Deliverable N.5.1

Strategies for XAI in aviation

Authors:
Narek, Minaskan (DFKI)

Abstract:
Task 5.1 of HAIKU focuses on explainable AI and defines implementation strategies for the Use Cases. In this deliverable a general overview of XAI together with strategies for the Use Cases is presented.

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	5.1
Deliverable Title	XAI strategies
Version	1.0
Status	Draft
Responsible Partner	DFKI
Contributors	DBL, ENAC, THALES, LFV, LiU, Skyway, Suite5, Embraer, ECTL
Contractual Date of Delivery	August 31st, 2023
Actual Date of Delivery	August 30th, 2023
Dissemination Level	Public

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	31.07.2023	First outline	Minaskan N. (DFKI)	Request of partners' review
0.2	10.08.2023	In review	Arrigoni V. (DBL)	Review by DBL
0.3	18.08.2023	In review	Pozzi S. (DBL)	Review by DBL
0.4	22.08.2023	In Review	Olivio P. & Villani A. (EMBT)	Review by EMBT
0.5	24.08.2023	In review	Minaskan M. (DFKI)	Review Integration
1.0	25.08.2023	Final version	Arrigoni V. (DBL)	Final quality check & Final version

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
AI	Artificial Intelligence
AOCC	Airport Operation Control Center
ATC	Air Traffic Controller
CLT	Construal Level Theory
DARPA	The Defense Advanced Research Projects Agency
DL	Deep Learning
DNN	Deep Neural Network
ETA	Estimated Time of Arrival
FMS	Flight Management System
HAIT/HAT	Human-AI Teaming
HCI	Human-Computer Interaction
IA	Intelligent Assistant
ISA	Intelligent Sequence Assistant
KPI	Key Performance Indicator
LIME	Local Surrogate Models
ML	Machine Learning
SHAP	SHapley Additive exPlanations
SVM	Support Vector Machine
SWIM	System-wide Information Management
TWR	The Tower Controller
UAM	Urban Air Mobility
UAS	Unmanned Aircraft Systems

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

UATM	Urban Air Traffic Management
UX	User Experience
XAI	Explainable Artificial Intelligence
XUI	Explainable User Interface

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

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

Task 5.1 of HAIKU aims to generate strategies of XAI (Explainable Artificial Intelligence) for the use cases defined in the scope of the project. The first step to achieve this goal was to look into concepts and methods of XAI in a broader scope and gather implementation methodologies. When designing an AI-powered system, the requirement analysis for XAI is usually conducted in activities related to human-AI teaming (HAIT/HAT). The variability in needs of XAI, makes it almost impossible to have a universal methodology for requirement analysis, rather each case must be analyzed and evaluated on its own.

That said, it is not impossible to provide methodologies and grouping for XAI, especially in similar applications across domains. Two main categories which were investigated are:

- Interaction with XAI: it focuses on interaction between a human and a potential AI system,
- Methods of building XAI models: it studies the currently available methods of training and building XAI models. This kind of categorization provides an overview of existing methods and trends which can be adopted for any case.

Nevertheless, this approach does not provide a requirement analysis for use cases to generate strategies. To that end, construal level theory (CLT) was adopted to provide a detailed insight into the use cases. The results were described in a story-based structure for each use case.

All the results together provide a sufficient pool of information to generate strategies, implementation plans, and evaluation methodologies for the use cases.

As part of Task 5.1, one webinar on XAI was organized, to build shared understanding in the consortium.

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 to explainable AI (XAI)

Perhaps one of the earliest research attempts in XAI goes back to The Defense Advanced Research Projects Agency (DARPA) XAI program (DW, 2019) which was focused on developing AI systems that can provide meaningful explanations for their decision-making process. Several objectives for XAI were mentioned in the program such as enhancing trust, transparency, and accountability in AI systems by enabling human users to understand the rationale behind AI-generated outputs. In high-stakes scenarios where the consequences of the AI decisions can have significant impacts, explainability becomes crucial for operational confidence and effective decision-making in unpredictable situations, perception and reasoning, judgment and recognition

An example of XAI, is perhaps easiest to come by in visual classification and object detection. Vasu et al. (Vasu et al., 2021) used saliency maps (or heat maps) to highlight regions of interest for the users and incorporating user feedback (human-in-the-loop) (Figure 1) in the form of a classifier which re-ranks the results.

