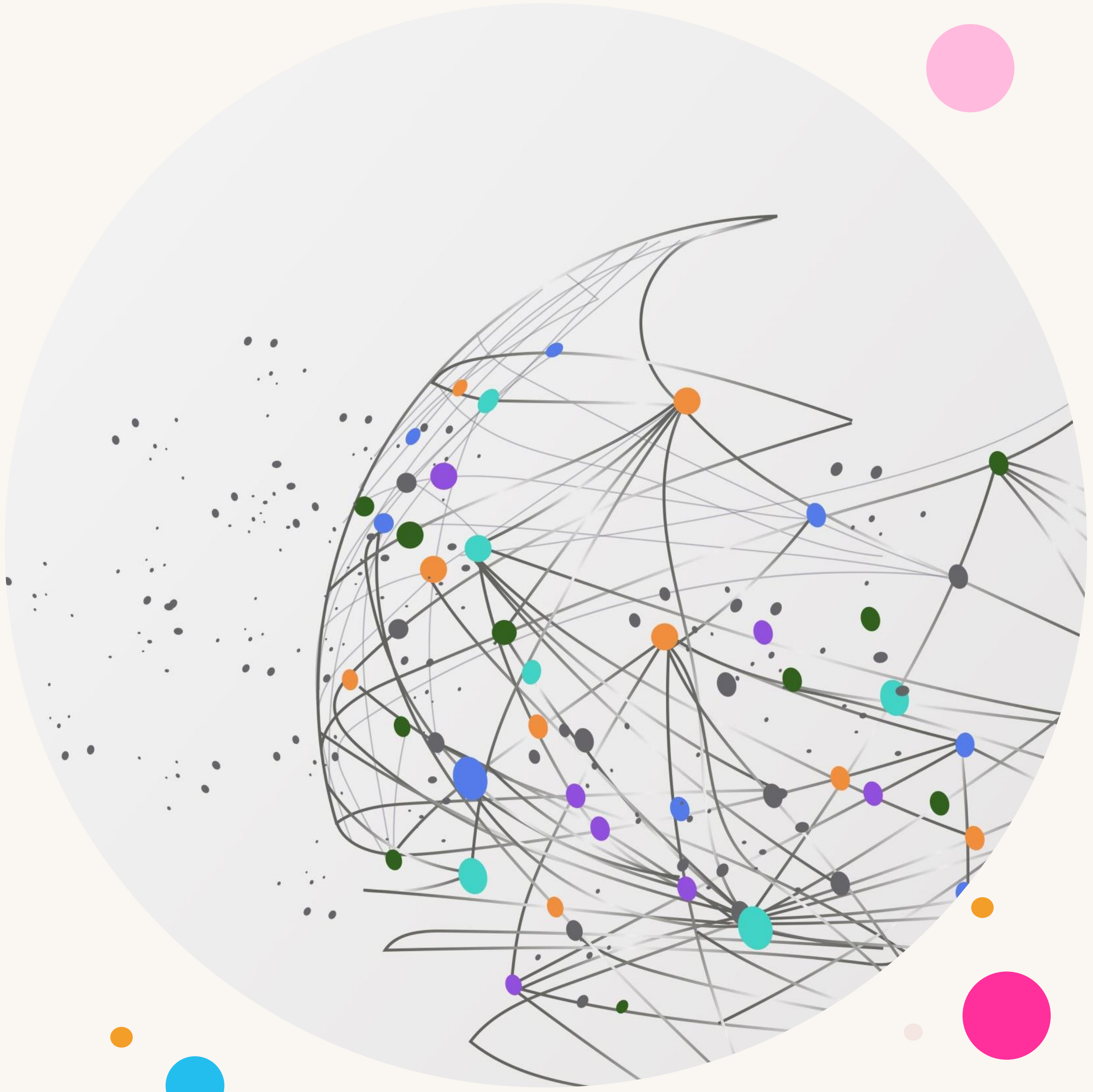
Private AI: Machine Learning on Encrypted Data

Kristin Lauter

Facebook AI Research

University of Washington



Privacy problem with AI?

