The *COFE* Ecosystem

GaNDLF

Generally Nuanced Deep Learning Framework

gandlf.org

ML Commons

S. Pati, et al., *Nature Communications Engineering*, 2(23), 2023

OpenVINO

Model optimization for inference on low-resource environments

openvino.ai

intel

A. Demidovskij, et al.; *ICCV Workshop*, 783-787, 2019

HF Hub

Model deployment across multiple platforms & ecosystems

hf.co

S.M. Jain, *Introduction to Transformers for NLP*, 51-67, Berkeley, 2022

OpenFL

Federated Learning Library

openfl.io

P. Foley, et al., *Phys Med Biol (ITCR Special Issue)*, 67(21), 214001, 2022

MedPerf

Governance & Orchestration

medperf.org

ML Commons

A. Karargyris, et al.; *Nature Machine Intelligence* 5:799-810, 2023

MedPerf

An Open Benchmarking Platform for Medical Artificial Intelligence

Hasan Kassem

Software Engineer
MLCommons

ML Commons

MLCommons is a global community

Founding Members

Members

Academics from educational institutions including:

Harvard University
Indiana University
Polytechnique Montreal
Peng Cheng Laboratory
Stanford University
University of California, Berkeley
University of Toronto
University of Tübingen
University of York, United Kingdom
Yonsei University

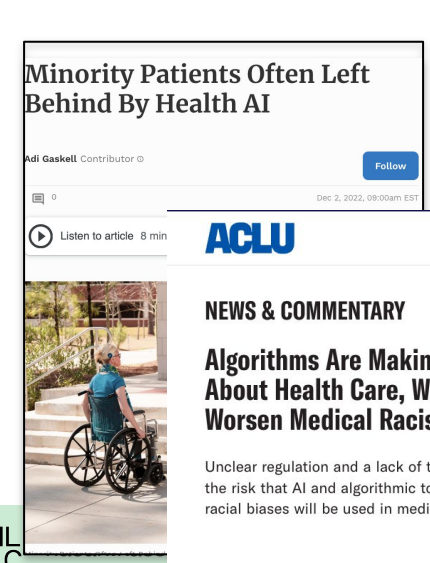
MLCommons' Medical working group

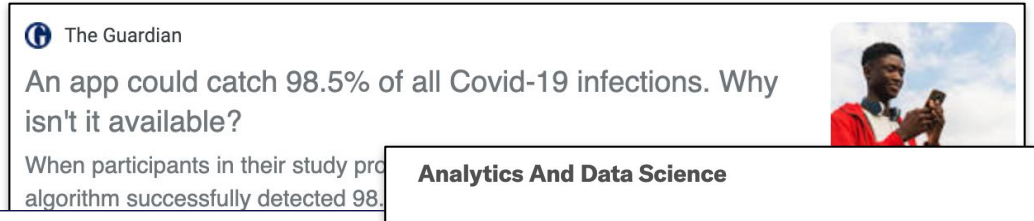
Comprised of professionals from 20+ companies, 20+ academic institutions and 10 hospitals, across 13 countries and 5 continents

