AKADEMIA
FORMATION

PROGRAMME DE FORMATION
RÉF. N15

INTELLIGENCE ARTIFICIELLE | NIVEAU AVANCÉ

IA locale d'entreprise : *serveur, RAG & administration.*

Passer de l'IA locale sur un poste individuel à un service d'entreprise mutualisé. En deux jours, vous concevez l'architecture de votre déploiement — dimensionnement chiffré selon vos utilisateurs, vos usages et votre budget, choix de solution motivé — et vous montez un pilote complet sur l'environnement de travaux pratiques fourni : serveur d'inférence mutualisé, comptes et droits d'accès, base documentaire interrogeable (RAG) évaluée sur un jeu de questions test, plan d'administration, de supervision et de conformité (ANSSI, CNIL, RGPD).

DURÉE	FORMAT	PUBLIC	CERTIFICATION
2 jours	**Inter · Intra**	**DSI, référents IT & data**	**Attestation**
14 heures	Présentiel ou distanciel		

À PROPOS DE LA FORMATION

IA locale d'entreprise : serveur mutualisé, base documentaire (RAG) et administration.

Passer de l'IA locale sur un poste individuel à un service d'entreprise mutualisé. En deux jours, vous concevez l'architecture de votre déploiement — dimensionnement chiffré selon vos utilisateurs, vos usages et votre budget, choix de solution motivé — et vous montez un pilote complet sur l'environnement de travaux pratiques fourni : serveur d'inférence mutualisé, comptes et droits d'accès, base documentaire interrogeable (RAG) évaluée sur un jeu de questions test, plan d'administration, de supervision et de conformité (ANSSI, CNIL, RGPD).

DURÉE	FORMAT	NIVEAU	PÉDAGOGIE	CERTIFICATION
2 jours	**Inter · Intra**	**Avancé**	**Active**	**Attestation**
14 heures	Présentiel ou distanciel		Petits groupes	Attestation délivrée

Objectifs pédagogiques

- Dimensionner un serveur d'IA mutualisé à partir des usages réels (nombre d'utilisateurs, profils d'usage, modèles cibles, mémoire GPU, budget matériel) en produisant une note de dimensionnement chiffrée et argumentée, appuyée sur des jalons matériels vérifiés et sur une mesure de charge réalisée en travaux pratiques.
- Choisir une solution de service d'IA mutualisé adaptée à son contexte (moteur d'inférence conçu pour la forte concurrence, solution légère exposée au réseau interne, interface multi-utilisateurs) en comparant les options à l'état de l'art sur des critères explicites — concurrence d'accès, gouvernance et pérennité du projet, licence, compétences internes, matériel — et motiver ce choix dans le dossier d'architecture.
- Mettre en service, sur l'environnement de travaux pratiques fourni par Akademia, un service d'IA mutualisé opérationnel : moteur d'inférence servant un modèle, interface multi-utilisateurs raccordée, comptes, groupes et droits d'accès différenciés, rattachement à un annuaire d'entreprise.
- Construire une base documentaire interrogeable (RAG) : préparer et ingérer un corpus, paramétrer le découpage et la recherche, puis évaluer la qualité des réponses sur un jeu de questions test et documenter les limites constatées (couverture du corpus, citations des sources, questions hors périmètre).
- Établir le plan d'administration du service au quotidien : cycle de vie des comptes et des droits, mises à jour du serveur et des modèles, sauvegardes et test de restauration, supervision de la charge et des files d'attente — en estimant honnêtement la charge d'administration associée.
- Contrôler la sécurité et la conformité du déploiement en appliquant les recommandations publiées : hébergement de confiance et cloisonnement (guide de l'ANSSI relatif à la sécurité des systèmes d'IA générative), journalisation des accès et durées de conservation (recommandations CNIL), protection des données personnelles du corpus (RGPD) et littératie IA prévue par l'AI Act (Règlement (UE) 2024/1689, art. 4, applicable depuis le 2 février 2025 ; cf. module 4) — formalisés dans le volet conformité du dossier d'architecture et dans une charte d'usage.

Public visé

Directeurs et responsables des systèmes d'information (DSI, RSI) de PME et d'ETI, administrateurs systèmes et réseaux, référents IT, data ou IA chargés de mettre à disposition des équipes une IA interne mutualisée, responsables techniques d'organisations soumises à de fortes exigences de confidentialité (professions réglementées, industrie, secteur public local). La formation est technique et concrète — architecture, mise en service, administration — sans être une formation de développeurs : aucun développement n'est réalisé au-delà du paramétrage et de la configuration.

