Deliverable N. 4.3

AI model performance report

Authors:
Narek Minaskan (DFKI)

Abstract:
This deliverable is a report on activities carried out and results achieved in Task 4.5: "AI engine development, training and testing". It provides a report on the selection of AI models, their architecture, data processing, and evaluation for the Use Cases.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Information Table

Deliverable Number	4.3
Deliverable Title	AI model performance model
Version	1.0
Status	Final
Responsible Partner	DFKI
Contributors	ENAC, TAVS, EMBRAER, LiU, LFV, Skyway, DBL, Suite5, ECTL, ENG, CERTH
Contractual Date of Delivery	31.08.2024
Actual Date of Delivery	30.08.2024
Dissemination Level	PU

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Document History

Version	Date	Status	Author	Description
0.1	27.07.2024	Submitted for review	Narek Minaskan	First version
0.2	31.07.2024	In review	Evmorfia Biliri	Reviewed by Suite5
0.2	01.08.2024	In review	Nicola Durante	Reviewed by ENG
0.3	06.08.2024	In review	Paula Olivio	Reviewed by EMBRAER
0.4	23.08.2024	In review	Narek Minaskan	Integrated reviews
0.5	30.08.2024	Final Review	Roberto Venditti	Final Quality Check

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

List of Acronyms

Acronym	Definition
DNN	Deep neural network
ECG	Electrocardiogram
EMG	Orbicularis oculi electromyogram
FPS	Frames per second
GFT	Genetic fuzzy tree
GSR	Galvanic skin response
IPM	Inverse perspective mapping
LSTM	Long short-term memory
mAP	Mean average precision
MLP	Multi-layer perception
MSE	Mean squared Error
NLP	Natural language processing
ODD	Operational design domain
PSR	Pupil size response
R-CNN	Region-Based convolutional network
SCR	Skin conductance response
SSD	Single shot detector
VAL 1	Validation 1 study
VAL 2	Validation 2 study
XGBR	eXtreme Gradient boosting regression

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Table of contents

Introduction	8
1. Use case 1: Flight deck startle response	10
2. Use case 2: Flight deck route planning/replanning	18
3. Use case 3: Urban Air Mobility	34
4. Use Case 4: Digital and Remote Tower	40
5. Use case 5: Airport safety management	45
6. Use case 6: Airport virus spread prevention	47
Conclusion	56
References	57

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Executive Summary

Deliverable 4.3 presents a report of the current results of Task 4.5 “AI engine development, training and report”. First, the introduction presents the overall status of AI development across all the Use Cases. Then, for each Use Case, there is a detailed outlook of the state of the art, requirements analysis, training and data process, model evaluation, and future planning.

This project has received funding by the European Union’s Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Introduction

This document is a report on Task 4.5: AI engine training and testing. Each use case in the project deals with a different aviation segment, and consequently, the Artificial Intelligence (AI) components required and developed for each Use Case are different.

The following table gives an overview of the status of AI across the UCs.

Use Case	**Segment**	**AI Component**	**What does it to**
UC1	Cockpit	Decision fusion classifier. (Classification).	Detects an event and classifies it as surprising/startling.
UC2	Cockpit	Fuzzy logic trees. (Supervised learning)	Maps pilots intentions to KPIs and suggests a route.
UC3	UAM	-	-
UC4	ATM	Prediction (regression) and optimization	Predicts the ETA for upcoming flights. Generates an optimal sequence for inbound-outbound aircraft.
UC5	Airport	-	-
UC6	Airport	Classifier, Tracking, NLP	Classification of air quality. Safest route recommendation in the airport. Chatbot service, for communication.

Table 1: AI status overview across Use Cases.

At present, use cases 3 and 5 do not incorporate an AI component due to challenges related to data quality. Specifically, in use case 3, the absence of a defined human role and a corresponding system has resulted in a lack of necessary data, making it difficult to develop a viable AI solution. Similarly, in use case 5, the available airport data requires extensive time for thorough analysis and filtering, and even then, it may not be sufficient for effective AI training. As a result, for these use cases, the focus has remained on discussing the potential ideas for AI, rather than moving forward with actual development. For the remaining use cases that include an AI component, the focus was placed on several key elements: the context, the state of the art,

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

requirements, and the model and data processing strategies. Where possible, an Evaluation was conducted to assess the effectiveness of the AI solutions. All the details regarding how the AI was trained and tested are in the following paragraphs.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

1. Use case 1: Flight deck startle response

1.1. Context

The FOCUS Intelligent Assistant aims to detect the startle and surprise reaction of pilots to be able to support them in recovering as quickly as possible. It aims to help pilots to regain control of the aircraft and a good situational awareness to be able to stay "ahead of the aircraft".
The AI component must reliably detect startle and surprise of pilots in the cockpit based on physiological inputs. "Reliably" means that the delay between event and detection shall be minimized while maximizing the precision of the classification, in such a time and safety critical scenario.

1.2. State of the art

To define benchmark parameters and methods, we conducted an analysis of existing work in the field. This analysis led to the identification of several publicly available datasets addressing startle detection in pilots:

- Reducing Commercial Aviation Fatalities by [1].
- PsPM-SMD: SCR, EMG, ECG, and respiration measurement in response to auditory startle probes by [2].
- PsPM-PCF2: PSR, SCR, ECG, respiration and startle-eyeblink EMG measurements in a delay fear conditioning task with 4 CS and different reinforcement rates [3].
- PsPM-FR: SCR, ECG and respiration measurements in a delay fear conditioning task with visual CS and electrical US by [4].

The datasets provide physiological data of pilots during startle/surprise situations, however, none of the datasets evaluated the situation in a flight or simulated flight scenario. As for the stimuli used to startle the participants, the data sets by Khemka et al., Ojala et al. and Tzovara et al. [2–4] report using auditory startle probes while Johnson et al. [1] do not specify what stimuli they use.

1.3. Requirements analysis

During our evaluation of existing datasets and the development of the startle and surprise detection component, several challenges were identified. These challenges are anticipated to arise under real-world conditions as well, thereby setting certain requirements for the system to be developed.

One significant challenge is **cross-user classification**. The goal is to develop an algorithm that does not require training on data from each individual user. Instead, the system should be applicable to new users without the need for recalibration. Current approaches struggle with predicting outcomes on unseen data, which could necessitate training sessions for every new user. This would be impractical, especially

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

for pilots, who cannot realistically engage in continuous training and calibration sessions. Therefore, the targeted solution must work across users without requiring specific retraining for each individual.

Another challenge is the **scarcity of data**, particularly in the targeted scenario of flight situations in the cockpit. The available data is also inconsistent in terms of the types and formats of recorded biosignals. Different datasets have been recorded with varying hardware, leading to discrepancies in signal specifications and data streams. There is no standard setup for physiological signals across the existing datasets; each dataset is a custom mix depending on the hardware available at the recording facility.

While this inconsistency poses a challenge for developing a classification algorithm, it is anticipated as a given condition in potential real-world applications. Depending on the outcomes of this project, we will recommend a specific set of physiological signals for the optimal performance of startle detection with the developed system. However, not all airlines and pilots will be able to utilize the full set of sensors due to varying circumstances. Therefore, **adaptation of the startle detection component to different conditions** is a crucial part of this project, which will be addressed through cross-dataset training.

Based on our analysis, the following key requirements were identified for the startle detection component:

- **Detection Accuracy**:
    - The component must detect startle with minimal delay.
    - It must maximize precision in its classification.
- **Adaptability**:
    - The component should function across different users without the need for retraining or calibration (cross-user approach).
    - It must adapt to varying sets of available sensors for classification.
    - It should be capable of utilizing training data from multiple datasets, even when those datasets have different setups and statistical properties.

Given the requirements, we developed a startle detection model inspired by the work of Harrivel et al. [5], who studied the induction and classification of different mental states, including startle/surprise, in 24 commercial aviation pilots within a flight simulation using multiple modalities. While Harrivel et al.'s use of a Late Fusion classifier provided a solid foundation, their approach was not entirely suitable for our application. In contrast to Harrivel et al.'s [5] Late Fusion approach, which fuses modalities during feature extraction, we opted for a **Decision Fusion classifier**. This approach is better suited to our needs, where it is unrealistic to expect every modality to be consistently available to pilots. In the Decision Fusion classifier (Figure 1):

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

- A separate classifier is trained for each modality.
- The final prediction is determined based on the collective predictions from all classifiers.

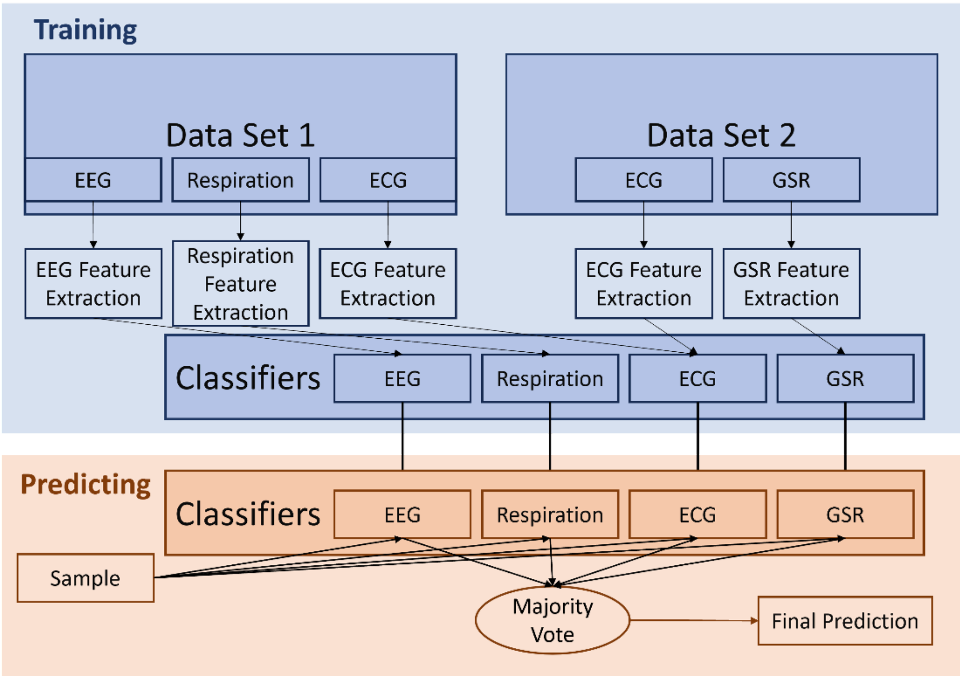


Figure 1: A diagram of an example Training and Prediction Process of the Startle Detection Component. Multiple classifiers are trained on each modality of different data sets. A prediction on a sample is made by obtaining predictions of each classifier with the final prediction chosen by a majority vote.

1.4. Training and data process

The datasets from Khemka et al., Ojala et al., and Tzovara et al. [2–4] were utilized to train the system. Since these datasets contained more non-startle than startle samples, we balanced the number of samples for each class to prevent the classifier from overfitting. Given the use of multiple datasets, a key challenge was normalizing the data across different modalities. Each dataset was separately preprocessed and resampled to 500Hz to maintain uniform sampling frequency.

For feature extraction, we adopted the methodology from Harrivel et al. [5], using statistical features tailored to each modality. Specifically, the following modalities and features were employed:

- **Respiration, ECG, and GSR**:
    - Extracted features included mean, variance, skewness, kurtosis, several quantiles, variation, interquartile range, Shannon entropy, and the area under the curve.
- **GSR (Additional Features)**:
    - We also computed the average slope and drop score.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

A **Random Forest classifier** was trained for each modality using a "warm" training approach, where the training was extended by iterating on the features from different datasets. The final prediction was determined through a Majority Vote across all modalities.

1.5. Evaluation

The developed component was initially trained and tested on publicly available datasets and evaluated in VAL1 using a pseudo-online procedure. This approach was chosen to avoid interference with other evaluation parameters in VAL1, primarily concerning the digital assistant's appearance and functionality. The startle detection was correctly triggered based on the evoked startle in the experiment, allowing for a focused evaluation of the digital assistant, while the startle detection component was evaluated pseudo-online afterward. A real online evaluation of the final component is planned for VAL2.

The offline evaluation was performed using the datasets from Khemka et al. [2] and Tzovara et al. [4], with testing was conducted on data provided by Ojala et al.[3], implementing a cross-user and cross-dataset approach. The classifier followed the same scheme as in the pseudo-online analysis, utilizing GSR, EEG, Respiration, and ECG, since all these data streams were available in the dataset provided by Ojala et al. [3]. The test set from this dataset was balanced (representation of class labels), resulting in:

- **Accuracy**: 70%
- **Precision**: 67%
- **Recall**: 79%

The offline analysis will be further expanded in a master thesis (Autor: Ganavi Basavaraju, Title: Startle Detection in Pilots on the Flight Deck: Utilizing Multiple Source Domain Adaptation with Various Physiological Signals, Supervision: DFKI and ENAC), focusing on cross-user evaluation and multi-dataset evaluation.

The pseudo-online evaluation was conducted using a simulation based on a recorded dataset from **VAL1**, which included data from three subjects at the point of evaluation. The data was segmented into 1-second windows with a 75% overlap. Each window was independently labeled and predicted using the classifier described earlier. The results of this evaluation are illustrated in Figure 2, which shows the confusion matrices for all three subjects evaluated using the classifier. The accuracy for each subject was as follows:

- **Subject 1**: 26% accuracy
- **Subject 2**: 66% accuracy
- **Subject 3**: 61% accuracy

This resulted in a mean accuracy of 51% and a median accuracy of 61%.

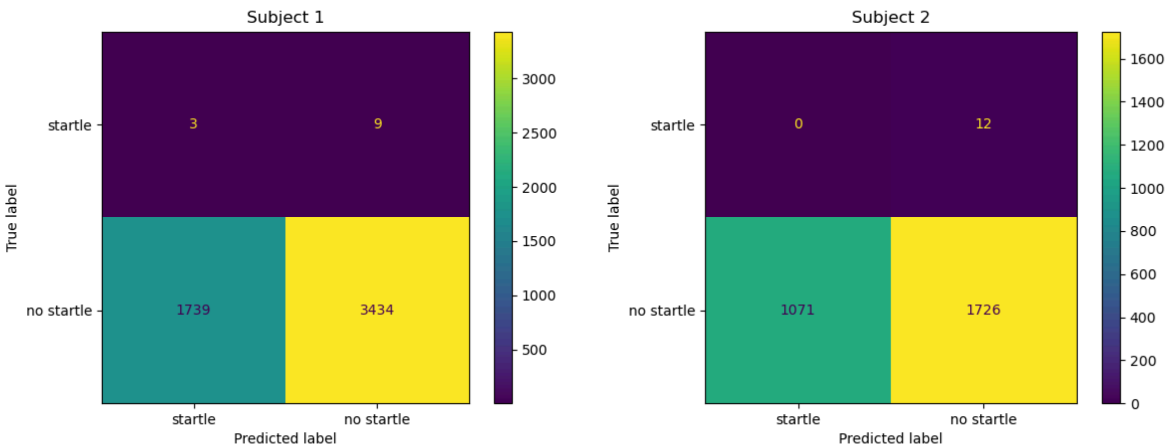
This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Given the imbalance in the dataset—where non-startle samples significantly outnumber startle samples—precision and recall provide a more meaningful measure of the classifier's performance. In this context:

- **True Negatives (TN)** and **False Positives (FP)** refer to non-startle events.
- **True Positives (TP)** and **False Positives (FP)** refer to startle events.