- Artificial Intelligence: uses (ML) machine learning algorithms to make useful predictions
  - Input: your data
  - Output: some recommendation, decision, or classification
- **Privacy problem**: you have to input **your data** in order to get the valuable prediction!
- Typical AI services are hosted in the cloud, run by a "smart agent" (e.g. Cortana, Siri, Alexa)

New mathematical tool:

Protect privacy and security of your data through encryption

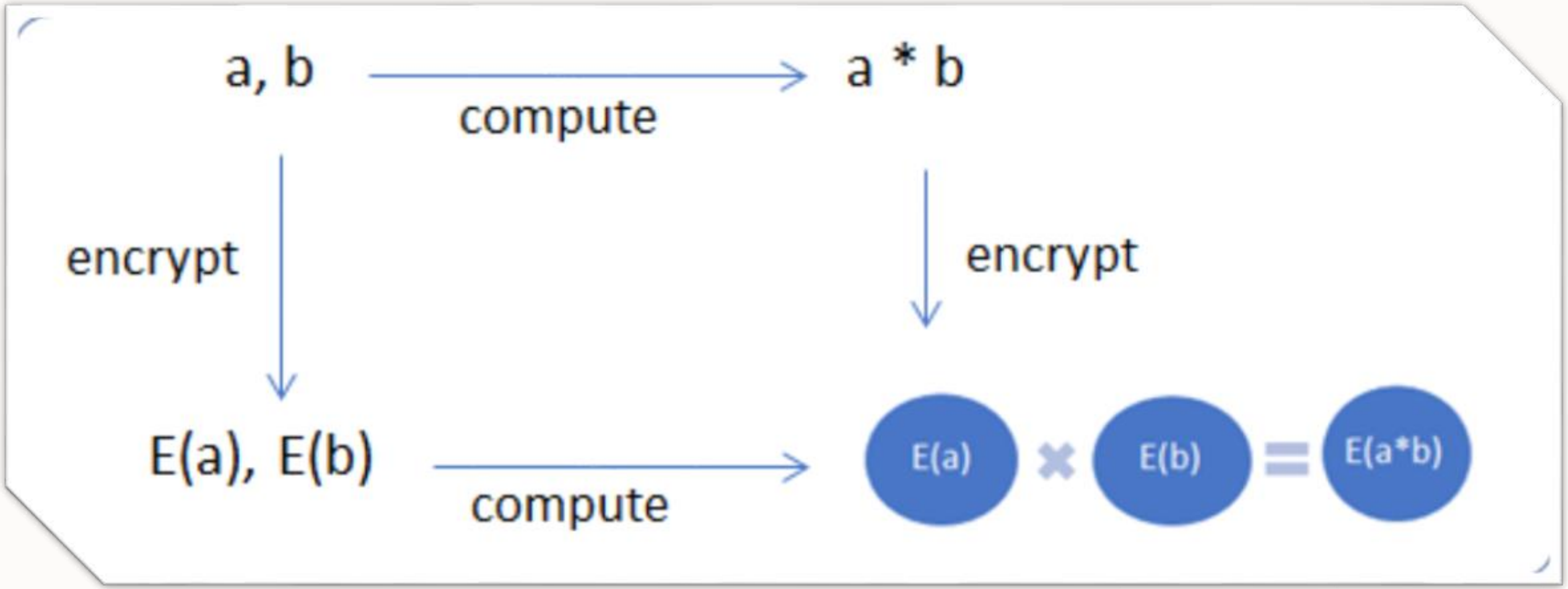
Need encryption which you can compute on:

Homomorphic Encryption!

Protecting Data via Encryption: Homomorphic Encryption

1. Put your gold in a locked box.
2. Keep the key.
3. Let your jeweler work on it through a glove box.
4. Unlock the box when the jeweler is done!

Homomorphic Encryption



Homomorphic Encryption (HE)

- Computation on encrypted data without decrypting it!
- 2009: Considered impractical
- 2011: Surprise breakthrough at Microsoft Research [BV11] [LNV 11]
- Practical encoding: 4 orders of magnitude speed-up
- 2016: **CryptoNets** evaluates neural net predictions on encrypted data!