Prérequis

Avoir suivi la formation « IA locale et gratuite : installation, souveraineté et confidentialité des données » (N14) ou disposer d'une pratique équivalente : un modèle local déjà installé et utilisé sur son poste — l'installation individuelle n'est pas ré-enseignée ici, un rappel de 30 minutes maximum ouvre la session. Être à l'aise avec l'administration de base d'un système d'information (notions de serveur, de réseau interne et de ligne de commande) ; aucune compétence en développement n'est requise. Chaque participant vient avec son contexte de déploiement : nombre d'utilisateurs visé, cas d'usage pressentis et, s'il le souhaite, un corpus documentaire propre — préalablement autorisé et anonymisé, échantillonné en amont : 20 à 50 documents et 100 Mo au maximum (au-delà, le corpus type est utilisé en séance et la méthode se réplique ensuite sur site) ; à défaut, Akademia fournit un corpus documentaire type. Matériel : un ordinateur portable par participant et un accès internet ; l'environnement serveur de travaux pratiques est intégralement fourni par Akademia (en distanciel : caméra, micro et, idéalement, double écran).

PROGRAMME DÉTAILLÉ

Quatre modules progressifs pour monter en compétences.

JOUR 1

Jour 1 — Dimensionner et mettre en service

De l'architecture cible chiffrée au service d'IA mutualisé opérationnel sur l'environnement de travaux pratiques.

MODULE

01.

3H30

Du poste individuel au serveur d'entreprise : architectures et dimensionnement

> **OBJECTIF OPÉRATIONNEL**
>
> *« Établir la note de dimensionnement chiffrée et argumentée de son déploiement (utilisateurs, usages, modèles cibles, mémoire GPU, budget matériel), appuyée sur des jalons matériels vérifiés et sur une mesure de charge réalisée en travaux pratiques. »*

CONTENU PÉDAGOGIQUE

- Rappel express des acquis du niveau poste de travail (30 minutes maximum — prérequis N14 ou pratique équivalente) — et pourquoi le poste individuel ne suffit plus dès que l'équipe doit partager modèles, base documentaire, comptes et historique.
- Architectures de mutualisation : ce que change l'accès simultané de plusieurs utilisateurs sur un serveur interne — et ce qu'il faut en déduire pour la file d'attente et la latence perçue par les équipes.
- Jalons matériels vérifiés : quels ordres de grandeur de modèles s'exécutent sur quelles classes de matériel — présentés et réactualisés en session, à croiser avec ses usages, jamais des promesses de capacité.
- La méthode de dimensionnement en cinq étapes, des profils d'usage au budget matériel en fourchette indicative, variable selon les configurations et les prix constatés.
- Honnêteté d'arbitrage : ce qu'un serveur interne apporte et ce qu'il ne promet pas — les capacités réelles d'une configuration se mesurent sur son propre couple modèle / usage.
- Règle d'usage de l'environnement de travaux pratiques, posée dès l'ouverture : aucune donnée réelle non autorisée n'est chargée sur les instances de TP.

MISE EN PRATIQUE

Atelier « Dimensionnement » : mesure de charge guidée sur l'environnement de TP (plusieurs requêtes simultanées : observation de la file d'attente et de la latence), puis chaque participant établit, sur la trame fournie, la note de dimensionnement chiffrée de SON déploiement — utilisateurs, profils d'usage, modèle(s) cible(s), mémoire GPU, budget en fourchette ; restitution flash de deux dimensionnements challengés par le groupe.

LIVRABLE

Note de dimensionnement chiffrée et argumentée de son déploiement — première pièce du dossier d'architecture.

MODULE

02.

3H30

Choisir sa solution serveur et mettre en service le service mutualisé

OBJECTIF OPÉRATIONNEL

« Comparer les solutions de mutualisation à l'état de l'art sur des critères explicites, motiver un choix adapté à son contexte, puis mettre en service sur son instance de travaux pratiques un service complet : moteur d'inférence, interface multi-utilisateurs, comptes, groupes et droits. »

