

# A Fully Funded PhD Scholarship

in machine learning and large language models (LLMs)

We invite highly motivated candidates to apply for a fully funded PhD scholarship to join PD. Dr Nazanin Sédille' (<https://www.chuv.ch/fr/laboratoires/dl-home/le-departement-en-bref/unites/service-de-chimie-clinique>) and Professor Oliver Y. Chén's teams ([www.oliverychen.com](http://www.oliverychen.com)). We work on projects related to: (a) building new machine learning and statistical methods for studying large-scale biological and medical data; (b) disease prediction; (c) digital health. For this PhD scholarship in particular, please see details below. The student will have joint affiliations with the Lausanne University Hospital (CHUV) and the University of Lausanne.

## I. Contexte: What does our groups do?

We develop new machine-learning and statistical methods and study large-scale data in health and disease. Our data are from diverse sources from the Service of Clinical Chemistry (<https://www.chuv.ch/fr/laboratoires/dl-home/le-departement-en-bref/unites/service-de-chimie-clinique>).

Our focus is threefold. (a) Building new, methodologically exciting models to address real-world problems; (b) using these methods to (i) study the interplays between large-scale multimodal, multivariate, high-dimensional features, and when/how they may be associated with diseases cross-sectionally and longitudinally and (ii) identify markers that support patient diagnosis and prognosis; (c) translating our algorithms into clinical decision support and patient health management apps.

## II. Mission

With this full scholarship, the PhD student will primarily work on three interlinked projects in collaboration between CHUV and UNIL on data related to cardiovascular and metabolomic diseases.

- Building better biomarkers for cardiovascular and metabolomic diseases.** Large language models (LLMs) designs oftentimes rely on machine learning and biology in relative isolation. Here, we aim to design LLMs to discover biomarkers for cardiovascular and metabolomic diseases using insights from both machine intelligence and biology. Using these new methodological frameworks, we aim to identify cardiovascular biomarkers (e.g., ECG and echocardiogram) related to heart disease risk factors and clinical outcomes such as disease onset or progression. In parallel, we aim to identify diabetic biomarkers (e.g., glucose metabolomic) related to, for example, renal insufficiency, and clinical outcomes such as diabetic onset or progression. Finally, we aim to quantify the disease pathways from risk factors to clinical outcomes via the discovered biomarkers.
- Disease prediction using multimodal data.** Large language models (LLMs) for disease prediction using single-modal data may overlook the comprehensive disease landscapes underpinned by multiple data sources. Here, we will design LLMs to identify, from multivariate, multimodal, potentially high-dimensional biomarkers, that can together improve the overall disease prediction accuracy as well as clinical explanation of the biomarkers.
- Longitudinal data analysis and early disease prediction.** Here, knitting expertise from machine learning and clinical chemistry, we aim to develop new methods that can (1) unveil the longitudinal trajectories of the disease profile, (2) forecast future disease progression, and (3) inform targeted and more timely disease management, treatment and prevention.

- The students will, if interested, [collaborate](#) with colleagues in other projects within and across teams.
- The students [have the freedom](#) to propose and develop [independent studies](#) within the broader aims of this Scholarship and collaborate with or visit other teams.
- The students will work in an [interdisciplinary](#), [multicultural](#) environment.
- The positions, once filled, may start immediately.

### III. Profile: What are we looking for?

#### Minimum qualifications:

- A [master's degree](#) and an [undergraduate degree](#) in disciplines relevant to (applied) mathematics, computer science, engineering, machine learning, or statistics.
- An interest in developing [new methods](#) and [applications](#) and employing them to address [real-world healthcare-related problems](#).
- An interest in [data visualisation](#).
- A [team player](#).
- The working language of the group is [English](#).

#### Desired qualifications:

- Strong programming skills related to machine learning, longitudinal methods, and large language models (LLMs).
- Experience in machine learning (LLMs), statistical modelling, and version control.

### IV. Nous offrons: What do we offer?

- Full scholarships that cover your [tuition](#) plus an [annual salary](#) (SNF salary scale).
- Joint affiliations with the [Lausanne University Hospital \(CHUV\)](#) and the [University of Lausanne](#).
- An [interdisciplinary](#) environment, and a [supportive](#) team. We strive for [equality](#), [diversity](#), and [inclusion](#). Our team is interdisciplinary and multicultural, and we encourage underrepresented students to apply.
- Possibility to collaborate with and visit [external colleagues](#) at Johns Hopkins University, KU Leuven, University of Bristol, University of Oxford, University of Pennsylvania, Vrije Universiteit Brussel, and Yale University.
- Access to [courses](#) from the CHUV and the University of Lausanne.

### V. Contact et envoi de candidature: How to apply?

Please send PD. Dr. Nazanin Sédille ([nazanin.sedille@chuv.ch](mailto:nazanin.sedille@chuv.ch)) and Professor Oliver Y. Chén ([olivery.chen@chuv.ch](mailto:olivery.chen@chuv.ch)) the following.

1. A motivation letter (no more than one page).
2. A CV.
3. Copies of your undergraduate and master's theses.
4. Contact information for three references.