

# Hackathon : AI for Impact - Deep Learning & Data Stories

**Objectif** : Concevoir une solution IA *end-to-end* combinant **Deep Learning**, **Data Science** et **Data Visualisation** pour résoudre un problème de santé (prédiction de maladies à partir d'images médicales).

**Durée** : 02 jours

## Jour 1 – Exploration, Modélisation & Pipeline

### 1. Kick-off & Équipes

- Présentation des défis, outils recommandés (Python, TensorFlow/PyTorch, Streamlit/Power BI).
- Formation des équipes (4–5 personnes) avec rôles :
  - *Data Engineer* (préparation des données),
  - *Data Scientist* (analyse statistique),
  - *Deep Learning Engineer* (modélisation),
  - *Visualization Expert* (dashboard).

### 2. Définition du Problème & Data Preparation

- Choix du dataset.
- Nettoyage : Gestion des missing values, feature engineering.
- Analyse exploratoire : PCA, clustering, statistiques descriptives...
- *Livrable intermédiaire* : Un notebook Jupyter avec l'EDA et le preprocessing.

### 3. Développement du Modèle DL/ML

- Entraînement : Modèle de Deep Learning ou ML avancé.
- Validation : Cross-validation, optimisation hyperparamètres (Optuna, GridSearch).
- Intégration : Création d'un pipeline reproductible (MLflow/DVC).

## Jour 2 – Visualisation, Raffinement & Pitch

### 1. Data Visualisation & Storytelling

- Dashboard interactif : Outils comme Streamlit, Plotly Dash, ou Tableau.
- Storytelling : Créer un récit autour des insights (ex : "Comment notre modèle réduit les erreurs de diagnostic de 30%").
- *Astuce* : Intégrer des explications xAI (SHAP, LIME) pour la transparence.

### 2. Préparation du Pitch

- Slides : Problème, méthodologie, résultats (précision, impact sociétal).
- Démo live : Présenter le modèle et le dashboard en action.
- Repo GitHub : Code, documentation, licence.

### 3. Présentations

- **Critères d'évaluation :**
  1. **Innovation** (40%) : Complexité technique et créativité.
  2. **Visualisation** (25%) : Clarté et interactivité.
  3. **Performance** (25%) : Précision du modèle.
  4. **Collaboration** (10%) : Répartition équilibrée des rôles.
- **Attributions :**
  - Meilleur modèle DL,
  - Meilleure narration visuelle,

### Stack Technique Recommandée

- **Data & ML** : Python (Pandas, Scikit-learn), TensorFlow/Keras, PyTorch, MLflow.
- **Visualisation** : Streamlit/Plotly (pour l'interactivité), Matplotlib/Seaborn (pour les graphs statiques).
- **Cloud** : Google Colab (GPU gratuit), AWS SageMaker.
- **Collaboration** : GitHub, Notion/Miro pour le suivi de projet.

### Datasets

- [Cataract dataset](#)
- [Sperm Morphology Image Dataset \(SMIDS\)](#)
- [Nail Disease Detection](#)
- [Brain Tumor Detection](#)
- [Blood Cell images for Cancer detection](#)