AMERICAN Scientist

Alice and Bob in Cipherspace

data

Science Home News Journals Topics Careers

Call for Nominations: Sept. 1 - Dec. 1, 2017 The Kavli Prize honors scientists in Astrophysics, Nanoscience and Neuroscience. For more information, click here. THE KAVLI PRIZE

Log in | My account | Contact us

Become a member
Renew my subscription
Sign up for newsletters

SHARE

News from *Science* has introduced metered access. Full access to all news content is included in AAAS membership.

Genetic gold. Each spot in a DNA microarray, such sensitive genetic information.

How to Hide Your Ger

By Thomas Sumner | Feb. 16, 2014, 6:45 PM

CHICAGO, ILLINOIS—As the cost of g

ENHANCING EDUCATION GLOBAL OUTREACH POLICY & ADVOCACY PRESS & PUBLIC ENGAGEMENT SUPPORTING CAREERS Member Login

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

AAAS

NEWS JOURNALS MEMBERS CAREERS PROGRAMS GIVING EVENTS ABOUT SEARCH

Read More News Stories ›

New Technique Can Encrypt Genetic Data, Protecting Privacy While Assisting Biomedical Research

17 February 2014 Meagan Phelan

PNAS What do you want to read next? Discover more. See what We recommend... PNAS Alerts Receive an Email When New Content Publishes

Contact | Feedback | Submit | Subscribe Advanced Search »

Proceedings of the National Academy of Sciences of the United States of America PNAS

CURRENT ISSUE // ARCHIVE // NEWS & MULTIMEDIA // AUTHORS // ABOUT COLLECTED ARTICLES // BROWSE BY TOPIC // EARLY EDITION // FRONT MATTER

Current Issue > vol. 112 no. 28 > Robert Frederick, 8515–8516, doi: 10.1073/pnas.1507452112

Check for updates

Core Concept: Homomorphic encryption

Robert Frederick, *Science Writer*

Extract Full Text Authors & Info Figures Metrics Related Content PDF

It turned out to be pretty easy for Yaniv Erlich to identify people who had donated their genomic data to science, even despite researchers' attempts to make that data anonymous. Erlich, a former computer hacker hired by businesses to test the security of computer systems, only needed his tailor-made computer program and an Internet connection to access publicly available databases.

This Issue

July 14, 2015
vol. 112 no. 28
Masthead (PDF)
Table of Contents

PREV ARTICLE NEXT ARTICLE

Don't Miss

Thinking of submitting your next paper to PNAS? Learn tips in our PNAS tutorial videos.

Navigate This Article

Article Tools

High-level Idea

- **Encryption:**
  - Encryption adds noise to a “secret” inner product
  - Decryption subtracts the secret inner product and the noise becomes easy to cancel

- **Hard Problem:**
  - Hard problem is to “decode” noisy vectors
  - If you have a short basis, it is easy to decompose vectors

- **Homomorphic property:**
  - Lattice vectors $\rightarrow$ coefficients of polynomials
  - Polynomials can be added and multiplied

SEAL

- **Simple Encrypted Arithmetic Library (SEAL)**
  - Public release by Microsoft Research in 2015
  - Widely adopted by teams worldwide
  - OSS release of Microsoft SEAL 2018
- **Standardization of HE:**
  - November 2018

WIRED — To Protect Genetic Privacy, Encrypt Your DNA

TO PROTECT GENETIC PRIVACY, ENCRYPT YOUR DNA

CRUNCH NETWORK

Changing the security landscape for entrepreneurs

Posted Aug 17, 2017 by Bob Ackerman Jr.

The Register

Software

Microsoft researchers smash homomorphic encryption speed barrier

Here's how Apple, Google, and Microsoft are trying to get inside your genes

Daniela Hernandez

Not satisfied by having our emails, chats, status updates, search histories, clicking behaviors, and shopping preferences, some of Silicon Valley's most powerful tech titans are in an arms race to get access to your most personal information: your DNA.

