Une image contenant Police, Graphique, graphisme, logo

Description générée automatiquementUne image contenant Police, logo, Graphique, texte

Description générée automatiquement

POLYTECH TOURS

64 avenue Jean Portalis

37200 TOURS, FRANCE

Tél +33 (0)2 47 36 14 14

[www.polytech.univ-tours.fr](http://www.polytech.univ-tours.fr/)

Rapport de stage de 4e année (2023-2024)

**Exploration des grands modèles de langages pour réaliser des prédictions et détections d'anomalies dans des séries temporelles**

**Entreprise :**

EVERDYN

**Une image contenant Police, Graphique, capture d’écran, logo

Description générée automatiquement**10 Rue Aristide Briand  
37390 Notre Dame D’Oé  
France

**Tuteur Entreprise :**

FREMONT Thomas

Responsable Agence Centre Ouest

**Étudiant :**

MERCIER Titouan

Promo 2025

**Tuteur académique :**

RAGOT Nicolas

# Sommaire

[Sommaire 1](#_Toc174724769)

[Table des illustrations 2](#_Toc174724770)

[Introduction 3](#_Toc174724771)

[1 Environnement 4](#_Toc174724772)

[1.1 Structurel 4](#_Toc174724773)

[1.2 Relationnel 5](#_Toc174724774)

[1.3 Contextuel 6](#_Toc174724775)

[2 Synthèse du travail effectué 7](#_Toc174724776)

[2.1 Recherche sur le Domaine 7](#_Toc174724777)

[2.1.1 Familiarisation avec le Domaine 7](#_Toc174724778)

[2.1.2 Fonctionnent des LLM 8](#_Toc174724779)

[2.1.3 Les techniques pour adapter un LLM 9](#_Toc174724780)

[2.2 Réalisation 12](#_Toc174724781)

[2.2.1 Faire un apprentissage 12](#_Toc174724782)

[2.2.2 Création des dataset 14](#_Toc174724783)

[2.2.3 Évaluation des performances 15](#_Toc174724784)

[3 Analyse réflexive de l’expérience 17](#_Toc174724785)

[3.1 Mon organisation 17](#_Toc174724786)

[3.2 Comparatif prévision/réalisation effective 21](#_Toc174724787)

[3.3 Les compétences acquises et consolidées 22](#_Toc174724788)

[Bilan de l’expérience 23](#_Toc174724789)

[Table des annexes 24](#_Toc174724790)

# Table des illustrations

**Aucune entrée de table d'illustration n'a été trouvée.**

# Introduction

Curieux et désirant évoluer dans ma pratique de l’informatique, je souhaite clôturer ma 4e année de formation d’ingénieur en réalisant un stage en entreprise. Ce stage me permettrait de confronter pour la première fois mes connaissances et compétences informatique au sein d'un environnement professionnel. Il sera également l’occasion d’avoir un potentiel aperçu de mon futur professionnel.

En effet je souhaite travailler dans le domaine de la cybersécurité. Pour tendre à cet objectif, je me suis par ailleurs engagé l'année prochaine à poursuivre, en double diplôme, une maîtrise en cybersécurité. J'ai donc initié mes recherches de stage vers ce domaine afin d’acquérir ma première expérience en cybersécurité. Je suis également resté ouvert à toute autre proposition de stage dès l'instant quelle correspondait à ma formation d'ingénieur.

Cet état d'esprit qui m'a permis d'avoir l'opportunité de réaliser mon stage au sein de l'antenne tourangelle d’EVERDYN. Cette entreprise m’a permis de passer mes 9 semaines de stage autour du domaine de l'intelligence artificielle. En effet j'ai dû *Explorer les grands modèles de langages pour réaliser des prédictions et des détections d'anomalies dans des séries temporelles*.

Cela peut à première vue sembler être éloigné de mon désir professionnel. Mais à la vue de la proposition de celui-ci je me suis rendu compte que le domaine de l'IA m’est relativement obscur. En effet durant ma formation d'ingénieur mes choix de cours et de projets non pas eu l’occasion de tendre vers le domaine de l'IA. Au vu de l'engouement actuel et de son imbrication grandissant au sein de la société et des différents domaines de l’informatique, je trouvais très intéressant de passer 9 semaines à mieux le comprendre.

Je vous propose à travers ce rapport de vous présenter dans un premier moment l’environnement de mon stage. Nous pourrons mieux connaitre l’entreprise et les collaborateurs avec qui j’ai pu évoluer du 17 juin au 16 aout, puis nous comprendrons le contexte et l’objectif de mon stage.

Nous verrons dans un second moment le travail accompli. Dans cette partie je détaille chaque étape qui importe, et évoque les choix fait face aux difficultés rencontrées. Nous finirons avec une analyse réflexive de mon expérience. Ici je vous présente et critique mon organisation, je fais également un point sur les connaisseurs et compétences acquérir durant mon stage.

# Environnement

## Structurel

Fondée en 2004, EVERDYN est une PME de 18 salariés domicilié à Aix en Provence mais présente dans toute la France grâce à son réseau de 5 agences. Ces agences ne sont pas réparties en centre de profit, il n’y a donc pas de concurrence entre elle.

EVERDYN évolue dans deux domaines d’activité. La première, leur principale activité présente au sein des 5 agences, est l’intégration, réparation et supervision de système. Ils sont capables de concevoir, programmer, démarrer la plupart des systèmes automatisés du monde industriel.

Leur second domaine d'activités, présent seulement au sein de l’agence d’Aix-en-Provence, est le développement de logiciels. Ils développent des logiciels industriels pour les utilisateurs d'AVEVA Software. Leur produit majeur est le logiciel et Interface de programmation d’application ALPANA. Elle permet de faire du « Dashboard », c’est-à-dire transformer des données en information par l’intermédiaire de représentation graphique. Elle est capable de s'interfacer avec toutes les sources de données possibles et permet aux industriels de monitorer leurs chaines de production, de partager et visualiser des tableaux de bords sur tout type de support électronique.

Crée en 2007, l’agence de Tours, où j'ai effectué mon stage, travaille uniquement dans l'intégration, la supervision et la réparation de systèmes. L'ensemble des 6 collaborateurs de Tours sont des automaticiens.

