SAND2021-7968C

Sandia National Laboratories

Exceptional service in the national interest

Privacy-Preserving AutoML

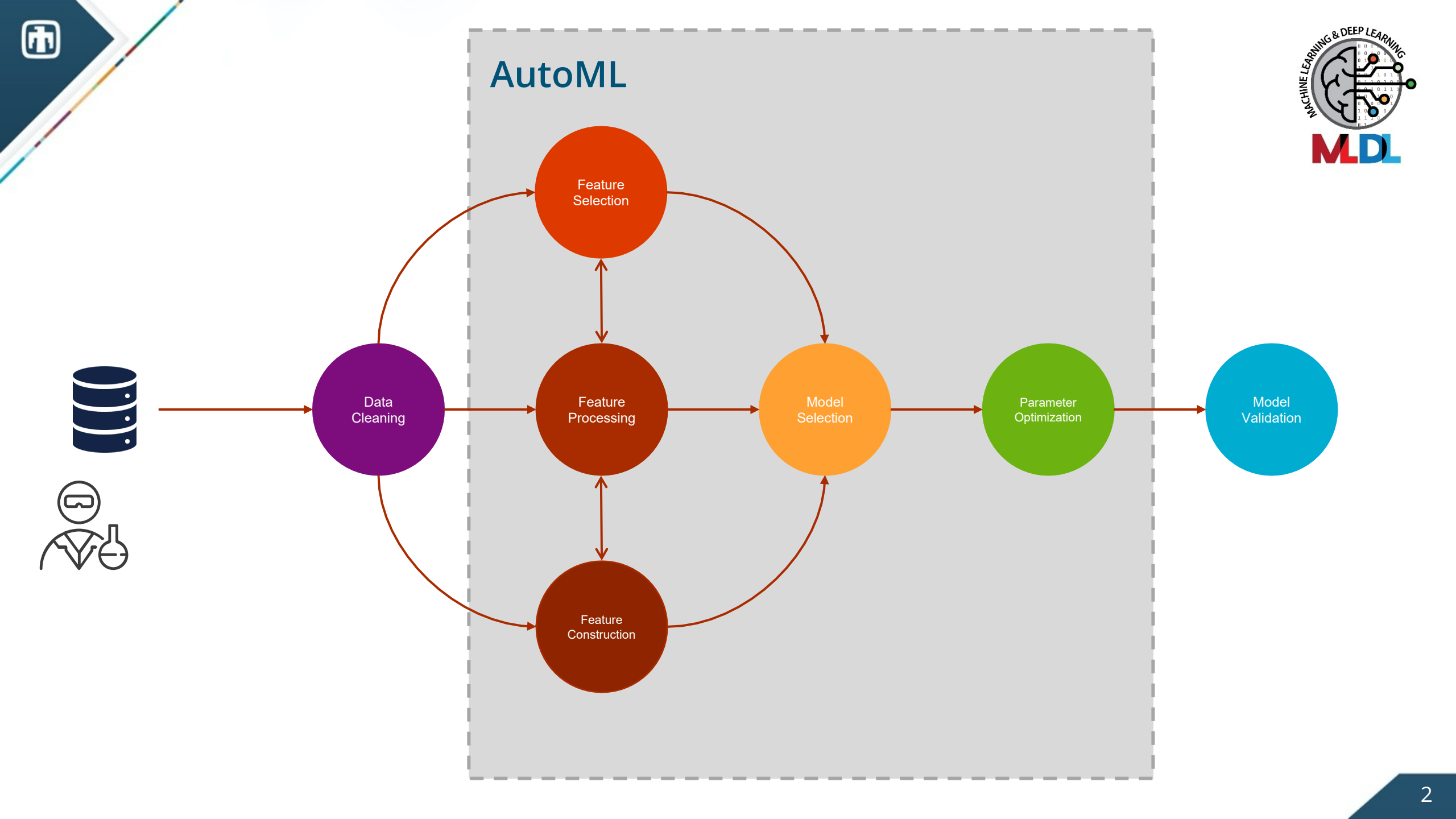
Machine Learning and Deep Learning Conference 2021

