Privacy-Enhancing Technologies

PRIZE CHALLENGES

Pandemic Response and Forecasting Technical Brief

Transforming Pandemic Response and Forecasting through Federated Learning with End-to-End Privacy

Background

Federated learning (FL), or more generally collaborative learning, shows huge promise for machine learning applications derived from sensitive data. FL enables training on distributed datasets without raw data being shared amongst the participating parties.

The goal of this prize challenge is to mature federated learning approaches and build trust in adoption by accelerating the development of efficient privacy-preserving federated learning solutions that leverage a combination of input and output privacy techniques to:

- Drive innovation in the technological development and application of novel privacy-enhancing technologies
- Deliver strong privacy guarantees against a set of common threats and privacy attacks
- Develop technology that is capable of generating effective models for predicting the risk of infectious disease for individuals over a period of time.

This challenge use case is focused on enhancing cross-organization, cross-border data access to support efforts to improve public health-related forecasting in order to bolster pandemic response capabilities. Participants are asked to develop innovative, privacy-preserving FL solutions to enable forecasting of infection risk for individuals in the context of a pandemic.

The COVID-19 pandemic has taken an immense toll on human lives and has had an unprecedented level of socioeconomic impact on individuals and societies around the world. The pandemic has highlighted the importance of developing better methods for harnessing the value of data, whilst protecting the privacy of sensitive information about individuals and groups. Such privacy-preserving data sharing and analytics can be beneficial for managing the ongoing COVID-19 pandemic, and will be critical for preparing for future pandemics or other public health emergencies. The scale of the challenge is clear: the World Health Organization estimates there were 14.9 million excess deaths associated with COVID-19 globally in 2020 and 2021. Developing tools to better understand the evolution of future pandemics could therefore save countless lives.

For the purposes of this Challenge, a privacy-preserving solution is defined as one which is able to ensure that sensitive attributes in the datasets remain confidential to the respective data owners across the machine learning lifecycle. This requires that access to the raw data is protected (input privacy), and that sensitive information cannot be reverse-engineered during model training or inference (output privacy).

This is a high-impact and exciting use case for novel privacy-enhancing technologies. Forecasting individual-level risk during a pandemic is challenging as it depends on multiple factors including the prevalence of the disease in the community, interactions the individual has had with those already infected, and sociodemographic and health-related attributes that may make an individual more susceptible to infection. There are currently challenging trade-offs between enabling sufficient access to data to build tools for effectively modeling infection risk, whilst ensuring that sensitive health and mobility data are kept confidential.

The use case has been based on real world experiences of challenges of public health data sharing. However, it is worth emphasizing that this is a model exercise designed primarily to generate novel ideas and solutions for protecting privacy, not to fully replicate real world scenarios. Deploying analytical approaches such as this in the real world requires a broad set of legal and ethical considerations, including assessments of trade-offs between different ethical imperatives, and careful assessment of public attitudes towards data use. Such considerations are outside of the primary scope of this prize challenge. The intent is that successful privacy solutions developed through this challenge can unlock new opportunities for effective, ethical data use by reducing the privacy impacts of analysis across large federated datasets.

Therefore, though novel innovation for this use case alone could achieve significant real-world impact, the challenge is designed to incentivize development of privacy technologies that can be applied to other use cases where data is distributed across multiple organizations or jurisdictions, both in public health and elsewhere. The best solutions will deliver meaningful innovation towards deployable solutions in this space, with consideration of how to evidence the privacy guarantees offered to data owners and regulators, but also have the potential to generalize to other situations.

FL produces a global model that aggregates local models obtained from distributed parties. As data from each participating party does not need to be shared, the approach provides a baseline level of privacy. However, privacy vulnerabilities exist across the FL lifecycle. For example, as the global federated model is trained, the parameters related to the local models could be used to learn about the sensitive information contained in the training data of each client. Similarly, the released global model could also be used to infer sensitive information about the training datasets. Protecting privacy across the FL pipeline requires a combination of privacy-enhancing technologies and techniques that can be deployed efficiently and effectively to preserve privacy while still producing ML models with high accuracy and utility. The core of this challenge is to develop FL approaches that provide such end-to-end privacy, in accordance with the privacy threat profile detailed in the *Privacy Threat Profile* section of this document.

Structure

The Challenge is split into three phases:

- **Phase 1: White paper** (also referred to as a Concept Paper, with the two terms used interchangeably). You will develop a technical white paper that describes your proposed approach
- **Phase 2: Solution development.** You will build and develop the solution proposed in your white paper
- **Phase 3: Red Teaming.** The top solutions will be tested by competing red teams.

Further details about the phases are provided below, and on the challenge website.

A range of support and opportunities will be provided to participants during the Challenge:

- Funding (see separate UK and US details on challenge website)
- Opportunities to engage with data protection regulators from public sector organizations operating in public health, namely:
  - UK Information Commissioner's Office (ICO)
  - Data and Analytics Research Environments UK (DARE UK)
  - NHS England
  - Staff of the US Centers for Disease Control (CDC)
- Technical support and guidance from the University of Virginia's Biocomplexity Institute (UVA-BI), including workshops during Phase 1 of the challenge detailing the use case and how to work with the synthetic population dataset.

The organizers plan to offer opportunities to showcase the best solutions in front of a global audience at the second Summit for Democracy, to be convened by President Joe Biden, in the first half of 2023.

The Challenge

Objective

The objective of the Challenge is to ***develop a privacy-preserving federated learning (PPFL)*** solution that is capable of training a model that predicts infection risk for individuals in a pandemic, while providing a demonstrable level of privacy against the defined threat profile.

This PPFL solution should aim to:

- Provide robust privacy protection for the collaborating parties
- Minimize loss of accuracy in the model, as compared to a centralized model
- Minimize additional computational resources (including CPU, memory, communication), as compared to a centralized model.

In addition to this, the evaluation process will reward participants who:

- Display a high degree of innovation
- Demonstrate how their solution (or parts of it) could be applied or generalized to other use cases
- Effectively prove or demonstrate the privacy guarantees offered by their solution, in a form that is comprehensible to data owners and regulators
- Consider how their solution, or a future version of it, could be applied in a production environment.