Il travaille pour des vendeurs de matériel industriel. Leur rôle est, une fois le matériel livré à l'entreprise acheteuse, intégrer l’automatisation du nouveau système et gérer l’interfaçage avec le reste de la chaine de production. Dans ce cadre EVERDYN travaille régulièrement pour CIAM PAKE, spécialiste dans la conception de systèmes de convoyage de contenants. Actuellement ils travaillent ensemble à l'intégration d’un système de convoyeurs à l’entreprise Laiterie de Saint-Denis-de-l'Hôtel.

Ils ont également des clients de l’on terme, c'est-à-dire qu'ils travaillent chez eux sur de plus longe période et directement dans leur entreprise pour superviser et développé l'automatisation de leur système. On peut citer par exemple l’entreprise SKF ou STMicroelectronics.

## Relationnel

Durant mon stage j'ai majoritairement interagi avec 3 personnes.

M. FREMONT Thomas, mon maître de stage :

Automaticien de métier, il est actuellement Responsable de l’agence de Tours. Il s'occupe de la gestion du personnel, de la planification des projets et de la répartition des effectifs chez les différents clients. Cela occupe environ 10% de son activité. Le reste du temps, il exerce son métier d’automaticien et conseille éventuellement les autres collaborateurs grâce à l’expérience qu’il a pu accumuler depuis son arrivée au sein de la société en 2008.

Il a été là pour m’introduire dans l’entreprise, me fournir le matériel nécessaire à mon travail ainsi que répondre à mes différentes questions sur l’organisation de l’entreprise. Il a été également là pour vérifier que je n’étais pas en difficulté dans la réalisation de mon stage.

M. LE NY Nicolas, Gérant de EVERDYN :

Il travaille au siège social à Aix-en-Provence. Il a défini mon sujet de stage et qui a contrôlé mon avancement. C'est également vers lui que je devais me tourner pour toutes les questions relatives à la définition du besoin ou aux attentes de l’entreprise sur la réalisation de mon stage. Nous avons exclusivement échangé par email et téléphone.

Mme. LE NY Cécile, Assistante Administrative :

Elle s’est occupée de mon accueil administratif, vérifiant la validité de ma convention de stage et en s'assurant que j'ai eu à disposition tous les documents légaux obligatoires. Elle m’a aussi aidé à obtenir toutes les informations à propos des politiques de qualité de vie au travail mise en œuvre au sein de la société.

Les différents collaborateurs de l’agence de Tours :

Rarement dans les locaux de l’entreprise car souvent en déplacement chez les clients, cela étant liés à la nature de leur travail. Ils sont au nombre de 6, dons mon maître de stage. J’ai eu le temps de les croiser tout au long de ma présence au sein de EVERDYN. Nous avons pu échanger lors de moments plus ou moins formels à propos de mon sujet de stage et de leur métier. cela ma périmiez de mieux comprendre les différente faséte du métier d’automaticien ainsi que les diffèrent partie qui pouvais être amènes a participer a l’arrivée d’un nouveau système au sein d’une entreprise.

## Contextuel

M. LE NY Nicolas a pour ambition de créer un outil informatique permettant d’anticiper les comportements des différents systèmes de l’industrie. Il permettra aux industriels de prédire les futurs dysfonctionnements des systèmes ou de prédire des variables liées à celui-ci comme par exemple la consommation énergétique ou le rendement.

Cet outil sera proposé aux entreprises souhaitant prédire le comportement de leurs systèmes sans avoir à investir dans la création d'un modèle de connaissance de celui-ci, qui peut s'avérer couteux.

Cet outil se veut générique et facilement adaptable afin d’être déployé dans tous types d'entreprise et pour tous types de système.

Pour fonctionner, cet outil se basera sur les données et covariables des systèmes sur lesquelles il devra effectuer des prédictions et une détection d'anomalie sur les données produites par des systèmes. Dans l'industrie ces données sont nombreuses et déjà stockées de manière chronologique dans des bases de données. Cette solution devra donc être capable d'assimiler les différentes données des systèmes d’une entreprise spécifique afin que chaque solution vendue soit peaufinée pour l'acheteur. Cette solution devra également être capable de s'exécuter localement dans chaque entreprise, car la majorité des entreprises ne veulent pas que leur données, soit exfiltrées de leur système. Les résultats de ces prédictions et détections seront ensuite utilisés par l'entreprise pour faciliter ces futures prises de décision.

Nous pouvons résumer cette solution en un mini cahier des charges de 2 fonctionnalités et une contrainte.

Mini cahier des charges :

FC1 : Prédire les futures valeurs d'une série temporelle

FC2 2 : Détecter une anomalie sur des valeurs de séries temporelles

CT1 : l'outil devra fonctionner en local au sein de l'entreprise pour lequel il fonctionne

Périmètre de mon stage :

L'objectif qui m’a été fixé par M. LE NY est d'explorer le potentiel de la technologie LLM pour la réalisation des prédictions et détections d’anomalie sur des séries temporelles. Cela lui permettra de voir s'il y a un intérêt à utiliser les LLM pour le développement de son nouvel outil. Son désir d'aller regarder vers cette technologie a été mû principalement par la lecture de l'article « Large Language Model Performance in Time Series Analysis » publié le 1er mai 2024 sur medium.com, un site de blog qui a une dimension de partage de contenu scientifique par la présence de nombreux articles sur divers types de technologie.

Il souhaitait également mieux comprendre cette technologie qui connait actuellement un grand engouement. ( faudra donc privilégier des LM peu volumineux capable de fonctionner sur de localement dans les entreprises. Ce qui a principalement pousser Monsieur le a envisagez d'explorer les solutions la M et la lecture de cet article … . Une autre contrainte est d'essayer de se tourner principalement vers des technologies françaises notamment celle de mistrale IA. Cela pourra faciliter l'acceptation de cette nouvelle technologie par les clients. )

# Synthèse du travail effectué

## Recherche sur le Domaine

Durant les 2 premières semaines de mon stage j'ai commencé par étudier le sujet en réalisant un mini état de l'art. J'ai commencé par me renseigner sur les différentes techniques qui permettent de réaliser des prédictions sur des séries temporelles. J’ai résumé ces recherches dans la sous-partie suivante « familiarisation avec le domaine ». Ensuite je me suis renseigné plus précisément sur la technologie LMM qui est au cœur du sujet de mon stage et également regarder les différentes méthodes existantes pour adapter leur comportement à une tâche spécifique. J’ai retranscrit ces résultats dans les parties 2.1.2 et 2.1.3

### Familiarisation avec le Domaine

**C’est quoi une série temporelle ?**

Une série temporelle et définie comme étant une suite d'observations d'une variable mesurer de façon ordonnée dans le temps. Ici la série contient des observations pour plusieurs variables elle est dite multivariée. Dans le cas contraire il s'agit d'une série temporelle uni variée. Elle peut être aussi continue ou discrète. Durant mon stage j'ai manipulé seulement des séries temporelles discrètes.

