Institut National Polytechnique

Flybot

***Projet Long* - Report**

Flight Booking Chatbot

Group 5 :
Amine Charifi
Othmane El Fetnassi
Youssef Bendagha
M'hand Glilah
Marouane Labyad
Ismail Moussaoui

Clients :
Florian Garibal *(Flybot)*
Sandrine Mouysset *(IRIT)*

Supervisor :
Younés Bensallam *(Airbus)*

Table Of Contents

1 Introduction

The *Projet Long* is an academic and industrial project, a key element of ENSEEIHT's engineering curriculum. It takes place in the beginning of the second semester of the third and last year of the Engineer Curriculum, and lasts between 8 and 10 weeks. It is an ideal opportunity for students to practice both their technical skills and project management skills on solving a real, business-world case for a client before starting their end-of-course internships. This year, the *projet long* took place between the 14th of January and the 8th of March.

Throughout the project, each team will meet their client on several occasions, and will be supervised and coached by an experienced manager. An oral presentation will take place at the end of the project, during which each team will present their work, their methodology and their main takeaways from this experience.

This report summarises our team's work and contains a presentation of our team and the context of our project, a description of the solution we implemented and the performance tests we carried out, a detailed presentation of our project management process and the key takeaways from this experience. Every referenced document can be found in the deliverables file.

2 Presentation of the project

2.1 Context and specifications

Our client is a startup called **Flybot**. Today, Flybot's main vocation is helping travelers make better and faster flight reservations by using a conversational AI, also known as a chatbot. Through engaging users in conversations, using natural language, the chatbot will try to understand the user's preferences and criteria so as to make a customised list of flight suggestions. Hence, the chatbot's main mission is to interpret natural language and extract relevant information and entities.

Up until now, the client uses **Facebook's wit.ai** as their chatbot. The client wishes to internalise this part of their service, and develop their own chatbot to extract and classify relevant information from users' messages. Our team will be in charge of this transformation.

In broad terms, the main goals of the project are :

— Training, testing and validating a flight booking conversational AI (chatbot), which will be built upon an existing open-source solution. The chat-

bot should be **at least as performant as the currently used solution (wit.ai)**.
— Build an API so as to be able to send text requests and receive the outcome produced by the developed chatbot

After two client meetings, these goals were elaborated and clear quantifiable targets established. Our team was then able to write a functional and technical specifications document, which was approved by the client. The said document can be found in the deliverables file under `Cahier_Des_Charges.pdf`.

During the project, our contact at Flybot was Florian Garibal, Flybot's Chief Technical Officer, and Sandrine Mouysset, a researcher at the IRIT laboratory.

2.2 Team members and supervisor

Our team is composed of six ENSEEIHT engineering students, all majoring in a specialization of the computer science and applied mathematics department. The team members and their respective majors are :
— Youssef Bendagha *(Computer Imagery)*
— Amine Charifi *(HPC Big Data)*
— Othmane El Fetnassi *(HPC Big Data)*
— M'hand Glilah *(HPC Big Data)*
— Marouane Labyad *(HPC Big Data)*
— Ismail Moussaoui *(Software)*

Throughout the entire project, our team was supervised and coached by M. Younés Bensallam, a project manager at Airbus.

3 Natural Language Processing : a brief presentation

Natural language processing (NLP) is a subfield of computer science and artificial intelligence focused on enabling computers to understand and process human natural language. In particular, NLP seeks to formalize ways to program computers to process and analyze large amounts of natural language data.

4 Our solution

4.1 Choosing an Open-Source base

As agreed upon with the client, the solution to be implemented by our team was going to be built upon an existing open source tool. Thereby, our first task was conducting desk research on state-of-the-art conversational AIs (chatbots).

After this first step, our team selected **RASA NLU** as a base tool. RASA NLU is an open-source natural language processing tool for intent classification and entity extraction in chatbots. As backend, RASA NLU uses **SpaCy**, an NLP library, as well as other famous machine learning libraries such as **TensorFlow** and **Scikit-Learn**.

RASA NLU is widely used in the industry by small and big companies alike, it is open-source and thus continuously and publicly updated and improved and it is as powerful as most mainstream proprietary tools, according to several comparative studies. These are the main arguments that helped our team make this decision.

4.2 Defining clear targets

To ensure client satisfaction and help guide our team's efforts, defining clear targets was imperative. Thus, it was important to define quantifiable metrics and agree on thresholds to reach.

