Imperial College London

Presentation for AIChain

October 2020

2CP

Decentralised protocols to transparently evaluate contributivity in Blockchain Federated Learning environments

Harry Cai, Daniel Rueckert & Jonathan Passerat-Palmbach

Imperial College London

Introduction

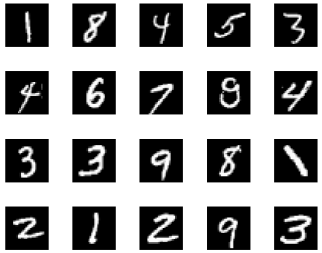
- Federated Learning harnesses data from multiple sources to build a model.
- As a model is trained using Federated Learning, how does its ownership change?
  - Can we use blockchains to determine the evolving ownership of such a model?
- *2CP = Crowdsource Protocol + Consortium Protocol*
- These are two new protocols for *blockchained Federated Learning*

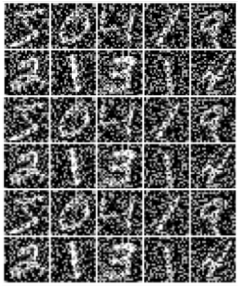
[Hartmann, 2018]

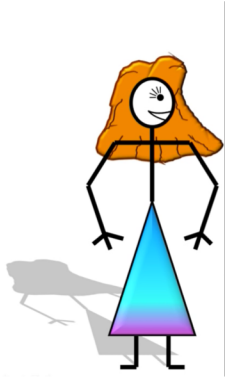
Imperial College London

Data Contributivity

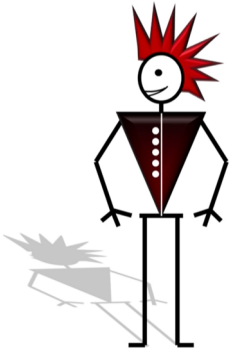
- Contributors may need incentives to participate
- Those with more to contribute may need larger incentives
- But how do we quantify contributivity?



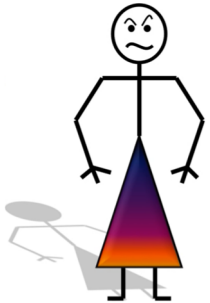




Carol

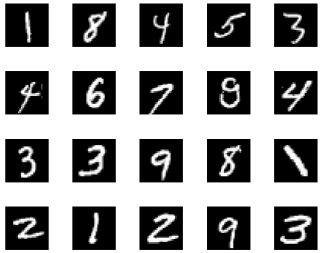


David

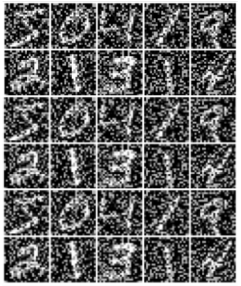


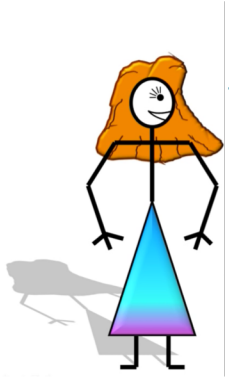
Eve

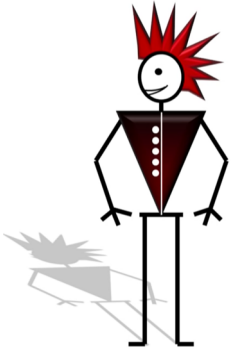
[Artwork: Creature Mann]

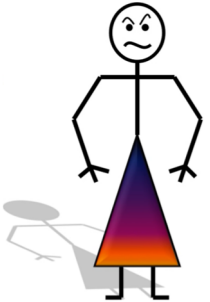


How will we split the profits?





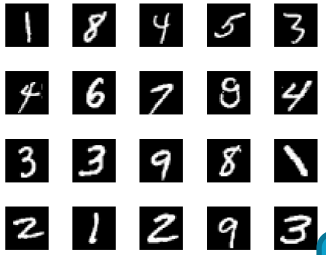




Carol

David

Eve

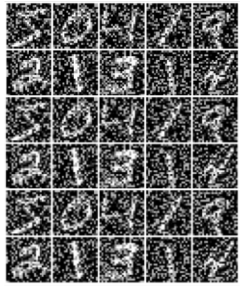


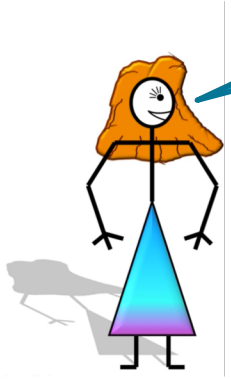
1/3 each!

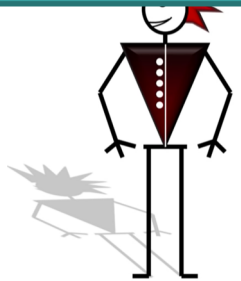
No, let's split by quantity!

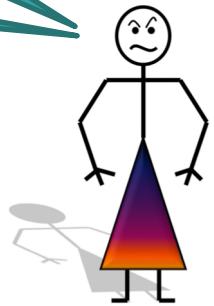
No, your data is bad quality!

My data is *unique* and *novel*!







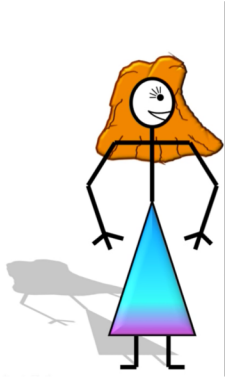


Carol

David

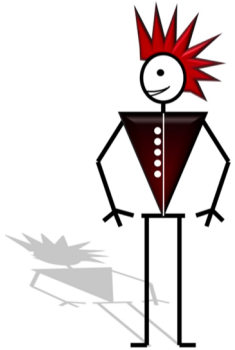
Eve

?



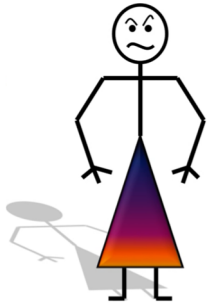
Carol

??



David

???