Une série temporelle peut être découpée en 3 composantes :

* La tendance : Elle traduit le comportement ”moyen” de la série, C’est à dire qu’elle représente sont évolution à long terme.
* La saisonnalité : Elle correspond à un phénomène qui se répète à intervalles de temps réguliers (c’est un phénomène saisonnier).
* Le bruit : Il correspond à des fluctuations irrégulières, en général de faible intensité mais de nature aléatoire.

**Comment faire des prédictions et de l’analyse d’anomalie dans les données ?**

La prédiction et détection d'anomalie sur des valeurs à partir de série temporelle n’est pas un sujet nouveau si les raiçant avancé en intelligence artificielle ont permis des progrès notamment en pouvant prendre en compte un nombre élargie des facteurs qui peuvent influencer le comportement d'une série temporelle. Les premiers travaux de recherche à ce sujet ont été initiés dans le domaine de la statistique et de l'analyse de données. C’est technique sont encore très performent et largement utiliser. On peut Sitter par exemple les deux méthodes suivantes :

* Lisage exponentielle : Cette méthode attribuant des poids dégressifs aux observations passées. Ce processus est contrôlé par un paramètre de lissage qui varie entre 0 et 1. Plus on et proche de 1, plus le modèle réagit rapidement aux changements récents dans les données. À l'inverse, plus on et proche de 0, plus la série sera davantage lissé, réduisant la sensibilité aux fluctuations rapides. Cette technique est particulièrement efficace pour des séries chronologiques qui présentent une tendance ou une saisonnalité, en permettant d’obtenir des prédictions plus stables et plus précises sur le court terme.
* Modèle ARIMA/SARIMA : modèles sont largement utilisés en économie, en finance et dans de nombreuses autres disciplines pour produire des prévisions précises basées sur des données historiques. ARIMA fonctionne grâce à l'auto-régression (AR), qui prend en compte les valeurs passées de la série pour prévoir les futures, l'intégration (I), qui rend la série stationnaire en éliminant les tendances, et la moyenne mobile (MA), qui lisse les fluctuations en utilisant des moyennes mobiles. Le modèle SARIMA étend ARIMA en ajoutant des composantes saisonnières pour traiter les séries présentant des patterns cycliques. Mes pours utiliser c’est model il faut les adaptes pour chaque utilisation car chaque modèle car il y a défaire paramètre a adapté (p,d,q, P, D, Q) . Et pour avoir une bonne adaptation des paramètres il faut analyser les données utilisées.
* PCA : l'analyse en composante principale est une très bonne méthode pour détecter des anomalies au sein la série temporelle multivarié car cette technique permet de réduire la dimension du problème. Les valeurs anormales en observant. Cette technique nécessite l'interprétation humaine afin de d'identifier les valeurs qui s'écartent au sein des composantes principales.

Avec l’arrivée des réseaux de neurone, de l’apprentissage automatique puit de l’apprentissage profond de novelle technique son arriver pour réaliser des prédiction sur des série temporelle :

* LSTM : (Long Short-Term Memory) sont un type de réseau de neurones récurrents (RNN) conçu pour traiter et modéliser des données séquentielles tout en surmontant les limitations des RNN traditionnels qui ont le problème du gradient évanescent, ce qui les rend inefficaces pour apprendre à partir de longues séquences, car l'information tend à se perdre au fil des couches. C’est LLM sont souvent utiliser pour réalise des détection d’anomalies pour les données et peuve également être utilise pour de la prédiction de valeur.
* T-GCN : (Temporal Graph Convolutional Networks), apparu en 2022 cette technologie permet de prédire des séries chronologiques non stationnaire. Combine des réseaux de neurones convolution els (CNN) et des graphes.
* Isolation Forest : cette méthode est très performante pour détecter les anomalies dans les dataset notamment avec les séries temporelles multivariée. Cette méthode fonctionne grâce à l'apprentissage automatique non supervisée. Cette technique fonctionne en créant un ensemble d'arbres de décision Elle crée ses arbres de décision en sélectionnant aléatoirement les variables, puis en réalisant un des partitionnement dans le jeux de données afin de trouver des valeurs isolées.

### Fonctionnent des LLM

Les LLM, grands modèles de langage, sont des systèmes d'intelligence artificielle. Ils sont majoritairement utilisés dans le cadre du traitement du langage naturel. En effet les LLM sont conçu pour le traitement de donnée séquentielle. Ils sont entrainés sur de très grand Corpus de texte afin d’être en capaciter de prédire la probabilité d’une séquence de mots en fonction des mots précédant. Ce fonctionnement permet de les utilisées pour de nombreuse application, par exemple la traduction de texte, Résumé de texte, etc. C’est notamment cette technologie qui a permis l’arrivée d’outillé comme tchat gpt.

C’est LLM fonctionne grasse a une architecture de réseaux de neurones profond, et plus particulièrement sur le récente modèle d'architecture « Transformer ». Elle a été introduite en 2017 dans l’article « Attention Is All You Need », comme son nom l'indique les Transformers fonctionnent uniquement avec un mécanisme d'attention. Ce mécanisme d'attention permet au LM de prendre conscience du contexte de la phrase c'est-à-dire lorsqu'il fait sa prédiction il peut prendre en compte les prédictions passées et peut évaluer la probabilité du nouveau mot en fonction de chaque mot de la phrase précédente. Hé par exemple pour la phrase « 3 » il va prédire le mot «…3 », et non pas « … ».plu exactement le mécanisme d'attention permet de sélection les parties importantes de l’entrer en attribuant des poids aux différents tokens un l'entrée du réseau de neurones.

Plus shématique ment le LLM fonctione en 4 étap :

Input : D'un texte en entrée

Token : récupération d'un texte en entrée

Hembendi igne : Chaque token est converti en vecteur numérique

passage du vecteur dans le trensformeurs

Hembendi igne : convertion du vecteur en texte.