Expertise in public health and infectious disease is not a prerequisite for entering the challenge and the assessment process will not focus on detailed understanding of the use case itself.

Datasets

Organizers will provide synthetic population datasets to participants via a secure method. The data includes:

- a social contact network, capturing when and where any two people come into contact, and the duration of the contact
- demographic attributes of individuals
- observations of individuals' health state (i.e., whether they are infected or not).

Two synthetic population datasets are provided: the first covers the population of the UK, and the second the population of the state of Virginia, USA. These datasets have been generated by the UVA-BI team using an outbreak simulation that created 63 days-worth of data. Table 1 below provides details of the approximate sizes of the resulting datasets.

	Virginia dataset	UK dataset
Population size	7.7 million	62 million
Number of social contacts	181 million	722 million
Number of disease state records (upper bound based on 1 reading per individual per day)	430 million	3.5 billion

Table 1: details about the size of the synthetic datasets

Participants will be provided both datasets during the development phase. The two datasets have identical structures and schemas (detailed below), but differ in size and scale. During phase 2 evaluation, submissions to the US challenge will be evaluated against a sequestered Virginia dataset, and submissions to the UK challenge will be evaluated against a sequestered UK dataset.

Note: The challenges are based on synthetic data to minimize the security burden placed on participants during the development phase; of course, the intent of the challenge is that privacy solutions are developed that would be appropriate for use on real datasets with demonstrable privacy guarantees. However, participants must adhere to a data use agreement (see Annex A).

Dataset structure and schema

Each synthetic population dataset can be considered as a relational database consisting of eight tables. These tables are described in Table 2 below. Tables 3-10 describe the fields that are present in each of the tables. Each of the eight tables will be provided to participants as a separate CSV file.

Table 2: Various records of the synthetic population dataset		
	Data Record	**Description**
1	Person data	Information about individuals in the dataset
2	Household data	Information about individual households
3	Residence location data	Location information about the residence
4	Activity location data	Information about activity and where it took place
5	Activity location assignment data	Information about which activity a person was involved in and timing information. Activities are repeated every day, as if it were like the film *Groundhog Day*
6	Population contact network data	Information about contact network including time and duration. This table is generated from the Activity location assignment data table, and also repeats every day.
7	Disease Outcome data (training)	Information about the health status on a particular day
8	Disease Outcome data (test)	Information about the health status

Table 3: Person File

Field	Description
Household ID (hid)	An integer identifying a household
Person ID (pid)	A unique integer identifying the owner of the house
Person number (person_number)	Sequence ID of a person in the household. Household of size 3 has people with person_numbers 1, 2, and 3.
Age (age)	Age of person
Sex (sex)	Indicating the gender of Person

Table 4: Household File

Field	Description
Household ID (hid)	A unique integer identifying the household
Residence ID (rlid)	An integer identifying the residence
Admin 1	ADCW[1] ID for admin1 region (for UK) or state FIPS code (Virginia)
Admin 2	ADCW ID for admin2 region; Equals to admin1 if for UK; 3 digit county FIPS code (Virginia)
Household Size (hh_size)	Number of persons in family

Table 5: Residence Locations File

Field	Description
Residence ID (rlid)	A unique integer identifying the residence
Longitude	the longitude of the location
Latitude	the latitude of the location
Admin 1	See household file description
Admin 2	See household file description

Table 6: Activity Locations File

Field	Description
Activity location ID (alid)	Unique integer identifying the location where activity took place
Longitude	Longitude of the location
Latitude	Latitude of the location
Admin 1	See household file description
Admin 2	See household file description
Work	Does location support work activity (Value is 0 or 1)
Shopping	Does location support shopping activity (Value is 0 or 1)
School	Does location support school activity (Value is 0 or 1)
Other	Does location support other activity (Value is 0 or 1)
College	Does location support college activity (Value is 0 or 1)
Religion	Does location support religion activity (Value is 0 or 1)

Table 7: Activity Location Assignment File

[1] ADCW is the ADC WorldMap: https://www.adci.com/adc-worldmap

Field	Description
Household ID	Household ID of the person
Person ID	Person ID of the person
Activity number	Number of the activity in the activity sequence to which it belongs
Activity type	Activity type (see above)
Start Time	Start time of activity in seconds since midnight Sunday/Monday
Duration	Duration of the activity in seconds
Location ID	Location ID of the location where the activity takes place (rlid or alid)

Table 8: Population Contact Network

Field	Description
Person ID 1 (pid1)	Person ID number 1 of this edge
Person ID 2 (pid2)	Person ID number 2 of this edge
Location ID (lid)	Location ID where contact (edge) arises
Start time	Start time of the contact between Person ID 1 and Person ID 2 measured in seconds since midnight of Sunday/Monday
Duration	Duration of the contact measured in seconds
Activity ID 1 (activity1)	Activity type of Person ID 1 at time of contact (see above)
Activity ID 2 (activity2)	Activity type of Person ID 2 at time of contact (see above)

Table 9: Disease outcomes file (training data)

Field	Description
Day	Simulation day
Person ID	Id of the person in the Person data or the Contact Network data
Disease state	Disease related state of the person on the day with possible values S for "susceptible", I for "infected", and R for "recovered". See *Disease states and asymptomatic infection* section for details.

Table 10: Disease outcomes file (test data)

Field	Description
Person ID	Id of the person in the Person data or the Contact Network data
Infected	Binary variable with 1 if the person was infected in the final week of the simulation and 0 if otherwise

Figure 1 below shows the overall data model, capturing the relationships between the different tables.

