PROGRAMME DE FORMATION

# Développer des applications d'IA générative : API de LLM, RAG et agents

*Développement & Programmation*

| RÉFÉRENCE | VERSION | ÉDITION | DOMAINE |
|---|---|---|---|
| WF-0438 | V2026-007 | 3 novembre 2026 | Développement & Programmation |

| DURÉE | MODALITÉ | PROGRAMME | SANCTION |
|---|---|---|---|
| 32 h | Distanciel asynchrone (e-learning) | 8 modules<br>37 leçons | Attestation<br>et certificat de réalisation |

## Objectifs pédagogiques

Appeler une API de modèle de langage en Python, gérer messages, paramètres, streaming, erreurs et reprises, et estimer jetons, latence et coût.

Obtenir des sorties structurées validées par un schéma et faire appeler des outils (fonctions) au modèle de façon contrôlée.

Construire une chaîne RAG : découpage, embeddings, indexation, récupération, réordonnancement et réponse citant ses sources.

Programmer un agent outillé borné : boucle d'appel d'outils, état, limites, validation humaine des actions sensibles.

Évaluer une application de LLM sur un jeu d'essai et la protéger contre l'injection de consigne et la fuite de données.

Préparer la mise en production : configuration, secrets, journalisation, supervision, cache et limites de débit.

# L'essentiel de la formation

## Public visé

Développeurs et développeuses, ingénieurs logiciel, data engineers et responsables techniques qui doivent intégrer un modèle de langage dans une application : assistant documentaire, extraction de données, agent outillé.

## Prérequis

Programmer en Python à un niveau intermédiaire : fonctions, dictionnaires, exceptions, lecture et écriture de JSON, environnement virtuel ; à défaut, la formation Python (WF-0110) en donne les bases.

Connaître les notions d'API HTTP (requête, réponse, code de statut, en-têtes) et de format JSON.

Avoir déjà utilisé un assistant d'IA générative ; les bases de la rédaction de consignes sont rappelées en début de parcours (leçons réemployées de WF-0410).

Disposer d'un ordinateur avec une connexion internet et un navigateur récent ; aucun matériel particulier, aucune clé d'API ni aucun compte payant ne sont nécessaires.

## Modalités et déroulement

Formation à distance, en ligne, à rythme libre, sur la plateforme de l'organisme : sept modules de cours, puis un module d'évaluation finale. La première leçon comprend un autodiagnostic de positionnement, non noté, qui aide à repérer les notions à rafraîchir. Chaque leçon associe une vidéo animée présentée par une narratrice et de courtes vidéos d'étape ponctuées de questions, des notions illustrées d'exemples de code Python, un scénario à embranchements, des cartes de révision, des activités interactives (classer, ordonner, associer), une mise en pratique écrite (consigne ou schéma à rédiger, métriques à calculer, plan d'architecture à argumenter…) corrigée par l'IA de la plateforme sur une grille de critères avec un corrigé modèle, et un quiz. Chaque module se clôt par un QCM de fin de module. Les cas pratiques s'appuient sur des organisations et des données fictives. Aucune installation, aucune clé d'API ni aucun compte payant n'est nécessaire ; les exemples de code peuvent être repris sur son propre poste. Accompagnement par un formateur par messagerie (réponse sous 48 h ouvrées), correction individuelle des livrables sur demande.

## Méthodes pédagogiques

- Positionnement en début de parcours
- Apports en vidéo : explications et démonstrations
- Mises en situation et scénarios à décisions
- Études de cas
- Exercices d'application et mises en pratique corrigées
- Production de livrables professionnels
- Cartes de révision
- Quiz et QCM d'auto-évaluation
- Accompagnement et retours du formateur

Répartition du temps de formation : **21 %** théorique · **51 %** pratique · **28 %** évaluation

## Moyens pédagogiques et techniques

- Plateforme de formation en ligne, accès individuel à distance
- Vidéos pédagogiques
- Modèles et supports téléchargeables
- Quiz et exercices interactifs en ligne
- Correction des productions sur grille de critères
- Messagerie avec le formateur (réponse sous 48 h ouvrées)

## Modalités d'évaluation

Autodiagnostic de positionnement en début de parcours (non noté) ; dans chaque leçon, questions intégrées aux vidéos, mise en pratique écrite corrigée sur une grille de critères et quiz de 7 à 8 questions ; QCM de fin de module à la fin de chacun des sept modules (seuil de réussite : 70 %, trois tentatives) ; évaluation finale interne : étude de cas (mise en production de l'assistant des conseillers d'un assureur fictif) corrigée sur grille, puis examen final de 30 questions en 54 minutes, deux tentatives au plus. Seuil de réussite de l'examen final : 70 % des points. Une attestation de fin de formation précise les compétences évaluées ; aucune certification, ni d'éditeur ni d'État, n'est délivrée ni préparée.