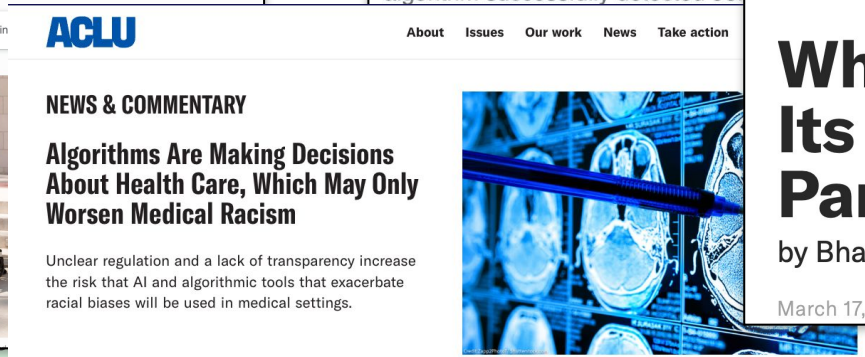
A*STAR, Singapore
Amazon, Seattle, WA
Brigham and Women's Hospital, Boston, MA
Broad Institute of MIT and Harvard, Cambridge, MA
Cisco, San Jose, CA
cKnowledge, Paris, France
Dana-Farber Cancer Institute, Boston, MA
Fast.ai, San Francisco, CA
Flower Labs, Hamburg Germany
Fondazione Policlinico A. Gemelli, Rome, Italy
German Cancer Research Center, Heidelberg, Germany
Google, Mountain View, CA
Harvard Medical School, Boston, MA
Harvard T.H. Chan School of Public Health, Boston, MA
Harvard University, Cambridge, MA
Hugging Face, New York, NY
IBM Research, San Jose, CA
IHU Strasbourg, Strasbourg, France
Intel, Santa Clara, CA
John Snow Labs, Lewes, DE
Landing.AI, Palo Alto, CA
Lawrence Livermore National Laboratory, Livermore, CA
Meta, Menlo Park, CA
Microsoft, Redmond, WA
MIT, Cambridge, MA
MLCommons, San Francisco, CA
Nutanix, San Jose, CA
NVIDIA, Santa Clara, CA
OctoML, Seattle, WA
Perelman School of Medicine, Philadelphia, PA
Red Hat, Raleigh, NC
Rutgers University, New Brunswick, NJ
Sage Bionetworks, Seattle, WA
Stanford University School of Medicine, Stanford, CA
Stanford University, Stanford, CA
Supermicro, San Jose, CA
University of Cambridge, Cambridge, UK
University of Heidelberg, Heidelberg, Germany
University of Pennsylvania, Philadelphia, PA
University of Queensland, Brisbane, Australia
University of Strasbourg, Strasbourg, France
University of Toronto, Toronto, Canada
University of Trento, Trento, Italy
University of York, York, UK
Vector Institute, Toronto, Canada
Weill Cornell Medicine, New York, NY
Tata Medical Center, Kolkata, India

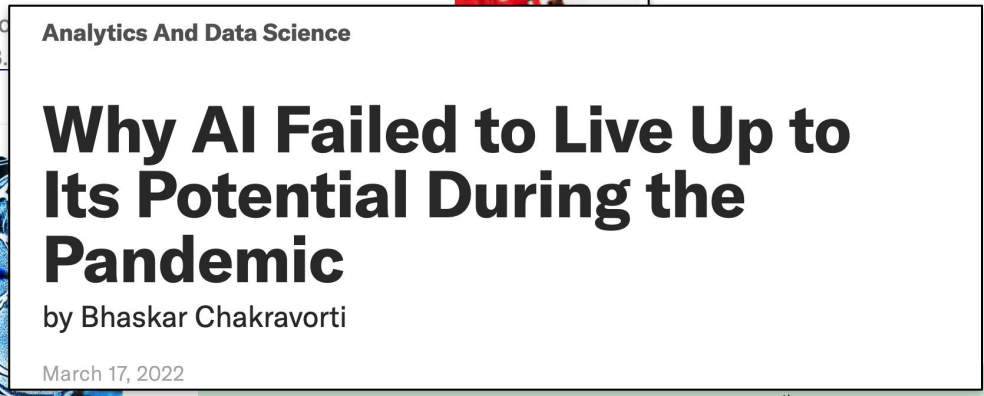
Why real-world validation?

- AI models necessitate more extensive validation
- AI models need diverse data, but healthcare data sharing is difficult.