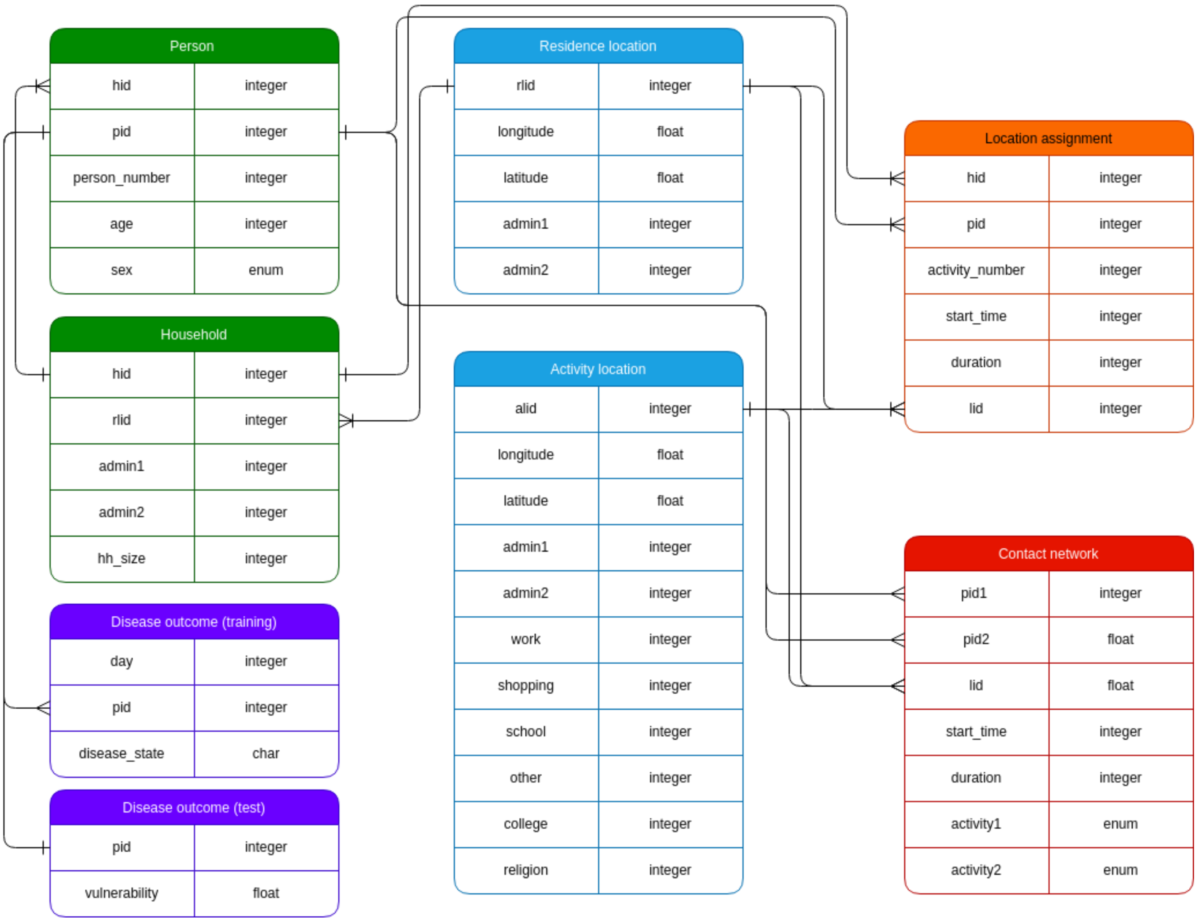


Figure 1. Overall data model of the synthetic population data. The arrows indicate dependencies between files; for example, the hid element in the person file should be constrained to the values available in the household file

Evaluation dataset

The datasets being provided are intended for local development use in both Phase 1 and Phase 2. Each dataset has been split in time - the first 56 days of the dataset is the training set, and the final 7 days of the dataset is the forecast target data. The prediction task, as detailed in the *Challenge Scenario* section below, is to make predictions for the final 7 days of the dataset. The ground truth is provided for the forecast target period for the development datasets.

In Phase 2, a separate and held-out dataset will be used for solution evaluation. You can expect the Phase 2 evaluation dataset to be close in size and statistical distributions to the development Virginia dataset (if you are participating in the US Challenge) or to the development UK dataset (if you are participating in the UK Challenge), but have a different contact network and an independent disease outbreak simulation. In Phase 2, you will submit code for your solution to a code execution environment. The code execution runtime will run cold-start federated training on the new dataset's training split and then run inference to generate predictions for the forecast target period. Your solution's performance will be measured by evaluating its predictions against the ground truth for the new dataset's target period.

Challenge scenario: developing privacy-preserving FL models for forecasting individual risk of infection

The key analytical objective of the PPFL task is to effectively train a model that can predict the risk of infection for individuals in a population over a period of time in a privacy-preserving manner. Such a predictive capability is important from the perspective of an individual; if they know they are likely to be infected in the next week, they may choose to take additional preventative measures (such as wearing a face covering, reducing social contacts, or taking antivirals prophylactically). Additionally, this capability can help inform the measures public health authorities implement to respond to such outbreaks.

The dataset will be horizontally partitioned into local datasets belonging to different federation units. This is intended to mirror how data is distributed across different health districts, hospitals, etc. These federation units have access to the full data tables described above. The dataset assumes that people are in precisely one federation unit, so a health district can see everywhere that one of its individuals goes, but does not have any access to information about other people who reside outside of the health district. As a result, the contacts between individuals from two different federation units are not represented in the population contact network in either of the local datasets.

The federated units work jointly to train a global FL model, and can take a common approach to technical design, infrastructure, etc., but are **not able to access each other's raw data.** In the real world there are a number of barriers that might prevent this; the parties and the datasets they contribute may be subject to a variety of privacy, competition and sector-specific regulations (e.g., HIPAA), may be operating in different jurisdictions, and may have legitimate commercial or ethical reasons for not sharing data with other parties.

The key task of this challenge is to design a privacy solution so that the collaborating parties can jointly train and deploy such a model without compromising the privacy requirements (more details on the requirements, and an associated threat model, are described below).

The 63 days of simulated disease outbreak data is split into 56 days of training data and 7 days of target data. Participants will be provided with the synthetic population data and the first 56 days of the simulated outbreak datasets for model training. The analytical task is to predict a risk score for the binary disease state (infected or not infected) of each individual in the final week of the simulation. That is, you should predict the risk of whether each person in the population will be in an infected state on any day in the last 7 days.