The precision and recall for each subject were as follows:

- **Precision**:
  - Subject 1: 0.002
  - Subject 2: 0
  - Subject 3: 0.003
  - Mean Precision: 0.002
  - Median Precision: 0.002
- **Recall**:
  - Subject 1: 0.5
  - Subject 2: 0
  - Subject 3: 0.6
  - Mean Recall: 0.36
  - Median Recall: 0.5



This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

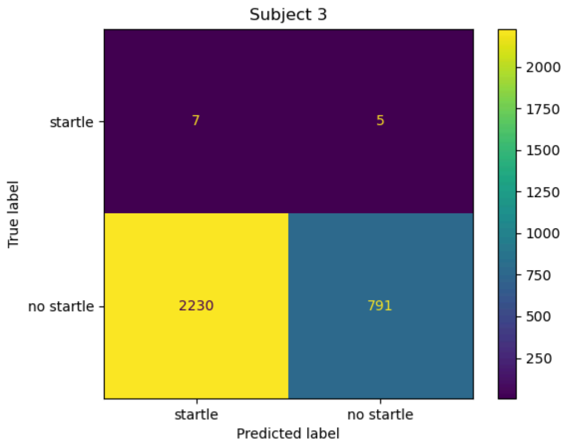


Figure 2: Confusion Matrices for the prediction of startle events in a pseudo-online evaluation on the data of VAL1 for all three subjects.

These results highlight the challenges in detecting startle events due to the imbalance in the dataset. While the accuracy for some subjects was reasonable, the low precision values indicate a high number of false positives, suggesting that the classifier may struggle with distinguishing startle events from non-startle events in this specific setup.

1.6. Discussion around model evaluation

The pseudo-online evaluation provided insights into the performance of the startle detection component. The accuracy, while varying across subjects, highlighted the challenges posed by the imbalanced dataset. The low precision values indicate a significant number of false positives, suggesting that the classifier struggled to differentiate between startle and non-startle events. However, the recall for subjects 1 and 3 was reasonable, demonstrating that the classifier was somewhat effective in detecting true startle events.

The results also showed that the current model is capable of functioning beyond random guessing, as indicated by the median accuracy of 61%. This suggests that the classifier is learning relevant patterns in the data, albeit with room for improvement. The low precision was expected given the imbalance in the dataset, where non-startle samples vastly outnumber startle samples.

In summary, while the pseudo-online evaluation did not meet the desired accuracy levels, it did confirm that the classifier is operational and can detect startle events with some degree of reliability. The offline analysis outperformed the pseudo-online evaluation, likely due to the more consistent nature of the data used in the former. The higher number of non-startle samples in the pseudo-online case contributed to the lower precision observed.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

The performance differences between offline and pseudo-online evaluations suggest that the model's robustness could be improved, particularly when faced with more varied and less structured data. These findings underscore the importance of further refining the classifier, especially in terms of improving precision and ensuring consistent performance across different datasets and scenarios.

1.7. Conclusion and future planning

The current state of the startle detection component meets the requirement of minimal delay in classification, allowing for timely intervention by the digital assistant in time-critical scenarios. The system is adaptable, able to utilize varying numbers of modalities based on available devices, and it does not require users to train the system themselves. However, there are still limitations that need to be addressed.

The precision of the model remains an area of concern, particularly in real-world applications where a high rate of false positives could be problematic. While the offline analysis showed promising results, the pseudo-online evaluation revealed that there is still a significant gap to bridge before the model can be reliably used in practice. The imbalance in the dataset, the variability of physiological signals across users, and the differences in data collection methods all contribute to these challenges.

For future work, several steps are planned to improve the model:

- Data Expansion: To create a truly cross-user model, more data will be needed to generalize the problem better. This includes incorporating more data from simulator environments and other experiments similar to VAL1.
- Unbalanced Training: Implementing unbalanced training strategies to better utilize the non-startle samples could help improve the precision of the model. This might involve creating multiple balanced classifiers or using models that penalize the non-startle condition during training.
- Multiple Source Domain Adaptation (MSDA): To address the variability across datasets, MSDA techniques like CoDATS and Multi-EPL will be explored. These approaches could help improve the model's ability to adapt to different data sources, thereby enhancing its generalizability.
- Continuous Learning: Although the current model operates in a cross-user approach, continuous retraining and calibration will be necessary in future iterations. This will be implemented using a reinforcement learning approach, where feedback from the pilot is used to tailor the model to individual users.

In the next phase, a real online evaluation will be conducted in VAL2, where the model will be tested in a live environment. This will provide further insights into its performance and help refine the system for eventual real-world deployment. The

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

continuous learning aspect, where the system evolves based on real-time feedback from pilots, will be a critical component of this future work.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

2. Use case 2: Flight deck route planning/replanning

2.1. Context

Use Case 2 focuses on an assistant solution to reduce the cognitive workload of pilots during mission management tasks. This is done by establishing an enhanced, bilateral communication process between the Human and the System to present, understand and evaluate options for courses of action. ComBi, a ML-based solution patented by Thales, is employed for this purpose. The possibilities proposed by the assistant is an enhanced bilateral communication between the Human and the System.

The human user is considered central to the system: it is a Human Centered Design where the user can communicate with the assistant through high-level operational metrics, that may be evaluated in terms of technical KPIs if needed. That is to say, the pilot does not have to think in a technical point of view and dealt with the low-level complexity of the solver.

For Haiku's UC2, the scenario is a flight from Marseille to Munich. The destination airport is closed due to meteorological degradation. The pilot asks the assistant to suggest rerouting flight plans (FP) to the best alternative airport following the best alternative trajectory respecting his/her preferences.

2.2. State of the art

Supervised learning is a type of machine learning where a model is trained on a labeled dataset, meaning each training example is paired with an output label. The goal is for the model to learn the relationship between the input data and the output labels so that it can accurately predict labels for new, unseen data. This approach is widely used in tasks like classification and regression, where the model learns from examples with known outcomes and then applies this learned knowledge to make predictions on future data.

Here, we aim to translate between two distinct spaces: the operationally understandable space, where human intentions reside, and the technical space, where low-level parameters are defined. This bidirectional translation process is crucial for aligning human intentions with technical parameters and vice versa. However, capturing the nuances of these translations is complex, as the lexical representation of intentions can be fuzzy and varies slightly between pilots. To address this, we use fuzzy logic models to accommodate the lack of a clear ground truth.

The translation process leverages expert knowledge from pilots, gathered through interviews and annotations, to create a labeled dataset. This dataset consists of samples with multiple float inputs (representing technical parameters) and

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

corresponding float outputs (reflecting human intentions). With this labeled data, we can apply supervised learning to model the relationships between these spaces. The model used in this project is ComBi, the Bidirectional Communicator using Genetic Fuzzy Tree (GFT), a frugal and explainable AI technology developed by Thales, which excels in handling such complex, fuzzy translation tasks within the framework of supervised learning.

2.3. Requirements analysis

In this context, the pilot's high-level operational intentions are reduced to three key objectives: pilot's cognitive comfort, passenger's comfort, and airline profitability.
Pilots can communicate their preferences by ranking these intentions in order of importance through an interface connected to the ComBi system (Figure 3).

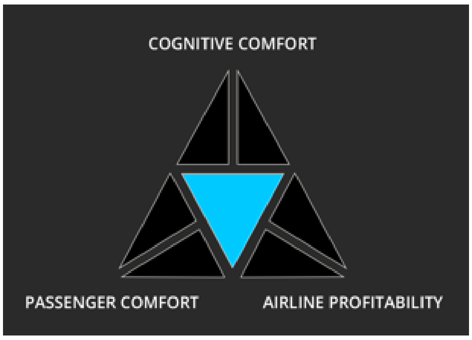


Figure 3: Downward translator interface.

The ComBi assistant will then translate these ranked preferences into technical parameters to construct a multicriteria costmap, which will be used by a trajectory solver to propose potential flight plans. Each proposed flight plan will be explained using the ComBi Upward xAI model, where each trajectory is mapped to Key Performance Indicators (KPIs) and translated back into the pilot's operational metrics (Figure 4).

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

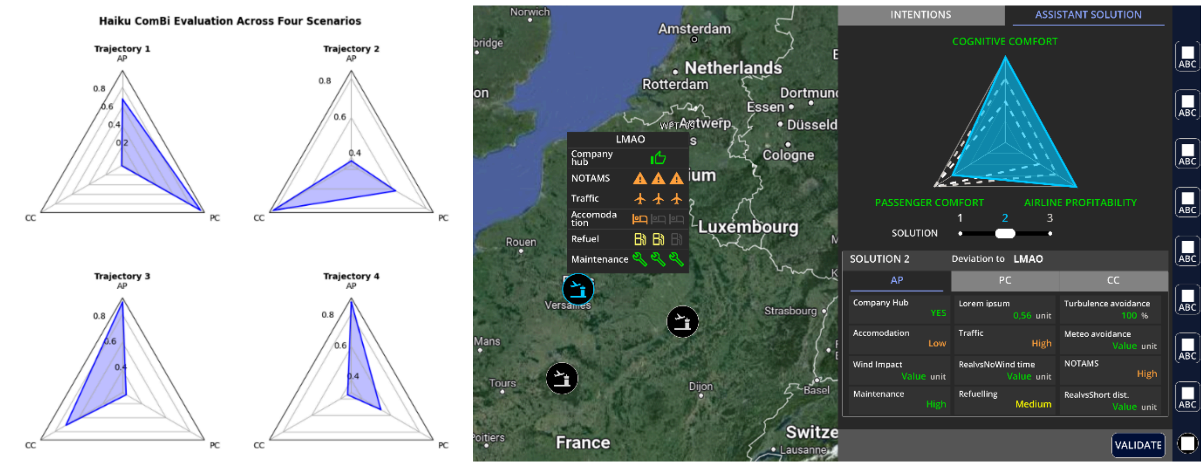


Figure 4: Trajectories suggested to the pilot.

2.4. Training and Data processing

The model training and development follows a Thales internal process, based on EASA guidelines for Dal D certification. This process allows for data collection and analysis, as well as model design and training. The model architecture (Figure 5) is described in the solution architecture in Deliverable 4.2: Intelligent assistant architectures.

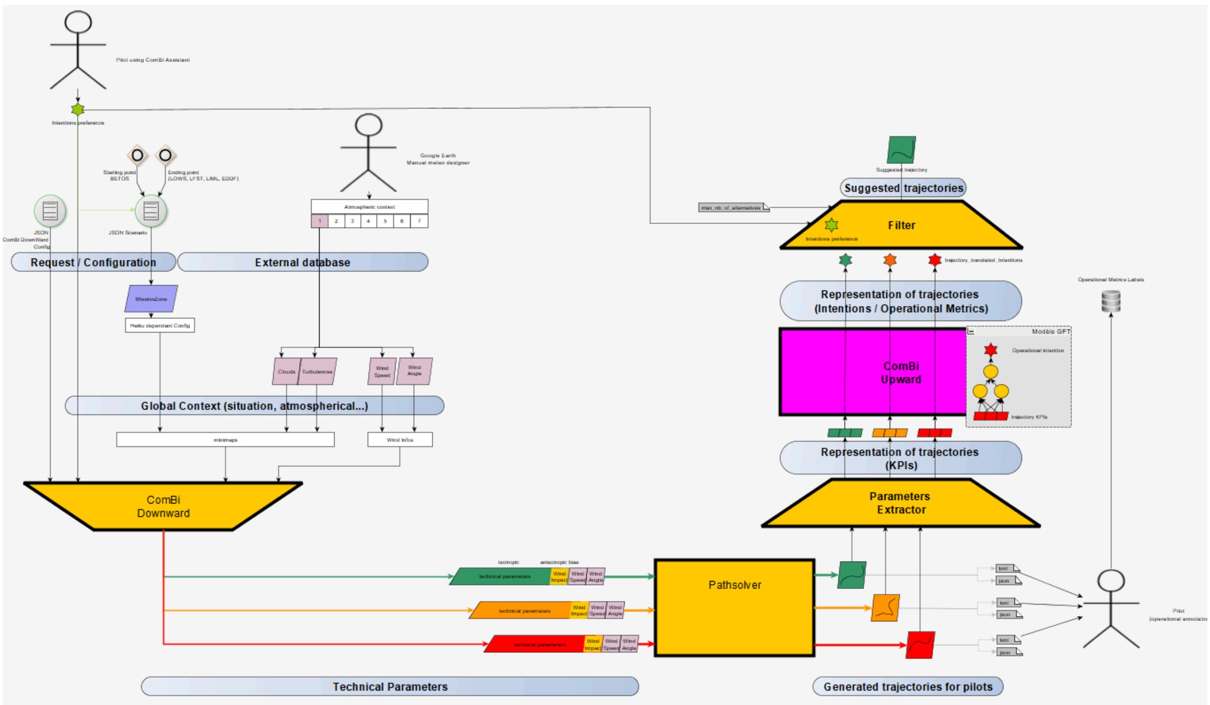


Figure 5: Model architecture. Source: Deliverable 4.2.

2.4.1. Airport KPIs

The development of the ComBi Upward translator relies on a robust process of data collection, particularly focusing on flight trajectories in relation to specific situational contexts. To train the ComBi Upward model within the HAIKU project, we manually

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

generated seven distinct meteorological contexts. These contexts included minimaps for clouds, turbulence, and wind, which were then translated into cost values. Clouds and turbulence were modeled using isotropic costs, while wind information was represented as anisotropic costs. These costs provided the foundational data necessary for context generation and situation evaluation.

For airport representation (Figure 6) in reroute decision-making scenarios, we simplified the airport data into a 6-float vector format. This vector included a binary value for whether the airport is a hub, along with categorical levels for NOTAMS, traffic, accommodation, refueling, and maintenance. This approach led to a KPI space comprising 486 unique vectors.

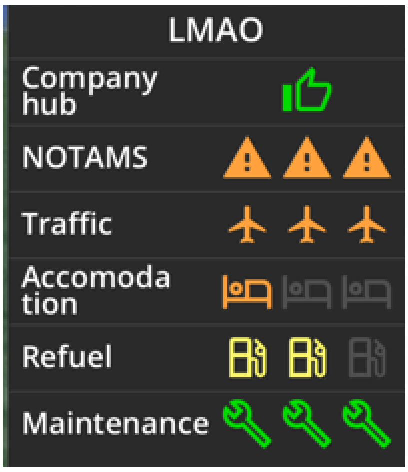


Figure 6: Airport representation

To enrich the dataset, we conducted interviews with seven pilots, during which 85 vectors were selected—primarily at random—from the set of 486 for annotation, with a few extreme cases deliberately included. Overall, 595 vectors were shown to the pilots, leading to a dataset that covers 42.6% of the entire airport KPI space, with 207 vectors shown and 279 remaining unseen. On average, each vector was presented to a pilot 1.22 times for annotation.

This structured data collection process ensures that the ComBi Upward translator is trained on a diverse and representative set of data, enabling it to accurately translate and explain flight trajectories in relation to operational contexts.

2.4.2. Trajectory KPIs generation

In this phase of the project, multiple flight trajectories were computed by aggregating cost maps through a grid search approach and utilizing a state-of-the-art path solver. This process, applied across seven distinct meteorological contexts and four rerouting airports, resulted in the generation of 85 unique trajectories.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

To facilitate analysis, these trajectories were converted into KPI vectors using a coded parameter extractor, resulting in an (85, 5) representation—85 trajectories each characterized by five KPIs. The chosen KPIs to represent our trajectories are as follows:

- **Turbulence Avoidance:** This KPI measures how effectively a trajectory avoids various types of turbulence (e.g., minor, major). It computes a ratio based on the proximity to the center of the turbulence's dangerous area, divided by the maximum possible distance from the turbulence.
- **Cloud Avoidance:** Similar to turbulence avoidance, this KPI evaluates how well the trajectory avoids different types of clouds (e.g., scattered, storms). It calculates a ratio of the proximity to the center of the storm's dangerous area, divided by the maximum possible distance from the storm.
- **Wind Impact:** This KPI assesses how well the trajectory optimizes wind effects. It computes a ratio of the average ground speed gain during the flight to the maximum ground speed gain achieved, considering the aircraft's commanded motor speed.
- **Real vs. Shortest Distance Ratio:** This KPI represents the efficiency of the flight path in terms of distance. It calculates a ratio between the actual distance flown and the shortest possible distance (if the trajectory ignores factors like wind, storms, or turbulence), divided by the real trajectory distance.
- **Duration:** This KPI reflects time consumption, which is also indicative of fuel consumption. It calculates a ratio of the actual time spent on the trajectory to the maximum allowable time (max_playtime) for the reroute.