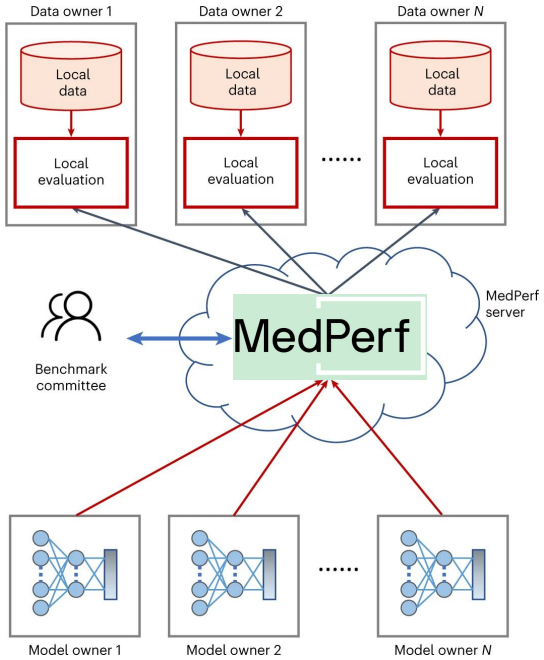






Our approach to medical AI benchmarking

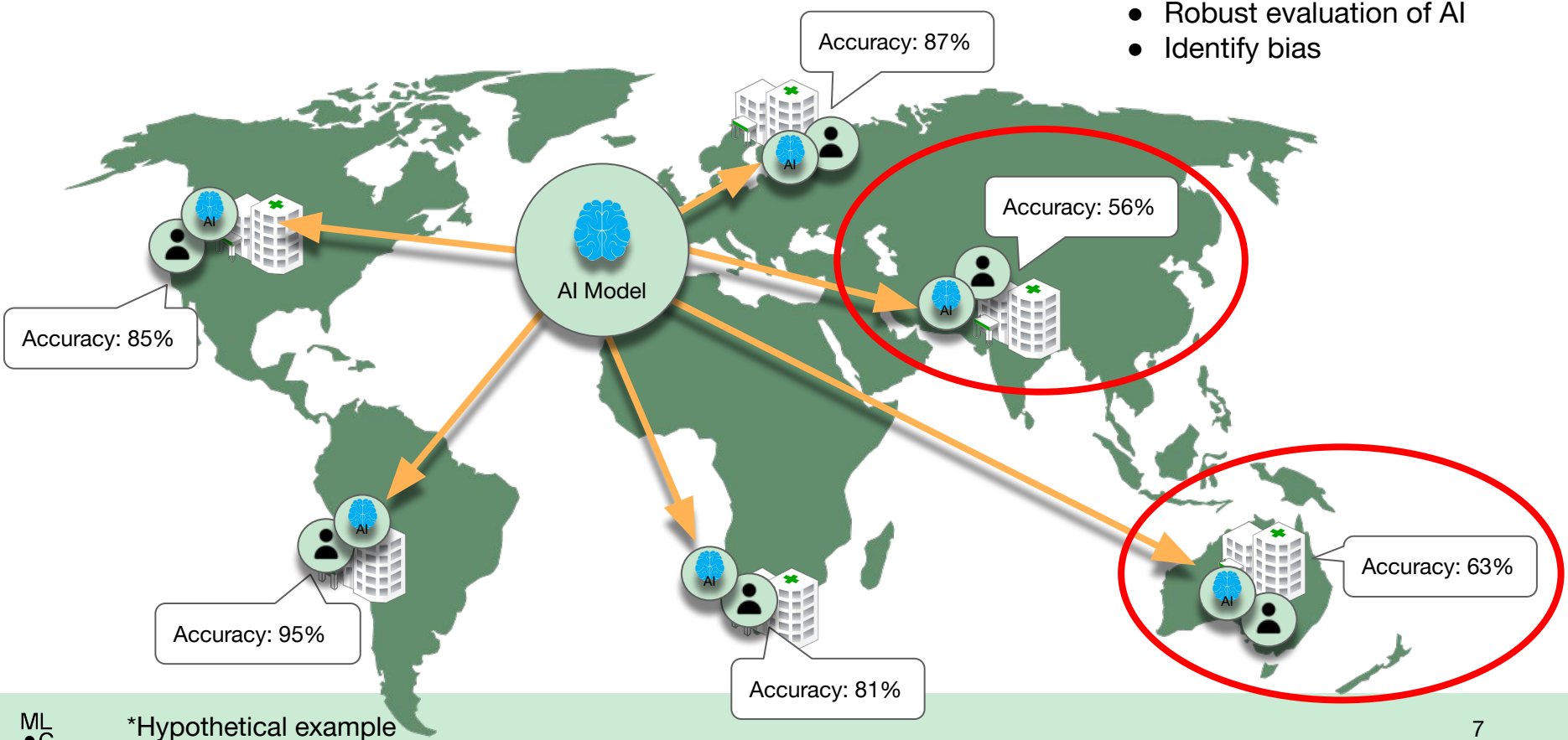
- MedPerf is an open community software platform that evaluates AI models on diverse real-world data for clinical impact
- Based on federated evaluation of AI
  - Driven by **stakeholders** (aka **benchmark committee)**
  - Runs on **real-world data**
  - Patient data is **never** shared. Models are remotely deployed and evaluated within the **premises of data providers** (i.e., hospitals)
  - Approach alleviates **data privacy concerns**
- Integrates human accountability and transparency from experts in the loop



