NOISE: ROBUSTNESS IN FEDERATED LEARNING MODELS

UCSC SIP | CSE-08

SCIENCE INTERNSHIP PROGRAM

01 CORE PARABLE

Garbage in, garbage out.

02 TECHNOLOGY

Privacy, but at what cost?

03 METHODOLOGY

A battery of attacks.

04 CONCLUSIONS

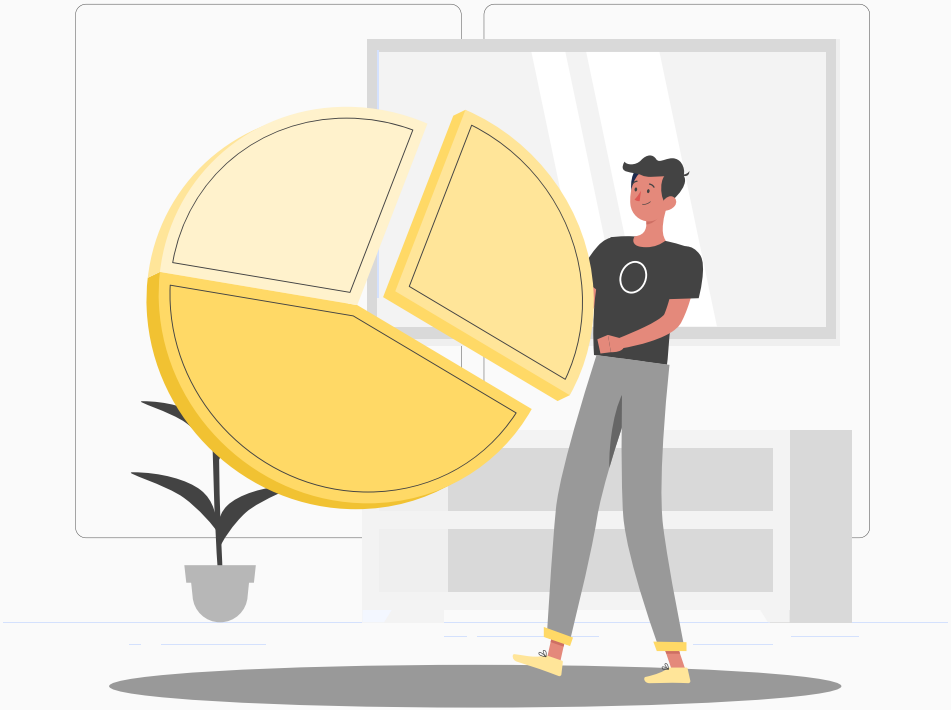
Is federated really safe?

"In deep learning, **there's no data like more data**. The more examples of a given phenomenon a network is exposed to, the more accurately it can pick out patterns and identify things in the real world."

