Privacy-Enhancing Technologies

PRIZE CHALLENGES

Lessons learned from running a privacy tech challenge

Dave Buckley *Centre for Data Ethics and Innovation*

gov.uk/cdei

linkedin.com/in/davidbuckley3

david.buckley@cdei.gov.uk

Who we are

The Centre for Data Ethics and Innovation (CDEI) **leads the UK Government's work to enable trustworthy innovation** using data and AI.

It works to **facilitate responsible and trusted innovation** to improve the lives of citizens and support growth.

To build public trust, **the CDEI works with partners across the public sector, industry and academia,** in the UK and internationally, to identify and tackle barriers to responsible innovation, and to scale these tools and methodologies to other organisations.

Barriers to ethical innovation

- **2021 CDEI survey:** 86% of AI vendors highlighted data availability and fragmentation as barriers to innovation
- Many valuable datasets exist across many sectors, but access is heavily restricted
- Data ethics issues can also be compounded by data access challenges: e.g. accessing demographic data for bias audit and mitigation

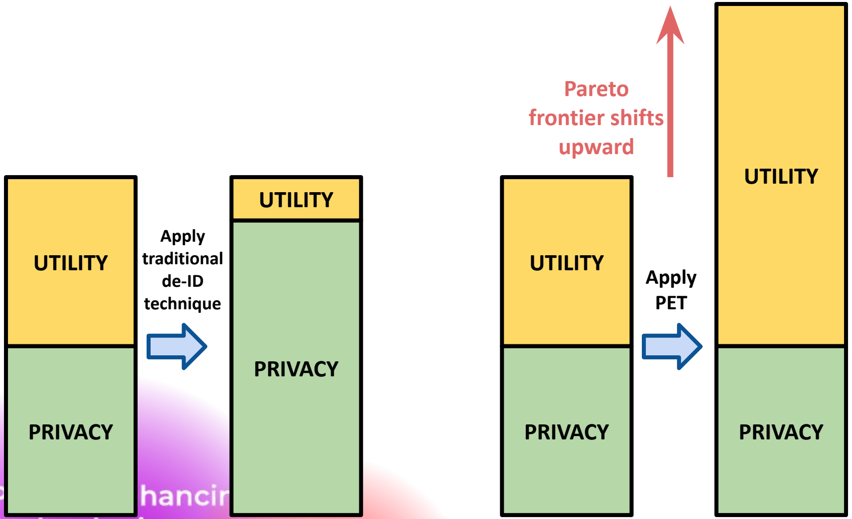
=> **Responsible Data Access programme** to tackle these challenges in practice

Strong focus on PETs: transformative set of technologies that can maximise value of data whilst maintaining privacy. Workstreams to drive development, and drive adoption to deliver data-driven innovation in the public sector

Why PETs? Have your cake and eat it

Royal Society, 2023:

“PETs are an emerging set of technologies and approaches that enable the derivation of useful results from data without providing full access to the data.”



UTILITY

PRIVACY

Apply traditional de-ID technique

UTILITY

PRIVACY

Pareto frontier shifts upward

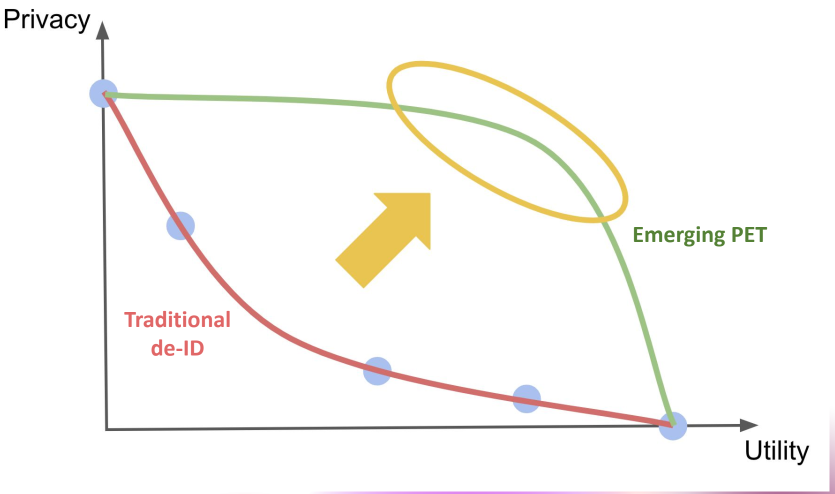
UTILITY

PRIVACY

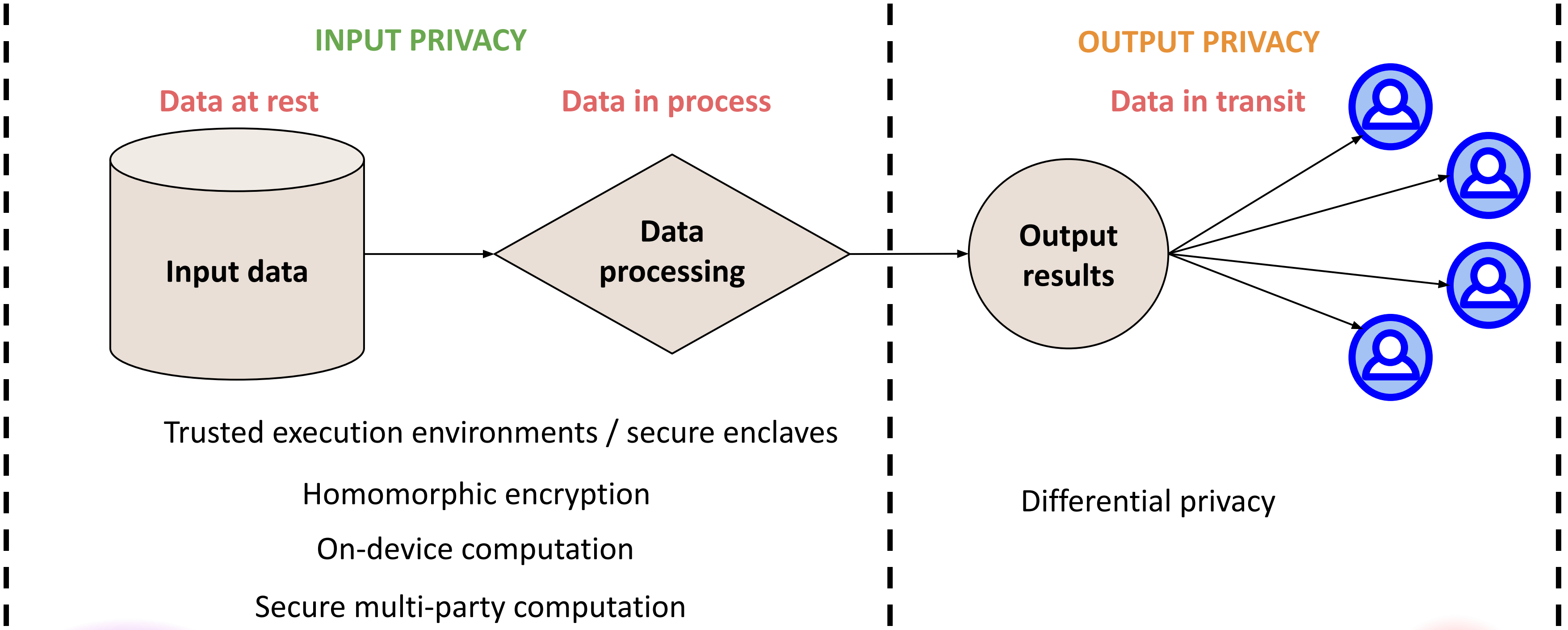
Apply PET

UTILITY

PRIVACY



PETs across the data processing lifecycle



Trusted execution environments / secure enclaves

Homomorphic encryption

On-device computation

Secure multi-party computation

Differential privacy

https://cdeiuk.github.io/pets-adoption-guide/what-are-pets

The need in healthcare

The Royal society report “From privacy to partnership The role of privacy enhancing technologies in data governance and collaborative analysis” [1] states that

> *“Recent* ***advances in medical imaging, audio and AI*** *have led to unprecedented possibilities in healthcare and research. This is especially true of the UK, where the public health* ***system is replete with population-scale electronic patient records****. These conditions, coupled with strong academic and research programmes, mean that the* ***UK is well positioned*** *to deliver timely and impactful health research and its translation to offer more effective treatments, track and prevent public health risks, utilising health data to improve and save lives [2].*
>
> *….*
>
> *Anonymous data is not covered by current data protection law in the UK and EU. However, it is* ***difficult to be certain that health data is anonymous****, particularly in biometric and other non-textual data. Health data is subject to specific legal requirements in the UK, as well as the common law duty of confidentiality.”*

<u>Three high level use-cases</u>

One-off Access for Research & Analysis

Example - Federated Learning across several trusts’ secure data

Continual Access to User

Example - Synthetic data in Secure Data Environment (SDE) or Regular API requests to data with privacy accountant

Auto workflow (built with biometric data flows)

Example - Homomorphic Encryption of biometric data to central processing

[1] https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/From-Privacy-to-Partnership.pdf?la=en-GB&hash=4769FEB5C984089FAB52FE7E22F379D6

[2] HM Government (Life sciences industrial strategy update). See https://www.gov.uk/government/publications/lifesciences-industrial-strategy-update (accessed 15 March 2022).

