Workshop on **Systems Challenges in Reliable and Secure Federated Learning**
Oct 25, 2021

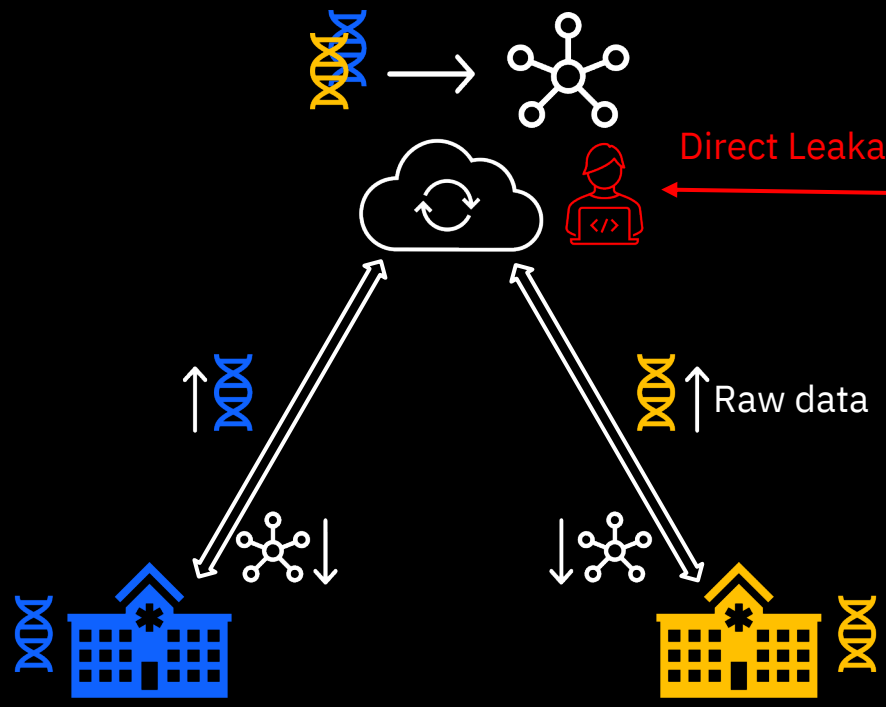
IBM **Research**

Separation of Powers in Federated Learning

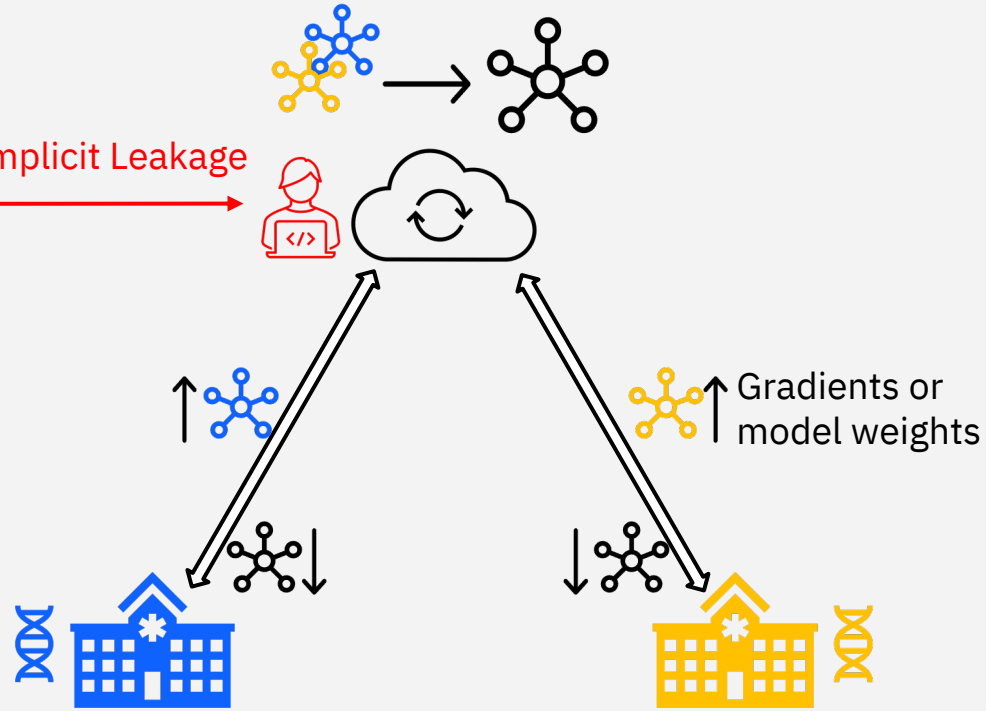
Pau-Chen **Cheng**
Kevin **Eykholt**
Zhongshu **Gu**
Hani **Jamjoom**
K.R. **Jayaram**
Enriquillo **Valdez**
Ashish **Verma**