http://sealcrypto.org

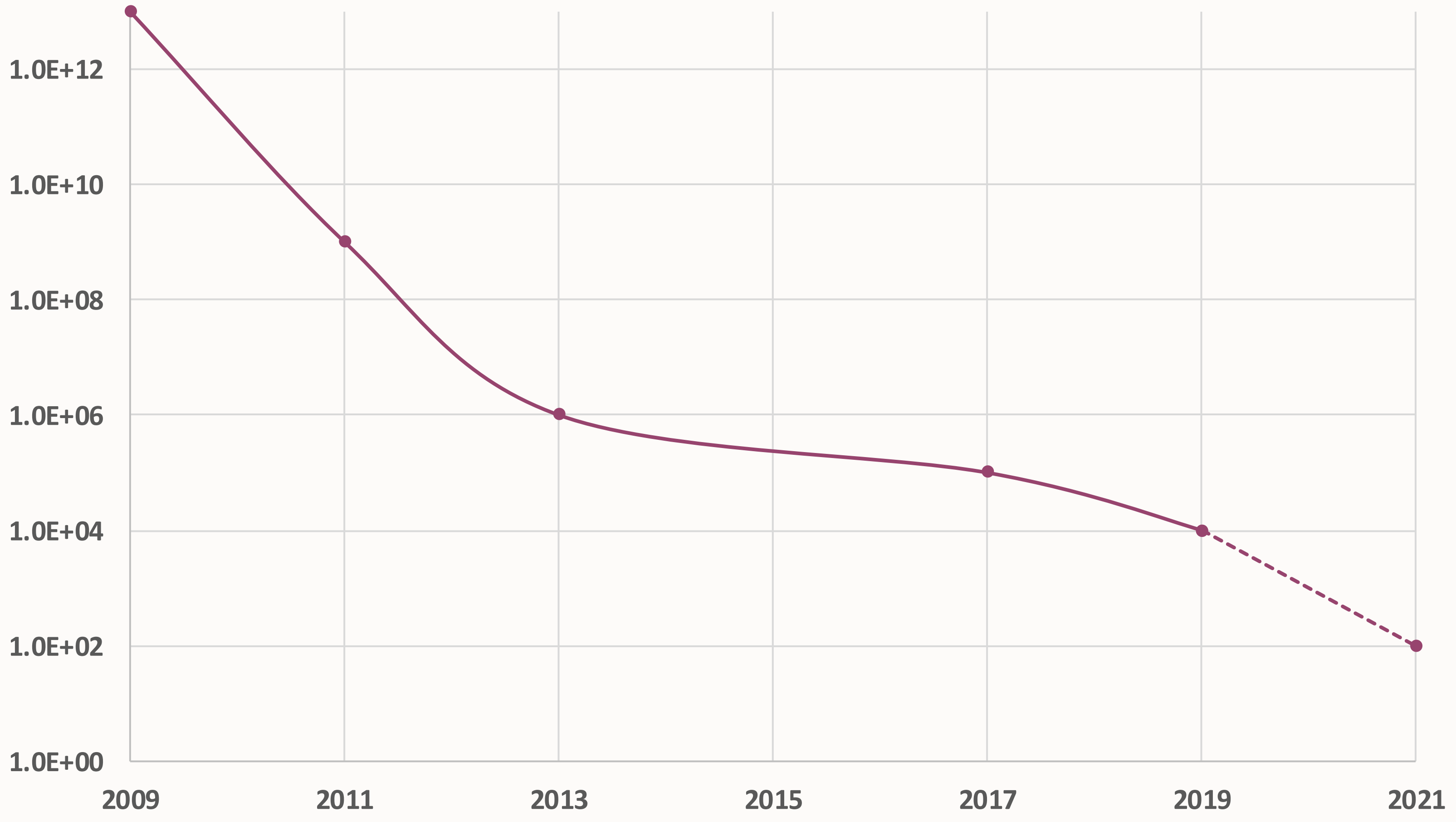
SEAL Team: KL, Kim Laine, Hao Chen, Wei Dai, Radames Cruz Moreno, Yongsoo Song, Ran Gilad-Bachrach, Shabnam Erfani, Sreekanth Kannepalli, Hamed Khanpour. Plus many interns…

Standardization: HomomorphicEncryption.org

- MSR launched in 2017
- Four workshops: Microsoft, MIT, U Toronto, Intel
- HES 1.0 Standard, November 2018
- Royal Society (PET) Report 2019

- Applications in regulated industries require standardization
- Standardization creates trust
- Open standards highly preferred in cryptography

Performance Improvement over Time (log scale!)



