

# Projet ANR AT2TA

## Descriptif détaillé du poste d'ingénieur de recherche

CDD 18 mois

### Objectifs

Le travail consistera en l'implémentation d'une plateforme logicielle pour le projet ANR AT2TA. Cette plateforme sera composée

- d'un site web pour diffuser les résultats (publications, datasets, benchmarks, leaderboard) et les informations du projet
- des services web de démonstration des technologies qui seront développées par les partenaires du projet.

Ce travail aura lieu dans le cadre du projet ANR AT2TA<sup>1</sup>, projet de recherche sur la résolution et l'utilisation du raisonnement analogique en data science. Plusieurs plateformes existent déjà (ANNA<sup>2</sup> pour les analogies morphologiques, KGPrune<sup>3</sup> pour l'extraction de sous-graphes de connaissances) et serviront de base pour continuer de développer la plateforme AT2TA.

#### Mission 1 : Développement web

- Maintenir le site web pour le projet
- Intégrer au site plusieurs éléments de visibilité : présentation du projet, publications, datasets, benchmarks, leaderboard, code

#### Mission 2 : Intégration des démonstrations en tant que web services

- Maintenir et étendre les plateformes de démonstration existantes : ANNA et KGPrune
- Accompagner les membres du projet pour développer des démonstrations associées à leur travaux en adoptant et étendant le framework de plateforme mis en place
- Participer au développement de telles démonstrations, finaliser l'adaptation si besoin et intégrer les démonstrations fournies par les membres du consortium afin de les exposer sur le site du projet sous la forme de web services
- Interfacer les démonstrations développées avec le site pour les datasets et avec un cluster (MBI ou G5K) pour les calculs des démos
- Étendre les démos avec de nouveaux modèles de Machine Learning si nécessaire

#### Mission 3 : Participer à la préparation et la réalisation d'une shared task international

- Aider à la préparation et la mise à disposition des jeux de données de référence
- Aider à la mise en place d'un leader board
- Aider à l'évaluation des participants
- Répondre aux questions des participants

#### Mission 4 : Reporting et suivi des activités autour de la plateforme

- s'assurer que les contenus (statiques et web services) sont régulièrement mis à jour sur la plateforme
- interagir pour cela avec les membres du consortium pour leur expliquer, les relancer si besoin
- faire un reporting régulier au WP2 et project leaders

#### Activités associées :

- rédaction de certains livrables techniques du projet AT2TA liés à la plateforme
- étude des composants de la plateforme déjà installée ANNA
- se former dans certaines technologies non maîtrisées qui pourraient s'avérer nécessaires

---

<sup>1</sup> <https://at2ta.loria.fr>

<sup>2</sup> <https://anna.loria.fr>

<sup>3</sup> <https://kgprune.loria.fr>

# Connaissances liées au poste

## Compétences opérationnelles

Maîtrise technique et expertise scientifique :

- Développement web : HTML, CSS, Javascript, React js, Bootstrap, JQuery
- Langage de Programmation et DevOps: Python, Fastapi, Docker, Git, Gitlab, Singularity Container, Github,
- Frameworks de Deep learning : Tensorflow et PyTorch pour développer, entraîner et évaluer des modèles
- Documentation technique: Rédaction de documents techniques clairs et détaillés, incluant des manuels d'utilisation, capacité de rédiger des rapports de projets etc)

## Compétences relationnelles

- Capacité à mener des projets de manière autonome, en gérant son emploi du temps et ses priorités efficacement
- Aptitude à prendre des décisions éclairées et judicieuses dans l'exercice de ses attributions, en tenant compte des objectifs et des contraintes
- Recherche et Développement : engager des recherches pour rester à jour avec les dernières avancées technologiques et méthodologiques, et intégrer ces nouvelles connaissances dans le travail quotidien
- Compétences en communication interpersonnelle, facilitant des interactions claires et constructives avec l'équipe. En particulier, des présentations aux membres du consortium sur la structure et potentialités de la plateforme en termes de gestion et déploiement de composantes (« micro-services »).

## Tendance d'évolution du Poste

- Compétences Techniques : maîtrise des frameworks de deep learning tels que Tensorflow, Pytorch, et python.
- Expérience pratique dans l'entraînement et l'évaluation de modèles de langage, y compris les modèles de type BERT, et LLAMA.
- Compétence en traitement automatique du langage naturel (NLP), notamment dans la résolution d'analogies morphologiques et sémantiques.
- Développement et Implémentation, Gestion de projets, Analyse et Résolution de Problèmes, Analyse et Résolution de Problèmes, Communication et Collaboration
- Collaboration avec des équipes de recherche pour publier des travaux scientifiques ou techniques.
- Participation à des conférences et à des ateliers pour rester à jour sur les dernières avancées en NLP et en IA.

## Livrables attendus (à titre indicatif)

Les principaux livrables logiciels sont :

- Le site web du projet et une V2 service web modulaire pour les tâches liées aux analogies
- Mise en place d'un benchmark pour une campagne d'évaluation dédiée aux analogies : ce service diffusera les corpus d'évaluation, mettra à disposition les protocoles et les outils d'évaluation, le leaderboard, etc.
- Interface web service / cluster Grid5000 pour passage à l'échelle des modèles de langage
- Packaging et diffusion de la plateforme sous la forme d'un code open-source auto-hébergeable

Ces livrables sont destinés à être diffusés en open-source.

## Contacts

Pour plus d'information, contacter :

- Miguel Couceiro [miguel.couceiro@loria.fr](mailto:miguel.couceiro@loria.fr) / [miguel.j.couceiro@tecnico.ulisboa.pt](mailto:miguel.j.couceiro@tecnico.ulisboa.pt)
- Pierre Monnin [pierre.monnin@inria.fr](mailto:pierre.monnin@inria.fr)
- Cherif-Hassan Nouradine [cherif-hassan.nouradine@loria.fr](mailto:cherif-hassan.nouradine@loria.fr) / [nouradinecherifhassane@gmail.com](mailto:nouradinecherifhassane@gmail.com)