—KAI-FU LEE, AUTHOR OF *AI SUPERPOWERS*

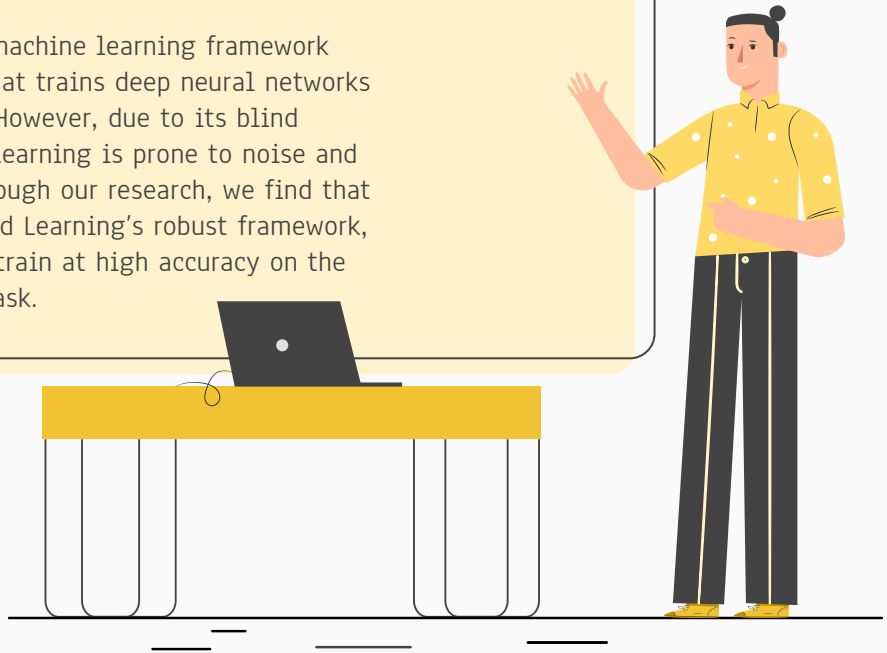
01.
CORE PARABLE

Garbage in, Garbage out



INTRODUCTION

Federated Learning is a novel machine learning framework focused on privacy preservation, that trains deep neural networks through decentralized data. However, due to its blind aggregation methods, Federated Learning is prone to noise and data/model poisoning attacks. Through our research, we find that despite this truth, due to Federated Learning's robust framework, even small neural networks can train at high accuracy on the main task.



CORE PARABLE

CONCERNS WITH FEDERATED LEARNING

Due to Federated Learning complex system, it is prone to many accuracy deteriorates/attack surfaces:

- Label/Image noise from annotation errors
- Model/Data poisoning attacks
- Malicious backdoor attacks

This causes many doubts in developers' minds to implement this framework

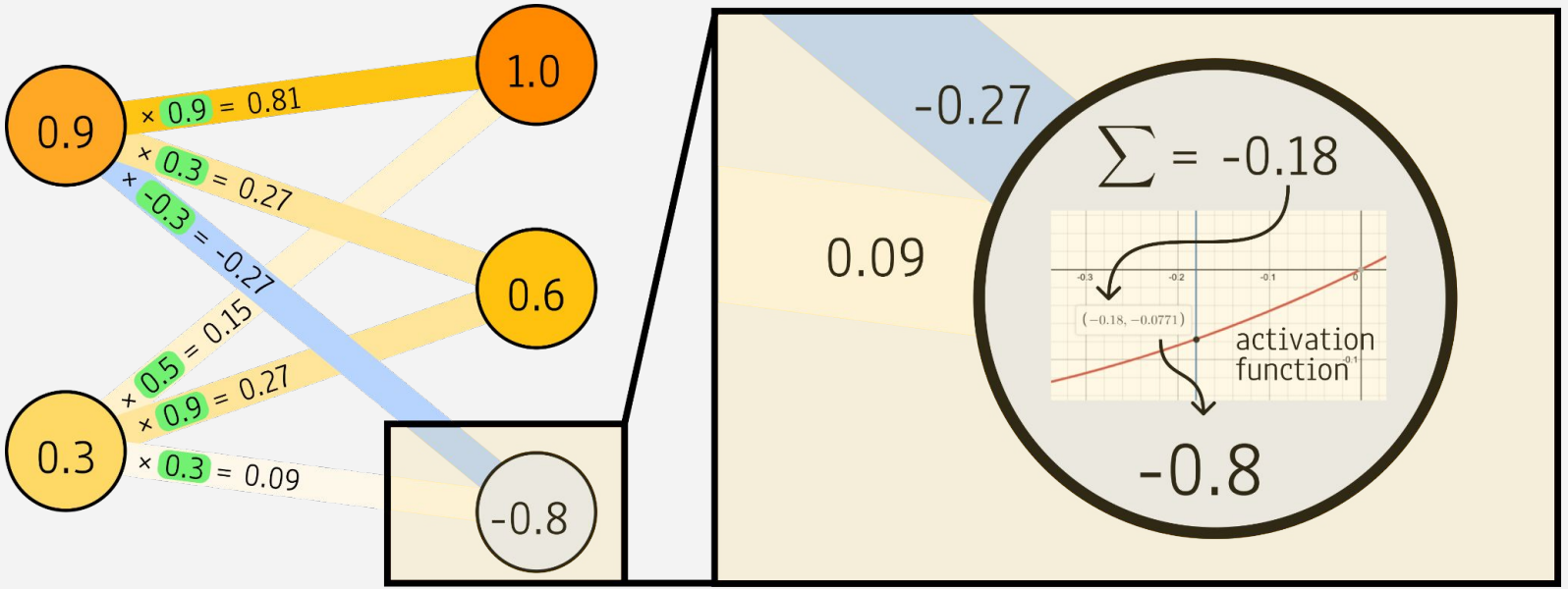
REASONS TO DISMISS THESE CONCERNS

- We found that even with a relatively small model, the accuracy was fairly high for the main tasks when trained with a noisy dataset
- This is possible due to the shear robustness of the Federated Learning architecture
- Important to note: We did not train for backdoor attacks (see Future Works)

