

Appel d'offres – Développeur IA – Contenus & LLMOps (industrialisation) (régie / forfait possible)

Réf. AO-V9-010 • Version publication • Date : 22/01/2026

Avis de consultation

La consultation est structurée par profil afin de permettre une sélection fine ; la CPME pourra retenir un ou plusieurs prestataires. Les modalités de remise des offres sont rappelées ci-dessous.

- Contact de consultation : llageois@cpme.fr.
- Date limite de remise des offres : 28 février 2026 à 17h00 (heure de Paris).
- Questions / réponses : selon modalités précisées lors de la publication.
- Format de réponse : PDF ou DOCX, en français.

Modalité privilégiée / alternative	Régie privilégiée. Une réponse en forfait est recevable si elle précise la solution, l'équipe mobilisée, le plan de mise en œuvre, les hypothèses et les engagements (qualité, sécurité, performance, réversibilité).
Lieu d'intervention	Paris / La Défense / Hybride. Des déplacements ponctuels peuvent être requis, notamment vers des antennes locales.
Démarrage souhaité	Dès que possible.
Durée indicative	3 à 6 mois renouvelables selon la phase, avec perspective long terme (build + run).
Possibilité d'embauche	La CPME souhaite conserver la possibilité de transformer certaines prestations en embauche ; les modalités devront être détaillées (conditions, processus, calendrier, clauses éventuelles et transition).

Contexte

La CPME nationale engage la mise en œuvre d'un Système d'Information unifié, articulé autour d'un socle CRM/ERP, d'un CMS léger et d'un Hub IA. La solution s'adresse à la structure nationale ainsi qu'à ses antennes locales. L'ambition est de bâtir une plateforme souveraine, sécurisée et industrialisée, capable de soutenir durablement l'exploitation (MCO), les corrections et l'évolution continue du produit.

Socle cible

Afin de cadrer la réponse et d'éviter la dispersion, le socle cible est volontairement formulé de manière concise. Les propositions devront s'y conformer ou expliciter clairement les écarts, leur justification et leurs conséquences (coûts, risques, exploitation).

- Stack cible : Symfony 7 + API Platform + PostgreSQL + Doctrine + Messenger.
- Front : Angular (prioritaire) ; CI/CD et Docker constituent le socle commun.
- Organisation : structure nationale et antennes locales.

Objet et mission

Le développeur IA « contenus & LLMOps » industrialise le Hub IA : gestion des prompts et templates, qualité, coûts, déploiement, monitoring et exploitation. Le rôle est distinct de l'orchestration/agents : il se concentre sur l'opérabilité, la maîtrise des coûts et la qualité mesurable, avec une capacité à déployer, monitorer et améliorer en continu. Les certifications sont appréciées (« nice to have ») mais ne constituent pas un prérequis.

Exigences transverses (qualité et sécurité)

Quel que soit le profil, la CPME attend une contribution active à la qualité du produit et à une sécurité applicative pragmatique. Ces exigences conditionnent la pérennisation (build + run) et s'inscrivent dans un cadre commun.

- Qualité intégrée : une Definition of Done commune s'applique à tous (non-régression, quality gates, documentation utile). La stratégie de tests et l'orchestration de la qualité sont pilotées par le profil QA/Test ; chaque contributeur participe selon son périmètre.
- Pratiques d'équipe : PR de taille maîtrisée, revues systématiques, conventions partagées, et usage de feature flags lorsque nécessaire.
- Sécurité applicative pragmatique : gestion rigoureuse des secrets, validations côté serveur et côté client, journaux et traces d'audit ; chiffrage lorsque cela s'avère nécessaire au regard des risques.

Exigences spécifiques au rôle

Les éléments ci-dessous précisent les attendus distinctifs de la mission, au-delà du socle commun.

- Gestion des prompts : versioning, templates, rollback et A/B testing lorsque pertinent.
- Optimisation des coûts : caching, routing de modèles, quotas et mesures de consommation par profil.
- Monitoring : latence, erreurs, coûts, dérive ; alerting et runbooks d'exploitation.

Livrables et critères de succès

Les livrables attendus doivent permettre une exécution fluide, une exploitation fiable et une réversibilité maîtrisée. Ils seront précisés lors du cadrage de démarrage, sans diminuer l'exigence de documentation utile.

- Un dispositif de gestion des prompts industrialisé (versioning, rollback, templates) et documenté.
- Des mécanismes de maîtrise des coûts (caching, routing, quotas), pilotés par des indicateurs.
- Un monitoring complet (qualité, latence, erreurs, coûts, dérive) et des runbooks opérationnels.
- Des quality gates IA : golden sets, tests automatiques, seuils d'acceptation et suivi dans le temps.

Compétences et expériences attendues

La CPME recherche des profils capables de produire, d'explicitier leurs choix, de documenter et de transmettre, dans un contexte multi-acteurs et multi-fournisseurs.