2.4.3. Discussion over generated KPIs

The exploratory data analysis conducted on the trajectory inputs has provided valuable insights into potential areas for improvement, such as the use of data augmentation. The analysis of our KPI histograms reveals a favorable trade-off between operational situation representativeness and the balance of our distributions, which, while not typically regular, are well-suited to our specific use case.

One notable observation is the presence of missing values in the duration KPI, particularly in the range between 0.3 and 0.4. Despite this, we do have samples covering low, median, and high values, indicating that the overall dataset still retains valuable information.

Boxplots of the KPIs (Figure 7) indicate strong coverage, especially for the Clouds and Turbulence Avoidance KPIs, suggesting that these aspects are well-represented in the dataset. The Wind Impact KPI shows comprehensive coverage across the entire range, with a higher concentration of data in the lower values. This skew towards lower

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

values is expected, as it is uncommon to encounter tailwinds throughout a flight in degraded meteorological conditions. Nevertheless, our meteorological contexts and scenario design have ensured that the dataset covers the entire design domain, including extreme and intermediary values, which is crucial for model training.

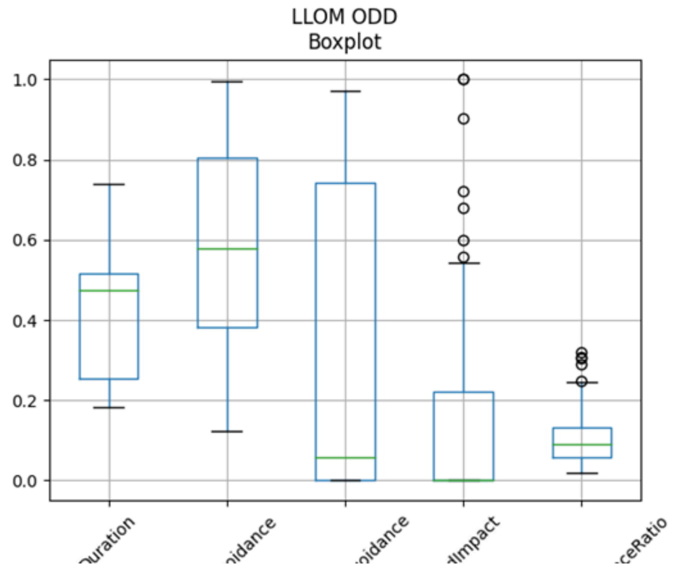


Figure 7: Trajectory KPIs inputs boxplots.

However, the Real vs. Shortest Distance Ratio KPI is less diverse, covering only 35% of the theoretical values. This limitation can be attributed to the specific scenarios for which we are training the model. Given that this assistant is intended to suggest rerouting options in situations where an airport is closed, and the aircraft is nearing the end of its flight with potentially low fuel, the dataset naturally emphasizes scenarios that avoid the 65% of trajectories that would involve significant deviations, frequent course changes, and strong impacts on passengers and crew.

In conclusion, the Low-Level Operational Metrics dataset does not exhibit a perfectly regular distribution, reflecting the specificities of our use cases and the inherent conditions of using the ComBi Assistant. The KPIs have their own unique distributions, which could be balanced further through techniques like undersampling or synthetic data generation. However, considering the notional Operational Domain Design, the dataset is sufficiently qualitative to allow the model to learn relevant patterns and rules, maximizing the value of this dataset for our application.

2.4.4. Data annotations and operational metrics

To refine the ComBi assistant's ability to align trajectories with pilot intentions, meteorological contexts paired with suggested trajectories were presented one by one to seven pilots. Each pilot analyzed the situation and then annotated and graded their agreement, on a scale from one to five, with how well the trajectory fit each of the

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

three primary intentions: Pilot Cognitive Comfort (CC), Passenger Comfort (PC), and Airline Profitability (AP).

During the evaluation, the pilots were asked to rate all 85 context-trajectory combinations. They responded to specific affirmations about each combination, indicating their level of agreement or disagreement. A score of 1 represented strong disagreement, while a score of 5 indicated strong agreement. This scoring system is used as a hypothesis to represent the operational metric linked to each intention. The goal for the ComBi assistant is to learn how to reproduce these grades based on the annotations, so it can assess and grade any trajectory within the Operational Design Domain (ODD) using the input space KPIs.

The interviews resulted in each of the seven pilots annotating all eighty-five trajectories, creating a dataset of 595 annotations—essentially, seven annotations per trajectory. This dataset highlights the broad range of concepts annotated by the pilots, reflecting the inherent fuzziness in how different pilots perceive and express the intentions of Cognitive Comfort, Passenger Comfort, and Airline Profitability.

During the interviews, pilots followed a consistent annotation process to evaluate the suggested trajectories and their associated airport KPIs. To ensure accurate assessments, we explained the different airport KPIs to the pilots, providing a clear explanation of the annotation protocol. Additionally, explanatory notes were kept next to the icons throughout the process, helping the pilots to correctly identify and evaluate the airports based on how well they aligned with the three intentions: Airline Profitability (AP), Pilot Cognitive Comfort (CC), and Passenger Comfort (PC).

However, a variance in evaluations emerged among the pilots. Due to differences in experience, pilots did not always assess the same trajectories and destinations in the same way. This variability highlights the fuzzy nature of how intentions are interpreted, as each pilot's understanding and evaluation of the concepts may differ slightly. Since there is no clear binary truth in these evaluations, fuzzy logic models are employed to account for these differences and better capture the range of pilot interpretations.

2.4.5. Training and evaluation sets

To prevent overfitting and ensure a reliable assessment of our model's performance during the training phase, the cleaned dataset was divided into two subsets: a training dataset and a testing dataset.

- **Training Dataset:** This subset comprises 75% of the full dataset, amounting to 446 samples. Each sample includes 5 KPIs annotated with 3 operational metrics. This portion of the data is used to train the model, allowing it to learn the underlying patterns and relationships.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

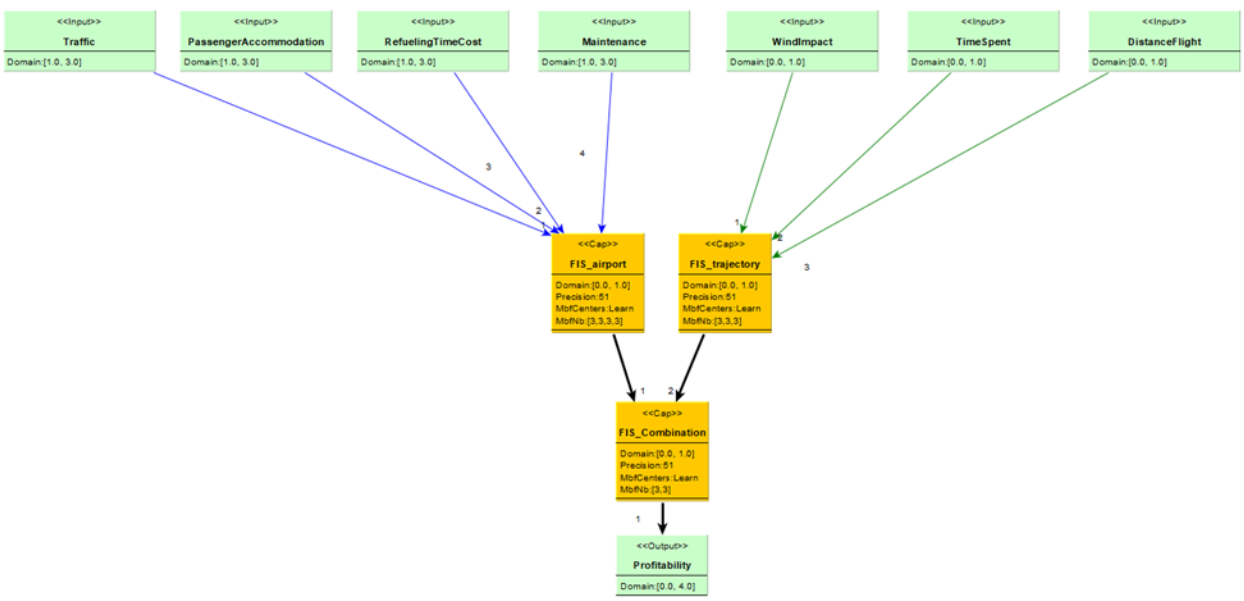
- **Testing Dataset:** The remaining 25% of the dataset, consisting of 149 samples, forms the testing dataset. Like the training set, each sample includes 5 KPIs annotated with 3 operational metrics. Importantly, the data in this subset has not been used during training, ensuring that the model's parameters and hyperparameters are evaluated on entirely unseen data.

This splitting approach helps in accurately evaluating the model's generalization ability, ensuring that the performance metrics obtained during testing provide a realistic measure of how well the model will perform on new, unseen data.

2.5. Upward translator model architecture

The ComBi Upward translator can be conceptualized as three fuzzy trees, with each tree dedicated to predicting and inferring the operational metric value for one of the three key intentions: Airline Profitability (AP), Pilot Cognitive Comfort (CC), and Passenger Comfort (PC). The model architecture for all three intentions follows a consistent design pattern, organized into two main levels that drive the decision-making process: airport consideration and trajectory consideration.

In Figure 7 the relevant KPIs for each intention are displayed, with airport-related KPIs on the upper left and trajectory-related KPIs on the upper right. These KPIs, or Low-Level Operational Metrics, are grouped and connected to two fuzzy processing units. One unit computes an independent score related to the airport, while the other computes a score related to the trajectory. These scores are then combined in a mitigation-processing unit to infer the operational metric corresponding to the specific intention.



This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

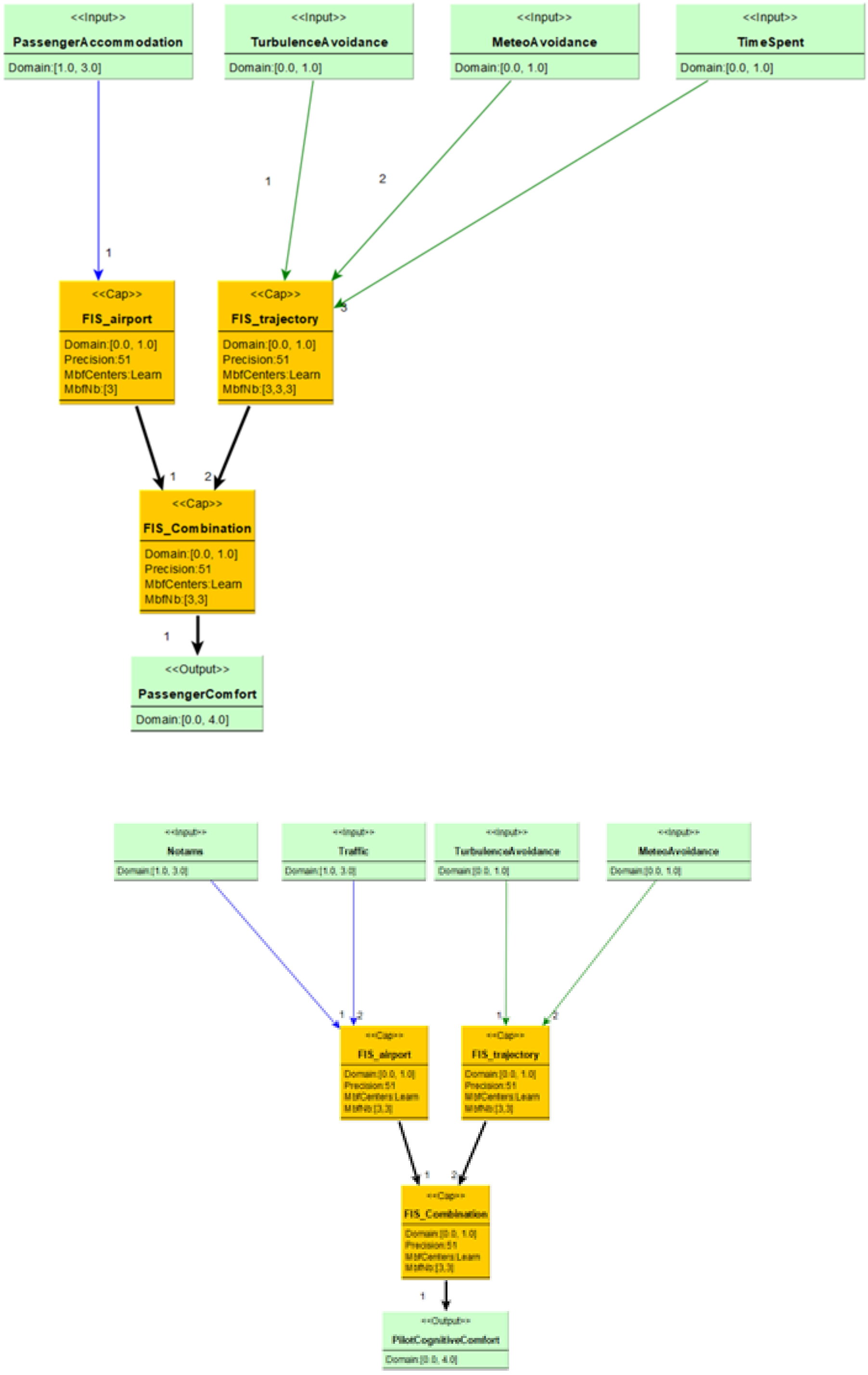


Figure 8: KPI mappings. Top: Airline Profitability (AP). Middle: Passenger Comfort (CM). Bottom: Pilot Cognitive Comfort.

The initial step in designing this model involved analyzing insights from the gathered dataset. Several model designs and paradigms were considered, with KPIs grouped into two categories: trajectory-related KPIs and airport-related KPIs.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

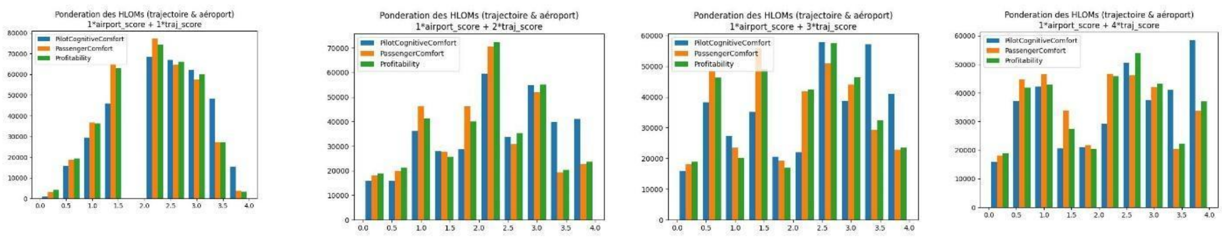
- **Single Input Models:** In this approach, each model predicts high-level operational values by using individual KPI groups (either airport or trajectory).
- **Multi-Input Models:** These models predict high-level operational values by combining all KPI groups, integrating both airport and trajectory KPIs.
- **Hierarchical Models:** This approach predicts high-level operational values by assigning different importance levels to each KPI group. A Fuzzy Computation unit infers the operational value for the airport, and another unit does the same for the trajectory. A final inference unit then combines both scores (trajectory and airport) in a nonlinear and explainable manner to infer the overall high-level operational metric.

This architecture leads to a model that combines the airport score and the trajectory score as annotated by the interviewed pilots, ensuring that the inferred operational metrics are both accurate and aligned with pilot assessments.

2.6. Insights and obtaining optimal model parameters

Using Principal Component Analysis (PCA), we visualized the dataset by reducing it to two constructed dimensions. The complexity of the underlying function became apparent, highlighting the need for an estimator capable of managing nonlinear, high-dimensional relationships.

The training process employed a genetic algorithm, where a population of chromosomes evolved according to Darwinian principles. Each chromosome represented a specific configuration, with parameters that were fine-tuned throughout the training. The Mean Squared Error (MSE) was used as the fitness function, guiding the optimization by minimizing the impact of outliers and improving model performance.
To enhance the statistical distribution, annotations provided by pilots were aggregated. The results of this aggregation are shown in Figure 9, with the corresponding statistical distribution depicted in Figure 10.



