

Type d'offre : Offre en entreprise

Date de publication : 03.03.26

Formel AI

AI Engineer

Informations générales

Type de contrat : CDI

Durée du contrat : CDI

Niveau d'expérience : 3+ years

Contact :

careers@formel.ai

Date de prise de poste : jeu 01/10/2026 - 12:00

Formel AI :

Chez **Formel AI**, nous combinons la créativité de l'IA générative avec la fiabilité des **méthodes formelles mathématiques**. En nous appuyant sur les capacités de vérification des méthodes formelles, telles que **Lean 4**, nous développons une nouvelle génération de **grands modèles de langage (LLM)** à la fois fiables, transparents et économes en ressources. Nous avons eu l'honneur de remporter le **2^e prix de l'Innovation de l'École normale supérieure** en novembre 2025.

Formel AI a été fondée par **Sylvain Combettes** (CEO) et **Antoine Mazarguil** (CTO), qui se sont rencontrés pendant leur doctorat à l'ENS Paris-Saclay. En février

2026, nous avons bouclé un **tour de financement auprès de business angels sursouscrit**, soutenu par plus de 30 investisseurs de premier plan, parmi lesquels les fondateurs de Datadog, Dataiku, Nabla et Artefact. Notre conseil scientifique compte également des professeurs issus de l'École polytechnique, de l'ENS Paris-Saclay et de l'École des Ponts.

Nous accueillons désormais nos premiers **partenaires pilotes (design partners)** et constituons l'équipe qui nous permettra de concrétiser cette ambition.

Rejoignez-nous !

Adresse :

Formel AI
75000 Paris
France

Détail de l'offre (poste, mission, profil) :

À propos du poste

En tant qu'**AI Engineer**, vous serez au cœur de l'innovation en intelligence artificielle et travaillerez sur une technologie de pointe combinant les grands modèles de langage (LLM) et les méthodes de vérification formelle. Vous rejoindrez une équipe fondatrice composée d'experts de la deep tech, qui s'attaque à l'un des défis majeurs de l'IA : développer des LLM transparents, fiables et économiquement efficaces.

Il s'agit d'un poste mêlant **recherche appliquée et ingénierie**, avec une forte dimension opérationnelle. Vous concevrez et développerez la prochaine génération de systèmes d'IA formellement vérifiés. Vous travaillerez sur de nouvelles architectures, des protocoles d'orchestration et des méthodes d'entraînement visant à éliminer les hallucinations tout en réduisant considérablement les coûts de calcul.

Vous rejoindrez Formel AI à une étape charnière de son développement : le passage des avancées de recherche au déploiement en conditions réelles, aux côtés de partenaires pilotes dans des domaines critiques tels que la santé et les logiciels d'entreprise, où la fiabilité de l'IA est indispensable.

Missions principales

- **Concevoir et développer des systèmes d'IA** : piloter l'architecture et l'implémentation de LLM intégrant des méthodes de vérification formelle.
 - **Développer des protocoles d'orchestration** : concevoir et mettre en œuvre des frameworks d'orchestration incluant l'appel d'outils (*tool calling*), la génération structurée et des workflows de vérification.
 - **Contribuer à l'innovation scientifique** : participer activement à la définition de la feuille de route de recherche de l'entreprise, en explorant de nouvelles architectures, méthodes d'entraînement et techniques de vérification, en collaboration avec le conseil scientifique.
 - **Déployer les solutions auprès des partenaires pilotes** : travailler étroitement avec 2 à 3 premiers clients dans des secteurs critiques afin de mettre en œuvre des preuves de concept démontrant des améliorations mesurables en matière de fiabilité et de maîtrise des coûts.
 - **Optimiser l'entraînement et l'évaluation** : développer des pipelines avancés de génération de jeux de données, des processus de fine-tuning (DPO, RLHF) et des protocoles d'évaluation rigoureux afin d'améliorer continuellement les performances des modèles.
 - **Construire une infrastructure réutilisable** : créer de la documentation, des modèles de code et des bonnes pratiques d'ingénierie permettant à l'équipe de déployer ses solutions à plus grande échelle au fil de sa croissance.
 - **Faire le lien entre théorie et pratique** : transformer les avancées de la recherche en IA en systèmes robustes, prêts à être déployés en production.
-

Profil recherché

Qualités essentielles

- **Curiosité orientée recherche** : vous aimez explorer de nouvelles applications et architectures d'IA et repousser les limites de ce qui est possible.
- **Excellence technique et approche opérationnelle** : vous êtes capable de concevoir, développer et déployer de manière autonome des systèmes d'IA prêts pour la production, de l'entraînement des modèles à leur orchestration.

- **Expertise technique approfondie** : vous maîtrisez les architectures de LLM, les méthodes d'entraînement et l'optimisation de l'inférence.
- **Capacité à faire le lien entre recherche et application** : vous savez transformer des concepts de recherche complexes en implémentations concrètes.

