A Design Evaluation of Text and Graphical Explanations from a Conceptual Intelligent Assistant in Urban Air Traffic Management

Jenny Söderman(✉), Ludwig Halvorsen, Jekaterina Basjuka, and Carl Westin

Linköping University, Linköping 581 83, Sweden
jenso406@student.liu.se

Abstract. This study explores the design of two explanation formats for a conceptual AI-based assistant in low-level urban drone traffic management, to support effective Human-AI collaboration. The design solutions aim to present decision alternatives and explanations in time-critical situations, crucial for enhancing situation awareness and understanding of the AI systems. Using research through design, a three-part design process was used to develop two explanation formats, which were then evaluated with three air traffic controllers and three novices. Each participant completed two trials with explanations presented using a text-based or graphic-based format in response to a static scenario involving an unpiloted air taxi transporting a passenger experiencing a medical emergency. After each trial, participants completed the Explanation Satisfaction Scale to evaluate aspects such as understanding, clarity, detail, completeness, and effectiveness. Their answers were then discussed using open-ended questions to derive qualitative insights into their evaluation of the explanations. Results revealed differing preferences: controllers favored the text-based presentation, appreciating its detail, while novices preferred the graphical version, appreciating its ease of processing information. The study concludes that a hybrid approach that combines and leverages respective strengths of text-based and graphical explanations could potentially offer a more versatile solution for explaining AI system behaviors that accommodates a broader range of user experience levels and preferences.

Keywords: Human-AI Teaming · Intelligent Assistants · Explainability · Transparency · Research Through Design · Urban Air Mobility

1 Introduction

Researchers and industry professionals are increasingly exploring the potential of AI-based intelligent assistants as digital colleagues in safety-critical domains, with the aim of improving operational efficiency and enhancing safety practices. The growing scientific interest reflects the recognition of AI's capacity to augment human capabilities in complex, high-stakes environments. Given the safety standards required in these domains, it is essential that AI systems are designed to support effective and efficient collaboration with humans [1–3]. This raises important questions about how to design AI

D. Harris and W.-C. Li (Eds.): HCII 2025, LNAI 15777, pp. 167–183, 2025.
https://doi.org/10.1007/978-3-031-93721-7_13

systems to facilitate such collaboration. The challenges inherent in effective Human-AI teaming underscore the need for systems that are transparent and explainable, supporting human situation awareness (SA) and enhance understanding of the AI agent's actions and decision-making processes [1].

Research on transparency and explainability aims to support human awareness and understanding of system behavior [4, 5] and facilitate trust and acceptance [6–9]. This includes concepts like machine learning interpretability [10] and explainable AI (XAI) [11]. Understanding depends on both the content and presentation of explanations [8, 12]. However, much of the research on automation transparency and XAI have focused on deriving the explanations, not on how to best present them to the user and decision maker, which is particularly critical in safety-sensitive, time-constrained environments such as traffic management and vehicle operation [13–15].

To address this research gap in how to best present explanations, this paper investigates how different explanation formats can be designed. The study used a Research Through Design approach, applying Schlatter and Levinson's [16] design principles for digital applications as a framework to support the creation of user-friendly explanations. The principles guided the design of two explanation formats for comparison: a text-based solution and a graphics-based solution. The two research questions were:

1. How can text-based and graphics-based explanations of an intelligent assistant's decision alternatives be designed within a map-based interface for low-level air traffic management to enhance user understanding and support decision-making, using Schlatter and Levinson's [16] design principles for digital applications?
2. How are text-based and graphics-based explanations perceived and understood by air traffic controllers and novices?

This study focused on how to present explanations, as opposed to developing the underlying AI system or XAI model. In the user evaluation, a conceptual AI system was introduced to participants as an intelligent assistant. The study was set within the context of a future low-level urban airspace traffic management (U-space) scenario, were a new human role, the UAM Coordinator, and AI system collaborate to manage drones and air taxis safely [17, 18]. The UAM Coordinator role builds on traditional air traffic control but extends to the management of U-space traffic operations, handling high-level airspace service tasks like dynamic airspace management, conformance monitoring, and stakeholder coordination during emergencies. The AI system, supervised by the UAM Coordinator, is envisioned to handle routine traffic monitoring and assist during emergencies, reducing the UAM Coordinator's workload, enhancing SA, and support decision-making. The study is part of the European Commission-funded HAIKU (Human AI Teaming Knowledge and Understanding for Aviation Safety) project that investigates human-centered AI-based intelligent assistants in aviation [19].

The next section presents the theoretical foundation, covering Human-AI teaming, transparency, XAI, and interface design. Section 3 outlines the design and evaluation of two presentation formats. Section 4 presents the results of the evaluation with air traffic controllers and novices. Section 5 discusses the findings in relation to the research questions. Finally, Sect. 6 concludes with recommendations for future work.