Photo credit: Dhilung **Kirat**

Collaborative Learning



Federated Learning



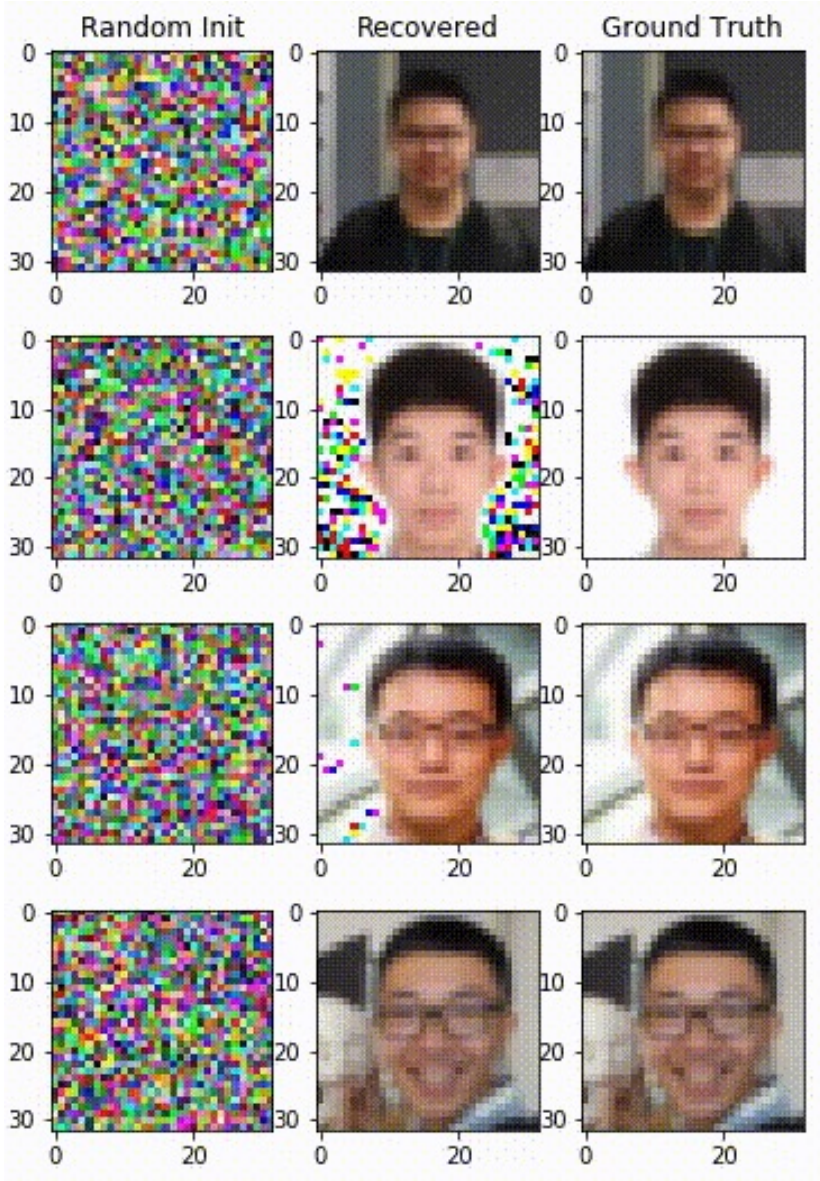
Information Leakage at **Aggregation**

Leakage in Model Updates

- Attribution leakage
- Content leakage

Examples

- DLG: Deep Leakage from Gradients (NeurIPS 2019) https://arxiv.org/pdf/1906.08935.pdf
- iDLG: Improved Deep Leakage from Gradients https://arxiv.org/abs/2001.02610
- IG: Inverting Gradients (NeurIPS 2020) https://arxiv.org/abs/2003.14053



https://github.com/mit-han-lab/dlg

Key **Insights**

Data Concentration

- All model updates in a central fusion server
- Single point of failure
- Leak complete and intact model updates