Expérience

- Plus de **3 ans d'expérience pratique** dans la conception et le déploiement de systèmes d'IA/ML, avec une solide expertise en Python, API de LLM et infrastructures cloud.
- Solides connaissances des fondamentaux du machine learning, des architectures de modèles et des pratiques modernes d'ingénierie en IA.
- Excellente connaissance des dernières avancées de la recherche en IA et des systèmes déployés en production.

Compétences requises

- **Maîtrise de Python** : vous écrivez un code Python propre et efficace et maîtrisez les principes de la programmation orientée objet.
- **Bases solides en développement full-stack** : vous comprenez les principes du développement frontend et backend, êtes capable d'intervenir sur l'ensemble de la stack si nécessaire et maîtrisez Python pour les systèmes backend.
- **Expertise LLM** : vous avez une expérience concrète des grands modèles de langage, notamment en intégration d'API, *prompt engineering* et frameworks d'orchestration.
- **Outils d'inférence LLM** : vous avez une expérience pratique d'au moins un framework majeur d'inférence LLM (vLLM, llama.cpp, TensorRT-LLM ou équivalent) et maîtrisez les techniques d'optimisation de l'inférence.
- **Entraînement et fine-tuning** : vous avez déjà mis en œuvre des workflows de fine-tuning (DPO, RLHF ou équivalent) et maîtrisez la génération de jeux de données et l'évaluation des modèles.
- **Architectures de modèles** : vous avez une connaissance approfondie des architectures Transformer et des techniques modernes d'optimisation de l'inférence.

- **Expérience en production** : vous avez déjà déployé des systèmes d'IA dans des environnements de production et maîtrisez l'ensemble du cycle de vie ML, de l'entraînement au déploiement.
 - **Culture de la recherche** : vous suivez les dernières avancées de la recherche en IA et êtes capable de prototyper et de valider rapidement de nouvelles idées.
-

Avantages

Rémunération et avantages

- **Salaire fixe compétitif** et intéressement au capital (**BSPCE**) d'une entreprise à fort potentiel.
- **Mutuelle Alan** pour vous et votre famille (conjoint et enfants).
- **Titres-restaurant Swile** au quotidien.
- **Organisation hybride**, avec une forte place accordée au travail en présentiel — généralement 3 à 4 jours par semaine au bureau — afin de favoriser la cohésion d'équipe, les échanges spontanés et les temps de travail approfondi en commun.

Ce qui rend ce poste unique

- **Impact au sein de l'équipe fondatrice** : rejoignez l'entreprise parmi ses tout premiers salariés et contribuez directement à définir la trajectoire d'une startup deep tech particulièrement ambitieuse.
- **Une startup portée par une mission** : contribuez à relever certains des principaux défis de l'IA — l'une des technologies les plus transformatrices de notre époque — en rendant les LLM fiables, transparents et économes en ressources, avec des bénéfices directs en matière de durabilité environnementale.
- **Travailler sur des technologies de pointe** : développez des systèmes combinant LLM et vérification formelle, un domaine émergent à la pointe de la sûreté et de la fiabilité de l'IA, où de nouvelles avancées apparaissent chaque semaine.
- **Une équipe de haut niveau** : collaborez avec un conseil scientifique composé de chercheurs issus d'institutions de premier plan (École polytechnique, ENS Paris-Saclay, etc.) ainsi qu'avec des business angels expérimentés issus

d'entreprises technologiques majeures (Datadog, Dataiku, Nabla, Artefact, etc.).

- **Des clients stratégiques** : travaillez avec des entreprises de premier plan dans la santé, les logiciels d'entreprise et d'autres secteurs critiques.
 - **La flexibilité d'une startup en phase de démarrage** : une grande autonomie pour définir votre manière de travailler et vos priorités.
-

Processus de recrutement

1. **Candidature** : envoyez votre CV via le formulaire Google indiqué dans l'offre.
2. **Premier échange - 30 min** : discussion avec le CTO afin d'échanger sur la vision de l'entreprise et les attentes mutuelles.
3. **Entretien technique approfondi** : exercice à réaliser chez soi, suivi d'une discussion de 60 minutes. L'objectif est de déterminer si la collaboration peut fonctionner dans de bonnes conditions.
4. **Entretien technique - 60 min** : échange approfondi portant notamment sur votre maîtrise des outils que vous serez amené à utiliser.
5. **Entretien avec les fondateurs - 45 min** : échange final avec le CTO et le CEO autour de la stratégie, de l'adéquation avec la culture de l'entreprise et de votre évolution à long terme au sein de Formel AI.

URL de l'offre :

<https://docs.google.com/forms/d/e/1FAIpQLSfvLZeeEiMIhtFWlmd2IbaQ9ghPhYw3r0nSRVHZejb>

Lien vers l'offre sur le site dataia.eu : <https://da-cor-dev.peppercube.org/node/1535>