2 Theoretical Foundation

This section presents research on Human-AI teaming and critical role of automation transparency and explainability in fostering human understanding of machine systems. Additionally, it overviews interface design guidelines, including Schlatter and Levinson's [16] design principles for digital applications that serve as the foundation for the designed interface solutions for explanations.

2.1 Human-AI Teaming

The recent rapid maturement and development of AI technologies has sparked a surge of research on human-autonomy and Human-AI teaming (both terms are being used interchangeably), largely related to military and safety-critical working environments [5, 8, 9]. Human-AI teaming generally refers to one or more humans and one or more AI systems working together to accomplish a task [1]. This teamwork can take various forms; for example, an AI system can function as an assistant, or a human and an AI system can collaborate by sharing tasks and responsibilities. Specifically in aviation, EASA [2, 3] classifies Human-AI teaming as AI-based systems capable of cooperating or collaborating with humans to achieve a goal. In cooperative systems, the AI follows a directive approach, assisting the human through predefined task allocation without requiring shared SA, while the human maintains full control. Conversely, collaborative systems adopt a co-constructive approach, pursuing shared goals with flexible task distribution, granting both the human and AI some autonomy, and relying on shared SA. In this setup, the human primarily acts as a supervisor. EASA emphasizes that transparency is crucial for maintaining SA, noting that a lack of transparency can undermine user trust and the ability to verify system behavior and outputs.

The success of synthetic teams are considered to depend on the human's ability to understand and predict the system, developing appropriate trust, making accurate decisions, and exerting control over the system [1]. The literature outlines several factors considered to impact human-AI teaming, among which key factors are the interconnected system transparency/explainability and SA (both individual and team awareness) [1, 5, 8, 9]. According to Endsley [9], it is a state of knowledge about dynamic situations, divided into three levels: First, the *perception* of relevant information through senses or system interfaces (Level 1). Second, the *comprehension* or understanding of the information's meaning and how it relates to goals (Level 2). Third, the *projection* or anticipation of how a situation may develop (Level 3). Endsley [9] argues that cognitive processes like memory, attention, and mental models shape SA, which varies by individual experience and training. As a key feature of AI systems, system design is essential for supporting the human operators' SA and understanding, particularly in how effectively interfaces present dynamic information. Designing interfaces to afford transparency and explainability for clearly communicating AI behaviors, actions, and recommendations, can strengthen both individual and team SA [1, 6, 9, 12].

2.2 Transparency and Explainability

Transparency and explainability of AI operations and decision makings processes are considered key system characteristics in Human-AI teaming, as they can improve users' understanding, trust, SA and interaction with the AI system [5, 8, 9, 12]. Although related, transparency is a broader concept that encompasses automation in general, including explainability, which specifically refers to AI systems [1]. Westin et al. [7] define automation transparency as the automation's ability to afford understanding and predictions about its behavior (p. 202). Similarly, explainability generally helps users understand the decision-making process of complex or opaque prediction models (e.g., black box) by explaining how a specific recommendation, decision, or action was derived [20]. Closely related and often used interchangeably is interpretability, which focuses on understanding the prediction model.

Endsley [9] argues that explainability approaches have focused on why a system did something in a certain way (e.g., explaining outputs or actions), while transparency approaches have emphasized what the system is doing (i.e., behavior). While transparency is essential for real-time SA, explainability indirectly supports SA by providing retrospective information that helps humans develop accurate mental models of AI systems. For aviation AI applications, EASA [2] refers to *operational explainability* as "the need to provide end users with 'understandable' information on how the AI-based system came to its results" (p. 98). They provide examples of target audiences such as airborne operations, maintenance, and air traffic control.

Expanding on Endsley's model, Chen et al. [6] introduced the Situation Awareness-Based Agent Transparency (SAT) model for designing interfaces that aid autonomous agent mission supervision. Level 1 SAT explains the agent's actions and behavior, drawing from Lee and See's [21] model that describes how a system's process, purpose, and performance influence user trust. Level 2 SAT describes the agent's reasoning and behavior using the Beliefs, Desires, Intentions (BDI) framework for structuring rational behavior [22]. Level 3 SAT predicts the agent's future actions and outcomes. Chen et al. [6] view these levels as distinct and adaptable, allowing designers to apply them selectively based on task goals and context. This flexibility is argued to support trust calibration and leads to more effective automation use.

Different transparency information has been identified as important to incorporate in design for supporting understanding at different levels of SA [1].

