

Stage : Data Scientist (H/F)

Speech2Text

Qui sommes-nous ?

Quantmetry c'est une centaine de personnes qui travaillent de concert pour accompagner nos clients dans leurs réflexions, P.O.C., industrialisation autour de la Data, l'Intelligence Artificielle, le Machine Learning et la Big Data. Nous sommes ce qu'on peut appeler un cabinet de conseil pure player en Data.

N'étant pas sectorisés, nous intervenons sur différents sujets (voici une petite liste non-exhaustive) :

- La maintenance prédictive,
- La segmentation clients, le scoring,
- La détection de fraude, de corruption, de blanchiment d'argent,
- Le développement des architectures de plateformes Data,
- L'analyse de textes et d'images dans le cadre de la lutte contre le cancer du sein.

Dans le cadre de notre incessante croissance, nous recherchons des personnes motivées pour nous accompagner et ainsi travailler sur des projets de recherche.

Préalable

Quantmetry propose ci-dessous le volet R&D d'une offre de stage. Tout stagiaire entrant chez Quantmetry, outre le travail de R&D qui lui est proposé et pour lequel il sera encadré, aura aussi pour objectif de participer à certaines missions de conseil chez des clients variés, lui permettant d'aborder le monde du consulting.

Contexte

Quantmetry se positionne depuis plusieurs années comme pionnier sur des sujets d'Intelligence Artificielle nécessitant une forte composante R&D. Afin de satisfaire au mieux les besoins de ses clients, Quantmetry se structure autour de pôles d'expertises, et en particulier le pôle NLP (Natural Language Processing).

Le Natural Language Processing (NLP) est une branche de l'informatique et de l'intelligence artificielle qui cherche à comprendre et à modéliser le langage. Plus précisément, cette discipline cherche à rendre le langage compréhensible par un ordinateur, de sorte à ce que ledit ordinateur puisse mimer la compréhension et le raisonnement d'un humain, de façon autonome, et plus rapidement.

Le langage oral étant le vecteur privilégié de communication depuis quelques (milliers d') années, c'est naturellement que sa transcription en texte écrit (ou Speech2Text) occupe aujourd'hui une place de choix au sein des grandes problématiques adressées par le NLP.

Après avoir réalisé avec succès missions et concours R&D sur le sujet, Quantmetry souhaite continuer d'accroître sa connaissance du Speech2Text avec l'ambition de réaliser des sujets aussi divers que le sous-titrage, la transcription de conversations cockpit, la transcription de rapports-terrains, du speech2text offline and on device, ... La popularité de cette tâche a largement cru au cours de ces 2 dernières années, notamment avec l'essor d'architectures telles que [DeepSpeech](#) ou bien [SMLTA](#). Le stage se propose donc d'internaliser ce type de technologies.

TÂCHES PROPOSÉES

La ou le stagiaire aura à produire un état de l'art sur les sujets de Speech2Text. La ou le stagiaire aura à entraîner et/ou fine-tuner des architectures à l'état de l'art, sur des données en anglais et en français.

RÉSULTATS ATTENDUS

Les résultats attendus sont le code source documenté accompagné d'une documentation scientifique et technique, comprenant une revue de l'état de l'art sur les sujets Speech2Text. Une ou plusieurs restitutions sous forme de présentation orale à des publics experts ou novices sont également à prévoir.

Compétences souhaitées

Nous recherchons quelqu'un.e de dynamique et curieux.se, qui saura s'intégrer et faire progresser notre ambitieuse expertise NLP !

Savoir-être

- Autonome
- Travail en équipe
- Proactif

Data Science

- Machine Learning
- Expérience en Deep Learning (stage, projet et/ou cours)

Engineering

- Python
- Bonnes pratiques de développement logiciel