

## **Mini project N° 2**

### **Information Extraction from Scientific Papers**

Build a system to extract key information (e.g., authors, institutions, and keywords) from scientific papers. To achieve this, you can use and train Named Entity Recognition (NER) models (you can choose any model you want).

Start by extracting text from PDFs, then segment it into structured sections (e.g., abstract, introduction, methodology). Afterward, clean the text by removing non-informative data such as references, citations, and figures. The preprocessing step comes after that.