How it works*

Real-world diverse data

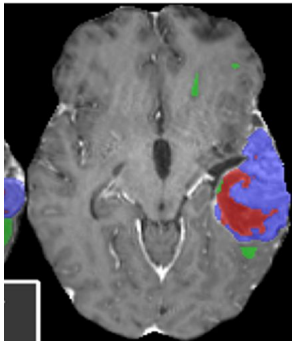
- Different patients and different clinical sites
- Robust evaluation of AI
- Identify bias



*Hypothetical example

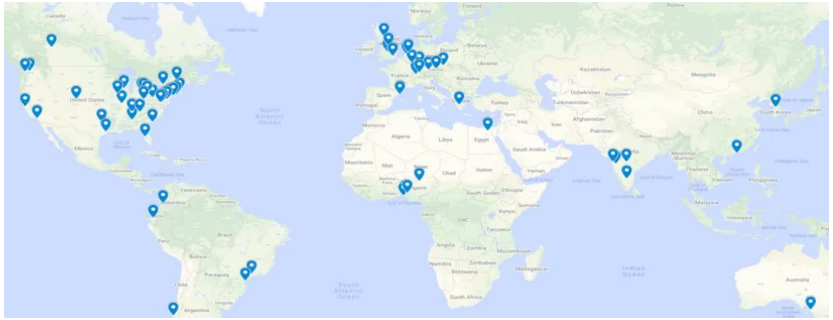
MedPerf Use Cases

- MedPerf team includes the FeTS/OpenFL researchers that ran a 71 hospital federated training and evaluation experiment in 2021[1]
- Currently supporting neuro-oncology federation of ~100 hospitals, led by Indiana University, Duke University and the Response Assessment for Neuro-Oncology Working Group [2]



SCHOOL OF MEDICINE

Duke

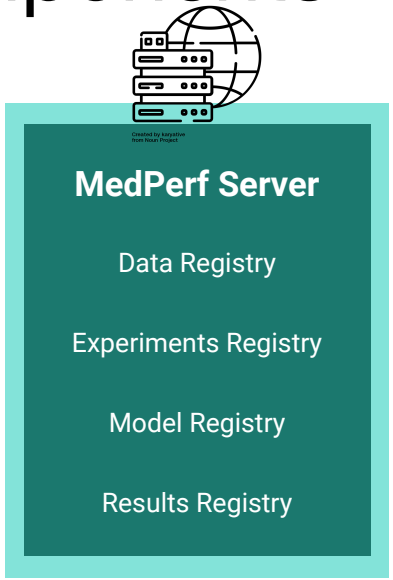


[1] https://www.nature.com/articles/s41467-022-33407-5
[2[https://fets-ai.github.io/FL-PoST/