Alycia N. Carey – 5629, Nicholas Pattengale - 5682

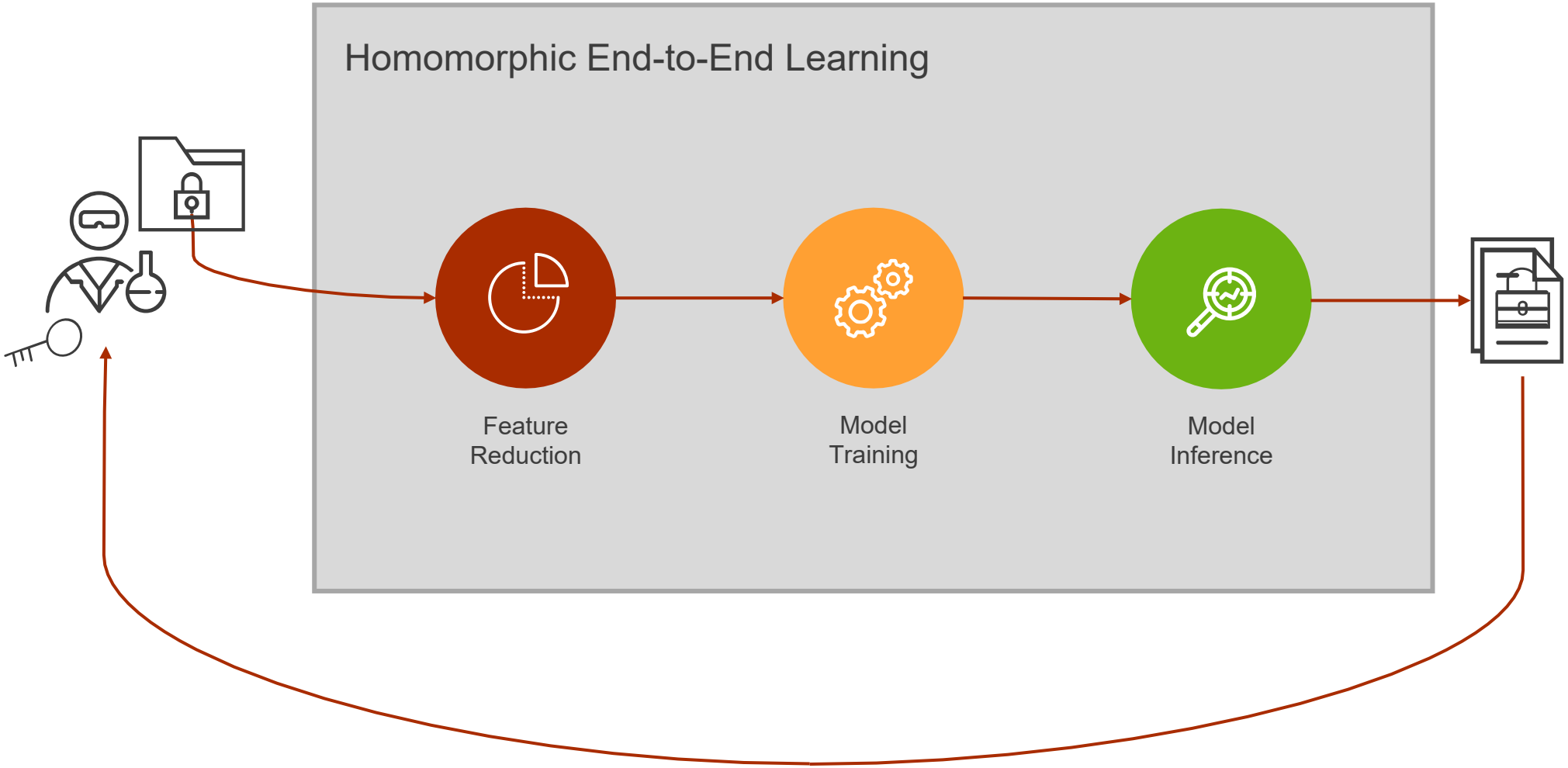
July 19 - 22, 2021

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.


AutoML
Feature Selection
Data Cleaning
Feature Processing
Model Selection
Parameter Optimization
Model Validation
Feature Construction

Motivation – Why Privacy-Preserving AutoML?



