

INSTITUT NATIONAL POLYTECHNIQUE



FLYBOT



---

## Projet Long - Chatbot Tourisme

### Méthodologie

---

*Groupe 5 :*

Amine CHARIFI  
Othmane EL FETNASSI  
Youssef BENDAGHA  
M'hand GLILAH  
Marouane LABYAD  
Ismail MOUSSAOUI

*Clients :*

Florian GARIBAL (*Flybot*)  
Sandrine MOUYSET (*IRIT*)

*Superviseur industriel :*

Younés BENSALLAM (*Airbus*)

# SOMMAIRE

<b>1</b>	<b>Rappel des Objectifs du projet</b>	<b>1</b>
<b>2</b>	<b>Déroulement global du projet</b>	<b>1</b>
<b>3</b>	<b>Phases du projet et Planning Prévisionnel</b>	<b>2</b>
<b>4</b>	<b>Conditions de réalisation et Détails des livrables</b>	<b>5</b>
4.1	Fournitures du client . . . . .	5
4.2	Lieux d'exécution des travaux . . . . .	5
4.3	Confidentialité . . . . .	5
4.4	Livrables . . . . .	5

# 1 Rappel des Objectifs du projet

L'objectif principal de ce projet est l'entraînement et la validation d'un chatbot dans le domaine du tourisme à partir d'un outil en NLP/NLU performant déjà existant pour la startup Flybot. Le chatbot devra produire des résultats similaires ou meilleures que ceux obtenus avec l'outil existant utilisé actuellement par le client. Une API REST sera ensuite mise en place pour permettre l'échange avec le chatbot développé et soumettre des phrases à celui-ci et en récupérer l'interprétation.

# 2 Déroulement global du projet

Afin de gérer et réaliser au mieux ce projet, notre équipe s'engage à respecter scrupuleusement un schéma de travail par étapes.

**Un chef de projet** a été désigné au préalable et sera l'interlocuteur privilégié du client. Il rédigera tous les documents nécessaires à la bonne réalisation du projet et fera l'intermédiaire entre le client et les membres de l'équipe en dehors des réunions où toute l'équipe est présente.

L'équipe sera accompagnée tout au long de ce projet par un **superviseur industriel** qui, de part son expérience dans le monde professionnel, jouera le rôle de conseiller en gestion de projet. Le superviseur industriel guidera l'équipe dans les choix des méthodes et outils de gestion et suivi de projet et partagera avec celle-ci les best practices, normes du monde professionnel.

Le projet sera réalisé en plusieurs phases, détaillées ci-dessous, et qui ont été obtenues après la réalisation du cahier des charges. Ce projet s'inscrivant principalement dans un cadre **d'intelligence artificielle/Machine Learning**, l'application des méthodes agiles *classiques* n'est pas parfaitement adaptée. Cependant, le développement itératif et incrémental de la solution sera mis en place, et les phases du projet peuvent ainsi être regroupées en deux grandes étapes : une première étape de recherche documentaire et d'études comparatives des outils existants et des possibilités d'utilisation de ceux-ci dans notre cas, et une seconde étape **d'entraînements et optimisation itératifs** du chatbot développé.

Pour assurer une bonne progression vers les performances souhaitées, à chaque itération **un responsable qualité** et **un responsable tests** seront désignés parmi les membres de l'équipe. Le responsable qualité proposera des métriques d'évaluation, adaptées à chaque étape, ainsi qu'un seuil de validation. Le respon-

sable tests sera quant à lui amené à établir un plan de tests et exécuter celui-ci, afin de mesurer les métriques proposées par le responsable qualité.

Tout au long de l'étude, le chef de projet informera le client de l'avancement de l'étude. Lorsque cela est utile, des réunions seront organisées entre le client et tous les membres de l'équipe.

### **3 Phases du projet et Planning Prévisionnel**

#### **Phase 1 - Choix d'un outil NLP/NLU existant**

- Dresser un état de l'art sur les outils existant en NLP/NLU (extraction et classification d'entité)
- Réaliser une étude comparative des meilleurs outils
- Réaliser une synthèse de l'étude comparative et choix d'une solution/un outil à retenir

Charge de travail estimée : 15 jours-hommes

#### **Phase 2 - Prise en main de l'outil retenu**

- Réaliser une documentation concernant l'outil retenu : des exemples, tutoriels d'utilisation, ressources nécessaires (bibliothèques, puissance de calcul), spécification sur le format des données d'entraînement/test
- Préparation de l'environnement de travail