CPU GPU/FPGA/ASIC

Cloud Compute Scenarios for HE

1. Private Storage and Computation
2. Private AI Prediction Services
3. Hosted Private Training
4. Private Set Intersection
5. Secure Collaborative Computation

Markets: Healthcare, Pharma, Financial, Government, Insurance, Auto, Manufacturing, Oil &Gas,…

Private AI Applications

- Finance
  - Fraud Detection
  - Automated Claims Processing
  - Threat Intelligence
  - Data Analytics
- Healthcare
  - Medical Diagnosis
  - Medical Support (Healthcare Bots),
  - Preventive Care,
  - Data Analytics
- Manufacturing
  - Predictive Maintenance, Data Analytics

Private AI Demos History

- 2014: Heart attack risk, personal health data ~ 1 second
- 2015: CryptoNets demo showing neural net prediction: MNIST data set ~80 seconds
- 2016: Genomics predicting flowering time from 200K SNPs ~ 1 second
- 2016: Pneumonia mortality risk: intelligible models ~ 8 seconds for 4,000 predictions

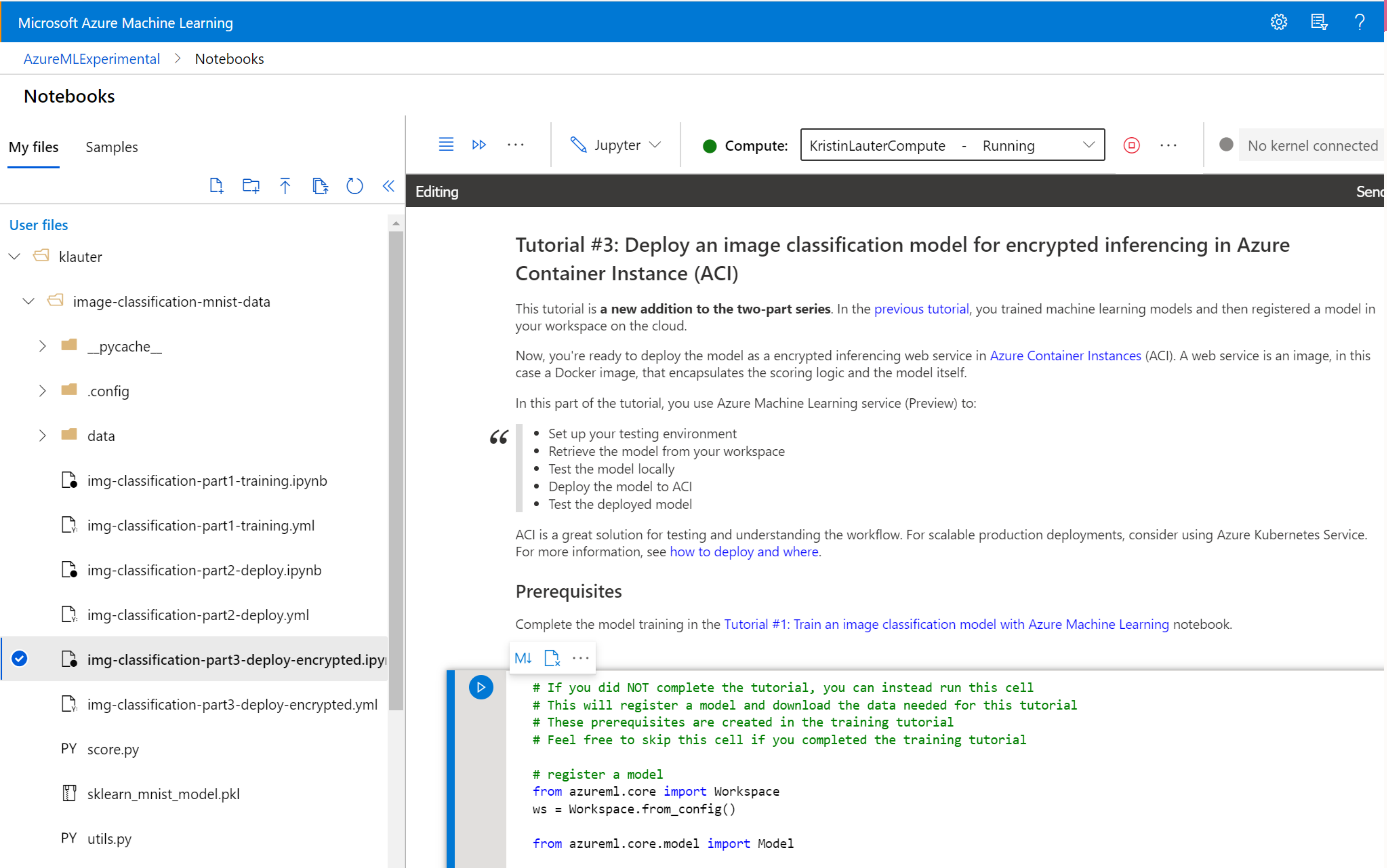
- 2018: Twitter sentiment analysis (150K text features) ~ less than a second
- 2018: cat/dog image classification ~ less than a second