- System Status (Level 1 SA): Includes information about the system's current state, goals, progress toward goals, factors considered, plans, and environmental constraints.
- Understandability (Level 2 SA): Encompasses the system's reasoning, decision logic, capabilities, limitations, and "confidence" in its assessments.
- Predictability (Level 3 SA): Relates to the system's ability to manage future situations, predict consequences, and "confidence" in its predictions

Hoffman et al. [23] introduce a framework for evaluating the explainability of an AI system, emphasizing the importance of determining whether the provided explanation is satisfactory. Key factors include whether the explanation is understandable, sufficiently detailed, perceived as complete, and assists the user in achieving their goal and deciding when to trust the system. To assess user feedback on these aspects, the authors proposed the Explanation Satisfaction Scale—a questionnaire with eight Likert-scale statements on a scale of 1–5, where 1 means strongly disagree and 5 means strongly agree. This scale was designed to gather users' judgments after interacting with a system and receiving one or more explanations.

2.3 Interface Design Guidelines

While considerable research has focused on deriving explanations and the content to explain, less attention has been given to how present these explanations and design user friendly experiences [14, 15]. Safety-critical working environments such as traffic control centers demand interfaces that support quick, accurate decision-making under pressure. Interface design guidelines provide designers with a structured approach to creating interfaces that are intuitive and easy to understand, while also promoting safe and efficient user interaction. Sanneman and Shah [24] highlight the importance of having the AI system explain its actions or decisions in relation to its goals. They argue that the information needed to understand causality may differ depending on the user and suggest that one way to identify relevant information is by considering the comparison object the user has in mind. Additionally, they emphasize the value of presenting contrasting information, e.g., by showing how the system would have acted under different conditions, which can improve understanding of its potential future behavior.

Dudley and Kristensson [15] outlined six principles for designing user interfaces in interactive machine learning. They emphasize: (1) clarifying task goals and constraints for the user, (2) helping users understand model uncertainty and confidence, (3) focusing on user intent (what they aim to achieve) rather than just user input (their actions), (4) enhancing user perception with effective data representations, (5) promoting interactivity and freedom to foster understanding, and (6) motivating users to actively engage with the task. However, they do not differentiate between machine learning applications in safety-critical environments, where time to act is limited. In real-time applications, the appropriateness of allowing the user exploration and interactivity with the system's operations is questionable [9].

There are various ways to present explanations. According to Cambria et al. [25], explanations can be categorized as text-, graphical-, or multimedia-based. *Text-based explanations* use written language (e.g., natural language), to mimic how people use language to describe and explain. This approach is believed to offer several advantages compared to graphical explanations, including increased efficiency, trust in the explanation, and better decision-making in uncertain situations. *Graphical-based explanations* involve visual representations (e.g., diagrams). Parasuraman et al. [28] note that graphics can help reduce operator mental workload, if the representation matches the operator's mental model of the system. *Multimedia-based explanations* combine multiple elements, such as text, graphics, images, and sound, to convey information.

In the context of digital applications, Schlatter and Levinson [16] argue that the three most important principles (i.e., meta-principles) affecting interface design are consistency, visual hierarchy, and personality. *Consistency* ensures a cohesive design by applying uniform rules for element placement and appearance, such as using the same typeface for similar components, making the interface more intuitive. *Visual hierarchy* highlights the relative importance of on-screen elements, guiding users in understanding available actions and their outcomes. This can be achieved by adjusting the size or adding white space to create contrast that naturally attracts user attention. *Personality* pertains to the user's appeal and satisfaction of the interface, influenced by their perception, usage, and engagement with it, and covers both the interface appearance (e.g., layout, color, type, imagery) and interactive aspects (e.g., controls and affordances). For this paper, layout, color, and icons hold particular importance. *Layout* involves the strategic arrangement of elements to ensure easy comprehension by users. For instance, grouping related elements together helps signal their connection. The strategic use of *color* can be applied to attract attention or trigger associations (e.g., red indicating "stop"). Lastly, clear, recognizable *icons* that visually represent concepts can replace or complement text to improve understanding. The advantage of using icons is that they quickly capture attention and convey meaning, provided they are easily identifiable.

3 Design Process and Evaluation

This section describes the design process and user evaluation of two interface solutions: a text-based explanation and a graphics-based using color-coded icons. The design process built on a Research-Through-Design approach aiming to fulfil Zimmerman et al.'s [26] four criteria for evaluating interaction design research. *Process* requires a clear description of the design work and method justification. *Invention* demands that the design proposal offers something new, often by combining ideas from different areas and reviewing existing literature. *Relevance* explains the goals of the proposal and its significance. Finally, *extensibility* ensures the results can be built upon for future work or applied to new design problems [26]. The design work followed Arvola's [27] method with three phases: conceptual, refinement, and detailing.

3.1 Scenario

The interface solutions were designed to provide explanations for a specific scenario. The scenario consisted of an unpiloted air taxi transporting a passenger experiencing a medical emergency. When the medical emergency occurs, the AI system proposes alternative hospitals where the air taxi can divert. The human operator (i.e., UAM Coordinator) must decide on which hospital to route the air taxi. The situation is presented to the UAM Coordinator in a map display of the city, with the flight plan of the air taxi.