Blockers

The main hurdle to adoption appears to be expectation. It is very difficult to balance the envisaged benefits versus the delivery (both below and above expectation issues) as current implementations often use bespoke combinations of techniques with very nuanced issues.

Algorithms

Many algorithms appear to be well developed but then are either insufficient for a particular application, or concerns have been highlighted around particular elements which undermine confidence.

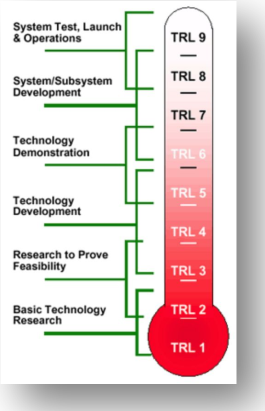
Hardware

Several technologies boast great opportunities but at a significant computational cost (either through high levels of communication or multiple computations of difficult algorithms).

Evaluation and standards

Often a balance between privacy, utility, computational cost, etc.. Is part of the solution. Without clear standards for acceptable levels or considerations that need to be addressed, it is difficult to implement these technologies on real data.

An assessment of combinations of techniques, e.g. Technology Readiness Levels (TRLs) would support adoption.

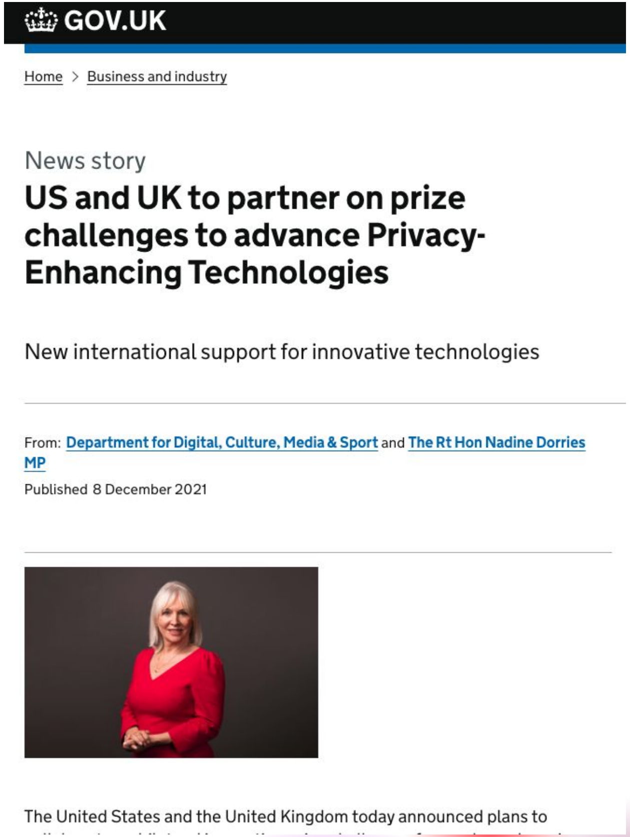

System Test, Launch & Operations
System/Subsystem Development
Technology Demonstration
Technology Development
Research to Prove Feasibility
Basic Technology Research
TRL 9
TRL 8
TRL 7
TRL 6
TRL 5
TRL 4
TRL 3
TRL 2
TRL 1

End-to-end demonstrations (including adherence to regulations) are required.

UK-US PETs prize challenges

PETs prize challenges

- Announced at the Summit for Democracy in December 2021
- Prize pot of ~$1.3m for researchers and engineers to develop innovative privacy-preserving systems to help tackle societal systems



GOV.UK

Home > Business and industry

News story

US and UK to partner on prize challenges to advance Privacy-Enhancing Technologies

New international support for innovative technologies

From: Department for Digital, Culture, Media & Sport and The Rt Hon Nadine Dorries MP

Published 8 December 2021

The United States and the United Kingdom today announced plans to

Privacy-Enhancing Technologies
PRIZE CHALLENGES

Challenge Partners

NHS

United States

- National Institute for Standards & Technology
- National Science Foundation
- White House Office of Science & Technology Policy

United Kingdom

- Center for Data Ethics & Innovation
- Innovate UK

Data Partners

- SWIFT
- UVA Biocomplexity Institute
- DrivenData

Independent assessors

Additional support from: NHS Transformation, DARE UK, Data Science Campus, FCA, FinCEN, NECC, ICO

Privacy-Enhancing Technologies
PRIZE CHALLENGES

Aims of the challenge

Technical goals

- **Drive technological innovation** in the state-of-the-art of privacy enhancing technologies
- Develop solutions that **deliver strong end-to-end privacy guarantees** against a set of common threats and privacy attacks
- Develop solutions that can **efficiently generate high-utility machine learning models**

Broader goals

- **Demonstrate the value-add of PETs** in "realistic" use cases
- **Build community** of PETs developers, would-be adopters, regulators, etc.
- **Deepen collaboration** between UK and US on tech innovation

Why privacy-preserving federated learning?

- Office hour sessions and targeted engagement with SMEs in early 2022 identified opportunity for meaningful innovation in PPFL:
  - Existing (vanilla) deployments of federated learning solutions often **do not truly protect confidentiality of data across the full ML lifecycle**
  - Cross-device FL common, cross-silo rare
  - Solutions are often bespoke, and not easily adaptable to different ML algorithms, different modalities of data, or different use cases
- **Opportunity:** encourage innovation that brings about efficient, performant, and adaptable federated solutions, that provide privacy guarantees across the ML lifecycle

Challenge Tracks

Track A: Transforming Financial Crime Prevention

- Solutions will leverage FL to enable analysis on synthetic datasets representing data held by the SWIFT payments network and datasets held by partner banks.
- Solutions will train a model to identify anomalous transactions, whilst preserving the confidentiality of various sensitive fields in the datasets.

Track B: Transforming Pandemic Forecasting

- Solutions will leverage FL to enable analysis on data partitioned across multiple units, which in the real world could be different hospitals, health districts etc.
- Solutions will train a model to predict an individual's risk to infection, whilst preserving the confidentiality of their health, demographic, and mobility data.