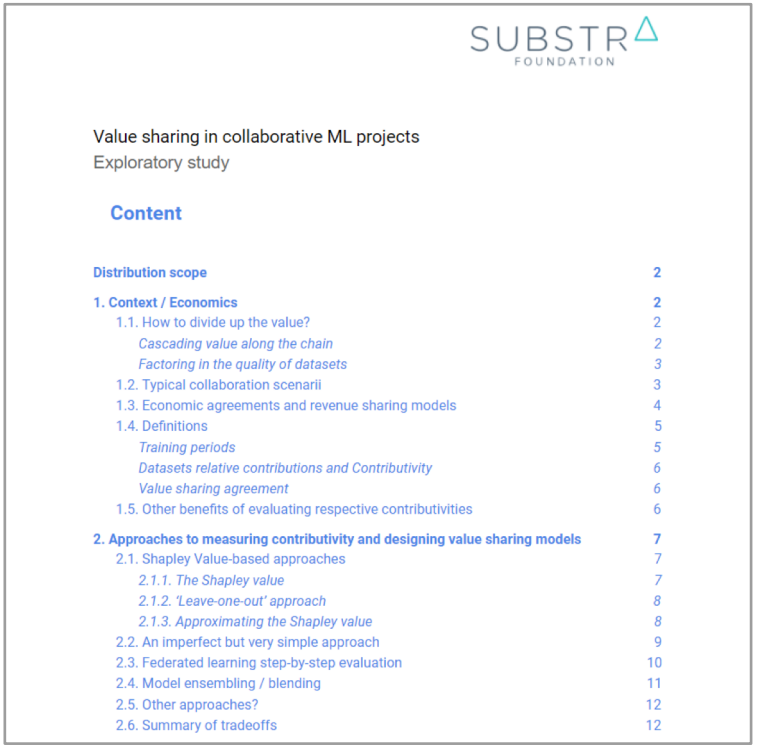


Eve

Imperial College London

Data Contributivity

- Related work
  - Heuristics
  - Shapley Value
  - Step by step evaluation



SUBSTRA FOUNDATION

Value sharing in collaborative ML projects

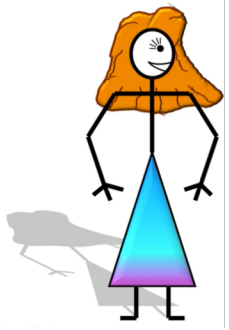
Exploratory study

Content

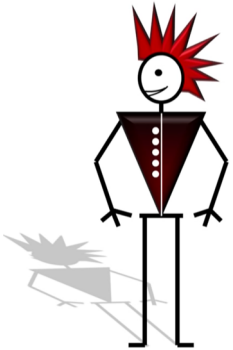
[Substra, Value Sharing in Collaborative ML Projects]

Accuracy

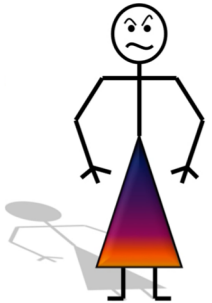
(none):	0
Carol:	0.96
David:	0.84
Eve:	0.11
Carol & David:	0.97
Carol & Eve:	0.93
David & Eve:	0.77
Carol & David & Eve:	0.99



Carol



David

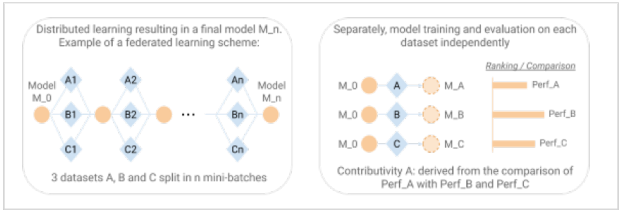


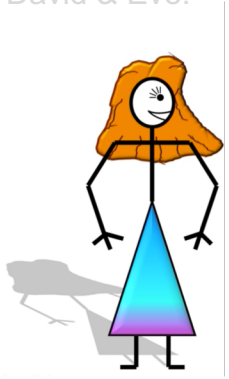
Eve

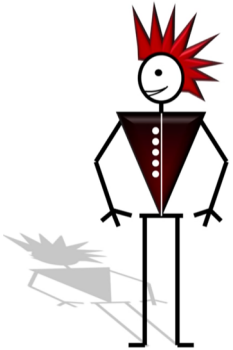
Accuracy

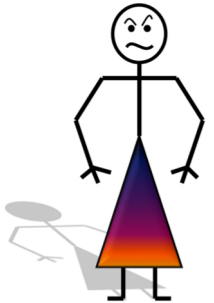
(none):	0
Carol:	**0.96**
David:	**0.84**
Eve:	**0.11**
Carol & David:	0.97
Carol & Eve:	0.93
David & Eve:	0.77
Carol & David & Eve:	0.99

‘Single partner run’









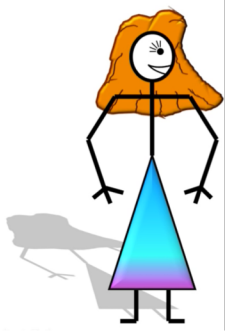
Carol
0.96

David
0.84

Eve
0.11

Accuracy

(none):	0
Carol:	0.96
David:	0.84
Eve:	0.11
Carol & David:	**0.97**
Carol & Eve:	**0.93**
David & Eve:	**0.77**
Carol & David & Eve:	**0.99**



Carol
0.22

‘Leave one out’

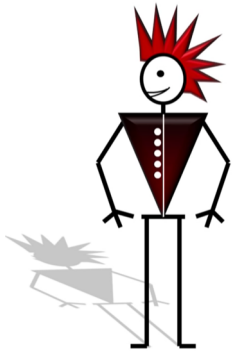
2.1.2. ‘Leave-one-out’ approach

A cheaper evaluation would be to consider only one subset instead of all subsets. For a given participant A, the value we compute would then be:

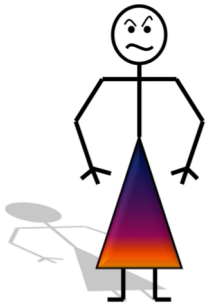
$$Contributivity\ C_A = Perf_{All} - Perf_{All \setminus A}$$

With such a contributivity measurement methodology, measure for one participant would cost one training on all participants but one to determine C_A (as we consider that $Perf_{All}$ is evaluated whatsoever, it is the core objective of the ML project).
In consequence the evaluation of all N participants respective contributivities measures would cost N times an $(N-1)$-large federated learning run. This is still quite expensive computationally-wise with regard to computational cost of the core objective of the project.

Although this simpler approach seems intuitive, **it is not satisfactory and doesn't reproduce the good properties of the Shapley value**. A good example to realise this is the case were in a given project, among the datasets used, two of them are (almost) identical. In such a case, the respective contributivities of these two will be (close to) zero.



David
0.06



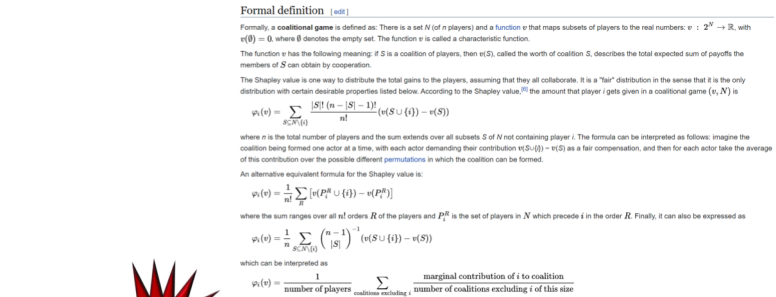
Eve
0.02

[Substra, Value Sharing in Collaborative ML Projects]

Accuracy

(none):	**0**
Carol:	**0.96**
David:	**0.84**
Eve:	**0.11**
Carol & David:	**0.97**
Carol & Eve:	**0.93**
David & Eve:	**0.77**
Carol & David & Eve:	**0.99**

Shapley Value



Formal definition [edit]

Formally, a **coalitional game** is defined as: There is a set N (of n players) and a function v that maps subsets of players to the real numbers: $v : 2^N \to \mathbb{R}$, with $v(\emptyset) = 0$, where $\emptyset$ denotes the empty set. The function v is called a characteristic function.