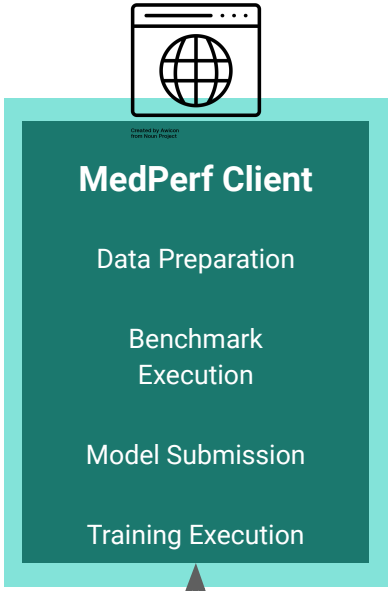
MedPerf

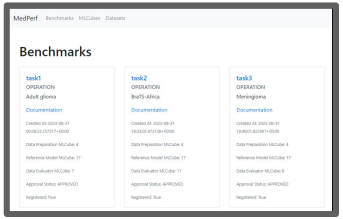
Technical Description

Core Components



Submit/Modify/Retrieve
AI workloads metadata

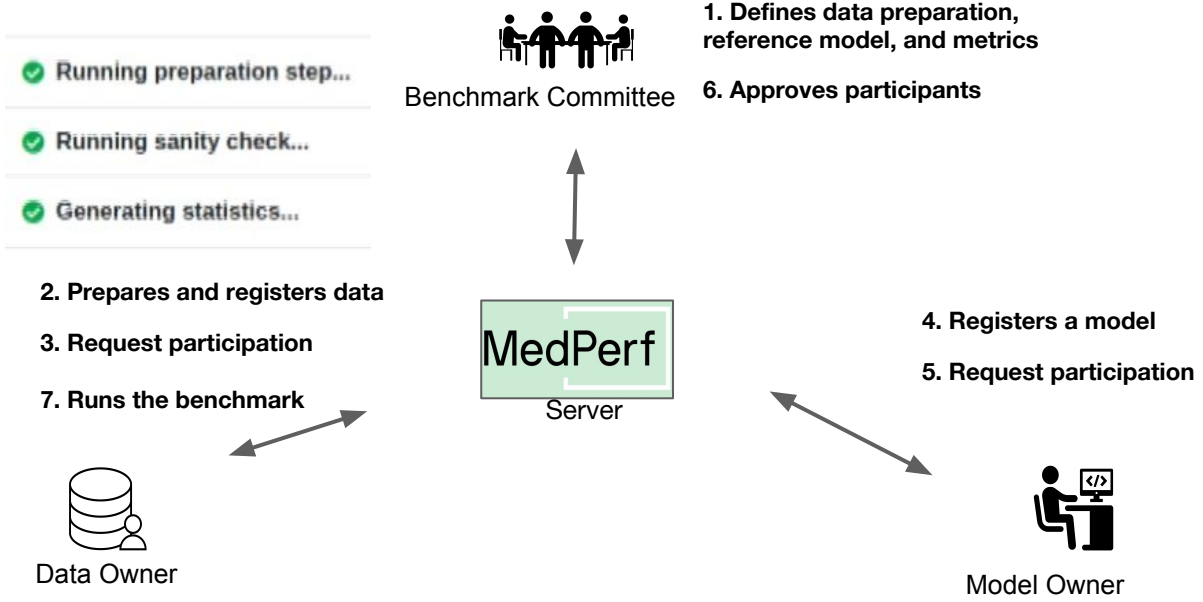