Algorithmic Properties

- Aggregation algorithms are bijective and coordinate-wise
- Aggregation *can* work on partitioned and out-of-order data
- Attacks *cannot* work on partitioned and out-of-order data

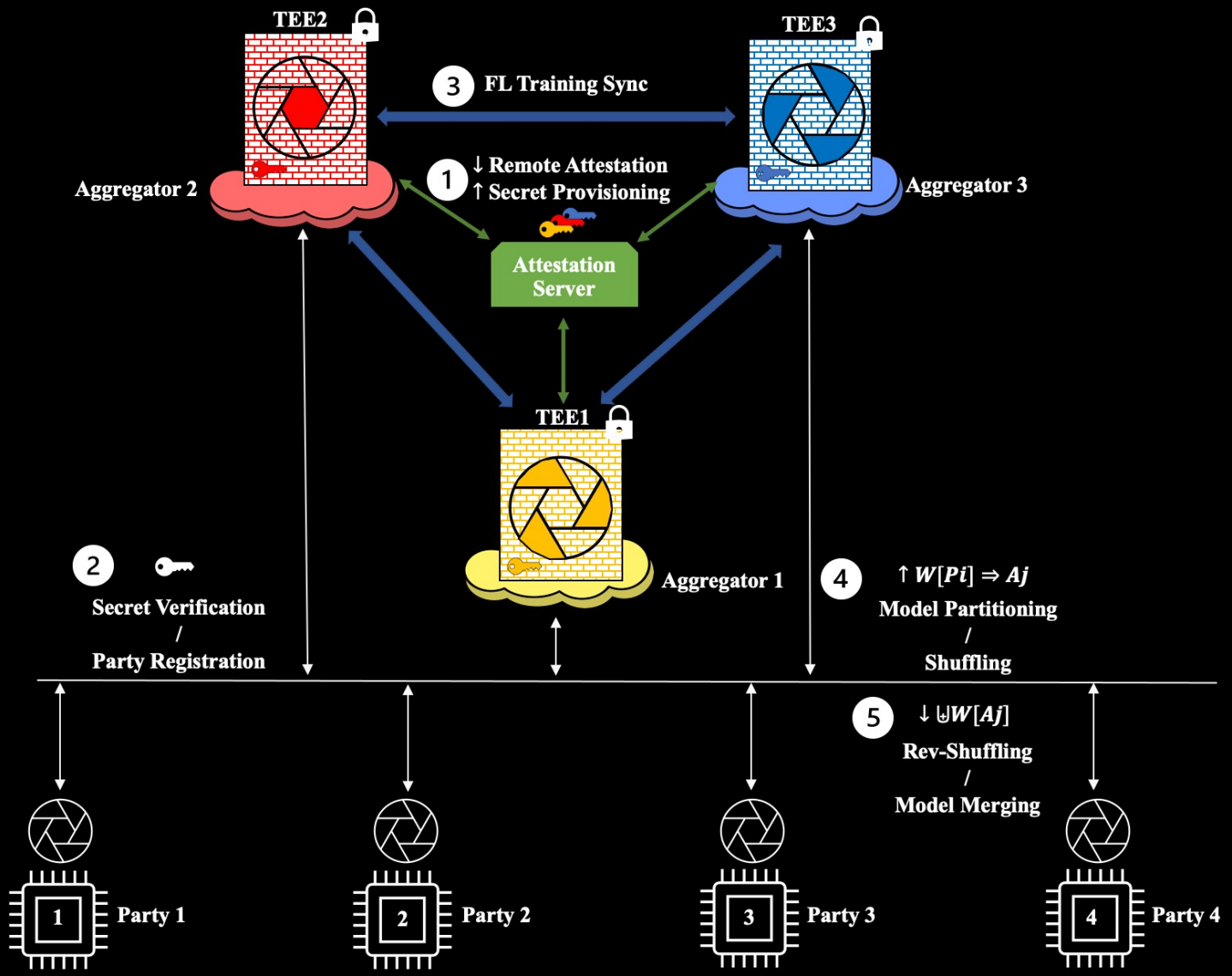
Confidential Execution

- Confidential execution with runtime memory encryption
- Remote attestable
- End-to-end data protection