This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

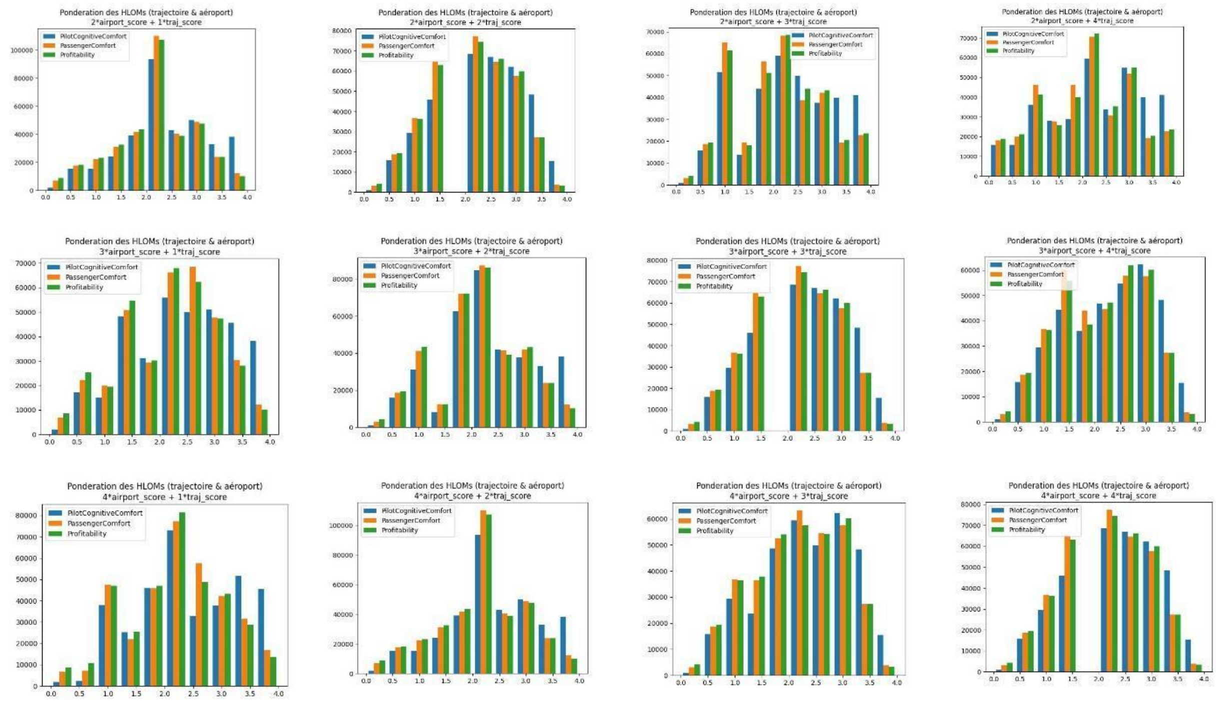
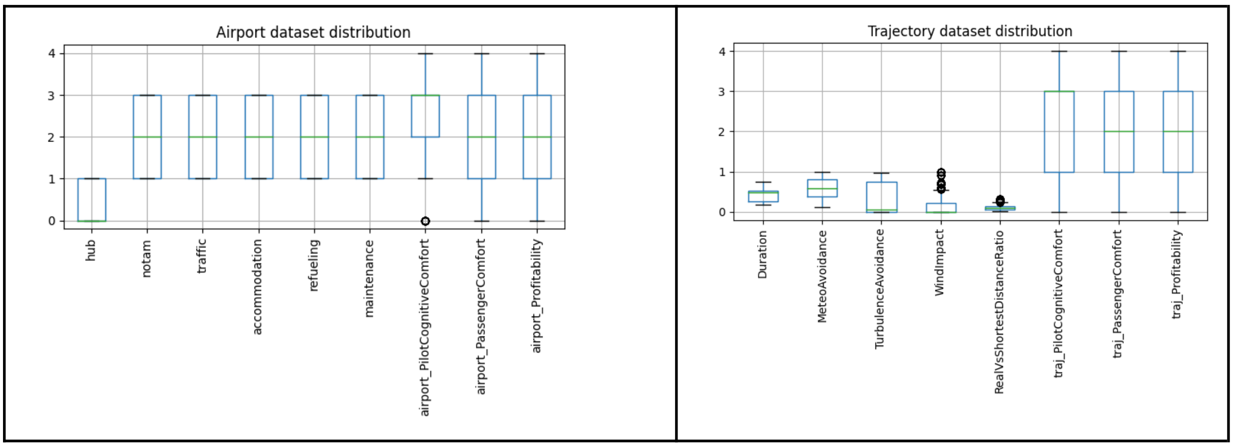
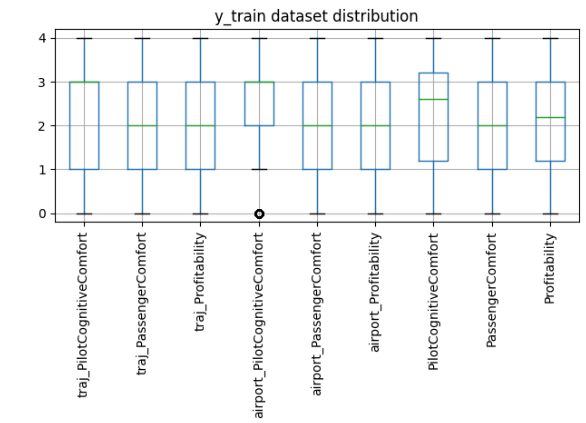


Figure 9: Impact of aggregation on the combined annotated score distribution.

The decision to aggregate data for predicting combined intentions was based on statistical considerations, with a focus on achieving a regular and balanced distribution of output labels. This linear aggregation approach was chosen after analyzing the sixteen subplots in Figure 9, which illustrates the impact of aggregation on the distribution of combined annotated scores.



This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332



	Valeurs différentes (round 2)
Duration	30
MeteoAvoidance	49
TurbulenceAvoidance	30
WindImpact	30
RealVsShortestDistanceRatio	24
hub	2
notam	3
traffic	3
accommodation	3
refueling	3
maintenance	3

Figure 10: Training dataset distribution for combined training from a boxplot point of view

The dataset analysis reveals that it is well-balanced and representative of various scenarios. Figure 10 provides statistical insights into the KPIs related to both Airports and Trajectories within the collected data:

- **Upper left plot**: Demonstrates the distribution of the airport dataset, showcasing factors such as the presence of a hub, the significance of NOTAMs, traffic congestion, refueling facilities, and maintenance capabilities. These are represented by categorical values (1, 2, or 3).
- **Upper right plot**: Distribution of the trajectory dataset, using similar categorical values.
- **Bottom left plot**: displays the distribution of labels via boxplots, representing the three operational intentions—Pilot Cognitive Comfort, Passenger Comfort, and Airport Profitability—each in relation to the specific concept being considered (trajectory, airport, or a combination of both).
- **Bottom right plot**: highlights the diversity of values within each KPI, offering statistical insights into the range and variability present in the dataset.

The same dataset analysis is presented from a histogram perspective in Figure 11, with Correlation Matrix." This approach provides a more detailed view of how the data is balanced and its representativeness, particularly regarding missing values.

In addition, the correlation matrix for all KPIs is included to highlight the intrinsic correlations among the different KPIs. This analysis was crucial for identifying any KPIs that might carry redundant information, ensuring that the input set is both efficient and informative for model training.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

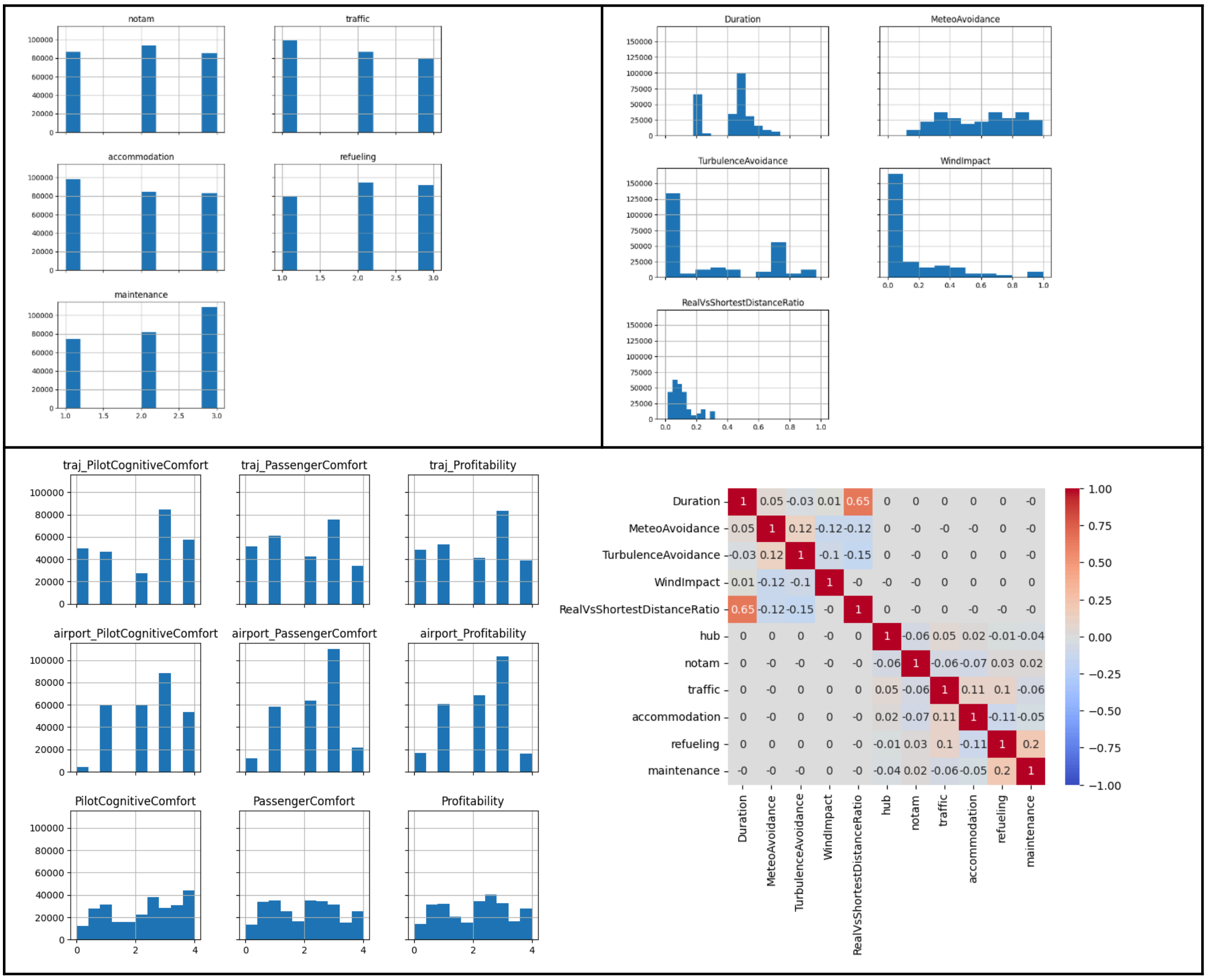


Figure 11: Training dataset distribution for combined training from an histogram point of view, with correlation matrix.

2.7. Model performance and overfitting control

In the context of fuzzy logic trees and supervised learning, model fitness refers to how well a model captures the underlying patterns in the training data while being able to generalize effectively to new, unseen data. Fuzzy logic trees, which handle uncertainty and imprecision through fuzzy logic rules, require a balanced model that can adapt to various scenarios without losing accuracy.

Overfitting, however, is a common challenge where a model becomes too closely tailored to the training data, picking up on noise or irrelevant details that do not translate well to new data. In fuzzy logic trees, this could lead to overly complex rules that perform well on the training set but fail in real-world applications. To mitigate overfitting, techniques such as pruning, regularization, or cross-validation are often employed, helping to ensure that the model remains robust and capable of making reliable predictions even in uncertain and varied environments.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Figure 12 shows the genetic training with early stopping to avoid overfitting. Visualizations about our chromosome population evolution can be seen. A fitness score is attached to every chromosome. For both two figures presented, the upper subplot presents the fitness on the training set whereas the bottom subplot highlights the fitness on the test set.

Different ranges are used to analyze our population's evolution across the generations: we have represented the evolution of the different standard deviations of our population's fitness: we have the different major statistical metrics for our chromosomes such as the min, the max, mean and quartiles' fitnesses in our population.

The same observation can be made for Pilot's cognitive comfort (Figure 13) and Passenger's comfort (Figure 14).

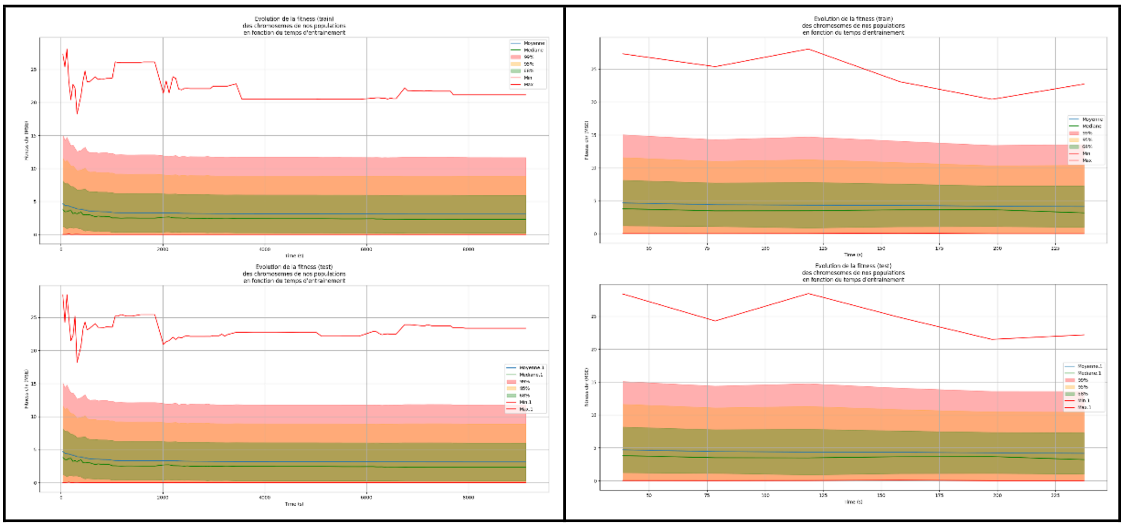


Figure 12: Representation of the optimization's steps performed during the Airline profitability's model training, with the genetic generations occurring, with a specific zoom on the chromosomes' using MSE (Mean Square Error).

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

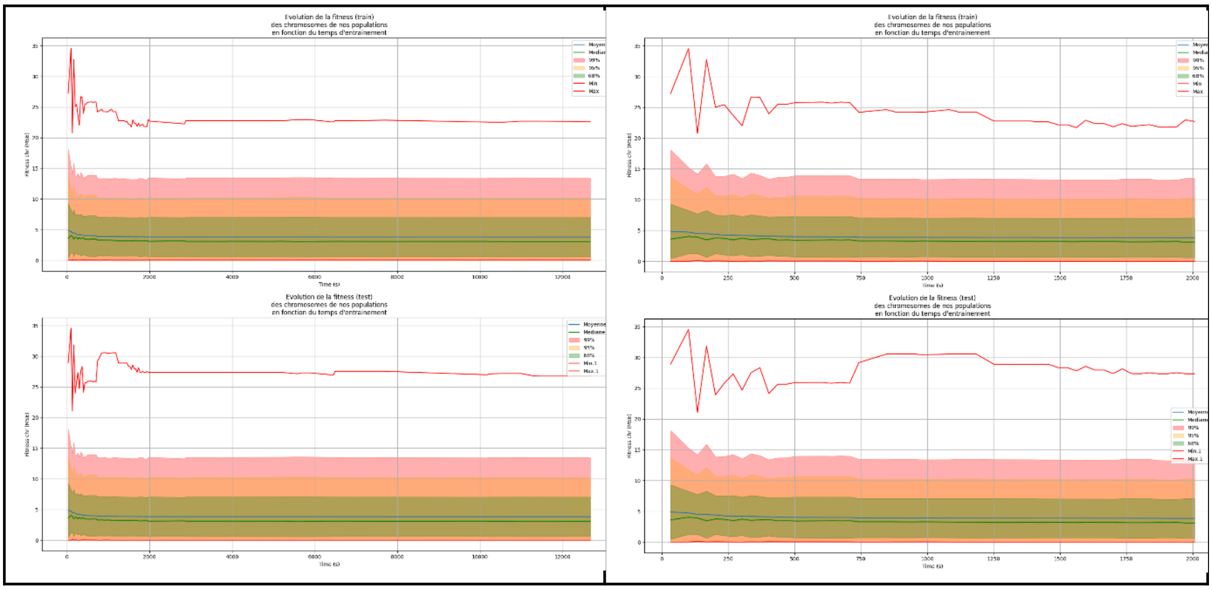


Figure 13: Representation of the optimization's steps performed during the Pilot's Cognitive Comfort's model training, with the genetic generations occurring, with a specific zoom on the chromosomes' using MSE.

