

Formation Advanced Generative AI Development on AWS (AIP-C01)

■ Durée :	3 jours (21 heures)
■ Tarif inter-entreprises :	2 775,00 € HT (Présentiel) 2 220,00 € HT (Distanciel)
■ Public :	Développeurs logiciels - Professionnels techniques
■ Pré-requis :	Avoir suivi les formations "AWS Technical Essentials" ou "Generative AI Essentials on AWS" - 2 ans ou plus d'expérience dans la création d'applications de niveau production sur AWS ou avec des technologies open source, ou une expérience générale en IA/ML ou en data engineering - 1 an d'expérience pratique dans la mise en œuvre de solutions d'IA générative
■ Objectifs :	À l'issue de la formation, le stagiaire sera capable d'utiliser les services clés de la plateforme AWS et contribuer efficacement à la conception ou à l'exploitation de solutions cloud et IA sécurisées et adaptées aux besoins métiers.
■ Modalités pédagogiques, techniques et d'encadrement :	<ul style="list-style-type: none">• Formation synchrone en présentiel et distanciel.• Méthodologie basée sur l'Active Learning : 75 % de pratique minimum.• Un PC par participant en présentiel, possibilité de mettre à disposition en bureau à distance un PC et l'environnement adéquat.• Un formateur expert.
■ Modalité d'évaluation :	<ul style="list-style-type: none">• Définition des besoins et attentes des apprenants en amont de la formation.• Auto-positionnement à l'entrée et la sortie de la formation.• Suivi continu par les formateurs durant les ateliers pratiques.• Évaluation à chaud de l'adéquation au besoin professionnel des apprenants le dernier jour de formation.
■ Sanction :	Attestation de fin de formation mentionnant le résultat des acquis

■ Référence :	CLO103004-F
■ Note de satisfaction des participants :	Pas de données disponibles
■ Contacts :	commercial@dawan.fr - 09 72 37 73 73
■ Modalités d'accès :	Possibilité de faire un devis en ligne (www.dawan.fr, moncompteformation.gouv.fr, maformation.fr, etc.) ou en appelant au standard.
■ Délais d'accès :	Variable selon le type de financement.
■ Accessibilité :	Si vous êtes en situation de handicap, nous sommes en mesure de vous accueillir, n'hésitez pas à nous contacter à referenthandicap@dawan.fr, nous étudierons ensemble vos besoins

Comprendre les enjeux du développement d'applications d'IA générative sur AWS

Situer l'IA générative dans l'écosystème AWS et dans les architectures cloud modernes

Identifier les cas d'usage professionnels : génération de texte, résumé, recherche augmentée, assistants, automatisation et exploitation de contenus multimodaux

Comprendre le rôle des foundation models, des embeddings, des bases vectorielles, des prompts, des agents et des pipelines de données

Distinguer expérimentation, prototype et solution industrialisable

Identifier les contraintes de sécurité, de qualité, de supervision, de coût et de gouvernance

Atelier fil rouge : analyser un besoin métier d'IA générative et définir les composants techniques nécessaires à une première architecture AWS sécurisée et exploitable

Évaluer et sélectionner les foundation models appropriés

Comprendre les caractéristiques des foundation models et leurs principaux critères de sélection

Comparer plusieurs modèles selon les usages attendus : génération, résumé, classification, extraction, conversation ou raisonnement assisté

Prendre en compte les performances, les coûts, la latence, les limites de contexte, les contraintes de sécurité et les possibilités de personnalisation

Évaluer les risques liés aux résultats produits : hallucinations, biais, incohérences, données sensibles ou réponses non conformes

Atelier fil rouge : comparer plusieurs options de modèles pour un cas d'usage donné et justifier le choix retenu selon les critères fonctionnels, techniques, économiques et de sécurité

Concevoir et implémenter des systèmes basés sur des foundation models

Structurer les appels au modèle depuis une application ou un composant backend

Gérer les entrées, sorties, paramètres d'inférence et formats de réponse

Concevoir une logique applicative robuste autour du modèle : validation, contrôle, filtrage, traitement des erreurs et traçabilité

Intégrer les résultats produits dans un workflow métier ou applicatif

Sécuriser les accès aux services mobilisés avec une logique de moindre privilège

Atelier fil rouge : mettre en œuvre un premier composant applicatif appelant un foundation model AWS et structurer les traitements d'entrée, de sortie, d'erreur et de contrôle

Construire des pipelines complets de traitement de données multimodales

Identifier les sources de données exploitables : textes, documents, images, fichiers structurés ou contenus applicatifs

Concevoir un pipeline de collecte, nettoyage, transformation et enrichissement des données

Gérer les formats, métadonnées, règles de qualité et contrôles de cohérence

Prendre en compte les données sensibles, les droits d'accès et les exigences de confidentialité

Préparer les données pour l'indexation, la recherche, l'interrogation ou la génération de réponses

Atelier fil rouge : préparer un pipeline de données destiné à alimenter une solution d'IA générative, depuis la collecte des contenus jusqu'à leur transformation en ressources exploitables