Web UI

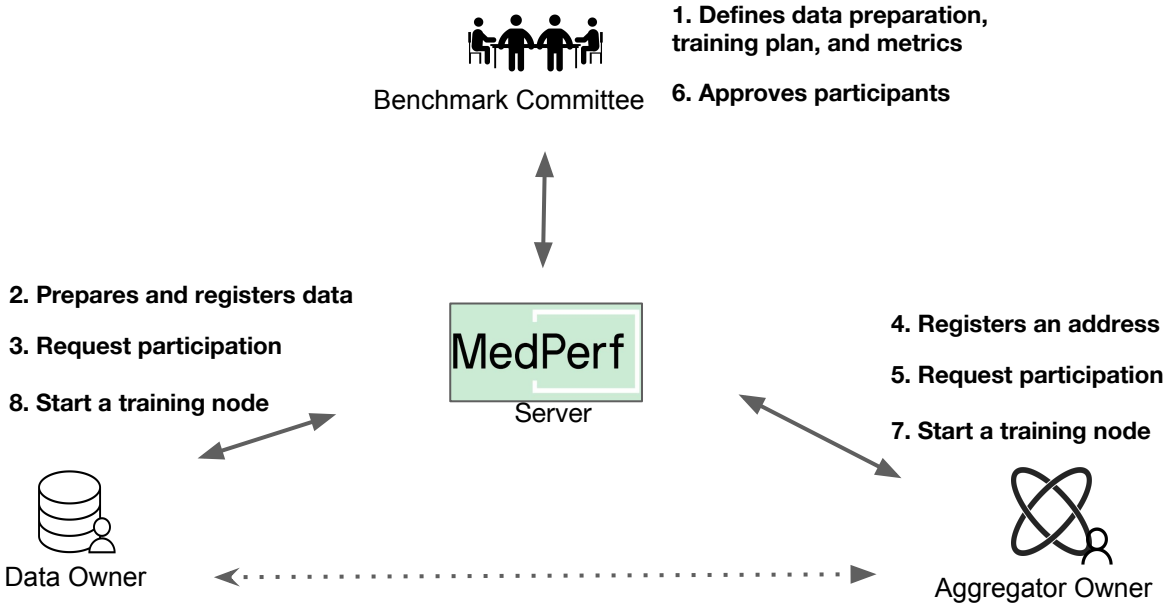
Evaluation Workflow

- MedPerf is an open community software platform that evaluates AI models on diverse real-world data for clinical impact



