

# Équipe Data — Tâches et Sous-tâches

Projet : Fader Networks

---

## 1. Objectif général

L'équipe Data est responsable de :

- la récupération du dataset CelebA,
- la création d'une structure de données propre et exploitable,
- le prétraitement complet des images et des attributs,
- la construction d'un pipeline PyTorch (DataLoader),
- la validation du pipeline sur un modèle Fader préentraîné,
- la documentation officielle des données.

## 2. 1. Récupération et organisation des données

### 2.1. 1.1 Téléchargement du dataset CelebA

- Téléchargement des ressources officielles depuis le site du MMLab.
- Fichiers indispensables :
  - Images alignées (Align&Cropped Images),
  - Fichier des attributs : `list_attr_celeba.txt`,
  - Fichier des partitions : `list_eval_partition.txt`.
- Fichiers optionnels :
  - `list_landmarks_celeba.txt`,
  - `identity_CelebA.txt`.

### 2.2. 1.2 Organisation de l'arborescence

- Création de la structure suivante :

```
Data/  
  images/  
  Anno/  
  Eval/  
  README.txt
```

- Vérification de l'intégrité des téléchargements.
- Déplacement des fichiers dans les dossiers adéquats.

## 3. 2. Vérification et inspection des données

### 3.1. 2.1 Cohérence images annotations

- Vérifier que les images sont nommées correctement :
  - 000001.jpg ligne 1,
  - 000002.jpg ligne 2.
- Vérifier le nombre total d'images ( 202 599).

### 3.2. 2.2 Inspection visuelle

- Afficher un échantillon aléatoire d'images.
- Vérifier :
  - qualité générale,
  - orientation correcte,
  - cohérence des attributs.

## 4. 3. Prétraitement des données

### 4.1. 3.1 Prétraitement des images

- Redimensionnement à 256 x 256.
- Normalisation des valeurs des pixels dans l'intervalle  $[-1, 1]$ .
- Conversion en tenseurs PyTorch.
- Contrôle qualité : visualisation après transformation.

### 4.2. 3.2 Prétraitement des attributs

- Chargement de `list_attr_celeba.txt`.
- Conversion des valeurs :
  - $-1 \rightarrow 0$ ,
  - $+1 \rightarrow 1$ .
- Sélection des attributs d'intérêt :
  - *Smiling*,
  - *Eyeglasses*,
  - *Male*,
  - *Young*,
  - etc.
- Construction d'un dictionnaire `image → attributs`.
- Vérification manuelle sur quelques images.

## 5. 4. Construction du DataLoader PyTorch

### 5.1. 4.1 Création de la classe CelebADataset

- Gestion du chargement des images.
- Gestion du chargement des attributs.
- Application automatique des transformations (resize, normalize, tensor).
- Chargement basé sur les partitions :
  - train,
  - validation,
  - test.

### 5.2. 4.2 Tests du DataLoader

- Affichage d'un batch d'images pour validation visuelle.
- Vérification des dimensions des tenseurs.
- Vérification de la correspondance image attributs.

## 6. 5. Validation avec un modèle préentraîné

### 6.1. 5.1 Téléchargement du modèle Fader Networks officiel

- Récupérer le modèle préentraîné depuis GitHub Facebook AI Research.
- Charger les poids dans PyTorch.

### 6.2. 5.2 Tests de reconstruction

- Passer les images dans le modèle :
  - encoder,
  - latent,
  - decoder.
- Vérifier que les images reconstruites sont cohérentes.

### 6.3. 5.3 Tests de modification d'attributs

- Modifier un attribut (ex : lunettes).
- Générer l'image modifiée.
- Vérifier le réalisme et la cohérence.
- Sauvegarder les images avant/après.

## 7. 6. Documentation finale

### 7.1. 6.1 Rédaction du README

- Décrire la structure du dossier Data/.
- Décrire les fichiers de CelebA.
- Décrire les étapes de prétraitement.

### 7.2. 6.2 Archivage des résultats

- Sauvegarde des figures générées.
- Sauvegarde des premières reconstructions du modèle officiel.
- Documentation des éventuelles difficultés rencontrées.

## Résumé des responsabilités de l'équipe Data

- Préparer un dataset propre, cohérent et utilisable par le modèle.
- Fournir un DataLoader fiable à l'équipe Implémentation.
- Valider le pipeline sur un modèle Fader officiel.
- Documenter l'ensemble du processus.