The function v has the following meaning: if S is a coalition of players, then $v(S)$, called the worth of coalition S, describes the total expected sum of payoffs the members of S can obtain by cooperation.

The Shapley value is one way to distribute the total gains to the players, assuming that they all collaborate. It is a "fair" distribution in the sense that it is the only distribution with certain desirable properties listed below. According to the Shapley value,[6] the amount that player i gets given in a coalitional game (v, N) is

$$\varphi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(n - |S| - 1)!}{n!} (v(S \cup \{i\}) - v(S))$$

where n is the total number of players and the sum extends over all subsets S of N not containing player i. The formula can be interpreted as follows: imagine the coalition being formed one actor at a time, with each actor demanding their contribution $v(S \cup \{i\}) - v(S)$ as a fair compensation, and then for each actor take the average of this contribution over the possible different permutations in which the coalition can be formed.

An alternative equivalent formula for the Shapley value is:

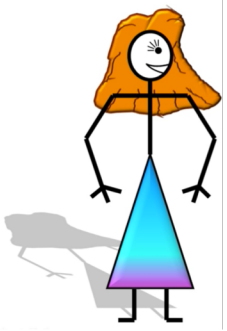
$$\varphi_i(v) = \frac{1}{n!} \sum_{R} \left[v(P_i^R \cup \{i\}) - v(P_i^R) \right]$$

where the sum ranges over all $n!$ orders R of the players and P_i^R is the set of players in N which precede i in the order R. Finally, it can also be expressed as

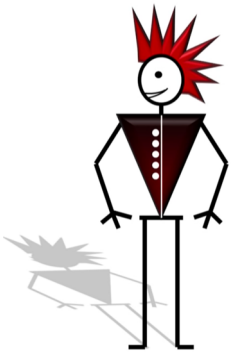
$$\varphi_i(v) = \frac{1}{n} \sum_{S \subseteq N \setminus \{i\}} \binom{n-1}{|S|}^{-1} (v(S \cup \{i\}) - v(S))$$

which can be interpreted as

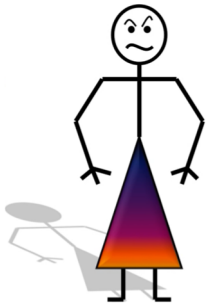
$$\varphi_i(v) = \frac{1}{\text{number of players}} \sum_{\text{coalitions excluding } i} \frac{\text{marginal contribution of } i \text{ to coalition}}{\text{number of coalitions excluding } i \text{ of this size}}$$



Carol
0.55



David
0.41

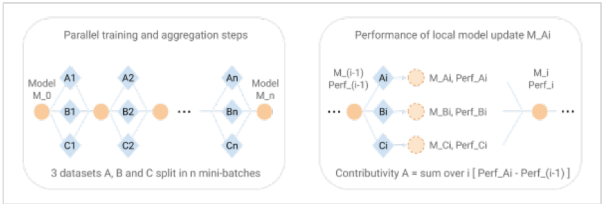


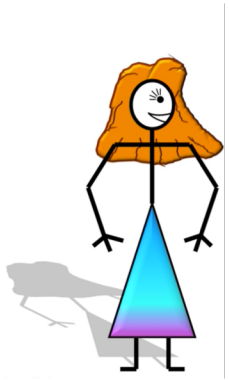
Eve
0.03

Accuracy change for ith iteration

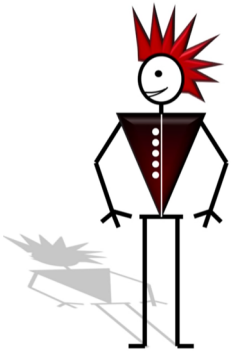
(none):	+0.
Carol:	+0.02
David:	+0.01
Eve:	-0.02

Step by step evaluation

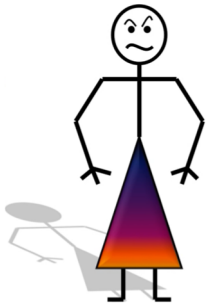




Carol



David



Eve

[Substra, Value Sharing in Collaborative ML Projects]

Imperial College London

Summary: data contributivity

- Can’t determine trainers’ contributivity before training.
- Instead calculate it afterwards, by analysing the effect of their model updates on performance.
- We advocate *step-by-step evaluation* as our contributivity metric, as it fits naturally into the Federated Learning process.

Imperial College London

Blockchain for decentralised Federated Learning

- Problems with FL:
  - Centralised
  - Lack of incentive
- But with blockchain:
  - Decentralised – no single point of failure
  - Trustless, immutable ledger where contributions can’t be censored
  - Automatically allocate shares in the final model

Imperial College London

Blockchain for decentralised Federated Learning

- Common features in related work
  - *BlockFL* architecture: key ideas
    - Updates as transactions
    - New training round each block / each *n* blocks
    - Use IPFS, store CIDs on-chain
    - Clients perform model aggregations themselves
  - Smart contracts to manage FL process

[Kim] [Nagar] [Weng] [Zhao]

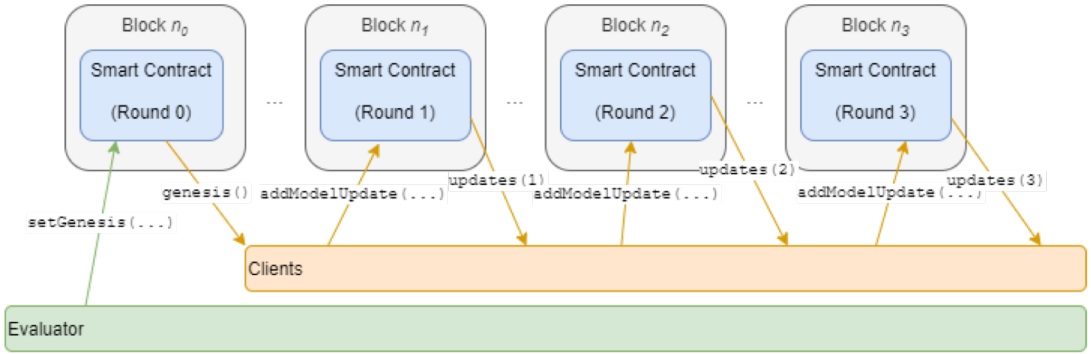
Imperial College London

Crowdsource Protocol

- For the Crowdsource Setting
  - Central organisation (‘Alice’)
  - Own a high quality data set
    - Ideal as a test set
    - But not large enough to train a model
  - Draw on data from outside sources (‘Bob, Carol et al’)
  - Evaluates each contributor using their test data