CONTENU PÉDAGOGIQUE

- Panorama vérifié des solutions serveur (posture multi-éditeurs, veille tenue à jour par le formateur) : moteurs d'inférence conçus pour la mutualisation, solutions légères exposées au réseau interne, interfaces multi-utilisateurs.
- Les critères de choix d'une DSI : concurrence d'accès, gouvernance et pérennité du projet, licence, compétences internes, matériel existant — grille de décision remplie sur son propre contexte.
- Lire la licence de chaque brique avant de déployer : toutes les briques « ouvertes » ne se valent pas juridiquement — cas concret lu en séance.
- Mise en service pas à pas sur l'environnement de travaux pratiques fourni : moteur d'inférence, modèle servi, interface raccordée, premiers tests à plusieurs comptes simultanés.
- Comptes, groupes et droits : rôles différenciés et rattachement à l'annuaire d'entreprise.
- Exposition et cloisonnement : service limité au réseau interne par défaut, séparation des environnements — conformément aux recommandations publiées (ANSSI), approfondies au module 4.

MISE EN PRATIQUE

Atelier « Mise en service » : chaque participant (ou binôme) met en service son instance de TP — moteur d'inférence démarré, modèle servi, interface raccordée — puis crée trois comptes aux rôles différenciés, vérifie les droits effectifs de chacun et raccorde l'annuaire de démonstration ; le choix de solution est motivé, critère par critère, et consigné dans le dossier d'architecture (gabarit guidé).

LIVRABLE

Service d'IA mutualisé opérationnel sur l'instance de TP (modèle servi, comptes et droits vérifiés) et choix de solution motivé, consigné dans le dossier d'architecture.

JOUR 2

Jour 2 — Base documentaire et administration

Du pilote RAG monté et évalué au plan d'administration, de sécurité et de conformité.

MODULE

03.

3H30

Monter la base documentaire interrogeable (RAG) et en évaluer la qualité

OBJECTIF OPÉRATIONNEL

« Monter un pilote RAG complet sur son instance de travaux pratiques — préparation du corpus, ingestion, découpage, recherche — puis en évaluer la qualité sur un jeu de questions test et en documenter honnêtement les limites. »

CONTENU PÉDAGOGIQUE

- Le principe de la génération augmentée par la récupération (RAG) : répondre à partir des documents de la base en citant ses sources — ce qu'un RAG sait faire et ne sait pas faire.
- La chaîne d'ingestion, de la collecte des documents à l'indexation : les réglages qui pèsent réellement sur la qualité des réponses.
- Recherche et restitution : les modes de recherche, le dosage des passages restitués au modèle et les citations des sources dans la réponse.
- Les briques de base documentaire selon la volumétrie — critère directeur : capitaliser sur ce que la DSI sait déjà exploiter.
- Qualité et limites, dites honnêtement : la réponse vaut ce que vaut le corpus ; méthode d'évaluation par jeu de questions test et droits d'accès aux collections alignés sur les habilitations internes.

MISE EN PRATIQUE

Atelier « Pilote RAG » : chaque participant ingère un corpus dans son instance de TP (corpus documentaire type fourni par Akademia ou son propre corpus, préalablement autorisé et anonymisé), règle le découpage et la recherche, interroge la base, puis évalue la qualité des réponses sur un jeu de dix questions test — dont des questions hors corpus — et renseigne sa fiche qualité (réussites, échecs, limites) ; restitution « ce que mon RAG sait et ne sait pas répondre ».

LIVRABLE

Pilote RAG monté et évalué sur l'instance de TP, avec fiche qualité du corpus (jeu de questions test, résultats, limites documentées).

MODULE

04.

3H30

Administrer, sécuriser et mettre en conformité le service au quotidien

OBJECTIF OPÉRATIONNEL

« Formaliser le plan d'administration et le volet sécurité / conformité de son déploiement — comptes, mises à jour, sauvegardes, supervision, journalisation, RGPD — en réalisant les gestes d'administration clés sur son instance de TP, puis consolider et défendre le dossier d'architecture final. »

CONTENU PÉDAGOGIQUE

- L'administration au quotidien, sans angélisme : cycle de vie des comptes, mises à jour, sauvegardes et test de restauration, supervision — et l'estimation honnête de la charge d'administration récurrente.
- Sécuriser le système d'IA interne selon les recommandations publiées (guide ANSSI) : hébergement de confiance, cloisonnement, maîtrise de l'exposition réseau — déclinés sur l'architecture de chaque participant.
- Journaliser sans sur-conserver : le bon niveau de granularité et les durées de conservation recommandées (CNIL) — traçabilité des accès sans conservation excessive du contenu des requêtes.
- Usage responsable et cadre réglementaire : secret des affaires, protection des données personnelles du corpus et des journaux (RGPD, recommandations CNIL), littératie IA prévue par l'AI Act (art. 4, applicable depuis le 2 février 2025) — la formation y contribue sans constituer une prestation de mise en conformité ; le formateur tient cette veille à jour.
- La charte d'usage du service IA interne : ce que les équipes peuvent y charger et y demander, statut des réponses, circuit de demande d'accès et de signalement — le document qui conditionne l'adoption autant que la technique.
- Après le pilote : passage au service régulier et plan d'action des 90 premiers jours.