### Les techniques pour adapter un LLM

**Optimisation de prompte :**

Objectif : formuler au mieux les requêtes soumises au modèle afin d'obtenir la meilleure réponse possible.

Fonctionnement :

Il est généralement conseillé de donner un rôle à l'IA sur tout pour répondre à des question complexe. Il est également conseillé des détaille le contenue des pronte et d’eviter le acronime ou abréviation pour éviter le serreur de contexte. Il est aussi intéressent de détailles les étapes que l’on souhaite que l’ia suive, cela permet d’avoir des réponse plus détailler est structurer.

Lorsque l'on souhaite une génération de textes conséquence il faut plutôt privilégier les questions ouvertes et avec une formulation positive pour éviter de contraindre la sortie. Tournée des limites h pour orienter cette génération. Les formulations les plus courtes comme des questions fermées ne sont pas à proscrire pour autant, elles peuvent permettre aussi d'avoir hé des réponses plus courte.

Intérêt / limite :

L’un des principaux avantages de cette technique et qu'elle est très facile à mettre en œuvre et ne coûte rien.

Il n'y a pas besoin de modifier le LLM, en fait on cherche à l'utiliser de manière optimale. H cependant il peut être parfois un peu long de trouver un bon modèle de prompt. L'inconvénient majeur est que nous dépendons entièrement de la base d'apprentissage du modèle initial toutes les générations heure en fait à base de ces apprentissages.

Exemple : récolte de coquillage

Comme on peut le voir cet exemple mais en évidence l'importance du contexte en effet hé il y a confusion entre couteaux le coquillage et l'objet de cuisine. Le fait également de détailler plus précisément ce que l'on souhaite cela permet d'avoir une réponse plus détaillée et structurer.

**Génération augmentée de récupération (RAG) :**

Objectif : Enrichir les connaissances du modèle avec des informations externes.

Fonctionnement : Le LLM est combinais avec un système de recherche d'informations. Lors de la génération, le modèle peut accéder à une base de connaissances externe pour récupérer des informations pertinentes et les intégrer à sa réponse. Cette base de connaissances est préalablement construite et est accessible grâce au système de recherche d’informations. Les information sont stockée de manière vectorisée, c'est à dire que les données d’origine ont été passés dans une fonction « embeddings » pour transformée le texte en vecteur numérique.

Certain système de RAG permet également de site les Sours qu’il a utilisé pour produire une réponse.

Intérêt / limite :

Cette méthode permet donc de créer des assistants capables de répondre avec précision aux questions sur un domaine particulier. L'avantage est également lorsque les données de ce domaine sont amené à changer il n'y a pas besoin de refaire tout un apprentissage il faut juste changer les données dans la base vectorielle.

Cependant c’est système sont relativement complexes à mettre en place, il est difficile de créer un bon récupérateur de fichiers, point central de ce système car si la récupération de fichier et non pertinent cela engendre des hallucinations, c'est-à-dire que le LM répond à la question mais inventant les éléments de réponse. Les temps de réponse sont également allongé car le processus intègre une étape supplémentaire avant la génération «  la récupération de fichier ».

Exemple schématique RAG : ici les source pourrait être l’ensemble des livres de spécialité du réseaux Polyte, les diffèrent règlement intérieur, Brochures du Réseau Polytech, ect. Ce qui permettrait de créer un agent de réponse à toutes les questions lies a l’organisation et le fonctionnement des école Polytech.

Une image contenant texte, capture d’écran, diagramme, Police

Description générée automatiquement

**L’apprentissage (fin Fine-Tuning) :**

Objectif : Augmenter et orienter la compréhension des promptes du llm afin de le spécialiser dans une tâche spécifique.

Fonctionnement : pour réaliser un apprentissage enfin nous avons donc besoin dans jeu de données avec un prompt et la réponse attendue, il y a également d'autres techniques qui permettent d'intégrer en plus de cela des réponses que l'on ne souhaite pas. Lors de cet apprentissage les poids LM son ajuster pour que les promptes données qui lui sont données ce rapproche au plus des réponses attendues.

Intérêt / limite :

Cette technique permet donc d'adapter un LLM à une tâche spécifique en utilisant un jeu de données restreint. Elle requière également une plus faible puissance de calcul. Elle nécessite par contre d’avoir accédé au fichier de poids d’un modèle de base déjà entraîné de façon générique sur une très grande quantité de text.

Il faut être encore plus précautionneux sur les données choisies pour l'apprentissage fin, car comme le jeux de données est plus petit le modèle y est beaucoup plus dépendant. Il peut avoir un risque de surapprentissage, c'est-à-dire que le model père ça capacité de généralisation et se comporte anormalement lorsque que les données s’écarte de celles l'apprentissage. L'apprentissage fin peut donc h dégrader les performances dans LMM.

Faire du L’apprentissage et également assez complexe d'un point car il faut avoir une bonne expérience et être en mesure de réaliser plusieurs apprentissages différents pour pouvoir choisir correctement les paramètres de l'apprentissage.

Exemple :

Si l’on souhaite adapte un LLM pour reconnaitre les sentiment globale des utilisateurs d'un produit ou service on peut grasse à l'apprentissage fin adapter un llm à la reconnaissance les émotions sur des commentaires. Pour cela il faut de lui fournir un dataset avec une colonne regroupant les commentaires et une une 2nde colonne stockant les label représentant le sentiment du texte.

## Réalisation

Après c’est deux semaines de recherche et une relecture des attentes de Nicola, j’ai choisie de d’essayé d’adapter un LLM pour la prédiction de valeur. Avec l’accord de Nicola je j’oisie de commencé dans un premier temps de me consacrer sur la prédiction de valeur car apriori pour une simple détection d’anomalie il y a déjà des techniques très performante comme l’Isolation Forest.

J'ai choisi pour adapter le comportement des LLM de cumuler 2 techniques, l’apprentissage fin et l’optimisation de prompt. J’ai privilégié l’apprentissage fin car ici je souhaite que l’IA développe une nouvelle capacité, la prédiction de valeur. Pour développer cette nouvelle capacité, la technique RAG n’est pas adapté. En effet je veux que le LMM soit capable de générer de nouvelle valeur et non d’aller chercher les données qu’il souhaite prédire dans une base de données extérieure.