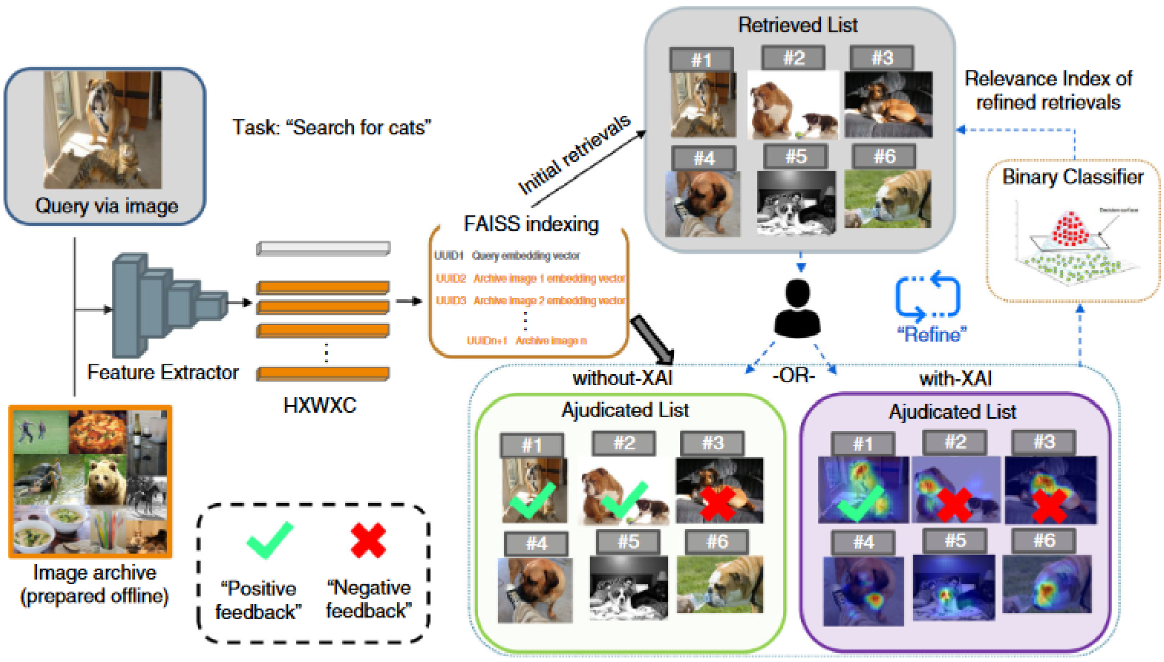

Figure 1: Saliency maps for image classifications proposed by Vasu et al.

However, as task complexity grows, so do the requirements for XAI. In essence, it is important to identify the answers to the following questions:

- What is being explained?
- Who is being explained to?
- How is it being explained?

The answers seem easy to find, however with increased complexity of use cases and special conditions surrounding each (e.g. Human-AI teaming principles etc.) it is not

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

possible to find universal answers for these questions. Regardless, some of DARPA's key takeaway messages were useful as general principles as to what qualities are good to have in the answers:

- For the explanation to improve performance and safety objectives during the development and operations, the task must be difficult enough that the AI explanation helps.
- User performance may be hindered by the cognitive load required to interpret explanations.
- Measures of explanation effectiveness can change over time.
- Advisability can improve user trust significantly over explanations alone.
- Users prefer systems that provide decisions with explanations over systems that provide only decisions. Tasks where explanations provide the most value are those where a user needs to understand the inner workings of how an AI system makes decisions.

Close collaboration across multiple disciplines is crucial for the effective development of XAI techniques. This collaborative effort involves disciplines such as computer science, machine learning, artificial intelligence, human factors, psychology, safety assessment, development process assurance and more. However, it can be challenging to foster such collaboration, as researchers often tend to focus on their individual domains and may need encouragement to work across disciplines.
The field can be categorized into two broad categories: model transparency and post-hoc explanations (Xu et al., 2019)(Figure 2).

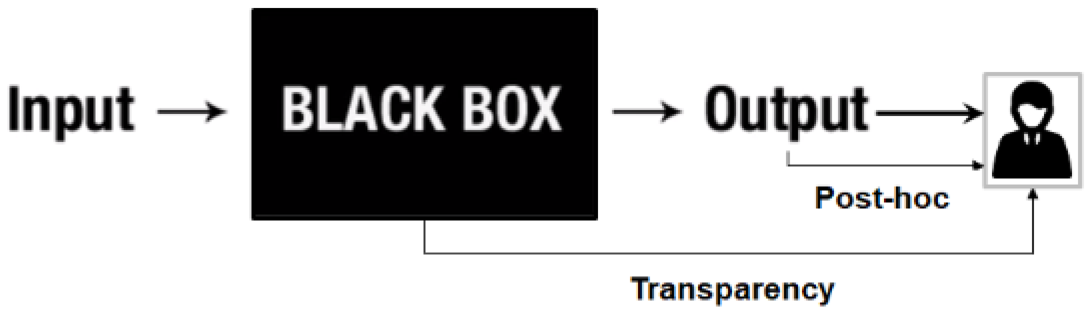

Figure 2: Two groups of Explainable AI: transparency and post-hoc

Model transparency methods aim to address the challenge posed by black box models, which lack interpretability and explainability. These methods focus on designing AI models that inherently provide interpretable and explainable outputs, making the decision-making process transparent from the model's architecture and internal mechanism. It can be argued that transparency method often involves using simpler models, such as decision trees or rule-based systems, that offer clear and understandable reasoning.

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

On the other hand, **post-hoc explanations** involve providing explanations for complex AI models that are already in use, without modifying their internal structure. These techniques analyze the model's behavior and generate explanations retrospectively, offering insights into why certain decisions were made.
Both approaches have practical utility and can be implemented in real-world scenarios, depending on the specific requirements and constraints of the application. By understanding these different categories, researchers and practitioners can explore and deploy the most suitable XAI techniques to enhance transparency, trust, and interpretability in AI systems while tackling the black box model problem.

