

# Labellisation automatique de documents pédagogiques en vue d'une adaptation accessible

Lucas Mariétan

## Contexte et problématique

Dans le but d'assister la transcription de documents pédagogiques pour permettre aux personnes à déficience visuelle d'avoir des supports accessibles, comment automatiser la labellisation de ces documents à l'aide de modèle de deep learning générique ?

## Démarche

### Compréhension métier

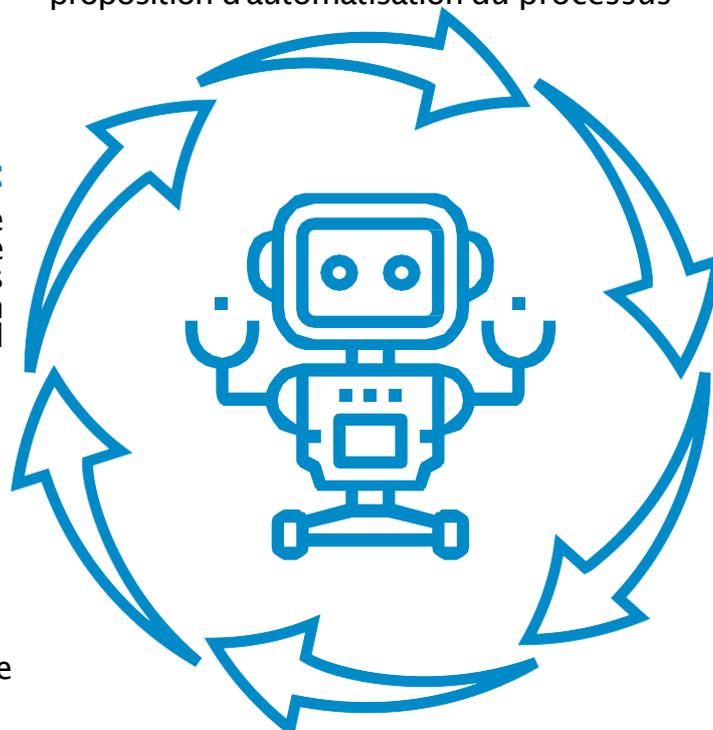
Labellisation manuelle énergivore → proposition d'automatisation du processus

### Compréhension données

Analyse de différents documents pédagogiques et compréhension de leur structure

### Préparation données

Création d'un jeu de données d'images représentatifs de la réalité selon des critères de sélection établi en amont



### Déploiement

Intégration de l'intelligence artificielle au début du processus de labellisation actuellement manuel

### Evaluation

Sélection du modèle Faster RCNN de PubLayNet et analyse par classe avec les métriques mAP, AR, courbe PxR

### Modélisation

Création de la Ground Truth, test sur différents modèles de deep learning (PubLayNet, PrimaLayout, HJDataset, NewspaperNavigator)

## Résultat

Automatisation partielle de la labellisation avec une précision moyenne de 20% avec une intersection sur union de 50%