Distributed Machine Learning and the Semblance of Trust

Dmitrii Usynin, Alexander Ziller, Daniel Rueckert, Jonathan Passerat-Palmbach, Georgios Kaissis

Imperial College London

Problem Statement

- The utilisation of large and diverse datasets for machine learning (ML) at scale is required to promote scientific insight into many meaningful problems.
- However, due to data governance regulations such as GDPR as well as ethical concerns, the aggregation of personal and sensitive data is problematic. This prompted the development of alternative strategies such as distributed ML (DML).
- Techniques such as Federated Learning (FL) allow the data owner to maintain data governance and perform model training locally without having to share their data. FL and related techniques are often described as **privacy-preserving**.
- We explain why this term is not appropriate and outline the risks associated with over-reliance on protocols that were not designed with formal definitions of privacy in mind.

The Structured Transparency framework (ST)

- The ability to train a model collaboratively without revealing the input data to other contributors (**input privacy**);
- The protection of private information which can be learned from the results (i.e. the output) of the computation (**output privacy**);
- The ability to verify the origin of the computation's result i.e. that the model update is not submitted by an unauthorised party (**input verification**);
- The capability to guarantee the correctness of the output and that the processing of inputs is honest (**output verification**);
- The ability to control data ownership and exercise governance over it (**flow governance**).

Conclusions

- We deduce that most DML protocols cannot be relied upon unless accompanied by additional mechanisms enhancing trust between participants.
- As evident from the comparison between the components of the ST framework and the current abilities of published DML systems, most frameworks can currently only satisfy a subset of ST requirements.

Term	Definition
Governance	Framework that defines the way data should be handled from the perspectives of access, ownership and auditability.
Privacy	The ability to control how much can be learned from the data about an individual.
Secrecy	Trait of the computation that implies that the sensitive data cannot be seen by anyone other than the data owner.
Security	Property of a protocol or system where the protocol or system as a whole cannot be threatened by an adversarial actor.
Accountability	The ability to track the source of the computation's results.
Verifiability	The capability to guarantee an honest processing of the input data and the integrity of the computation.

Recommendations

- Future research in the field of trustworthy AI should agree upon and adhere to terminological guidelines. For instance, neither systems without formal privacy guarantees (e.g. FL) nor systems only offering input privacy (e.g. encryption) should be haphazardly termed **privacy-preserving** (or similar).
- Only systems adhering to all aforementioned principles should be termed **trustworthy**, to avoid negative consequences associated with over-reliance on protocols that were not designed with fundamental privacy requirements in mind. Such systems should undergo external auditing in order to verify their correctness e.g. through the means of formal network certification.
- Most existing DML must be augmented with additional mechanisms in order to be able to guarantee both the privacy of the participants and the robustness of the jointly trained models. We emphasise the importance of **education** pertaining to these systems, both for experts and laypeople, in this regard.