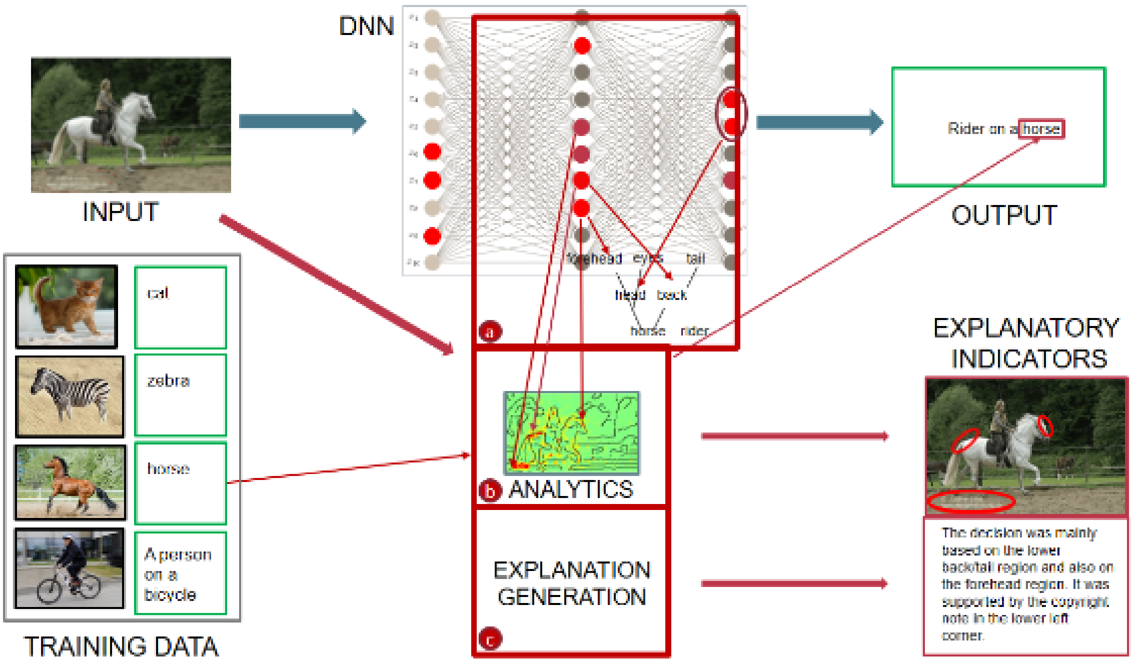

Figure 3: Three types of explanation for a DNN. a) Networks model transparency b) Semantics of the network component, neuron activation and object parts c) Human readable explanations

Figure 3 demonstrates the abstract ideas of transparent and post-hoc methods in a Deep Neural Network (DNN). The network's functionality and black-box structure is displayed in the form of tree-like structures and neuron activation effect whereas the explanation about the output is described in text format to the user.
There are other ways of interpreting XAI and the concept of explanation. For instance Baber et al. (Baber et al., 2021a, 2021b) developed a model for XAI in which it demonstrates the occurrence of different types of explanations, each requiring distinct forms of support. In essence, an explanation involves a consensus between the explainer and explainee on the features in data sets or a situation that demands attention and why these features are relevant. To capture the levels of relevance, three tiers are proposed:

- 'Cluster,' where a group of features typically co-occur
- 'Belief,' defining reasons for the occurrence of such a cluster;

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

- 'Policy,' justifying the belief and linking it to action.

The agreement on features and relevance relies heavily on the knowledge and experience of both the explainer and explainee, necessitating alignment during the explanation process. Thus, 'Explanation' becomes the process through which common ground is established and upheld. Based on their framework, the following guidelines are suggested:

- Explanations should emphasize Relevance by illustrating the relationship between features of a situation and the event being explained, remaining plausible according to a concept of Relevance agreed upon by the Explainer and Explainee.
- Explanations should encompass relevant Features by reaching a mutual understanding between the Explainer and Explainee regarding key features of the situation.
- Explanations should be Framed to suit the audience, aligning the explanation with the explainee's comprehension of the situation and objectives.
- Explanations should be interactive, actively involving the explainee in the explanation process.
- Explanations should be (where appropriate) actionable, providing the explainee with information that can be used to enhance future actions and behaviors.

After examining the resources available for XAI and considering the diverse use cases, our initial focus was on identifying methods with practical utility. This approach led us to uncover concepts and interaction methods with XAI, along with adaptations to existing ML/DL frameworks for building explainable AI solutions, which were presented and discussed in a webinar. Moreover, to adopt a unifying storytelling approach for requirement analysis across all use cases, **Construal Level Theory (CLT)** was utilized. By applying CLT, we described the function of XAI in each use case based on the nature of the operation, providing valuable insights into tailoring XAI solutions to specific application domains.

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

Methodologies for XAI

The contents in this section were presented in detail in the XAI webinar conducted within the HAIKU Project to provide a better understanding of the concept and build a common basic knowledge on XAI within the consortium members. It was mainly focused on practical aspects of XAI, namely interaction and existing frameworks. XAI as a concept in human-AI teaming, can be analyzed in several ways, each focusing on the one or several aspects of a use case. The focus in Task 5.1 was kept on operational explainability, which excludes the black-box model explainability.

Interaction with XAI

Interaction with XAI refers to the process of engaging with and utilizing the explanations generated by AI systems to gain insights into their decision-making process. When AI systems provide explanations for their decisions or predictions, users can interact with these explanations to better understand the underlying factors and reasoning behind the AI's output. The interaction concepts are derived from the classic HCI (Human-Computer Interaction)(Hornbæk and Oulasvirta, 2017) and adopted to AI (Chromik and Butz, 2021):

Interaction as transmission focuses on maximizing the flow of information through a noisy channel. The interaction involves choosing the most suitable message for transmission from a set of available messages. As an example, Alqaraawi et al. (Alqaraawi et al., 2020) explores the effectiveness of saliency map explanations for convolutional neural networks through a user study. The paper specifically focuses on evaluating the quality and usefulness of saliency map explanations from a user perspective. It examines how well users can interpret and trust these explanations, as well as their impact on users' understanding of CNN predictions.

Interaction as dialogue describes a cyclical communication process between a computer and a human. The interaction occurs in stages, with the aim of ensuring a correct mapping between user intentions and computer functions. It acknowledges that a single explanation may not be enough for full understanding, and instead, emphasizes providing natural and accessible explanations. An example would be Kim et al. (Kim et al., 2020) in which they introduce an Explainable User Interface (XUI) that allows users to ask factoid questions about charts using natural language.

Interaction as control tries to achieve a fast and steady convergence of the human-computer system towards a desired state. Inspired by control theory, the interaction aims to adjust a control signal to reach a specific level and continuously update its behavior based on feedback. It has two separate target groups: AI Experts and AI Novices. For AI experts, explanations are provided on an abstract level using numbers and visualizations and XUI is a standalone application that facilitates this interaction whereas XUIs aim to effectively communicate technical features of the ML model to non-technical users.