Charge de travail estimée : 10 jours-hommes

#### **Phase 3 - Tests et analyse des performances de l'outil existant (wit.ai)**

- Prise en main de l'outil existant : tests manuels, analyse des anciens logs
- Définition des éléments à retenir dans la nouvelle solution
- Mesure des performances selon des métriques basiques (*accuracy*, *recall*, *F-score*)

Charge de travail estimée : 10 jours-hommes

#### **Phase 4 - Prétraitement des données fournies**

- Etude exploratoire du fichier logs
- Etude de la distribution des données textuelles, élimination des données erronées et séparation en données train/test/validation

- Prétraitement du jeu de données

Charge de travail estimée : 20 jours-hommes

#### **Phase 5 - Définition des métriques d'évaluation**

- Réaliser une recherche et une synthèse sur les métriques utilisées dans l'industrie pour l'évaluation des chatbots
- Définition d'autres indicateurs utiles

- Analyser pour chaque métrique l'accessibilité, c'est-à-dire la facilité de mesure

- Dresser une synthèse contenant le choix des métriques retenues pour l'évaluation des performances du chatbot développé

Charge de travail estimée : 12 jours-hommes

#### **Phase 6 - Implémentation de l'API REST (1)**

- Implémentation (sans chatbot) de l'API REST

Charge de travail estimée : 1 jour-homme

#### **Phase 7 - Premier entraînement et établissement d'un benchmark**

- Premier Entraînement du chatbot avec le fichier logs (après prétraitement) comme base d'entraînement

- Avec les métriques retenues, établissement d'un benchmark de performances. Les itérations suivantes auront pour objectif d'améliorer ces performances par le biais d'optimisation de paramètres/hyperparamètres incrémentale

Charge de travail estimée : 20 jours-hommes

### **Phase 8 - Itérations successives**

Après l'établissement du benchmark (valeur de performance de départ), et à chaque nouvelles itération :

- Définition des paramètres à modifier et des pistes à explorer (data augmentation, data cleaning, modification de la durée d'entraînement, etc)
- Implémentation des modification et nouvel entraînement du chatbot
- Analyse des résultats et validation ou rejet des modifications apportées
- Etablissement d'un nouveau plan de développement/test/validation pour la prochaine itération

Charge de travail estimée : 60 jours-hommes

### **Phase 9 - Implémentation de l'API REST (2)**

- Intégration du chatbot développé
- Tests de l'API (requête - retour obtenu)
- Optionnel : implémentation d'un module statistique pour collecter des informations sur les entités

Charge de travail estimée : 5/10 jours-hommes

### **Phase 10 - Rédaction de la documentation**

- Rédaction d'une documentation complète : manuel utilisateur, documentation des codes sources (signatures des modules/fonctions implémentées)

Charge de travail estimée : 25 jours-hommes

Au total, **183 jours-hommes** (répartis sur les 6 membres de l'équipes), auxquels se rajoutent les travaux de rédactions de rapports académiques, les différentes réunions de travail et la préparation des présentations finales (pour un total de **240 jours-hommes**).

## 4 Conditions de réalisation et Détails des livrables

### 4.1 Fournitures du client

Le client accepte de fournir à l'équipe de travail, au début de l'étude ou au besoin :

- Le fichier `logs` contenant les échanges entre les utilisateurs de Flybot et le chatbot utilisé jusqu'à présent. Ce fichier servira de base d'entraînement pour l'outil développé ;
- Un fichier supplémentaire de données structurées généré par le client, pour l'entraînement et la validation du chatbot développé ;
- L'accès à un ordinateur à hautes performances ;
- Un exemple d'API REST implémentée sous python avec le framework Flask ;

### 4.2 Lieux d'exécution des travaux

Le projet sera réalisé au sein de l'ENSEEIH. Les réunions avec le client peuvent être tenues dans les locaux de celui-ci.

### 4.3 Confidentialité

Les membres de l'équipe s'engagent à garder le secret sur leurs travaux liés à ce projet. En particulier, les données et fichiers partagés par le client seront soumis à une confidentialité absolue.

### 4.4 Livrables

A l'issue du projet, l'équipe remettra au client **tous les codes sources** qui ont servi au prétraitement des données, à l'entraînement du chatbot, à la configuration de celui-ci et à la réalisation de tests unitaires, fonctionnels et intégraux. Les codes sources seront accompagnés d'une **documentation complète** (manuel utilisateur, signatures complètes des modules/fonctions implémentés) permettant au client la prise en main de l'outil développé. En particulier, un fichier **configuration** sera remis et permettra au client de rajouter/modifier facilement, par modification de lignes de code, les entités à détecter par le chatbot développé.