

# Maîtriser les techniques de génération et d'Augmentation de données

## INFORMATIONS GÉNÉRALES

**Type de formation :** Formation continue

**Éligible au CPF :** Non

**Domaine :** IA, Big Data et Bases de données

**Action collective :** Oui

**Filière :** IA

**Code ACO :** CISIA

**Rubrique :** Certification ATLAS : CISIA (Actions co.)

**Code de formation :** CISIA-TTD

## € Tarifs

**Prix public :** 2000 €

### Tarif & financement :

Financement possible via les Actions Collectives ATLAS ou le Plan de Formation.

## PRÉSENTATION

### Objectifs & compétences

Maîtriser les techniques de génération et d'augmentation de données pour l'IA.  
Comprendre les méthodes avancées de transformation et enrichissement des données.  
Appliquer ces techniques pour améliorer la diversité et la qualité des jeux de données.  
Évaluer les impacts des modifications de données sur l'entraînement des modèles IA.

### Public visé

Professionnels de l'IT, Data scientists, ingénieurs en données, et professionnels de l'IA impliqués dans la préparation et le traitement des données.

### Pré-requis

Connaissances de base en science des données  
Familiarité avec les techniques de traitement des données et les concepts d'IA  
Connaissance de base en programmation (Python, R, etc.)

## Lieux & Horaires

**Durée :** 12 heures

**Rythme :** 9h30-12h30 et 14h-17h

### Délai d'accès :

Jusqu'à 8 jours avant le début de la formation, sous condition d'un dossier d'inscription complet

## Prochaines sessions

Consultez-nous pour les prochaines sessions.

## PROGRAMME

### CISIA-TTD01 : Documenter les Flux et le Cycle de Vie des Données – 12H00

#### 1. Introduction aux Techniques de Génération de Données (4 heures)

- **Concepts de Génération de Données**

- Données synthétiques
- Confidentialité différentielle
- Techniques de génération basée sur des modèles

- **Applications et Avantages**

- Cas d'usage typiques
- Avantages pour la diversité des données et la protection de la confidentialité

- **Cas Pratique : Génération de Données Synthétiques**

- Mise en œuvre de données synthétiques à l'aide de bibliothèques Python
- Démonstration de techniques de confidentialité différentielle

**Lien avec la Compétence C3 :** Préparation des données pour renforcer leur intégrité et pertinence, en tenant compte des techniques de génération adaptées aux besoins métiers et cas d'usage.

#### 2. Techniques d'Augmentation de Données (4 heures)

- **Concepts d'Augmentation de Données**

- Méthodes classiques (répétition, rotation, transformation, etc.)

- Techniques avancées (augmentation basée sur des GANs, augmentation pour la vision par ordinateur, etc.)

- **Applications et Avantages**

- Amélioration des modèles d'apprentissage
- Gestion des déséquilibres dans les jeux de données
- **Cas Pratique : Mise en Œuvre des Techniques d'Augmentation**
- Utilisation de bibliothèques comme imgaug ou Albumentations pour

l'augmentation d'images

- Démonstration de l'impact de l'augmentation sur les performances du modèle

**Lien avec la Compétence C3** : Préparation des données en utilisant les techniques d'augmentation pour répondre aux besoins métiers, renforcer l'intégrité des données, et améliorer les performances du modèle.

### **3. Intégration des Techniques de Génération et d'Augmentation de Données dans les Flux de Travail (2 heures)**

- **Intégration dans les Pipelines de Données**

- Mise en œuvre dans les processus de prétraitement des données
- Intégration avec les outils d'IA et les environnements de

développement

- **Cas Pratique : Intégration des Techniques dans un Pipeline de Données**

- Démonstration de l'intégration des techniques de génération et d'augmentation dans un pipeline de traitement de données

**Lien avec les Compétences C1 et C4** : Identifier et choisir les techniques appropriées pour répondre aux besoins métiers et aux cas d'usage, et évaluer leur pertinence pour les modèles d'IA.

### **4. Évaluation des Techniques et Retours d'Expérience (2 heures)**

- **Évaluation des Méthodes et Outils**

- Critères d'évaluation de l'efficacité des techniques de génération et d'augmentation
- Discussion sur les retours d'expérience et les meilleures pratiques

- **Questions-Réponses et Synthèse**

- Révision des concepts clés et des défis rencontrés
- Discussion interactive sur les applications pratiques et les futures améliorations

**Lien avec la Compétence C9** : Adopter une démarche d'amélioration continue en évaluant et en ajustant les techniques en fonction des retours et des évolutions des besoins utilisateurs.

Outils utilisés?: Python, TensorFlow, PyTorch, Scikit-learn, Jupyter Notebook.

## **MODALITÉS**

### **Modalités**

L'ensemble du parcours est accessible en présentiel, à distance ou mode hybride.

**Présentation théorique** en présentiel.

**Atelier pratique** avec exercices en ligne et en présentiel.

**Études de Cas** : Analyse d'applications réelles des techniques de génération et d'augmentation.

**Discussion Interactive** : Échange sur les meilleures pratiques, les défis rencontrés et les retours d'expérience.

## **CERTIFICATIONS**

A l'issue du parcours (10 modules), les candidats pourront passer le jury de certification ATLAS :

**Concevoir et implémenter une solution d'IA**