# Maël Houbre

*First year PhD student in Natural Language Processing*

## Education

2022–2025 **PhD Student**, *LS2N, University of Nantes*,
PhD student in NLP and information retrieval.

2016–2019 **M.S in Computer science (Engineer School)**, *Telecom Nancy*,
French computer science engineering school. Speciality in embedded software.

2014–2016 **Classes préparatoires aux grandes écoles**, *Brest Navy High School*,
Two-year undergraduate intensive course in mathematics and physics to prepare for French engineer schools' competition exams.

## PhD studies

Title *Unsupervised absent keyphrase generation for scientific indexing*

Supervisors Béatrice Daille, Florian Boudin

Description Online publishing platforms made scientific publications much easier to share, resulting in a rapid growth of the number of articles available to the community. However this increasing number of publications makes if difficult for researchers to comprehensively find relevant articles for their research. When using keyphrases as a document expansion for indexing, research have showed that keyphrases, especially keyphrases that are not in the source text, that is, absent keyphrases, have a beneficial impact on document retrieval. However, state of the art generative approaches need large datasets to be trained which makes it difficult to have efficient models for several domains and languages. The work of this thesis is therefore focused on how to generate absent keyphrases in an unsupervised manner.

## Publications

- **Maël Houbre**, Florian Boudin, Béatrice Daille: A Large-Scale Dataset for Biomedical Keyphrase Generation. *Accepted at LOUHI 2022: The 13th International Workshop on Health Text Mining and Information Analysis*

## NLP related work experience

Nov 2019–Mar 2022 **NLP Engineer**, *French Ministry of Defence*
As an NLP engineer, I was essentially in charge of the evaluation of the various algorithms and technical reports that we received during projects.
Missions:
- Develop baseline algorithms
- Prepare training and testing datasets, evaluation scripts for each task
- Evaluate models' performances and write reports

+336 17 84 30 66 • mael.houbre@univ-nantes.fr • mhoubre
https://mhoubre.github.io

Oct 2018–Mar 2019 **Graduation Project with Sopra Steria**, *Automatic paraphrase generation for data augmentation*

In my senior year in Telecom Nancy, I took part in a project with the company Sopra Steria. We were a team of three students who were in charge of developing a deep learning based paraphrase generation model for Sopra Steria's data science department.
Missions:

- Learn about word/document embedding librairies (FastText, gensim) and compare them using different similarity measures
- Develop a LSTM based generation model

## Computer science skills

- Programming languages: Python, C/C++, R
- Machine Learning librairies: PyTorch, Hugging face transformers, Scikit-learn
- Dataset librairies: Pandas, Hugging face datasets
- Scientific skills: Signal processing, Linear algebra, Probability and statistics, Functional analysis