

# IA Extraction de données structurées

## Financement

Formation professionnelle continue  
Non conventionnée / sans dispositif

## Organisme responsable et contact

EVOLUTION5  
Jean-Denis Coindre  
06.29.78.66.25  
contact@evolution5.fr

## Accès à la formation

### Publics visés :

Demandeur d'emploi  
Jeune de moins de 26 ans  
Personne handicapée  
Salarié(e)  
Actif(vé) non salarié(e)

### Sélection :

Dossier

### Niveau d'entrée requis :

Sans niveau spécifique

### Conditions d'accès :

Aucune

### Prérequis pédagogiques :

Non renseigné

Contrat de professionnalisation possible ?

Non

## Objectif de la formation

• Comprendre la différence entre données non structurées et structurées. • Identifier les principales sources de données non structurées. • Apprendre les méthodes et outils pour transformer les données. • Intégrer les bonnes pratiques de qualité, d'éthique et de sécurité. • Mettre en pratique sur des cas génériques applicables à différentes spécialités : marketing, ingénierie, médical, financier...

## Contenu et modalités d'organisation

**Module 1 – Comprendre les données structurées et non structurées** Définir les notions de données structurées et non structurées Illustrer par des cas concrets en entreprise (rapports, emails, images, notes, enregistrements, etc.) Expliquer pourquoi transformer les données en informations exploitables est essentiel à la performance **Module 2 – Identifier les sources et formats de données non structurées** Analyser les textes libres (emails, contrats, rapports PDF, comptes rendus) Explorer les données visuelles et multimédia (images, vidéos, schémas techniques) Examiner les données issues de capteurs, logs systèmes ou monitoring Rechercher et exploiter les informations issues de la documentation, de la veille ou de bases externes **Module 3 – Utiliser les techniques et outils d'extraction** Appliquer les méthodes classiques (règles, dictionnaires métiers, taxonomies) Introduire le NLP (traitement automatique du langage naturel) Exploiter les grands modèles de langage (LLMs, ex. GPT) Mettre en œuvre l'OCR pour traiter les documents scannés Transformer les données extraites en formats structurés (CSV, JSON, bases relationnelles, etc.) **Module 4 – Garantir la qualité, la fiabilité et la validation** Détecter les erreurs fréquentes (OCR, ambiguïtés, doublons) Valider les extractions par relecture humaine (human-in-the-loop) Mesurer la performance avec des métriques adaptées (précision, rappel, cohérence) Adopter les bonnes pratiques pour fiabiliser les résultats **Module 5 – Respecter l'éthique, la réglementation et la sécurité** Protéger la confidentialité et la sensibilité des données Appliquer les cadres légaux (RGPD, normes sectorielles) Mettre en œuvre l'anonymisation et la pseudonymisation Assurer un hébergement sécurisé conforme aux standards internationaux **Module 6 – Mettre en pratique sur un cas concret** Extraire des informations clés d'un rapport, d'un contrat ou d'un document technique Convertir un document PDF ou une source brute en tableau structuré exploitable

Parcours de formation personnalisable ? **Oui** Type de parcours **Mixte**

## Validation(s) Visée(s)

> Attestation de fin de formation

## Et après ?

Suite de parcours

Non renseigné

## Calendrier des sessions

Numéro Carif	Dates de formation	Ville	Organisme de formation	Type d'entrée	CPF	Modalités
00625096	du 01/09/2025 au 31/12/2025	Mérignac (33)	EVOLUTION5		Non éligible	