Challenge structure

Phase 1

White paper

You will develop a technical white paper that describes your proposed approach

Phase 2

Solution development

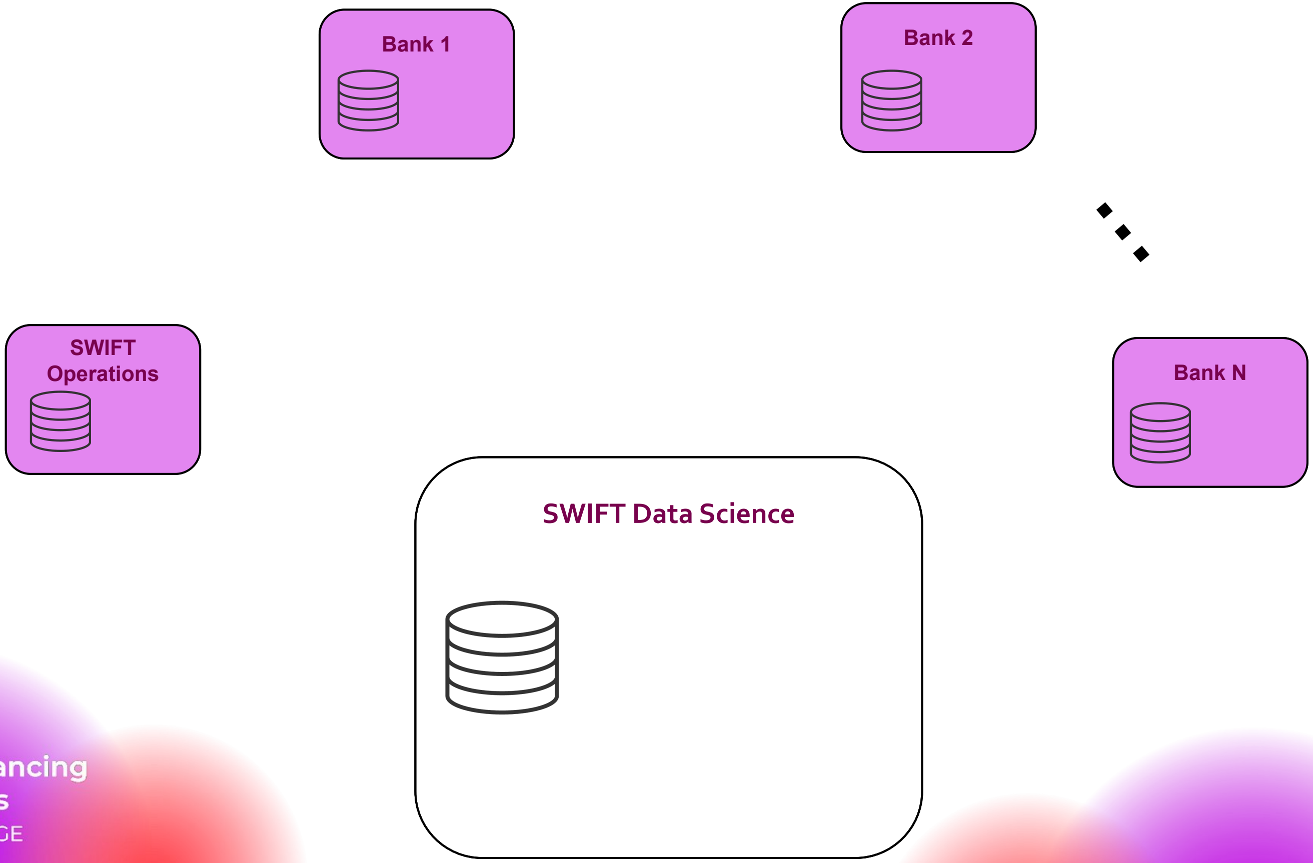
You will build and develop the solution proposed in your white paper