Algorithmic Approach

1. Survey

- Catalog published techniques
- Implementation exists or not?
- Open source or not?
- Maturity/Quality

2. Demonstration

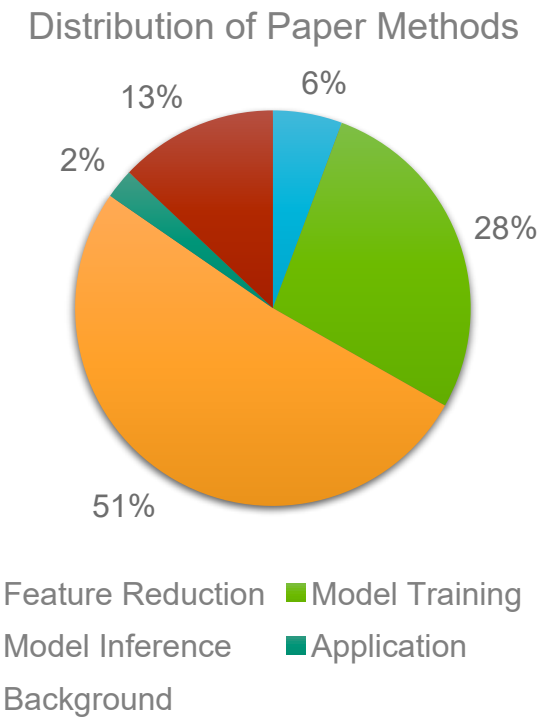
- Use only open source data and modules
- Demo/articulate utility of technique mashup under:
  - Same FHE scheme
    - A PP Distributed Architecture for DLaaS
    - Only model training and model inference
  - Mixed FHE schemes
    - CHIMERA
    - Multi-encrypt/decrypt process