3.2 Conceptual Phase

In the conceptual phase, insights from literature helped define the design objectives, generate ideas, and evaluate these ideas to determine a narrower design focus.

Design Objectives. Based on the literature review, it was decided to develop two solutions: text-based explanations, which can enhance efficiency and trust under uncertainty [25], and graphics-based explanations, which may ease the cognitive burden [28]. A reference scenario and prototype map interface was used to simulate the UAM Coordinator's future environment, as envisioned in the HAIKU project [19]. The solution proposals were designed to convey the following: (1) a medical emergency on an unpiloted air taxi carrying a passenger, (2) the AI assistant's decision alternatives (nearby hospitals to divert), and (3) explanations of how each option was derived. The content of explanations was based on four key factors elicited from consulting three U-space and air traffic management experts: (1) travel time to the hospital, (2) hospital waiting times, (3) doctor availability, and (4) landing availability. Both solutions would also display the available decision time and highlight the drone and hospital locations on the map.

Idea Generation. A brainstorming activity was carried out to explore ideas for the layout and presentation of explanations. For the text-based solution, ideas included a side menu, everyday language, bullet points, continuous text, and a time scale (short/medium/long). Bullet points and a time scale were considered to simplify comparisons under time pressure. For the graphics-based explanation, ideas included placing explanations on the map, using diagrams, images, color coding, and traffic sign-inspired icons. The layout was considered to help users understand the spatial positioning of hospitals and their distances. Color coding and icons were considered suitable for their ability to draw attention and quickly convey meaning [16].

Evaluation of Ideas. Each design idea was assessed based on its suitability to the scenario. For the text-based solution, the *side panel*, *bullet points*, and *time scale* were chosen to organize the information clearly and support fast comparison under time pressure, while *natural language* was used to enhance understanding and build trust [25]. For the graphical solution, the ideas of *placing options on the map* near the hospitals, *color coding*, and *icons* (inspired by traffic signs) were chosen to enable quick interpretation and capture attention [16]. Overlaying explanations on the map aimed to help users easily understand the distance to each option.

3.3 Refinement Phase

The refinement phase refined selected key ideas by outlining the design and layout in paper sketches (Figs. 1 and 2). Inspiration was taken from an early HAIKU project design for presenting decision options in medical emergencies [16] where options were displayed as a bullet list within a dialogue window overlaid on the map view. Decision options were labeled A, B, and C to reflect the AI system's ranking, with travel time shown next to hospital names in the text-based design.

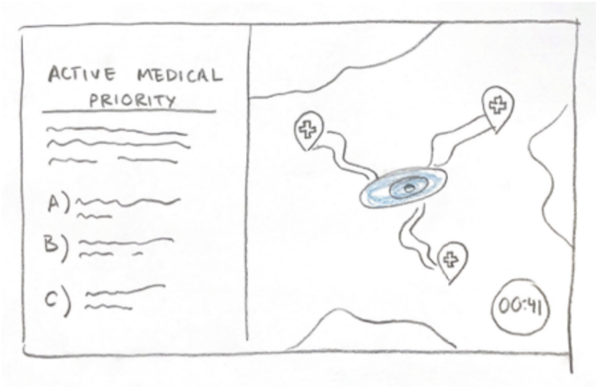

Fig. 1. Overall design of the text-based explanation.

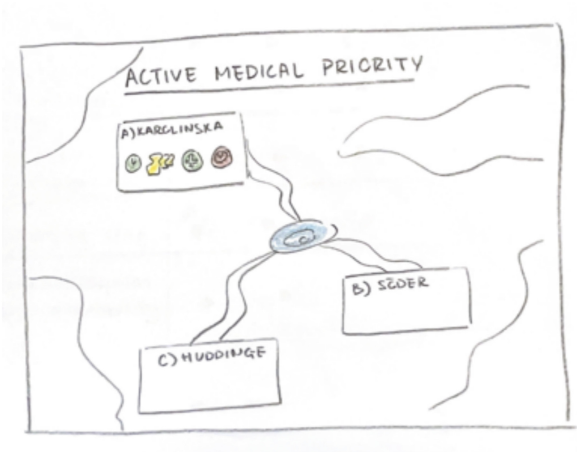

Fig. 2. Overall design of the graphics-based explanation.

3.4 Detailing Phase

In the detailing phase, selected sketches from the development stage were implemented as detailed interface prototypes in Figma (https://www.figma.com). The map featured in both solutions was adapted from customizing Google Map images in Snazzy Maps (https://snazzymaps.com).