## Sanction

Une attestation de fin de formation et un certificat de réalisation sont remis au participant sous réserve de la réalisation du parcours.

Cette formation n'est pas présentée comme préparant à une certification enregistrée : elle ne comprend ni inscription ni présentation à des épreuves de certification, et n'ouvre droit à aucune délivrance de certification ni au financement par le CPF.

## Modalités et délais d'accès

La demande d'inscription peut être effectuée par téléphone, par courrier électronique ou au moyen du formulaire en ligne.

Un entretien ou un questionnaire de positionnement permet de vérifier l'adéquation de la formation avec les besoins, le niveau initial et les prérequis du candidat.

L'inscription devient définitive après validation du dossier administratif, acceptation des documents contractuels et, le cas échéant, confirmation du financement.

Les accès à la plateforme sont généralement activés sous 48 à 72 heures ouvrées après validation complète du dossier.

Les formations à distance sont proposées en entrées et sorties permanentes, sous réserve des délais ou dates imposés par le financeur ou le certificateur.

## Accessibilité aux personnes en situation de handicap

Vidéos sous-titrées et transcrites ; notions, énoncés, corrigés et exemples de code en texte lisible par les lecteurs d'écran ; navigation au clavier ; rythme libre. Référent handicap joignable à handicap@will-formation.com pour étudier toute adaptation (durée, supports, outils, modalités d'évaluation, dont la durée de l'examen final).

## Tarif de la formation

Prix : **2 320 €**

Financement : prise en charge possible par les organismes financeurs, selon les critères d'éligibilité applicables.

Qualiopi processus certifié

La certification qualité a été délivrée au titre de la catégorie d'action suivante : ACTIONS DE FORMATION. Certificat QUA25070009. WILL FORMATION — Société par actions simplifiée à associé unique au capital de 30 000 € — NDA 11757059675 (cet enregistrement ne vaut pas agrément de l'État) — SIRET 893 713 180 00014 — 01 85 09 25 99 — info@will-formation.com — www.will-formation.com

# Programme détaillé

8 modules, 37 leçons — durée totale 32 h.

### 01 Appeler une API de modèle de langage

1. Positionnement — Python, API et modèles de langage
2. Rappels : du modèle de langage à la consigne programmée
3. Appeler une API de chat en Python : messages, rôles et paramètres
4. Streaming, erreurs, reprises et maîtrise des jetons
5. Atelier de synthèse — Appeler une API de modèle de langage
6. QCM de fin de module

### 02 Sorties structurées et appel d'outils

1. Obtenir une sortie JSON et la valider par un schéma
2. Appel d'outils : déclarer des fonctions et exécuter les appels
3. Consignes système et gabarits versionnés
4. Atelier de synthèse — Sorties structurées et appel d'outils
5. QCM de fin de module

### 03 Embeddings et recherche sémantique

1. Représentations vectorielles et similarité
2. Découper et indexer un corpus documentaire
3. Base vectorielle : recherche, filtres et recherche hybride
4. Atelier de synthèse — Embeddings et recherche sémantique
5. QCM de fin de module

### 04 Construire une chaîne RAG

1. Assembler récupération et génération avec citations
2. Améliorer la récupération : réécriture et réordonnancement
3. Cas limites : réponse absente, sources contradictoires, hors périmètre
4. Atelier de synthèse — Construire une chaîne RAG
5. QCM de fin de module

### 05 Programmer des agents outillés

1. La boucle d'agent : choisir un outil, exécuter, observer
2. Planifier, mémoriser l'état et borner l'agent
3. Humain dans la boucle et actions à validation
4. Atelier de synthèse — Programmer des agents outillés
5. QCM de fin de module

### 06 Évaluer et sécuriser une application de LLM

1. Jeu d'essai et métriques d'une application de LLM
2. Évaluer un RAG : récupération, fidélité et LLM-juge
3. Sécurité : injection de consigne, fuite de données et sorties non sûres
4. Atelier de synthèse — Évaluer et sécuriser une application de LLM
5. QCM de fin de module

### 07 Mettre en production une application de LLM

1. Architecture d'un service : API, configuration et secrets
2. Journalisation, traces et supervision
3. Cache, limites de débit, montée en charge et repli
4. Atelier de synthèse — Mettre en production une application de LLM
5. QCM de fin de module

### 08 Évaluation finale du parcours

1. Étude de cas finale