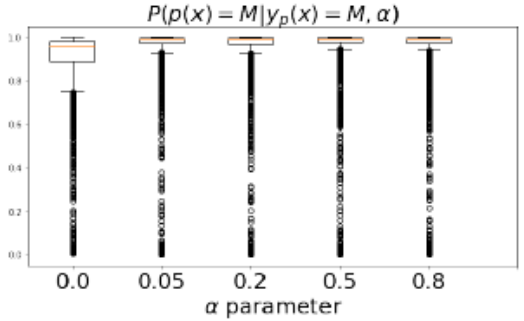


(d) *Top-5 Categorical accuracy, stochastic model, plug-and-play approach*

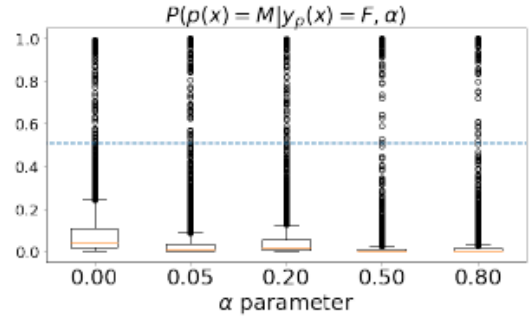
Experimental Results:
Performance on unfiltered data



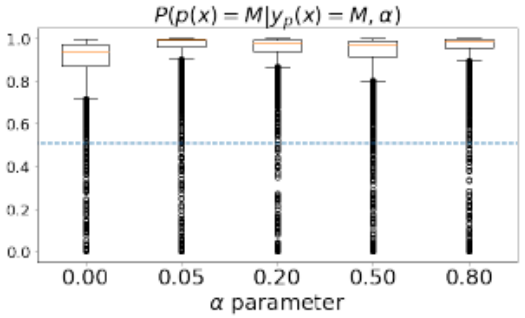
(a) $P(p(x) = M|y_p(x) = F, \alpha)$, *Deterministic model*



(b) $P(p(x) = M|y_p(x) = M, \alpha)$, *Deterministic model*

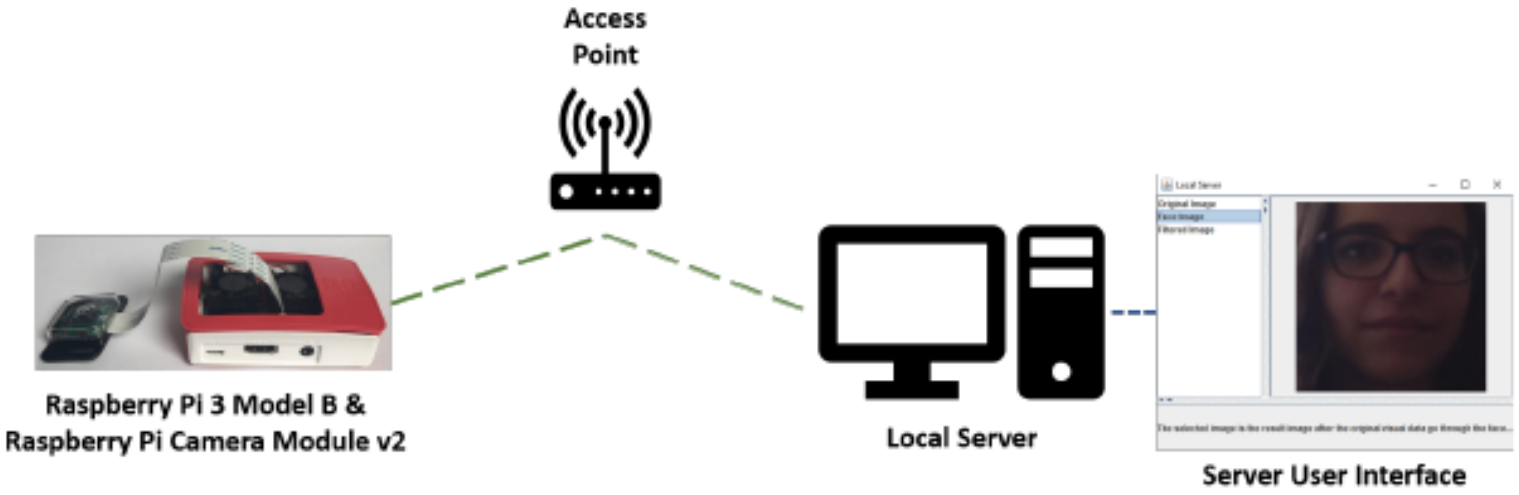


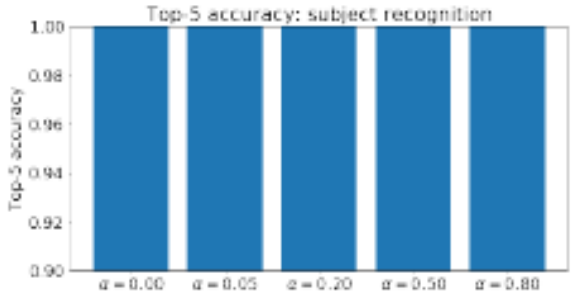
(c) $P(p(x) = M|y_p(x) = F, \alpha)$, *Stochastic model*



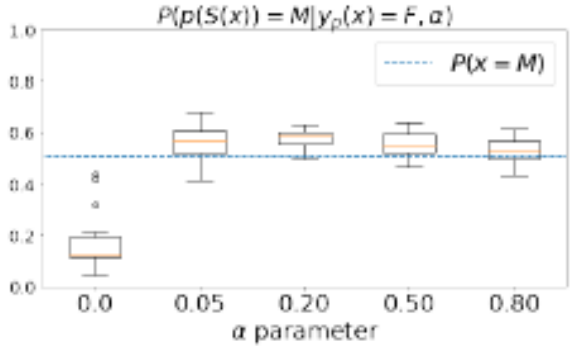
(d) $P(p(x) = M|y_p(x) = M, \alpha)$, *Stochastic model*