( Les contrainr …)

(annonce plans )

### Faire un apprentissage

Pour réaliser mes différents apprentissages j'ai choisi d'utiliser l’environnement de « Hugging Face », une start-up franco-américaine qui permet de partager des modèles de LLM en open source. Elle fournit également une bibliothèque python « transformer» qui faciliter la manipulation des LLM des outils permettant de télécharger inféré et former facilement des modèles pré entraînés.

**Choix des modèles :**

Mistral 7B : Ce modèle comprends 7,3 Billion de paramétré et occupe un place de 30 Go. Je l’ai choisi car l’entreprise qui la crée, « Mistral IA », le rendu disponible en open source et on peut également trouver les fichiers de ce modèle sur Huggung Face. De plus ce modèle et très performant car il dépasse pour un modèle équivalent les performances Llame qui sont les modèles open source proposés par « Meta ». De plus mon maitre de stage souhaiter que je regarde les modèles proposés par Mistral IA car cette entreprise et française et cela hé serait un argument de vente supplémentaire pour l'outil final.

[LiteLlama-460M-1T](https://huggingface.co/ahxt/LiteLlama-460M-1T) :

Afin d’avoir un modelé plus facilement manipulable et surtout beaucoup plus léger j’ai choisie de faire des assai avec « [LiteLlama-460M-1T](https://huggingface.co/ahxt/LiteLlama-460M-1T) », ce modèle est une version allégée de LLaMa 2 il contient seulement 4700 million de paramètres. Il a été proposé par un utilisateur sur hugging face en ce modèle a une certaine popularité car il a été quand même téléchargé 20 mile fois le moi dérnier. Il faudra peut-être s'attendre à avoir de moins bonnes performances que le modèle mistral.

**Choix pour l’apprentissage :**

Afin de Réaliser les apprentissages je me suis tourné vers deux technologies l’API auto-train de Hugging Face et l’application H2o LLM studio. Ces 2 ces 2 applications ont le même rôle à savoir réaliser les apprentissages fins sur des LM pour les 2 applications on peut utiliser les modèles hugging face. Hé j'ai choisi de tester hé ces 2 pour pouvoir ensuite réaliser des apprentissages sur ces 2 applications et Hé comparer pour ensuite c'est préconiser laquelle serait le mieux utilis si l'entreprise prend la décision de continuer à travailler avec les LM à la suite de mon stage.

Au final je n'ai pas réussi faire des apprentissages en local avec ces 2 applications. Pour les 2 j'ai rencontré des difficultés au niveau des GPU. Au début je développais sur un ordi ou les GPU étaient de marque AMD. Ces 2 applications open-source fonctionnent avec les bibliothèques CUDA compatible seulement avec les GPU Nvidia. J’ai donc changer de PC pour avoir des GPU Nvidia.

Ensuit une fois les installation fait et les application fonctionnelle j’ai réalisé un apprentissage test avec un modèle léger, un très petite dataset et les paramètre de basse. L’objectif de cela est de voir si tous et fonctionnelle. Il s’est avéré au final que mémoire des GPU était trop petite. Cuda chercher à alloués de la mémoire supplémenter se qui générer une erreur et randée imposible de.

J'ai essayé de voir si il était possible de réduire le nombre d'instances hé parallélisé hé d'utiliser des paramètres en moins du modèle afin qu'il et besoin de moins de place pour s'exécuter mais ces recherches sont restées infructueuses.

J’ai utiliser seullement laplication autotrain car elle perrmetter de louer des puissance de calcule.

Auto-train :

Voici l’interface :

Une image contenant texte, capture d’écran, nombre, Police

Description générée automatiquement

### Création des dataset

Pour la création des différentes dataset qui ont servi aux apprentissages je me suis basé sur 2 sources de données.

Source 1 : *data.gouv.fr*

data.gouv.fr et une plateforme Française qui à l’ambition de réunir toutes les sources de donnée ouverte française. Sur ce site on trouve une grande quantité de donnée de diverse. Pour créer mes série temporelle j’ai choisie d’utilise la consommation électrique brute française. Pour celas j’ai formater les donne afin d’avoir un fichier plus structure et mon volumineux. En effet le jeu de données de basse présente les courbes de consommation d’électricité (par demi-heure en MW) et de gaz (par heure en MW [PCS](https://opendata.reseaux-energies.fr/pages/glossaire/) 0°C) ainsi que des métadonnée. J’ai donc fais une moyenne pour chaque jours de la consommation brute électrique sur la période du 12/01/2012 au 31/12/2020.

J’ai stoker c’est information dans de format, l’un respectent la convention Json et l’autre en limitent le nb de carter pour être plus facilement utiliser dans les prompt. Car en effe lorsque l’on utilise un LLM nous somme limiter par la taille des prompte.

Voici un extrè des deux jeux de donnée :

Source 2 : système de refroidissement d’un client de Everdyn.

Cette source de donné une série temporelle multi varier, elle correspondons à l’évolution temporelle des variables permettant de superviser le fonctionnement d’un échangeur thermique.

Les 5 variable sont :

* Débit d’eau : cette variable représente le débit de l'eau en mètre cube au sein du système cette valeur et sensé être constante.
* Pourcentage de commande des pompes : Cette variable correspond à la vitesse en pourcentage des variateurs qui commande les 3 pompes qui alimente le système en eau.
* Température départ : c'est la température en °C d'eau à l'entrée de l'échangeur
* Température d'eau glacée : cette valeur représente la température en °C de l'eau la sortie de l'échangeur. Elle doit être à une température de 5°+-0.2
* Puissance de l’échangeur : se mesurent en Kilo watt cette variable est une résiste résultante des variables précédentes elle se détermine grâce à la formule suivante .

Ces données dans son dupliqués dans 2 fichiers au format CSV. Le premier fichier retrace l'évolution des variations de valeur par paramètre c'est-à-dire à chaque fois qu'un paramètre change une nouvelle ligne est ajoutée au CV avec sa nouvelle valeur et la date à laquelle ces produits cet événement. Le 2nd fichier retrace également l'évolution des variables mais cette fois-ci indique la valeur de chaque paramètre sur chaque ligne c'est-à-dire si un changement se procure en même temps il les valeurs des variable apparaîtra sur la même ligne et il peut donc y avoir des colonnes sans valeur si le paramètre n'a pas évolué.