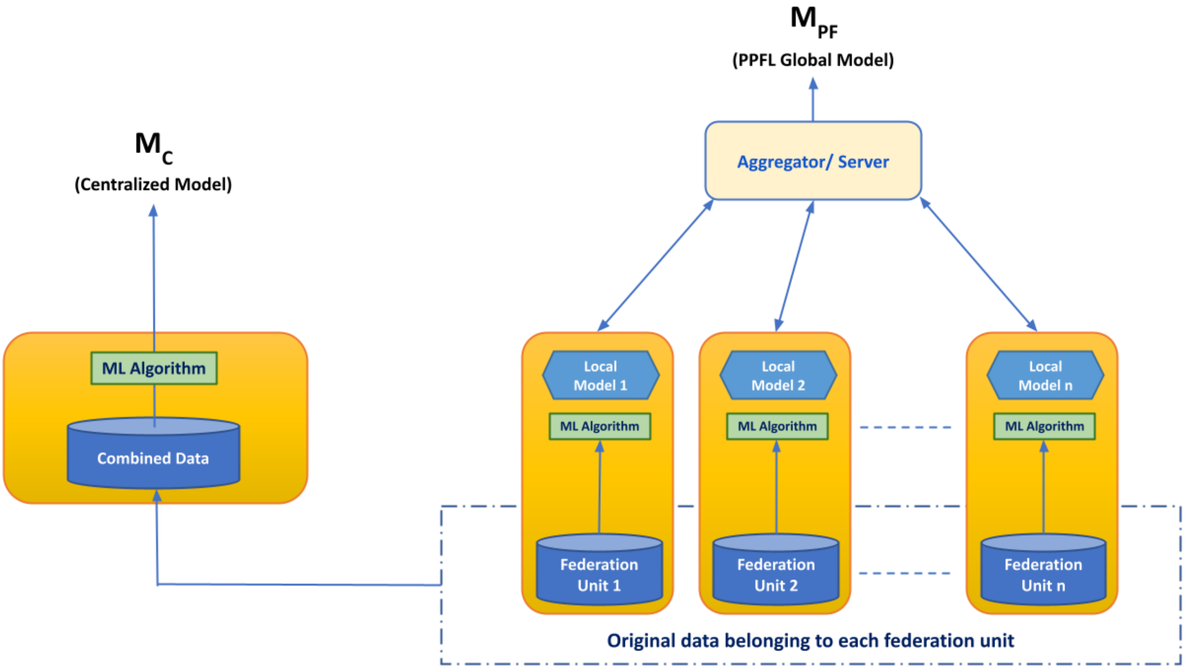


Figure 2. Simple illustration of a Centralized ML model and a federated learning model

For the purposes of the challenge, participants should demonstrate their solution by training two models:

- M_C = a centralized model trained on the datasets in a non-privacy-preserving way.
- M_{PF} = a privacy-preserving federated learning model trained using their privacy solutions.

Participants will be provided an example centralized model developed by the UVA-BI team, as well as the code used to train this model. Participants are permitted to use this example code as the basis of their PPFL solution, or take an entirely different analytical approach.

In either case, the core of the evaluation will be assessing the comparison between a centralized model M_C (trained without consideration of privacy), and an alternative model M_{PF} that combines a federated learning approach with innovative privacy-preserving techniques.

Details of the evaluation criteria are given below which, at a high level, consider:

- The ability of the solution to deliver (and evidence) relevant privacy properties
- The accuracy of model M_{PF} compared to M_C
- The performance/computational cost of training M_{PF} compared to M_C
- The scalability, usability, and adaptability of the solution.

It is important to note that the **accuracy and performance measurements are comparative**; the challenge is designed to reward strong privacy solutions which minimize accuracy loss and can be run with acceptable compute, memory, storage and communication costs. Privacy solutions which can support more effective machine learning approaches are encouraged (and will likely score higher in some areas), but the overall accuracy of the centralized model M_C is not a key factor in scoring.

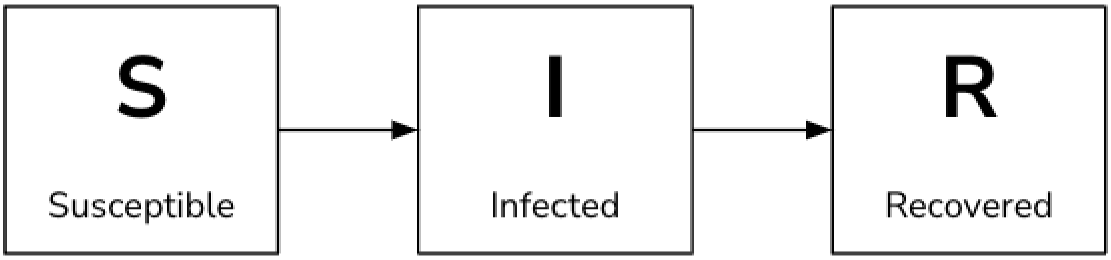
Participants are free to determine the set of privacy technologies they use, with the exception of specialized or bespoke hardware. This exclusion is to ensure a fair baseline for Phase 2 evaluation. Solutions will be evaluated in a common technical environment, with each solution running on identical (virtualized) hardware, with access to the same compute, memory, storage and network infrastructure. We are therefore unable to provide access to specialist hardware such as secure enclaves/trusted execution environments in this challenge. However, the challenge organizers retain a deep interest in hardware-based privacy technologies, and encourage researchers or companies working in this space to engage with us to explore how we could collaborate in future to advance research and adoption of such technologies.

There are no restrictions on the software that challenge participants use in their solutions. We anticipate that proposals may leverage various de-identification techniques, differential privacy, cryptographic techniques, or combinations thereof. But this is not a prescriptive list, and we highly encourage submissions that propose novel technological approaches, or innovative application of existing technologies.