Interaction as Experience attempts to understand human expectations towards a computer system. It is closely linked to user experience (UX), which includes person's

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

emotions, feelings, and thoughts formed before, during, or after interacting with the system. Yin et al (Yin et al., 2019) for instance, demonstrates that a user's trust is influenced by prior information about the AI's predictive accuracy, even after multiple interactions.

Interaction as optimal behavior centers on adapting user behavior to align with their tasks and goals during system interactions. It acknowledges the constraints and tradeoffs users face in these interactions and is based on the principle of bounded rationality, in order to avoid seeking solutions due to cognitive limitations rather than optimal ones.

Interaction as tool use enhances the user's abilities beyond the capabilities of the tool itself. AI can serve as a tool for learning, particularly in the context of XAI. This interaction concept enables humans to discover concealed patterns and valuable insights in domain-specific data. To support this learning process, explanations play a crucial role and are necessary. Paudyal et al. (Paudyal et al., 2020) presents an interactive XUI for a computer-vision-based sign language AI where the textual explanations provide learners with feedback on the location, shapes, and movements of their hands.

Interaction as Embodied action is mainly reflected as a symbiotic relationship with autonomous systems. Real-time communication of capabilities and intentions becomes essential for achieving common goals. XUIs enable bidirectional influence between humans and AI agents. Tabrez et al. (Tabrez et al., 2019) introduce an AI agent that analyzes a human collaborator's game decisions in a collaborative game. The AI agent verbally interrupts the human if the common goal becomes unattainable due to a wrong move. It dynamically builds a theory of mind of the human and offers tailored explanations to correct their understanding of the game situation.

XAI models

In practice, XAI models are a collection of ML/DL models trained with varying network structures. They help understand the decisions or predictions made by the models and provide insight into the relationship between input and output of the models. Therefore, it is important to have terminology which describes the characteristics of such models.

Interpretability and explainability are two key aspects of understanding AI models and their decision-making processes.

- **Interpretability** refers to a model's ability to associate a cause with an effect, providing insight into why a certain decision was made. A model is considered interpretable when it can consistently produce the same outputs for the same inputs, allowing users to comprehend the relationships between input features and the resulting predictions. On the other hand, explainability goes beyond interpretability and involves identifying the specific features within the interpretable domain that contributed to a particular decision for a given

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

example. It encompasses the collection of features that played a role in producing the decision outcome.

- **Explainability** delves into the finer details of the decision-making process, providing a more comprehensive understanding of which features influenced the model's output for a specific instance.

That said, certain models are considered to have intrinsic interpretability (white-box models), which means that they possess inherent interpretability due to their simple and transparent structure. These models are designed in a way that facilitates human comprehension of their decision-making process, making them inherently interpretable without the need for additional post hoc analysis. The simplicity of their architecture allows for a clear understanding of the relationships between input features and output predictions, enabling users to gain insights directly from the model's design.
An example for an intrinsic interpretable model are decision trees (Figure 4). These models follow a specific set of rules, creating a tree-like structure with nodes representing decision points and edges representing the flow of decisions. Decision trees are inherently explainable due to their transparent structure, making it easier to understand how they arrive at specific predictions. One way to visualize the interpretability of decision trees is through tree decomposition, which involves breaking down the decision paths and tracking the rules followed at each node. This visualization helps users comprehend the sequence of decisions made by the model to reach the final prediction. Additionally, decision trees provide a feature importance metric, calculated using variance analysis.

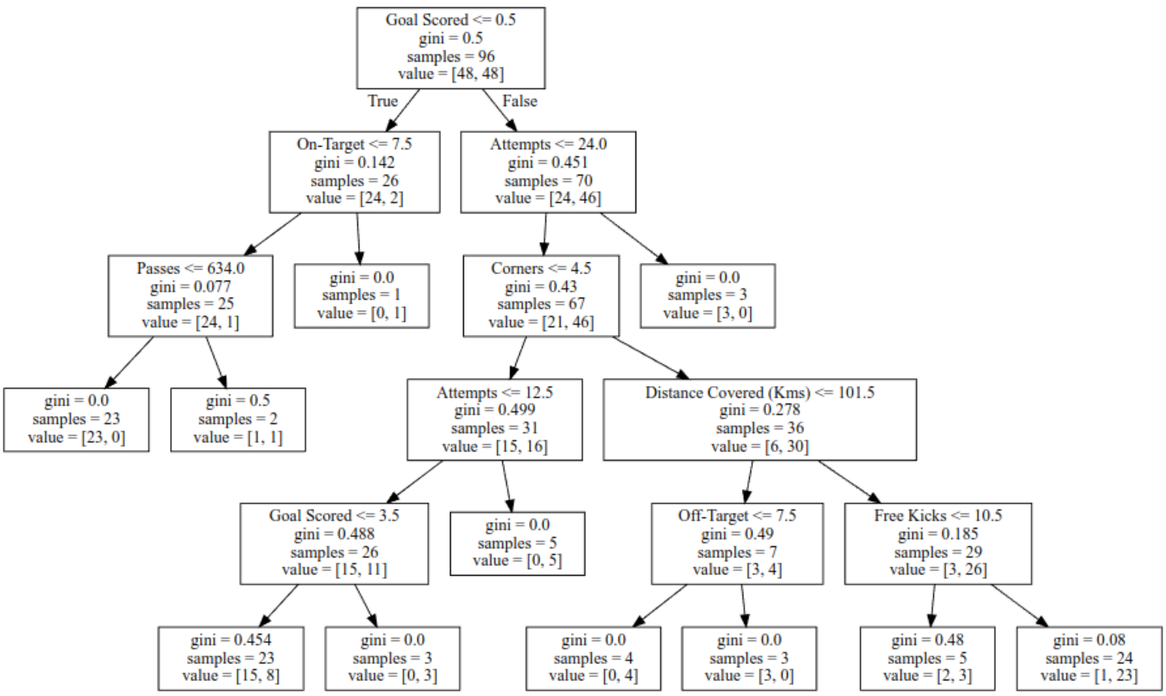

Figure 4: A decision tree example.