- Compétences LLMOps : évaluation continue, monitoring et gestion des déploiements.
- Culture produit : qualité mesurable et réduction des sorties non fiables ; règles de citation appliquées.
- Capacité à intégrer l'IA dans un CI/CD et à produire une documentation utile.
- Sécurité : journaux/audit, politiques, gestion des données sensibles.

Modalités d'évaluation

Le dispositif de sélection est homogène afin de comparer efficacement des candidatures issues de fournisseurs différents. Les modalités exactes pourront être adaptées selon les contraintes calendaires, sans modifier les principes ci-dessus.

- Un entretien unique, couvrant à la fois les dimensions techniques, méthodologiques et de collaboration (incluant motivation, communication et projection dans la durée).
- Au besoin : une étude de cas (pour certains rôles) et/ou un exercice pratique sur repo starter pour des rôles techniques (MR/PR et restitution). Indication : décrire un dispositif prompts versionnés + métriques qualité/coûts + monitoring et rollback opérationnel.

Réponse attendue

Afin de faciliter l'analyse et la comparaison des offres, merci de structurer votre réponse autour des éléments suivants. En cas de réponse en forfait, les compléments doivent être suffisamment précis pour permettre un chiffrage et une exécution sans ambiguïté.

- Un CV détaillé (ou une ressource nominative) précisant le rôle exact, la séniorité, ainsi que des références comparables.
- Les modalités d'intervention : disponibilité, présence (hybride), capacité de déplacement, et organisation proposée.
- Les conditions financières : TJM, conditions de facturation, modalités de remplacement ; ou prix forfaitaire et conditions associées.
- Une approche synthétique (1–2 pages) : organisation, principaux risques, transfert de compétences, contribution au run (MCO).

- Les modalités de transformation en embauche : conditions, processus, calendrier, clauses éventuelles, et modalités de transition (si applicable).
- Si réponse en forfait : solution technique, macro-planning, staffing, livrables, hypothèses, engagements (qualité/sécurité/performance), et modalités de réversibilité.

Confidentialité et propriété

Les informations transmises par la CPME dans le cadre de la consultation sont confidentielles. Les livrables et développements réalisés dans le cadre de la mission doivent être pérennisables : documentation, transfert de compétences et réversibilité sont attendus.

Annexes

A. Cadre organisationnel et gouvernance

- Principe directeur : un seul propriétaire de décision par sujet afin d'éviter les renvois de responsabilité.
- PO : porte la vision, arbitre, décide et assure la validation finale (sign-off).
- PPO : opère le backlog au quotidien (US/AC/DoR), prépare les sprints et la recette ; escalade au PO les sujets structurants (périmètre, délais, règles majeures).
- Delivery Manager : sécurise la livraison (planification, dépendances, risques), coordonne les acteurs et le multi-fournisseurs, et pilote le suivi sécurité & conformité (cadre, preuves, arbitrages, escalades).
- UX/UI : conçoit les parcours et l'interface, formalise les standards d'ergonomie et d'accessibilité, et organise le handoff avec l'équipe de réalisation.
- QA/Test : définit la stratégie de test, met en place l'automatisation et les quality gates, et anime l'amélioration continue de la qualité.
- Lead Tech CRM/CMS : garantit les standards, l'architecture et les quality gates. Il porte également le rôle de Scrum Master côté développement (ou le délègue à un développeur senior désigné).
- Lead Tech DevOps : définit et opère la plateforme (CI/CD, Docker, orchestrateur), les environnements et les exigences PRA/PCA.
- Équipe de réalisation : développeurs back/front/fullstack, data/ETL, IA, selon les besoins.

B. Rituels et instances

- Comité stratégique (mensuel) : suivi budgétaire, risques majeurs, décisions structurantes.
- Comité produit (hebdomadaire) : priorisation, périmètre, arbitrages (décision portée par le PO).
- Comité delivery, sécurité & conformité (hebdomadaire) : pilotage des dépendances, suivi des risques, état des preuves et sujets de conformité (pilotage Delivery Manager).
- Comité technique & qualité (hebdomadaire) : architecture, dette, quality gates, incidents et améliorations transverses (Lead Tech + QA).
- Rituels d'équipe : daily, planning, review, rétrospective (animation Scrum Master côté dev).

C. Standards communs (qualité, sécurité, pérennisation)

- Definition of Done : non-régression, quality gates, documentation minimale, instrumentation (logs/metrics) et revue de code.
- Mécanique de livraison : PR petites et fréquentes, revues obligatoires, conventions, feature flags si besoin.
- Sécurité : secrets, validations, audit/logs, chiffrement si nécessaire ; suivi des preuves dans le cadre « delivery, sécurité & conformité ».
- Pérennisation : runbooks, transfert de compétences et réversibilité intégrés dès la phase build.