After listing the metrics that are used the most in the industry to evaluate chatbot performance, we narrowed down on three metrics, using the **actionability** and **accessibility** criteria. Actionable metrics are indicators which contain enough clear information so as to be able to make decisions. Accessible metrics are indicators that are easily and accurately measurable by our team. This work is synthesized in the document entitled `Synthèse: métriques d'évaluation d'un chatbot`, which can be found in the deliverables file.

The three metrics that we agreed on are : **precision, recall and f1-score**. These metrics will be measured on a dataset that is relevant with regards to the flight booking industry.
As for the threshold to reach, the developed solution should have **similar or better performance, as measured by these three metrics on the same dataset and with the same computing power, to that of the currently used solution**.

4.3 Data

The client agreed to share with the team the logs file containing all the messages received up to this point by Flybot, and the corresponding output produced by wit.ai. The file contained roughly 240 000 *annotated* messages, which means a message and its corresponding extraction and classification of intents and entities.

Our team carried out a first exploratory study of this file to assess its quality and whether or not it would be a suitable training base for our model. The details of this study and its results can be found in the document : `Synthèse: étude exploratoire du fichier logs`.

Our team reached the conclusion that this dataset is not enough to train the model to reach the targeted performance. In fact, less than 2 % of the logs file, about 3800 annotated utterances, can be used in the training of our chatbot. All other messages are either too short (a word or a single character) or incorrectly annotated.

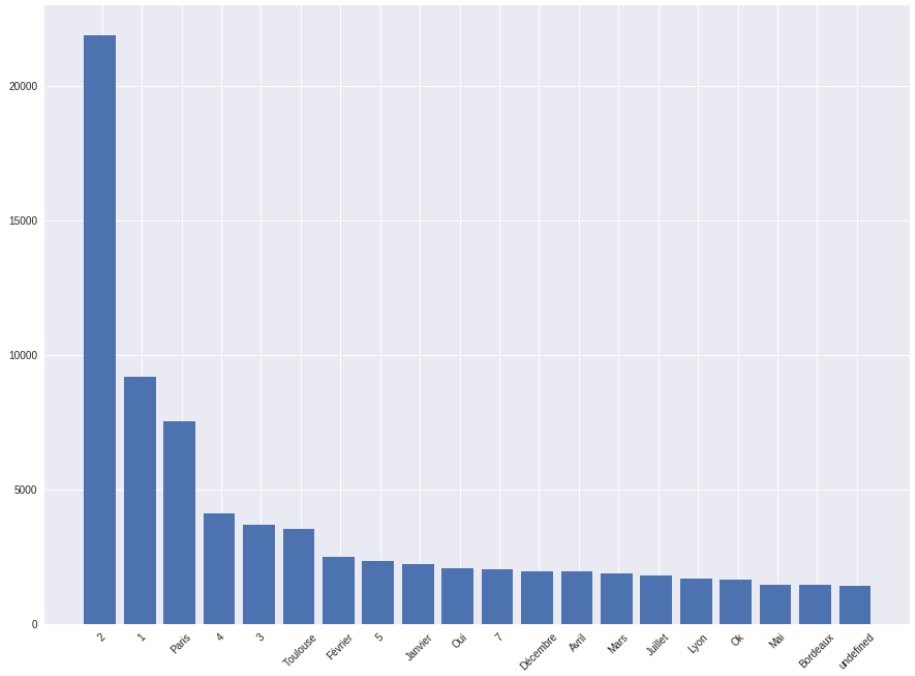


FIGURE 1 – The messages that were most received by the chatbot up to this point are too short (a single word) and are thus not suitable for training

In addition to the examples from the logs file, our team worked on generating

more training data using an algorithm that generates sentences and their corresponding annotations from a list of words or expressions. The used program is included with the deliverables.

4.4 Training iterations

Several training iterations took place once we had our training data ready and preprocessed and our model ready to be trained. The first iteration was run with a basic model, and enabled our team to establish a performance benchmark. The goal of the iterations that followed was to make specific modifications and improve the performance of the chatbot from the baseline towards the targets. Each iteration can be described as a loop, which was at the core of our project. The following figure summarises our training loop.