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

Post hoc interpretability on the other hand, also known as model-agnostic interpretability, refers to a set of interpretable methods that analyze machine learning models after they have been trained. These methods are versatile in their application, as they can be used with both intrinsically interpretable (white-box) models and complex black-box models. Regardless of a model's inherent interpretability, post hoc interpretability treats the machine learning model as a black box, probing its internal workings and generating explanations for its decision-making process. By analyzing the model retrospectively, post hoc interpretability techniques offer insights into how the model arrived at specific predictions, promoting transparency and understanding even in cases where the underlying model may be highly complex and difficult to interpret directly. This introduces the local and global terms for the model. Where local methods aim to explain individual predictions made by machine learning models, global methods aim to explain the overall behavior and characteristics of the entire machine learning model. A summary table (Table 1) of the methods and their features is provided below.

Interpretable Models	**Global Model-Agnostic**	**Local Model-Agnostic**
Decision Trees	Partial Dependence Plot	Local surrogate models (LIME)
Linear Regression	Permutation Feature Importance	Counterfactual explanations
Logistic Regression	SHAP	SHAP
Support Vector Machines (SVM)		
Naive Bayes		
K-nearest Neighbors		

Table 1: A summary of explainable models

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

Construal Level Theory

Proposed in 2012 by Trope and Liberman (Trope and Liberman, 2012), CLT is a psychological framework that provides insights into how individuals mentally represent and interpret events, objects, and information based on their psychological distance. CLT argues that people's construal or mental representations of different aspects of their experiences are influenced by factors such as temporal distance (how far into the future an event is perceived), spatial distance (how physically close or distant an object or event is perceived), social distance (how emotionally close or distant a person or group is perceived), and hypotheticality (how likely or certain an event is perceived to occur).

At the core of CLT is the idea that individuals can construe objects and events at different levels of abstraction, ranging from concrete and specific to abstract and general. At higher levels of construal, people focus on the broader and more abstract aspects of a situation, enabling them to identify general patterns and principles. On the other hand, at lower levels of construal, individuals pay attention to specific details and immediate consequences.

CLT helps understanding decision-making processes, risk perception, persuasion techniques, intergroup relations, and goal pursuit. The theory provides a framework to comprehend how individuals perceive and process information, make judgments, and behave based on their mental representations of events and objects at different psychological distances.

McDermott and Folds (McDermott and Folds, 2022), applied CLT to design of informational systems by developing a language called "RECITAL" to facilitate the information flow in distributed human-machine teams, modeled as a control hierarchy. The objective is to define the data, services, and user interfaces that allow humans to create, edit, query, and understand complex operational tasks, such as rules, intent, decisional authority, and related control actions, while interacting with each other and intelligent machines. The research explores three foundational concepts: formalizing the “RECITAL" language based on intent, rules, and delegated authority; applying it to established models of human-machine distributed teams; and using construal level theory from social psychology to guide information abstractions. The ultimate goal is to conceptualize a standard information model to support intentional design of human-machine teams.

It is important to note that the authors (McDermott and Folds, 2022) also exploit the hierarchical information flow (similar to military) in human-machine teaming scenarios. As it is derived from social psychology, it examines people's preference for abstract or concrete information based on psychological distance, which can be related to time, space, task relevance, or other interests. The level of information abstraction versus concreteness may be reflected in the comprehensiveness and level of detail of the presented information elements.

For informational systems, a six-layer model (Table 2) was designed which links construal levels to related information abstractions. It is described as a progressive

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

measurement information model, where increasing CLT numbers denotes progressively increasing detail of information.

Level of abstraction	Type of information/construal	Content/measurement model	User information needs and consumption time
CLT 1	Executive summary/main claim	This key outcome was/will be achieved as shown by these key indicators	Provide further definition of entities in the plan; provides success or failure indicators of the plan/ ~10s
CLT 2	Executive mission review/The main reason	CLT1 + because of these key causal effects	+ provide backstories for key entities, spatial and temporal aspects of the mission; drill down the additional details for the critical selected mission aspects/ ~30s
CLT 3	Mission summary/The justification	CLT2 + because in the full causal model these paths are of greatest importance (magnitude)	+ provide advisories and alerts related to changing context of mission as related to tasks at hand/~5min
CLT 4	Mission brief/the basis of justification	CLT3 + because here is the full measurement model	+ provide links to alternative planning and tasking if determined by context/~30min
CLT 5	Mission plan or report/A full summary of the data	CLT4 + because here is the time-step history of all the measurements	As supplied situationally to one of the levels above/~60min
CLT 6	Mission details/all the data	CLT5 + and here are all the anomalies and alternatives considered	As supplied situationally to one of the levels above/>60min

Table 2: six-layer model of CLT for informational systems.

A full description of the levels as well as their application is provided in the paper (McDermott and Folds, 2022).

Applicability

The operations described in HAIKU use cases and in general aviation, resemble operations in military, where information is presented on the basis of need and time constraints to the user. Our approach was to apply the CLT layers to capture the XAI needs related to the use cases and build an overview of operation for each use case. Time constraints vary significantly in each use case. As an example, in use case 1 the pilot may have less than 30 seconds or around 1 minute for each interaction 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

information consumption, whereas in use case 5 the user can choose how much information they would like to consume in each interaction.
CLT allows for an uniform approach for this problem and addresses the needs for abstraction level and time constraints. With the insights obtained from how each operation is performed, including the kind of information transfer needed, associated abstraction levels and time constraints, the XAI strategies may be outlined. Additionally, the content which reflects the interaction and ML/DL models for XAI, can then be used later for implementation and validation. Altogether, these methods provide a basis for strategies for each use case.