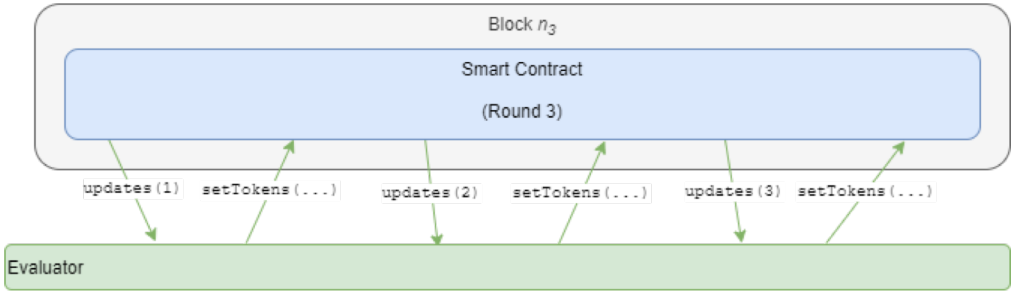
Imperial College London

Crowdsource Protocol



Imperial College London

Crowdsource Protocol



Imperial College London

Crowdsource Protocol

- Only works if the test set is high quality and representative
- But allows organisations to use their entire dataset as the test set

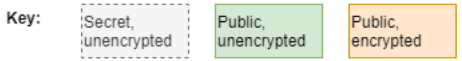
Imperial College London

Verified Evaluation

- Problem: results are only fair if Alice follows the protocol honestly
- Solution: *verified evaluation*

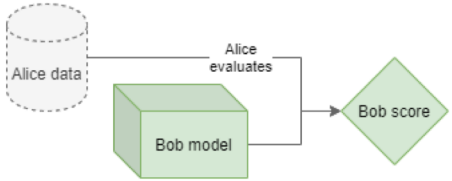
Imperial College London

Verified Evaluation

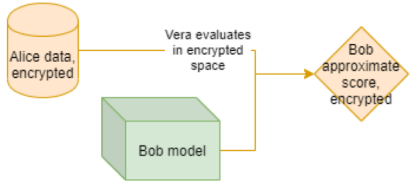


Step 1. Before evaluating anything, Alice publicly commits to her dataset.
Using a fully homomorphic encryption scheme, she encrypts her data and publishes the encrypted data.

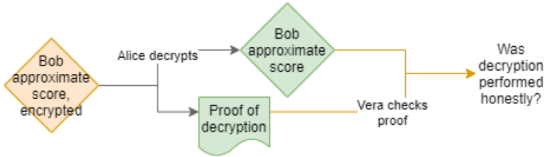
Step 2. Alice evaluates Bob's model using her private data, and publishes the result. Suppose Bob disputes the result.



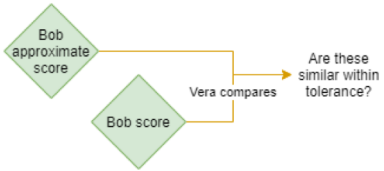
Step 3. Vera, an independent and trusted third party, evaluates Bob's model using Alice's encrypted data in the encrypted space. This is possible because the encryption scheme is fully homomorphic.



Step 4. Alice takes Vera's result and decrypts it, providing a ZKP of correct decryption. Vera checks the ZKP to ensure that the decryption was performed honestly.



Step 5. If Vera finds a match, Alice decrypted honestly so we have a genuine approximate score. If the approximate score is within a set tolerance of the original, then the original score is deemed honest.



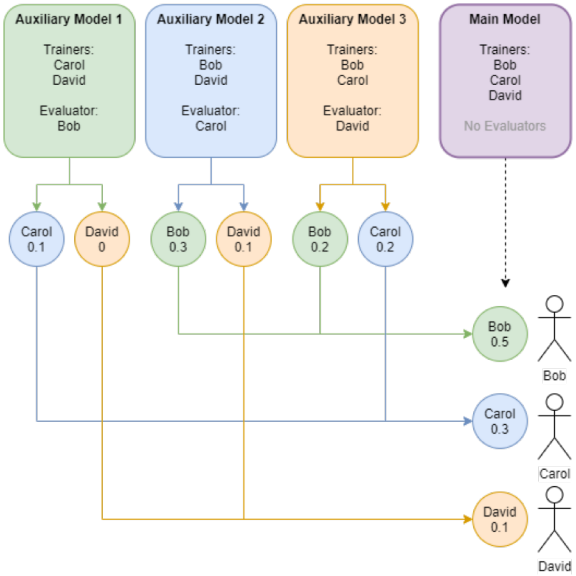
Imperial College London

Consortium Protocol

- For the Consortium Setting
  - Multiple, equal organisations (‘Bob, Carol et al’)
  - Collectively train a model
  - Collectively evaluate each others’ contributions

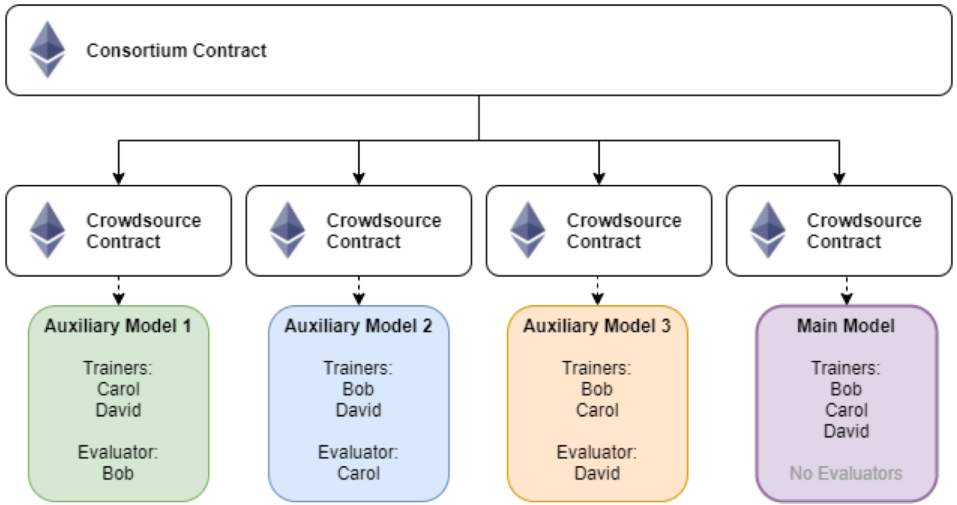
Imperial College London

Consortium Protocol



Imperial College London

Consortium Protocol



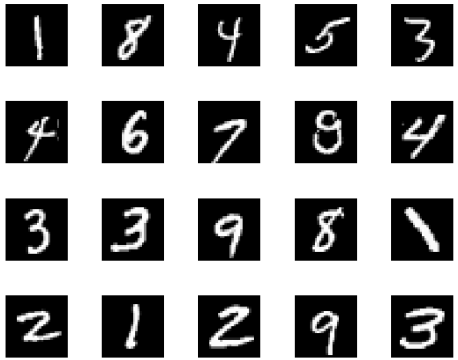
Imperial College London

Consortium Protocol

- Limitations
  - Closed process, participants determined beforehand
  - Must train *N+1* models for *N* clients
    - *N* ‘wasted’ models

Imperial College London

Experiments (MNIST)



