

Title: Postdoctoral Researcher in Artificial Intelligence and Natural Language Processing (SCAI/BnF research program)

Body: 12-month postdoctoral contract, renewable)

Attachment: UMR 7222 ISIR

Keywords: machine learning, explainability, databases, computer science, applied mathematics, statistics, natural language processing, recommendation

Who are we?

Sorbonne University is a multidisciplinary research university created on January 1, 2018 by merging the universities Paris-Sorbonne and UPMC. Deploying its training to 54,000 students including 4,700 doctoral students and 10,200 foreign students, It employs 6,300 teachers, teacher-researchers and researchers and 4,900 library, administrative, technical, social and health staff. Its budget is 670 M€. Sorbonne University has a first-rate potential, mainly located in the heart of Paris, and extends its presence in more than twenty sites in Île-de-France and in the regions. Sorbonne University is organized into three faculties: Humanities, Science & Engineering and Medicine. These faculties have significant autonomy to implement the university's strategy within their own boundaries, based on a contract of objectives and resources. University governance is primarily devoted to promoting the university's strategy, steering, developing partnerships and diversifying resources.

Presentation of the project

In a national and international context marked by competition around artificial intelligence, Sorbonne University has created the "Sorbonne Center for Artificial Intelligence" (SCAI), which brings together in a single location, located in the heart of the Latin Quarter, a strategic range of disciplines in modern artificial intelligence. The ambition of SCAI is to contribute significantly to the excellence of interdisciplinary research in artificial intelligence by promoting exchanges between professors, researchers, teachers, students and industrialists.

The research project described below is part of the strategic partnership between Sorbonne University and the BnF, which brings together the expertise of the MLIA team of ISIR at the BnF in order to develop a joint research work on the subject of recommender systems.

The Bibliothèque nationale de France (BnF) is one of the largest heritage libraries in the world. Its mission is to collect, catalog, preserve, enrich and communicate the national documentary heritage. For many years now, BnF has been involved in ambitious digitization programs for its collections, to which we can now add the massive entry of natively digital collections. BnF is constantly enriching its digital heritage, the mass, diversity and rate of growth of which require new processing and consultation tools. To enable as many people as possible to discover and appropriate this heritage, BnF has been involved in artificial intelligence (AI) technologies for several years.

Main activities

Gallica, the digital library of the BnF, contains nearly 10 million digitized documents that are freely accessible online (18.5 million visits per year). However, most users do not know that Gallica contains not only printed documents, but also photographs, sound recordings, videos, and 3D objects. In satisfaction surveys, only a minority of users consider the search engine's answers to be relevant and a majority would like to be better guided in their searches. A recommendation system should be able to help users find their way through the mass of collections and improve the visibility of the least known. In this project, BnF is committed to adopting a resolutely ethical approach. The exploitation of user logs must respect their privacy and guarantee both the relevance and transparency of the algorithms, avoiding the risk of filter bubbles. The interface design is also at the heart of the approach: a trustworthy system relies on a good user experience and on the diversity and relevance of the proposed recommendations. Three lines of thought emerge:

- 1) based on the available data, including both user logs and collection descriptions, how to develop predictive algorithms?
- 2) how to integrate diversity in the recommendation algorithm while leaving the choice to the user to moderate his serendipity threshold?
- 3) how to build user trust in algorithm design and audit?

Main missions

This project consists in working on information access in the Gallica library, from the point of view of machine and deep learning techniques. The research axes concern (1) the analysis and indexing of textual documents as well as (2) the analysis of user traces and (3) recommendation systems. We are particularly interested in multimodal techniques that allow contextualizing a document or a query based on user interactions.

The successful candidate will be responsible for:

- Implementing models to learn the semantics of textual data for the purpose of indexing them.
- Developing algorithms based on representation learning methodologies to effectively blend text and user traces.
- Reporting and presenting development work in a clear and effective manner, both for discussion with BnF experts and writing machine learning publications.

The printed book collection will be the primary focus of the program described above, but an extension to other collections with textual descriptors (in particular iconographic collections) may be considered.

Education:

A PhD degree in Computer Science or equivalent is required, as well as a strong scientific record, particularly in NLP and/or Recommender Systems and/or Information Retrieval. Experience with international research projects and applications in SHS would be an asset.

General information:

Location: Pierre and Marie Curie campus of Sorbonne University and Datalab of the BnF

Contract: 12-month fixed-term contract with the possibility of an extension

Expected hiring date: as soon as possible

Workload: full time

Desired experience: 1 to 3 years

Salary according to experience

Main contacts:

Laure Soulier, MCF in computer science at Sorbonne University, MLIA team, ISIR.

Emmanuelle Bermès, Scientific and Technical Assistant to the Director of Services and Networks at BnF.

Jean-Philippe Moreux, Scientific expert of Gallica at the BnF.

Supervision: NO

Project management: YES

Knowledge and skills

A strong background in natural language processing or text analysis is essential, and good programming skills are required. Experience with recommender systems is assumed. An understanding of the ethical issues of such systems is also expected.

Language: knowledge of French is not required but is strongly preferred

Applications (CV + motivation + references) should be sent by email to xavier.fresquet@sorbonne-universite.fr with a copy to philippe.chevallier@bnf.fr