

## Titre

Stage M2 sur l'apprentissage fédéré avec des données non-IID

**list**  
cea tech

## Présentation du laboratoire : CEA LIST / LS2D

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 LS2D (Laboratoire Sciences des Données et de la Décision) travaille sur les algorithmes et méthodologies de l'intelligence artificielle et du traitement du signal. Les recherches et avancées technologiques du laboratoire sont guidées par des applications variées, pour lesquelles les spécificités et contraintes sur les données ou l'environnement d'exécution nécessitent une conception fine des IA et de leur intégration comme briques unitaires de systèmes complexes.

## Description du projet

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 (voir Fig.1). Au lieu de centraliser les données, seuls les paramètres du modèle sont agrégés sur le serveur central ce qui permet de préserver la confidentialité des données et de limiter les coûts de communication.

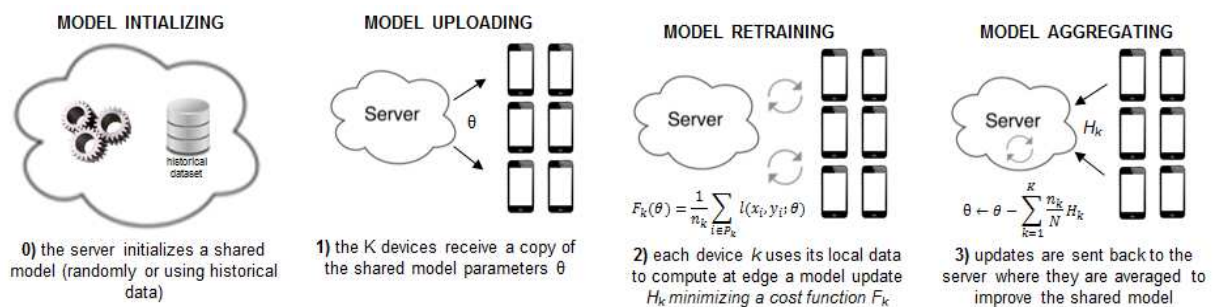


Fig-1 : illustration de l'apprentissage fédéré

L'apprentissage d'un réseau de neurones nécessite un ensemble de données iid (indépendantes et identiquement distribuées) afin de garantir que le gradient stochastique est une estimation non biaisée du gradient entier. Contrairement à un apprentissage centralisé, il est impossible dans le cadre de l'apprentissage fédéré d'assurer que les données locales de chaque utilisateur soient toujours iid. L'objectif du stage consiste à explorer des pistes de recherche afin d'adapter l'apprentissage des modèles locaux en fonction de la distance entre la distribution des données locales et celles de l'ensemble de la population. Cette personnalisation des modèles devrait homogénéiser les performances at edge indépendamment de la distribution des données locales.

Les cas d'applications envisagées concernent des données de consommations électriques, des données de véhicules autonomes ou des données d'assistants vocaux.

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

## Profil recherché

---

Le stage s'adresse à un(e) étudiant(e) du cycle ingénieur/universitaire cherchant un stage M2 et manifestant l'envie de travailler dans le milieu de la recherche.

Idéalement, le candidat suit actuellement une formation en lien avec le domaine de l'Intelligence Artificielle. Des connaissances en Machine Learning et statistiques sont nécessaires. La maîtrise de Python est fortement recommandée.

Le/la candidat(e) devra être capable d'apporter ses idées novatrices, son enthousiasme, sa rigueur et devra faire preuve d'un esprit d'équipe prononcé.

La durée du stage est de 6 mois et pourra se poursuivre par une thèse de doctorat. Le stage est rémunéré.

## Informations administratives et contacts

---

Nature du contrat de travail : stage

Type du contrat de travail : Droit privé

Délai administratif pour début de contrat : environ 5 mois

Envoyer CV et lettre de motivation à :

Aurélien Mayoue ([aurelien.mayoue@cea.fr](mailto:aurelien.mayoue@cea.fr))

Cédric Gouy-Pailler ([cedric.gouy-pailler@cea.fr](mailto:cedric.gouy-pailler@cea.fr))