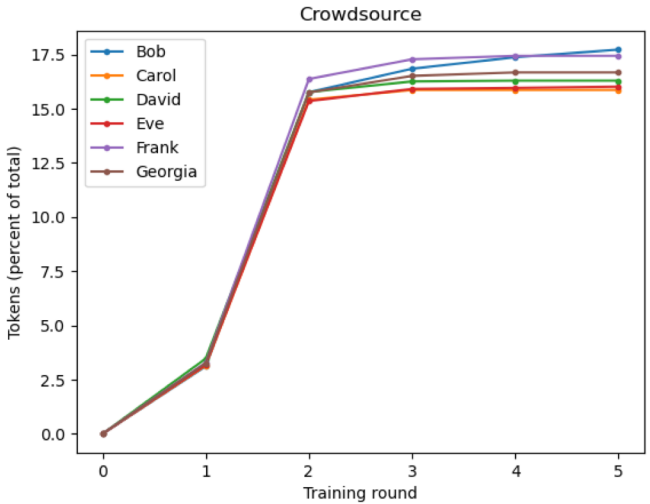
Imperial College London

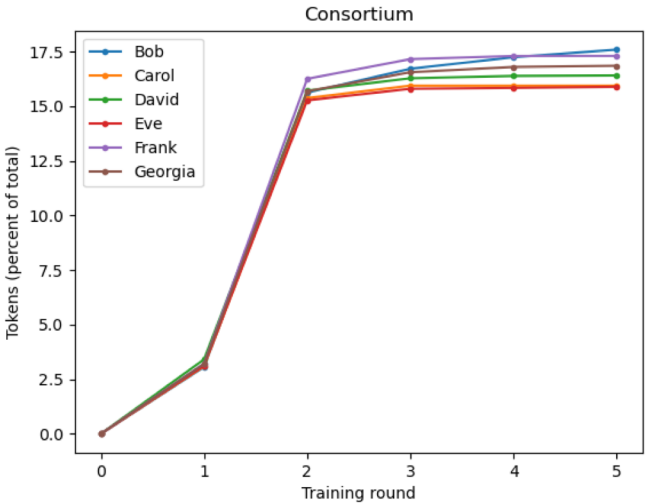
Experiments (MNIST)

- Used MNIST train set (x60,000) and test set (x10,000)
- Crowdsource Protocol:
  - The test set belongs to Alice, who is the evaluator
  - The train set is split between Bob et al, who are the trainers
- Consortium Protocol:
  - Alice and the test set are not used
  - Bob et al split the train set and form the consortium

Imperial College London

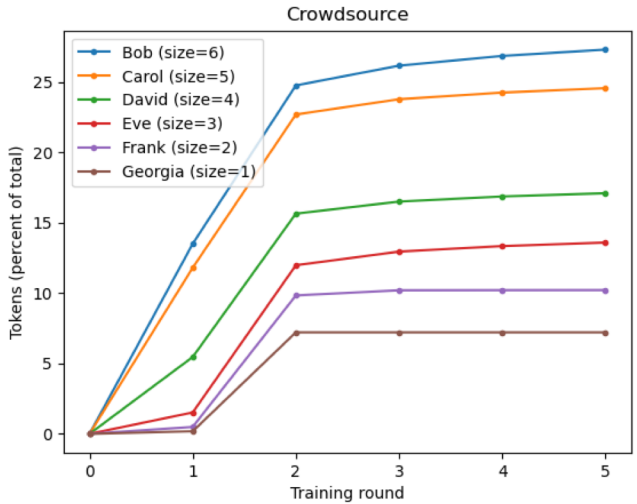
MNIST Test A (‘equal’)

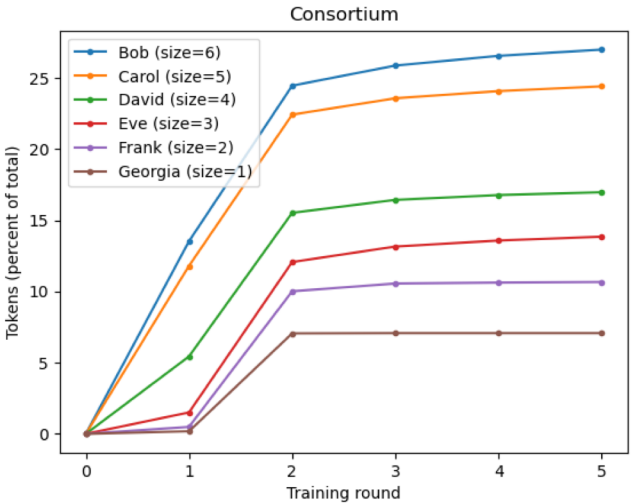




Imperial College London

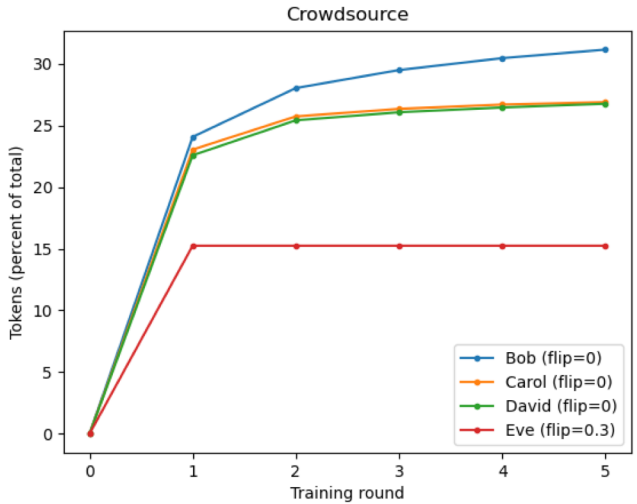
MNIST Test B (‘size’)

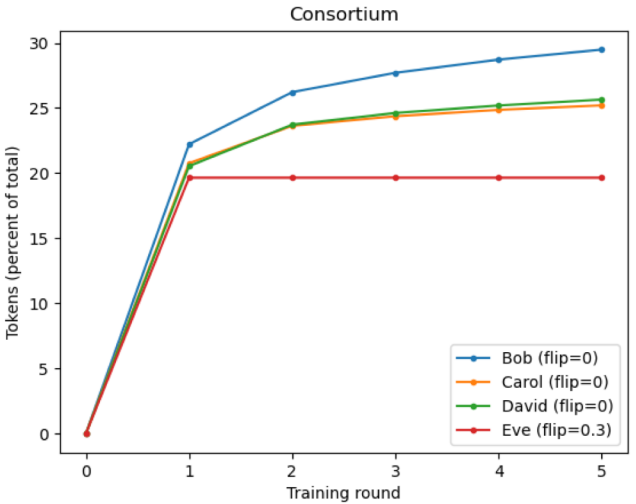




Imperial College London

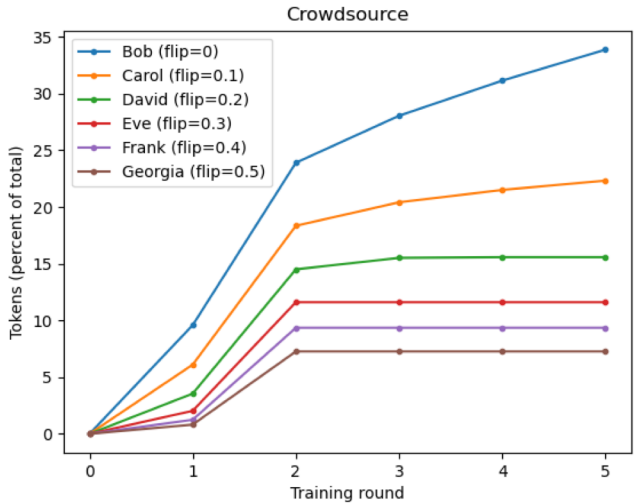
MNIST Test C (‘flip’)