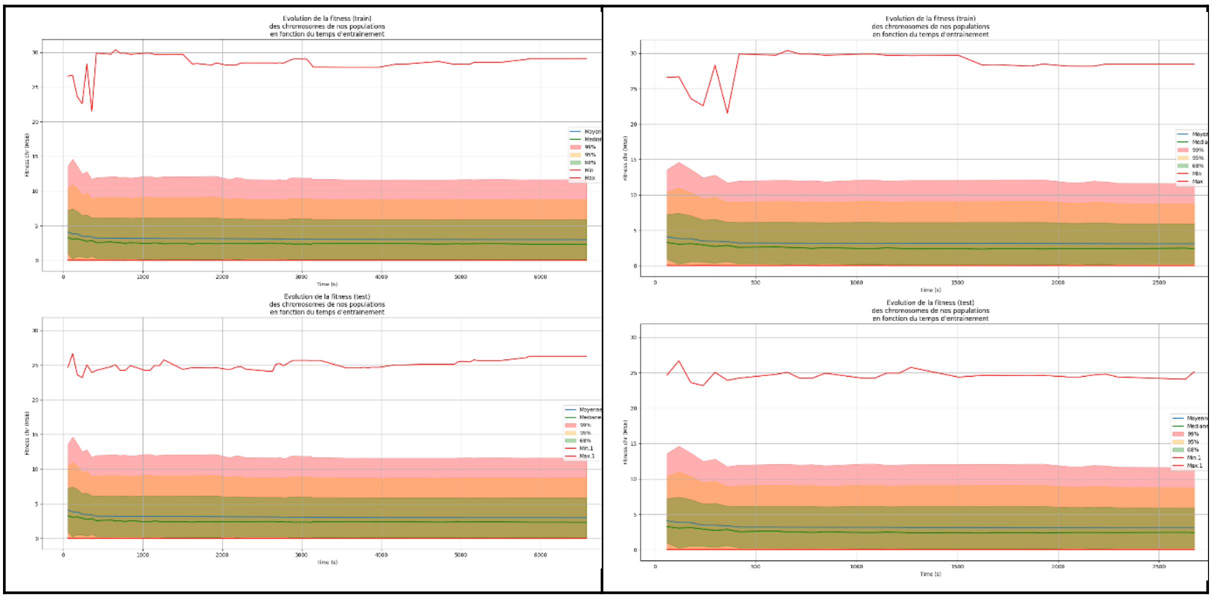


Figure 14: Representation of the optimization's steps performed during the Passenger's Comfort's model training, with the genetic generations occurring, with a specific zoom on the chromosomes' using MSE.

To enhance efficiency, the genetic training was adjusted to focus on optimizing a single factor, such as trajectory assessment, rather than airport-related factors. This adjustment reduced the overall training time and increased the model's robustness to errors and outliers. Without parallelization, the total training time for one model is currently approximately 3 hours—2 hours dedicated to loading the training and testing datasets, and 1 hour for the genetic training optimization.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

The Absolute Error metrics for the three high-level operational intentions—Passenger Comfort, Pilot Cognitive Comfort, and Airline Profitability—remain below a mean threshold of 1. This low error value is notable given the variability in pilot annotations.

Overall, the results from training and testing indicate that the models perform well. Additionally, results visualized using a Google Earth view of the ComBi-assistant are shown in Figure 12. These results suggested a flight plan adjustment, recommending rerouting from the BETOS waypoint to LOWL (Linz airport) instead of continuing to EDDM (Munich airport).

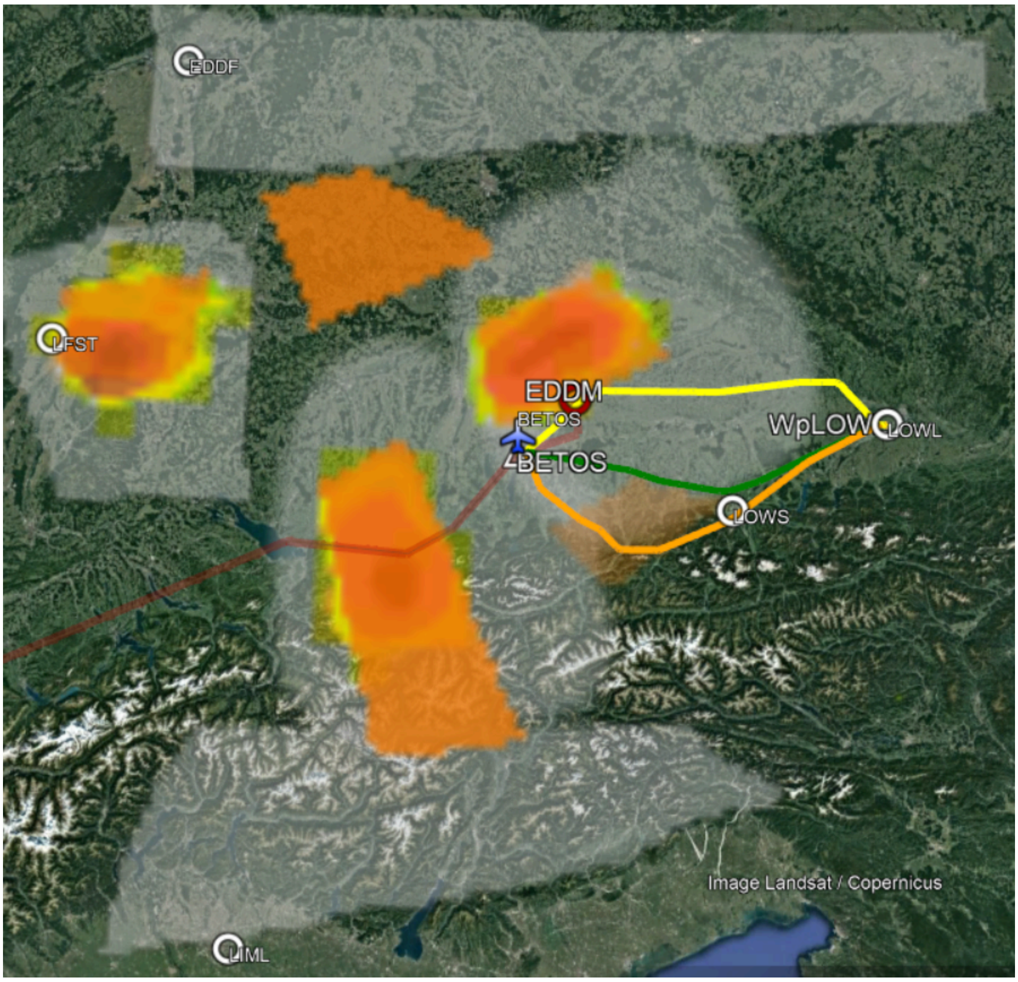


Figure 15: ComBi assistant suggesting 3 flight plans. These 3 flight plans have been evaluated by the trained models and they are the best ones in terms of operational intentions.

2.8. Conclusion and future planning

Based on the good performance of the models, we do not anticipate making any immediate changes. However, we will continue to monitor their effectiveness and consider future enhancements. One potential area for improvement is the integration of continuous learning, which could involve gathering feedback from pilots to refine the models' trajectory predictions and airport suggestions over time.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

3. Use case 3: Urban Air Mobility

3.1. Context

This use case is a futuristic concept which explores an operational domain (management of urban airspace, or U-space) with end users who do not exist yet, like the UAM Coordinator. Given that the concept is still at its early stages, UC3 doesn't have an already built and tested AI component. Nonetheless, the plan is for the DUC Intelligent Assistant to help the future UAM Coordinator in overseeing U-space operations. To do so, DUC would handle most routine tasks, such as traffic and conformance monitoring, and providing flight and weather information. Additionally, the DUC would direct the UAM Coordinator's attention to specific situations or events as necessary, using visual cues in the interface. To perform these functions effectively, DUC would need advanced capabilities that can be enhanced by AI applications.

These would be examples of what the DUC should do - and it should leverage AI technologies. However, given that the concept is futuristic in nature, we couldn't just develop an AI to be tested like other Use Cases, and had to go a different route. First, we had to find an overall approach to identify how AI could be concretely used on the basis of our vision of DUC. The IA UC3 has a much broader scope than in other Use Cases because the high-level tasks for the human and IA are not specifically defined or limited to a particular function. The IA's capabilities are envisioned within a futuristic workplace and environment, requiring an initial development of a comprehensive understanding and concept of the overall working environment. Second, we have to select a few feasible applications that we can actually work on. For this Use Case, the contribution of this deliverable addresses the second step.

3.2. Approach

When considering AI applications that DUC could incorporate, we started by exploring problems and challenges to be solved in relation to the high-level tasks that DUC and the UAM Coordinator would work on (Figure 13). We have distinguished two areas of high-level tasks. The first group of tasks relate to the process and problems on which the DUC and UAM Coordinator work collaboratively. These high-level tasks relate to the U-space services that the DUC and UAM Coordinator provide to achieve their shared high-level goal: U-space safety, efficiency, and security to all U-space users. The second group of tasks pertains to the capabilities DUC needs to function effectively as a team member and contribute to a well-functioning team. These tasks relate to the four pillars of our theoretical human-AI teaming framework.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

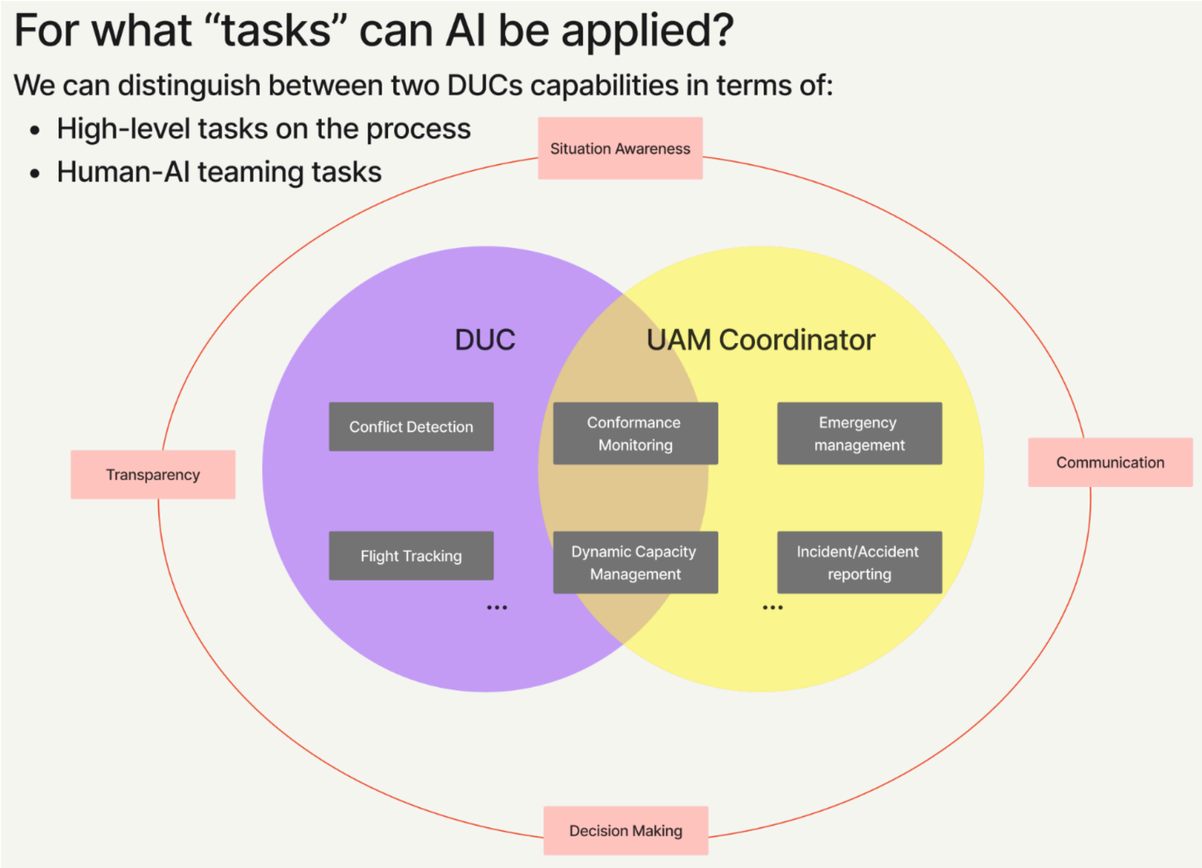


Figure 16: Framework for potential AI applications in relation to high-level task.

To generate ideas about potential AI applications for this use case, an online workshop was conducted with the members of the consortium. The members were divided into two groups and were asked to identify high-level task for which AI can be used and what problems can the AI solve.
The first group identified the following as a high-level task.

High-level task	Comments
Emergency Management (instructed to discuss)	Dynamically/tactically update flight plans in emergency situations to achieve "continue to safe landing" objective In an abnormal situation the human has a great ability to analyze the situation. AI/automation ability to process large amounts of data. Time is often an important factor in emergency situations, so the use of AI would be great to scan through several sources of data that could be of importance and the present a condensed version highlighting important factors to the human to base the decisions on.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

	Data: U-space user registration and network data, U-plans and contingency plans, aircraft and battery status (endurance, range)
Dynamic Capacity Management	Depending on what sources of information that will be available, the AI could perform dynamic capacity management based on factors on ground in the city. Also, if we get IAM in the air, it will fly at higher altitudes. AI could be good at analyzing the current traffic situation within ATC and then balance how U-space operations can operate regarding ATS/ATC in the same area. Prediction and identification of traffic hotspots. Risk assessment based on uncertainties (deviations) in flight plans Provide replanning options in case of higher risk identified. Dynamic Management taking into consideration external factors (like on the ground events) DUC could predict / early detect saturated zones and bottlenecks from updated flight plans (and associated uncertainties) and propose rescheduling options with global and individual outcomes for the UAM coordinator to evaluate
Tactical Conflict Prediction and Resolution	DAA is supposed to be the last layer in a safety net. AI could assist in making sure DAA is not triggered. Or maybe AI could help in presenting a better avoidance maneuver than today's TCAS that only works with vertical separation. Vertical adjustment is not perfect to use close to the ground as well.
Tracking (flight tracking)	Basic on historical data it will be possible to predict movements
Vertiport/cargo-hub dynamic information	AI is able to present to the UAM Coordinator what actions to take given the most updated vertiport info / status, e.g. contact Vertiport A, contact flight A.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Incident/Accident reporting	DUC could help the UAM Coordinator to summarize the relevant information available from data sources, while the UAM Coordinator checks if complementary information must be added. This could help to make more agile reports, while information is still "fresh" in the UAM Coordinator's mind. These reports may be structured to form databases that may bring understanding on causal relationships and feed predictive/decision models to enhance incident prevention policies. Using ML techniques, it is one of the easy tasks to reach and to get results.