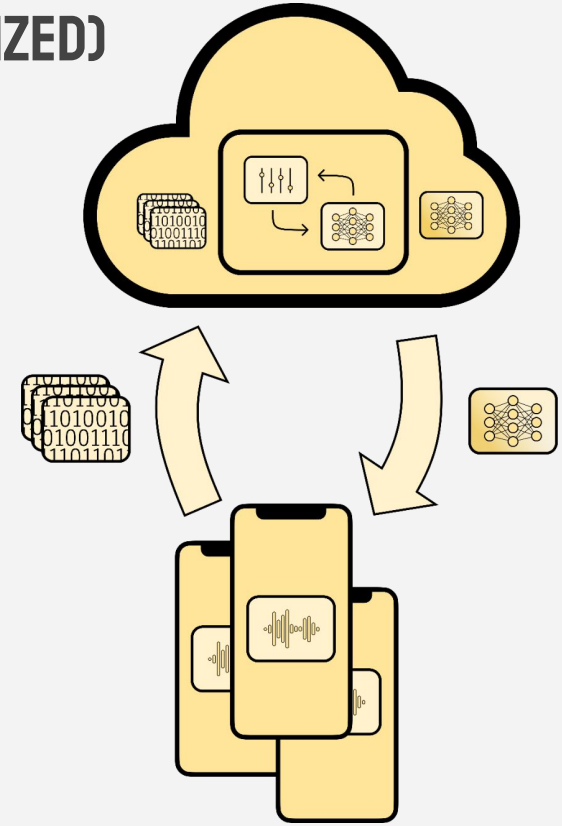
02
TECHNOLOGY

Privacy, but at what cost?