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

Strategies for use cases

Use case 1: Flight deck startle response

The startle response assistant aims to detect the startle and surprise reaction of pilots to be able to support them in recovering as quickly as possible. It will help pilots to regain control of the aircraft and a good situational awareness to be able to stay “ahead of the aircraft”. The startle response assistant will play an active role when an unexpected event occurs and when a startle or surprise response is detected. Support will be provided as long as necessary. Generally, the adverse effect of a startle or a surprise does not last more than 5 minutes.

CLT Level	Applicability	Description
CLT 1	During operation	• On Startle detection, inform the pilot about their startle situation. • Before startle even, inform the pilot of a potential event ahead. • On startle detection, display the most immediate action (depending on the situation) to be done.
CLT 2	During Operation	CLT 1 + : • Possible quick startle responses. • Preemptive startle task list before event .
CLT 3	During Operation	CLT 2 + : • Advisory action points/dialogue for stabilizing the aircraft. • Possible startle recovery procedures.
CLT 4	Post-OP briefing	CLT 3 + : • How pilot was supported to recover from startle • Why certain aircraft stabilization and advisory outputs were given (e.g. go-around)
CLT 5	Post-OP briefing	CLT 4 + : • Full measurements for startle detection and aircraft stabilization.
CLT 6	Post-OP briefing	CLT 5 + : • Possible anomalies that were not considered in startle. • Possible alternative approaches to startle and stabilize.

Table 3: CLT levels for use case 1, startle detection.

Use case 2: Flight deck route planning/replanning

During flights, pilots must manage complex situations involving numerous factors such as bad weather, complicated terrain, dense traffic, technical failures, human errors,

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

etc. As Human cognitive resources in the cockpit are limited, pilots can sometimes fail to correctly assess the risks associated with such conditions, especially when several combine. For use case 2, the decision making is of strategic nature, under a non-emergency situation, allowing for a reasonable time to make the decision. The situation addressed by the IA (Intelligent Assistant) is of medium to high complexity ("fuzzy" meteo event, high density airspace, etc). Other concurrent activities are happening in the cockpit, demanding attention from pilots.

A scenario where CLT can be applied in use case 2 is meteorological threat. IA detects sensitive changes in mapped weather threats using SWIM services and/or meteorological radar data. Flight plan is re-evaluated according to operational intentions and replanning options with respective KPI impacts are presented. Crew evaluates and selects a replanning option. IA supports negotiation of changes with ATC/AOCC. IA implements agreed flight changes in FMS.

First operation is as follows:

Decision point: Is there a significant weather threat that demands an alternative route?

Key outcome: Evaluation of threat alert presented by the IA (mandatory/safety-related? recommendable? As expected?)

CLT Level	Applicability	Description
CLT 1	During operation	Impact to safety; KPIs with current route (map + pictograph with values)
CLT 2	During operation	CLT 1 + : • Affected segment (e.g. Area to avoid, Risk colormap, Differentiating safety threat to operational intention threat)
CLT 3	During operation	CLT 2 + : • Detailed weather information (data fusion) (e.g. cloud type, height Information source and quality)
CLT 4	During operation	CLT 3 + : • Comparison to forecasted information used in planning
CLT 5	During operation	CLT 4 + : • Detailed weather information for each sensor/source
CLT 6	Post-OP	Detailed assessment of IA states and behavior

Table 4: CLT levels for weather threat evaluation in use case 2.

The next operation:

Decision point: Select alternative route due to weather threat.

Key outcome: Evaluate possible routes based on weather data.

CLT Level	Applicability	Description
CLT 1	Operational, in-flight	How well the recommended route fulfills the set of intentions (map + pictograph / index / 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

CLT 2	Operational, in-flight	CLT 1 + : • KPIs of the recommended route (map + pictograph with values)
CLT 3	Operational, in-flight	CLT 2 + : • Show alternative route options to fulfill intentions with CLT1 and CLT2.
CLT 4	Operational, in-flight	CLT 3 + : • Highlight uncertainties related to data quality/availability
CLT 5	Operational, Training / Coaching	CLT 4 + : • Show assessed performance vs calculated optimal, highlighting potential improvement points
CLT 6	Post-OP	Demonstrate robust, intended behavior assessing the IA states Evaluate root cause of incidents assessing the IA states.

Table 6: CLT for alternative route selection in use case 2.

Use case 3: Urban Air Mobility

The emergence of UAM is expected to lead to a significant increase in low-level airspace traffic over cities in the coming decades. To manage this airborne traffic safely and efficiently, UATM solutions are necessary. These solutions should focus on optimized airspace usage, adaptable airspace structures, and shared situation awareness for all stakeholders. The central human role, known as the UAM Coordinator, will be responsible for providing real-time strategic and tactical U-space services to UAS (Unmanned Aircraft Systems) and UAM operators and stakeholders. To ensure safety and efficiency in managing the increased traffic and coordinating ground and airborne activities, the UAM Coordinator will be supported by intelligent assistant (IA) capable of monitoring all air and ground traffic within the city airspace. Additionally, these assistants will monitor ground events and city activities with potential impacts on trajectory planning.

CLT Level	Applicability	Description
CLT 1	During Operation	Drone request notification
CLT 2	During Operation	CLT 1 + : • Route planning (e.g. annotated maps, displayed to the UAM coordinator)
CLT 3	During Operation	CLT 2 + : • Highlight of the route blocks or causes of route change (e.g. annotated map)
CLT 4	During Operation/Post-OP	CLT 3 + : • Full sensory data availability to the UAM for analysis
CLT 5	Post-OP	CLT 4 + : • Assessed performance of the route calculation

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

CLT 6	Post-OP	Full analysis of the performance of the assistant system, planned route, and anomaly handling

Table 7: CLT for use case 3.