Phase 3

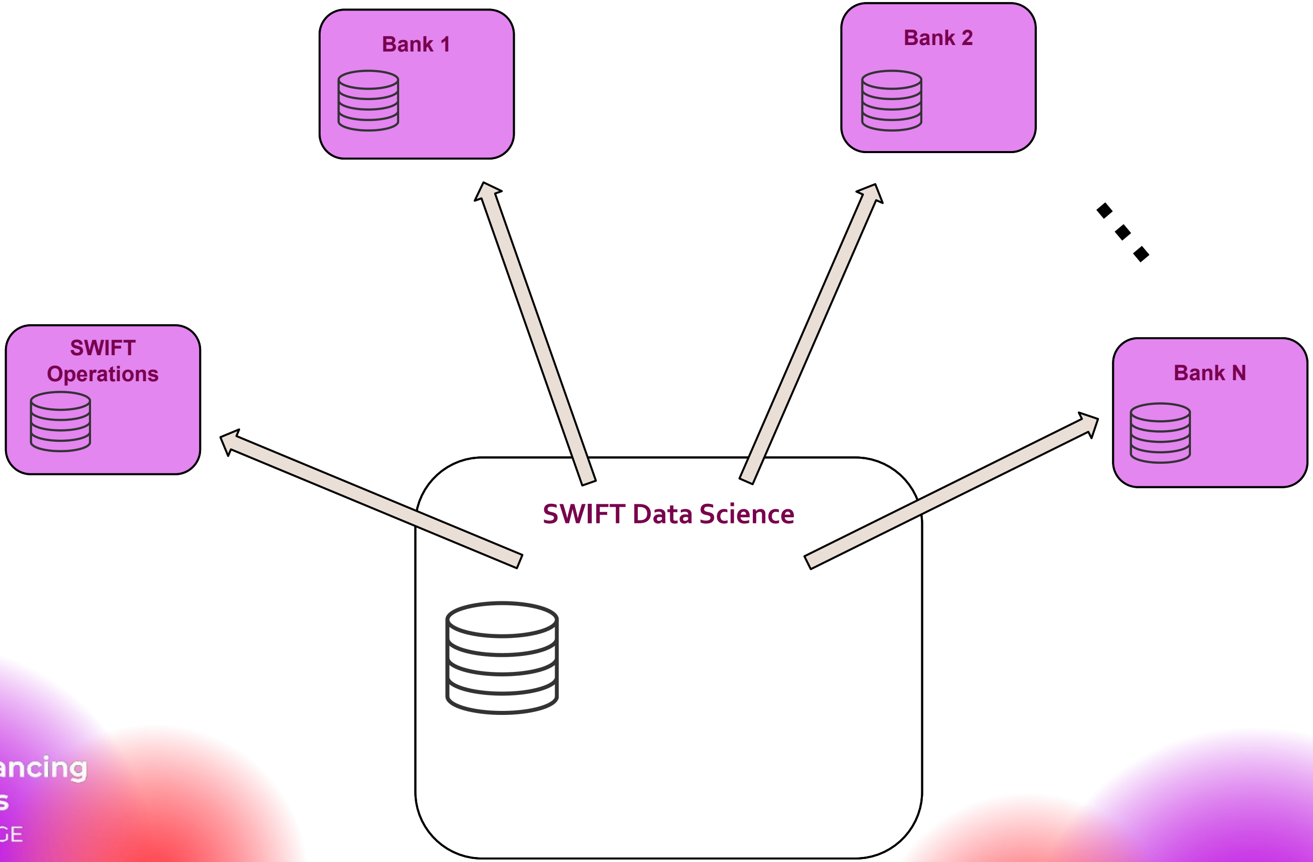
Red Teaming

The top solutions will be tested by competing red teams

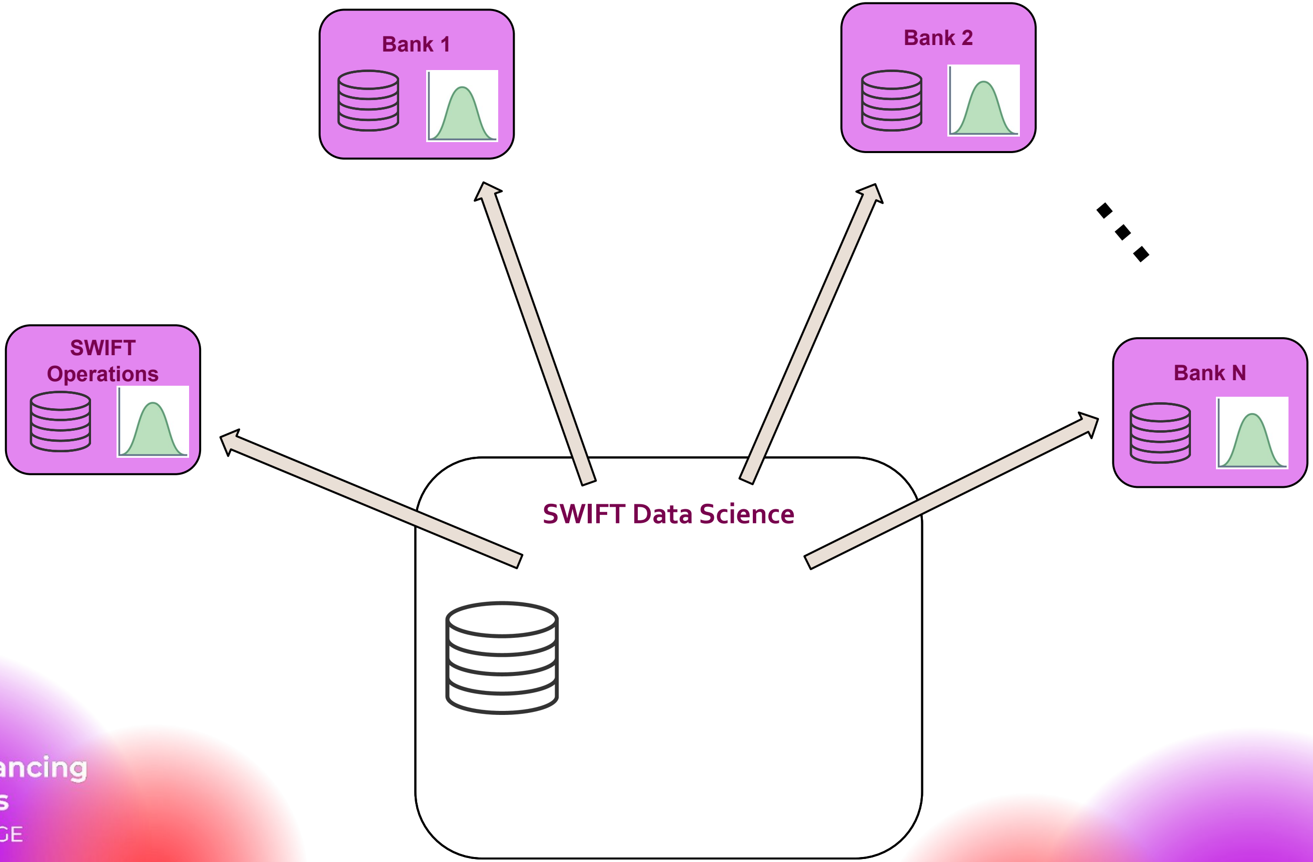
Federated learning…with SWIFT



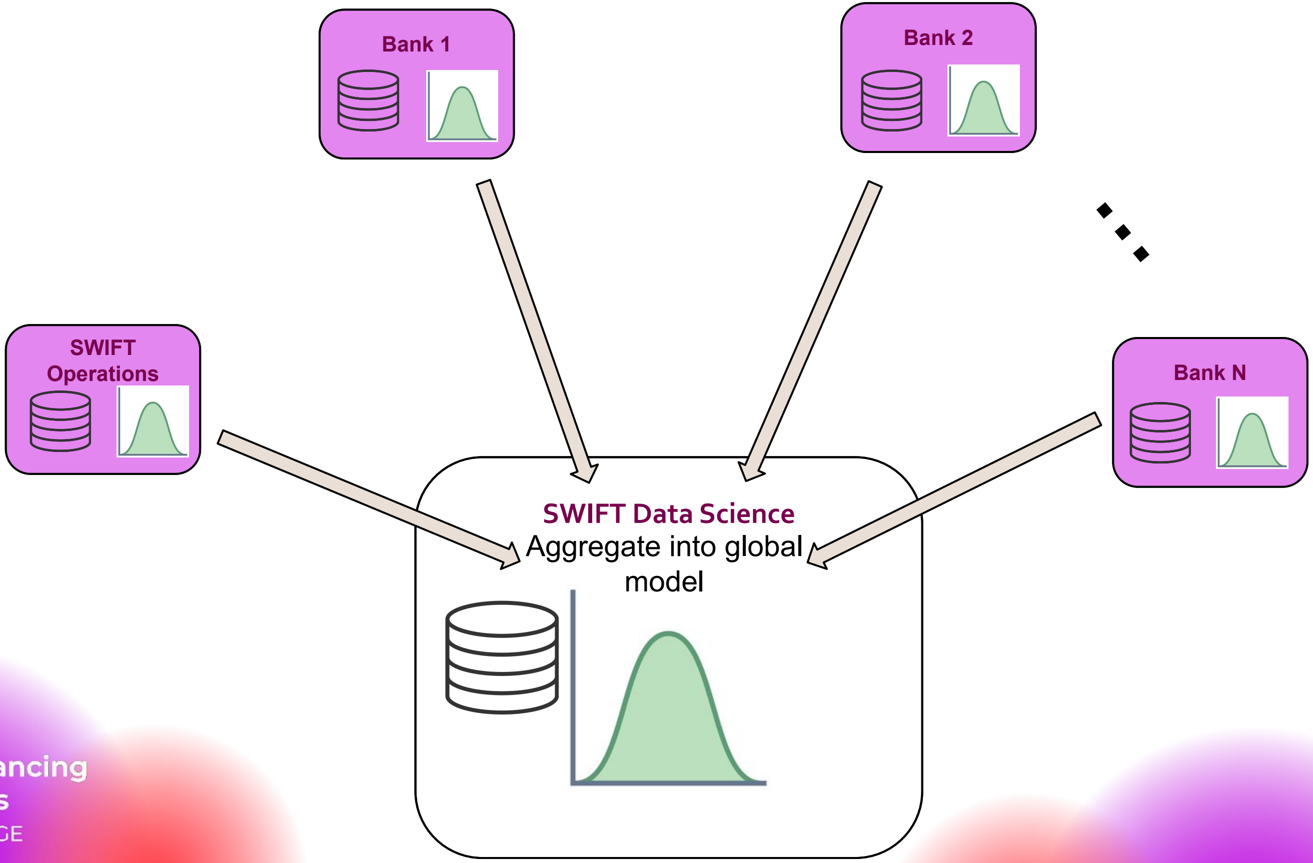
Federated learning...with SWIFT



Federated learning…with SWIFT



Federated learning…with SWIFT



Bank 1

Bank 2

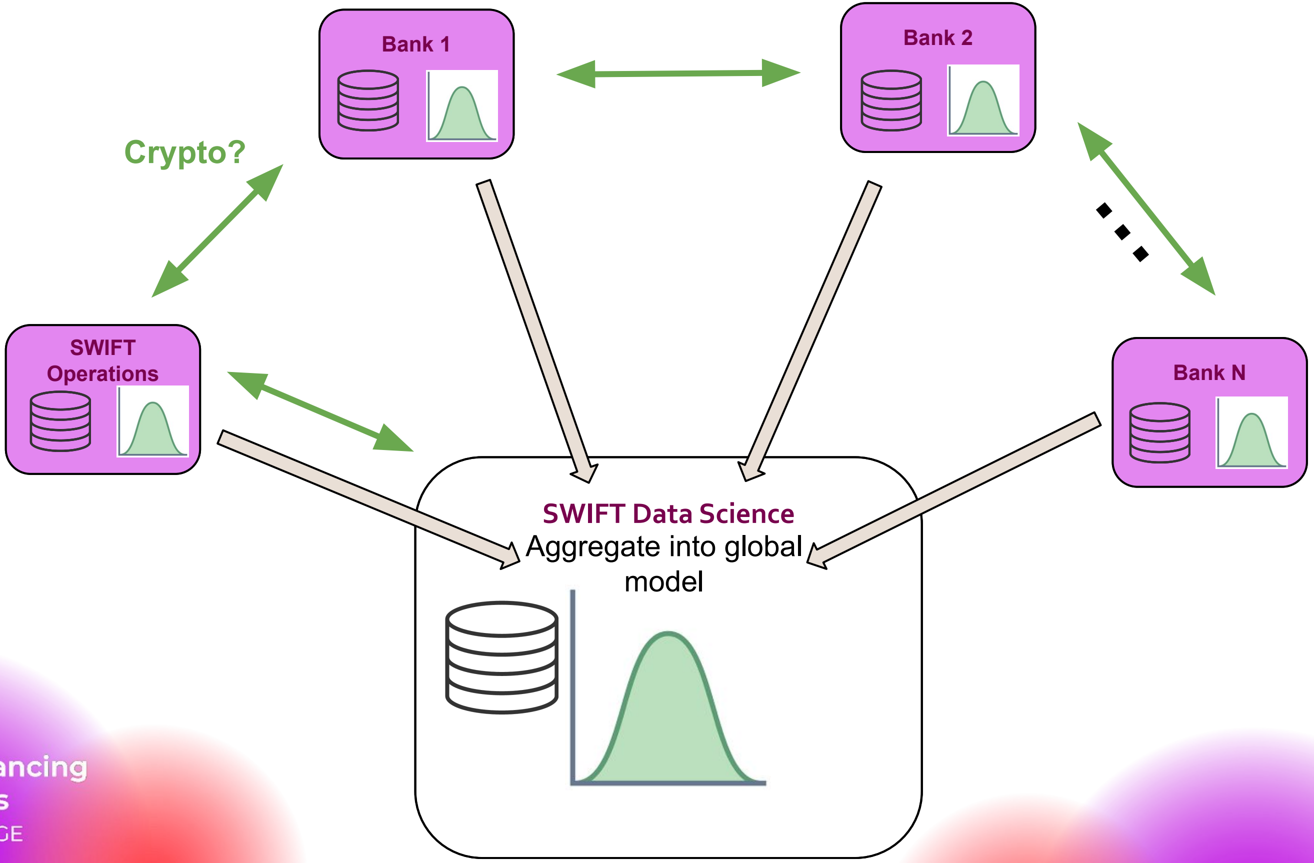
SWIFT Operations

Bank N

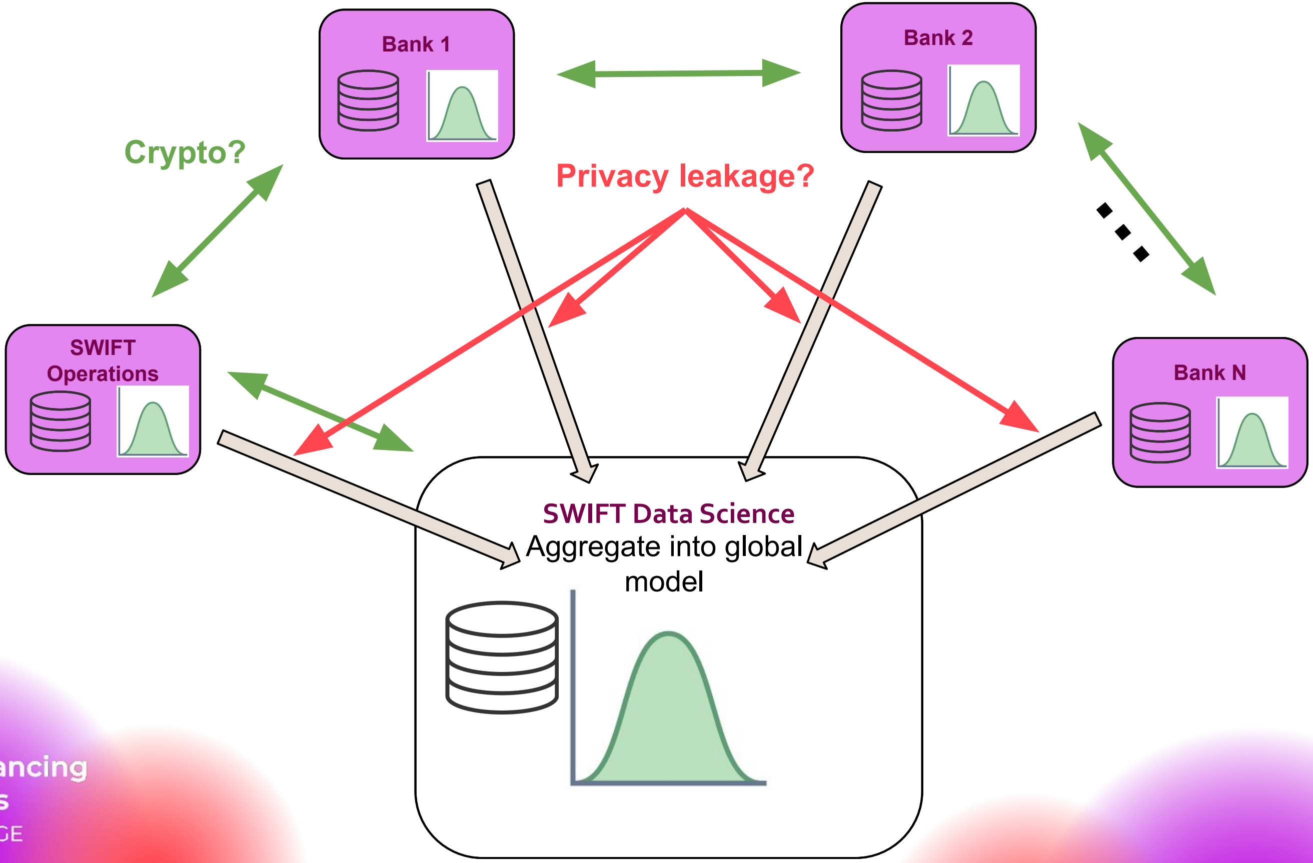
SWIFT Data Science

Aggregate into global model

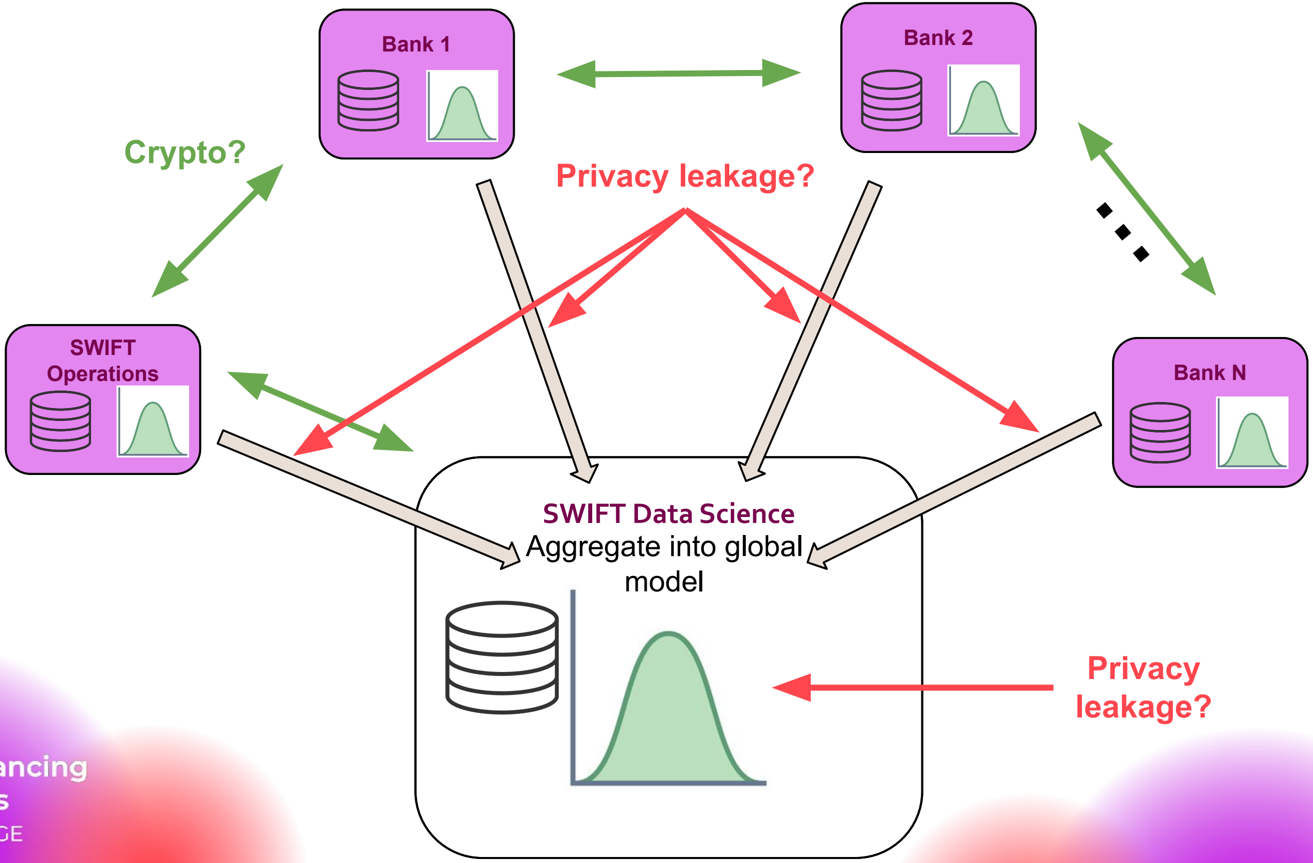
Federated learning…with SWIFT



Federated learning...with SWIFT



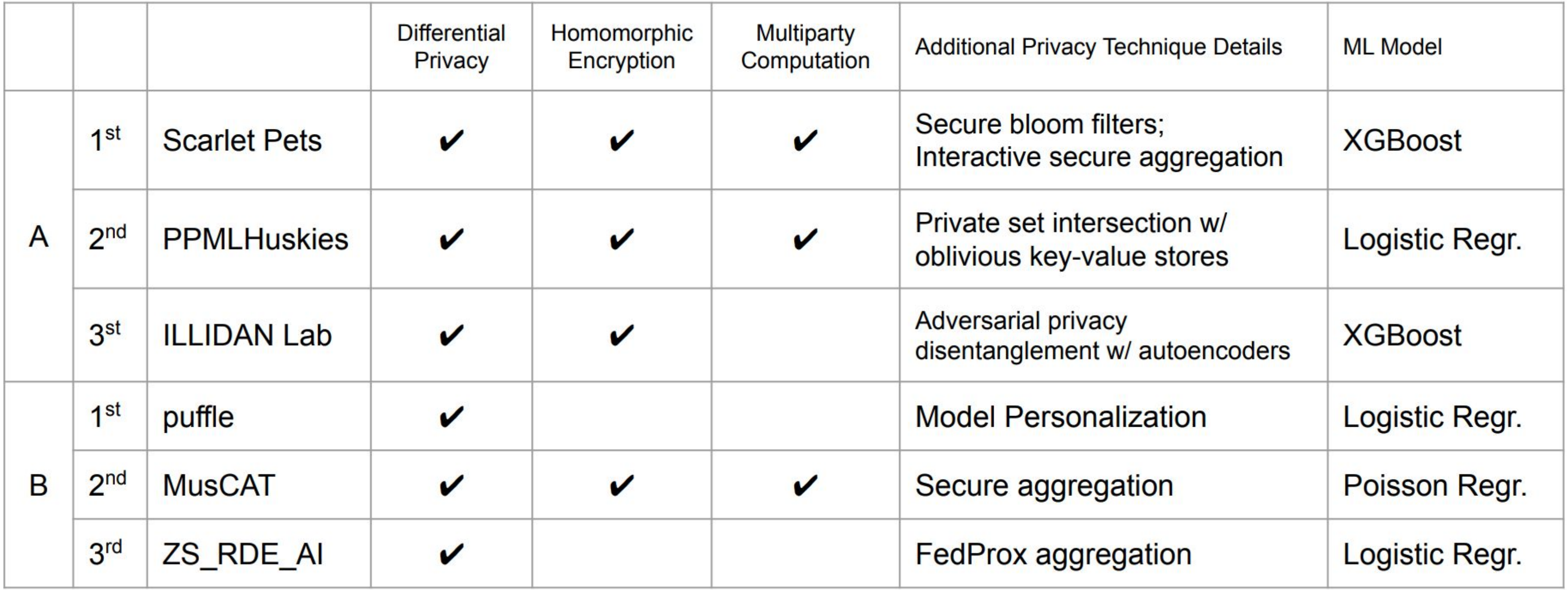
Federated learning...with SWIFT



Submissions

- **Phase 1**
    - 51 white paper submissions in finance
    - 39 white paper submissions in health
- **Phase 2:**
    - 12 solutions developed in finance
    - 7 solutions developed in health

Submissions



			Differential Privacy	Homomorphic Encryption	Multiparty Computation	Additional Privacy Technique Details	ML Model
A	1st	Scarlet Pets	✔	✔	✔	Secure bloom filters; Interactive secure aggregation	XGBoost
	2nd	PPMLHuskies	✔	✔	✔	Private set intersection w/ oblivious key-value stores	Logistic Regr.
	3st	ILLIDAN Lab	✔	✔		Adversarial privacy disentanglement w/ autoencoders	XGBoost
B	1st	puffle	✔			Model Personalization	Logistic Regr.
	2nd	MusCAT	✔	✔	✔	Secure aggregation	Poisson Regr.
	3rd	ZS_RDE_AI	✔			FedProx aggregation	Logistic Regr.