Training Workflow

- MedPerf has been extended to orchestrate federated training experiments



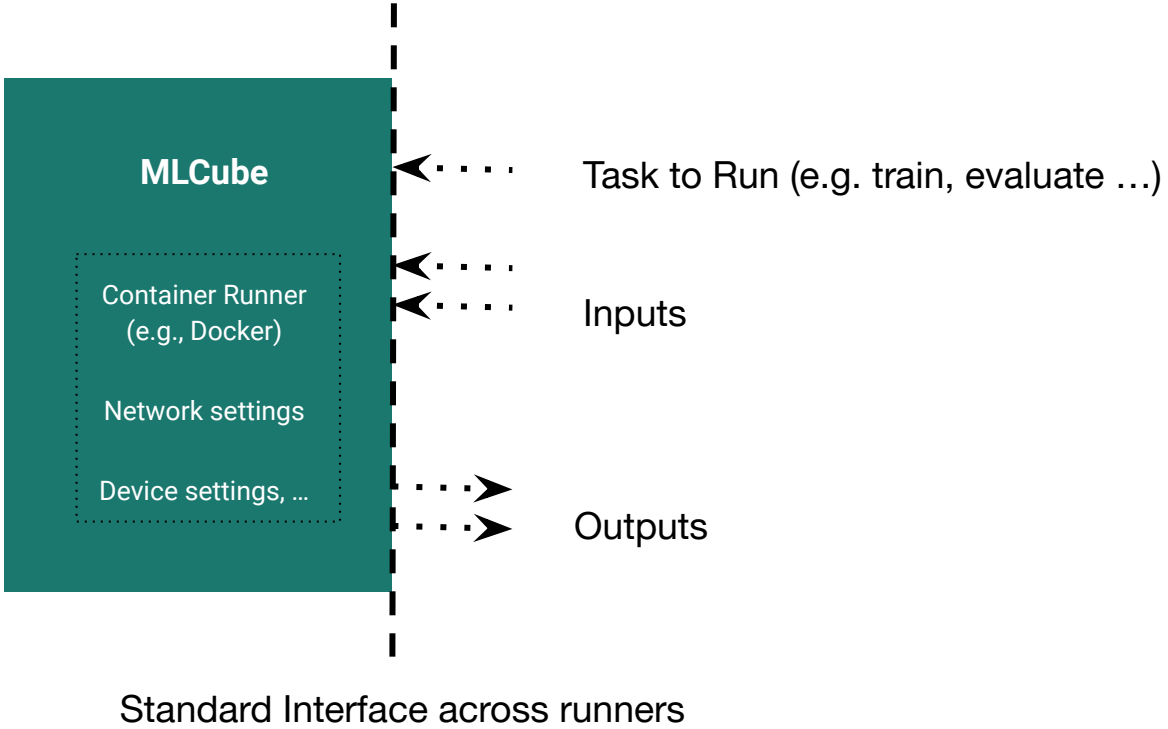
MedPerf

The OpenFL-GaNDLF MLCube

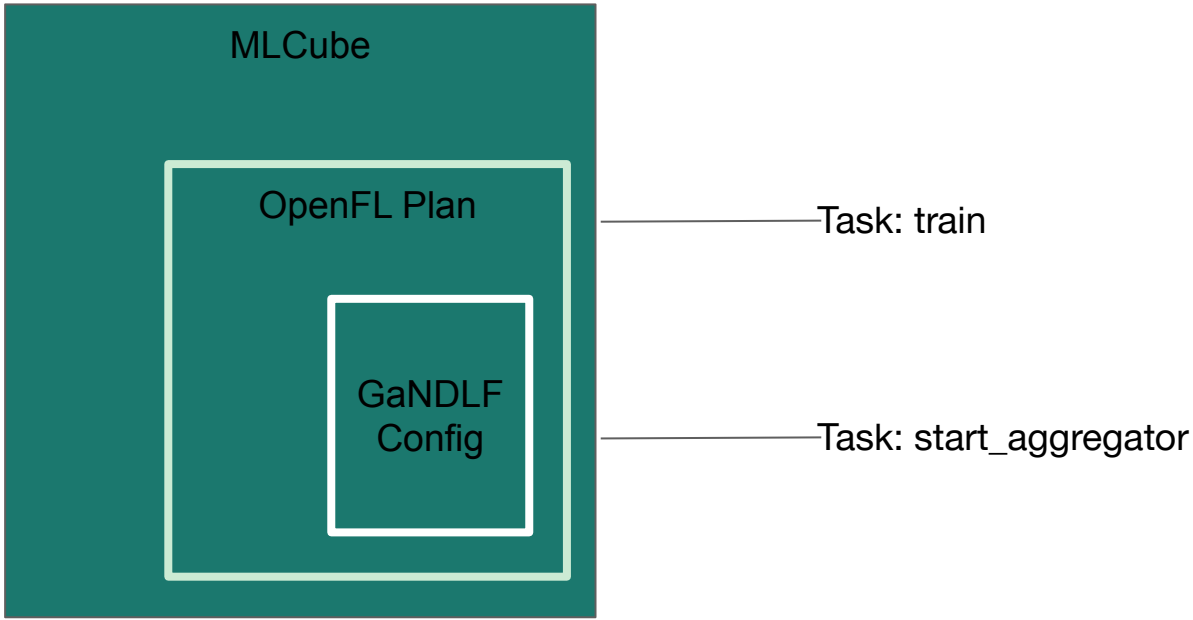
What is an MLCube?

Benchmark Committee

1. Defines data preparation, reference model, and metrics



The OpenFL-GaNDLF MLCube



MedPerf

What's Coming Next?