Disease states and asymptomatic infection

Each individual in the population will have a disease state for each day of the dataset. The disease state acts like a finite-state machine with the following states:

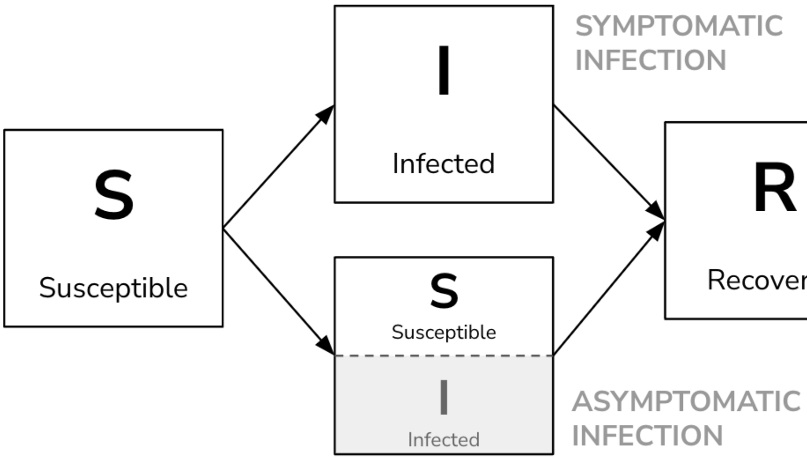
- S: indicates "susceptible". This individual is susceptible to infection.
- I: indicates "infected". This individual has been infected and is infectious. They can infect other individuals (who are susceptible) through contact.
- R: indicates "recovered". This individual has recovered and is no longer infectious or infectable.



State transition diagram for disease state in challenge dataset.

An individual disease state can only progress from S to I to R. Once an individual is in the recovered R state, they will not be able to become infected again within this simulation.

This dataset also features an additional layer of complexity around asymptomatic infection. Some individuals in the simulation will be asymptomatically infected and will transition to I states, but the I state is hidden from the data. Such an individual is still infectious while in the I state but will appear in the data as being in the S state before transitioning to the R state. This is in contrast to individuals with the normal symptomatic infection whose disease state is transparently visible in the data. Individuals in both normal symptomatic infection and asymptomatic infection states will transition to a visible R state upon recovery.



State transition diagram for disease state in challenge dataset showing both symptomatic and asymptomatic infections. Asymptomatically infected individuals will have a hidden infected state in the simulation, but appear as susceptible in the dataset. For this reason, asymptomatic infections may appear to transition directly from susceptible to recovered.

Privacy Threat Profile

Overview

Participants will design and develop end-to-end solutions that preserve privacy across a range of possible threats and attack scenarios, through all stages of the machine learning model lifecycle. Participants should therefore carefully consider the overall privacy of their solution, focusing on the protection of sensitive information held by all parties involved in the federated learning scenario. The solutions designed and developed by participants will include comprehensive measures to address the threat profile described below. These measures will provide an appropriate degree of resilience to a wide range of potential attacks defined within the threat profile.

Scope of Sensitive data

Participants' solutions must prevent the unintended disclosure of sensitive information in the population dataset to any other party, including other insider stakeholders (i.e., the other federation units) and outsiders.

The following information in the dataset should be treated as sensitive:

- All personally identifiable information, including the identity of a person, their housing information, their demographics such as age, etc.
- Location activities of an individual
- The health state of an individual
- Social contact information, including the location of any social contact, and the identities of who was involved in the social contact.

Lifecycle

Participants will consider risks across the entire lifecycle of a solution including, in particular, the following stages:

- Training
  - Raw training data should be protected appropriately during training
  - Sensitive information in the training data should not be left vulnerable to reverse-engineering from the local model weight updates.
- Prediction/inference

- Sensitive information in the training data should not be left vulnerable to reverse-engineering from the global model. The privacy solution should aim to ensure that those with access to the global model cannot infer sensitive information in the training data for the lifetime of the model's production deployment.

Actors and intention

Participants will consider threat models that range from honest-but-curious[2] to malicious (aggregators and participating clients) and propose solutions accordingly. While participating organizations can be trusted, such threat models help capture a broad spectrum of possible risks, such as the outsourcing of computation to the untrusted cloud; and, in the event trusted private cloud infrastructure is used, the remaining possibility that malicious external actors could compromise part of that infrastructure (for example, one or multiple participating federation units), leading to a potential reduction in the trustworthiness of components within the system.

Privacy attack types

Any vulnerabilities that could lead to the unintended exposure of private information could fundamentally undermine the solution as a whole. Participants will therefore consider a range of known possible privacy attacks, and any new ones relevant to the privacy techniques employed or to this specific use case. Participants will primarily be expected to consider inference vulnerabilities and attacks, including the risks of membership inference and attribute inference. In the white paper, participants should clearly indicate the threat models considered and any other assumptions.

Participants will be expected to address the risks associated with the considered threat model through the design and implementation of technical mitigations in their solutions, and to explain in their white paper how solutions will mitigate against these. Participants will be expected to consider whether technical innovations introduced in their proposed solution may introduce novel privacy vulnerabilities and to clearly articulate potential privacy attacks and mitigations. Throughout both the white paper and solution development phases, participants should also take into account established privacy and security vulnerabilities and attacks, and corresponding best practice mitigations.

[2] An honest-but-curious party is a legitimate party in the federated learning scenario who will not attempt to deviate from the defined protocol but will attempt to learn all possible information from data received legitimately from other parties.

Challenge Phases & Evaluation

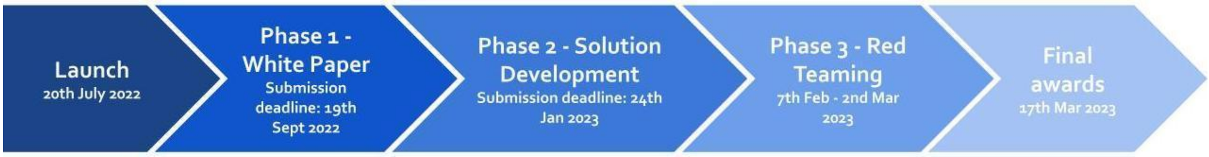


Figure 2. Timeline for the challenge phases

Phase 1: White Paper

In Phase 1, participants are asked to produce a technical white paper setting out their proposed solution, and provide information to support initial decisions about funding eligibility and Phase 1 prize awards. It is expected that some initial implementation/prototyping activity may be already underway when participants submit white papers, but code does not need to be submitted at this point.

White papers should be a **maximum of 10 pages** (excluding references), with size 11 font.

In addition to the white paper, participants need to respond to a number of other questions to establish their eligibility to participate in the prize challenge. The details of these, and process for submitting them, are different for UK and US participants - please see information on the challenge website.