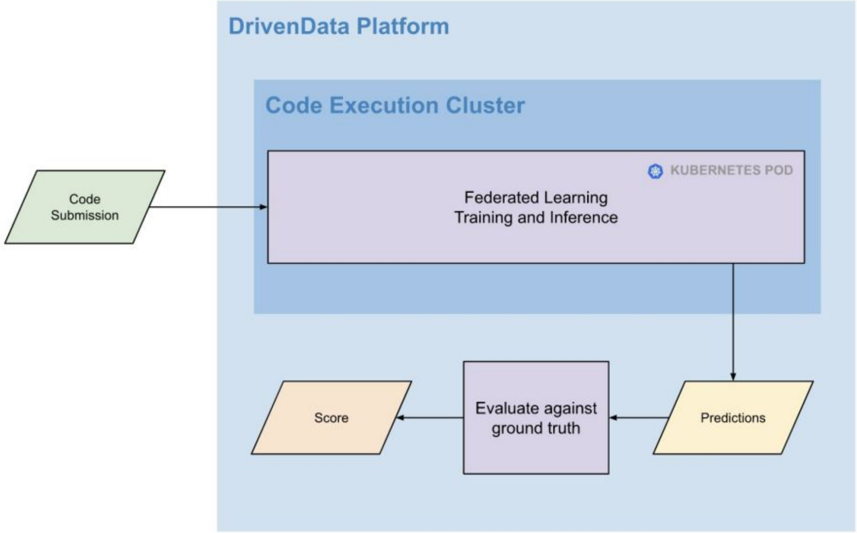
Privacy-Enhancing Technologies
PRIZE CHALLENGES

Challenge setup

Design of evaluation environment

Motivation

- Demonstrate that solutions have real, working software implementations
- Empirically measure solution runtime performance in a comparable way on common hardware (e.g., run time, memory usage, network usage)
- Allow red teams to apply attacks to actual implementation



Design of evaluation environment

Containerized execution in a Kubernetes cluster

- Azure Standard_NC6 nodes: 6 CPU cores, 56 GiB RAM, 1 Nvidia Tesla K80 GPU, 12 GiB GPU memory, 340 GiB disk

Simulated federated learning: runs on one node with parties simulated by multiprocessing

Participants supplied implementations of clients and aggregation strategies following API spec of the Flower federated learning framework

- Designed to help make solutions more comparable, including measurement of runtime performance metrics

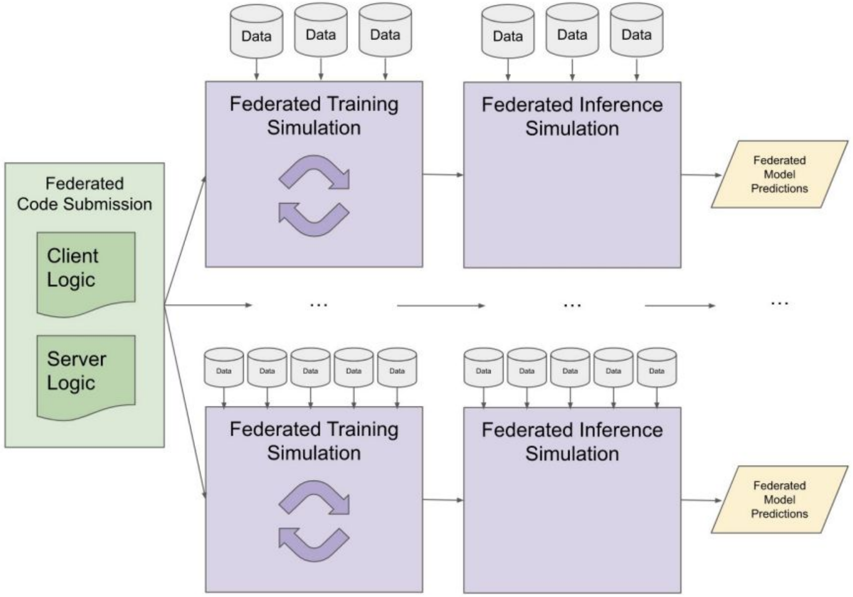

Federated Code Submission
Client Logic
Server Logic
Data
Federated Training Simulation
Federated Inference Simulation
Federated Model Predictions
...

Diagram showing code execution workflow, with evaluation on multiple federated partitioning scenarios.

Blue and Red teams

Blue teams received:

- Detailed problem description
- Access to training and test datasets
- Example code for a simple classifier
- Support through Slack and the DrivenData forum

Blue teams submitted:

- Their code
- A 10-page report describing their solution, experimental results etc.

Red teams received:

- Docker container of blue team’s solution
- Blue team’s code
- Blue team’s report

Red teams submitted:

- Their code
- A report describing their attacks, detailing results, and suggesting potential mitigations

Privacy-Enhancing
Technologies
PRIZE CHALLENGES

Solutions

Challenge Results

Privitar empowers organisations to use their data safely and ethically. Their modern data provisioning solution builds collaborative workflows and policy-based data privacy and access controls into data operations. Privitar combines technology, regulatory expertise, and best practices to support modern data innovation initiatives while navigating regulations and protecting customer trust.

University of Cambridge – this team comes from CaMLSys – the Cambridge Machine Learning Systems lab, based in the Computer Science and Technology department. The team is composed of Professor Nicholas D. Lane, Senior Research Associate, Pedro Porto Buarque de Gusmão, and PhD and Masters students

Faculty is a founder-led company. Since the beginning, their mission has been to bring the benefits of artificial intelligence to everyone. They have worked on some of the biggest and most difficult challenges faced by major organisations, using 'decision intelligence' to help organisations make better decisions on the things that matter.

