# Offer #2024-07680

# PhD Position F/M Private and Byzantine-Robust Federated Learning

**Contract type :** Fixed-term contract

**Level of qualifications required :** Graduate degree or equivalent

**Fonction :** PhD Position

## About the research centre or Inria department

The Inria centre at Université Côte d'Azur includes 37 research teams and 8 support services. The centre's staff (about 500 people) is made up of scientists of different nationalities, engineers, technicians and administrative staff. The teams are mainly located on the university campuses of Sophia Antipolis and Nice as well as Montpellier, in close collaboration with research and higher education laboratories and establishments (Université Côte d'Azur, CNRS, INRAE, INSERM ...), but also with the regiona economic players.

With a presence in the fields of computational neuroscience and biology, data science and modeling, software engineering and certification, as well as collaborative robotics, the Inria Centre at Université Côte d'Azur is a major player in terms of scientific excellence through its results and collaborations at both European and international levels.

## Context

This PhD position is a collaboration between two Inria research teams: PreMeDICaL and Magnet. The position is funded by the IPoP project, a large interdisciplinary project on privacy. The hired PhD student will be mainly based in PreMeDICaL (Montpellier, France) but will have the opportunity to make regular visits to Magnet in Lille.

The PhD student will be jointly supervised by Aurélien Bellet, Nirupam Gupta, Batiste Le Bars and Marc Tommasi. Together, they gather a world-leading expertise in all three key aspects of the topic: federated learning, privacy and robustness.

This project will stimulate existing and emerging collaborations with other research groups on themes at the intersection between machine learning, privacy, robustness and decentralized algorithms. For instance, there will be opportunities to collaborate with other members of the IPoP project, the members of FedMalin (a large Inria project on federated learning), the members of the SSF-ML-DH project (on secure, safe and fairness machine learning for health), and the Distributed Computing Lab at EPFL led by Rachid Guerraoui.

In terms of concrete applications, both PreMeDICaL and Magnet have ongoing collaborations with hospitals and other clinical partners. These collaborations will provide opportunities to apply the approaches developed during the PhD to concrete use-cases, for instance to run multi-centric decentralized medical studies while preserving the confidentiality of the datasets held in each institution and providing robustness guarantees.

## Assignment

The increasing size of data generated by smartphones and IoT devices motivated the development of Federated Learning (FL) (Kairouz et al. 2021), a decentralized learning framework for on-device collaborative training of machine learning models. FL algorithms like FedAvg (McMahan et al. 2017) allow clients to train a common global model without sharing their personal data. FL reduces data collection costs and can help to mitigate data privacy issues, making it possible to train models on large datasets that would otherwise be inaccessible. FL is currently used by many big tech companies (e.g., Google, Apple, Facebook) for learning on their users' data, but the research community envisions also promising applications to learning across large data-silos, like hospitals that cannot share their patients' data (Rieke et al. 2020).

While they mitigate privacy concerns by not exchanging raw data, FL does not in itself offer rigorous privacy guarantees, and FL algorithms can be attacked by malicious participants. For these reasons, in recent years a large body of the literature has focused on the design of decentralized algorithms that are more privacy-preserving, using different variants of differential privacy (Noble et al. 2022; Cyffers et al. 2022). On the other hand, another line of research has focused on decentralized learning algorithms that are robust to the presence of malicious individuals in the system (Byzantine agents) (Farhadkhani et al. 2022, 2023; Guerraoui et al. 2023). Nevertheless, the design and analysis of algorithms that are both robust and privacy-preserving is far less considered and understood. Recently, it has been shown that in

the case where the server is honest-but-curious, the combination of differential privacy and robustness induces an additional error term, making them at odds with each other (Allouah et al. 2023). Specifically, we face a utility-privacy-robustness trilemma (on top of the conventional privacy-utility and robustness-utility trade-offs). Conversely, in the case of a trusted server, some studies (Hopkins et al. 2023) have shown that privacy and robustness can actually be mutually beneficial. A key question then arises: in what contexts are these two notions really good for each other?

