

## MASTER INTELLIGENCE ARTIFICIELLE

# Reconnaissance d'Objets avec Sortie Vocale pour Personnes Malvoyantes Etat D'avancement



**Realise par**

Abderahman EL-HAMIDY  
AIT DIHIM NASSIM  
Mustapha MENSOURI

**Rapporte par**

Pr. Issam Qaffou

1<sup>er</sup> Semestre, AU : 2025-2026

## Table des matières

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Semaine 1 : Introduction &amp; Cadrage</b>                            | <b>2</b> |
| 1.1      | Présentation du sujet de recherche . . . . .                             | 2        |
| 1.2      | Revue de littérature et projets similaires . . . . .                     | 2        |
| 1.3      | Définition des objectifs précis . . . . .                                | 2        |
| 1.4      | Organisation de l'équipe et choix des outils . . . . .                   | 3        |
| 1.5      | Planning et organisation . . . . .                                       | 3        |
| <b>2</b> | <b>Semaine 2 : Préparation des Données, Prétraitement et Exploration</b> | <b>3</b> |
| 2.1      | Téléchargement et sélection du dataset . . . . .                         | 3        |
| 2.2      | Nettoyage et organisation des données . . . . .                          | 3        |
| 2.3      | Analyse des données . . . . .  | 4        |
| 2.4      | Visualisation et exploration . . . . .                                   | 4        |
| 2.5      | Prétraitements appliqués . . . . .                                       | 4        |
| 2.6      | Code de prétraitement . . . . .  | 5        |
| <b>3</b> | <b>Prochaines Étapes</b>   | <b>5</b> |
| <b>4</b> | <b>Conclusion</b>  | <b>5</b> |

# 1 Semaine 1 : Introduction & Cadrage

## 1.1 Présentation du sujet de recherche

Notre projet vise à développer un système de reconnaissance d'objets avec sortie vocale en temps réel destiné aux personnes malvoyantes. Face à la difficulté rencontrée par 2,2 milliards de personnes souffrant de déficience visuelle dans le monde pour identifier les objets du quotidien, nous proposons une solution d'assistance technologique pour renforcer leur autonomie et leur sécurité dans les activités quotidiennes.

Le système intégrera une caméra légère pour l'acquisition d'images en continu, un module de détection d'objets capable d'identifier meubles, produits, signalisation et objets courants, ainsi qu'une synthèse vocale pour fournir des descriptions audio compréhensibles et concises adaptées au contexte.

## 1.2 Revue de littérature et projets similaires

Plusieurs solutions commerciales existent déjà sur le marché, notamment :

- **Seeing AI** (Microsoft) : application mobile utilisant l'intelligence artificielle pour décrire l'environnement
- **Google Lookout** : assistant visuel pour personnes aveugles ou malvoyantes
- **Be My Eyes** : plateforme de volontariat connectant personnes malvoyantes et voyantes

Ces projets s'appuient sur des modèles de détection d'objets performants tels que YOLO (You Only Look Once) et Faster R-CNN, implémentés via des frameworks comme TensorFlow et PyTorch, avec intégration de systèmes de synthèse vocale (TTS - Text-to-Speech).

Toutefois, ces solutions présentent certaines limitations que notre projet cherche à surmonter :

1. **Dépendance à Internet** : la plupart nécessitent une connexion permanente
2. **Coûts élevés** : barrières financières à l'accessibilité
3. **Latence importante** : délais dans la génération des descriptions

## 1.3 Définition des objectifs précis

**Objectif général** : Développer un système de reconnaissance d'objets avec sortie vocale fonctionnant en temps réel.

**Objectifs spécifiques** :

- Détecter avec précision les objets courants (meubles, produits alimentaires, signalisation)
- Fournir des descriptions vocales immédiates et contextualisées
- Assurer un fonctionnement hors-ligne pour garantir l'autonomie
- Minimiser la latence pour une utilisation fluide
- Proposer une solution open-source à coût réduit
- Respecter la vie privée des utilisateurs (traitement local des données)

**Impact attendu** : Augmentation significative de l'autonomie des utilisateurs malvoyants dans leurs activités quotidiennes.