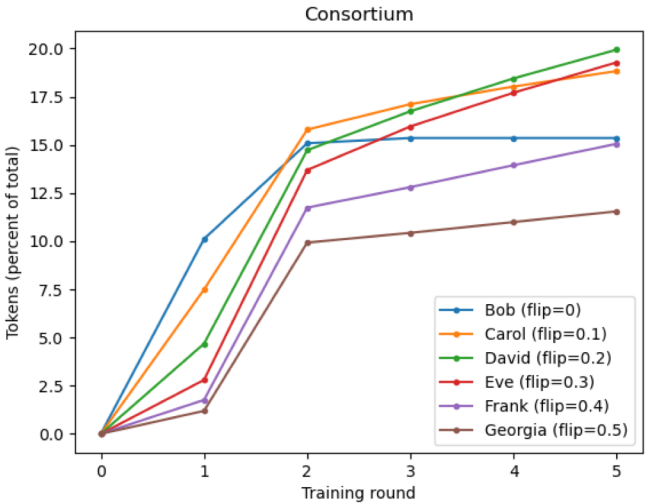




Imperial College London

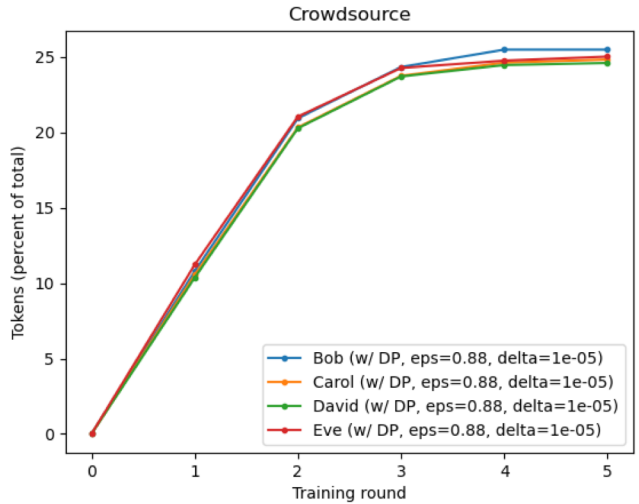
MNIST Test C (‘flip’)

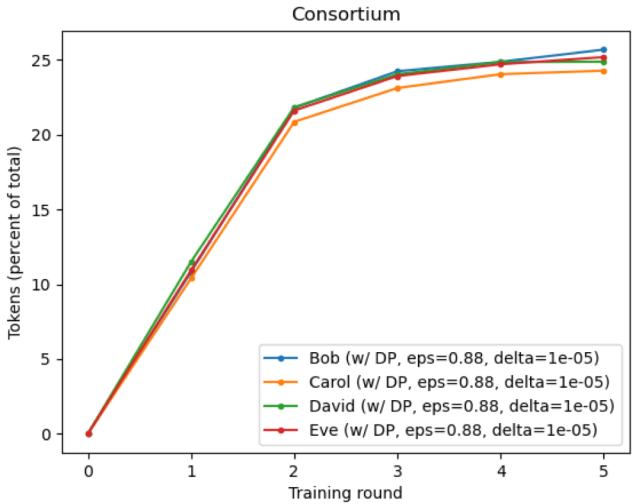




Imperial College London

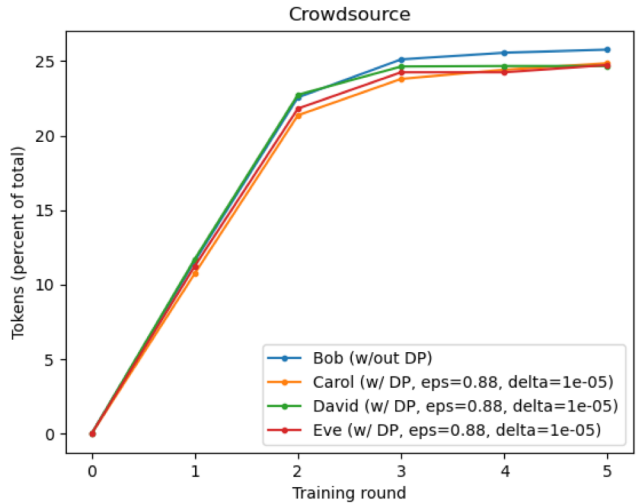
MNIST Test D (‘dp’)

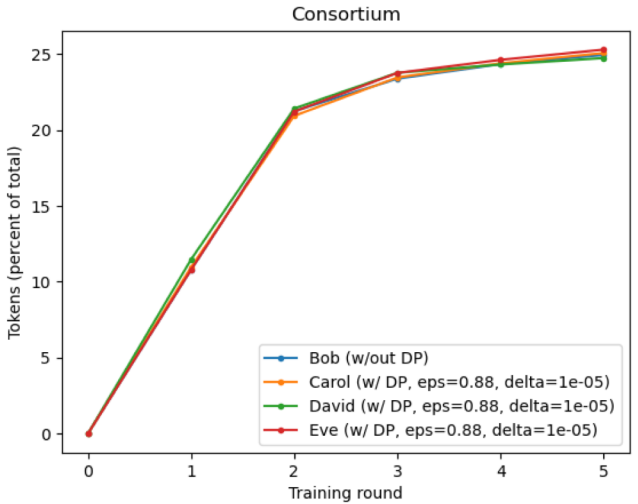




Imperial College London

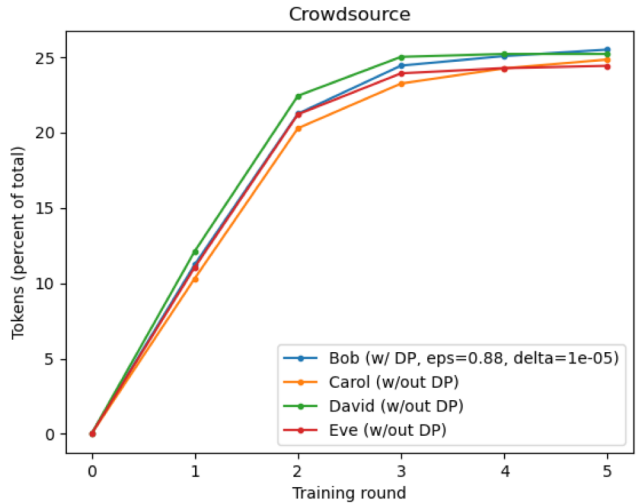
MNIST Test D (‘dp’)