Design Choices for the Text-Based Solution. *Layout* was a key focus, with decision alternatives grouped in a side panel to indicate related information. Within the panel, alternatives and their explanations were separated into distinct white boxes to clarify connections and create white space for contrast and visual emphasis. *Visual hierarchy* was maintained by enlarging and bolding headings for each decision alternative (hospital name and travel time). *Color* was also applied strategically by featuring the side panel in a slightly transparent gray background to distinguish it from the map, while buttons and map pins were colored to contrast with the background for clarity. *Consistency* was ensured by using a single typeface and a limited color palette throughout the design. Bullet points were used to enhance readability and simplify comparison between options. The terms "short", "medium", and "long" described travel and queue times for clarity. "Available" and "unavailable" were chosen to clearly indicate doctor and landing site availability, ensuring straightforward comparison.

Design Choices for the Graphics-Based Solution. In this design, each decision alternative was placed in a box on the map at the corresponding hospital location that contained the relevant explanation, represented in color-coded icons. This *layout* aimed to clearly distinguish each alternative and associated explanation. *Color* played a key role in the design, with boxes given a slightly transparent gray background to make them stand out against the map (similar to the text-based proposal). Red, blue, and black were used to create a strong contrast with the background. For the explanations, *colors* and *icons* were inspired by traffic lights. Colors were associations, with green, yellow, and red icons used to signal travel times, hospital queue durations, and the availability of doctors and landing sites. Icons were designed to be simple, easily recognizable, and intuitive, such as using a traffic jam sign to represent the hospital queue. *Consistency* was maintained by using a single typeface, a limited color palette, and consistent rounding of elements.

3.5 User Evaluations

A quantitative and qualitative user evaluation was conducted to compare the two explanation formats in terms of their impact on users' understanding and satisfaction. The study included three experienced air traffic controllers (ages 41–55, avg. 47.3) with 16–30 years of experience (avg. 21) and three novice participants (Linköping University students, ages 23–29, avg. 25.3). All volunteered, recruited through a convenience sampling. Controllers were familiar with U-space and the HAIKU project.

Explanation Conditions. The interface featured a map of Stockholm City, Sweden with an enroute air taxi. In response to a medical emergency onboard, a conceptual AI system provided three decision alternatives, each corresponding to a potential hospital for diversion. Explanations revealed the contribution of four key factors to each alternative. Two explanation formats were used. In the layout of the *text-based explanation* shown in Fig. 3, emergency details, decision alternatives, and their explanations are organized in a side panel overlaying the map. Each alternative appears in a white box with the hospital name, travel time, four bullet points reflecting the four key factors, and a blue selection button. The distance-related factor is labeled as short, medium, or long. "Cardiological dept. available" confirms the needed specialist is present and "Vertiport available/unavailable" shows landing availability. The map visualizes the air taxi as a blue circle and hospitals as red pins labeled A, B, or C. A countdown timer for the validity of the suggestions is provided in the bottom right corner.

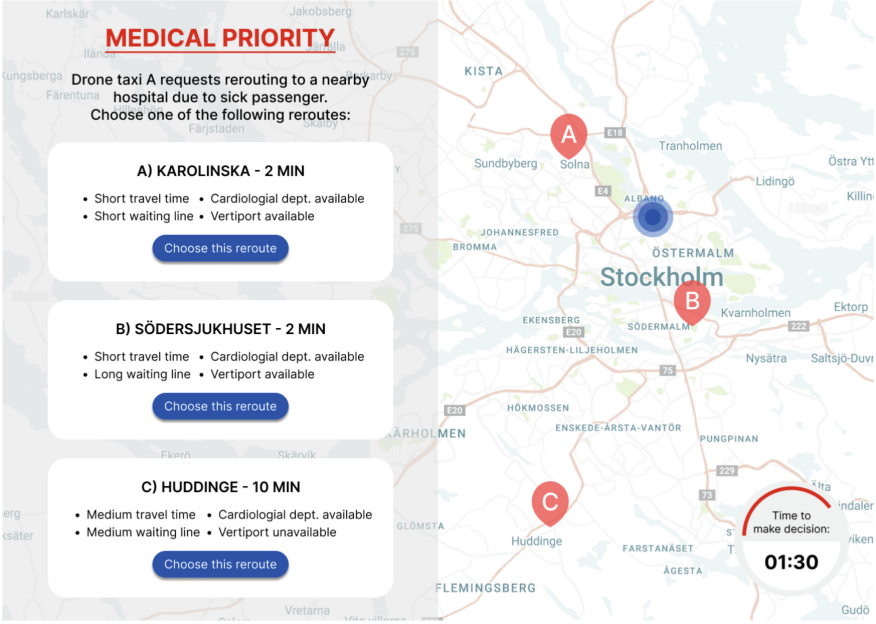

Fig. 3. Text-based explanation with a bullet-point list detailing how each of the four factors influences the proposed decision. Hospital positions are shown as red pins, each labeled with a letter corresponding to a decision alternative.