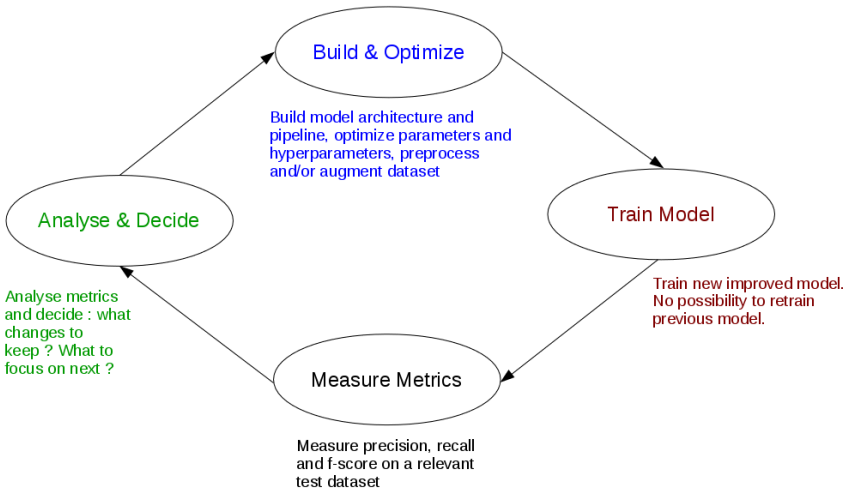


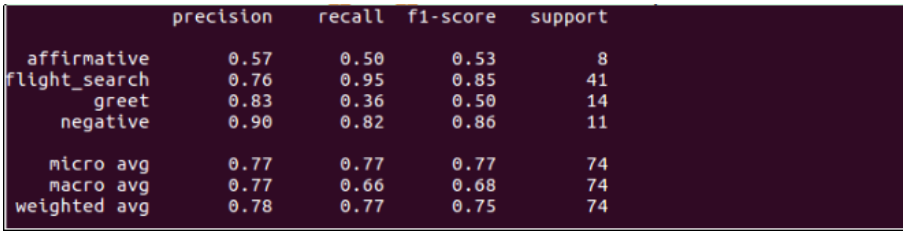
FIGURE 2 – Steps of a training iteration

After building our model or optimising and tuning the previous one, the model is trained again and the evaluation metrics measured. Then, our team would then analyse these values and decide what modifications to keep and what modifications to add for the next iteration.

4.5 Performance evaluation

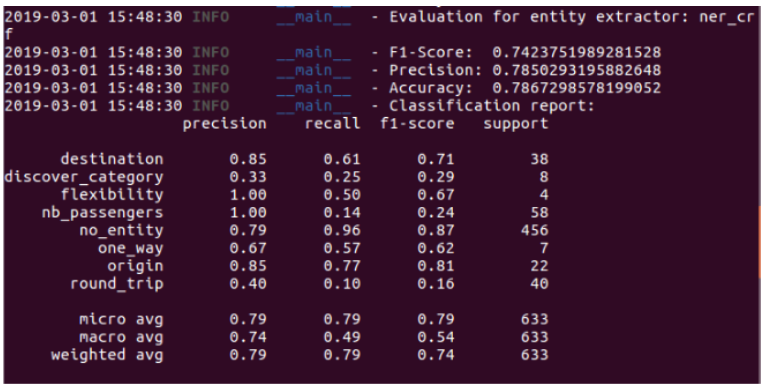
Each week, a quality manager from the team assessed the progress we were making towards our targets. This was mainly done through measuring the three metrics we chose to evaluate our chatbot. The role of the quality manager and the quality process we developed are detailed in the next section (Project Management).

The following is a summary of our final chatbot's performance, measured on a test dataset that is relevant with respect to the flight reservation industry.



	precision	recall	f1-score	support
affirmative	0.57	0.50	0.53	8
flight_search	0.76	0.95	0.85	41
greet	0.83	0.36	0.50	14
negative	0.90	0.82	0.86	11
micro avg	0.77	0.77	0.77	74
macro avg	0.77	0.66	0.68	74
weighted avg	0.78	0.77	0.75	74