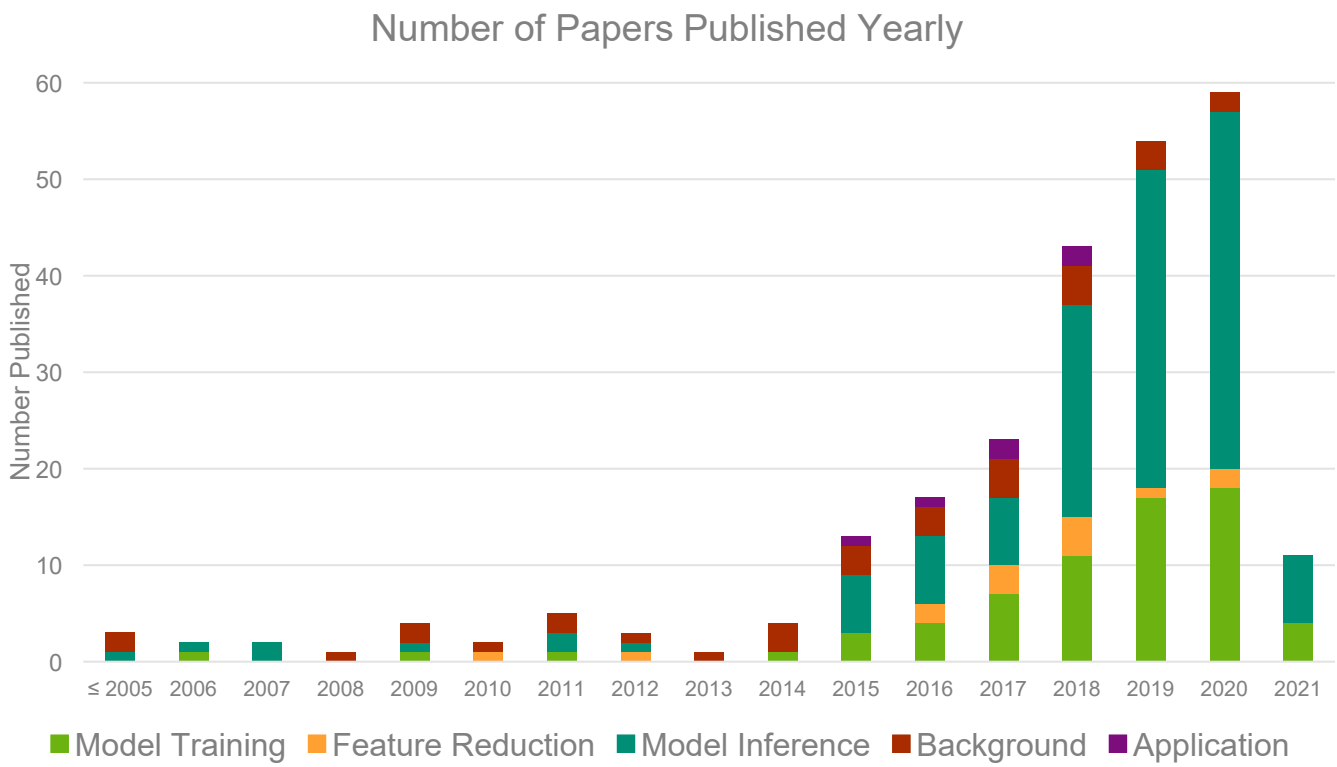
Paper Corpus

Papers were collected from ArXiV, IEEE, IACR, ACM, and Google Scholar using keywords such as: *homomorphic feature reduction, homomorphic model training,* and *homomorphic inference*

Total paper count: 198

- Feature Reduction: 14
- Model Training: 68
- Inference: 127
- Application: 6
- Background: 32





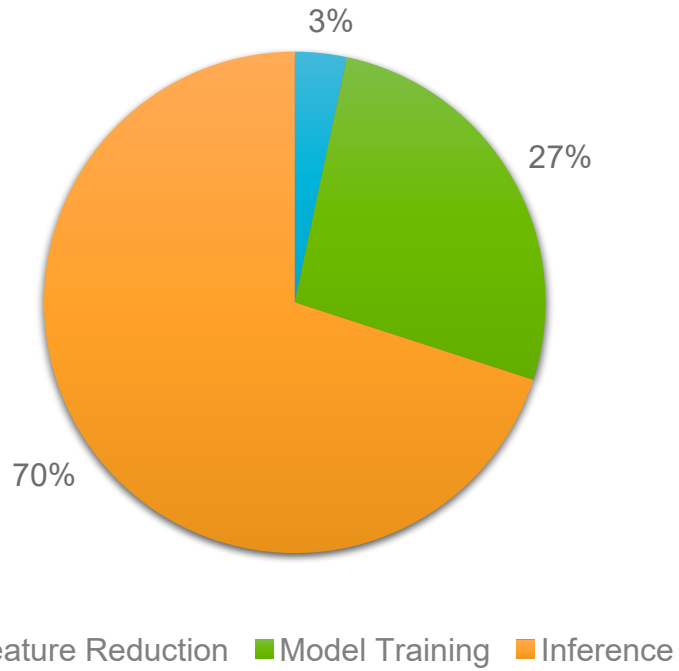
Note: The representation of background and application papers may not be representative of the current landscape. This is due to focus on collecting feature reduction, model training, and model inference publications.