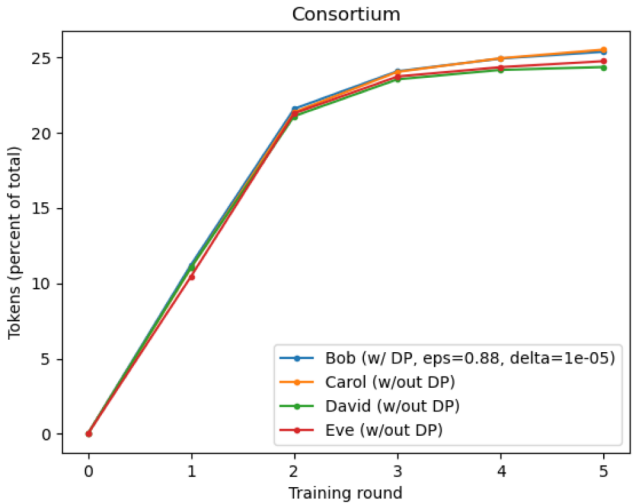




Imperial College London

MNIST Test D (‘dp’)





Imperial College London

Experiments (COVID)

	Age	SaO2	O2	Leucocytes	Lymphocytes	CRP	Creatinine	D-dimer	LDH	CK	IL-6	Thrombocytes	INR	aPTT	Severity-Score	Both-Lungs	R-Lung	R·
516	0.294610	0.461649	-0.475755	0.226097	0.446651	-0.262501	-0.041238	-1.174693	-0.322516	1.053417	-0.484239	-0.303732	0.070015	0.213049	-0.389628	0.589920	0.309100	0.240
124	-0.262037	0.457053	-0.475755	-0.800407	0.336498	-0.131505	1.551536	-0.569964	-0.769618	0.591823	0.013788	-0.618253	-0.119378	-0.594905	0.892822	-1.265645	-1.403146	-1.411
524	0.085959	0.204272	-0.475755	-0.194185	0.950674	-1.425409	-0.061001	-0.910159	-1.229510	-0.309380	-2.313850	0.912633	-0.094943	-0.698224	-1.424851	1.186554	1.309053	1.106
155	0.789180	-1.384443	1.414923	0.829902	-0.299654	1.008849	0.426462	0.172021	1.592793	1.944876	0.053151	-0.055964	0.648672	1.023122	-0.115119	-0.664704	-0.946409	-1.561
704	-0.385239	-0.471036	1.436715	0.132408	0.036286	0.734315	-0.919808	0.787109	1.170379	0.564035	1.818868	1.196934	-0.368682	0.243616	0.600053	-1.155214	-1.016000	-0.935
579	-1.238621	1.063703	-0.475755	-0.155291	0.254153	-0.390332	0.237443	-0.404238	-0.652436	0.193210	-1.168836	-0.573240	0.046485	0.714710	-0.389628	1.041906	0.873517	1.014
382	1.029456	0.757809	-0.475755	0.385242	0.729222	-0.933915	0.035156	-0.667147	-0.769008	0.646299	-1.159318	-0.357469	0.484057	0.485760	-0.389628	0.487110	0.165589	0.199
491	-0.563192	0.066535	-0.475755	0.175956	0.121934	0.431503	-0.180767	-1.454216	0.237524	1.438660	0.217076	-0.178366	-0.368682	-0.138156	-0.286884	0.604463	0.434125	0.319
220	-0.723811	-0.713397	0.703144	1.459584	-1.326525	0.410287	0.616528	-0.123736	0.136420	0.144704	0.599154	0.899923	-0.368682	-0.520259	0.058866	-0.021353	0.100153	-0.149
45	-0.372659	-0.162758	-0.475755	0.890652	-1.722782	1.210179	-0.649942	0.782687	1.045866	0.787862	0.013788	-0.950380	0.513278	0.825552	1.635109	-1.919590	-2.044440	-1.940

ression_0.0	Immunosuppression_1.0	ICU
1.0	0.0	0
1.0	0.0	1
1.0	0.0	0
1.0	0.0	1
1.0	0.0	0
1.0	0.0	0
1.0	0.0	0
1.0	0.0	0
1.0	0.0	1
1.0	0.0	1

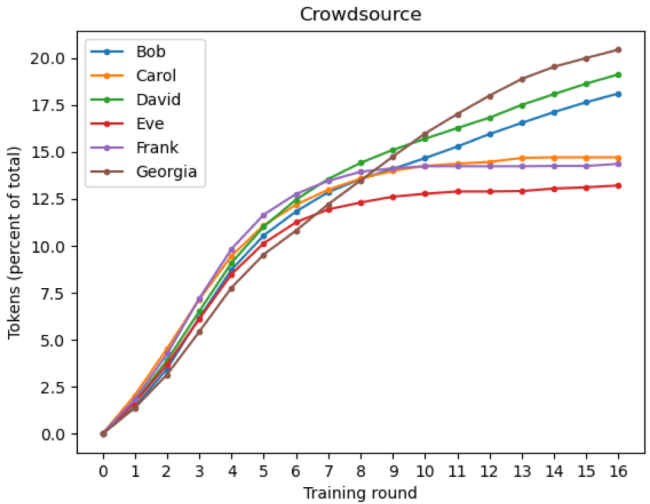
Imperial College London

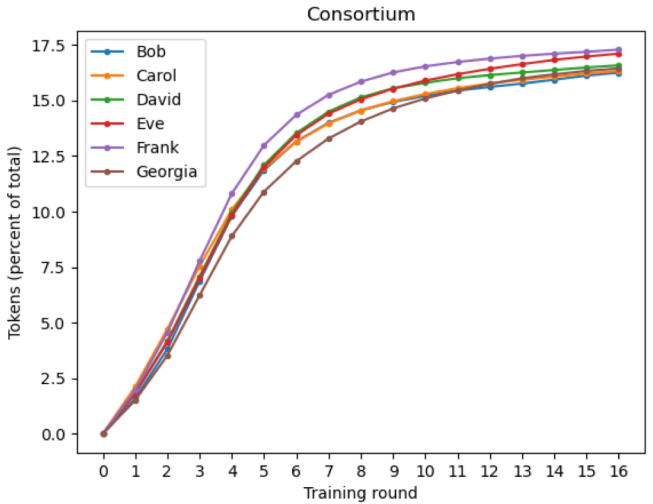
Experiments (COVID)

- Used synthetic COVID outcomes dataset, split into train set (x572) and test set (x143)
- Crowdsource Protocol:
  - The test set belongs to Alice, who is the evaluator
  - The train set is split between Bob et al, who are the trainers
- Consortium Protocol:
  - Alice and the test set are not used
  - Bob et al split the train set and form the consortium