**Research Objectives**

The main goal of this PhD is to answer the previous question on the basis of new theoretical analyses, and to design decentralized algorithms that are both robust and differentially private. Several lines of research could be investigated.

A natural direction to seek better trade-offs between privacy, robustness and utility is to relax the notions of privacy and/or robustness. One may consider the general framework of Pufferfish privacy (Kifer & Machanavajjhala, 2014; Pierquin et al., 2023), which allows to relax differential privacy by considering more specific *secrets* to protect and by constraining the prior belief that the adversary may have about the data. Similarly, while Byzantine robustness has been shown to be at odds with local differential privacy (Allouah et al. 2023(a)), it is possible to consider weaker threat models, such as the hidden state model (Ye & Shokri, 2022), the shuffle model (Cheu et al., 2019) or the network model (Cyffers et al., 2022). Regarding robustness, current approaches are designed to ensure protection against Byzantine users that can misbehave arbitrarily (Guerraoui et al., 2023). However, such robustness is too stringent and leads to conservative learning performance in practice when no user is fully adversarial. For example, in the case of medical applications it is safe to assume that all the users, usually hospitals, clinics or pharmacies, are honest by intention, but misbehavior could occur due to mistakes like *mislabelling* (Allen-Zhu et al., 2020). Refining (or designing new) Byzantine-robust schemes to weaker adversaries is crucial to fully realize the benefits of robust decentralized learning in real-world applications.

Another line of investigation is to reconsider the notion of utility. In the majority of the aformentioned work, the key quantity to control (utility) is the optimization error of the empirical risk. However, in (decentralized) machine learning one is often interested in also controling the generalization error (Bassily et al., 2020; Le Bars et al., 2024), namely the error that will be made on unobserved data points. In this case, it will be interesting to study how robustness and privacy can jointly improve algorithm stability and thus help generalization, e.g., by studying the connections between *gradient coherence* (Chatterjee, 2019) and *robust aggregation* (Yin et al., 2018; Allouah et al., 2023(b).

A last direction of research is to consider model update *compression* or *sparsification* techniques (Stich et al., 2018) that have been independently shown to help privacy (e.g, see Rui et al., 2023). Whether the benefits of these scheme hold true when aiming for robustness along with privacy remains unclear. Some technical challenges are as follows. (i) While sparsification (in decentralized learning) improves the overall privacy-utility trade-off, the same need not be true for the privacy-robustness trade-off. (ii) The compression noise can be amplified in the presence of malicious clients in the system (Rammal et al., 2024).

**References**

- Kairouz, P. et al. Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1-2), pp. 1-210, 2021.

- McMahan, B., Moore, E., Ramage, D., Hampson, S. and Aguera y Arcas, B. Communication efficient learning of deep networks from decentralized data. AISTATS 2017.

- Rieke, N., Hancox, J., Li, W. et al. The future of digital health with federated learning. npj Digit. Med. 3, 119, 2020.

- Noble, M., Bellet, A., and Dieuleveut, A. Differentially private federated learning on heterogeneous data. AISTATS 2022.

- Cyffers, Bellet, A. Privacy amplification by decentralization. AISTATS 2022

- Farhadkhani, S., Guerraoui, R., Gupta, N., Hoang, L. N., Pinot, R., & Stephan, J. Robust Collaborative Learning with Linear Gradient Overhead. ICML 2023.

- Guerraoui, R., Nirupam G., and Rafael P. Byzantine Machine Learning: A Primer. ACM Computing Surveys, 2023.

- Farhadkhani, S., Guerraoui, R., Gupta, N., Pinot, R., and Stephan, J. Byzantine machine learning made easy by resilient averaging of momentums. ICML 2022.

- Allouah, Y., Guerraoui, R., Gupta, N., Pinot, R., & Stephan, J. On the privacy-robustness-utility trilemma in distributed learning. ICML 2023(a).