Voici les colonnes pour chaque fichier :

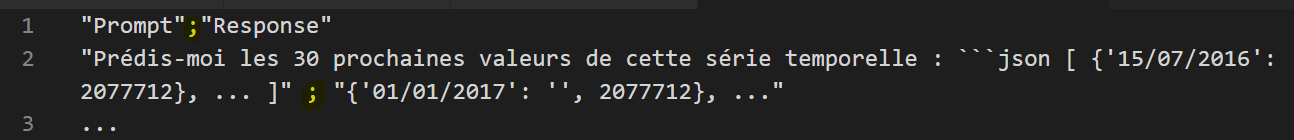
Fichier 1 : Date, Nom paramètre, valeurs, Nouvelle valeur

Fichier 2 : Date, Débit d’eau , Pourcentage de commande des pompes, Température départ , Température d'eau glacée , Puissance de l’échangeur

**Création des datasse pour l’apprentissage :**

Comme vu dans la partie « Faire un apprentissage » nous avons utilisé la technique SFT, pour que cette technique fonctionne correctement il faut fournir au LLM un ensemble de plusieurs promptes avec la réponse attendue. J'ai donc utilisé la première source de données afin de créer un LM capable de prédire les futures consommations énergétiques de la France en fonction d’une période de consommation précédente.

Voici 2 exemples de data set qui ont pu être utilisés pour les apprentissages :



Une image contenant texte, Police, capture d’écran

Description générée automatiquement

Ici les 3 petits points en ligne 2 représentent la suite des séries temporelles. Pour la partie prompte la longueur de ces séries temporelles variées entre 2 mois et un an et demi. Pour la partie réponse elle variait entre un mois et 15 jours. En effet j'ai entraîné le HLM pour qu'il puisse prédire au maximum les 30 prochaines valeurs d'une série temporelle.

Les 3 petits points en ligne 3 symbolise les autres exemples de prompt en général mes dataset contenaient entre 30 à 50 lignes. Pour créer c'est 30 à 50 lignes j'ai essayé de limiter les chevauchements Données pour éviter un problème de surapprentissage à chaque jour de l'année ne se retrouvait jamais plus de 2 fois dont un prompt. Hé pour arriver à cela j'ai découpé une première fois par année hé je faisais prédire le mois de janvier suivant au LLM. R j'ai découpé le dataset en une succession de périodes de 2 à 4 mois puis dans la partie réponse je m'étais les 15 jours qui suivait à chaque fois les périodes 2 à 4 mois.

Il est également important de noter que j'ai exclu des c’est l'année 2019 et le mois de novembre et décembre 2014 qui serviront de base de validation.

Après plusieurs essais il s'est avéré que le 2e type de prompt était plus performant. Avec le premier type lorsque l’on réaliser un inférence la sortie du LLM n'était pas la suite de la série temporelle mais plutôt un texte pour expliquer une méthode à suivre pour prédire ses valeurs. Comme le second type de prompt contenait moins de textes et une seule instruction claire je pense que le LLM compléter plus facilement la série de données en général seulement les valeurs h

### Évaluation des performances

**Méthode d’évaluation :**

Afin d'évaluer les performances de mes modèles je gardais à chaque fois une partie des données de mes sources en guise de base de validation. Ces données n'étaient donc pas utilisées pour créer les datasets utile à l'apprentissage. Cela permet de ne pas masquer un éventuel problème de surapprentissage effectivement le LLM peut être très performant si les données sont celles du dataset D'apprentissage mais dès qu’elle s'en éloigne il peut perdre en performance.

Pour comparer la série temporelle prédite par rapport à la série temporelle réel (celle de la base validation), j'ai écrit un script python qui calcule des métriques de performent et qui permettait de visualiser les deux courbes sur un même graphiques. Les métriques que j’ai choisie étais la comparaison des moyenne, médiane, écart-type des deux série temporelle. Je calculer également l’erreur moyen de prédiction de chaque valeur et l’erreur maximum et minimum, voici la formule utilise :

**Résulta :**

Premier laisser : Dans un premier temp j’ai effectué des tests avec les versions gratuites disponible en ligne des LLM Claude, Mistral, GPT3. L'objectif était d'avoir une première idée des performances des LLM hé n'ayant pas encore subi d'apprentissage spécifique pour la prédiction de série temporelle.

Mistral7B :

[LiteLlama-460M-1T](https://huggingface.co/ahxt/LiteLlama-460M-1T) :

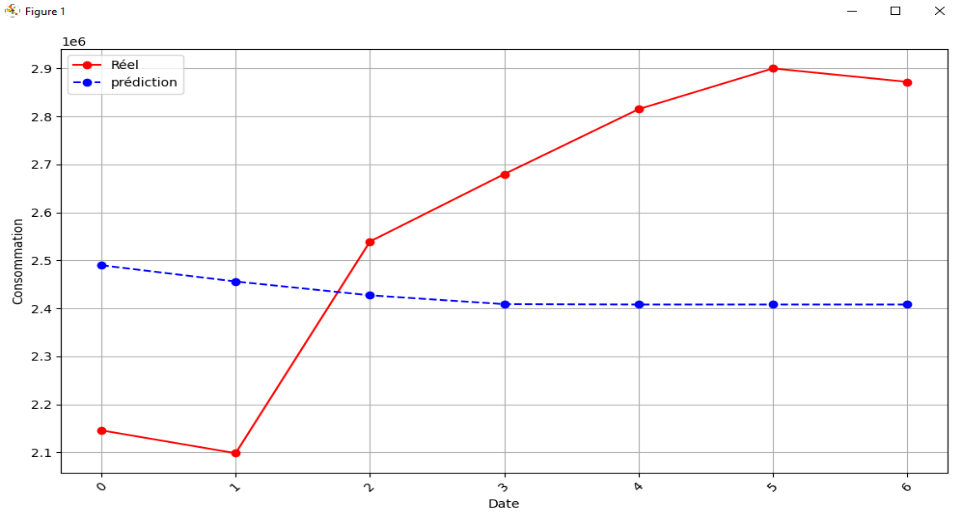
* Une image contenant texte, capture d’écran, Police, nombre

  Description générée automatiquementUne image contenant texte, Police, capture d’écran

  Description générée automatiquementUne image contenant texte, ligne, Tracé, diagramme

  Description générée automatiquementPremière génération : Avec une prompte issue d’un dada set d’entraine
* Une image contenant texte, Police, capture d’écran, typographie

  Description générée automatiquementUne image contenant texte, capture d’écran, Police, nombre

  Description générée automatiquementDeuxième génération : Avec le prompte de test

# Analyse réflexive de l’expérience

## Mon organisation

Durant mon stage, j'ai dû m’organisé seul pour planifier les différentes tâches à effectuer pour aboutir à la résolution de mon sujet. J’ai mis en place différente technique, que nous verrons si après, pour être efficace et contrôler mon avancement. Dans cette mise en place j’ai dû prendre en compte seulement deux contraintes organisationnelles imposées par l'entreprise. À savoir : les horaire de travail, être présent 7 h par jour dans les locaux de l’entreprise du lundi au vendredi et la rédaction d'un rapport d’activité hebdomadaire.