## 1.4 Organisation de l'équipe et choix des outils

La collecte des données a été répartie entre les membres de l'équipe pour assurer une grande diversité du dataset couvrant des situations quotidiennes variées :

- **Membre 1** : Mobilier (chaises, tables, lampes)
- **Membre 2** : Alimentation (bouteilles, fruits, produits alimentaires)
- **Membre 3** : Signalisation (feux de circulation, panneaux)

### Outils et technologies sélectionnés :

- **Détection d'objets** : YOLOv8 (choix prioritaire pour sa rapidité en temps réel), Detectron2 (alternative pour segmentation précise), MobileNet-SSD (option pour appareils mobiles)
- **Frameworks** : PyTorch et TensorFlow
- **Traitement d'images** : OpenCV
- **Synthèse vocale** : gTTS (qualité cloud) et pyttsx3 (fonctionnement hors-ligne)
- **Annotation** : Label Studio
- **Traitement de données** : Python, Pandas

## 1.5 Planning et organisation

Le projet a été structuré en phases successives permettant un développement itératif avec validation continue des résultats. Le sujet a été validé et le périmètre clairement défini lors de la première semaine, avec une revue rapide des outils disponibles et des projets similaires existants.

# 2 Semaine 2 : Préparation des Données, Prétraitement et Exploration

## 2.1 Téléchargement et sélection du dataset

Nous avons choisi d'utiliser des datasets publics reconnus pour leur qualité et leur diversité, complétés par une collecte de données personnalisée :

### Datasets publics utilisés :

- **COCO (Common Objects in Context)** : 80 classes, 330 000 images annotées avec bounding boxes, offrant une grande variété d'objets dans des contextes réels
- **Open Images** : 600 classes, 9 millions d'images, permettant d'enrichir la couverture des catégories d'objets

Ces sources présentent l'avantage d'une licence permissive et d'annotations de qualité professionnelle, garantissant une base solide pour l'entraînement de nos modèles.

### Collecte de données personnalisée :

En parallèle, chaque membre de l'équipe contribue à la constitution d'un dataset complémentaire en photographiant des objets spécifiques selon la répartition établie (mobilier, alimentation, signalisation). Cette approche permet d'adapter le système aux contextes d'utilisation réels et d'améliorer la pertinence des détections pour notre public cible.

## 2.2 Nettoyage et organisation des données

Le processus de préparation des données a suivi plusieurs étapes méthodiques :

### 1. Nettoyage :

- Suppression des images dupliquées pour éviter le surapprentissage
- Renommage systématique des fichiers selon une nomenclature cohérente
- Vérification de l'intégrité des fichiers images

## 2. Organisation :

- Structure du dataset selon les bonnes pratiques du machine learning :
  - **Training set** : 70% des données pour l'entraînement des modèles
  - **Validation set** : 15% pour l'ajustement des hyperparamètres
  - **Test set** : 15% pour l'évaluation finale des performances
- Utilisation de scripts Python et de la bibliothèque Pandas pour automatiser l'organisation

## 3. Annotation :

- Utilisation de Label Studio pour créer et vérifier les bounding boxes
- Annotation des données collectées personnellement selon le format requis par YOLO
- Vérification de la cohérence des annotations existantes dans les datasets publics

## 2.3 Analyse des données

### Statistiques du dataset initial :

- **Volume total** : 4 840 images annotées constituant une base robuste
- **Nombre de classes** : 50 catégories couvrant mobilier, alimentation, signalisation et électronique
- **Répartition moyenne** : environ 100 images par classe assurant une représentation équilibrée

### Analyse de la qualité :

Les datasets COCO et Open Images offrent des images de résolutions variées capturées dans des conditions réelles (éclairage naturel et artificiel, différentes perspectives, occlusions partielles). Cette diversité est essentielle pour la robustesse du modèle en situation d'utilisation réelle.

