

Ingénieur junior - Déploiement et exploitation d'un cluster IA hétérogène (H/F)

Localisation

ENSEIRB-MATMECA
Avenue des Facultés
33400 Talence

Employeur

Bordeaux INP
Avenue des Facultés
33 400 Talence
+33 5 56 84 61 00
www.bordeaux-inp.fr

Type de recrutement

CDD de 12 mois (non renouvelable)

Composante : ENSEIRB-MATMECA

Date de vacance d'emploi : 1^{er} octobre 2026

Date limite de candidature : 4 septembre 2026

Date de prise de poste : 1^{er} octobre 2026

Cadre emploi

E2C45 - Ingénieur / Ingénieure
en ingénierie logicielle
(Catégorie A)

Supérieur hiérarchique

Denis Barthou

Télétravail possible

Non

Rémunération

Environ 2500€ brut mensuel (30K€ brut annuel)

Vos missions en quelques mots

L'ENSEIRB-MATMECA est l'une des 6 écoles d'ingénieurs de Bordeaux INP. Elle compte près de 1200 étudiants, 90 personnels enseignants et enseignants-chercheurs, 25 personnels administratifs et techniques. Pour assurer ses missions, l'école s'appuie sur plus de 190 personnels (enseignants-chercheurs, enseignants, personnels administratifs et techniques) et près de 400 vacataires issus du monde académique et industriel.

La mission proposée consiste à installer et administrer un cluster IA très hétérogène, dans le cadre du projet ANR CAP-IA. L'ingénieur / l'ingénieure l'installera, le rendra accessible à différents publics (étudiants et étudiantes, enseignants et enseignantes, stagiaires de formation continue principalement). L'objectif est de proposer une expérience proche d'un cloud IA privé, dans un cadre académique maîtrisé, orienté formation, expérimentation et reproductibilité.

Activités

Dans le cadre du projet ANR CAP-IA, l'ENSEIRB-MATMECA déploie un cluster hétérogène dédié à l'intelligence artificielle. Le cluster intégrera différents types de machines, d'accélérateurs et de configurations logicielles, depuis des équipements orientés IA embarquée ou frugale jusqu'à des serveurs plus puissants pour l'inférence de grands modèles. Il sera également instrumenté pour mesurer la consommation énergétique en conditions réelles, afin d'étudier les compromis entre performance, énergie et usages de l'IA. Cette infrastructure vise à offrir un environnement expérimental, pédagogique et scientifique autour des infrastructures IA modernes.

Vous participerez au déploiement, à l'administration et à l'évaluation du cluster IA afin de le rendre utilisable par différents publics.

Dans le cadre de ce projet, vos missions principales seront les suivantes :

- Installer, configurer et administrer les machines du cluster ;
- Mettre en place les systèmes, pilotes, bibliothèques et outils nécessaires à l'utilisation des différents accélérateurs ;
- Déployer des environnements logiciels adaptés aux usages IA, notamment via Docker et éventuellement Kubernetes ou une solution équivalente ;
- Gérer les comptes, droits d'accès, ressources utilisateurs, stockage, jeux de données, modèles et résultats ;
- Mettre en place une interface d'accès simple, permettant de lancer des environnements IA à la demande, dans un esprit "cloud-like" ;
- Permettre l'exécution de notebooks, conteneurs d'inférence, tests de modèles et services IA temporaires ;
- Contribuer à une suite de benchmarks reproductibles pour mesurer les performances et la consommation énergétique des différentes configurations ;
- Suivre des métriques telles que débit d'inférence, latence, temps de chargement, occupation mémoire, consommation et efficacité énergétique ;
- Documenter l'installation, les procédures d'utilisation et les cas d'usages pédagogiques ;
- Accompagner les utilisateurs dans la prise en main de la plateforme.

La mission pourra également inclure un volet plus avancé d'optimisation de l'inférence : comparaison de moteurs d'inférence, utilisation d'outils tels que PyTorch, Hugging Face, vLLM, ONNX Runtime ou TensorRT, étude de la quantification, du batching, du streaming ou du parallélisme, et analyse des compromis entre performance, précision, mémoire, latence et énergie.

Profil recherché

Le poste s'adresse à une personne de profil junior et motivée par les infrastructures IA, les systèmes, le cloud et les questions de performance. Une expérience très approfondie n'est pas indispensable, mais une bonne capacité d'apprentissage, de l'autonomie et un intérêt marqué pour les technologies IA modernes sont attendus.

Compétences souhaitées :

- Bonne connaissance de Linux et de l'administration système, notions en réseau, stockage et environnements multi-utilisateurs ;
- Intérêt ou expérience avec Docker et les environnements conteneurisés ;
- Connaissances en programmation Python, notions de programmation GPU ;
- Intérêt pour l'IA, l'inférence de modèles et les architectures matérielles.

Savoir-faire :

- Rédiger et mettre à jour une documentation fonctionnelle et technique ;

- Evaluer les solutions logicielles pour l'IA, par l'expérimentation et des benchmarks et les analyser ;
- Capacité à documenter clairement une installation ou une procédure ;
- Goût pour l'expérimentation, la mesure de performance et l'analyse technique.

Savoir être :

- Capacité de proposition de nouvelles solutions/adaptations ;
- Capacité de raisonnement analytique pour l'analyse des résultats d'évaluation ;
- Adaptation technique à la diversité des matériels.

Travailler à Bordeaux INP, les +

- Entre 47 et 52 jours de congés annuels
- Participation employeur à la complémentaire santé
- Prise en charge à 75% de l'abonnement aux transports en commun
- Forfait « mobilités durables » (vélo, covoiturage) pour le trajet domicile/travail
- Accès au parking du personnel

Éléments de candidatures

Documents à transmettre

- Lettre de motivation et CV détaillé

Les documents sont à déposer **avant le 4 septembre 2026** sur la plateforme :

<https://gestionrh.bordeaux-inp.fr/recrutement-contractuel/>.

Pour information, les candidatures ne seront pas traitées entre le 24 juillet et le 24 août 2026.

Contacts

Contact RH : Sabrina Corpas – sabrina.corpas@bordeaux-inp.fr

Contact Métier : Denis Barthou - denis.barthou@enseirb-matmeca.fr

Dans le cadre de sa démarche en faveur de l'égalité professionnelle et de la diversité, Bordeaux INP s'engage contre toute forme de discrimination et encourage tous les candidats et toutes les candidates éligibles à transmettre leur candidature.

Bordeaux INP, en quelques chiffres

2600

étudiantes et étudiants

254

enseignants et
enseignants-chercheurs

Près de **200**

personnels administratifs

Qui sommes-nous ?

Bordeaux INP est un Établissement Public à Caractère Scientifique, Culturel et Professionnel (EPSCP), constitué sous la forme d'un grand établissement. Fédérateur des écoles d'ingénieurs en Nouvelle-Aquitaine, il propose une offre de formation scientifique et technique de haut niveau adossée à une recherche d'excellence et à une forte capacité à produire et à transférer l'innovation. Bordeaux INP regroupe 6 écoles d'ingénieurs publiques, une classe préparatoire intégrée « La Prépa des INP » et un incubateur étudiant « INPulse ».