Table 2: High-level tasks identified by the first group.

The second group identified the following high-level tasks:

High-level task	**Comments**
Emergency Management	Several empty notes with no comments [reflects that task was considered important but participants did not have time to discuss it]
Dynamic Capacity Management (instructed to discuss)	AI applications: ● Determining capacity - Noise limit (risk of complaints) - Residual risk of collision - Availability of vertiports for emergency landings ● Determining demand - Predicting traffic ● Resolving demand / capacity imbalances
Tactical Conflict Prediction and Resolution	Too High volume of traffic for human resolution Conflicts anywhere Unknown intent for some GA flight Maybe should do some calculations on collision risk - small UASs will have a very low probability of collision. This may guide us in how much work is focused on this activity.
Vertiport/cargo-hub dynamic information	Several empty notes with no comments [reflects that task was considered important but participants did not have time to discuss it]

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Tracking (flight tracking)	Several empty notes with no comments [reflects that task was considered important but participants did not have time to discuss it]
Conformance monitoring	One empty note with no comments [reflects that task was considered important but participants did not have time to discuss it]
Geo-awareness	One empty note with no comments [reflects that task was considered important but participants did not have time to discuss it]
Weather information	AI scans weather sources for example, TAF, and communicates through a dialogue window. Could also alert for unstable weather towering cumulus, thunderstorms. It can check multiple sources and also collect data from drone sensors- and give feedback to ATM. AI interprets the information and communicates with the human on an appropriate level. The human can also ask questions regarding forecasted weather. Implications discussed.
Geographical information	AI monitor and logs irregularities, and updates. Cross checks the latest updates installed. Checks with NOTAM for example cranes or other temporary obstacles. Could also keep track of obstacle clearance heights (OCH) margins and updates in charts approaches etc. Ie related to airports and if crossing "below"
Population density map	Interesting and applicable, to predict activation of Restriction zones, congestion areas, presence of police and ambulance traffic.
Drone aeronautical information	One empty note with no comments [reflects that task was considered important but participants did not have time to discuss it]

Table 3: High-level tasks identified by second group.

For possible AI applications relevant to DUC, the first group discussed emergency management and suggested that AI can be leveraged for data processing to build up proper situation awareness in case of emergency. In addition, "Continue safe flight and landing" (safe landing at any time) was discussed and concluded that AI can be used for rerouting aircraft in an emergency.

The second group discussed dynamic capacity management and established that this is a synthesis of other high-level tasks. For example, AI can be used for emergency landing sites for addressing dynamic capacity (i.e. predicting capacity constraints)

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

based on monitoring several data parameters. Taken together, the following applications were identified:

- U-space capacity measures (who to reroute or delay)
- traffic hotspot/congestion predictions
- traffic flow optimization
- forecasting models for determining aircraft reroute and diversion solutions.
- conformance monitoring in terms of anomaly detection
- natural language processing for bi-directional voice communication
- attention guidance for supporting situation awareness sharing (building on monitoring UAM Coordinator using eye-tracking)
- priority determination in decision making/recommending actions.

3.3. Conclusion and future planning

The DUC IA concept of operations is ambitious in that it encompasses a wide variety of high-level tasks that can leverage AI technologies. These high-level tasks comprise a spectrum of functions ranging from autonomous (e.g., conformance monitoring and conflict detection and resolution), shared (e.g., checklist management, information retrieval, providing status reports) to assisting capabilities (support in problem solving by e.g., making recommendations). The work conducted thus far in UC3, including the described workshop in this deliverable and other activities for defining the UAM environment and UAM Coordinator role, has provided us with an initial understanding of the DUCs role, what high-level tasks we should focus on, and how AI can be leveraged to achieve these.

Future work is dedicated to continue narrowing down on candidate AI techniques to develop for DUC. In doing so, we plan to continue exploring applications for establishing shared situation awareness and capabilities in relation to the following high-level tasks:

- Emergency management;
- Dynamic capacity management.

Thereto, two activities are planned:

- A follow-up workshop with a narrower focus limited to identifying concrete AI applications suitable for a subset of high-level tasks.
- Follow-up in-depth discussion with HAIKU AI develop partners to explore possibilities for developing a prototype AI application for DUC.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

4. Use Case 4: Digital and Remote Tower

4.1. Context

UC4 is developing ISA (Intelligent Sequence Assistant) to support tower controllers during high traffic, by optimizing runway usage and ensuring safe and efficient operations. The IA can suggest traffic sequences, optimize inbound and outbound flows, and reduce stress and the risk of errors.

In this use case, the AI component is divided into two main categories: Aircraft sequencing and Arrival time estimation. As these two are fundamentally two different problems to solve, each section will be discussed separately.

4.2. Aircraft sequencing

4.2.1. Context

The Aircraft Sequencing Problem in this use case, involves determining the optimal sequence in which aircraft should land or take off at an airport to maximize efficiency and safety while minimizing delays. It is a well-known problem in air traffic management and scheduling, which can be very complex to solve due to the dynamic and unpredictable nature of air traffic. There are several aspects that need to be considered to find the optimal solution. These aspects include numerous safety requirements (e.g., ensuring that there is a safe distance between aircraft during landing and take-off to avoid collisions and turbulence issues), runway capacity, maintaining a “fair” order of service and minimizing delays and waiting times, environmental considerations and other. It should be highlighted that what constitutes an optimal sequence is also not strictly defined in the sense that the optimal sequence in terms of runway utilization may seem unsafe to some ATCOs depending also on the environment and overall context or even personal preference, making aircraft sequencing a challenging problem.

4.2.2. State of the art

Traditional methods for sequencing aircraft rely on deterministic rules and algorithms based on first-come-first-served principles, often managed manually by air traffic controllers [6]. These methods can be effective but may struggle to optimize sequencing under varying traffic conditions and dynamic airspace environments. A common approach to the aircraft sequencing problem involves combinatorial optimization using Mixed-Integer Linear Programming (MILP), where binary decision variables determine aircraft pairwise order alongside continuous variables representing arrival times. Efforts to mitigate the computational complexity of combinatorial optimization include data-splitting techniques and analytical solutions specific to MILP [7]. Recent advancements [8,9] utilize heuristics and real-time Automatic Dependent Surveillance - Broadcast (ADS-B) data for dynamic sequence optimization. Dynamic programming approaches are explored in [10] to address fairness constraints, whereas stochastic formulations [11] introduce uncertainty into

This project has received funding by the European Union’s Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

arrival-time predictions, typically addressed through two-stage stochastic programming.

ML algorithms, such as Reinforcement Learning (RL), have been explored to learn optimal sequencing strategies in real-time, considering factors like aircraft type, speed, trajectory, and weather conditions. RL models can adapt and improve their strategies based on feedback from the environment, potentially leading to more adaptive and responsive sequencing decisions [12,13].
Hybrid approaches that combine ML with traditional optimization algorithms, such as Genetic Algorithms (GA) are also gaining traction [14] in cases where computational needs make the use of mathematical optimization inefficient. These hybrids leverage the strengths of both approaches to balance computational efficiency with predictive accuracy.

In HAIKU the selected approach is the use of mathematical programming. The formulation described in [15], which according to [16] is the most cited MIP model for aircraft sequencing, is adapted according to the operational constraints applicable to the use case and the Alicante airport. The cost function is also changed since the runway utilization and not the deviation from the defined CTOT/EOBT is the priority, notwithstanding the need to respect the defined schedule. Relaxation parameters are added to ensure that the problem never results in infeasibility, since this is the expectation in real life scenarios.

4.3. Arrival time estimation

4.3.1. Context

Predicting the landing time of aircraft approaching an airport is a crucial aspect of air traffic management. Accurate predictions of how much time an approaching aircraft needs until touch down are essential for efficient sequencing and scheduling of aircraft, minimizing delays, and ensuring safety. Various factors affect the arrival time estimation (ETA), including the flight trajectory (there is no unique trajectory), the altitude and speed profiles along the trajectories, the type of aircraft, the actual speed and position of the specific aircraft and the way these change over time, the weather, the airspace congestion, the estimated time to vacate the runway and others.

In the current context, ATCO instructions (e.g. based on current runway occupancy and schedule) that affect landing times are not in scope, i.e. the goal here is to predict the landing time if the aircraft is allowed to approach as it normally would without being instructed otherwise. The prediction of aircraft arrival times is a critical component of modern air traffic management systems, aiming to enhance the efficiency and safety of airspace operations.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

4.3.2. State of the art

State-of-the-art approaches leverage a combination of traditional modeling techniques and advanced machine learning algorithms. Traditional methods include physics-based models that simulate the aircraft's descent profile and arrival procedures, considering factors such as weather conditions, air traffic control (ATC) instructions, and aircraft performance characteristics. These models often utilize statistical techniques and Kalman filters [17] to update predictions based on real-time data. While effective, these approaches can struggle with the complexities and dynamic nature of real-world air traffic environments. It should be highlighted that research also focuses on integrating data from diverse sources, including weather forecasts, ATC communications, and aircraft sensor data, to provide comprehensive and accurate arrival time predictions [18].

In recent years, numerous machine learning (ML) techniques have emerged as powerful tools for improving arrival time predictions. Models such as Gradient Boosting Machines (e.g., XGBoost) and deep learning approaches (e.g., RNNs, LSTMs) are increasingly employed to capture non-linear relationships and temporal dependencies in the data [19–21]. Hybrid approaches that combine physics-based models with ML techniques [22] have shown promising results in enhancing prediction accuracy.

4.3.3. Requirements analysis

In UC4, the aircraft arrival time estimation is used to provide the necessary input to the aircraft sequencing system that was discussed previously. In particular, the time until arriving aircraft will reach the runway (if not instructed to enter a holding pattern/ go-around maneuver) constitutes information that ATCOs need to have in order to decide on the optimal sequence. Similarly, as described above, the aircraft sequencing system developed within HAIKU to help ATCOs with the sequencing requires this information as input. This information is not directly available, as in real life situations ATCOs mentally perform this calculation and the same is true for the simulator software used to implement and validate the HAIKU solution, i.e. the simulation software does not provide ETA until the aircraft is very close to landing, at which point in time the sequence has been already decided by the ATCO.
Thus, the system developed for arrival time estimation should:

- Predict arrival time in time granularity of seconds with minimal error. It should be noted that the error in this case should not be measured as a general average, since it is important that it decreases as aircraft approach the airport, e.g. an error of 30 seconds may be acceptable when the aircraft is 10 minutes from landing (in which case it will most probably not be part of the active sequence yet) but not when the aircraft is 1 minute from landing. At the same time, it is important to ensure that slight variations of the trajectories and speed patterns seen during model training will be properly handled during production use of the system, i.e. overfitting should be avoided.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

- Predict arrival time almost instantly after the updated position of an aircraft is received, since this new estimation will be both shown to ATCOs and used to update the aircraft sequence, which should be done quickly to ensure that ATCOs remain aware of the current system status.

From the state-of-the-art analysis it is evident that machine learning methods can be effectively employed to address the problem of predicting the aircraft arrival time. Although the training process might be time consuming, the application of the trained model to retrieve a prediction can happen instantly, hence the second requirement above is met by any machine learning model. To identify the most performant model, experiments were performed with linear regression, XGBoost and MLP. Based on the experiment results, XGBoost, known to be very powerful approach in implementing regression systems, was chosen.

4.3.4. Training and data process

The dataset used for training the model is proprietary, derived from simulator exercises used in ATCO training, specifically focused on Alicante airport. The dataset comprises approximately 750 arrival trajectories, which were split into 600 flights for the training set and 150 flights for the validation set. After data cleaning to remove incomplete trajectories or those with go-around maneuvers, the data was preprocessed into a tabular format.

Each row represents a point within an arrival trajectory, with a corresponding "label" indicating the time (in seconds) until the aircraft lands. Features used in the model include position (x, y, z), track, true airspeed, aircraft model, and flight rules (IFR, VFR). Additional features were engineered to capture the aircraft's distance from the runway, the distance from a specific convergence point in the airspace, and an indicator of approximate aircraft acceleration.

4.3.5. Evaluation

To test the developed models, approximately 150 flights from the simulator exercises were extracted and preprocessed similarly to the training and validation data. An important consideration is that as an aircraft approaches the runway, the "seconds until landing" labels decrease, and the need for accurate predictions becomes more critical—up to a point where the aircraft is only seconds away, and the ETA is no longer needed by ATCOs. Thus, it is crucial to not only measure the average error but also to understand the error distribution as the aircraft nears the runway. This is visualized in Figure 14, which displays the average Mean Absolute Error (MAE) across various ground distance bins. The average error is close to 25 seconds for aircraft that are approximately 12 minutes from landing and drops to under 10 seconds when aircraft are within five minutes from landing.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

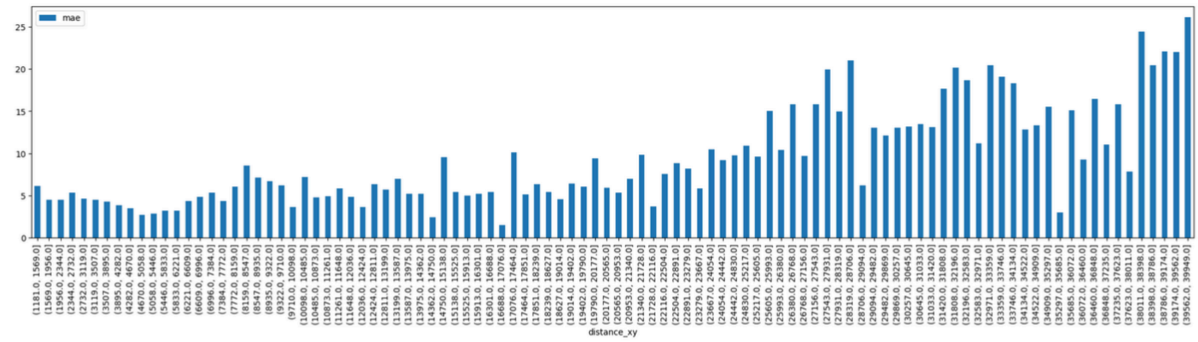


Figure 17: MAE across distance

4.4. Conclusion and future planning

The experimental results from the aircraft arrival time estimation system look promising both in terms of average error and in the way the error follows the trajectory (decreases along the trajectory as we move closer to the airport runway). Some simulated end to end runs (using simulated exercises) were performed to see how the results from the ETA system feed the sequencing problem, in order to gain some insights into the real-life validation of the complete system which is the one the ATCOs will interact with. These end to end tests will be the focus of the next steps when it comes to validation activities, and insights gained through them will be leveraged to further improve the system. Trajectory variations in future simulation exercises, as well as more challenging time-wise scenarios, will allow us to test and further improve the system's performance. Additionally, the current calculation method for taxi times (time it takes aircraft to move from a stand to the runway) will be evaluated. Taxi times are important inputs to the main aircraft sequencing system and affect the calculation of the optimal sequence. Currently these calculations are based on statistical computations, but the development of a learning system (to account for different behaviour depending on the contextual information) will be considered in case the current static results obtained per aircraft type and stand are not accurate enough and need to be improved.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

5. Use case 5: Airport safety management

5.1. Context

The goal of the assistant in this use case is to help identify key causal factors, crucial for risk reduction. London Luton Airport (LLA) anticipates that the visualization tool developed will enhance safety, data collection, categorization, analysis, and visualization.

Developing an intelligent digital assistant for safety data analysis in aviation encompasses several advanced technologies and methodologies. Different methodologies can be applied like data integration and interoperability. Modern systems emphasise the integration of data from multiple sources, ensuring comprehensive analysis. This includes data from various stakeholders at LLA, each contributing unique insights. Interoperable systems facilitate seamless data sharing and collaboration among the over 70 stakeholders operating airside at LLA.

