

64 - Implémentation locale d'agents IA optimisés pour répondre à des cas d'usages concrets

Partenaire extérieur : Wavestone, 6 Rue Nathalie Sarraute, 44200, Nantes, 0149032383

Correspondant administratif du partenaire extérieur : Axel BOUCHEROT, Responsable relation école, axel.boucherot@wavestone.com

Tuteur entreprise qui réalisera le suivi du groupe (si différent du correspondant administratif du partenaire extérieur) : Anaël YAHY, Consultant IA, anael.yahi@wavestone.com

Mots clés du projet : Intelligence artificielle, LLM, Agents

- PRESENTATION DU PARTENAIRE, CONTEXTE DU PROJET ET POSITIONNEMENT VIS A VIS DU PARTENAIRE :**

Wavestone est un cabinet de conseil en transformation numérique qui rassemble aujourd'hui plus de 5500 collaborateurs dans 17 pays. Au sein du bureau de Nantes, nous sommes 150 consultants, consultants seniors et managers à intervenir auprès d'entreprises de référence implantées dans toute la région Grand Ouest. Nous les accompagnons en proximité dans leurs transformations digitales et leurs décisions stratégiques.



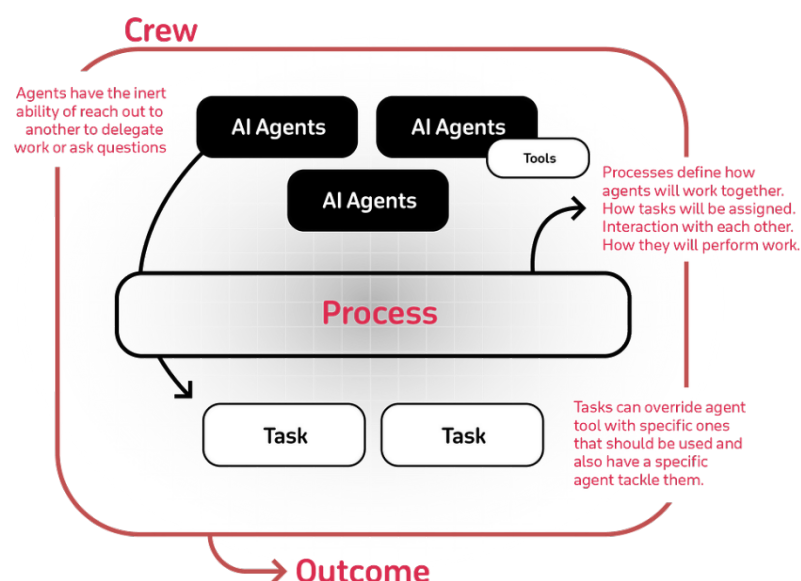
Une offre qui se développe rapidement aujourd'hui est l'aide au déploiement de solutions d'IA générative auprès de nos clients. A cette fin, le cabinet étudie et développe de multiples technologies pouvant répondre à des cas d'usage variés.

Parmi elles, nous nous intéressons aux technologies dites « d'agents GPT », qui peuvent être combinés afin de réaliser des tâches complexes nécessitant successivement des actions et compétences différentes (recherche web, analyse de documents, rédaction de code...).

- **DESCRIPTIF SUCCINCT DU PROJET :**

Pour rester au plus proche des évolutions technologiques, Wavestone s'interroge sur les capacités réelles des agents LLM, qui sont présentés comme étant des outils permettant la réalisation d'actions successives et complexes (recherche web, analyse de documents, rédaction de code...).

En particulier, nous étudions l'outil « CrewAI » permettant d'implémenter ces équipes d'agents d'intelligence artificielle dans des environnements LLM locaux (Ollama) pour garantir la confidentialité des données utilisées.



Votre objectif sera donc d'identifier des cas d'usages pertinents pour plusieurs grandes entreprises et/ou collectivités territoriales et d'étudier l'implémentation de crews d'agents IA optimisés pour répondre à ces cas d'usages.

Un élargissement avec une réflexion autour des autres frameworks d'agents LLM (comme AutoGen) pour comparer les performances et cas d'usages pourrait également être envisagé.

- **LIVRABLES :**

- Etude bibliographique sur l'état de l'art des agents LLM (On-premise, Cloud...) – 20%
- Construction d'une liste de cas d'usages pertinents (sur lesquels l'évaluation de la performance des agents sera faite) – 10%
- Dossier d'analyse reprenant les performances d'agents LLM par cas d'usages (contenant notamment les configurations des agents les plus performantes) – 60%
- Construction d'un corpus de configurations d'agents (sous git par exemple) contenant les paramètres permettant d'avoir les performances maximales par cas d'usages – 10%

- **DONNEES D'ENTREES FOURNIES PAR LE PARTENAIRE**

- Listes de besoins collectés qui seront à associer à des cas d'usages spécifiques (besoins clients, besoins Wavesone...)
- Supports de présentation internes Wavesone sur les LLM, Ollama et thématiques IA Générative

- **RESSOURCES PARTICULIERES**

Les agents LLM sont consommateurs en ressource et seront difficilement exécutables sur des PCs standards. Quelques modèles légers peuvent fonctionner mais nous sommes en réflexion sur la fourniture d'un PC IA plus puissant (matériel ou via service Cloud) utilisable pour exécuter les agents (max 1 PC pour tout le groupe d'étudiant)

- **COMPETENCES MISES EN ŒUVRE :**

- Langage Python
- Machine Learning (modèles LLM)
- Appels API
- Analyse de cas d'usage