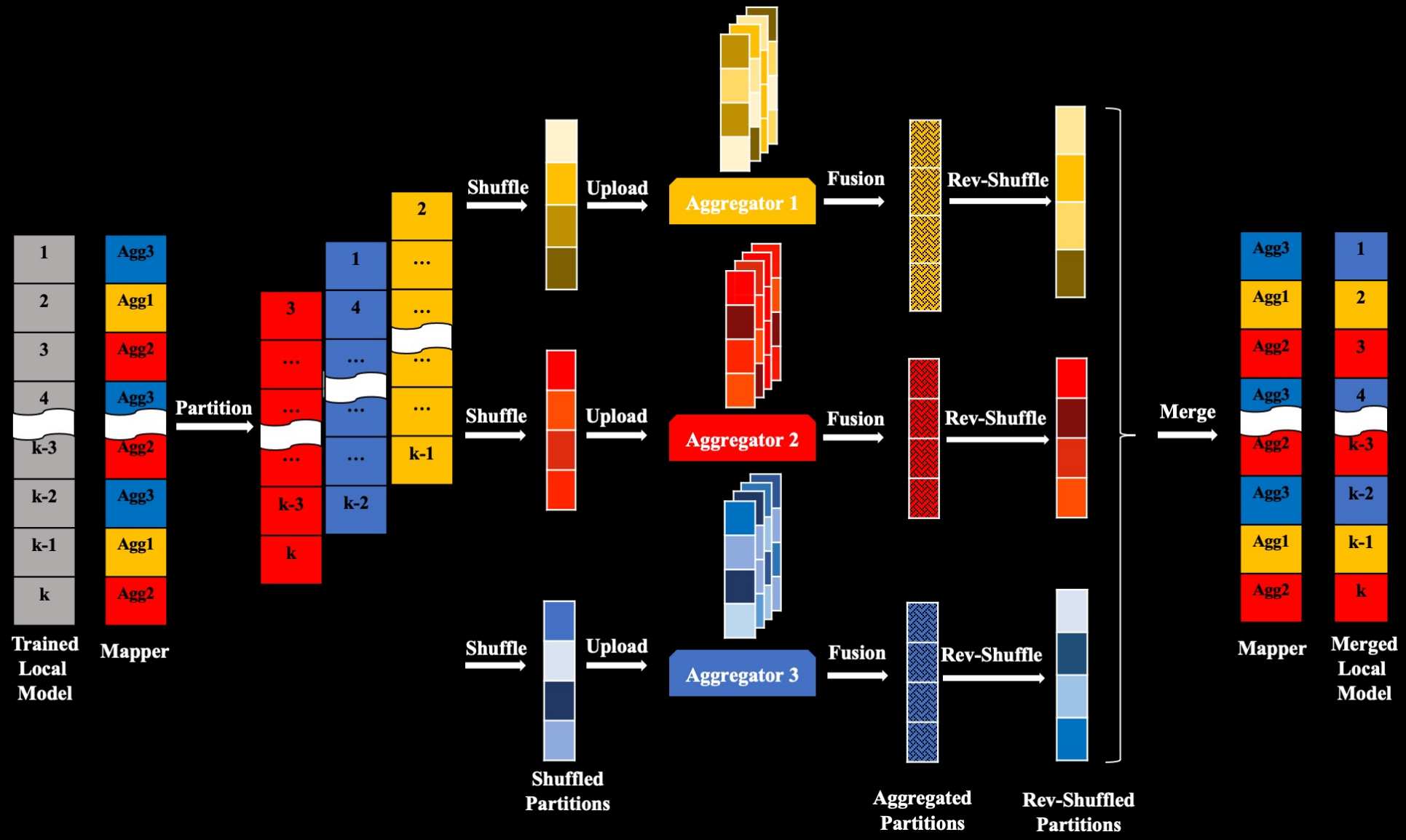
TRUDA: **Defense-in-Depth**

- Decentralized Aggregation
  - Multiple decentralized aggregators → Prevent single point of failure
  - Randomized model partitioning → Fragmentary view of model updates
- Shuffled Aggregation
  - Data shuffling at parameter-level → Out-of-order view of model updates
  - New shuffling every training round → Misaligned model update order across consecutive rounds
- Trustworthy Aggregation
  - SEV to protect FL aggregation → Confidential execution and remote attestation
  - Trusted aggregator authentication → Verifying trustworthiness of aggregators before training

