



Offre n°2025-08941

Post-Doctorant F/H Preuves de programmes flottants sur GPU

Type de contrat : Fixed-term contract

Contrat renouvelable : Oui

Niveau de diplôme exigé : PhD or equivalent

Fonction : Post-Doctoral Research Visit

Niveau d'expérience souhaité : Recently graduated

A propos du centre ou de la direction fonctionnelle

Le centre de recherche Inria de Saclay a été créé en 2008. Sa dynamique s'inscrit dans le développement du plateau de Saclay, en partenariat étroit d'une part avec le pôle de l'Université Paris-Saclay et d'autre part avec le pôle de l'Institut Polytechnique de Paris. Afin de construire une politique de site ambitieuse, le centre Inria de Saclay a signé en 2021 des accords stratégiques avec ces deux partenaires territoriaux privilégiés.

Le centre compte 39 équipes-projets, dont 27 sont communes avec l'Université Paris-Saclay ou l'Institut Polytechnique de Paris. Son action mobilise plus de 600 personnes, scientifiques et personnels d'appui à la recherche et à l'innovation, issues de 54 nationalités.

Le centre Inria Saclay - Île-de-France est un acteur essentiel de la recherche en sciences du numérique sur le plateau de Saclay. Il porte les valeurs et les projets qui font l'originalité d'Inria dans le paysage de la recherche : l'excellence scientifique, le transfert technologique, les partenariats pluridisciplinaires avec des établissements aux compétences complémentaires aux nôtres, afin de maximiser l'impact scientifique, économique et sociétal d'Inria.

Contexte et atouts du poste

Avec l'Intelligence Artificielle, les GPUs deviennent de plus en plus spécialisés pour l'accélération de Réseaux de Neurones Profonds. De ce fait, une grande partie de la capacité de calcul de GPUs récents est basée est réservée aux calculs de produits de matrices.

L'objectif est de mener une recherche sur les garanties que l'on peut obtenir lorsque des programmes flottants sont exécutés sur des GPUs.

Des déplacements en conférence sont prévus, les frais de déplacements seront pris en charge dans la limite du barème en vigueur.

Mission confiée

Avec l'aide de Sylvie Boldo, la personne recrutée sera amenée à essayer de prouver des programmes sur GPU en partant des propriétés qu'on peut inférer du hardware. La principale difficulté est le manque de standards. Les formats de calculs utilisés par ces applications (FP8, FP16, BF16) ne sont pas tous décrit par le standard IEEE et donc ne sont pas contraints. Un nouveau standard pour les tout petits flottants est en cours d'être rédigé, mais sera probablement assez laxo sur les questions de précision. Et de toute façon, l'industrie avance plus vite que IEEE, avec de nombreuses modifications en parallèle et est souvent secrète sur la précision et les algorithmes utilisés.

[1] IEEE, 754-2019 - IEEE Standard for Binary Floating-Point Arithmetic, juin 2019.

[2]N. Wang, J. Choi, D. Brand, C.-Y. Chen et K. Gopalakrishnan, "Training Deep Neural Networks with 8-Bit Floating Point Numbers," in 32nd International Conference on Neural Information Processing Systems (NIPS), 2018.

[3]Intel, BFLOAT16 – Hardware Numerics Definition Revision 1.0, nov. 2018.

[4]I. S. P. W. Group, Interim Report on Binary Floating-point Formats for Machine Learning,

<https://github.com/P3109/Public/blob/main/Shared%20Reports/IEEE%20WG%20P3109%20Interim%20Report.pdf>, sept. 2023.

[5]O.-C. Project, OCP 8-bit Floating Point Specification (OFP8),

<https://www.opencompute.org/documents/ocp-8-bit-floating-point-specification-ofp8-revision-1-0-2023-12-01-pdf-1>, juin 2023.

[6]O.-C. Project, OCP Microscaling Formats (MX) Specification,

<https://www.opencompute.org/documents/ocp-microscaling-formats-mx-v1-0-spec-final-pdf>, sept. 2023.

Principales activités

Le candidat devra étudier les GPU, à la fois en lisant les articles et brevets, mais aussi possiblement en faisant du rétro-engineering pour deviner les précisions et calculs effectués.

Cette analyse fera ensuite l'objet d'une formalisation, permettant son usage dans la vérification de programmes.

Compétences

- bon niveau d'anglais (écrit, oral)
- une connaissance des GPU ou de Rocq est souhaitée

Avantages

- Restauration subventionnée

- Transports publics remboursés partiellement
- Congés: 7 semaines de congés annuels + 10 jours de RTT (base temps plein) + possibilité d'autorisations d'absence exceptionnelle (ex : enfants malades, déménagement)
- Possibilité de télétravail (après 6 mois d'ancienneté) et aménagement du temps de travail
- Équipements professionnels à disposition (visioconférence, prêts de matériels informatiques, etc.)
- Prestations sociales, culturelles et sportives (Association de gestion des œuvres sociales d'Inria)
- Accès à la formation professionnelle
- Sécurité sociale

Rémunération

Salaire : 2.788 euros brut/mois

Informations générales

- **Thème/Domaine** : Proofs and Verification
Software engineering (BAP E)
- **Ville** : Gif sur Yvette
- **Centre Inria** : [Centre Inria de Saclay](#)
- **Date de prise de fonction souhaitée** : 2025-10-01
- **Durée de contrat** : 11 months
- **Date limite pour postuler** : 2025-09-30

Contacts

- **Équipe Inria** : [TOCCATA](#)
- **Recruteur** :
Boldo Sylvie / Sylvie.Boldo@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 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 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 200 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

Il est nécessaire d'avoir de la curiosité et de l'appétence sur les aspects à la fois matériels et formels de l'arithmétique des ordinateurs.
Une thèse dans l'un de ces domaines serait un atout.

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.

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.