

# MODÉLISATION STATISTIQUE POUR LES DONNÉES COMPLEXES ET LE BIG DATA

*Mise en œuvre d'un agent conversationnel pour l'apprentissage de la programmation*

**Adrien Guille - SAE6.EMS.01 - Université Lumière Lyon 2**

## Objectif

L'objectif est de mettre au point un agent conversationnel pour aider les étudiants du BUT SD dans le cadre des modules "Bases de la programmation" des semestres 1 et 2. Cet agent devra répondre à leurs questions en s'appuyant sur les supports utilisés en cours selon le principe du *Retrieval Augmented Generation* (RAG). Il devra s'exécuter localement, sans requêter aucune API distante.

## Composants du système

Le système à développer devra coordonner plusieurs composants :

- Prétraitement des supports (e.g. extraction du contenu des fichiers pdf, conversion des notebooks en texte brut)
- Découpage, vectorisation et indexation des supports
- Génération des réponses

## Questions

Le développement soulèvera diverses questions méthodologies et technologiques, entre autres :

- Comment découper les supports ?
- Comment vectoriser et indexer les supports ?
- Comment formater le prompt ?
- Quel modèle de langue utiliser pour générer les réponses ?

De nombreux outils et de nombreux modèles existent, parmi lesquels choisir.

# Livrables et évaluation

## Système

Chaque groupe déposera avant le **18 juin à 13h00**, sur Moodle, le **code** et la **documentation** nécessaire pour permettre l'exécution du système mis au point. La qualité de la réalisation sera notée à travers l'évaluation des réponses apportées par le système à un jeu de questions prédéfinies.

*Nota bene* : toute difficulté rencontrée pour exécuter le système entraînera une pénalité proportionnelle.

## Soutenance

Chaque groupe présentera le système développé lors d'une **soutenance**, le **19 juin entre 9h00 et 12h00**. Le **support** utilisé durant la présentation sera déposé sur Moodle avant le **19 juin 12h30**. La compréhension du système mis au point sera évaluée à travers la pertinence de la présentation et l'échange qui suivra.

*Nota bene* : les notes de soutenances seront individualisées si nécessaire.