Move Away HIPAA & GDPR, Here Comes CrypTFlow – Running AI without Data Sharing

Arjun Soin, Pratik Bhatu, Rohit Takhar, Nishanth Chandran, Divya Gupta, Javier Alvarez-Valle, Rahul Sharma, Vidur Mahajan and Matthew P Lungren

Microsoft Research | Stanford AIMI | CARING

TEACHING ACTIVITY

At the conclusion of this activity, participants will be able to:

- Understand the basics of privacy as it pertains to AI model inferencing
- Learn the basics of encryption and secure multi-party computation, especially the CrypTFlow package
- See some early experimental results

BACKGROUND

Currently, running Artificial Intelligence (AI) algorithms on medical images requires either the sharing of medical images with developers of the algorithms, or sharing of algorithms with the hospitals. Both these options are sub-optimal since there is always a real risk of patient privacy breach or intellectual property (IP) theft. We elaborate upon a novel secure multi-party computation methodology which enables the running (and thereby validation) of AI algorithms without the need for sharing of data or IP.

THE BASICS OF ENCRYPTION AND MPC

What is Encryption?

Encryption is the process of converting data into a "secret code" which can only be de-coded using a "secret key".

What is the problem with encryption?

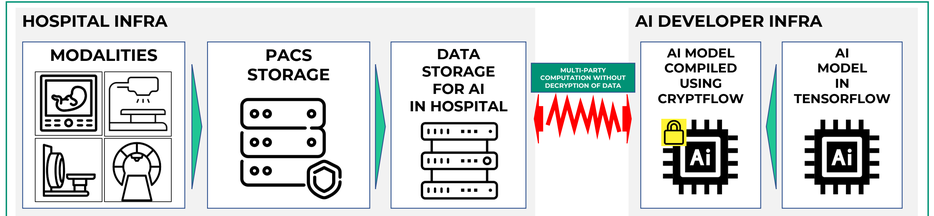
When running AI algorithms on medical data, today, it is essential to share the "secret key" with the AI developer, which makes this process insecure.

What is Secure Multi-Party Computation?

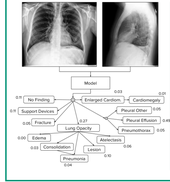
Secure Multi-Party Computation is a subfield of cryptography with the goal of creating methods for parties to jointly compute a function over their inputs while keeping those inputs private. Essentially, it means that a mathematical operation can be performed on data, in a way where none of the involved parties find out what the data is, or what the mathematical operation is.

What is CrypTFlow?

CrypTFlow is a system that converts the inference code of AI algorithms created with TensorFlow into secure Multi-Party Computation protocols at the push of a button. In a nutshell, it allows an AI algorithm to run on medical data without having to share the data with the AI model developer, or having to share the AI model details with the owner/keeper of the medical data



EARLY EXPERIMENTS WITH CRYPTFLOW



Data	Finding	Performance	Time
CheXpert CARING500	CrypTFlow-based model inference **matches** native deployment	Both pipelines yield the **same AUROC** (performance) & **Brier Scores** (calibration)	Per X-ray image secure inference takes **~900s** while insecure inference takes **~0.3s**

Contact: asoin@stanford.edu, rahsha@microsoft.com, rohit.takhar@caring-research.com