In the *graphics-based explanation* shown in Fig. 4, a top-center box labels the situation as a "medical priority." Decision alternatives and their explanations are placed directly on the map at each hospital's location. Explanations comprise four color-coded icons: travel time, hospital queue, Cardiologist availability, and vertiport availability. Green, yellow, and red represent short, medium, and long times, while cardiologist and landing availability are shown in green (available) or red (unavailable). Similar to the text-based explanation, the air taxi is marked with a blue circle, connected to hospitals by blue lines, and a countdown timer appears in the bottom right corner.

Procedure. Evaluations with controllers were conducted remotely via Microsoft Teams, while tests with novices were in person. Each session lasted 30–45 min and audio were recorded via Microsoft Teams or an iPhone voice memo app to aid subsequent analysis. Participants were welcomed, informed about the study, and briefed that the purpose was to compare two explanation formats. Consent was obtained orally. Participants were instructed that they could withdraw their consent at any time, that the test would be recorded, and their data would remain anonymous. Each participant completed two trials with decision alternatives and explanations presented using either the text-based or graphic-based format (the order was counterbalanced).

Participants received the following instruction: *A drone is transporting a passenger from Globen to Täby Centrum in Stockholm. During the journey, the passenger suddenly experiences a medical emergency and requires urgent care. Using various sensors, the drone detects the passenger's condition and alerts the intelligent assistant. The assistant*

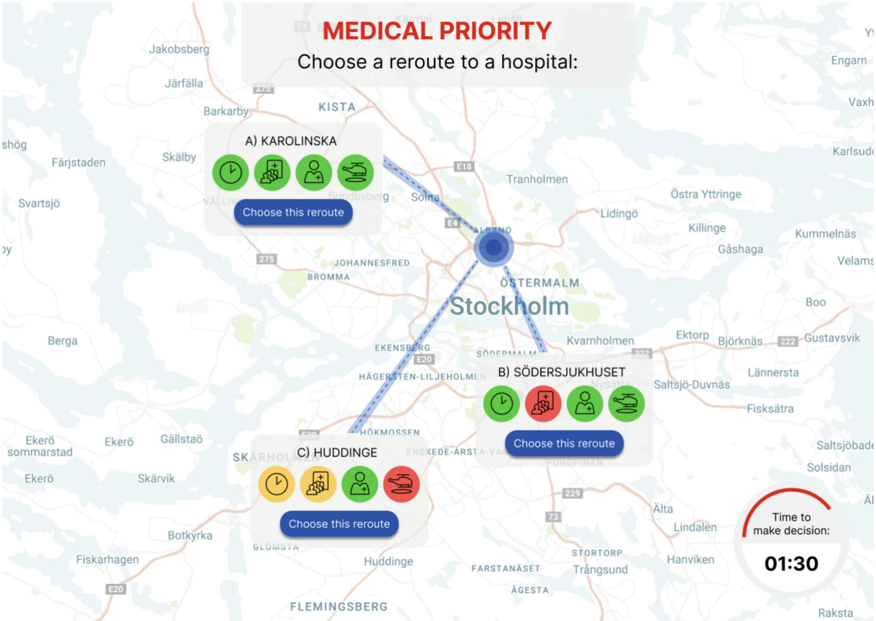

Fig. 4. Graphics-based explanation with the four factors influencing the proposed solutions displayed through color-coded icons, positioned on the map at the locations of the hospitals.

then generates several route options to different hospitals and presents them to the human, who is responsible for selecting the most suitable option. Participants could review each explanation format freely before deciding on one of the three alternatives. After each trial, participants completed the Explanation Satisfaction Scale [23], which includes eight 5-point Likert-scale statements, and a demographics questionnaire. Participants' answers were then discussed open-ended. Quantitative data were summarized with means and standard deviations, while qualitative open-ended responses were thematically analyzed using Braun and Clarke's [29] thematic analysis.

4 Results

The following section presents the questionnaire responses, and the themes identified through thematic analysis in relation to open-ended questions.

4.1 Questionnaire Results

Table 1 shows questionnaire results using Hoffman et al.'s [23] scale. Each question has a maximum score of 5, and a summarized mean total of maximum 4.0. Results indicated that controllers preferred text-based explanations (average total rating of 4.3 for all eight statements) over graphics-based explanations (average 3.4). Conversely, novices favored graphics-based explanations (average 4.3) over text-based ones (average 4.0). For the statement regarding trust in the system, controllers rated the text-based explanation higher (*M* = 3.7, *SD* = 0.6) than novices (*M* = 2.3, *SD* = 0.6). In contrast, novices rated graphic-based explanations more favorably (*M* = 3.7, *SD* = 1.2) than controllers *(M* = 2.7, SD = 1.5). Controllers generally rated text-based explanations high (*M* = 4–5), except for system accuracy (*M* = 3.3, *SD* = 0.6). Graphics-based explanations received lower scores on several aspects, including completeness (*M* = 3.0, *SD* = 2.0), accuracy (*M* = 2.7, SD = 1.5), and trust (*M* = 2.7, *SD* = 1.5). Controllers' ratings for text-based explanations were more consistent (*SD* = 0.6–1.0) than for graphics-based explanations (*SD* = 1.5–2.0), indicating greater individual variability of the latter.