Hardware Prototype

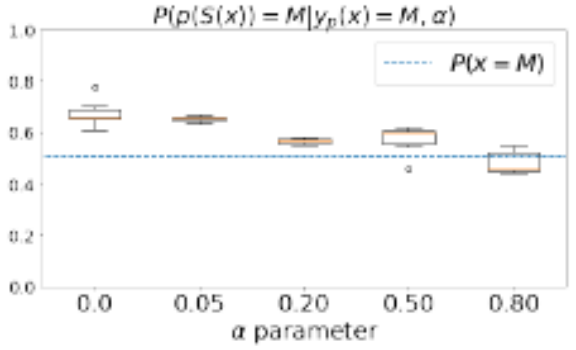




(a) *Top-5 categorical accuracy, stochastic model.*

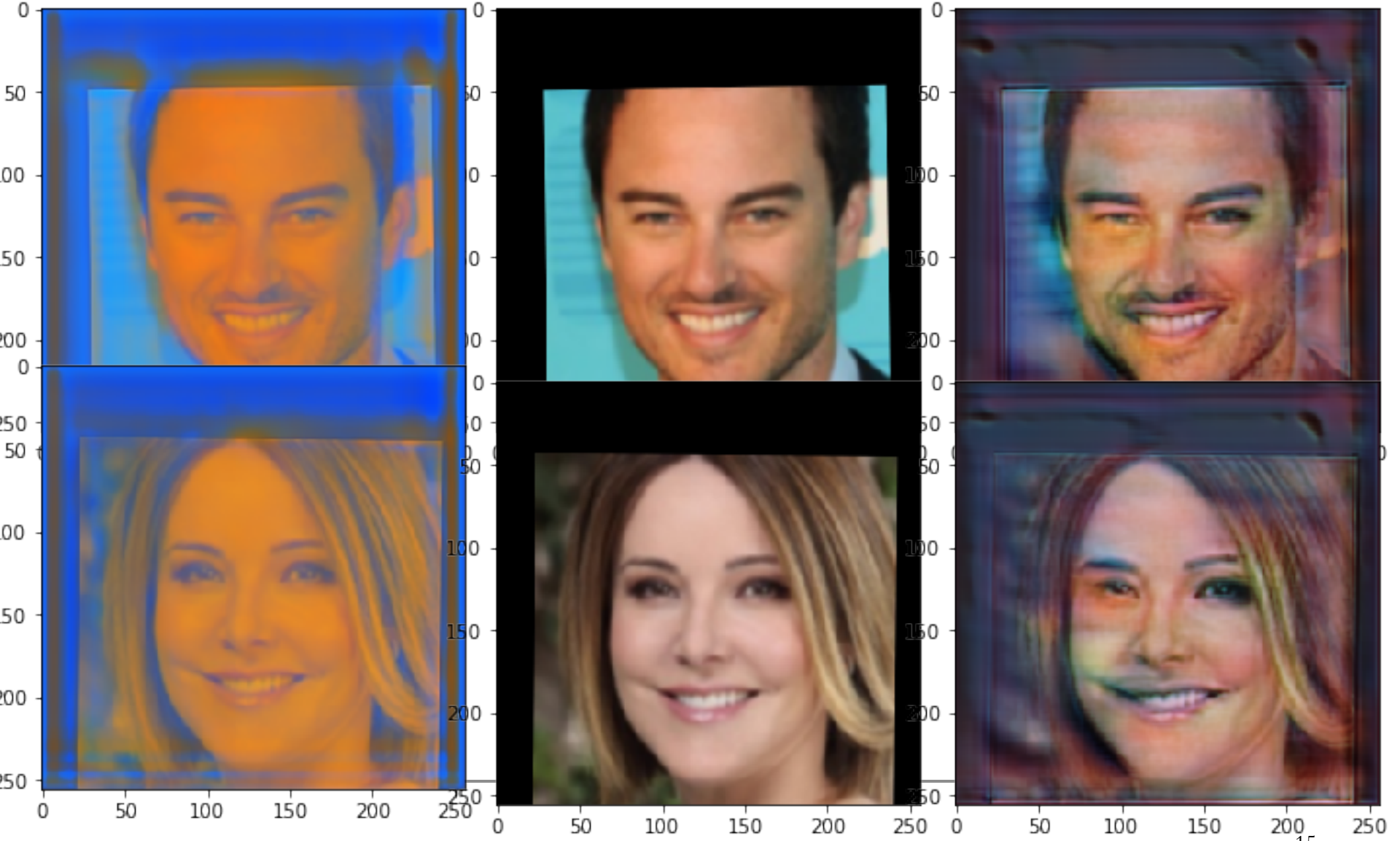


(b) $P(p(S(x)) = M|y_p(x) = F, \alpha)$, *Stochastic model*



(c) $P(p(S(x)) = M|y_p(x) = M, \alpha)$, *Stochastic model*

How do They Look Like?



Related Work

- Differential privacy
  - Utility loss; non-collaborative
- Information bottleneck and privacy funnel
  - No design mechanism
- Adversarial examples
  - Targeted to make a system fail
- Removing nuisance/attributes
  - No for utility and privacy; closer to fairness
- Protecting training data

Discussion

- Privacy in a collaborative environment
- Mathematics of privacy connected to other ML tasks
- Mathematics of privacy connected to other fields like information theory