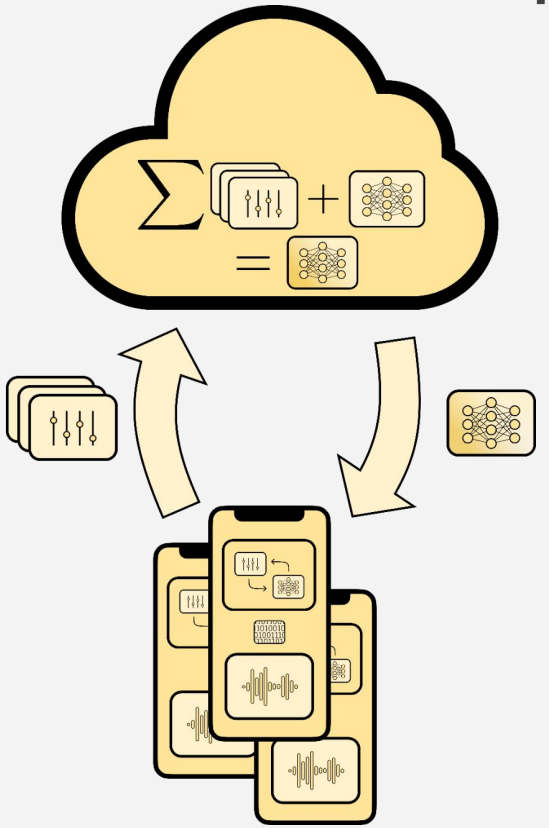
NEURAL NETWORKS



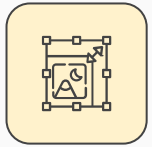
STANDARD (CENTRALIZED) LEARNING



FEDERATED LEARNING



TYPES OF NOISE



FEATURE NOISE

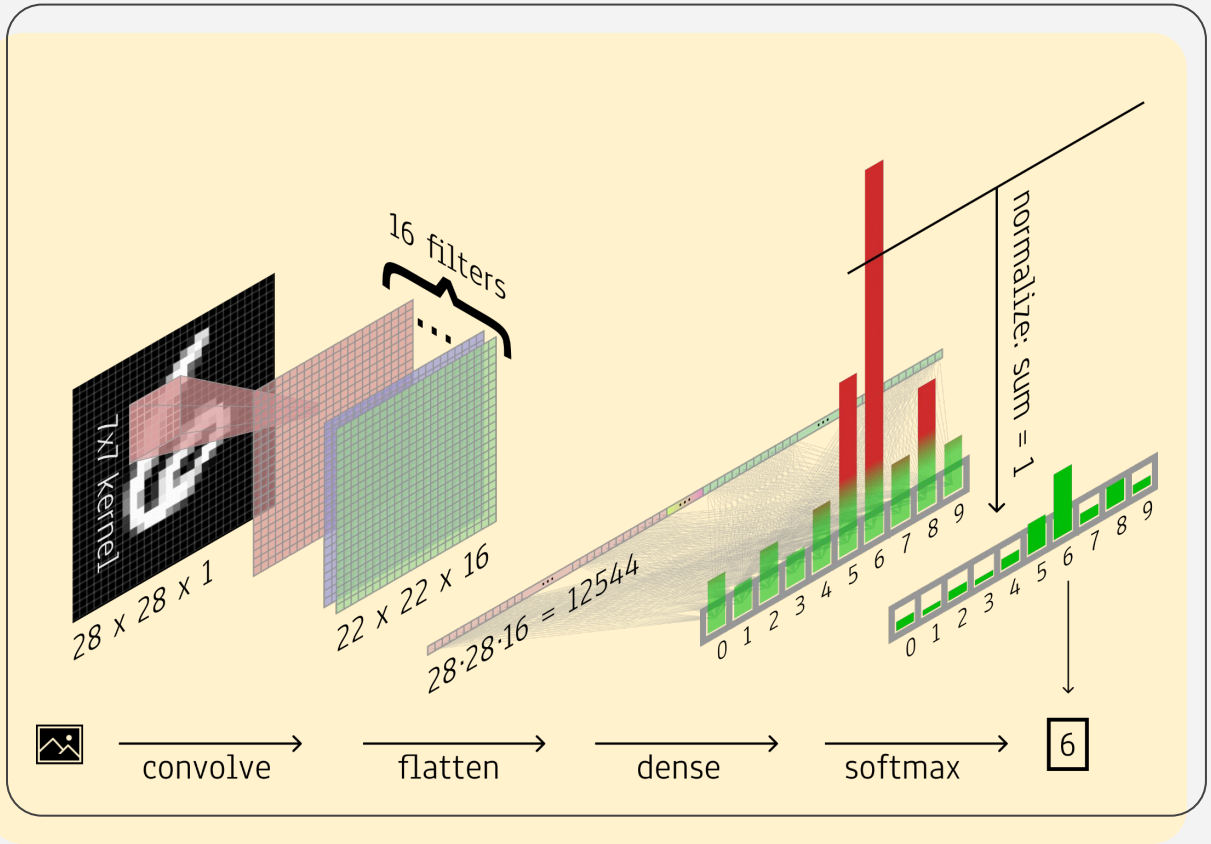
- Blur
- Replaces pixels with the average of their neighbors
- Masking
- Removes part of the image

LABEL NOISE

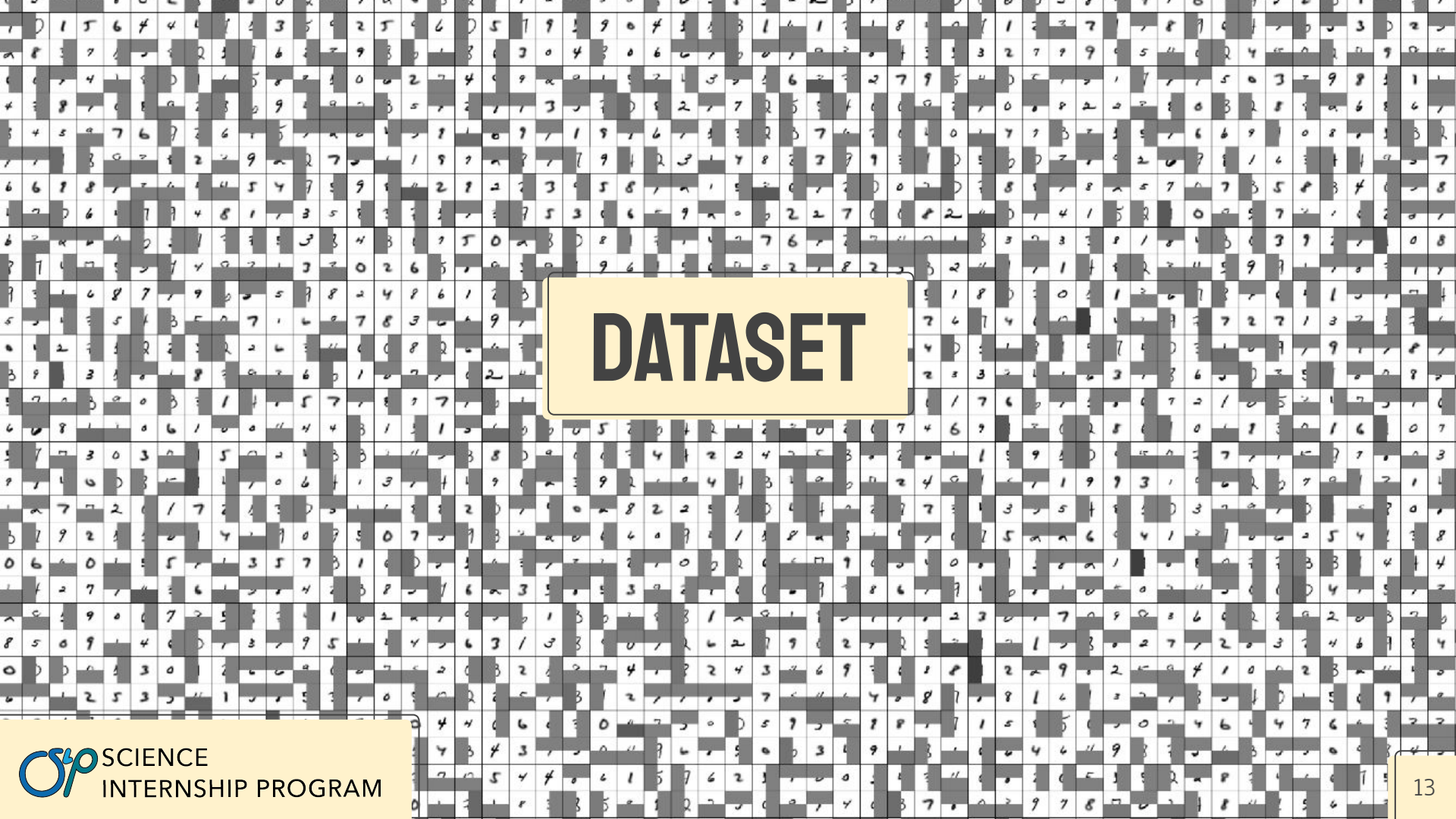
- Swapping labels
- Can be full or partial

