

Postdoctoral Position – Mixed Effects Neural Networks for Genome Interpretation

We are looking for a motivated postdoctoral researcher to join the [AI for Genome Interpretation \(AI4GI\)](#) group at the IGMM (CNRS, Montpellier) for **18 months**. The project is a collaboration between IGMM and IMAG, at the interface of genetics, bioinformatics, statistics, machine learning and deep learning.

The project

Motivation: Interpreting the genome means modeling the relationship between genotype and phenotype, which is the fundamental goal of biology. Achieving this could revolutionize genetics, medicine, and agricultural technology, leading for example to the development of better crops, able to face the challenges posed by global warming.

Objectives: This project is an interdisciplinary effort at the frontier between Biology (Genetics, Genomics), Bioinformatics, Artificial Intelligence (Neural Networks) and Statistics (LMMs). The aim is to join the Bioinformatics expertise of Dr. Raimondi on the development of GI NN methods and their application to relevant biological problems with the expertise of Dr. Bry and Dr. Trottier on the statistical inference of Linear Mixed Models (LMMs).

The project's goal is to develop a new breed of Mixed Effects Neural Networks (MENN) for Genome Interpretation that take the best from both worlds, merging the flexibility and power of NNs with the ability of LMMs to robustly learn from structured and noisy (non i.i.d.) data, applying them on the prediction of both **plants** and **human phenotypes**.

These models will combine the flexibility of neural networks with the statistical robustness of linear mixed models to tackle one of biology's most fundamental questions: *how do genetic variants determine phenotypes?*

The postdoc will:

- Start by familiarizing with existing research and methods for genome interpretation (GI NNs, LMMs, GWAS).
- Familiarize with the sequencing data
- Develop and benchmark MENN prototypes on sequencing datasets (WES/WGS), starting first from model organisms and then working on disease risk prediction in humans.

Candidate profile: We are looking for a **motivated** and curious candidate, with a **strong passion for science** and for scientific discovery through the use and creation of new neural networks and machine learning methods.

Bioinformatics and Genome Interpretation are multi-disciplinary and rapidly evolving fields. Therefore, the candidate is expected to 1) be eager to **continuously learn** new skills, methods and concepts, and 2) to **enjoy finding new solutions** in the face of new and unforeseen difficulties.

The ideal candidate has very good 1) python programming skills, 2) understanding of the **mathematical foundations and principles of Machine Learning, Linear Algebra** (vectorial and matricial operations, optimization), with a particular focus on Neural Networks, 3) **problem solving skills**, 4) familiarity with **GNU/Linux** environment.

A good understanding of the basic concepts of Bioinformatics is not necessary but welcome. The project will consist in developing un-orthodox Neural Network models with Pytorch.

At least the B2 level of English is required.

Skills required

We are looking for someone with:

- **Strong background in neural networks, machine learning,** linear algebra and an understanding of statistics.
- **Solid programming skills in Python** and in **scientific computing** (PyTorch, scikit-learn, numpy, etc).
- Familiarity with **GNU/Linux**.
- **Problem solving skills.**
- Good communication and teamwork skills.
- Knowledge of linear/mixed models is a plus.
- Familiarity with GWAS, population genetics, or bioinformatics pipelines are a plus.
- Experience with the processing of genomic biological data (whole exome or genome sequencing) is a plus

Practical details

- Location: IGMM, Montpellier (with joint supervision at IMAG).
- Duration: **18 months**.
- Starting date: flexible, but the candidate must be selected **before** the end of 2025.

If you're interested in working at the crossroads of AI, statistics, and genomics—and in developing new methods rather than just applying existing ones—we'd like to hear from you.

Applications should be made at this link:

<https://emploi.cnrs.fr/Offres/CDD/UMR5535-SARADE-091/Default.aspx?lang=EN>

Post-doctorat – Réseaux de Neurones à Effets Mixtes pour l’Interprétation du Génome

Nous recherchons un chercheur postdoctoral motivé pour rejoindre le groupe *AI for Genome Interpretation (AI4GI)* à l’IGMM (CNRS, Montpellier) pour 18 mois. Le projet est une collaboration entre l’IGMM et l’IMAG, à l’interface de la génétique, de la bioinformatique, des statistiques, de l’apprentissage automatique et de l’apprentissage profond.