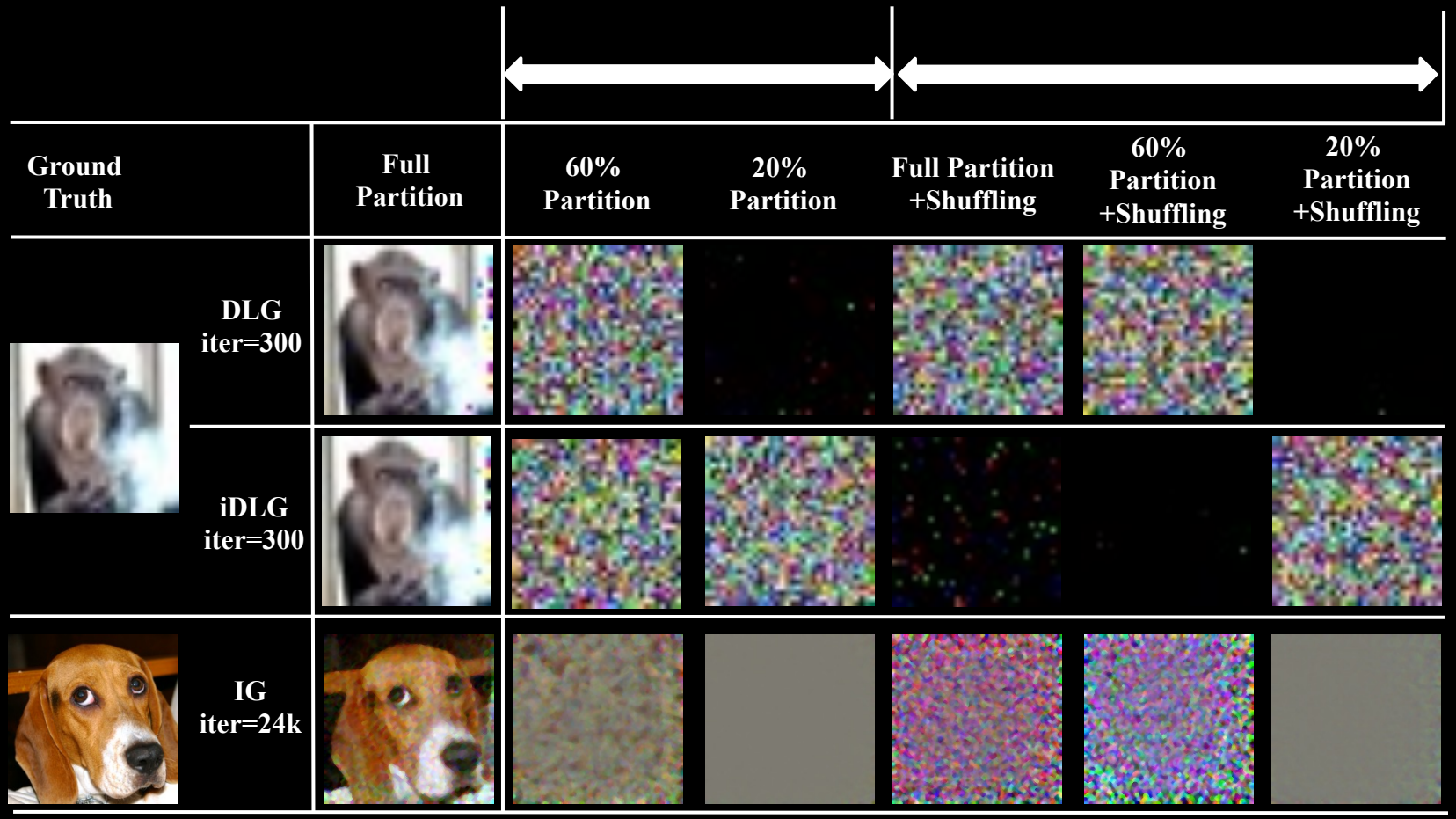
TRUDA **Architecture**



Model **Partitioning and Shuffling**



Initial **Results**



Conclusion

- Breaking data concentration in aggregation
- Exploiting computational properties of aggregation algorithms
- Defense-in-Depth with decentralized, shuffled, and trustworthy aggregation

Q & A

zgu@us.ibm.com