Mettre en œuvre des solutions avancées de bases vectorielles et de RAG

Comprendre les principes des embeddings et de la recherche vectorielle

Identifier le rôle d'une base vectorielle dans une architecture de génération augmentée par récupération

Comprendre les étapes d'un pipeline RAG : ingestion, découpage, vectorisation, indexation, recherche, enrichissement du prompt et génération

Contrôler la pertinence des résultats retournés par la recherche

Réduire les risques d'hallucination en s'appuyant sur des sources maîtrisées

Atelier fil rouge : construire une chaîne RAG sur AWS à partir d'un corpus documentaire, avec indexation vectorielle, recherche de contexte et génération de réponse contrôlée

Créer et gérer des frameworks avancés d'ingénierie de prompts

Structurer des prompts adaptés aux usages : question/réponse, résumé, extraction, classification, reformulation ou génération contrôlée

Utiliser des consignes système, exemples, contraintes de format et règles de validation

Créer des bibliothèques de prompts réutilisables et maintenables

Mettre en place une démarche de versionnement et de test des prompts

Identifier les risques : prompt injection, fuite d'information, réponse non conforme ou contournement des règles

Atelier fil rouge : concevoir, tester et améliorer un ensemble de prompts pour un assistant IA métier, en intégrant contraintes de format, sécurité, contrôle des réponses et critères de qualité

Expliquer et utiliser les composants des frameworks d'Agentic AI

Comprendre le principe des agents IA et leur rôle dans l'automatisation de tâches complexes

Identifier les composants d'un système agentique : objectif, mémoire, outils, planification, exécution, contrôle et évaluation

Relier les agents IA aux services, API ou sources de données disponibles dans une architecture cloud

Mettre en place des garde-fous pour éviter les actions non maîtrisées ou les boucles d'exécution

Contrôler les appels aux outils et les droits associés

Atelier fil rouge : concevoir le scénario d'un agent IA capable d'interroger des ressources, d'appeler des outils contrôlés et de produire une réponse argumentée dans un cadre sécurisé

Implémenter la sécurité, la sûreté de l'IA et la gouvernance des solutions génératives

Comprendre les risques spécifiques aux solutions d'IA générative : hallucinations, biais, injection de prompt, fuite de données, réponses toxiques ou non conformes

Appliquer des contrôles de sécurité sur les identités, permissions, données et accès aux services AWS

Mettre en œuvre une logique de moindre privilège pour les composants applicatifs et les services mobilisés

Comprendre les mécanismes de filtrage, modération, contrôle des entrées et contrôle des sorties

Prendre en compte la conformité, la traçabilité et la responsabilité dans l'exploitation des solutions IA

Atelier fil rouge : sécuriser une solution d'IA générative en définissant les permissions, contrôles de données, garde-fous applicatifs et règles de sûreté nécessaires à un usage professionnel

Optimiser les performances, les coûts et la scalabilité

Identifier les facteurs influençant le coût d'une solution d'IA générative

Comprendre les effets du choix du modèle, du volume d'appels, de la taille des prompts, du contexte, du stockage et de la recherche vectorielle

Adapter l'architecture aux besoins de performance, de latence et de disponibilité

Prévoir des stratégies de limitation, de mise en cache ou de réduction des traitements inutiles

Analyser les compromis entre qualité des réponses, coût, rapidité et robustesse

Atelier fil rouge : analyser une architecture d'IA générative, identifier les postes de coûts et de latence, puis proposer des optimisations mesurables

Concevoir le monitoring, l'observabilité, les tests et la validation

Identifier les métriques à suivre : erreurs, latence, coût, usage, qualité des réponses, pertinence du contexte et incidents de sécurité

Mettre en place une logique de journalisation et de traçabilité des interactions

Créer des jeux de tests représentatifs des cas d'usage attendus

Mettre en place des critères de validation des réponses et des comportements

Organiser un processus d'amélioration continue de la solution

Atelier fil rouge : définir un dispositif de supervision et de validation pour une application d'IA générative

Intégrer une solution d'IA générative dans un environnement d'entreprise

Connecter la solution à des sources de données, API, applications ou workflows métiers

Prendre en compte les identités, droits, réseaux, sécurité et règles de gouvernance

Préparer les conditions de déploiement et d'exploitation

Organiser la documentation technique et fonctionnelle de la solution

Définir les rôles entre développeurs, architectes, responsables sécurité, métiers et exploitants

Atelier fil rouge : finaliser l'intégration de la solution d'IA générative dans un scénario d'entreprise

Préparer l'examen AWS Certified Generative AI Developer - Professional

Comprendre la structure de l'examen et les domaines évalués

S'entraîner sur des questions de type certification portant sur les modèles, le RAG, les bases vectorielles, les prompts, les agents, la sécurité, les coûts et l'observabilité

Repérer les pièges fréquents et les logiques de réponse attendues

Identifier les points à réviser après la formation

Atelier fil rouge : réaliser un entraînement guidé sur des questions de niveau AWS

Certified Generative AI Developer - Professional et corriger les réponses de manière argumentée

...