Ph.D. research topic

- Title of the proposed topic: Incentives for Federated Learning
- Research axis of the 3IA: Core elements of AI, AI for smart and secure territories
- Potential co-supervisor (name, affiliation):
- The laboratory and/or research group: Inria, NEO team, https://team.inria.fr/neo/

Apply by sending an email directly to the supervisor.
The application will include:
- Letter of recommendation of the supervisor indicated above
- Curriculum vitae.
- Motivation Letter.
- Academic transcripts of a master’s degree(s) or equivalent.
- At least, one letter of recommendation.
- Internship report, if possible.

⇒ All the requested documents must be gathered and concatenated in a single PDF file named in the following format: LAST NAME of the candidate_Last Name of the supervisor_2023.pdf

- Description of the topic:

Context:

The increasing size of data generated by smartphones and IoT devices motivated the development of Federated Learning (FL) [1,2], a framework for on-device collaborative training of machine learning models. FL algorithms like FedAvg [3] allow clients to train a common global model without sharing their personal data; FL reduces data collection costs and protects clients' data privacy. In doing so it makes possible to train models on large datasets that would otherwise have been inaccessible.

FL is currently used by many big data companies (e.g., Google, Apple, Facebook) for learning on their users' data, but we envision also promising applications to learning across large data-silos, like hospitals that cannot share their patients' data [4].
Research goal:

One of the main scientific challenges of FL, in comparison to other forms of distributed learning, is statistical heterogeneity, i.e., the fact that clients' local datasets are in general drawn from different distributions. Statistical heterogeneity for example slows down the convergence of FL algorithms [5].

In this thesis we are interested on how statistical heterogeneity may affect users' choices. In particular, if a client's dataset distribution is quite different from the other distributions, the client may prefer to train a local model autonomously. The dissatisfied client may then abandon the training procedure (or refuse to join it in the future), impoverishing the aggregate pool of data and then the quality of the final model. Defections of some clients can then potentially trigger a cascade of defections as clients are less and less satisfied with the model learned by FL algorithms.

The candidate will investigate FL adoption in a setting where users can decide to opt out from the federation. Insights can come from a game-theoretic study of the stability of the federation [6-9], leading to the design of economic incentives for users [10-12]. New FL algorithms can also promote clients' participation by directly maximizing the fraction of clients who benefit from using the global model [13], or by allowing the client to learn a personalized model adapted to its local distribution [14-17].

This study will need to address open issues related to quantifying statistical heterogeneity across clients [18, 19] and quantifying the value of each client's dataset [20-22]. These issues are highly relevant in the developing data economy where multiple online data exchange platforms, such as AWS data exchange [23] and Dawex [24]. The complexity of these issues is amplified in the FL setting, where each participant has only access to its own data.

Candidate profile:

The candidate should have a solid mathematical background (in particular on optimization) and in general be keen on using mathematics to model real problems and get insights. He should also be knowledgeable on machine learning and have good programming skills. Previous experiences with PyTorch or TensorFlow is a plus.

Useful Information/Bibliography:

VTS Asia Pacific Wireless Communications Symposium (APWCS), 2019.