FIGURE 3 – A summary of our chatbot's performance in intent classification. The chatbot does best with the *flight search* intent, the one we are most interested in.



```
2019-03-01 15:48:30 INFO     __main__   - Evaluation for entity extractor: ner_cr
f
2019-03-01 15:48:30 INFO     __main__   - F1-Score:  0.7423751989281528
2019-03-01 15:48:30 INFO     __main__   - Precision: 0.7850293195882648
2019-03-01 15:48:30 INFO     __main__   - Accuracy:  0.7867298578199052
2019-03-01 15:48:30 INFO     __main__   - Classification report:
                  precision    recall  f1-score   support

      destination       0.85      0.61      0.71        38
discover_category       0.33      0.25      0.29         8
      flexibility       1.00      0.50      0.67         4
    nb_passengers       1.00      0.14      0.24        58
        no_entity       0.79      0.96      0.87       456
          one_way       0.67      0.57      0.62         7
           origin       0.85      0.77      0.81        22
       round_trip       0.40      0.10      0.16        40

        micro avg       0.79      0.79      0.79       633
        macro avg       0.74      0.49      0.54       633
     weighted avg       0.79      0.79      0.74       633
```

FIGURE 4 – A summary of our chatbot's performance in extracting and classifying entities

Overall, these results were satisfying. The performance of the chatbot exceeds the established targets with some intents and entities, and further training with more data can lead to similar outcomes with the other intents and entities.

4.6 API integration

An API was developed and the final chatbot integrated to it. Users with a token can send a text request, which is a sentence, to the API and get the chatbot outcome : the intent of the message, the extracted entities and their classifications. The API was also linked to a database so as to enable tracking message sources in a way that is completely compliant with the GDPR. This was done in order to group messages so as to form full conversations and can also be very useful when conducting statistical studies.

5 Project Management

5.1 Methodology

In order to reach the targets established with the client, our team elaborated a meticulous work plan which we stuck to all along the project.

The project was executed in several phases, carefully defined after the elaboration of the specifications document. Theses phases can be categorized in two main steps : a first step of desk research, comparative studies and data generation and preprocessing, and a second step which consists of one single task, that of elaborating, training and testing a conversational AI until reaching the performance targets.

Because our project is mainly a machine/deep learning project, applying classical lean methods throughout all the phases of the project was not perfectly suitable. For phases of the first step, our team applied **the Scrum framework**. Sprints lasted a week, and daily sprint meetings of less than fifteen minutes were organised. For the second phase, our team agreed on a **highly iterative process to train and validate our chatbot**. The process consists of building and training a model, measuring its performance and then deciding as to what the next improvements should focus on. This process was repeated several times until reaching our performance targets. Thus, each iteration was the same as the previous one in terms of the tasks to be carried out.

The different phases of our project and their respective cost estimation in man-hours are described in great detail in the Methodologie.pdf document, which can be found in the deliverables file.

Work progress and task allocation were closely monitored by using a **Planning And Progress Dashboard**, which can be found in the deliverables file and which will be discussed in-depth in the next section.

A project manager was designated amongst the team members prior to the launch of the project, and was in charge of updating the client regularly on work progress and making sure the team was making steady progress towards the targets. Moreover, each week a quality manager, a member of the team, was in charge of defining key performance and development indicators, to be measured along with the technical metrics, to quantify the said progress. The team organisation and the different roles are described in detail in the *Team Organisation* section of this report.

Except on very rare occasions, team members met on a daily basis at ENSEEIHT, where they carried out most of their tasks.

5.2 Tools

Communication
Throughout the project, our team communicated through instant messaging platforms such as **Messenger** and **Slack**. These platforms enabled efficient, less formal and more fun team conversations.

File Sharing
To centralize our work and gain in efficiency, we used **Google Drive** to write, edit and store all input and output documents of this project. Our work directory was made of five separate main folders :

- **Input** : Guides and deadlines for the project
- **Output** : All output documents from the project, including the specifications document, the methodology document and all intermediary reports
- **Tools** : Contains the **Planning and Progress** dashboard and **the risk management** sheet
- **Meetings** : The minutes of all meetings with client or with the supervisor
- **Resources** : Academic papers and useful links