Table 1. Participants' responses to the Hoffman et al. [23] questionnaire for the text- and graphics-based explanations, presented in *mean (standard deviation).*

Statement	Text		Graphics	
	Controllers	Novices	Controllers	Novices
From the explanation, I understand how the system works	4.7 (0.6)	4.7 (0.6)	4.0 (1.7)	4.3 (0.6)
This explanation of how the system works is satisfying	4.7 (0.6)	3.7 (0.5)	3.7 (1.5)	4.0 (1.0)
This explanation of how the system works has sufficient detail	4.7 (0.6)	4.3 (1.5)	3.3 (2.1)	4.3 (0.6)
This explanation of how the system works seems complete	4.7 (0.6)	4.3 (1.5)	3.0 (2.0)	4.3 (0.6)
This explanation of how the system works tells me how to use it	4 (1.0)	4.7 (0.6)	4.0 (1.7)	4.7 (0.6)
This explanation of how the system works is useful to my goals	4.7 (0.6)	4.7 (0.6)	4.0 (1.7)	5.0 (0.0)
This explanation of the system shows me how accurate the system is	3.3 (0.6)	3.3 (0.6)	2.7 (1.5)	3.7 (0.6)
This explanation lets me judge when I should trust and not trust the system	3.7 (0.6)	2.3 (0.6)	2.7 (1.5)	3.7 (1.2)
Mean total (max value 4.0)	***3.4***	***3.2***	***2.7***	***3.4***

4.2 Thematic Analysis

Thematic analysis results are categorized by text-based and graphics-based explanations, highlighting whether themes were mentioned by controllers, novices, or both. Thematic analysis showed that explanations provided limited insight into underlying factors. Novices found text explanations difficult and slow, while color-coded graphics improved decision-making by highlighting optimal solutions. Participants favored text-based layouts for preserving map visibility and aligning alternatives side-by-side.

Text-Based Explanations Four themes emerged. The first theme was the *Level of detail in the explanation (controllers only).* Some controllers felt uncertain about the meaning of the short/medium/long scale and desired more specific, numerical details. They found the text-based explanations more detailed compared to the graphical ones. One participant noted, "*When you say the word short, medium and long … what does short mean? What does medium mean? What does long mean? … if the system could even provide numbers then it would be really good.*"

The second theme was related to the *Superficial understanding of the system (both controllers and novices).* Participants stated that the explanations provided only a limited understanding, mainly conveying that four factors were considered. One remarked, "*With the four points that are there, that's what I know. I don't know whether there's more underlying."*

The third theme was *Time-consuming/difficult to grasp the explanation (both controllers and novices).* Many, especially novices, found it difficult and time-consuming to process the information and differentiate between alternatives. One said, "*You had to read several times before seeing that the waiting times differ between A and B.*

The fourth theme was *Design-related feedback (both controllers and novices).* Some participants appreciated the layout. One appreciated that the decision alternatives were gathered in one place. Another participant appreciated that the map view was not obscured with options and explanations (as in the graphics-based interface), stating "*I think the first system [text-based interface) was clearer … they [hospitals] were all in the same place while in the other [graphics-based interface] I had to search for them.*"

Graphics-Based Explanations Four key themes were identified. The first theme *Superficial understanding of the system (both controllers and novices),* reflected participants' difficulties in understanding how the system determined the color classifications. One said, "*I don't really know how it comes up with the information… how do you know that there's a short queue at Karolinska, for example?*" One controller explained: "*Here, the assistant might as well just decide on A because I can't really assess it… I wouldn't say I'm stuck, but I just can't evaluate it. Why is it red or yellow? Where are the boundaries?*"

The second theme was Easy to judge using colors (both controllers and novices). Many, especially novices, found it easy to interpret the color codes because they had certain associations with them from before. The third theme was *Difficult to understand the icons (both controllers and novices).* While most participants understood the icons, several expressed that it can be difficult to understand what they mean if you have not seen them before. The fourth theme focused on *Design-related feedback (both controllers and novices).* One participant found the decision alternatives more challenging to locate on the map, compared to the text-based explanation. Another felt the map was unclear, as part of the route disappeared under the decision advisory box, "*Now you don't know how far it is, like Södersjukhuset could continue a bit further up underneath and become almost as long as Huddinge.*"

5 Discussion

This design orientated work explored different explanation presentation formats of a conceptual AI system's interface for U-space airspace traffic management. The first research question focused on how to design text-based and graphical explanations for different decision alternatives presented on a map-based interface. The two solutions developed offered distinct approaches to presenting explanations in a time-sensitive, high-stakes scenario, building on Schlatter and Levinson's [16] design principles. The text-based explanation provided information in bullet-list in a side panel, while the graphics-based explanation provided information in separate boxes spatially distributed on the map view next to the alternative (i.e., hospital) they referred to.

