# 2022-05132 - Post-Doctoral Research Visit F/M Energy-Aware Federated Learning

**Type de contrat :** CDD
**Niveau de diplôme exigé :** Thèse ou équivalent
**Fonction :** Post-Doctorant

## A propos du centre ou de la direction fonctionnelle

The Inria Université Côte d'Azur center counts 36 research teams as well as 7 support departments. The center's staff (about 500 people including 320 Inria employees) is made up of scientists of different nationalities (250 foreigners of 50 nationalities), engineers, technicians and administrative staff. 1/3 of the staff are civil servants, the others are contractual agents. The majority of the center's research teams are located in Sophia Antipolis and Nice in the Alpes-Maritimes. Four teams are based in Montpellier and two teams are hosted in Bologna in Italy and Athens. The Center is a founding member of Université Côte d'Azur and partner of the I-site MUSE supported by the University of Montpellier.

## Contexte et atouts du poste

Deep neural networks have enabled impressive accuracy improvements across many machine learning tasks. Often the highest scores are obtained by the most computationally-hungry models [1]. As a result, training a state-of-the-art model now requires substantial computational resources which demand considerable energy, along with the associated economic and environmental costs. Research and development of new models multiply these costs by thousands of times due to the need to try different model architectures and different hyper-parameters.

A recent paper [2] has estimated the amount of energy and the corresponding CO2 emissions required to train different models.

For example, the full neural architecture search described in [1] to train a big transformer model for machine translation is estimated to have consumed 650 kWh and generated the equivalent of 284 tons of CO2.

As a comparison, the average American citizen produces 16 tons of CO2 per year and a New York City-San Francisco round-trip flight of a Boeing 777 with 300 passengers produces 260 tons. As the role of AI becomes more pervasive in our society, its sustainability needs to be addressed.

The development of new low-energy hardware accelerators is an important direction to explore, and neuromorphic hardware for spiking neural networks [3] or new light-based hardware [4] are definitely interesting solutions. But in this project, we investigate a more algorithmic/system-level approach to reduce energy consumption for distributed ML training over the Internet.

- References:


[1] D. R. So, C. Liang, and Q. V. Le, The evolved transformer, 36th Intl. Conference on Machine Learning (ICML), 2019.

[2] E. Strubell, A. Ganesh, and A. McCallum, Energy and Policy Considerations for Deep Learning in NLP, Annual Meeting of the Association for Computational Linguistics (ACL), 2019.

[3] M. Davies et al., Loihi: A Neuromorphic Manycore Processor with On-Chip Learning, in IEEE Micro, vol. 38, no. 1, pp. 82-99, January/February 2018.

[4] LightOn, AI for all, everywhere. NLP Extreme scale AI usable through the Muse API (lighton.ai), see their list of publications.

[5] Chuan Xu, Giovanni Neglia, and Nicola Sebastianelli. Dynamic Backup Workers for Parallel Machine Learning, under submission, available at Dynamic Backup Workers for Parallel Machine Learning - Archive ouverte HAL (archives-ouvertes.fr)

[6] A. Nedic and A. Ozdaglar, Distributed Subgradient Methods for Multi-Agent Optimization, in IEEE Trans. on Automatic Control, vol. 54, no. 1, pp. 48-61, Jan. 2009.

[7] M. El Chamie, G. Neglia, and K. Avrachenkov, Reducing Communication Overhead for Average Consensus, IFIP 12th Intl. Conference on Networking (IFIP Networking), 2013.

[8] G. Neglia, G. Calbi, D. Towsley, and G. Vardoyan, The Role of Network Topology for Distributed Machine Learning, IEEE Conference on Computer Communications (INFOCOM), 2019.

[9] G. Neglia, C. Xu, D. Towsley, and G. Calbi, Decentralized gradient

methods: does topology matter?, Intl. Conference on Artificial Intelligence and Statistics (AISTATS), 2020.

[10] G. Neglia, M. Sereno, and G. Bianchi, Geographical Load Balancing across Green Datacenters: a Mean Field Analysis, Greenmetrics workshop, in conjunction with ACM Sigmetrics/IFIP Performance, 2016.

[11] I. Dimitriou, S. Alouf, and A. Jean-Marie, A Markovian Queueing System for Modeling a Smart Green Base Station, 12th European Performance Evaluation Workshop (EPW), 2015.

[12] F. Prieur, A. Jean-Marie, and M. Tidball, Growth and Irreversible

Pollution: Are Emission Permits a Means of Avoiding Environmental and Poverty Traps?, Macroeconomic Dynamics, Volume 17, Numéro 2 , pp. 261-293, mars 2013.

[13] A. Jean-Marie, M. Tidball, and M. Moreaux, Optimal carbon sequestration when reservoirs are leaky, 12th Viennese Workshop on Optimal Control, Dynamic Games and Nonlinear Dynamics, 2012.

[14] G. Di Bella, L. Giarré, M. Ippolito, A. Jean-Marie, G. Neglia, and I. Tinnirello, Modeling Energy Demand Aggregators for Residential Consumers, IEEE 52nd Annual Conference on Decision and Control (CDC), 2013.


## Informations générales

- **Thème/Domaine :** Optimisation, apprentissage et méthodes statistiques Système & réseaux (BAP E)
- **Ville :** Sophia Antipolis
- **Centre Inria :** CRI Sophia Antipolis - Méditerranée
- **Date de prise de fonction souhaitée :** 2022-10-01
- **Durée de contrat :** 2 ans
- **Date limite pour postuler :** 2022-08-31

## Contacts

- **Equipe Inria :** NEO
- **Recruteur :**
  Neglia Giovanni / Giovanni.Neglia@inria.fr

## A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3500 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 180 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

## L'essentiel pour réussir

We are looking for a candidate with a strong background on optimization or energy-aware computing platforms.

## Consignes pour postuler

**Sécurité défense :**
Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

**Politique de recrutement :**
Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.

**Attention**: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

[15] I. Tinnirello, G. Neglia, L. Giarré, G. Di Bella, A. Jean-Marie, and M. Ippolito, Large Scale Control of Deferrable Domestic Loads in Smart Grids, IEEE Transactions on Smart Grid, vol. 9, no. 2, 2016.

## Mission confiée

The question we ask ourselves is:

given a set of available geographically distributed computing units with different energy efficiency and a different mix of energy sources to exploit, how many resources should we allocate and where?

These decisions need to be taken dynamically, as the availability of renewable energies from the sun or the wind changes over short timescales and the amount of resources needed may be a function of the current algorithmic progress of the optimization algorithm (see e.g. [5]).

In particular, we will consider consensus-based distributed optimization approaches [6]. They differ from the usual parameter server framework because each computing unit

1) keeps updating a local version of the parameters and

2) broadcasts its updates only to a subset of nodes (its neighbors).

The remarkable advantages of consensus methods are their flexibility to select the communication topology and to allow some computing nodes to participate only occasionally in the training. These features allow us to reduce the energy footprint of ML training by reducing the amount of communications and activating some computing units only when it is needed.

Giovanni Neglia has started exploring the trade-off between convergence time of consensus methods and communication requirements in [7,8,9] and load balancing among micro-datacenters powered by renewable energy sources in [10]. He has also worked on the control of electrical loads in smart grids [14,15].

## Principales activités

Working toward publications.

It is possible to be involved in PhD and master students' supervision.

This offer is part of a collaboration between the NEO research team and the company Accenture Labs based in Sophia Antipolis. The candidate will be co-supervised, and hosted mainly at Accenture Labs for the duration of the project.

## Compétences

The working language is English.

## Avantages

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

## Rémunération

Gross Salary: 2653 € per month