03.
METHODOLOGY

A battery of attacks.



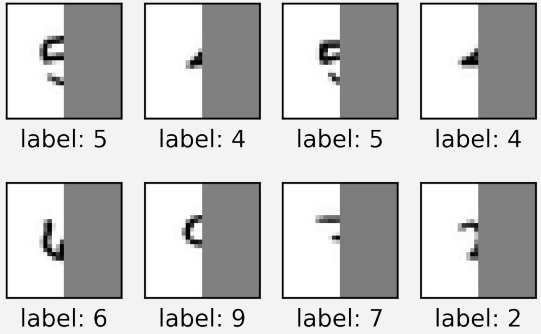
MODEL ARCHITECTURE


DATASET

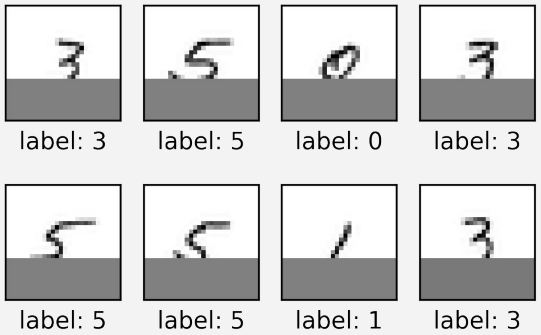
FEATURE NOISE

- Noise in the features, aka the input images
  - Mask noise
  - Blur noise
  - These examples are **harder**, but not wrong