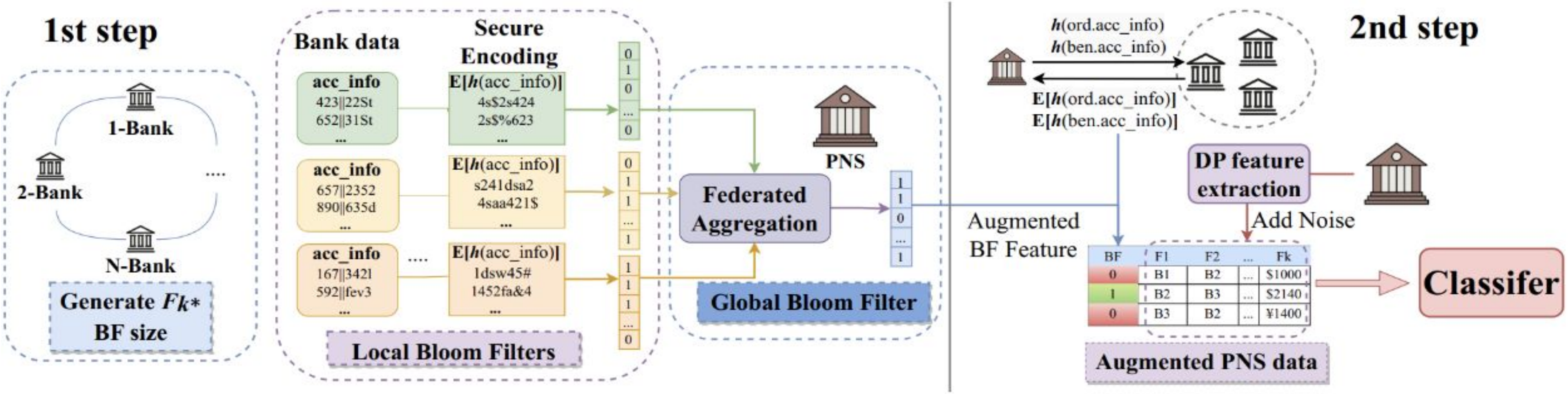
Trūata is a PET company that specialises in quantifying the privacy risk in datasets and ensuring analytical outputs meet required privacy thresholds. Trūata's red team for this challenge contains six data scientists with specialties in data privacy, privacy engineering, machine learning and statistical analysis. Along with a vast amount of experience in executing adversarial attacks on outputs created using privacy-enhancing technologies.

https://github.com/usnistgov/PrivacyEngCollabSpace/tree/master/tools/de-identification

Challenge Solutions – Bloom Filters

Step 1: Privacy preserving feature mining through hashing encryption. This is followed by creating bloom filters which allow efficient lookups as to whether a feature is present. All done at the local level.

Step 2: Aggregation of bloom filters to central model. Classifier able to train on these rule-based features



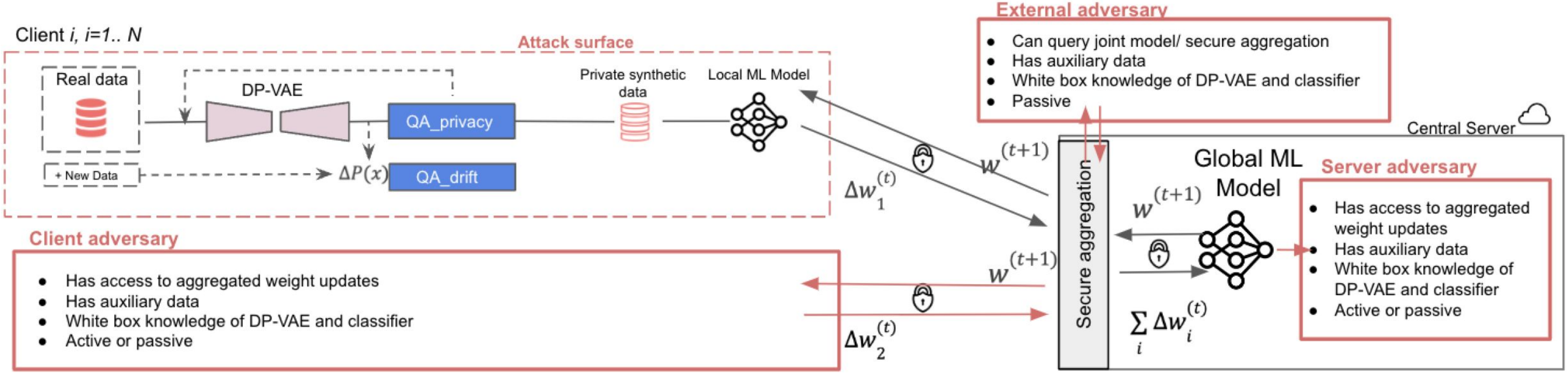
BF	F1	F2	...	Fk
0	B1	B2	...	$1000
1	B2	B3	...	$2140
0	B3	B2	...	¥1400

https://github.com/idsla/Scarlet-PETs
https://rutgers.app.box.com/s/q84zjo3edv5d1e1eu67ypihiw8cb2djq

+ Little drop in accuracy between centralised and federated approaches.
- Sensitive to the features mined.

Challenge Solutions – Synthetic Data

Step 1: A Variational Autoencoder with differential privacy is applied to a training subset of the data.
Step 2: A Privacy and drift gateway implement adversarial attacks and monitor the impact of new data upon the generator
Step 3: A local model is trained on the synthetic data
Step 4: SecAgg and FedOpt is used to average the model weights in an iterative cycle where the weights are passed back and forth between local and global models till a convergence is reached.



External adversary
- Can query joint model/ secure aggregation
- Has auxiliary data
- White box knowledge of DP-VAE and classifier
- Passive

Server adversary
- Has access to aggregated weight updates
- Has auxiliary data
- White box knowledge of DP-VAE and classifier
- Active or passive

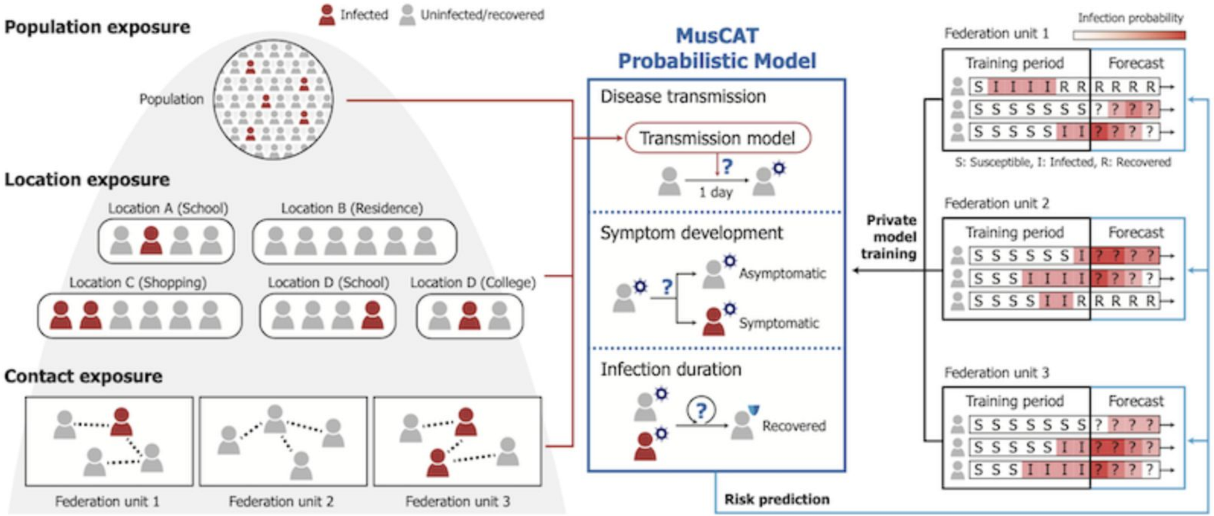
Client adversary
- Has access to aggregated weight updates
- Has auxiliary data
- White box knowledge of DP-VAE and classifier
- Active or passive

+ Modular and Extensible
- Dependent on quality of synthetic data privacy-utility balance

Challenge Solutions – HE and Differential Privacy

This solution uses a multi-level SIR model using DP-SGD and the CoinPress algorithm to guarantee the training is DP.

SecAgg using Multiparty Homomorphic Encryption is used for aggregating information between the local and global models



Federation unit 1

Training period							Forecast			
S	I	I	I	I	R	R	R	R	R	R
S	S	S	S	S	S	S	?	?	?	?
S	S	S	S	S	I	I	?	?	?	?