Open-Source Implementations

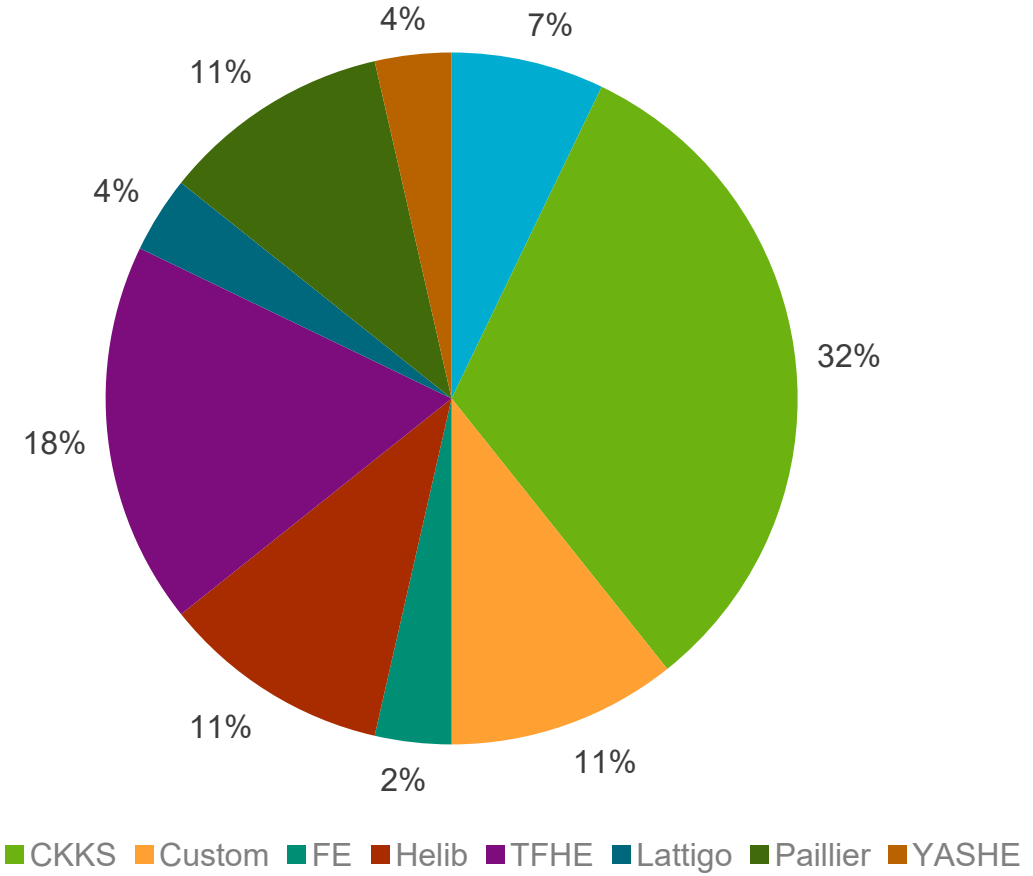
The open-source implementations were collected from papers that listed a public Github or Gitlab repository.

- Total open-source count: 24
  - Feature Reduction: 1
  - Model Training: 8
  - Inference: 21