Use case 4: Digital and Remote Towers

Intelligent Sequence Assistant (ISA) is designed to bolster and augment decision-making capabilities for ATCs. By focusing on optimizing runway utilization at single-runway airports, ISA offers real-time sequence suggestions for both arriving and departing aircraft. This dynamic and immediate assistance empowers Tower ATCs to handle traffic flow more efficiently by ensuring timely and accurate forecast updates. The envisioned advantages include enhanced decision-making, optimized runway usage, improved operational efficiency, and a safer, more streamlined air traffic management system. ISA achieves this by considering preset restrictions, prioritization rules, and inputs from various events, enabling it to make well-informed and effective sequence recommendations. With ISA's support, ATCs can enhance their situational awareness and manage air traffic with greater precision, ultimately leading to a more seamless and secure air traffic environment.

CLT Level	Applicability	Description
CLT 1	Pre-Op/During operation	Overview of the expected traffic (arrivals and departures, workforce during peak hours). If Pre-Op, can be used to brief ATCOs before work shift. KPIs with current situation
CLT 2	During Operation	CLT 1 + : • Display immediate solution for a sequence change (e.g. maps, updated flight maps etc.)
CLT 3	During Operation	CLT 2 + : • Provides complete information for the changes, and the differences with the initial KPIs. • Provides necessary alerts for the ongoing event.
CLT 4	Post-OP	CLT 3 + : • Provides useful information regarding the event (changed sequence, essential sensor data) for debriefing for the next shift.
CLT 5	Post-OP	CLT 4 + : • Full sensory data with the impact of the targeted changes due to the event is available at this level (e.g. traffic workload per X minutes, correlation between events and delays etc.)
CLT 6	Post-OP	CLT 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

		• Additional data (historical analysis, airlines, airplanes etc.) is available for further detailed analysis of the event. Could be used for training purposes.

Table 8: CLT for use case 4.

Use case 5: Airport Safety Watch

Airport safety watch relies heavily on analysis of safety and operational parameter data from Luton Airport (airport operator, airlines, handling services etc,), and has two main goals. First, analysis of persistent incident types (e.g. pushback errors, wrong taxiway selection etc.) and identifying solutions which would decrease the occurrence of these incidents. Next goal, aims to predict the risk of certain operations on particular days according to risk factors.

CLT Level	Applicability	Description
CLT 1	Operation/Senior level	A series of key risks and impacts are provided (maps, graphs, etc.).
CLT 2	Operation/ directors, department heads	CLT 1 + : • Possible risks (predicted risks) are provided (maps, highlighted areas, possible causes in areas).
CLT 3	Operation/ stack partners	CLT 2 + : • Possible solutions to the current ongoing incident • Preventative measures for possible immediate incidents (risks).
CLT 4	Operation/ Stack partners, Operational and safety managers	CLT 3 + : • Specific incident risk is selected, possible solutions and preventive measures are displayed. • Live update of day-to-day risks.
CLT 5	Operation/ Airside operational safety leadership	CLT 4 + : • The full analysis of the risk data will be provided to the partners and their effectiveness will be taken as feedback
CLT 6	Operation/ Operational personnel (airside)	CLT 5 + : • Heightened operational risks with precise data are distributed to operational actors. • Possible tailoring to specific risk areas (runway problems, icing, etc.) • Impact and correlation between incidents

Table 9: CLT for use case 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

Use case 6: Airport spreading Virus prevention

COVID 19 and airborne diseases pose a great impact to individuals' health. In particular, in indoor environments, crowded places are often susceptible to more frequent infections. As such they need to be avoided. In airports, people move with specific patterns toward the duty-free shops and restaurants, as well as the gates. There is a necessity in routing the crowd in such a way that they do not overcrowd shops in the airport. For instance, a passenger may go to the perfume shop a bit later during her stay if the maximum number of people reside in the shop. There is a definition of the number of people per a particular number of square meters that is defined by the Ministry of Health of a country. By being able to count the people that go to airport shops and coffee places/restaurants we can minimize the infection probability. This can be done from the time that the people disembark the plane or check in until they reach the common areas of the airport. Moreover, the air quality also needs to be measured whereby CO2, temperature, humidity and total Volatile Organic Compounds (tVOC) indexes are better.
There is a need for a digital assistant that will inform the passengers and the shop administrators regarding the available capacity as well as the air quality of the common places in airports, in order to avoid crowding them, with the risk of infection. Scheduling and routing the passengers is essential since it will spread the time of visit to common places. An integrated system for indoor environments is essential which will encapsulate the use of wireless networking and AI to provide an efficient digital assistant to avoid COVID infection.

CLT Level	Applicability	Description
CLT 1	Operation	Visual and text information regarding the COVID recommendation response. The duration of this message will be immediately outputted to the screen of the passenger (is not a lengthy narration) showing the level of crowdedness and air quality.
CLT 2	Operation	CLT 1 + : Since this is targeting the intent and related considerations for rules of engagement, this is the future recommendation of the system showing the infection probability as a variable and the level of people flowing towards the places of interest as well as the air quality. The initial routing decision is given to the passenger.
CLT 3	Operation	CLT 2 + : Here the fact that the duration of an operation is less than 5 minutes provides the passenger with the time to process the recommendation by the system of the COVID infection probability to airport common places. Detailed pathways information and air quality forecasts are given and the limitation of the infection is provided to show the success of the recommendation system in enhancing trust.