Le projet

Motivation : Interpréter le génome signifie modéliser la relation entre génotype et phénotype, ce qui constitue l’objectif fondamental de la biologie. Réussir cela pourrait révolutionner la génétique, la médecine et les technologies agricoles, menant par exemple au développement de meilleures cultures, capables de faire face aux défis posés par le réchauffement climatique.

Objectifs : Ce projet est un effort interdisciplinaire à la frontière entre Biologie (Génétique, Génomique), Bioinformatique, Intelligence Artificielle (Réseaux de Neurones) et Statistiques (LMMs). L’objectif est de joindre l’expertise en Bioinformatique du Dr. Raimondi sur le développement de méthodes de réseaux de neurones pour l’interprétation du génome et leur application à des problèmes biologiques pertinents avec l’expertise du Dr. Bry et du Dr. Trottier sur l’inférence statistique des modèles linéaires à effets mixtes (LMMs).

L’objectif du projet est de développer une nouvelle génération de *Réseaux de Neurones à Effets Mixtes (MENN)* pour l’interprétation du génome, qui prennent le meilleur des deux approches, en fusionnant la flexibilité et la puissance des réseaux de neurones avec la capacité des LMMs à apprendre de manière robuste à partir de données structurées et bruitées (non i.i.d.), en les appliquant à la prédiction de phénotypes végétaux et humains.

Ces modèles combineront la flexibilité des réseaux de neurones avec la robustesse statistique des modèles linéaires à effets mixtes pour aborder l’une des questions les plus fondamentales de la biologie : comment les variantes génétiques déterminent-elles les phénotypes ?

Le post-doctorant devra :

- Commencer par se familiariser avec les recherches et méthodes existantes pour l’interprétation du génome (réseaux de neurones GI, LMMs, GWAS).
- Se familiariser avec les données de séquençage.
- Développer et évaluer des prototypes de MENN sur des ensembles de données de séquençage (WES/WGS), en commençant par des organismes modèles puis en travaillant sur la prédiction du risque de maladies chez l’humain.

Profil du candidat :

Nous recherchons un candidat motivé et curieux, avec une forte passion pour la science et pour la découverte scientifique à travers l’utilisation et la création de nouveaux réseaux de neurones et de nouvelles méthodes d’apprentissage automatique.

La bioinformatique et l’interprétation du génome sont des domaines multidisciplinaires et en évolution rapide. Par conséquent, le candidat devra :

1. être désireux d'apprendre continuellement de nouvelles compétences, méthodes et concepts,
2. aimer trouver de nouvelles solutions face à des difficultés nouvelles et imprévues.

Le candidat idéal possède :

1. de très bonnes compétences en programmation Python,
2. une bonne compréhension des fondements et principes mathématiques de l'Apprentissage Automatique, de l'Algèbre Linéaire (**pytorch**, **scikit-learn**, **numpt**), avec un accent particulier sur les Réseaux de Neurones,
3. des compétences en résolution de problèmes,
4. une familiarité avec l'environnement GNU/Linux.

Une bonne compréhension des concepts de base en bioinformatique n'est pas nécessaire mais appréciée. Le projet consistera à développer des modèles non conventionnels de réseaux de neurones avec PyTorch.

Un niveau d'anglais d'au moins B2 est requis.

Compétences requises

Nous recherchons une personne avec :

- Une solide formation en réseaux de neurones, apprentissage automatique, algèbre linéaire et une compréhension des statistiques.
- De solides compétences en programmation Python et en calcul scientifique (PyTorch, scikit-learn, numpy, etc.).
- Une familiarité avec GNU/Linux.
- Des compétences en résolution de problèmes.
- De bonnes compétences en communication et en travail en équipe.
- La connaissance des modèles linéaires/mixtes est un plus.
- La familiarité avec les GWAS, la génétique des populations ou les pipelines de bioinformatique est un plus.
- Une expérience dans le traitement de données biologiques génomiques (séquençage d'exome complet ou de génome complet) est un plus.

Détails pratiques

- Lieu : IGMM, Montpellier (avec co-encadrement à l'IMAG).
- Durée : **18 mois**.
- Date de début : flexible, mais le candidat doit être sélectionné avant fin 2025.

Si vous êtes intéressé à travailler à l'intersection de l'IA, des statistiques et de la génomique, et à développer de nouvelles méthodes plutôt qu'à simplement appliquer celles existantes, nous aimerions avoir de vos nouvelles.