Federation unit 2

Training period							Forecast			
S	S	S	S	S	S	I	?	?	?	?
S	S	S	I	I	I	I	?	?	?	?
S	S	S	S	I	I	R	R	R	R	R

Federation unit 3

Training period							Forecast			
S	S	S	S	S	S	S	?	?	?	?
S	S	S	S	S	I	I	?	?	?	?
S	S	S	I	I	I	I	?	?	?	?

https://github.com/hhcho/muscat
https://www.dropbox.com/s/dzyc8himjtcu05j/PETsChallenge_MusCAT_Report.pdf?dl=0

+ Combination of many techniques
- Probabilistic modelling – limited overall forecasting accuracy

Red teaming

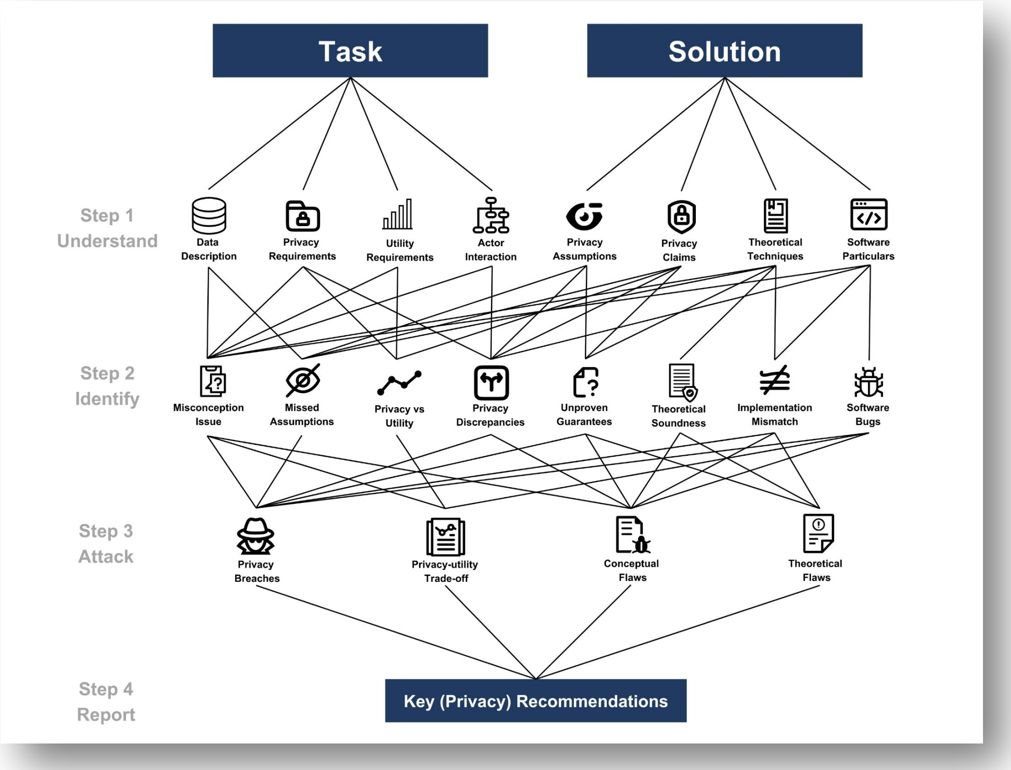
Approaching the red teaming task

- White-box attacks - red teams first studied the blue team code and reports in order to identify key privacy claims, and potential areas of weakness

- Successful red team attacks were based on:
  - Incorrect privacy claims
  - Incorrect assumptions
  - Incorrect implementations

- Many of the solutions fell down due to miss-representing their implementation or over-stating a privacy claim.

See: https://www.sri.inf.ethz.ch/blog/fedcomp



Reflections from an assessor

In the 2022/23 challenge one of the most successful elements was the red teaming phase as most solutions were found to have vulnerabilities and issues leading to privacy leakage.

A key statement from the US red team was “In fact, we were surprised to find how many solutions were already flawed due to misunderstandings regarding the precise requirements or the interactions between parties.” A key to the red team’s success was to go after the privacy claims and assumptions of the associated solutions.

A project implementing PETs needs to state an “initial” privacy-utility curve alongside a clear description of the data (including identified sensitive information) and infrastructure that the data sits within. Ideally the red team would have a say in the setup of the baseline data store before the PET is implemented in order to highlight techniques specific vulnerabilities.

A project implementing PETs should require compartmentalisation of different parts where possible to support the evaluation of individual aspects

Stating levels of privacy versus threat models would be useful e.g. for one sensitive variable this technique is safe but for two it’s not. Therefore, in normal times this can’t be used but if the privacy model is lowered then this technique would become appropriate.

Examples of shortcomings

- DP protecting against membership inference attacks at local node level, but not across the system - able to perform attacks with 74% accuracy
- Unrealistic threat models used by blue teams
- Privacy budgets set primarily to get sufficient accuracy

Every blue team solution was found to have issues. You wouldn’t know this from reading the blue team reports...

Privacy-Enhancing Technologies
PRIZE CHALLENGES

Lessons learned

Technical Learnings

- **Range of Solutions** - Level of variation makes it difficult to know where to focus
- **Innovation mainly in implementation rather than technique -** Mainly due to time but shows the infrastructure and context specific issues are more than the technical
- **Difficult to lift and shift current tooling -** Suggests techniques are not as robust/interoperable as hoped
- **Combination is key** - e.g. FedAvg with SMPC and DP
- **Differential Privacy is King** – However, often need to know rare cases so these can’t be noised out
- **Red Team attacks were successful –** emphasises the need for this step
- **Assessment needs data science, math, computer science, privacy, PETs, and domain knowledge** - very difficult to evaluate, and to get sign-off in a real world deployment setting
- **Often unclear responsibility** - Who is responsible for end-to-end (storage, infrastructure, technique, users, …)

Challenges of running a challenge

- Determining area of focus challenging
  - Engage with domain experts early and often

- Finding good data is hard
  - Identifying interesting datasets in health and finance challenging
  - Generating synthetic data at short notice led to data quality issues

- Hard to know a priori if the use case and dataset is amenable to federated learning
  - We had highly imbalanced datasets - mostly outliers, which makes it challenging to be privacy-preserving e.g. with DP you essentially need to bring outliers closer to the mean
  - Datasets were large, training times long

- Highly challenging to evaluate solutions fairly
  - Efforts to simplify evaluation potentially made things more challenging - may have been better embracing subjectivity

- Difficult to recruit red teams (small market), limited time for red teams to design and run their attacks
  - Implementation issues vs design issues

Final takeaways

- US applicants were generally a lot stronger
- Solutions submitted by academia were generally stronger than those submitted by industry
- Lots more work is needed! We need to:
  - build more tooling
  - develop better frameworks and standards
  - fund foundational research
  - build a thriving assurance and audit market that includes privacy audit and assurance techniques like red teaming
- Community building hugely successful and important - we had over 200 people attend the post-challenge Demo Day at the Royal Society in London in May 2023

PPMLHuskies and the
Financial Crime Track
Martine De Cock, Anderson Nascimento,
Sikha Pentyala, Steven Golob, Dean Kelley,
Zekeriya Erkin, Jelle Vos, Célio Porsius Martins,
Ricardo Maia
UNIVERSITY of WASHINGTON
TUDelft
UnB
THE ROYAL SOCIETY

Privacy-Enhancing Technologies
PRIZE CHALLENGES

Moving forward

- Technical blog series on privacy-preserving federated learning in collaboration with NIST **starting this Thursday**

- Continued UK-US collaboration through the Atlantic Declaration

- Privacy a key component of President Biden’s recent Executive Order on AI:
  - NIST mandated to “create guidelines for agencies to evaluate the efficacy of differential-privacy-guarantee protections”
  - NSF mandated to “fund the creation of a Research Coordination Network (RCN) dedicated to advancing privacy research and, in particular, the development, deployment, and scaling of PETs.”

- UK AI Safety Institute are working with OpenMined develop and deploy technical infrastructure that will facilitate AI safety research across governments and AI research organisations

Prize challenge resources

- Challenge home page
- Video announcing the results of the challenges (well worth a watch)
- UK winners' blog
- US winners' blog
- Open-sourced US solutions in NIST's Privacy Engineering Collaboration Space
- Synthetic population dataset used for pandemic track - available open access
- Technical briefs for financial crime track and pandemic response track

Privacy-Enhancing Technologies
PRIZE CHALLENGES

Thank you! Any questions?

gov.uk/cdei
linkedin.com/in/davidbuckley3
david.buckley@cdei.gov.uk