Distribution of implementation methods

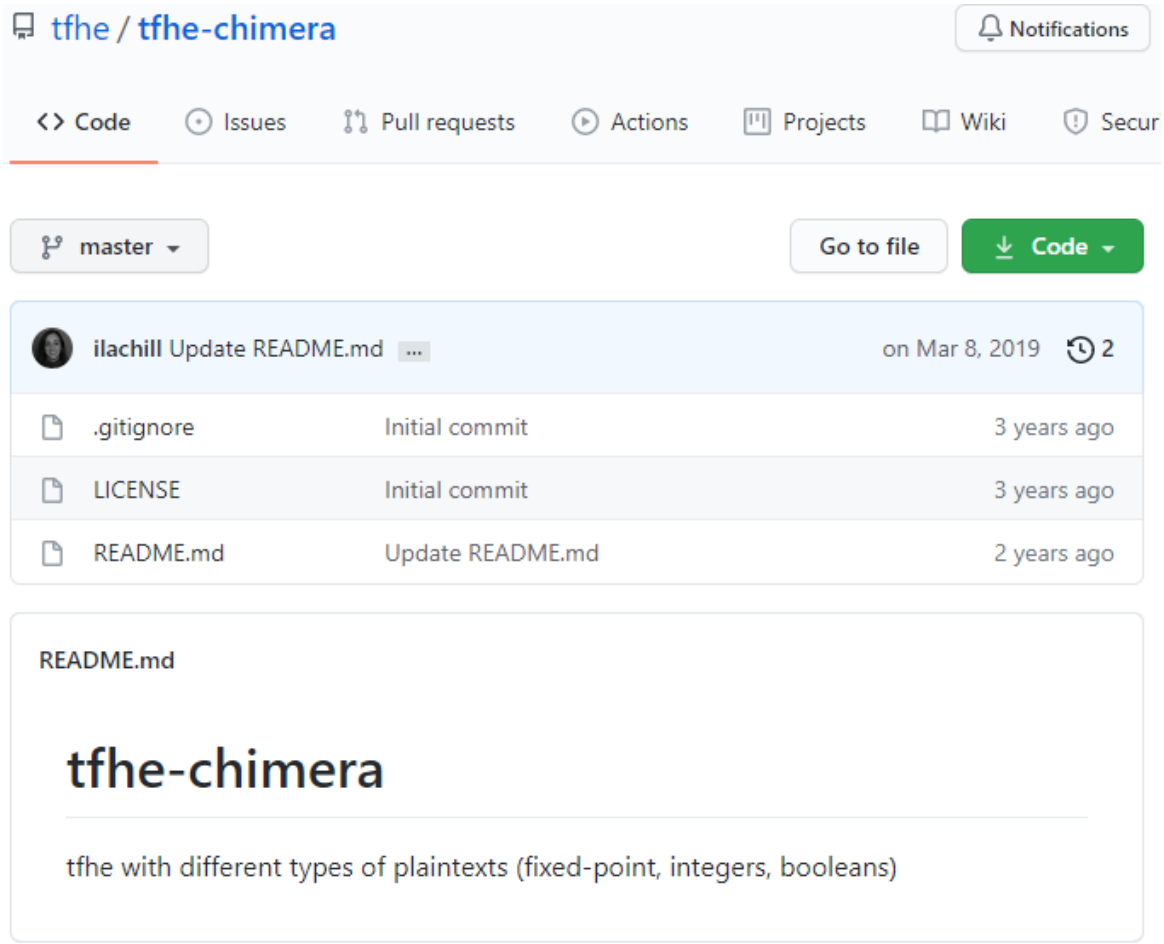


Distribution of FHE schemes in open-source implementations



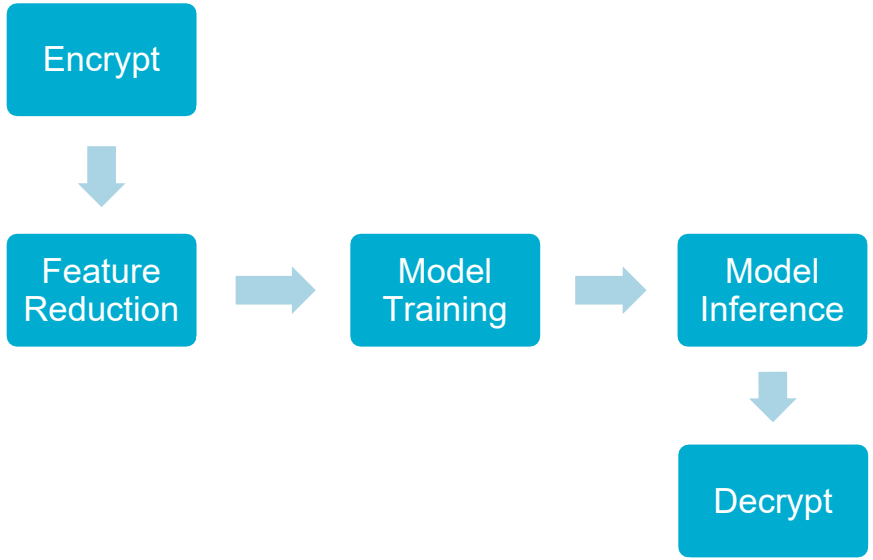
Main Findings

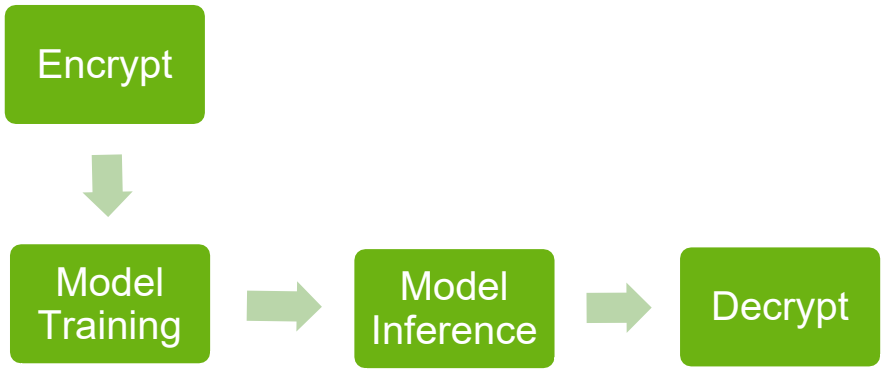
- Most papers that have no open-source implementation have a custom FHE scheme
- Most open-source implementations do not use the same FHE scheme (BFV, CKKS, …)
  - This makes creating an end-to-end pipeline difficult
- The one paper linking different schemes together (CHIMERA) has no open source implementation
- There is a massive lack of open source feature reduction implementations
- Currently, no end-to-end privacy-preserving AutoML pipeline has been released