L'équilibre entre les classes a été vérifié pour éviter les biais d'apprentissage. Les classes sous-représentées ont été identifiées pour une collecte supplémentaire ciblée.

## 2.4 Visualisation et exploration

Des notebooks d'exploration ont été développés pour :

- Visualiser des échantillons d'images annotées de chaque classe
- Analyser la distribution des dimensions d'images
- Identifier les variations d'échelle des objets dans les images
- Évaluer la qualité des bounding boxes existantes

Cette phase exploratoire a permis de mieux comprendre les caractéristiques du dataset et d'identifier les défis spécifiques à relever (objets de petite taille, occlusions, conditions d'éclairage difficiles).

## 2.5 Prétraitements appliqués

Les transformations suivantes ont été implémentées pour optimiser l'entraînement :

### Prétraitements de base :

- **Redimensionnement** : standardisation des images à la résolution d'entrée du modèle (typiquement  $640 \times 640$  pour YOLOv8)
- **Normalisation** : mise à l'échelle des valeurs de pixels pour faciliter la convergence
- **Conversion des formats** : harmonisation vers un format unique compatible avec les frameworks choisis

**Data augmentation :**

- Rotations légères ( $\pm 15^\circ$ ) pour améliorer l'invariance à l'orientation
- Variations de luminosité et contraste simulant différentes conditions d'éclairage
- Flips horizontaux pour augmenter la diversité des données
- Ajout de bruit subtil pour améliorer la robustesse

Ces techniques permettent d'augmenter artificiellement la taille du dataset et d'améliorer la capacité de généralisation des modèles.

## 2.6 Code de prétraitement

L'ensemble du pipeline de préparation des données a été documenté dans des notebooks Jupyter commentés, incluant :

- Scripts de téléchargement et extraction des datasets
- Fonctions de nettoyage et organisation automatisées
- Pipelines de transformation et augmentation des données
- Outils de visualisation et d'analyse statistique
- Générateurs de données compatibles avec PyTorch et TensorFlow

Le code est structuré de manière modulaire pour faciliter la maintenance et les ajustements futurs.

## 3 Prochaines Étapes

**Semaine 3 et au-delà :**

1. **Finalisation du dataset** : compléter les annotations, équilibrer les classes et définir les protocoles d'évaluation précis
2. **Entraînement des modèles** : comparer YOLOv8, Detectron2 et MobileNet-SSD sur les critères de latence, précision et robustesse ; optimiser pour l'inférence locale
3. **Intégration de la synthèse vocale** : développer un prototype avec pyttsx3 pour le fonctionnement hors-ligne
4. **Prototype utilisateur** : créer une démonstration avec webcam et organiser des tests utilisateurs ciblés pour mesurer l'utilité, la latence et la compréhension des descriptions

**Objectif final** : Proposer une solution d'intelligence artificielle accessible, open-source et respectueuse de la vie privée pour renforcer l'autonomie des personnes malvoyantes dans leurs activités quotidiennes.

## 4 Conclusion

Les deux premières semaines du projet ont permis d'établir des fondations solides avec un cadrage précis, une revue approfondie de l'état de l'art, et la constitution d'un dataset

de qualité combinant ressources publiques et collecte personnalisée. L'organisation méthodique de l'équipe et les choix technologiques réfléchis nous positionnent favorablement pour la phase d'entraînement et de développement du prototype.

Notre approche se distingue par son engagement vers une solution hors-ligne, open-source et optimisée pour le temps réel, répondant ainsi aux limitations identifiées dans les solutions existantes. La suite du projet se concentrera sur la concrétisation technique de cette vision avec un objectif constant : maximiser l'autonomie et la qualité de vie des personnes malvoyantes.