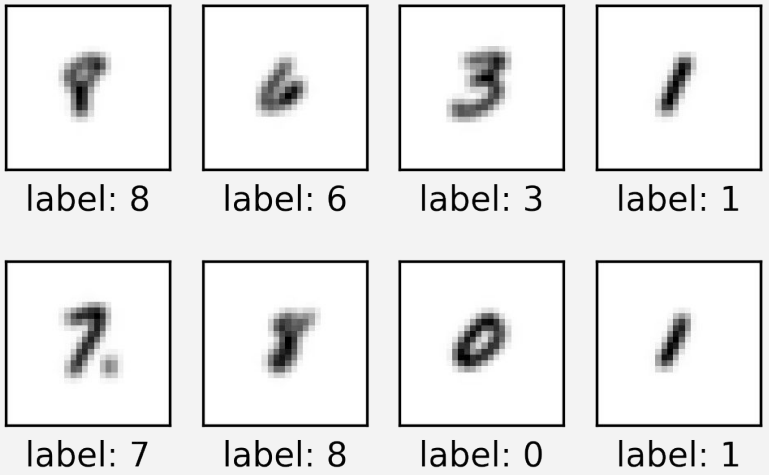
mask_right_half



mask_bot_third



gaussian_1_5

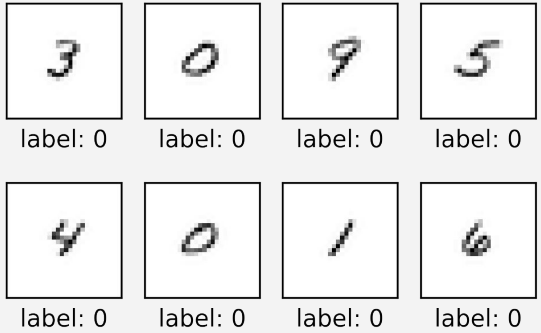


SCIENCE
INTERNSHIP PROGRAM

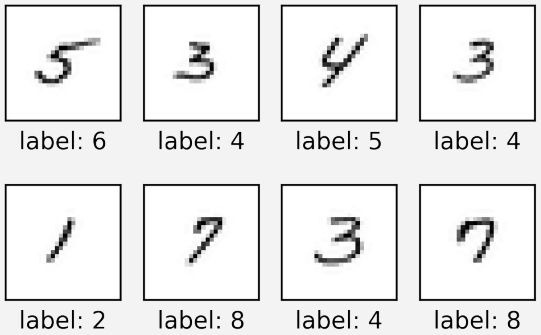
LABEL NOISE

- Noise in the labels for each example
  - Swapping two numbers
  - Setting all label to zero
  - These examples are **actively wrong**