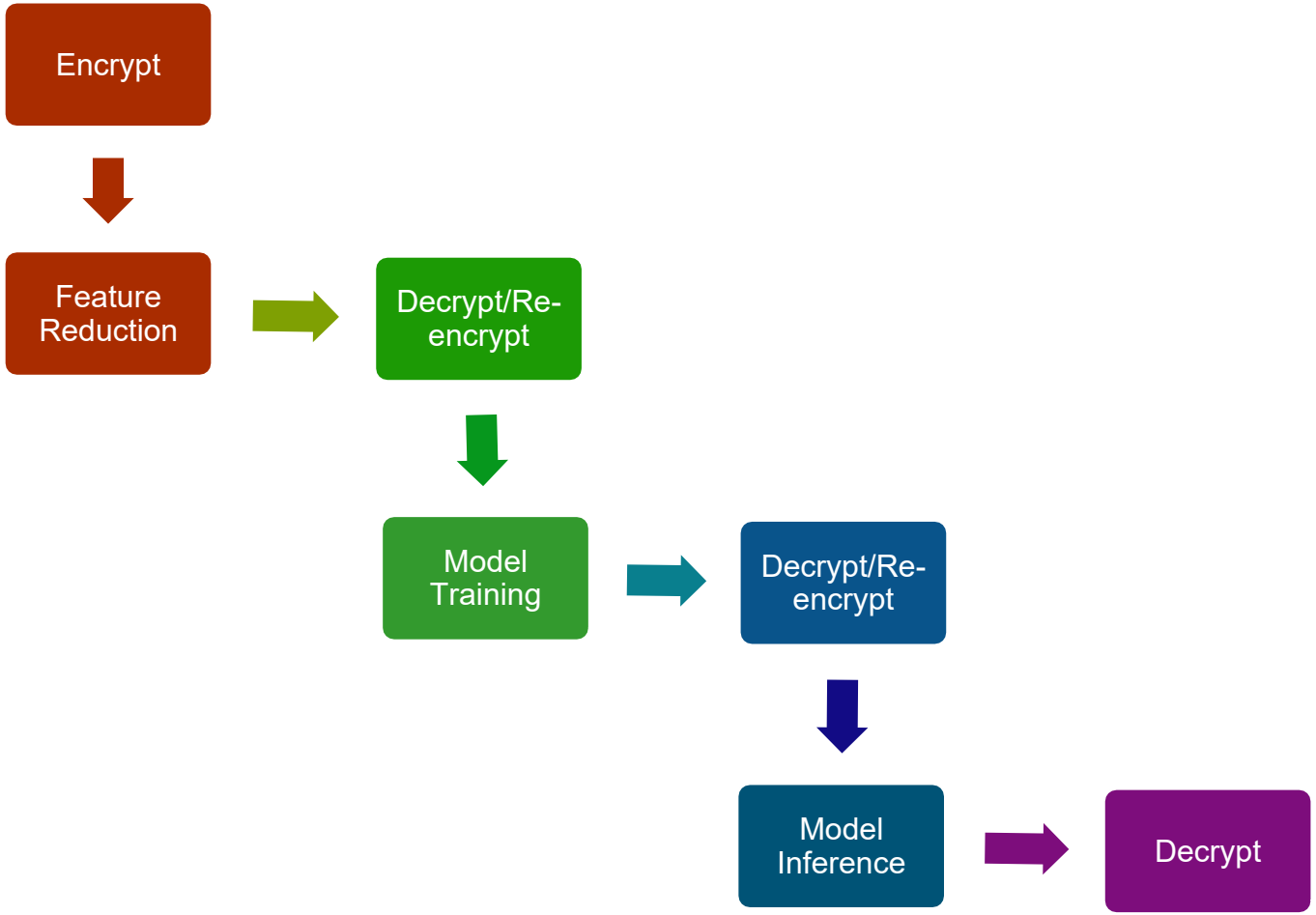
Next Steps - Demonstration

1. Under the same FHE scheme





2. Under different FHE schemes



Questions?

ancarey@sandia.gov