**Les outils et méthode de mon organisation :**

**Tableau de planification :** Dès la première semaine, j’ai découpé l’entièreté de ma période de stage en période ponctués par des dates clés. Vous pouvez retrouver cette planification en annexe B. Je détaille également son contenu et justifie les choix de planification dans le paragraphe suivant « La planification », page 10. Le rôle de cette planification est d’avoir une route à suivre et d’avoir une vision sur le moyen terme du déroulé de mon stage. Cette planification n’est pas rigide elle peut évoluer au fil des imprévues et apparitions de nouveaux évènements. Nous verrons d'ailleurs dans la sous partie suivante « Comparatif prévision/réalisation effective » qu’elle a été amenée à être modifié à plusieurs reprises.

**Liste des tâches à effectuer :** en complément du Tableau de Planification, j’ai créé une liste de tâche sur le bloc-notes de mon téléphone. Je mettais à jour cette liste de façon journalière. Cela m’a permis d’avoir une vision sur le court terme des différentes tâches à effectuer.

**Le Week :** C’est un rapport hebdomadaire sous la forme d’un PowerPoint, vous pouvez le trouver en Annexe A. Chaque employé, stagiaire et apprenti de l’entreprise doit le remplir et l’envoyer à son supérieur. La fonction de ce rapport est de suivre l’activité des salariés et faire remonter d’éventuelles difficultés qui pourraient apparaitre. Par exemple grâce à la section « Problème » ou « retour QHSE». Il permet également à l’employé de structurer l’exercice de son activité, notamment à travers les deux sections « réalisation » et « prévision ». Ces deux sections permettent également au supérieur d'avoir un aperçu de l'avancée des travaux effectués par le salarié. Pour ma part, je devais envoyer ce rapport à mon maître de stage et également aux patrons de la société Nicolas Le Ny.

**Cahier de stage :** J’ai choisi afin d’avoir une persistance dans les réflexions et les informations importantes que j’ai rencontré, de réaliser un genre de « cahier de bord ». Dans ce cahier, je note de façon chronologique toutes les notes que j'ai été amené à prendre j'effectuais également une séparation journalière de ces notes. Cela permet de faciliter le retour rétrospectif et de retrouver rapidement une information dont on aurait besoin et qui aurait été prise à un moment éloigné de son utilisation.

**Historique de recherche :** sous forme d’un document texte je sauvegarde tous les liens des sites internet que j’ai trouvé pertinents pour mon travail. Je rédige une petite description pour chaque lien, afin de les distinguer facilement entre eux et me remémorer plus rapidement leur contenu. Le rôle de ce document est de garder une trace de mon travail de recherche. Et ainsi pouvoir retrouver facilement des éléments qui pourraient m’être de nouveaux utiles.

**Git :** j’ai utilisé cet outil de versioning qui fonctionne avec un dépôt distant pour sauvegarder l’ensemble de mes productions écrites ( code, dataset, divers documents explicatifs, rapport de stage …). J’ai utilisé cet outil pour ne pas avoir de perte de donnée et faciliter ma mobilité de travail, car en effet j’ai été amené à travailler sur plusieurs pc différents.

Pour ce stage l’utilisation de ces outils et méthodes a bien fonctionné, un des points fort est que l’ensemble des outils sont simples et peut contraignant, ce qui fait que le temps passé à l’organisation est très faible par rapport au temps passé à la réalisation des différentes tâches.

Mon historique de recherchem’a été très utile pour la partie « 2.1 États de l’art » que j’ai rédigé en fin de stage, environ 6 semaines après la fin de mes recherches. Après réflexion malgré l’existence de cet historique de recherche il aurait été plus facile de rédiger, à minima une ébauche, de la partie « État de l’art » à la fin de ma phase de recherche car j’ai quand même du passer du temps à me remémorer/relire certaine partie de mes recherches.

Le cahier de stage ma été inspiré par la réalisation de mon dernier stage réalisé lors de l'été 2021, en effet durant ce stage les différentes prises de notes que j'ai été amené à prendre n'ont pas toujours été prises sur le même support et parfois sur des feuilles volantes. Ce qui m’a amener de pas toute les utilise en oubliande d’enconsidérré cette méthode de prise de note énergie est pas posée Hey problème de problème h si ce n'est que LA complexifié ma rétrospection de mon stage car il était plus con compliqué de les relire ou hé j'ai été amené à peut-être pas considéré certaines notes. Mais mon actuel stage étant plus long j'ai choisi cette prise de note structurée sous format qui est pour pour qu'elle soit tous réunis dans un même cahier hé cette notation chronologique permet aussi de facilement se rappeler certaines dates clés ou plus simplement à quel jour j'ai réalisé telle tâche.

Une future stage réutiliser le même fonctionnement et les mêmes outils. Il est hanté récent de considérer hé le travail de mon stage c'est déroulé en individuel et sans hé ahh non important d'interaction avec d'autres membres. Donc dans un autre environnement de stage il faudra peut-être adapter la stratégie de planification hé d'élaboration des listes des tâches hé il faudra peut-être allez faire de manière individuelle et que lancien Blue une partie des collaborateurs doivent être au courant de celle-ci.