5.2. Approach

At this stage there is no AI component development for the use case. Rather, the resulting Dashboard utilises Data Science techniques such as feature extraction to mine the LLA incident data for factors and look for meaningful correlations and clusters according to specific inquiries. The data from LLA includes information from over 900,000 flights between years 2016-2023. These include 563 unique incidents during the years 2016-2023 and information on those incidents concerning pushback errors, selection of wrong taxiway and holding-point busts, where full data reports for these incidents are provided. Moreover, information regarding the inbound/outbound flights data within the day are also provided. These data have been extracted from the LLA systems in a manual manner. Furthermore, weather conditions data, focusing on the LLA region, has been acquired from external sources.

At the current stage of the UC5 implementation, emphasis has been placed on the exploratory data analysis of the datasets coming from the LLA to better understand and to visualise the problem with possible patterns, around the safety incidents recorded at the airport. During the data ingestion step, data linking and matching has been performed that allowed data harmonisation.
As the amount of recorded incidents is expected to be increased over the next period by ingesting new data coming from LLA, UC5 will employ AI algorithms to continue the study related to pattern identification across the incidents and the context in which they happened, considering not only the raw data, as done at this point, but also the features created during the first phase, indicatively through clustering techniques. The final dashboard will include specific explainability aspects relevant to the incidents to better convey to safety experts, the main causes of the incidents as well as the overall surrounding conditions that may have had an impact to the realisation of the incident.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

5.3. Conclusion and future planning

There are three key future development plans for ASW:

1. Addition of ground handling (GH) collision incident data to the ASW system. This will add a fourth incident type, but will be a big change as GH factors and considerations are broader than for the three incident types already included, due to variety in the operations. This may enrich the system overall, possibly to the benefit of analysing the three existing incident types. Feature extraction may have to work harder for GH incident data.
2. Addition of a formal Human Factors taxonomy. At the moment the way human errors and contributory factors are described is variable and not always consistent across companies, nor 'deep' enough in incident causation terms to be useful in generating error incident reduction insights. In 2025 therefore, it is currently planned (pending Stack approval) to develop a common Human Factors taxonomy for GH, flight crew and ATC. The latte have their own taxonomies, so harmonisation will be required. Once this is implemented, it will be possible to search for error types or Human Factors elements both within and across the different incident types.
3. An API (Application Programming Interface) will be developed to enhance data ingestion. This will be implemented in 2025 by LLA.

These steps, as well as the addition of more incidents as time passes, will help pave the way for a future transition to a truer AI-based system for the airport.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

6. Use case 6: Airport virus spread prevention

6.1. Context

The Intelligent Assistant (IA) of the COVAID use case was developed to inform passengers about the likelihood of COVID-19 infection at the airport and to guide them through their visits to the airport's common areas. Beyond providing infection likelihood information, the IA also includes a chatbot that engages passengers in conversations about their journey to the airport both before the flight and during the routing process.

The core of the system involves an AI-driven algorithm that uses sensor data and passenger inputs to generate routing decisions aimed at reducing infection risks. Given the complexity of these decisions, it is vital to provide passengers with clear explanations of the routing logic. To enhance transparency and build trust, the system includes a chat application within the Android app, enabling passengers to ask questions and receive explanations about the AI's decisions.

Additionally, an air quality application is integrated to monitor indoor air quality, providing real-time statistics and classifications relevant to infection risks. This component ensures that airport personnel can maintain a healthy environment and that passengers are informed about the air quality.

6.2. State of the art

6.2.1. Routing and social distancing

[23] presents a deep learning framework designed to automate social distancing monitoring through surveillance video. The framework uses the YOLO v3 [24] object detection model to identify humans and distinguish them from the background, coupled with the Deepsort algorithm for tracking individuals. This combination outperforms other models like Faster R-CNN and SSD in terms of mean average precision (mAP) and frames per second (FPS), offering a robust solution for real-time monitoring. Additionally, a violation index based on the pairwise vectorized L2 norm quantifies non-compliance with social distancing protocols.

Further advancements are seen in [25], which introduces a hybrid model using YOLOv4 and Computer Vision techniques for people detection in crowds. This model, enhanced by inverse perspective mapping (IPM) and the SORT tracking algorithm, excels in both indoor and outdoor environments, achieving high precision and real-time processing speed. The model's online infection risk assessment scheme further aids in identifying high-risk zones for virus transmission, helping authorities redesign public spaces to mitigate risks.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

6.2.2. Chatbot Integration

In terms of chatbot technology, [26] focuses on AI-based tools for preliminary profiling in Human Resources, utilizing deep learning architectures like recurrent neural networks. Similarly, [27] discusses a web-based chatbot designed to answer inquiries about college facilities and policies. This model, implemented using the Flask framework, employs a retrieval-based approach to enhance user experience, showing a 20% improvement in performance and a 5% increase in accessibility after its integration.

6.2.3. Air quality classification

Air quality monitoring, essential for maintaining a healthy environment in airports, is explored in [28] and [29]. [28] introduces IndoAirSense, a framework that uses low-cost sensors and machine learning techniques like Multi-Layer Perceptron (MLP) and eXtreme Gradient Boosting Regression (XGBR) for real-time indoor air quality estimation. This approach achieves high accuracy in both estimation and forecasting.

Complementing this, [29] highlights the need for comprehensive air quality evaluations in densely populated areas like metro stations. The study proposes an innovative method combining optimal combination weighting with enhanced fuzzy comprehensive evaluation, overcoming the limitations of traditional methods. This approach proves more effective in accurately assessing air quality, particularly in complex environments.

6.3. Requirements analysis

The AI system architecture is designed with a modular approach, consisting of three independent yet interlinked sub-models. Each sub-system contributes to the overall functionality of the AI model, ensuring a cohesive and efficient operation.

Machine Learning Sub-Systems: Each sub-model operates independently, handling specific tasks such as data processing, model execution, and interaction with the Android application and database. These machine learning algorithms are implemented in Python and hosted on cloud infrastructure. The cloud environment not only stores pre-trained models but also manages data storage and facilitates seamless communication between the system's various components. In order for the COVAID tool to operate successfully the ML algorithms operate in a determined manner that handles the incoming data from the sensors and the passenger preferences.

Android Application: Serving as the user interface, the Android application is integral to the system, offering real-time notifications, facilitating chat interactions, and displaying air quality metrics to users. The HMI of the Android application is essentially the interface of the AI engine that resides in the cloud.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Air Quality Sensors: These sensors collect indoor environmental data, crucial for monitoring and maintaining optimal air quality levels.

Camera Sensors: Cameras are deployed to monitor occupancy and queue lengths within the airport, feeding data into the AI system for processing.

Passenger Mobile Apps: The system also gathers user preferences and inputs through the mobile applications used by passengers, personalizing the AI-driven responses and recommendations.

The cloud-hosted architecture ensures that all components—air quality sensors, camera sensors, and mobile apps—work together seamlessly, contributing to the system's holistic functionality. The architecture's modular design, as depicted in the generic architecture diagram (Figure 15), underscores its flexibility and scalability, enabling efficient data processing and real-time user interaction.

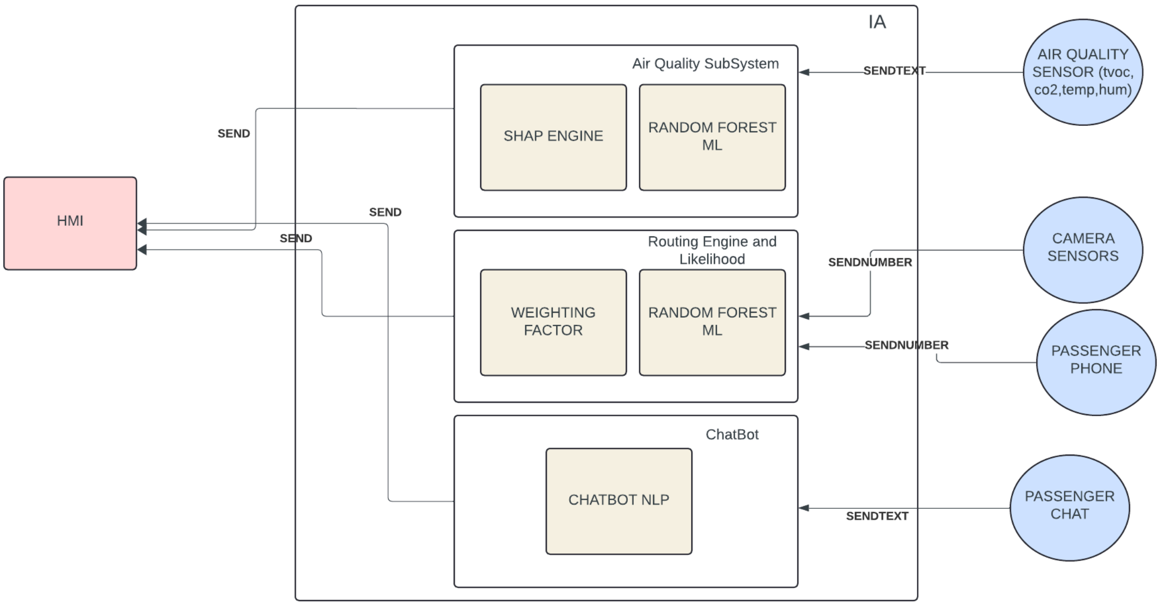


Figure 18: Generic architecture of modules

6.4. Data processes and testing

For the passenger application, an artificial dataset was designed to classify the likelihood of COVID-19 infection into two categories: high and low. The dataset consists of 27 features, corresponding to nine common areas within the airport,

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

including "Shop1," "Shop2," "Shop3," "Toilet1," "Toilet2," "Toilet3," "Coffee Shop," "Restaurant," and "Canteen." For each area, three specific features were generated:

- "<common_place>_coming": The number of people heading to the area.
- "<common_place>_queue": The number of people queuing.
- "<common_place>_in": The number of people currently inside the area.

Additional areas can be incorporated into the dataset as needed. The dataset, containing 2,000 entries in CSV format, was designed to simulate realistic conditions with patterns indicating high COVID-19 infection risk and includes a margin of error to ensure robustness.

The weighting factor within the model was rigorously tested, though not explicitly trained, by injecting known values into the dataset and verifying the correctness of the outputs. Multiple test datasets were created with specific injected scenarios, covering various conditions such as different levels of occupancy, queue lengths, and the number of people approaching common areas. The algorithm's outputs were compared against expected results to confirm correct identification of optimal routing sequences based on the inputs.

The chatbot's Natural Language Processing (NLP) capabilities were thoroughly evaluated by populating its knowledge base with a comprehensive set of Q&As. Various questions were posed to test the retrieval accuracy, and similar questions were used to assess the algorithm's ranking capability. This ensured the system could differentiate between similar queries, such as whether a toilet is full or not. Additionally, the time-sensitive nature of responses was tested, ensuring that the system could update its datetime field every 10 minutes to capture real-time responses. The chatbot algorithm was tested both as a standalone component and as part of the integrated AI system, successfully performing expected functions and updating rankings based on passenger interactions.

The air quality dataset utilized in the study was sourced from a publicly available dataset on Kaggle. To develop a robust model, the dataset was split into training and testing sets, with 80% allocated for training and 20% for testing. This approach allowed the model to learn from a diverse range of data points while retaining a portion of the data for performance evaluation on unseen cases. Throughout the process, the algorithm exhibited notable efficiency in processing, demonstrating quick convergence and effective use of computational resources. This ensured the model's ability to generalize and make accurate predictions based on real-time data.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

6.5. Results

In terms of the ML classifier for the COVID-19 infection likelihood at the airport he Random Forest model gave the following results, with 2000 samples training with 10 and 100 decision trees:

- Accuracy with 10 decision trees: 0.9984
- Accuracy with 100 decision trees: 0.9984

The visiualisation of the most important features of the algorithm are given in Figure 19. We expect this data visualisation to change with the creation of the dataset from the deployed devices and the Android application at the testbed.

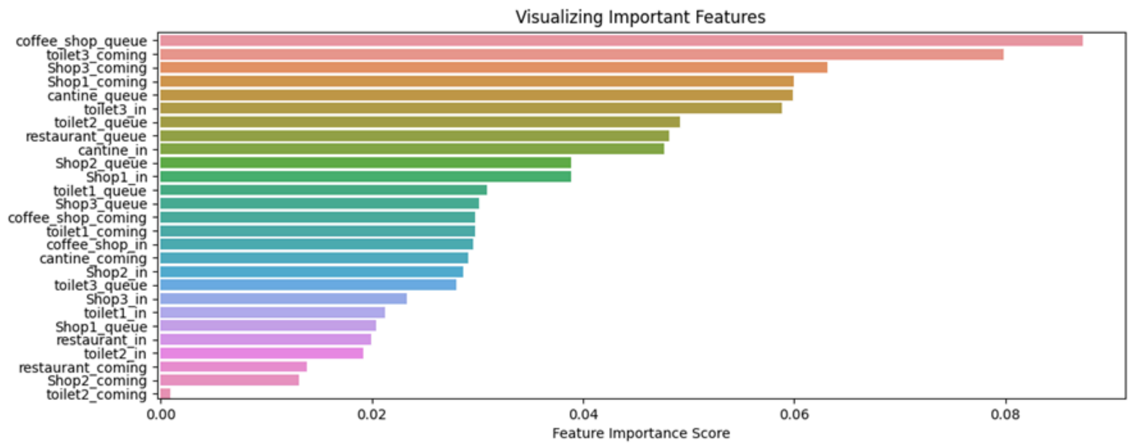


Figure 19:Infection ML algorithm importance of features

The air quality classifier is also based on the Random Forest model and it exhibited the following accuracy results using 10 and 100 decision trees as well. Note that the dataset used is available online in Kaggle encompassing 135,100 entries across seven key features: datetime, CO2 levels, humidity, PM10 and PM25 particulate matter concentrations, temperature, and VOC (volatile organic compounds).

- Accuracy with 10 decision trees: 1.00
- Accuracy with 100 decision trees: 1.00

The dataset has been clustered with k-Means clustering and using the elbow method we identified the optimal cluster of the data. The importance of the features is given in Figure 20. The unnamed is the datetime field.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

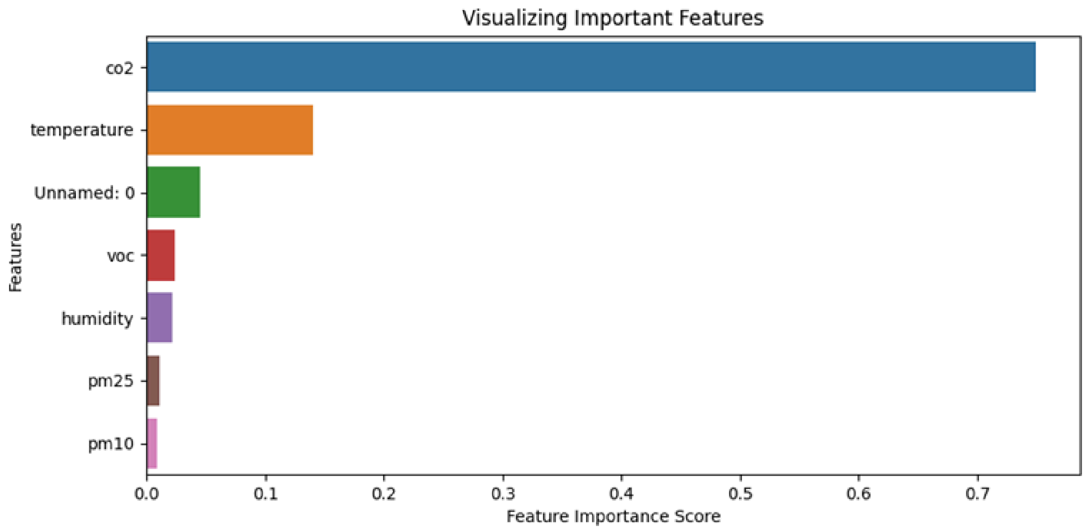


Figure 20: Air quality ML algorithm importance of features

For the routing algorithm, we show its proper functionality in the implementation deliverable. However, for clarity, we provide results using an online tool that performs HTTP GET requests to the web service of our AI engine. The input parameter is the string with 1s and 0s showing that the passenger pressed the respective button for a common place or not, respectively. Note that the output shows the weights from the smallest to the largest followed by the places from the first to be visited to the last. The result is given in Figure 23 in the CLT explanation. We omit the HMI display since it has been also given to other deliverables.