- 2019: Asure Run (Private Fitness App)
- 2019: Chest Xray diagnostics
- 2019: Secure Weather prediction

Azure ML Private Prediction

- Released May 2020, announced at MS Build Conference
- Allows you to deploy an image classification model for encrypted inferencing in Azure Container Instance (ACI)
- Built on Microsoft SEAL
- Access the tutorial and documentation at:

https://ml.azure.com
https://aka.ms/SEALonAML


Microsoft Azure Machine Learning
AzureMLExperimental > Notebooks
Notebooks
My files
Samples
User files
Jupyter
Compute:
KristinLauterCompute - Running
No kernel connected
Editing
Send
M↓

- klauter
  - image-classification-mnist-data
    - __pycache__
    - .config
    - data
    - img-classification-part1-training.ipynb
    - img-classification-part1-training.yml
    - img-classification-part2-deploy.ipynb
    - img-classification-part2-deploy.yml
    - img-classification-part3-deploy-encrypted.ipy
    - img-classification-part3-deploy-encrypted.yml
    - score.py
    - sklearn_mnist_model.pkl
    - utils.py

Tutorial #3: Deploy an image classification model for encrypted inferencing in Azure Container Instance (ACI)

This tutorial is **a new addition to the two-part series**. In the previous tutorial, you trained machine learning models and then registered a model in your workspace on the cloud.

Now, you're ready to deploy the model as a encrypted inferencing web service in Azure Container Instances (ACI). A web service is an image, in this case a Docker image, that encapsulates the scoring logic and the model itself.

In this part of the tutorial, you use Azure Machine Learning service (Preview) to:

> - Set up your testing environment
> - Retrieve the model from your workspace
> - Test the model locally
> - Deploy the model to ACI
> - Test the deployed model

ACI is a great solution for testing and understanding the workflow. For scalable production deployments, consider using Azure Kubernetes Service. For more information, see how to deploy and where.

Prerequisites

Complete the model training in the Tutorial #1: Train an image classification model with Azure Machine Learning notebook.

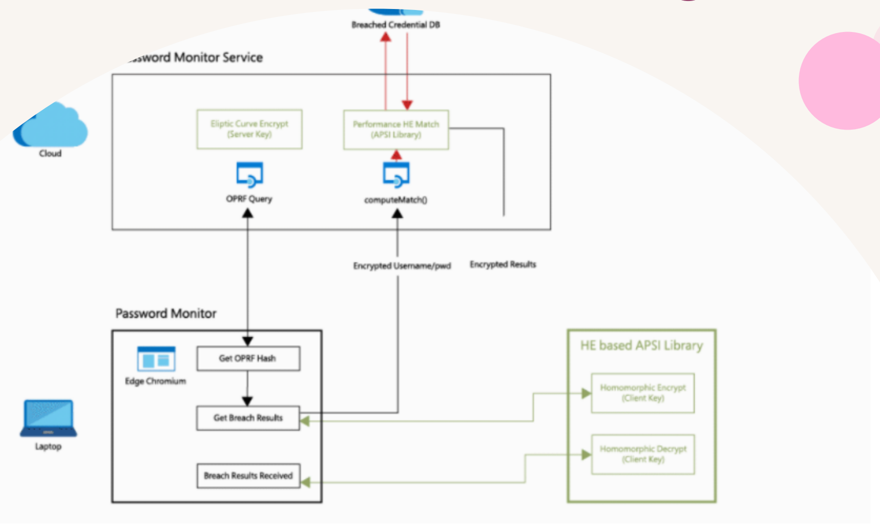
```
# If you did NOT complete the tutorial, you can instead run this cell
# This will register a model and download the data needed for this tutorial
# These prerequisites are created in the training tutorial
# Feel free to skip this cell if you completed the training tutorial

# register a model
from azureml.core import Workspace
ws = Workspace.from_config()

from azureml.core.model import Model
```