**La planification :**

Au cours de ma première semaine, j’ai constaté que j'étais laissé en grande autonomie pour réaliser mon stage. J’ai donc pris l’initiative de créer un planning afin d'avoir des grandes étapes à suivre pour aboutir à la réalisation du sujet. J’ai donc découpé mon stage en 4 grandes phases :

Première étape : « familiarisation avec le sujet »

Je l’ai fixé à une durée de 2 semaines. Cette première période devait me permettre de mieux comprendre les différentes technologies du domaine de l’IA et également de comprendre le fonctionnement des LLM.

La seconde étape : « choix des techno »

D’une durée de 1 semaine. Cette période a pour rôle de choisir les technologies et étapes à suivre pour utiliser et adapter un LLM pour réaliser des prédictions et des détections. Elle sera également l'occasion de réaliser les premières prédictions et ou détections d'anomalies en utilisant LLM disponible gratuitement en ligne, du type … .

La troisième étape : « réalisation des apprentissages »

D’une durée de 4 semaines. Cette étape est l'étape centrale du stage elle permet de réaliser des apprentissages et d'obtenir des résultats de prédiction et de détection d'anomalie. J'ai décidé d'utiliser un processus itératif pour réaliser cette étape. Ce processus itératif se découpe en 3 sous étapes hé devoir-être répéter de façon hebdomadaire.

* Réalisation des datasses : C'est au cours de cette étape qu'on choisit comment structurer les données pour qu'elle soit au format le plus approprié pour réaliser l'apprentissage en fonction de la stratégie d'apprentissage.
* Réalisation de l'apprentissage : c'est là où on met en place le modèle, choisir les différents paramètres de l'apprentissage.
* Évaluer les performances du modèle : grâce à un script préalablement réalisé, c'est lors de cette étape que l'on récupère les différentes métriques d'évaluation des performances de notre nouveau modèle. Cela nous permet ensuite de tirer des conclusions et de faire éventuellement de nouveaux choix de stratégie d'apprentissage, de modèles ou de paramètres d'apprentissage pour les prochaines itérations afin d'améliorer les prédictions et détections.

La quatrièmeétape : « clôture »

D’une durée de 2 semaines. Cette étape a pour rôle de mettre en forme les travaux que j'ai pu effectuer et réaliser une synthèse des résultats pour pouvoir faire un retour de mon travail à l'entreprise. Cette étape sera pour moi l'occasion de prendre du temps pour rédiger mon rapport de stage.

## Comparatif prévision/réalisation effective

## Les compétences acquises et consolidées

**Compétence technique :**

Réaliser des apprentissages avec l’API « Auto Train » :

Lors de la réalisation des apprentissages sur les modèle d’IA j’ai exclusivement utiliser Auto trais j’ai donc pu apprendre à :

* Installer l’API Auto Train en Locale.
* Maîtriser les différentes mises en forme possible des Dataset pour réaliser des Finetuning sur des LLM
* Comprendre le rôle des diffèrent méthode de Finetuning. ( SFT, ORPO, DPO, Reward)
* Comprendre le rôle des diffèrent paramètre liée à l’entrainement SFT.

Exécuter tes des modélle d’IA sur un serveur distant grasse à AWS :

Executé et réalise des infairence sur des modél d’IA grâce à la librérie trensformeus :

Écrire du code python en respectant la nome Pep8 :

Manipuler des Fichier volumineux :

Dans le cadre de mon stage j’ai été amener à manipuler des fichiers volumineux. Les principaux fichiers volumineux que j'ai eu à traiter sont les dataset stocké sur mon PC et un disque dur. Il y avait également les fichiers contenant les tenseur des modèles d'IA (extension « .safetensors ») , Il était tockez sur des repos distants.

Pour les fichiers stockés en local il n'y avait pas réellement de contraintes car mes disques dur était relativement grand.

Par contre pour manipuler les fichiers volumineux sur les dépôts distants j'ai dû apprendre à utiliser un nouvel outil : « Git Large File Storage », cette application s'utilise comme Git R cependant il y a un fichier supplémentaire intitulé « .gitattributes » où il faut signifier à l'intérieur les fichiers volumineux pour qui il sois g érer de fasson optimal lors de Commit, Push, Pull.

**Compétence douce :**

Lecture est comprenions de contenue technique en anglais :

Les difairen interacteur lors de projet dans l’industrie

Installer et configure de logicielle est apit peut documenter.

# Bilan de l’expérience

C'est 9 semaines de stages ont été très enrichissantes et m'ont permis de m'ouvrir au monde de l'intelligence artificielle qui était jusque-là plutôt obscur. Hé cela m'a permis de comprendre les différents technique appliquer à ce domaine et de pratiquer l'une d'elle les LLM

hé, mais j'ai également pu développer des compétences moins technique blabla

hé hé cette expérience aura donc certainement une influence très positive sur mon futur professionnel… h

Et très positive j'ai permis de très développer mon autonomie dans la gestion de projet et dans son aboutissement, mais bon j'aurais aimé quand même avoir puis pouvoir échanger qualifié dans mon domaine de stage avec afin d'obtenir des conseils/guider techniques afin d'accélérer ma réalisation et d'augmenter on apprentissage.

Ces 2 semaines de stage n'ont pas changé mon projet professionnel, mais m'a permis comme je le souhaitais avant d'arriver de me familiariser avec un nouveau domaine et tester des technologies de celui-ci.

# Table des annexes

A - Organigramme page : 19

Source : Intranet de l’entreprise

**Exploration des grands modèles de langages pour réaliser des prédiction et détection d'anomalies dans des séries temporelles**

**Résumé :**

Durant 9 semaines j'ai évolué au sein d'une entreprise d'intégrateur industriel qui ont une seconde  activité dans le développement de logiciel.  J'ai participer au prémisse de  leur projet de développement d'un outils de prédiction et détection d'anomalies dans des série temporelle.

Ma mission a été de tester et d’évalué le potentielle de la technologie LLM pour réaliser cette outil. Cela ma permit de voir les différente techno de LLM disponible en open source. Comprendre et réaliser un apprentissage sur un LLM afin de l’adapter à un tache particulière, sur un jeu de donnée particulier. Puit évaluer et comparer les performance de ces LLM.

Cet outil sera à destination des industriels qui souhaitent avoir de prédictions sur le valeur de retours de leurs systèmes en fonction des covaroble qui peuvent l'influencer.

**Mots-clés :**

3 à 8 mots-clés

LLM :

Série temporelle

**Abstract :**

*5 to 15 lines to provide an overview of your internship work*

Keywords :

3 to 8 keywords