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

CLT 4	Post-OP	CLT 3 + : This level essentially initiates the post-operation procedure that provides the airport personnel with information relevant to the application of the recommendation system and the routing process as well as the air quality. By providing the routing process statistics to a higher-level authority other than the passenger the system may be evaluated and the trust may be strengthened as well as its validity elaborated.
CLT 5	Post-OP	CLT 4 + : This is an envelope of the previous level enhanced with more parameters regarding the passengers' trends regarding the recommendation system. This targets the airport staff responsible for health and safety and it will provide the overall mission of the system to be deployed before being available again to the passengers. This will provide information to maintain and improve the system as a whole.
CLT 6	Post-OP	CLT 5 +: The details and logs will be updated with parameters on the fly, if applicable, as well as maintenance and other information. This again targets the airport personnel.

Table 10: CLT for use case 6.

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 and Summary

CLT provides the use cases with a high-level view on designing their Intelligent Assistants (IA) with respect to explainability attributes such as what (or how) an explainable assistant system would (should) behave for a certain scenario, similar to requirement analysis for classical software development. Together with interaction and models, they will next transform into implementation design in task 5.2 of HAIKU.

In the analysis of CLT (Table 11), it is important to understand that:

- Not all the stages of CLT are suitable for HAIKU use cases.
- The focus of strategies and implementation will be on levels that aim on operation (during operation) and less or none on Pre/Post-OP. However, for completeness, these stages are mentioned for each use case.
- The utility of Pre/Post-Ops mainly lay on training and operation analysis on organizational level.
- Regardless, a complete IA must be capable of providing the necessary information in all levels to corresponding stakeholders (e.g. a level 6 report for legal reasons or organizational purposes).

UC **CLT Lev.**	**1**	**2**	**3**	**4**	**5**	**6**
CLT 1	Operation	Operation	Operation	Operation	Operation	Operation
CLT 2	Operation	Operation	Operation	Operation	Operation	Operation
CLT 3	Opertaion	Operation	Operation	Operation	Operation	Operation
CLT 4	Post-OP	Operation	Operation/Post-OP	Post-OP	Operation	Post-OP
CLT 5	Post-OP	Operation	Post-OP	Post-OP	Operation	Post-OP
CLT 6	Post-OP	Post-OP	Post-OP	Post-OP	Operation	Post-OP

Table 11: Comparative view of operational stage for the use cases with regards to CLT levels.

- In UC3 CLT 4, the abstraction can be applied to both Operational and Post-OP. To distinguish and decide the operational stage, duration of information consumption must be taken into account.
- CLT levels 1-3 usually aim at short duration information consumption which is suitable for sensitive/time-restricted scenarios (e.g. UC1 startle response). Some use cases are more elaborate with time and information consumption.
- UC 5 is a good example where all levels are potentially in operation. However, the information in different levels is consumed by different stakeholders/roles.

The next steps in second year (Task 5.2) will be to implement fully or partially, an XAI interface for the use case. This will be realized for each use case by selecting an operational stage from the CLT table and implementing an interface based on the interaction paradigms, which will enable information consumption and interaction between ML/DL models and the user. Finally, this interface will be evaluated in validation studies.

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

Alqaraawi, A., Schuessler, M., Weiß, P., Costanza, E., Berthouze, N., 2020. Evaluating saliency map explanations for convolutional neural networks: a user study. Presented at the Proceedings of the 25th international conference on intelligent user interfaces, pp. 275–285.

Baber, C., McCormick, E., Apperly, I., 2021a. A human-centered process model for explainable AI. Presented at the Naturalistic decision making and resilience engineering symposium 2021, France.

Baber, C., McCormick, E., Apperly, I., 2021b. A framework for explainable AI.

Chromik, M., Butz, A., 2021. Human-XAI Interaction: A Review and Design Principles for Explanation User Interfaces, in: Ardito, C., Lanzilotti, R., Malizia, A., Petrie, H., Piccinno, A., Desolda, G., Inkpen, K. (Eds.), Human-Computer Interaction – INTERACT 2021, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 619–640. https://doi.org/10.1007/978-3-030-85616-8_36

EASA, 2023. A human-centric approach to AI in aviation. European Aviation Safety Agency 1.

Hornbæk, K., Oulasvirta, A., 2017. What Is Interaction?, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. Presented at the CHI '17: CHI Conference on Human Factors in Computing Systems, ACM, Denver Colorado USA, pp. 5040–5052. https://doi.org/10.1145/3025453.3025765

Kim, D.H., Hoque, E., Agrawala, M., 2020. Answering questions about charts and generating visual explanations. Presented at the Proceedings of the 2020 CHI conference on human factors in computing systems, pp. 1–13.

McDermott, T., Folds, D., 2022. Construal level theory in the design of informational systems. Frontiers in Physics 10.

Paudyal, P., Banerjee, A., Gupta, S., 2020. On evaluating the effects of feedback for sign language learning using explainable AI. Presented at the Proceedings of the 25th International Conference on Intelligent User Interfaces Companion, pp. 83–84.

Tabrez, A., Agrawal, S., Hayes, B., 2019. Explanation-based reward coaching to improve human performance via reinforcement learning. Presented at the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, pp.

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

249–257.

Trope, Y., Liberman, N., 2012. Construal level theory. Handbook of theories of social psychology 1, 118–134.

Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., Zhu, J., 2019. Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges, in: Tang, J., Kan, M.-Y., Zhao, D., Li, S., Zan, H. (Eds.), Natural Language Processing and Chinese Computing, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 563–574. https://doi.org/10.1007/978-3-030-32236-6_51

Yin, M., Wortman Vaughan, J., Wallach, H., 2019. Understanding the effect of accuracy on trust in machine learning models. Presented at the Proceedings of the 2019 chi conference on human factors in computing systems, pp. 1–12.

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