In terms of the NLP engine, the results are given in deliverable 4.4 in its updated version. For clarity we repeat some of the results in this deliverable to show the AI model. Here we show the Q&A between the chatbot and the passenger as well as some CLT levels that have been implemented. These can be seen in Figure 22 and 23 respectively.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

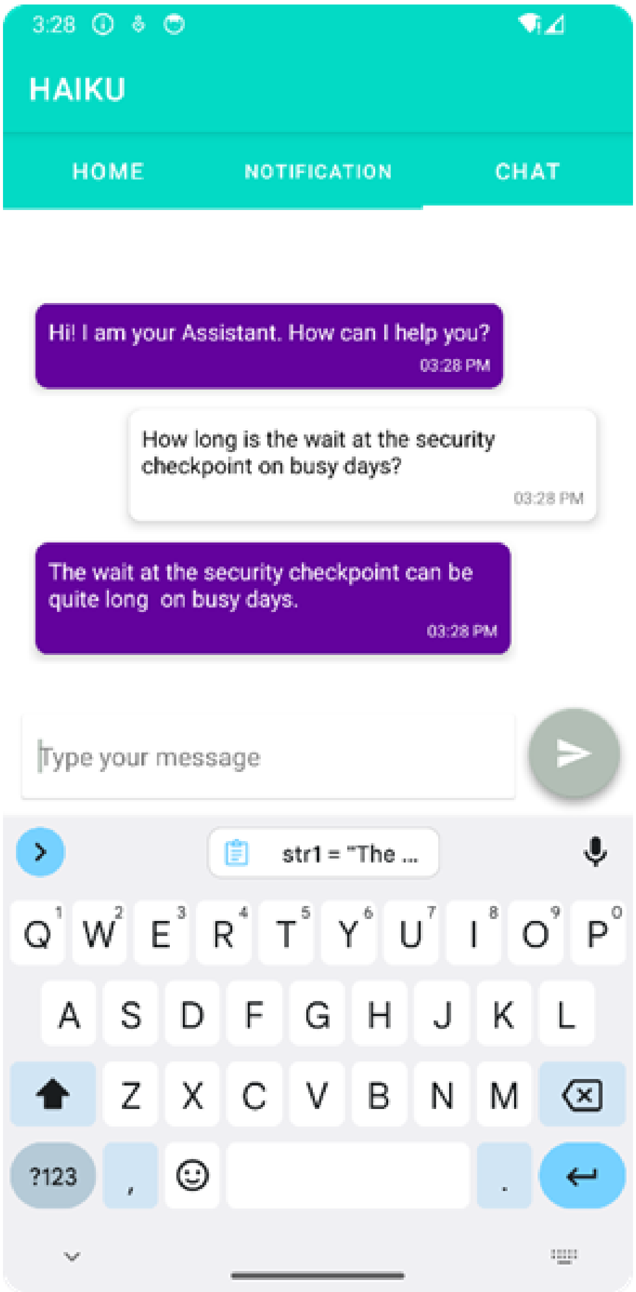


Figure 22: Chatbot and passenger Q&A

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

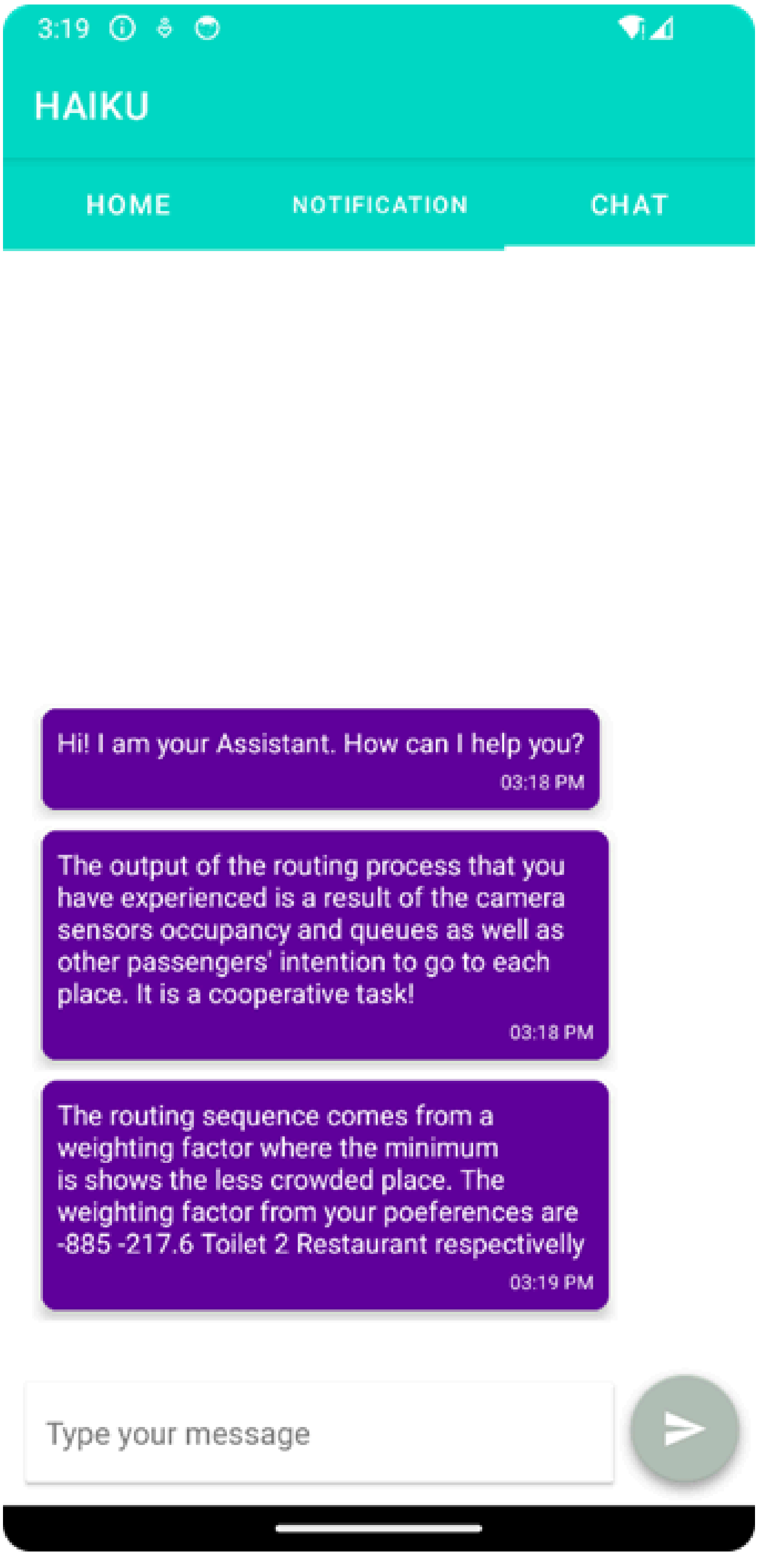


Figure 23: CLT2 and 3

6.6. Conclusions and Future Planning

The COVAID IA is essentially a set of integrated tools and hardware that operate in order to reduce the likelihood of infection by providing routing recommendations. The

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

tool comprises an AI engine that resides in the cloud, whereby the ML classification algorithm for COVID 19 infection likelihood and the intel;ligent routing algorithm reside. These tools get input from camera sensors and Android mobile phone applications to determine the optimal routing sequence and inform the passenger regarding the likelihood of infection. Moreover, the IA includes a NLP engine which engages with Q&As with the passenger to enhance the trust in the tool. The goal of the NLP engine is to provide not only answers but to provide sensor data as well for better information to the passenger and check for irrationality or even changing the response by the chatbot. The ML for the air quality classification together with the SHAP explainability plots have been tested with a given dataset and works satisfactory.
The future planning of the IA engine are the following:

- Validate the routing algorithm and the ML algorithm in a testbed at the Amygdaleona airport in Kavala, Greece,
- Populate the knowledge base with more Q&As that reflect the state of the common places that the passenger has or going to visit.
- Provide more data from the AI engine to the HMI in order to have routing algorithm metrics explanations to the passenger.
- Implement sensor data being added to the responses of the IA to the passenger by enhancing the NLP engine.
- Use further explanations to the operator regarding the air quality.
- Produce datasets from the COVAID tool based on the sensors.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Conclusion

As the Use Cases progress further, the AI models will be fine tuned and optimized accordingly. If necessary, new datasets may be introduced to improve the overall performance, which would require retraining of the existing models and performing new testing and validation. Meanwhile, for Use Cases 3 and 5, we will explore the opportunities of implementing an AI model, in areas where the greatest potential is evident.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

References

[1] Johnson E. Reducing Commercial Aviation Fatalities 2018. https://kaggle.com/competitions/reducing-commercial-aviation-fatalities (accessed July 19, 2024).
[2] Khemka S, Tzovara A, Gerster S, Quednow BB, Bach DR. PsPM-SMD: SCR, EMG, ECG, and respiration measurement in response to auditory startle probes 2019. https://doi.org/10.5281/zenodo.3430920.
[3] Ojala KE, Bach DR. PsPM-PCF2: PSR, SCR, ECG, respiration and startle-eyeblink EMG measurements in a delay fear conditioning task with 4 CS and different reinforcement rates 2020. https://doi.org/10.5281/zenodo.7313441.
[4] Tzovara A, Castegnetti G, Gerster S, Hofer N, Khemka S, Korn CW, et al. PsPM-FR: SCR, ECG and respiration measurements in a delay fear conditioning task with visual CS and electrical US. 2021. https://doi.org/10.5281/zenodo.5573765.
[5] Harrivel AR, Stephens CL, Milletich RJ, Heinich CM, Last MC, Napoli NJ, et al. Prediction of Cognitive States during Flight Simulation using Multimodal Psychophysiological Sensing. AIAA Information Systems-AIAA Infotech @ Aerospace, Grapevine, Texas: American Institute of Aeronautics and Astronautics; 2017. https://doi.org/10.2514/6.2017-1135.
[6] Kamo S, Rosenow J, Fricke H, Soler M. Robust optimization integrating aircraft trajectory and sequence under weather forecast uncertainty. Transportation Research Part C: Emerging Technologies 2023;152:104187. https://doi.org/10.1016/j.trc.2023.104187.
[7] Prakash R, Piplani R, Desai J. An optimal data-splitting algorithm for aircraft scheduling on a single runway to maximize throughput. Transportation Research Part C: Emerging Technologies 2018;95:570–81. https://doi.org/10.1016/j.trc.2018.07.031.
[8] Pradeep P, Wei P. Heuristic Approach for Arrival Sequencing and Scheduling for eVTOL Aircraft in On-Demand Urban Air Mobility. 2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC), 2018, p. 1–7. https://doi.org/10.1109/DASC.2018.8569225.
[9] Xue D, Hsu L-T, Wu C-L, Lee C-H, Ng KKH. Cooperative surveillance systems and digital-technology enabler for a real-time standard terminal arrival schedule displacement. Advanced Engineering Informatics 2021;50:101402. https://doi.org/10.1016/j.aei.2021.101402.
[10] Miyazawa Y, Wickramasinghe NK, Harada A, Miyamoto Y. Dynamic Programming Application to Airliner Four Dimensional Optimal Flight Trajectory. AIAA Guidance, Navigation, and Control (GNC) Conference, American Institute of Aeronautics and Astronautics; n.d. https://doi.org/10.2514/6.2013-4969.
[11] Khassiba A, Bastin F, Gendron B, Cafieri S, Mongeau M. Extended Aircraft Arrival Management Under Uncertainty: A Computational Study. Journal of Air Transportation 2019;27:131–43. https://doi.org/10.2514/1.D0135.
[12] Churchill A, Coupe WJ, Jung YC. Predicting Arrival and Departure Runway Assignments with Machine Learning. AIAA AVIATION 2021 FORUM, American Institute of Aeronautics and Astronautics; n.d. https://doi.org/10.2514/6.2021-2400.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

[13] Razzaghi P, Tabrizian A, Guo W, Chen S, Taye A, Thompson E, et al. A survey on reinforcement learning in aviation applications. Engineering Applications of Artificial Intelligence 2024;136:108911. https://doi.org/10.1016/j.engappai.2024.108911.
[14] Sylejmani K, Bytyçi E, Dika A. Solving aircraft sequencing problem by using genetic algorithms. Intelligent Decision Technologies 2017;11:451–63. https://doi.org/10.3233/IDT-170309.
[15] Beasley JE, Krishnamoorthy M, Sharaiha YM, Abramson D. Scheduling Aircraft Landings—The Static Case. Transportation Science 2000;34:180–97. https://doi.org/10.1287/trsc.34.2.180.12302.
[16] Ikli S, Mancel C, Mongeau M, Olive X, Rachelson E. The aircraft runway scheduling problem: A survey. Computers & Operations Research 2021;132:105336. https://doi.org/10.1016/j.cor.2021.105336.
[17] Lee J, Lee S, Hwang I. Hybrid System Modeling and Estimation for Arrival Time Prediction in Terminal Airspace. Journal of Guidance, Control, and Dynamics 2016;39:903–10. https://doi.org/10.2514/1.G001412.
[18] Ayhan S, Costas P, Samet H. Predicting Estimated Time of Arrival for Commercial Flights. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, New York, NY, USA: Association for Computing Machinery; 2018, p. 33–42. https://doi.org/10.1145/3219819.3219874.
[19] Basturk O, Cetek C. Prediction of aircraft estimated time of arrival using machine learning methods. The Aeronautical Journal 2021;125:1245–59. https://doi.org/10.1017/aer.2021.13.
[20] Dhief I, Wang Z, Liang M, Alam S, Schultz M, Delahaye D. Predicting Aircraft Landing Time in Extended-TMA Using Machine Learning Methods, 2020.
[21] Wang G, Liu K, Chen H, Wang Y, Zhao Q. A High-precision Method of Flight Arrival Time Estimation based on XGBoost. 2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT, 2020, p. 883–8. https://doi.org/10.1109/ICCASIT50869.2020.9368723.
[22] Strottmann Kern C, de Medeiros IP, Yoneyama T. Data-driven aircraft estimated time of arrival prediction. 2015 Annual IEEE Systems Conference (SysCon) Proceedings, 2015, p. 727–33. https://doi.org/10.1109/SYSCON.2015.7116837.
[23] Punn NS, Sonbhadra SK, Agarwal S, Rai G. Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques 2021.
[24] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement 2018. https://doi.org/10.48550/arXiv.1804.02767.
[25] Rezaei M, Azarmi M. DeepSOCIAL: Social Distancing Monitoring and Infection Risk Assessment in COVID-19 Pandemic. Applied Sciences 2020;10:7514. https://doi.org/10.3390/app10217514.
[26] Sheikh SA, Tiwari V, Singhal S. Generative model chatbot for Human Resource using Deep Learning. 2019 International Conference on Data Science and Engineering (ICDSE), 2019, p. 126–32. https://doi.org/10.1109/ICDSE47409.2019.8971795.
[27] Akkineni H, Lakshmi PVS, Sarada L. Design and Development of Retrieval-Based Chatbot Using Sentence Similarity. In: Nayak P, Pal S, Peng S-L, editors. IoT and Analytics for Sensor Networks, Singapore: Springer; 2022, p. 477–87. https://doi.org/10.1007/978-981-16-2919-8_43.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

[28] Sharma PK, Mondal A, Jaiswal S, Saha M, Nandi S, De T, et al. IndoAirSense: A framework for indoor air quality estimation and forecasting. Atmospheric Pollution Research 2021;12:10–22. https://doi.org/10.1016/j.apr.2020.07.027.
[29] Peng Z, Su P, Chen W. Indoor Air Quality Assessment of Metro Stations Based on the Optimal Combination Weight and Improved Fuzzy Comprehensive Evaluation. Journal of Environmental Engineering 2023;149:04023001. https://doi.org/10.1061/JOEEDU.EEENG-6944.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332