

# **Outsourcing Training without Uploading Data**

#### via Efficient Collaborative Open-Source Sampling



#### Junyuan Hong<sup>\*1</sup>, Lingjuan Lyu<sup>2</sup>, Jiayu Zhou<sup>1</sup> and Micheal Spranger<sup>2</sup>

<sup>1</sup>Michigan State University, <sup>2</sup>Sony AI

\*Work done during internship at Sony AI.



## Outsourcing training empowers edge intelligence



Machine Learning as a Service (MLaaS)





Amazon SageMaker

Wearables

**Cloud Machine** Learning Engine

¥,

#### Intelligent Edge Devices





Transportation

& Mobility



Home Control Appliances Gateway

Smart Home Devices

Factory Automation Building &







Healthcare



#### Outsourcing training compromises privacy



## New idea: Find a privacy-free proxy dataset



4

## New idea: Find a privacy-free proxy dataset



#### Efficient Collaborative Open-Source Sampling (ECOS)



6

#### Efficient Collaborative Open-Source Sampling (ECOS)



Sony Al

#### **Application: Selective Manual Labeling**





## Application: Selective Manual Labeling

 When the same number of OOD data (sampling budget) are labeled, samples by ECOS lead to better accuracy in semisupervised learning.



#### Conclusions

- New privacy-preserving training: We find public data in place of the client data for cloud training.
- **New sampling paradigm**: ECOS is communication- and computation-efficient and private.
- Flexible on multiple learning tasks: selective manual labeling, automated client labeling, and adaptive model compression.





# Sony Al

# Outsourcing Training without Uploading Data

Thank you for your time!