- Allen-Zhu, Z., Ebrahimianghazani, F., Li, J., & Alistarh, D. Byzantine-Resilient Non-Convex Stochastic Gradient Descent. ICML 2020.

- Hopkins, S. B., Kamath, G., Majid, M., & Narayanan, S. Robustness implies privacy in statistical estimation. STOC 2023.

- Bassily, R., Feldman, V., Guzmán, C., & Talwar, K. Stability of stochastic gradient descent on nonsmooth convex losses. NeurIPS 2020.

- Le Bars, B., Bellet, A., Tommasi M., Scaman K., Neglia, G. Improved Stability and Generalization Guarantees of the Decentralized SGD Algorithm. ICML 2024.

- Yin, D., Chen, Y., Kannan, R., & Bartlett, P. Byzantine-robust distributed learning: Towards optimal statistical rates. ICML 2018

- Chatterjee, Satrajit. Coherent Gradients: An Approach to Understanding Generalization in Gradient Descent-based Optimization. ICML 2019.

- Allouah, Y., Farhadkhani, S., Guerraoui, R., Gupta, N., Pinot, R., & Stephan, J. Fixing by Mixing: A Recipe for Optimal Byzantine ML Under Heterogeneity. AISTATS 2023(b).

- Kifer & Machanavajjhala. Pufferfish: A Framework for Mathematical Privacy Definitions. ACM Transactions on Database System, 2014.

- Pierquin et al. Rényi Pufferfish Privacy: General Additive Noise Mechanisms and Privacy Amplification by Iteration. ICML 2024.

- Ye & Shokri. Differentially Private Learning Needs Hidden State (Or Much Faster Convergence). NeurIPS 2022.

- Cheu et al. Distributed Differential Privacy via Shuffling. Eurocrypt 2019.

- Stich, Sebastian U., Jean-Baptiste Cordonnier, and Martin Jaggi. Sparsified SGD with Memory. NeurIPS 2018.

- Rui, H., Yuanxiong Guo, and Yanmin Gong. Federated Learning with Sparsified Model Perturbation: Improving Accuracy Under Client-level Differential Privacy. IEEE Transactions on Mobile Computing, 2023.

- Rammal, A., Gruntkowska, K., Fedin, N., Gorbunov, E. and Richtárik, P. Communication Compression for Byzantine Robust Learning: New Efficient Algorithms and Improved Rates. AISTATS 2024

## Main activities

Research

## Skills

The applicant is expected to have studied machine learning and/or optimization, and to have good mathematical skills. Some knowledge in distributed algorithms and broad interest for the topic of trustworthy AI is a plus.

## Benefits package

- Restauration subventionnée
- Transports publics remboursés partiellement
- Congés: 7 semaines de congés annuels + 10 jours de RTT (base temps plein) + possibilité d'autorisations d'absence exceptionnelle (ex : enfants malades, déménagement)
- Possibilité de télétravail (après 6 mois d'ancienneté) et aménagement du temps de travail
- Équipements professionnels à disposition (visioconférence, prêts de matériels informatiques, etc.)
- Prestations sociales, culturelles et sportives (Association de gestion des œuvres sociales d'Inria)
- Accès à la formation professionnelle
- Sécurité sociale

## Remuneration

Gross Salary per month: 2100€ brut per month (year 1 & 2) and 2190€ brut per month (year 3)

## General Information

- **Theme/Domain :** Optimization, machine learning and statistical methods
  Statistics (Big data) (BAP E)
- **Town/city :** Montpellier
- **Inria Center :** Centre Inria d'Université Côte d'Azur
- **Starting date :** 2024-10-01
- **Duration of contract :** 3 years
- **Deadline to apply :** 2024-07-09

## Contacts

- **Inria Team :** PREMEDICAL
- **PhD Supervisor :**
  Bellet Aurelien / aurelien.bellet@inria.fr

## About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

**Warning** : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

## Instruction to apply

**Defence Security :**
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST).Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

**Recruitment Policy :**
As part of its diversity policy, all Inria positions are accessible to people with disabilities.