MISE EN PRATIQUE

Atelier final « Dossier d'architecture » : chaque participant réalise les gestes d'administration sur son instance de TP (création puis désactivation d'un compte, sauvegarde et restauration test, lecture des journaux et de la supervision), consolide son dossier d'architecture — note de dimensionnement, choix de solution motivé, plan d'administration, volet sécurité et conformité, charte d'usage amorcée — puis le défend en soutenance flash de trois minutes devant le groupe ; évaluation sommative sur grille critériée et plan d'action des 90 premiers jours.

LIVRABLE

Dossier d'architecture complet et défendu (dimensionnement chiffré, choix de solution motivé, plan d'administration, volet sécurité / conformité, charte d'usage) et plan d'action des 90 premiers jours.

MÉTHODES PÉDAGOGIQUES

Apprendre par la pratique, avec un formateur expert à vos côtés.

- Pédagogie active et apprentissage par le faire : la pratique occupe la place centrale — de l'ordre de 45 % du temps en atelier individuel (ou en binôme sur les instances de travaux pratiques) accompagné sur le déploiement cible du participant (fil rouge) et plus de 60 % du temps consacré à la pratique au sens large en y ajoutant les démonstrations commentées, la mesure de charge et les restitutions appliquées à ce même déploiement ; le reste en apports méthodologiques cadrés.
- Méthode magistrale : apports structurés et cadrage méthodologique par le formateur, appuyés sur des supports visuels et des schémas d'architecture.
- Démonstrations en direct : le formateur manipule le serveur de démonstration devant le groupe (mise en service bout en bout, ingestion du corpus, lecture des journaux) ; la mesure de charge est, elle, réalisée par les participants en travaux pratiques collectifs guidés.
- Méthode active : ateliers individuels accompagnés sur instance dédiée, restitutions flash et soutenance finale du dossier d'architecture favorisant le regard critique.
- Accompagnement individualisé : le formateur adapte le niveau de soutien selon le profil (administrateur système aguerri ou référent data moins technique), sur la base du test de positionnement.
- Approche par compétences : chaque module produit un livrable directement réinvestissable dans le déploiement réel.

Profil du formateur

Formateur expert à double compétence : infrastructure et administration de systèmes d'IA locale (serveurs d'inférence GPU, déploiement en interne, supervision) et mise en œuvre de bases documentaires RAG en entreprise. Il justifie de déploiements mutualisés concrets en organisation et d'une pratique des référentiels applicables (recommandations ANSSI et CNIL).

Moyens & supports

- En présentiel : salle équipée d'un vidéo-projecteur, paperboard, connexion internet et un poste par participant.
- En distanciel : classe virtuelle synchrone via les outils Akademia (partage d'écran, sous-groupes, partage de fichiers).
- Environnement de travaux pratiques fourni par Akademia : serveur GPU de démonstration et instances dédiées (une par participant ou par binôme), accessibles pendant toute la formation pour les ateliers de mise en service, de base documentaire et d'administration — aucun matériel serveur n'est requis côté participant ; annuaire de démonstration (LDAP) et corpus documentaire type anonymisé inclus.
- Plateforme LMS Akademia (FormAI) : test de positionnement en ligne et mise à disposition de l'ensemble des ressources (supports, gabarits, pas-à-pas de mise en service, corpus type, jeu de questions test).
- Kit de gabarits remis à chaque participant : trame de note de dimensionnement, grille de décision, gabarit de dossier d'architecture, plan d'exploitation (checklist d'administration), fiche qualité corpus, modèle de charte d'usage, trame de plan d'action des 90 premiers jours.
- Outillage IA locale (catégories de capacités, posture multi-éditeurs) : moteurs d'inférence mutualisés à API standard de place, interfaces multi-utilisateurs avec comptes, groupes, droits et rattachement annuaire, briques de base documentaire (base intégrée, extension vectorielle d'une base existante ou base vectorielle dédiée) et outils de supervision — solutions vérifiées à l'état de l'art à la date de conception (juillet 2026), veille tenue à jour par le formateur.