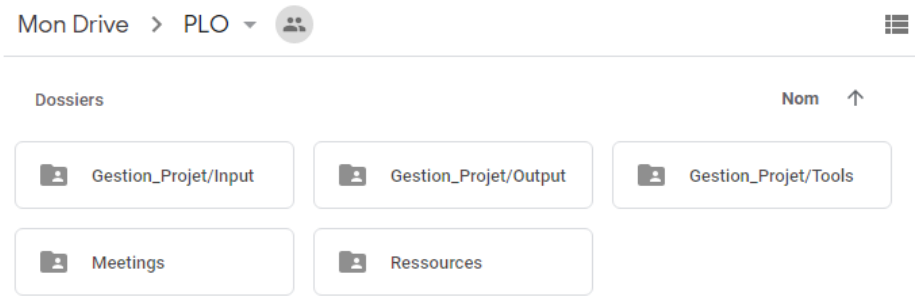
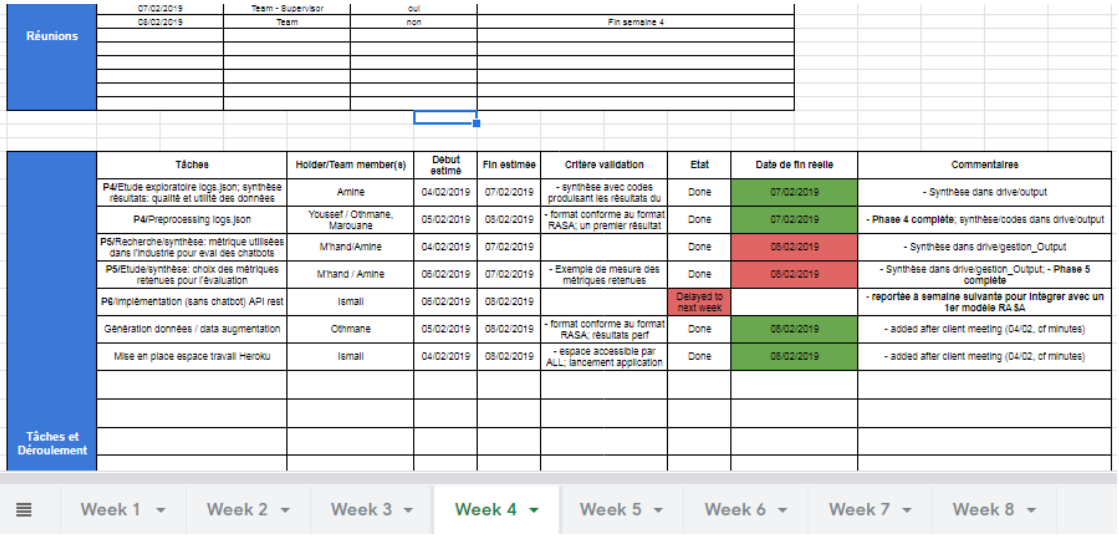


FIGURE 5 – Work directory in Drive

For code files, our team used **Git**, the most widely used code version software. Using Git enabled the team to work on several versions at the same time, easily detect and fix bugs and keep a record of all previous work.

Planning and Progress Dashboard

Our team custom made a planning and progress dashboard. Our dashboard was made of 8 separate sheets, each corresponding to a week of the project. For each week, the team organisation (Project manager, Quality manager and Test engineer for the week) and the week's meetings were displayed, as well as a task board for the week. The task board contained **the name and description of the task, its start and estimated finish date, its actual finish date, a status tracking column and a commentary column**. Each task had a holder, a team member in charge of the task. For most of the tasks, the holder worked with other team members. The figure below is taken from the said board.



Réunions				
	07/02/2019	Team - Supervisor	oui	
	08/02/2019	Team	non	Fin semaine 4

Tâches et Déroulement	Tâches	Holder/Team member(s)	Début estimé	Fin estimée	Critère validation	Etat	Date de fin réelle	Commentaires
	P4/Etude exploratoire logs.json; synthèse résultats: qualité et utilité des données	Amine	04/02/2019	07/02/2019	- synthèse avec codes produisant les résultats du	Done	07/02/2019	- Synthèse dans drive/output
	P4/Preprocessing logs.json	Youssef / Othmane, Marouane	05/02/2019	08/02/2019	- format conforme au format RASA; un premier résultat	Done	07/02/2019	- Phase 4 complète; synthèse/codes dans drive/output
	P5/Recherche/synthèse: métrique utilisées dans l'industrie pour eval des chatbots	M'hand/Amine	04/02/2019	07/02/2019		Done	08/02/2019	- Synthèse dans drive/gestion_Output
	P5/Etude/synthèse: choix des métriques retenues pour l'évaluation	M'hand / Amine	06/02/2019	07/02/2019	- Exemple de mesure des métriques retenues	Done	08/02/2019	- Synthèse dans drive/gestion_Output; - Phase 5 complète
	P6/Implémentation (sans chatbot) API rest	Ismail	06/02/2019	08/02/2019		Delayed to next week		- reportée à semaine suivante pour intégrer avec un 1er modèle RASA
	Génération données / data augmentation	Othmane	05/02/2019	08/02/2019	- format conforme au format RASA; résultats perf	Done	08/02/2019	- added after client meeting (04/02, cf minutes)
	Mise en place espace travail Heroku	Ismail	04/02/2019	08/02/2019	- espace accessible par ALL; lancement application	Done	08/02/2019	- added after client meeting (04/02, cf minutes)