MedPerf

Web UI

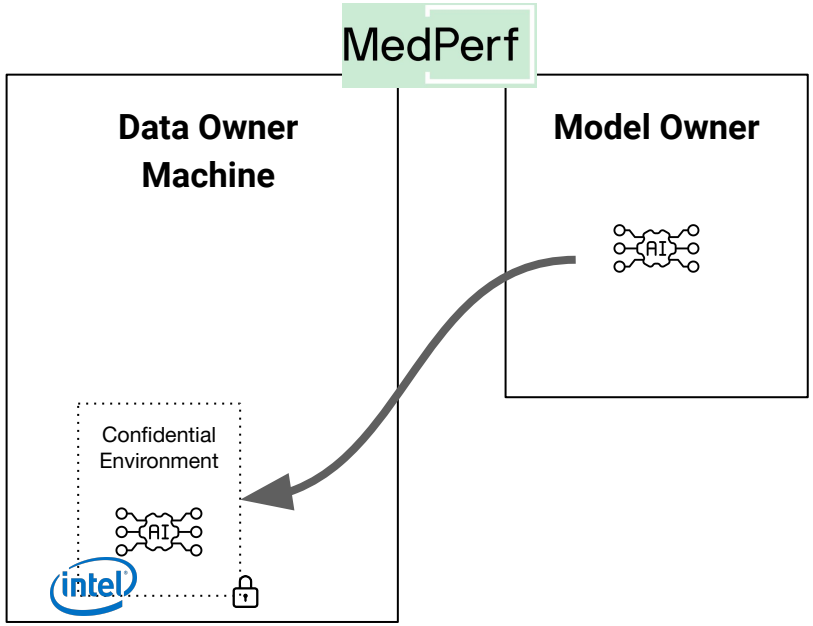
MedPerf

Security Features

Confidential Compute

- Run MedPerf workloads in a secure environment on the data owner machine.
  - IP is protected. No need to trust data owners.

- Our partners are helping us support running workloads with confidential compute.



Data Policies

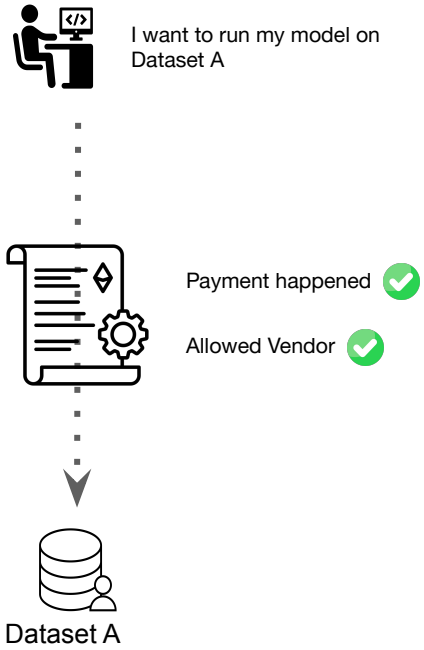
Integrating smart contracts to enforce policies in benchmarks between stakeholders

Value:

- Policies are designed and co-developed with community.
- Automation of policy evaluation
- With SGX, no one can tamper with the contract execution.

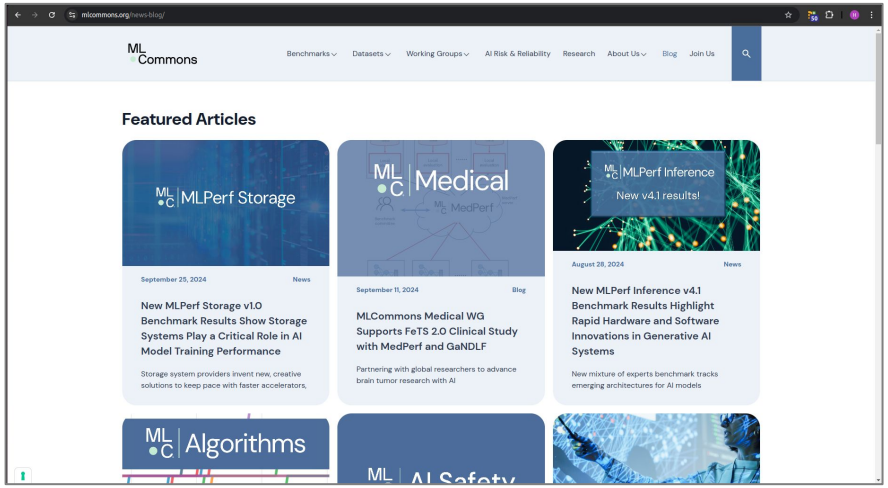
Examples:

- Which AI vendor can run a model on a certain dataset
- How many times a model is executed
- Membership agreements between stakeholders



We welcome people who want to make ML better.

- Join our mailing list
- Attend community events
- Become a member (free for academics)
- Participate in working groups



Join us at **mlcommons.org**

Visit medperf.org

Email us at **medperf-hello@mlcommons.org**

Try the hands-on session at **https://fl-tutorials.org/**