

PLAN DE FORMATION

Industrialiser l'IA Générative : les architectures de référence

Durée 2 jours (14 heures)**Ville** Rennes, Brest, Nantes, Paris, Montpellier, Lyon, Bordeaux, Angers, Lille, Toulouse▶ **PRÉSENTATION DE LA FORMATION**

La formation Industrialiser l'IA Générative donne les clés des architectures de référence de l'IA générative en entreprise pour concevoir des solutions robustes, sécurisées et industrialisables, en intégrant LLM, RAG, agents et le protocole MCP. La plus complète du parcours, elle se déroule sur deux jours mêlant apports d'architecture, études de cas et conception de schémas, en s'appuyant sur des briques open source de référence (LiteLLM, RAGFlow, OpenSearch, Agno, Cerbos, Langfuse) et sur les solutions du marché.

**Contact**

02 34 09 31 72
formation@nextdecision.fr
www.next-decision.fr

▶ **OBJECTIFS DE LA FORMATION**

- Cartographier les briques d'une architecture d'IA générative d'entreprise
- Choisir et dimensionner un LLM selon les contraintes : coût, latence, souveraineté
- Concevoir une architecture RAG industrielle
- Intégrer des serveurs MCP (Model Context Protocol) pour connecter modèles, outils et données
- Définir une architecture d'agents et les patterns d'industrialisation : sécurité, supervision, CI/CD

▶ **PRÉREQUIS DE LA FORMATION**

- Expérience confirmée du développement ou de l'architecture logicielle
- Connaissance des principes des LLM (Niveau 3 ou équivalent)

▶ **PUBLIC VISÉ**

Architectes solution, logiciel et data

Lead developers, ingénieurs IA/ML

CTO et responsables techniques chargés de cadrer ou bâtir des plateformes d'IA générative



▶ PROGRAMME DE LA FORMATION

- Socle LLM et architecture RAG industrielle
 - o Cadre et enjeux d'industrialisation
 - De l'expérimentation à la plateforme
 - Les briques d'une architecture d'IA générative d'entreprise
 - o Les LLM en production
 - Modèles propriétaires (OpenAI, Anthropic, Google) vs open-weight (Mistral, Llama, Qwen)
 - Inférence : moteurs auto-hébergés (vLLM, Ollama, TGI) vs SaaS externe (Azure AI Foundry) — arbitrage coût / souveraineté
 - Gateway LLM, routage, sécurité et budgets : LiteLLM
 - Prompting, fine-tuning, distillation : quand utiliser quoi
 - o Architecture RAG de référence
 - Pipeline d'ingestion, vectorisation, indexation : LlamaIndex, LangChain ou RAGFlow
 - Recherche hybride, reranking, gestion de la fraîcheur
 - Base vectorielle et scalabilité : OpenSearch ; alternatives pgvector, Qdrant, Weaviate
 - Étude de cas : schéma d'architecture RAG d'entreprise
- MCP, agents et industrialisation
 - o MCP Servers : connecter modèles, outils et données
 - Le Model Context Protocol : un standard ouvert, indépendant d'un fournisseur
 - Standardiser l'accès des modèles aux outils et sources de données (serveurs et clients)
 - Mise en œuvre : serveurs MCP via développement spécifique ou la fonctionnalité MCP de n8n
 - Sécurité, périmètres d'accès et gouvernance des connexions
 - Étude de cas : intégration de serveurs MCP dans un SI existant
 - o Architectures d'agents
 - Patterns d'agents : orchestration, planification, appel d'outils via MCP
 - Frameworks et runtime d'agents : LangChain / LangGraph, CrewAI, AutoGen ou Agno
 - Isolation de l'exécution du code des agents (sandboxing)
 - Habilitations et accès aux outils : Cerbos, OpenFGA ou Permit.io
 - Guardrails, supervision et limites de coût
 - o Industrialisation (LLMOps)
 - Observabilité et évaluation continue (ex. Langfuse, LangSmith) ; centralisation des logs (ex. Fluentd + OpenSearch)
 - Garde-fous natifs de la passerelle LiteLLM : masquage PII (Presidio), détection de secrets, anti-injection (ex. Lakera), modération
 - CI/CD pour l'IA ; sécurité et conformité
 - Maîtrise des coûts à l'échelle : suivi et budgets LiteLLM
 - o Synthèse
 - Conception d'une architecture cible de synthèse



▶ LES PETITS PLUS DE NEXT

Tarification et modalités de prise en charge

- Nos formateurs sont toutes et tous consultants formateurs, nos formations sont facturées à la journée du consultant et non pas à celle du stagiaire. Notre organisme est certifié Qualiopi (pour ses actions de formation), nos formations peuvent donc être prises en charge par votre OPCO.

Les moyens pédagogiques

- Les supports pédagogiques sont transmis par voie numérique par nos soins à chacun des stagiaires. Ce support est projeté via un vidéo projecteur afin d'animer la formation. Des exercices d'application ou études de cas sont prévus afin de valider les acquis des stagiaires.

Les modalités de suivi de l'exécution

- Les formateurs se rendent à la formation avec l'ensemble des documents de fin de session. À la fin de chaque stage, la feuille d'émargement est signée pour la validation de la participation des stagiaires via notre plateforme de signature numérique. Les absences sont justifiées et notifiées sur cette même plateforme.

Les évaluations à chaud sont envoyées électroniquement puis complétées par chacun des stagiaires. Les évaluations à froid sont envoyées électroniquement 6 mois après la formation

Les horaires de formation

- Les sessions de formation débutent à 09h00 et se terminent à 17h00 pour un total de 7 heures de formation obligatoires. Pour les stages animés dans vos locaux, les horaires peuvent être aménagés selon vos disponibilités.

▶ NOS PARTICULARITÉS

Formation Intra-Entreprise

Nous proposons des formations intra-entreprises durant lesquelles nos consultants formateurs pourront adapter le contenu de la formation à votre besoin: ils se déplaceront en amont des sessions de formation afin de s'imprégner du contexte du projet et de l'adapter à votre entreprise. Vous pouvez commander une formation catalogue et effectuer cette session dans vos propres locaux. Ainsi, seul le formateur se déplace à la rencontre de vos équipes.

Formation Inter-Entreprises

Dans nos locaux ou dans ceux de nos partenaires. Ce format présente l'avantage de partager et d'échanger avec d'autres professionnels venus d'autres horizons sur les sujets couverts par le stage, ce qui facilite l'application du contenu de la formation dans le contexte de l'entreprise au quotidien.