



Offre n°2025-08971

PhD Position F/M Private and Fair Machine Learning

Type de contrat : Fixed-term contract

Contrat renouvelable : Oui

Niveau de diplôme exigé : Graduate degree or equivalent

Fonction : PhD Position

A propos du centre ou de la direction fonctionnelle

The Centre Inria de l'Université de Grenoble groups together almost 600 people in 24 research teams and 9 research support departments.

Staff is present on three campuses in Grenoble, in close collaboration with other research and higher education institutions (Université Grenoble Alpes, CNRS, CEA, INRAE, ...), but also with key economic players in the area.

The Centre Inria de l'Université Grenoble Alpes is active in the fields of high-performance computing, verification and embedded systems, modeling of the environment at multiple levels, and data science and artificial intelligence. The center is a top-level scientific institute with an extensive network of international collaborations in Europe and the rest of the world.

Contexte et atouts du poste

Context. This PhD thesis is part of the ANR JCJC project [AI-PULSE](#) (Aligning Privacy, Utility, and Fairness for Responsible AI), coordinated by Héber H. Arcolezi. AI-PULSE, which started in March 2025, aims to design machine learning models that are both differentially private and fairness-aware.

The thesis will be conducted under a co-tutelle agreement between Inria Grenoble (France) and ÉTS Montreal (Canada), leveraging the complementary strengths of both institutions.

The envisioned plan is for the recruited PhD student to spend approximately two years at Inria Grenoble (Privatics team), followed by two years at ÉTS Montreal. This structure will provide a rich and balanced international training environment, enabling the student to benefit from diverse expertise, ecosystems, and research cultures.

The thesis will be co-supervised by:

- [Héber H. Arcolezi](#), Researcher at Inria and in-coming Assistant Professor at ÉTS Montreal (February 2026).
- [Claude Castelluccia](#), Researcher at Inria Grenoble and CNIL Commissioner.
- [Ulrich Aïvodji](#), Assistant Professor at ÉTS Montreal.

The PhD project will contribute to the core objectives of AI-PULSE, with a particular focus on advancing methodologies that jointly address privacy and fairness in machine learning under local differential privacy. The student will also benefit from the broader international collaborations and mobility opportunities enabled by the co-tutelle agreement, further strengthening the international dimension of their training and research.

Mission confiée

Assignment. Modern *Machine Learning* (ML) systems increasingly drive automated decision-making across sectors like healthcare, finance, and public policy. While these systems can offer remarkable benefits, they also raise critical concerns regarding individual privacy and algorithmic fairness.

Differential Privacy (DP) [1] has emerged as a gold-standard privacy notion for balancing the privacy-utility trade-off in data analytics. However, the central (server-side) approach to DP requires trust in a third party holding the raw data. Hence, there is growing interest in *Local Differential Privacy* (LDP) [2], which performs data obfuscation directly at the user's side, removing the need to trust a central server with unprotected personal information.

Meanwhile, *fairness in ML* [3] typically involves mitigating disparate impact among subgroups defined by sensitive attributes (e.g., race, gender). Yet, fairness interventions often require exactly the sort of sensitive information that privacy mechanisms try to hide. In addition, privacy and fairness can inadvertently impact each other, bringing new privacy-fairness-utility trade-offs.

Specifically, the interplay between privacy and fairness is both crucial and nuanced:

- Satisfying DP can unintentionally worsen fairness if certain subpopulations are more sensitive to noise [4].
- Enforcing fairness can unintentionally worsen privacy [5].

Therefore, the aim of this PhD thesis is to design, analyze, and implement differential privacy mechanisms that foster equitable ML outcomes while preserving model performance (utility). A key objective will then be to design a comprehensive open-source Python framework, offering ready-to-use building blocks for private and fair machine learning.

Research Objective. The goal of this PhD thesis is to advance the state of the art at the intersection of privacy and fairness by designing new representation learning techniques under LDP that enable fair and effective ML models. In particular, we aim to develop new locally private representations that preserve the utility of the

data for downstream tasks while facilitating fair learning, even when sensitive attributes are obfuscated or partially unavailable.

The thesis will focus on:

- Designing new LDP-based data representations or embeddings optimized for fairness-aware ML.
- Theoretically analyzing the privacy-fairness-utility trade-offs of these representations.
- Experimentally validating the proposed methods on benchmark datasets.

The expected outcomes of this thesis will contribute both to the core objectives of AI-PULSE and to the broader responsible AI community by offering practical tools and theoretical insights for privacy-preserving and fair ML in decentralized settings.

Selected references:

- [1] Dwork, Cynthia, and Aaron Roth. "The algorithmic foundations of differential privacy." *Foundations and Trends® in Theoretical Computer Science* 9.3–4 (2014): 211-407.
- [2] Yang, Mengmeng, et al. "Local differential privacy and its applications: A comprehensive survey." *Computer Standards & Interfaces* 89 (2024): 103827.
- [3] Barocas, Solon, Moritz Hardt, and Arvind Narayanan. *Fairness and machine learning: Limitations and opportunities*. MIT press, 2023.
- [4] Bagdasaryan, Eugene, Omid Poursaeed, and Vitaly Shmatikov. "Differential privacy has disparate impact on model accuracy." *Advances in neural information processing systems* 32 (2019).
- [5] Chang, Hongyan, and Reza Shokri. "On the privacy risks of algorithmic fairness." *2021 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 2021.

Principales activités

Main activities:

- Carry out the PhD research project on Differentially Private and Fairness-Aware Machine Learning.
- Collaborate with other team members and with project partners (e.g., UQAM, Federal University of Ceará, Inria, ÉTS Montréal).
- Disseminate research results through publications and presentations at international conferences.

Compétences

We are looking for a candidate with:

- Good programming skills in Python and good analytical skills.
- A good background in probability/statistics and deep learning is expected.

- Knowledge of differential privacy and/or fairness is a plus, but not necessary.
- The candidate should be fluent in English.

Avantages

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Rémunération

2200 euros gross salary /month

Informations générales

- **Thème/Domaine** : Security and Confidentiality
Statistics (Big data) (BAP E)
- **Ville** : Montbonnot
- **Centre Inria** : [Centre Inria de l'Université Grenoble Alpes](#)
- **Date de prise de fonction souhaitée** : 2025-10-01
- **Durée de contrat** : 2 years
- **Date limite pour postuler** : 2025-07-26

Contacts

- **Équipe Inria** : [PRIVATICS](#)
- **Directeur de thèse** :
Hwang Arcolezi Heber / heber.hwang-arcolezi@inria.fr

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation

contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 200 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

Attention: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

Consignes pour postuler

Applications must be submitted online via the Inria website. Processing of applications submitted via other channels is not guaranteed.

Sécurité défense :

Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.