Imperial College London

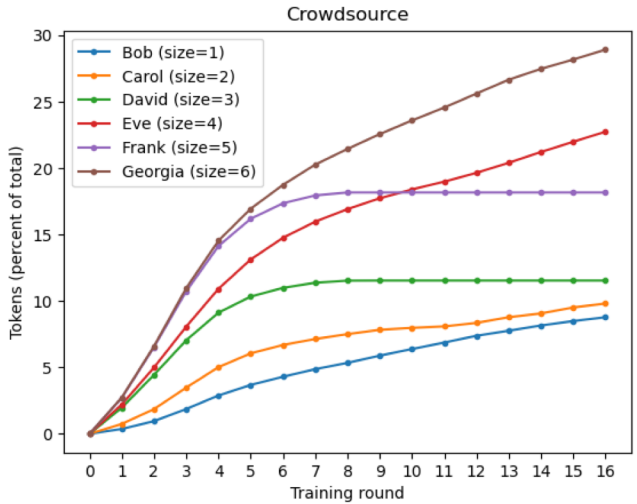
COVID Test A (‘equal’)

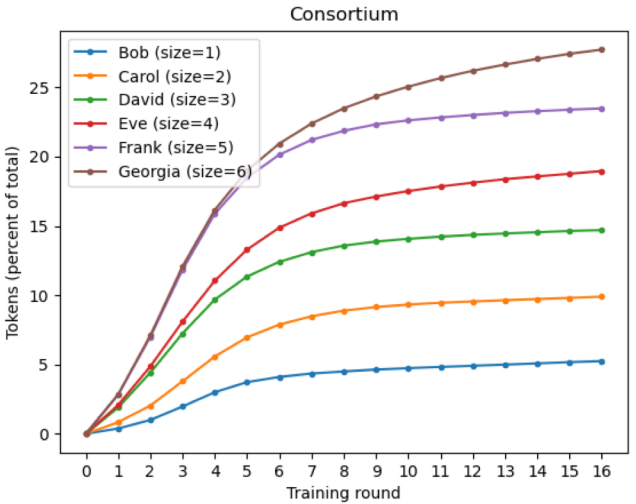




Imperial College London

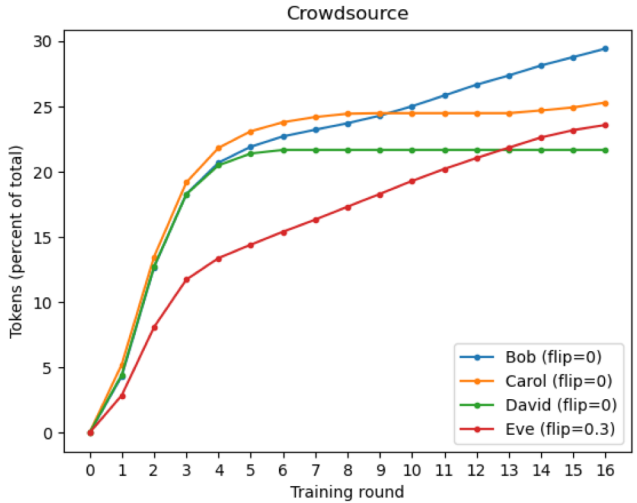
COVID Test B (‘size’)

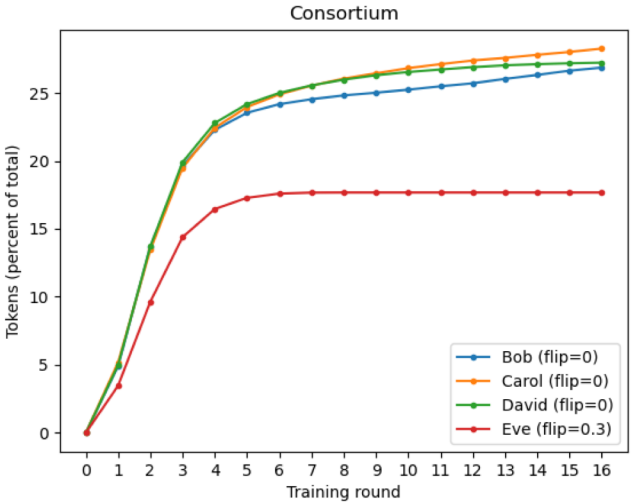




Imperial College London

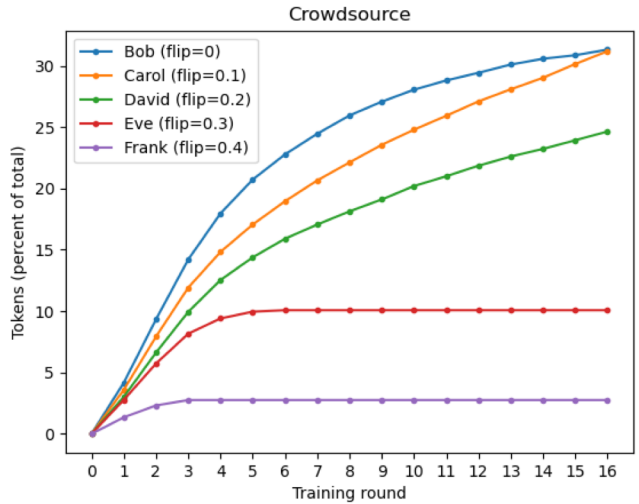
COVID Test C (‘flip’)

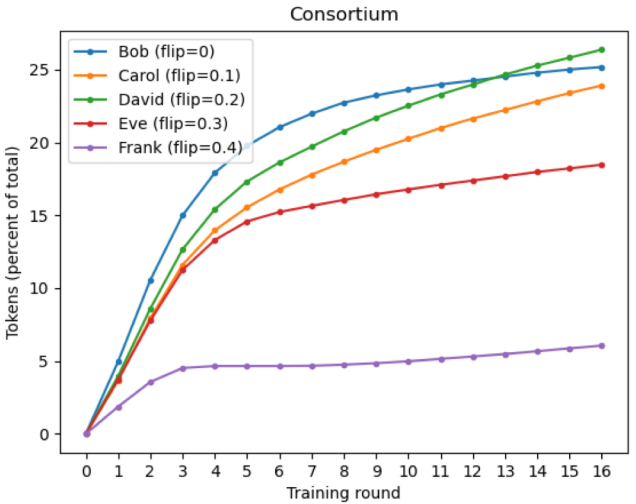




Imperial College London

COVID Test C (‘flip’)





Imperial College London

COVID Test D (‘dp’)

- Trainer datasets too small
- Cannot achieve a useful level of differential privacy

Imperial College London

Gas Costs

- Crowdsource Protocol:
  - Trainer: constant per round
  - Evaluator: rises linearly with number of trainers
- Consortium Protocol:
  - Setup: rises linearly with number of trainers
  - Trainer: rises linearly with number of trainers
  - Costs much more than Crowdsource
  - e.g. 16 rounds with 6 clients, as in COVID Test A = **390 USD**

Imperial College London

Conclusion

- *2CP = Crowdsource Protocol + Consortium Protocol*
- Both protocols are decentralised and give fair rewards to contributors
- Results from Crowdsource Protocol were sound (given a good test set)
- Results from Consortium Protocol were similar and sound, except under model poisoning attacks
  - Could be improved with a different aggregation scheme
- Consortium Protocol sometimes outperformed Crowdsource Protocol (if test set is not good)