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](#_Toc174348261)

[Table des illustrations 1](#_Toc174348262)

[Introduction 2](#_Toc174348263)

[1 Environnement 3](#_Toc174348264)

[1.1 Structurel 3](#_Toc174348265)

[1.2 Relationnel 4](#_Toc174348266)

[1.3 Contextuel 5](#_Toc174348267)

[2 Synthèse du travail effectué 6](#_Toc174348268)

[2.1 Recherche sur le Domaine 6](#_Toc174348269)

[2.1.1 Familiarisation avec le Domaine 6](#_Toc174348270)

[2.1.2 Fonctionnent des LLM 7](#_Toc174348271)

[2.1.3 Les techniques pour adapter un LLM 8](#_Toc174348272)

[2.2 Réalisation 9](#_Toc174348273)

[2.2.1 Faire un apprentissage : 9](#_Toc174348274)

[2.2.2 Création des datasheets 9](#_Toc174348275)

[2.2.3 Évaluation des performances : 9](#_Toc174348276)

[3 Analyse réflexive de l’expérience 10](#_Toc174348277)

[3.1 Mon organisation 10](#_Toc174348278)

[3.2 Comparatif prévision/réalisation effective 13](#_Toc174348279)

[3.3 Les compétences acquises et consolidées 14](#_Toc174348280)

[Bilan de l’expérience 15](#_Toc174348281)

[Table des annexes 16](#_Toc174348282)

# 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.

Annonce Plant

# 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 ». Elle est également 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 ces 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 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 pour des directs, c'est-à-dire qu'ils travaillent chez eux sur la supervision et le développement de l'automatisation de leur système. On peut citer par exemple l’entreprise SKF ou STMicroelectronics.

## Relationnel

Durant mon stage j'ai majoritairement interagit 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 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 période de stage. Nous avons pu échanger lors de moments plus ou moins formels à propos de mon sujet de stage et de leur métier. (Il mon permis ...)

## 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 de prédiction et détection sur des séries temporelles afin de voir s'il y a un intérêt aa l’utiliser pour le développement de cet 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

La prédiction sur des valeurs à partir de série temporelle n’est pas un sujet nouveau si les rossant avance en intelligence artificielle ont permis des progrès notamment en pouvant prendre en compte un nombre élargie des facteurs qui pouvaient peuvent influencer le comportement d'une série temporaire. Les premiers travaux de recherche à ce sujet ont été initiés dans le domaine de la statistique et de l'analyse de données.

### 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 ». Ce type de modèle a était 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. Contrairement aux précédentes technologies de réseau de neurone qui utiliser dés réseaux convolutife, … .

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

ce qui donne comme avantage …

Pour résumer plus simplement voici les différentes étapes qu'il se passe lorsque l'on fait une requête dans un LMM :

1. Le tesxe et découper en petite partie appeler token
2. C’est toquenne son convertien en vecteur de

### 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 :**

## Réalisation

Pour réaliser ces apprentissages j'ai donc choisi de cumuler 2 technique à la fois l’apprentissage fin et l'optimisation de prompte. J’ai privilégier l’apprentissage fin car ici je souhaiter que l’IA développe un nouveaux skillse (la prediction de valeur) et nom pas qu’il soit capable dallée cherche les donnée qu’il souait prédire dans un basse de donnée extérieur.

Je realiserai cette aprentisag en crees des datacete en des coupent un serire de donnée en plusire putie bout et lui demendait de prédire céest bout. Jexplique plus en détaille la création de c’est dataset dans la soupartie « … »

### Faire un apprentissage

Pour réaliser mes différents apprentissages j'ai choisi d'utiliser environnement Hugging Face, hé on est une start-up franco-américaine qui permet de partager des modèles via open source. ….

C'est donc utiliser l'API auto-train pour réaliser mes apprentissages :

Cette API pays permet différents types d'apprentissage :

Générique :

SFT :

ORPO :

DPO :

Reward :

### Création des datasheets

### Évaluation des performances

Hé hé afin d'évaluer les performances des prédictions j'ai comme nous l'avons vu dans l'état précédente à chaque apprentissage gardez une partie des datasets comme base de validation et donc pour ne pas avoir hé le problème de… cette base de validation n'était pas utilis dans le processus d'apprentissage. Hé donc j'ai donc créé un script python qui compare Hey de série temporel lune la série temporelle réelle et l'autre générée par h le LM.

Pour réaliser cette comparaison je crée un graphique j'ai affiché les 2 courbes afin de pouvoir de manière visuelle rapidement R si les prévisions sont plutôt bonnes ou mauvaises. Hé je calculais également différentes métriques. Je récupérer la distance quadratique moyenne. La plus grande distance quadratique est la plus petite. Je calculais ces métriques 2 fois une fois sur les séries temporelles normales et la 2nde sur les séries temporelles mais en les ayant normalisées.

Je veux calculer également une métrique appelée hé performance de la prédiction. Hé cette valeur est un pourcentage hé 100% correspond à une prédiction exacte et le seuil de 0 pour-cent était atteint lorsque la prédiction 2 fois supérieur à la valeur réelle.

Résulta :

Évaluation des performances des différente prédiction est détection réaliser pendant le stage :

Premier aisser : Dans un premier temp j’ai effectué des test avec les version gratuit disponible en ligne des LLM Claude, Mistral, GPT3

Résultat du premier apprentissage

Résultat premier apprentissage avec le modèle mistral7B :

# 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 :

# 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