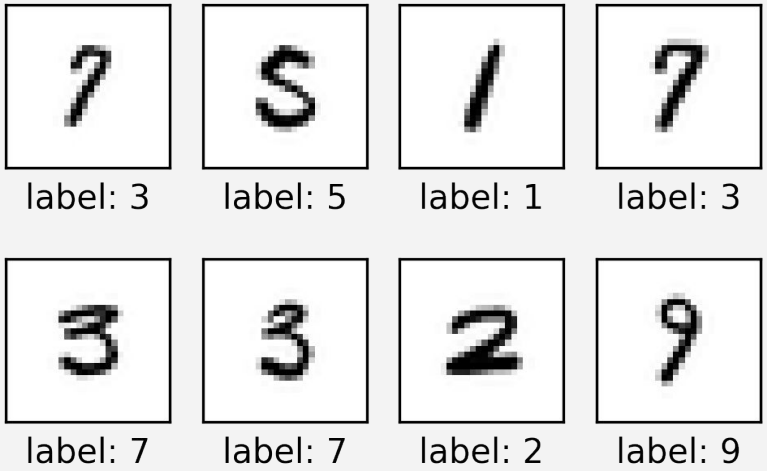
zero_labels



shift_label_up



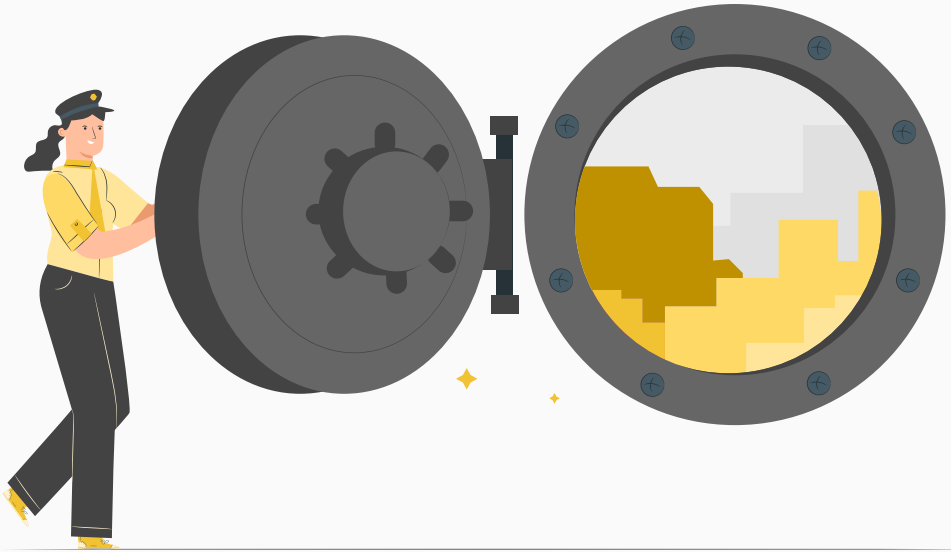
swap_three_seven

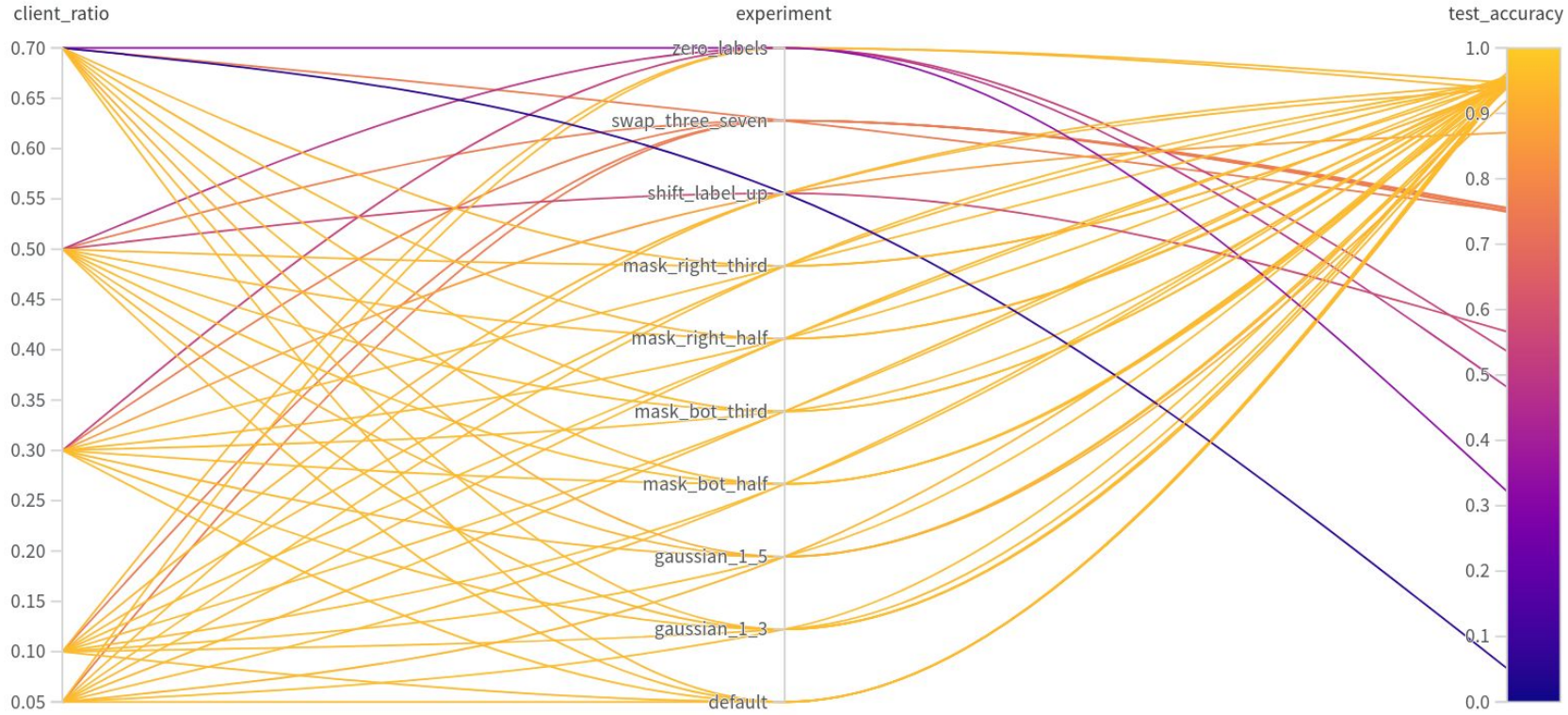


SCIENCE
INTERNSHIP PROGRAM

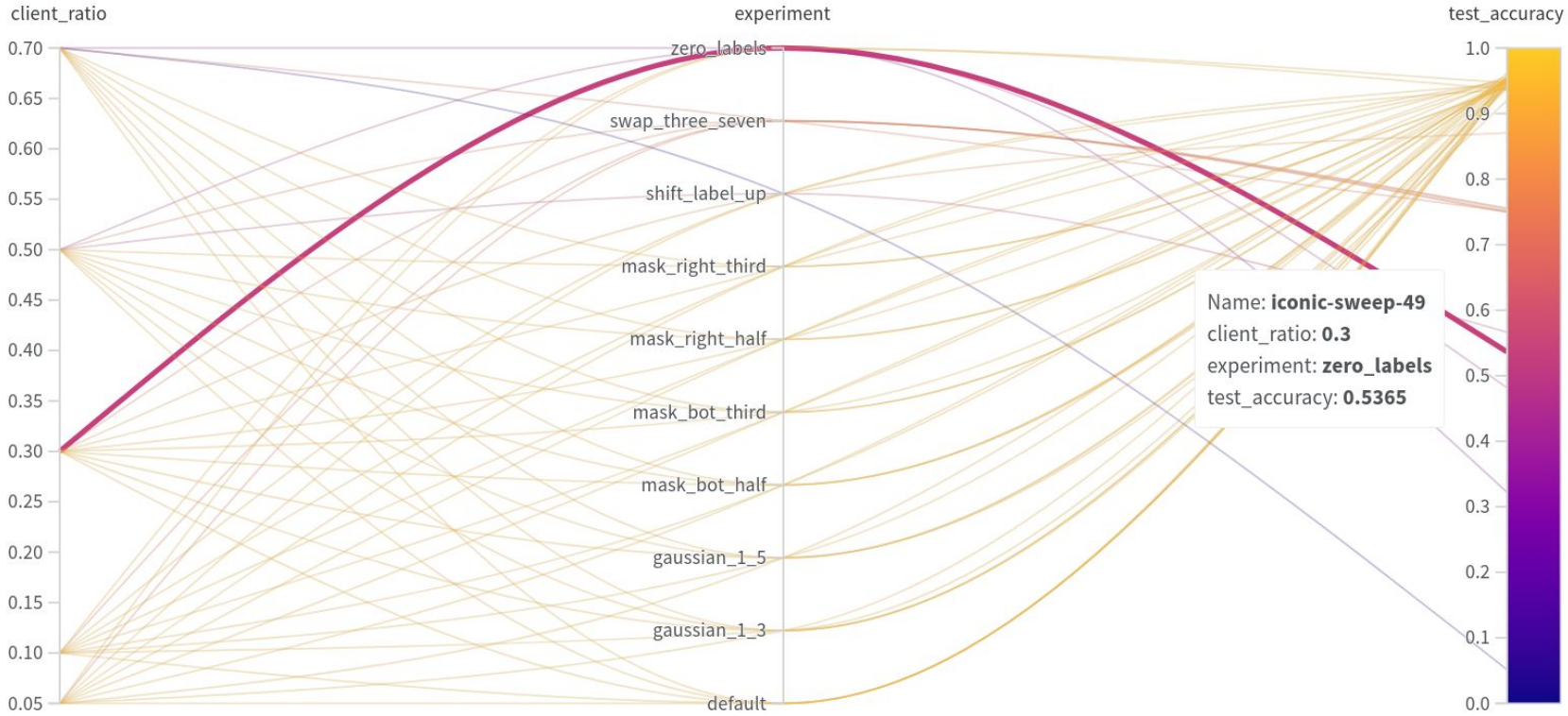
04.
CONCLUSIONS

Is federated really safe?





EXPERIMENTAL RESULTS



EXPERIMENTAL RESULTS

CONCLUSIONS

ANALYSIS & LESSON LEARNED

- Overall, the model Federated context was able to handle all iterations of malicious client ratios due to its robust aggregation protocol
- Sweeps with lower accuracy were due to noise making it impractical for the model to train because majority of the dataset taught the model wrong data
- Federated Learning is effective in most noisy and clean scenarios

FUTURE WORK

- Utilize Exponential Gradient Reweighting to do Robust Federated Aggregation on noisy datasets
- Utilize Differential Privacy to do anomaly detection against malicious backdoors

SELECTED REFERENCES

WARMUTH, MANFRED K., ET AL.	2021	*Exponentiated Gradient Reweighting for Robust Training Under Label Noise and Beyond*. arXiv
BAGDASARYAN, EUGENE, ET AL.	2019	*How To Backdoor Federated Learning*. arXiv
CHEN, XINYUN, ET AL.	2017	*Targeted Backdoor Attacks on Deep Learning Systems Using Data Poisoning*. arXiv
BHAGOJI, ARJUN, ET AL.	2019	*Analyzing Federated Learning through an Adversarial Lens*. International Conference on Machine Learning
CHEN, CHIEN-LUN, ET AL.	2020	*Backdoor Attacks on Federated Meta-Learning*. arXiv

THANKS

Does anyone have any questions?

SCIENCE
INTERNSHIP PROGRAM