Modalités d'évaluation

- Test de positionnement en ligne réalisé sur la plateforme LMS avant le début de la formation (pratique de l'IA locale au poste, administration système, contexte de déploiement), complété par un tour de table des attentes.
- Évaluation formative continue : les livrables de chaque module (note de dimensionnement, mise en service et droits vérifiés, pilote RAG et fiche qualité, plan d'administration) et leurs restitutions permettent au formateur de vérifier la progression sur chaque objectif et d'apporter une remédiation immédiate.
- Évaluation sommative : grille critériée d'atteinte des objectifs, renseignée au fil des ateliers puis consolidée et restituée en fin de session, appliquée aux productions réalisées sur le déploiement fil rouge (note de dimensionnement chiffrée ; choix de solution motivé ; service mutualisé mis en service sur l'instance de TP avec comptes, droits vérifiés et annuaire de démonstration raccordé ; pilote RAG évalué et fiche qualité du corpus ; plan d'administration et gestes d'administration réalisés ; volet sécurité et conformité du dossier d'architecture et charte d'usage), intégrant un critère d'usage responsable observable (aucune donnée réelle non autorisée chargée sur l'environnement de TP, corpus autorisé et anonymisé), complétée par un auto-positionnement de sortie reprenant les items du test de positionnement amont pour objectiver la progression.
- Évaluation de satisfaction à chaud en fin de session et évaluation à froid à distance, à 1 à 3 mois, mesurant le transfert en situation de travail : état d'avancement du déploiement (décision d'architecture, matériel, pilote interne, charte diffusée) et premiers usages constatés par les équipes, à titre indicatif et non garanti.

Documentation remise aux stagiaires

- Le support de formation complet
- La trame de note de dimensionnement chiffrée
- La grille de décision « choix de solution »
- Le gabarit de dossier d'architecture (dimensionnement, choix motivé, administration, sécurité et conformité)
- Le pas-à-pas de mise en service (moteur d'inférence, interface multi-utilisateurs, comptes et droits)
- Le corpus documentaire type anonymisé et le jeu de questions test
- La fiche qualité corpus (évaluation du RAG et limites documentées)
- Le plan d'exploitation : checklist d'administration (comptes, mises à jour, sauvegardes et restauration, supervision)
- L'aide-mémoire journalisation et conformité (recommandations ANSSI et CNIL, RGPD)
- Le modèle de charte d'usage du service IA interne et la trame de plan d'action des 90 premiers jours
- Attestation de fin de formation mentionnant les objectifs et le résultat de l'évaluation des acquis

Équipements à apporter

- Ordinateur portable

Accessibilité & handicap

Les besoins d'adaptation sont recensés dès l'inscription. Un référent handicap Akademia est identifié et joignable pour étudier, au cas par cas avec le participant, les aménagements possibles (rythme, supports, modalité). Les conditions d'accès sont vérifiées selon la situation.

Modalités & délais d'accès

Formation en petit groupe (4 à 6 participants), chaque participant ou binôme disposant de sa propre instance serveur sur l'environnement de travaux pratiques. Inscription en ligne ou auprès du service formation, entrée à date fixe selon le calendrier des sessions. Pour les financements OPCO, l'inscription doit intervenir suffisamment tôt pour respecter les délais d'instruction du

dossier ; Akademia accompagne le participant dans ses démarches.

Tarif

SESSION INTER-ENTREPRISES

1 990 € net de taxe

par participant · 2 jours (14 h)

Exonération de TVA · art. 261-4-4° a du CGI

SESSION INTRA-ENTREPRISE

Tarif sur devis

Session dédiée à vos collaborateurs, dans vos locaux ou à distance. Contactez-nous pour une proposition chiffrée personnalisée selon l'effectif et les modalités.

Prise en charge possible par votre OPCO ou France Travail. Nos équipes vous accompagnent dans le montage du dossier de financement.

PASSONS À L'ACTION

Construisons ensemble votre session sur-mesure.

Dites-nous vos contraintes (format, lieu, dates, nombre de participants) et recevez une proposition personnalisée sous 24 heures ouvrées.

Akademia Formation

SERVICE ADMINISTRATION DES VENTES

adv@akademiaformation.com

www.akademiaformation.com

Devis personnalisé

RÉPONSE SOUS 24 H OUVRÉES

Format inter · intra

Présentiel ou distanciel

FIN DU PROGRAMME