Offer #2024-07904

Doctorant F/H Foundation Models of human brain function

The offer description below is in French

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

Level of experience : Recently graduated

About the research centre or Inria department

The Inria Saclay-Île-de-France Research Centre was established in 2008. It has developed as part of the Saclay site in partnership with Paris-Saclay University and with the Institut Polytechnique de Paris since 2021.

The centre has 39 project teams , 27 of which operate jointly with Paris-Saclay University and the Institut Polytechnique de Paris. Its activities occupy over 600 scientists and research and innovation support staff, including 54 different nationalities.

Context

One of the major directions for future neuroscience is to build on the expertise accumulated in AI-powered
cognitive systems, such as architectures that process language or visual content, but in the future will also
include motor actions, planning and navigation. While both AI and neuroscience will benefit from comparing
brain activity data with AI systems, one difficulty is that the links between AI models and brain activity
have been made in very specific contexts [1, 2, 3] and may not generalise beyond a few standard situations
(static images, language, sounds).
A recent and beneficial shift in recent years has been the development and public sharing of large-scale brain
imaging datasets, whether performed on large populations [4, 5, 6] or on small groups of individuals but with
very large amounts of data available [7, 8, 9] and http://www.cneuromod.ca – a context known as deep
phenotyping. Given the availability of such data, which are only partially or inconsistently annotated [10],
the question is: can one identify core structures of these networks that would provide relevant primitives for
fitting AI models?

Assignment

In this PhD, we aim to build basic models of brain function using large-scale imaging resources, from which
a network of components would be identified.
• Components refer to dictionary-like decompositions of brain data, factorising the data into sparse but
structured regions as in [11]. These topographies should be multiscale [12] and, unlike previous work,

could be provided with a precise semantic specification [13, 14, 10] inherited from the annotations
associated with the brain data.
• Network refers to a graphical model underlying the properties of these components, which would de-
scribe the interactions that exist between them. The network structure can be learned together with
the above components.
• The model should be adaptive to new individuals (personalization): given some new brain imaging data,
it should be relaxed to accommodate idiosyncrasies of the new data, based on the known properties of
the components and the features visible in the data.
While dictionary learning has already proven to be a powerful tool, the semantic, network and adaptive
features of this decomposition are novel. For the adaptive part, we will rely on a optimal transport approach
recently developed in our group, which has been shown to be the most powerful in mapping inter-individual
variability [15].

Main activities

The work will start by collecting as many publicly available datasets as possible to enable large-scale learning.
We will then develop a multi-resolution dictionary learning strategy similar to [11] to 2k or even 4k components
(the most accurate models are currently only 1k). We will augment the learned model with contrastive
strategies that produce merging of the data, revealing statistical relationships between models. For this we
will draw inspiration fro [19, 16].
After this step, we will extend the model with the following features (these tasks can be performed in parallel
and are weakly interdependent).
• We will associate semantics with components by investigating stable predictive associations between
cognitive annotations of the data and the brain components whose signal predicts the occurrence of
these labels across datasets.
• We will learn the dependence structure of the model; this can be summarised by a classical covariance
structure, but embeddings produced by contrastive models actually contain an implicit dependence
structure that could be captured by conditional distributions of the signals in the model components.
• We will propose a scheme for adapting the model to new individual data based on the model in [15].
Unlike existing tools, this correspondence model should be able to account for all available information:
not only image domain signals, but also semantics and connectivity.
Technical developments To this end, we will revisit and extend the large-scale dictionary learning model
of [11, 12], and investigate non-linear contrastive variants of these models. We will use large-scale inference
tools to learn clean semantic associations from the data.
Validation on brain imaging analysis tasks Validation is a core step of the procedure, necessary to
validate any successive improvements to the initial model. It will consist of a selection of several core imaging
tasks that represent the expected virtues of the sought brain model: i) good representation of the existing
data, whether they come from public image databases not used to build the model, or from the existing
literature; ii) perform decoding at scale with good accuracy [10]; iii) fit of the AI model, as measured by the
amount of explained variance.

Skills

Compétences techniques et niveau requis : Maitre de Python scientifique, neuroimagerie, apprentissage automatique.

Langues : anglais

Compétences relationnelles : -

Compétences additionnelles appréciées :-

Benefits package

- 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 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

Remuneration

Minimum remuneration: 2,100 € gross/month

Order of August 29, 2016 setting the remuneration of contract doctoral students

General Information

- **Theme/Domain :** Computational Neuroscience and Medicine
  Biologie et santé, Sciences de la vie et de la terre (BAP A)
- **Town/city :** Palaiseau
- **Inria Center :** Centre Inria de Saclay
- **Starting date :** 2024-08-01
- **Duration of contract :** 3 years
- **Deadline to apply :** 2024-07-31

Contacts

- **Inria Team :** MIND
- **PhD Supervisor :**
  Thirion Bertrand / Bertrand.Thirion@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

The keys to success

The successful candidate will be interested in applications of machine learning and in understanding human
cognition. Note that the work will take place in a multidisciplinary environment (physics, neuroscience,
computer science, modelling, psychology), neuroscience, computer science, modelling, psychology).

Prior experience on deep model is a major asset, as it makes it easier for the candidate to understand
the concepts and tools involved. Knowledge of scientific computing in Python (Numpy, Scipy, Pytorch) is
required. All the work will be done in Python based on standard machine learning libraries and the Nilearn
library for neuroimaging aspects. The candidate will benefit from the numerous development of the Mind and Unicog teams for computational facilities and expertise in the various domains involved (machine learning,
optimization, statistics, neuroscience, psychology).

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

Instruction to apply

Defence Security :
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST).Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :
As part of its diversity policy, all Inria positions are accessible to people with disabilities.