## Titre

Violation de confidentialité dans le processus d'apprentissage fédéré: quelles sont les menaces, attaques et défenses en présence d'acteurs malveillants ?

## Présentation de l'unité d'accueil : CEA LIST / SID

Situé à Saclay, en Ile-de-France sud, le CEA LIST (http://www-list.cea.fr/) est un centre de recherche scientifique et technologique dédié au développement de logiciels, de systèmes embarqués et de capteurs pour des applications destinées à la défense, la sécurité, l'énergie, le nucléaire, l'environnement et la santé. Le CEA LIST fait partie de l'écosystème dynamique et stimulant de l'Université Paris Saclay - le plus grand pôle scientifique français comptant 60 000 étudiants. Il compte plus de 700 chercheurs se focalisant sur les systèmes numériques intelligents, centrés autour de l'intelligence artificielle, l'usine du futur, l'instrumentation innovante, les systèmes cyberphysiques et la santé numérique. Au sein de cet institut, le SID (Service d'Intelligence des Données), travaille sur les algorithmes et méthodologies avancées de l'intelligence artificielle et plus particulièrement sur les thématiques de l'IA explicative, l'IA de confiance, l'IA frugale et l'IA distribuée.

## Description du sujet de thèse

En 2016, Google publie les principes fondateurs de l'apprentissage fédéré [1] avec la promesse de créer des IA sans compromettre les données des utilisateurs. Cette méthode est en train de changer le paradigme actuel de l'IA centralisée, où construire de meilleurs modèles se résume souvent à collecter toujours plus de données personnelles et les centraliser sur un serveur. L'apprentissage fédéré est une approche collaborative où tous les utilisateurs d'un service participent à l'apprentissage du modèle sans transmettre leurs données personnelles mais uniquement les paramètres du modèle mis à jour localement. Au lieu de centraliser les données, seuls les paramètres du modèle sont agrégés sur le serveur central ce qui permet d'améliorer la confidentialité des données et de limiter les coûts de communication.

Notre sujet de thèse aborde un verrou fondamental de l'apprentissage fédéré qui est celui de la confidentialité. Par construction, l'apprentissage fédéré apparait comme une solution pour la confidentialité des données mais pas pour la confidentialité du modèle. Et même si les paramètres du modèle contiennent beaucoup moins d'informations à propos des clients que les données brutes, il est tout à fait envisageable d'inférer des informations clients à partir d'un modèle statistique. Cette menace est particulièrement présente dans le cadre fédéré où des acteurs malveillants (clients ou serveur) peuvent exploiter les paramètres transmis à chaque tour pour reconstruire de l'information à propos des individus. L'objectif de la thèse est de faire un tour d'horizon des attaques déployées dans le cadre fédéré pour ensuite proposer des solutions innovantes pour garantir la confidentialité des données face à un serveur ou des clients malveillants.

[1] Google AI blog: https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

## Profil recherché

La thèse s'adresse à un(e) étudiant(e) du cycle ingénieur/universitaire disposant d'un Master 2 dans l'un des domaines suivants : IA, Machine Learning, statistiques. Des notions de cryptographie seraient un plus.

Lieux du poste : Saclay (Ile-de-France)

Date de début envisagée : 11/2022

## Contact

Aurélien MAYOUE (aurelien.mayoue@cea.fr)