FIGURE 6 – Week 4 of our Planning and Progress Dashboard

This tool enabled us to have a clear overview of our progress, of whether or not clear progress was being made and deadlines met. It was also very useful in task allocation as it helped allocate tasks in a balanced way between team members. Moreover, it served as an action tracker, and another mean of communication between team members : the *Commentary* column contained updates regarding the tasks and actions to be carried next. Last but not least, this dashboard gave us a great overview of the progress we've made in terms of project management, between the start and the end of the project. In fact, the number of delays decreased significantly throughout the project, deadlines were met without any problems, and our team worked more efficiently. **This dashboard can be found in the deliverables file, under planning.xls**

.

5.3 Team organisation

Prior to the beginning of the project, a **Project Manager** was designated. The project manager's main role was to ensure a steady overall progression towards achieving all the goals of the project. To that end, the project manager was in charge of estimating the cost of the project in terms of man-hours, of organizing the project in several phases and allocating tasks. Along with the other team members, the project manager elaborated a **Project Management Process**, to guide the team's work throughout the project. The project manager was also the client's contact, regularly updated him on the progress made and organised weekly meetings with the client and with the supervisor.

The other five members of the team held several roles during this project. Each week, a **Quality Manager** and a **Test Engineer** were designated amongst the team. The remaining members formed the **Developoment team**.
The **Quality Manager** was in charge of establishing a list of relevant key performance and development indicators to be measured that week, along with the technical metrics, to assess that week's work and quantify the progress that was made.
The **Test Engineer** was in charge of elaborating a test plan, which execution would enable to collect and measure the indicators set by the quality manager.
The **Development Team** was in charge of the implementation and test of solutions, mostly algorithms and machine learning programs.

This organisation enabled each team member to have different roles and responsibilities during the project and proved to be very efficient.

5.4 Meetings

Client Meetings
During this project, we had six meetings with the client. The first meeting was dedicated to understanding the client needs and formulating clear targets and goals to achieve. During the second meeting, the specifications document and methodology document were presented, examined and approved by the client. The remaining meetings were dedicated to briefing the client on our progress and presenting demos. During the last meeting, a final demo was presented, and all deliverables handed over to the client.

Meetings with the supervisor
During this project, we had eight meetings with the supervisor. During the first two meetings, the supervisor coached the team in establishing a project management process and layed out several templates and frameworks that the team could use. During the following meetings, the supervisor assessed our work in terms of project management, shared his feedback with the team, answered any questions related to project management and gave us leads to explore. During the last meeting, our final demo was presented to the supervisor and our presentation rehearsed.

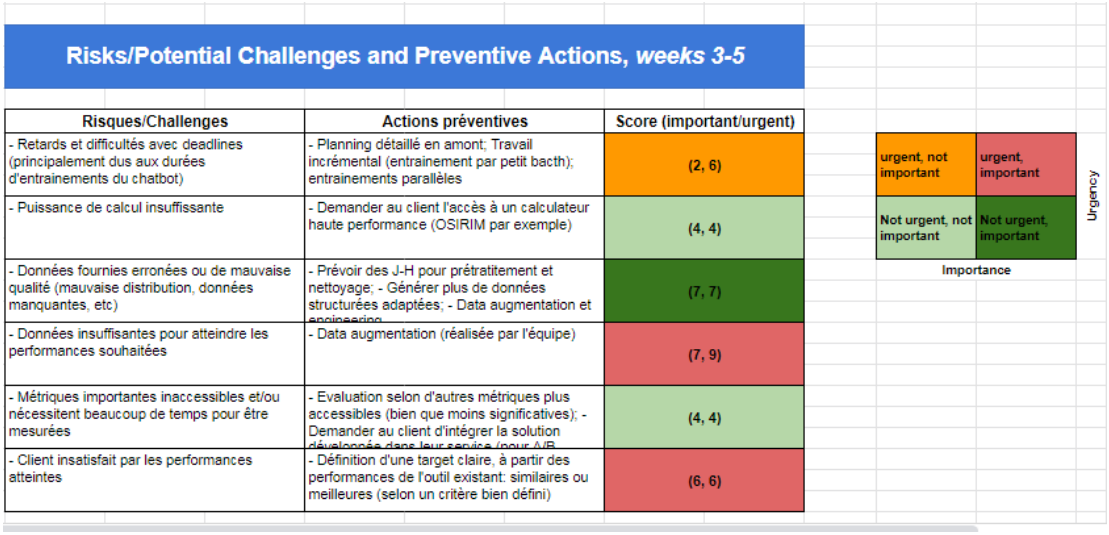
These meetings' minutes can be found in the deliverables file.

