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
- Supervisor (name, affiliation, email): Giovanni Neglia, Inria, giovanni.neglia@inria.fr, http://www-sop.inria.fr/members/Giovanni.Neglia/
- 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 vitæ.
- 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:

[1] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 37 (3):50-60, 2020.
[2] Peter Carouse, H Brendan McMahan, Brendan Avent, Aurelien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Keith Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. arXiv preprint arXiv:1912.04977, 2019.

[3] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication efficient learning of deep networks from decentralized data. In Artificial Intelligence and Statistics, PMLR, 2017.
[4] Rieke, N., Hancox, J., Li, W. et al. The future of digital health with federated learning. npj Digit. Med. 3, 119, 2020.
[5] Xiang Li, Kaixuan Huang, Wenhao Yang, Shusen Wang, and Zhihua Zhang, On the Convergence of FedAvg on Non-IID Data, in International Conference on Learning Representations, 2019.
[6] Xuezhen Tu, Kun Zhu, Nguyen Cong Luong, Dusit Niyato, Yang Zhang, and Juan Li. Incentive mechanisms for federated learning: From economic and game theoretic perspective. arXiv preprint arXiv:2111.11850, 2021.
[7] Kate Donahue and Jon Kleinberg. Model-sharing games: Analyzing federated learning under voluntary participation. In The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21), 2021.
[8] Kate Donahue and Jon Kleinberg. Optimality and stability in federated learning: A game-theoretic approach. In Advances in Neural Information Processing Systems, 2021.
[9] Avrim Blum, Nika Haghtalab, Richard Lanas Phillips, and Han Shao. One for one, or all for all: Equilibria and optimality of collaboration in federated learning. In International Conference on Machine Learning, 2021.
[10] Jingoo Han, Ahmad Faraz Khan, Syed Zawad, Ali Anwar, Nathalie Baracaldo Angel, Yi Zhou, Feng Yan, and Ali R. Butt. Tokenized incentive for federated learning. In Proceedings of the Federated Learning Workshop at the Association for the Advancement of Artificial Intelligence (AAAI) Conference, 2022.
[11] Jiawen Kang et al. "Incentive Mechanism for Reliable Federated Learning: A Joint Optimization Approach to Combining Reputation and Contract Theory". In: IEEE Internet of Things Journal 6.6 (2019), pp. 10700– 10714.
[12] Jiawen Kang, Zehui Xiong, Dusit Niyato, Han Yu, Ying-Chang Liang, and Dong In Kim. Incentive design for efficient federated learning in mobile networks: A contract theory approach. In 2019 IEEE
[13] Yae Jee Cho, Divyansh Jhunjhunwala, Tian Li, Virginia Smith, Gauri Joshi, To Federate or Not To Federate: Incentivizing Client Participation in Federated Learning, arXiv:2205.14840
[14] Othmane Marfoq, Giovanni Neglia, Aurelien Bellet, Laetitia Kameni, and Richard Vidal. Federated multi-task learning under a mixture of distributions, NeurIPS 2021
[15] Othmane Marfoq, Giovanni Neglia, Laetitia Kameni, Richard Vidal, Personalized Federated Learning through Local Memorization, ICML 2022.
VTS Asia Pacific Wireless Communications Symposium (APWCS), 2019.
[16] Valentina Zantedeschi, Aurelien Bellet, and Marc Tommasi. Fully decentralized joint learning of personalized models and collaboration graphs. volume 108 of Procexedings of Machine Learning Research, pages 864-874, 2020. PMLR.
[17] Tian Li, Shengyuan Hu, Ahmad Beirami, and Virginia Smith. Ditto: Fair and robust federated learning through personalization. In International Conference on Machine Learning, pages 6357-6368. PMLR, 2021.
[18] Yishai Mansour, Mehryar Mohri, Jae Ro, and Ananda Theertha Suresh, 2020. Three approaches for personalization with applications to federated learning. arXiv preprint arXiv:2002.10619.

[19] Matthieu Even, Laurent Massoulié, Kevin Scaman, On sample optimality in personalized collaborative and federated learning, NeurIPS 2022-36th Conference on Neural Information Processing System
[20] Anish Agarwal, Munther Dahleh, and Tuhin Sarkar, A marketplace for data: An algorithmic solution, in Proceedings of the 2019 ACM Conference on Economics and Computation, pp. 701–726, 2019.
[21] Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, Asu Ozdaglar, Too much data: Prices and inefficiencies in data markets, tech. rep., National Bureau of Economic Research, 2019.
[22] Shengzhong Liu, Zhenzhe Zheng, Fan Wu, Shaojie Tang, and Guihai Chen, Context-aware data quality estimation in mobile crowdsensing, in IEEE INFOCOM 2017- IEEE Conference on Computer Communications, 2017.
[23] Aws. https://aws.amazon.com/cn/financial-services/partner-solutions/xignite-market-data-cloud-platform/.
[24] Dawex data exchange. https://www.dawex.com/en/.