Password Breach Detection

[Password Monitor: Safeguarding passwords in Microsoft Edge - Microsoft Research]



One of the biggest pillars for Microsoft Edge is trust. Today, to further bolster that trust while keeping our customers safe, we introduce a new feature called Password Monitor. The feature notifies users if any of their saved passwords have been found in a third-party breach. All this is done while ensuring Microsoft doesn't learn the user's passwords. The underlying technology ensures privacy and security of the user's passwords, which means that neither Microsoft nor any other party can learn the user's passwords while they are being monitored.

his unique security feature is possible due to pioneering cryptography research technology incubation done here at Microsoft Research. The feature is a ation of our research on *homomorphic encryption* and its practical ns. It is the result of a collaboration between former research inc ptography and Privacy Research Group, and Edge produ ilt on the Microsoft SEAL homomorphic encry to bring Password Monitor to

Packages and Libraries

- GitHub - microsoft/SEAL: Microsoft SEAL is an easy-to-use and powerful homomorphic encryption library.
- GitHub - OpenMined/TenSEAL: A library for doing homomorphic encryption operations on tensors
- GitHub - Lab41/PySEAL: This repository is a fork of Microsoft Research's homomorphic encryption implementation, the Simple Encrypted Arithmetic Library (SEAL). This code wraps the SEAL build in a docker container and provides Python API's to the encryption library.
- GitHub - tf-encrypted/tf-seal: Bridge between TensorFlow and the Microsoft SEAL homomorphic encryption library

Resources:

- Introduction – Homomorphic Encryption Standardization
- Homomorphic Encryption - Microsoft Research

Ongoing research and challenges

1. Improvements to CryptoNets:
   - Classification and training deep neural nets on encrypted data
   - Many other ML models and tasks
2. Hardware acceleration
   - GPU acceleration (cuHE)
   - FPGA acceleration (HEAX)
   - DARPA DPRIVE Program: Intel/Microsoft
3. Reducing overhead
   - Storage blow-up
   - Energy costs
4. Improving functionality
   - Combining schemes for logical and arithmetic operations

Recent work

Homomorphic encryption security standard. Technical report, HomomorphicEncryption.org, Toronto, Canada, November 2018. https://eprint.iacr.org/2019/939

CHET: an optimizing compiler for fully homomorphic neural-network inferencing. In Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation, pages 142–156. ACM, 2019.

XONN: Xnor-based oblivious deep neural network inference. In 28th USENIX Security Symposium (USENIX Security 19), pages 1501–1518, Santa Clara, CA, August 2019.

HEAR: Human Action Recognition via Neural Networks on Homomorphically Encrypted Data, Miran Kim, Xiaoqian Jiang, Kristin Lauter, Elkhan Ismayilzada, Shayan Shams https://arxiv.org/pdf/2104.09164.pdf

THANK YOU!

Mathematics of Homomorphic Encryption

- **New hard problems** proposed (2004-2013)
  - ❑ Small Principal Ideal Problem, Approximate GCD,
  - ❑ Learning With Errors (LWE), Ring-Learning With Errors (RLWE)
  - ❑ Related to well-known hard lattice problems:

- **Lattice-based Cryptography:**
  - ❑ Proposed by Hoffstein, Pipher, and Silverman in 1996 (NTRU), Aijtai-Dwork
  - ❑ Compare to other public key systems: RSA (1975), ECC (1985), Pairings (2000)

- **Hard Lattice Problems:**
  - ❑ Approximate Shortest Vector Problem (SVP), Bounded Distance Decoding
  - ❑ 30 year history of Lattice-basis reduction (LLL, BKZ, BKZ 2.0, FpLLL, sieving, challenges)

- **Security:**
  - ❑ Best attacks take exponential time
  - ❑ Secure against quantum attacks (so far…)