

**Offre de stage 2024/2025**

<b>Titre</b>	<b>Comment faire de l'apprentissage distribué sur des flux de données ?</b>
<b>Niveau</b>	Master/Ingénieur
<b>Date de début/ fin</b>	Février-Juillet 2025
<b>Ville, Pays</b>	Annecy, France
<b>Laboratoire</b>	<a href="#">Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance</a> - LISTIC
<b>Description du sujet</b>	<p>Ce stage prend racine dans le cadre d'un projet exploratoire financé par l'université Savoie Mont Blanc. L'objectif de ce projet est, entre autres, de développer des algorithmes d'intelligence artificielle capables d'apprendre sur des flux de données. Durant ce stage, nous nous concentrons plus spécifiquement sur l'apprentissage fédéré décentralisé, une méthode d'apprentissage distribuée innovante.</p> <p>L'apprentissage fédéré est une méthode d'apprentissage récente développée par Google en 2016. Elle consiste en l'optimisation de modèles (souvent de deep learning) placés chez une multitude de clients et agrégés régulièrement sur un serveur pour gagner en capacité de généralisation. Cette méthode d'apprentissage est très utilisée dans l'embarqué ou dans des contextes spécifiques comme le médical pour optimiser des modèles généraux en utilisant les données produites localement et sans nécessiter leur déplacement vers un serveur central. Elle permet par ce biais de maintenir leur confidentialité. Ainsi, chaque client possède un jeu de données local qui lui permet de s'entraîner et partager ensuite sa connaissance. Ce partage permet l'agrégation des modèles, une étape critique qui consiste en la fusion des paramètres obtenus par chaque client pour créer un modèle global plus général et redistribué à tous les participants. Néanmoins, les méthodes d'agrégation issues de l'état de l'art prennent peu en compte l'aspect décentralisé des clients et la dynamique de leurs données. Ces aléas illustrent pourtant les cas d'utilisation réels et augmentent le risque de pénaliser les performances du modèle général.</p> <p>Nous nous concentrerons notamment sur deux aspects : (1) comprendre les effets de la dynamique des données sur l'apprentissage des modèles locaux et du modèle global pour de l'apprentissage fédéré centralisé (2) la décentralisation de l'apprentissage fédéré tout en maintenant les capacités d'apprentissage sur des flux de données.</p> <p>Le travail sera réalisé à l'aide d'un framework développé au laboratoire. Les tâches seront les suivantes :</p> <ul style="list-style-type: none"><li>- Prise en main et enrichissement du framework et des jeux de données.</li><li>- Etat de l'art sur l'apprentissage fédéré, les fonctions d'agrégation, l'apprentissage fédéré décentralisé et l'apprentissage continu.</li><li>- Proposition d'une solution pour améliorer l'apprentissage des modèles locaux et globaux dans l'apprentissage fédéré et l'apprentissage fédéré décentralisé avec un flux de données.</li><li>- Rédaction d'un article de recherche</li></ul>
<b>Compétences requis</b>	<ul style="list-style-type: none"><li>- Deep learning</li><li>- Apprentissage par renforcement et/ou apprentissage fédéré</li><li>- Python</li></ul>
<b>Gratification</b>	Selon législation en vigueur (~600€/mois)
<b>Tuteurs / Contacts</b>	Alexandre Benoit – <a href="mailto:alexandre.benoit@univ-smb.fr">alexandre.benoit@univ-smb.fr</a> Mickaël Bettinelli – <a href="mailto:mickael.bettinelli@univ-smb.fr">mickael.bettinelli@univ-smb.fr</a> Vincent Couturier – <a href="mailto:vincent.couturier@univ-smb.fr">vincent.couturier@univ-smb.fr</a>