The technical white paper should:

- Clearly describe the technical approaches and sketch out proof of privacy guarantees based on the threat model considered, including:
    - The design of any algorithms, protocols, etc. utilized
    - Formal or informal arguments of how the solution will provide privacy guarantees.
- Clearly list any additional privacy issues specific to the technological approaches used
- Justify initial enhancement or novelty compared to the state-of-the-art
- Articulate:
    - Expected efficiency and scalability of the privacy solution
    - Expected trade-offs between privacy and accuracy/utility

    - How the explainability of model outputs may be impacted by your privacy solution.
- Describe how the solution will cater to the types of data provided to participants, and articulate what additional work may be needed to generalize the solution to other types of data.

Evaluators will score the white papers against the weighted criteria outlined in the table below. Note that the different criteria are not fully independent of one another. For example, solutions will likely need to carefully consider trade-offs between privacy and accuracy, accuracy and efficiency, etc. Participants should take the weightings of the criteria into account when considering these trade-offs. Importantly however, proposals must demonstrate how acceptable levels of both privacy and accuracy will be achieved – one must not be completely traded off for the other (a fully privacy-preserving but totally inaccurate model is of little use to anyone). Proposals that do not sufficiently demonstrate how both privacy and accuracy will be achieved (as determined by an independent expert assessor) will not be eligible to score points in the remaining criteria.

Topic	Specific Criteria	Weighting (/100)
Technical Understanding	Does the white paper demonstrate an understanding of the technical challenges that need to be overcome to deliver the solution?	10
Privacy	Has the white paper considered an appropriate range of potential privacy attacks, and how the solution will mitigate those?	25
Accuracy	Is it credible that the proposed solution could deliver a useful level of model accuracy?	10
Efficiency and Scalability	Is it credible that the proposed solution can be run within a reasonable amount of computational resource (e.g., CPU, memory, storage, communication), when compared to a centralized approach for the same machine learning technique?	15

	Does the white paper propose an approach to scalability that is sufficiently convincing from a technical standpoint to justify further consideration, and reasonably likely to perform to an adequate standard when implemented? Solution scalability will be evaluated primarily for a) the number of connected federation units, b) volume of population data held by those units	
Adaptability	Is the proposed solution potentially adaptable to different use cases and/or different machine learning techniques?	5
Feasibility	Is it likely that the solution can be meaningfully prototyped within the timeframe of the challenge?	10
Innovation	Does the white paper propose a solution with the potential to improve on the state of the art in privacy-enhancing technology? Does the white paper demonstrate an understanding of any existing solutions or approaches and how their solution improves on or differs from those?	20
Usability and Explainability	Does the proposed solution show that it can be easily deployed and used in the real world, and provide a means to preserve any explainability of model outputs?	5

Phase 2: Solution Development

In Phase 2, participants are asked to develop working prototypes of their solutions. These solutions are expected to be capable of being used to train a model against the evaluation dataset, with measurement of relevant performance and accuracy metrics. However, we are not expecting fully productionized solutions; for example participants will be able to actively support deployment on any testing platforms, and any test runs against evaluation data.

The following section describes how we plan to evaluate solutions. Further details, including technical details for submission, will be supplied in an *Evaluation Methodology* document to be supplied to participants at the start of Phase 2. Though our intent is that the details below will remain unchanged, organizers reserve the right to make changes to specific evaluation criteria or weightings if we consider that this is necessary to fairly and efficiently evaluate the full range of solutions proposed, or for other reasons.

Evaluation

Participants will submit the following for evaluation:

1. Centralized model M_C:

- Code for training a centralized model M_C
- Documentation for how to train and make inferences from the centralized model, including a list of any dependencies (e.g., a requirements.txt)

2. Privacy-preserving federated learning model M_{PF}:

- Code for training the PPFL model, M_{PF}
- Documentation for how to train and make inferences from the PPFL model (e.g., a requirements.txt)

3. Key metrics:

- Self-reported privacy, accuracy, and efficiency metrics for the two models (these will not be used to evaluate solutions, but help to flag potential issues if assessors obtain very different metric values).

Submitted solutions will be deployed and evaluated on a technical infrastructure hosted by the challenge organizers. This infrastructure will provide a common environment for the testing, evaluation, and benchmarking of solutions in accordance with the Phase 2 evaluation criteria below.

Participants will be provided with the following in the Evaluation Methodology document to be supplied at the start of Phase 2:

- Hardware specifications (CPU/GPU, memory, data storage etc.) for the runtime environment that will be used to benchmark and evaluate solutions

- Software specifications for the runtime environment, including any fundamental requirements for code to successfully execute
- Upper limits for a) execution time, and b) resource requirements (CPU/GPU, memory, data storage etc.) for code execution during the model training stage
- Example code submissions.

An independent assessor will take the following steps to perform the evaluation, using a sequestered training and test dataset that participants have not had access to:

1. Train the centralized model M_C on a single node, with synthetic population data stored on local disk
    - Measure efficiency metrics during training
    - Measure accuracy metrics using test data on the resultant model.
2. Train the PPFL model M_{PF} across N+1 nodes, where one node is the server/aggregator and the remaining nodes are for the federation units. Predetermined values of N between 1 and 10 will be used to evaluate the scalability of the solutions.
    - Measure efficiency metrics during training
    - Measure accuracy metrics using test data on the resultant model
    - Measure privacy metrics during training (to assess risk of leakage from local model updates and any other exchanged information) and inference (to assess risk of leakage from the resultant global model).
3. Qualitative assessments will be made of the solution's adaptability, usability, explainability, and level of technical innovation.

Details of the specific criteria and how they will contribute towards overall scoring are given in the table below. We will provide more specific information on how these will be measured practically in the Evaluation Methodology document to be provided prior to the start of Phase 2.

As with the Phase 1 evaluation, the different criteria are not fully independent of one another, and the outcomes of trade-off considerations made in the white paper should be reflected in the developed solution. Importantly, solutions must meet a minimum threshold of privacy and accuracy (which will be quantitatively measured) to be eligible to score points in the remaining criteria.