Team Meetings
Our team had two formal meetings each week. During the first meeting, which usually took place on Monday mornings, the week's planning was reviewed, using

the planning and progress dashboard, and last modifications added. The second meeting, which usually took place on Friday afternoons, was dedicated to updating the progress dashboard, sharing feedback and takeaways between the team members and deciding as to what improvements can be made to our work process. In addition to these formal meetings, the team met daily for 15 minutes to share feedback and update all team members on the overall progress of the project.

5.5 Risk management

After establishing the specifications document with the client and organizing the project in different phases, our team carried out an analysis of potential risks to the good execution of our project. The first step was to list all potential risks. Afterwards, for each risk, a list of preventive actions was made, as well as a list of corrective actions. The preventive actions were taken into account when writing the methodology document, and most of them were implement in the first weeks of the project.
After this first step, our team needed to continuously monitor these risks, and order them by priority. To that end, we used the **important/urgent** matrix to give a score to each risk. What our team plotted on the important/urgent matrix is not the risk itself, but the actions which our team would need to take in case the risk materializes.
This scoring and continuous monitoring was done through a sheet which can be found in the deliverables file under Risk_Management.xls. The following figure is the risk management board for weeks 3 to 5.



Risks/Potential Challenges and Preventive Actions, *weeks 3-5*

Risques/Challenges	Actions préventives	Score (important/urgent)
- Retards et difficultés avec deadlines (principalement dus aux durées d'entrainements du chatbot)	- Planning détaillé en amont; Travail incrémental (entrainement par petit bacth); entrainements parallèles	(2, 6)
- Puissance de calcul insuffissante	- Demander au client l'accès à un calculateur haute performance (OSIRIM par exemple)	(4, 4)
- Données fournies erronées ou de mauvaise qualité (mauvaise distribution, données manquantes, etc)	- Prévoir des J-H pour prétratitement et nettoyage; - Générer plus de données structurées adaptées; - Data augmentation et engineering	(7, 7)
- Données insuffisantes pour atteindre les performances souhaitées	- Data augmentation (réalisée par l'équipe)	(7, 9)
- Métriques importantes inaccessibles et/ou nécessitent beaucoup de temps pour être mesurées	- Evaluation selon d'autres métriques plus accessibles (bien que moins significatives); - Demander au client d'intégrer la solution développée dans leur service (pour A/B	(4, 4)
- Client insatisfait par les performances atteintes	- Définition d'une target claire, à partir des performances de l'outil existant: similaires ou meilleures (selon un critère bien défini)	(6, 6)

FIGURE 7 – Risk management dashboard for weeks 3 to 5

6 Key takeaways

This project was an ideal opportunity for our team members, as well as other students, to practice both our technical skills and project management skills on solving a real, business-world case for a client before starting our end-of-course internships.

This experience also illustrated once again the importance of a good project management process to successfully execute a project, no matter what field it is related to. Moreover, it highlighted how building a project management process is not a static task that should be carried out once before the project. Rather, it is a continuous and highly iterative process that lasts throughout the project : it is crucial to assess what works well and where there is room for improvement. Project management is thereby a continuous learning process, and a dynamic system.

Every single team member enjoyed this experience, from start to finish. The project was very interesting from a technological point of view, the client was very understanding and encouraging, and the supervisor was always available for coaching and guidance. This is yet another proof that, when we love what we do and are passionate about it, and when there is a great work culture and team spirit, success follows naturally.