Topic	Factors	Weighting (/100)
Privacy	Information leakage possible from the PPFL model M_{PF} during training and inference, for a fixed level of model accuracy[3] Ability to clearly evidence privacy guarantees offered by solution in a form accessible to a regulator and/or data owner audience	35
Accuracy	Absolute accuracy of the PPFL model M_{PF} developed (e.g., F1 score) Comparative accuracy of PPFL model compared with a centralized model, for a fixed amount of information leakage	20
Efficiency and scalability	Time to train PPFL model M_{PF} vs comparison with the centralized model M_C Network overhead of model training Memory (and other temporary storage) overhead of model training Ability to demonstrate scalability of the overall	20

[3] For example: If differential privacy is employed to protect output privacy, and F_1 score is an appropriate accuracy metric, what is the smallest value for the privacy budget ε that can be configured to achieve an F_1 score that is a fixed amount less than the F_1 score of the centralized model M_C

	approach taken for additional nodes	
Adaptability	Range of different use cases that the solution could potentially be applied to, beyond the scope of the current challenge	5
Usability and Explainability	Level of effort to translate the solution into one that could be successfully deployed in a real world environment Extent and ease of which privacy parameters can be tuned Ability to demonstrate that the solution implementation preserves any explainability of model outputs.	10
Innovation	Demonstrated advancement in the state-of-the-art of privacy technology, informed by above-described accuracy, privacy and efficiency factors	10

Partitions

For local development, you are provided a full, unpartitioned dataset. In Phase 2 evaluation, the evaluation data will be partitioned along administrative boundaries, e.g., by grouped FIPS codes/counties. Any cross-partition edges will be dropped, where an edge is created between two people (nodes) that come into contact as specified by the population contact network table. That is, each partition will only have visibility into edges between people that reside within the same partition. Cross-partition edges will continue to affect the spread of infection, but will not be visible to any of the partitions. Any partitioning of the data that you might perform in your local development experiments should take this into account. Your solution should be able to handle any number of partitions, and in Phase 2, we may evaluate your solution with a number of partitions between 1 and 10.

Prediction Target and Evaluation Metric

The target variable for the modeling task is a risk score (between 0.0 and 1.0) for each individual in the population. That risk score corresponds to a confidence that that individual enters into an symptomatic infected I disease state at any time during the final week of the simulation (between days 57 and 63 inclusive).

Note that some individuals become infected asymptomatically. Those individuals count as negative cases (not infected) for the purposes of the ground truth.

The evaluation metric will be Area Under the Precision–Recall Curve (AUPRC), also known as average precision (AP), PR-AUC, or AUCPR. This is a commonly used metric for binary classification that summarizes model performance across all operating thresholds. This metric rewards models which can consistently assign positive cases with a higher risk score than negative cases. AUPRC is computed as follows:

$$\text{AUPRC} = \sum_{n} (R_n - R_{n-1}) P_n$$

where P_n and R_n are the precision and recall, respectively, when thresholding at the nth individual sorted in order of increasing recall.

Phase 3: Red Teaming/Testing

In Phase 3, red teams will plan and launch privacy attacks against the highest-scoring solutions developed in Phase 2. Solutions will be re-evaluated based on the outcomes of the red teaming attacks, and each solution will be assigned a final score which will be used to determine the allocation of prize awards. The criteria outlined in Phase 2 will be used for this re-evaluation, taking into account the impact of the red team attacks on the solutions. Most notably, it is expected that privacy scores will change according to how resilient the solution was to the red teams' privacy attacks.

Success of red team attacks will be assessed by a panel of judges using the criteria below, in order to evaluate the empirical results reported, the approaches taken and the severity of the flaws red teams are able to exploit. Details of the specific criteria and how they will contribute towards overall scoring are given in the table below.

Each red team will be assigned multiple solutions to test, with each solution therefore being tested by multiple red teams. Individual red teams will be scored, in part, by comparing the outcomes of attacks carried out against the same solutions by different red teams. Additionally, an individual red team's attacks against the solutions it was assigned to attack will be assessed for consistency, difficulty, novelty, rigor, and practicality.

Further details on red teaming and recruitment of red teams will be provided to participants in Autumn/Fall 2022.

Topic	Factors	Weighting (/100)
Effectiveness	How completely does the attack break the privacy claims made by the target solution? (e.g., what portion of user data is revealed, and how accurately is it reconstructed)?	40
Applicability / Threat Model	How realistic is the attack? How hard would it be to apply in a practical deployment?	30
Generality	Is the attack specific to the target solution, or does it generalize to other solutions?	20
Innovation	How significantly does the attack improve on the state-of-the-art?	10

Annex A: conditions of data use

To participate in Phase 2 of the challenge, participants are required to accept and comply with a lightweight data use agreement for one or both of the synthetic datasets. The datasets do not contain personal data, but their use is restricted to the purposes of this challenge. The datasets are provided by the University of Virginia Biocomplexity Institute and Initiative, Network Systems Science and Advanced Computing Division (UVA). Any and all rights to the datasets remain vested in UVA.

The process for accepting the data use agreement, and securely downloading the data, is different for UK and US participants. Please consult the UK and US challenge briefing materials on the challenge website for further details.

Challenge Participants are only allowed to use these datasets for the Challenge and in accordance with the following terms and conditions:

Item	Description
Data and ownership	The provider shall provide the dataset described for the purposes of the PETs Prize Challenges. Provider shall retain ownership of any rights it may have in the Data, and participant does not obtain any rights in the Data other than as set forth herein. The Data will not include information that can be used to distinguish or trace an individual's identity such as name, social security number, date and place of birth, mother's maiden name, or biometric records or any other information that is linked or linkable to an individual, such as medical, educational, financial, or employment information ("Personally Identifiable Information"). Participants will not use the Data, either alone or in concert with any other information, to make any effort to identify or contact individuals who are or may be the sources of Data without specific written approval from Provider and appropriate Institutional Review Board (IRB) approval, if required pursuant to applicable laws or regulations. Should Participant inadvertently receive identifiable information or otherwise identify a subject, Participant shall promptly notify the Organisers and follow Provider's reasonable written instructions, which may include return or destruction of the identifiable information. Participant shall promptly report to the Organisers any use or disclosure of the Data not provided for by this Agreement of which it becomes aware. Participant agrees to follow its own applicable institutional policies and applicable laws and regulations. The parties agree to take such action as is necessary to amend this Agreement, from time to time, in order for the Provider or Participant to remain in compliance with the applicable requirements.
Cost	The Data will be provided to the Participant at no cost.
Use	Participant shall not use the Data except as authorized under this Agreement. The Data will be used solely to conduct the Project and solely by Participant Scientist and Participant's faculty, employees, fellows, students, and agents, including any authorized third parties, ("Participant Personnel") that have a need to use, or provide a service in respect of, the Data in connection with the Project and whose obligations of use are consistent with the terms of this Agreement (collectively, "Authorized Persons"). For the purposes of this Agreement, Authorized Persons expressly includes Innovate UK and all participants in the Project. Participant agrees to use the Data in compliance with all applicable laws, rules, and regulations, as well as all applicable professional standards. Except as authorized under this Agreement or otherwise required by law, Participant agrees to retain control over the Data and shall not disclose, release, sell, rent, lease, loan, or otherwise grant access to the Data to any third party, except Authorized Persons, without the prior written consent of Provider. Participant agrees to establish appropriate administrative, technical, and physical safeguards to prevent unauthorized use of or access to the Data and comply with any other special requirements relating to safeguarding of the Data as may be set forth in these terms. In the event that any unauthorized use occurs, Provider may restrict Participant's or any Authorized Persons' use or possession of the Data.

Publication	In no event will Participant or any Authorized Persons publish or make publicly available in any way the Data. Participant and all Authorized Persons agree to recognize the contribution of the Provider as the source of the Data in all written, visual, or oral public disclosures concerning the Project using the Data, as appropriate in accordance with scholarly standards and any specific format that has been indicated in section Acknowledgement/Attribution below.
Term	Unless terminated earlier in accordance with this section or extended via an amendment in accordance with Section 10, this Agreement shall expire one (1) year from the Effective Date set forth above. Either party may terminate this Agreement with thirty (30) days written notice to the other party's Authorized Official to the address set forth above. Upon expiration or early termination of this Agreement, Participant shall follow the disposition instructions provided below, provided, however, that Participant may retain one (1) copy of the Data to the extent necessary to comply with the records retention requirements under any law.
No warranties	Except as provided below or prohibited by law, any Data delivered pursuant to this Agreement is understood to be provided "AS IS." PROVIDER MAKES NO REPRESENTATIONS AND EXTENDS NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED. THERE ARE NO EXPRESS OR IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE, OR THAT THE USE OF THE DATA WILL NOT INFRINGE ANY PATENT, COPYRIGHT, TRADEMARK, OR OTHER PROPRIETARY RIGHTS. Notwithstanding the foregoing, Provider, to the best of its knowledge and belief, has the right and authority to provide the Data to Participant for use in the Project.
Liability	Except to the extent prohibited by law, the Participant assumes all liability for damages which may arise from its use, storage, disclosure, or disposal of the Data. The Provider will not be liable to the Participant for any loss, claim, or demand made by the Participant, or made against the Participant by any other party, due to or arising from the use of the Data by the Participant, except to the extent permitted by law when caused by the gross negligence or willful misconduct of the Provider. No indemnification for any loss, claim, damage, or liability is intended or provided by either party under this Agreement.
Acknowledgment / Attribution:	Where use of the Data results in academic publications or presentations, Participants agree to recognize the contribution of the Provider as the source of the Data in all written, visual, or oral public disclosures concerning the Project using the Data, as appropriate in accordance with scholarly standards as follows: "This work is based on the dataset developed under NSF RAPID: COVID-19 Response Support: Building Synthetic Multi-scale Networks, NSF Expeditions in Computing and the University of Virginia Strategic Investment Fund (provided under the UK-US PETs Prize Challenges, available only to participants) by the University of Virginia Biocomplexity Institute and Initiative, Network Systems Science and Advanced Computing Division and such dataset is used with express permission under the UK-US PETs Prize Challenges Rules." In other non-academic publicity, advertising or news release about the Challenges, the Participants will agree to acknowledge the role of the Provider of the dataset as follows:

	Dataset provided by the University of Virginia Biocomplexity Institute and Initiative, Network Systems Science and Advanced Computing Division. Participants will not use the name of Provider in a manner that states or implies an endorsement from Provider for any products or services. No trademarks or logos of Provider may be used without express written permission from the Provider.
Disposition Requirements upon the termination or expiration of the Agreement	Within 30 days upon conclusion of the Project, Participants must delete any Data or related information provided under this Agreement, except as may be required by any applicable record retention laws or regulations. If any Data or related information is kept under this exception, it cannot be used for any other purpose.

Annex B: Version History

Version/Date	Changes
V1.0 (20th July)	Initial version at launch
V1.1 (17th Aug 2022)	Various improvements to technical content to better align with the final version of the data sets as released to participants, specifically: ● In the "Datasets" and "Evaluation datasets" sections, content added to describe that in Phase 2 submissions to the US challenge will be evaluated against a sequestered Virginia dataset, and submissions to the UK challenge will be evaluated against a sequestered UK dataset. ● In the "Challenge Scenario" section, content added to specify that the specific analytical task is to predict the risk of whether each person in the population will be in an infected state on any day in the last 7 days. ● New "Disease states and asymptomatic infection" added, detailing the different states of infection and the possible transitions between states. ● Additions to "Dataset structure and schema" section: ○ **Activity location assignment data:** appended "Activities are repeated every day, as if it were like the film Groundhog Day" ○ **Population contact network data:** appended "This table is generated from the Activity location assignment data table, and also repeats every day." ○ **Disease state:** appended "with possible values S for "susceptible", I for "infected", and R for "recovered". See *Disease states and asymptomatic infection* section for details."

	○ **Disease State / Vulnerability:** replaced with a Boolean “Infected” variable ● New “Partitions” subsection added, detailing how the data is partitioned and how partitioning impacts the edges of the social contact graph. ● New “Prediction Target and Evaluation Metric” subsection added, detailing how the Area Under the Precision–Recall Curve (AUPRC) will be used to assess model accuracy.