The second research question explored users' understanding and perceived satisfaction of the text-based and graphical explanations. Air traffic controllers rated the text-based explanation higher, considering it to offer more detail about the considered factors affecting the decision alternatives, while novices preferred the graphics-based explanation for being quicker and easier to understand. Controllers found the graphics-based explanation more difficult to assess, which negatively impacted their sense of control over the decision-making process. This may explain their lower trust ratings of the graphics-based explanations. Novices, on the other hand, valued the simplicity and speed the graphics-based explanation gave in time-pressured situations. This supports previous research by Parasuraman et al. (2000) that graphical formats can help reduce an operator's mental workload if it matches the operator's representation of the system.

Both groups found the color coding of the graphics-based explanation helpful, though controllers questioned their specific meaning. The text-based explanation was perceived as more detailed by controllers, but both groups found it time-consuming to process. Overall, results suggest a combination of elements from both design presentations could maximize the satisfaction and efficiency in processing the information.

5.1 Situation Awareness

Results indicated that both controllers and novices found both explanation interfaces effective in achieving the task goal. This suggests that both interface solutions supported the development of level 2 SA [9], which involves understanding the relevance of information to a goal. However, it may have been helpful to include the rationale behind the system's decisions in the explanation [1]. Although the interface solutions indicates that the system takes four factors into account, it could be beneficial to also explain why a factor is categorized as green, yellow, or red. As some participants suggested, this could be achieved by adding explanatory text alongside each icon. Neither interface solution addressed level 3 SA as they did not incorporate any of the factors identified considered as essential for achieving level 3 SA, such as presenting the system's "confidence" or its ability to manage future situations [1].

5.2 The Explanation Satisfaction Scale

Results related to Hoffman et al.'s [23] Explanation Satisfaction Scale show an overall positive experience among participants, although some caution is warranted. Some participants, especially novices, found it difficult to answer questions about the level of detail in the explanation, as they felt they lacked sufficient knowledge of the system. Additionally, the theme identified from the open-ended questions of "superficial understanding of the system" emerged, despite participants rating their understanding in the questionnaire as high. This could suggest varying interpretations of what constitutes "understanding" or that deeper comprehension was not necessary. Furthermore, the limited time participants had to view the explanation may have prevented them from fully grasping it. Overall, this indicates that the Explanation Satisfaction Scale may not have been a completely reliable measure of how satisfactory the explanation was, as the questionnaire responses may not reflect participants' true experience.

5.3 Study Limitations

This study had several limitations in its design and execution. The testing of only two designs (text-based and graphical) limited the exploration of potential variations within each category. The use of simplified explanations in time-sensitive scenarios may have omitted relevant information, focusing primarily on decision-making factors rather than other potentially important aspects of explainability. Additionally, a potential carryover effect could have impacted participants' understanding and evaluation of the second condition they encountered as the same scenario was used in both conditions. The inclusion of air traffic controllers as proxies for future UAM Coordinators may not fully represent the intended end-users. Furthermore, the small sample size, particularly of air traffic controllers, limits the generalizability of the findings.

6 Conclusions and Future Work

This study presents the design work and evaluation of two map-based interfaces for explaining rerouting recommendations in a safety-critical context of urban U-space traffic management. The purpose was to explore, through design using Schlatter and Levinson's [16] interface design principles, how explanations can be presented in different ways: using a text-based format and a graphics-based format. Findings from the evaluation show that the two presentation formats of explanations affected participants' perception and understanding of them in different way. Air traffic controllers preferred the text-based explanation, appreciating the level of detail, while novices preferred the graphical proposal for its simplicity and ease of understanding. The graphical explanations color coding made decision alternatives clearer, and was perceived both positively and negatively. Both groups noted benefits and limitations in each proposal, suggesting an optimal explanation might combine elements from both designs.

This study highlights the need to carefully consider how explanations are presented. Future research should explore tailoring explanation presentation to users' preferences, combining text and visuals for increased clarity, and reassess the value of using the Explanation Satisfaction Scale. Alternative design approaches and presentation format of explanations should be explored, including diagrams for the graphics-based explanation or continuous text, as opposed to bullet points for the text-based version. Additionally, exploring varying levels of human influence on the system could shed light on how different explanations formats impacts trust and satisfaction with the explanation.

Acknowledgments. This publication is based on work performed in the HAIKU Project which has received funding from the European Union's Horizon Europe research and innovation programme, under Grant Agreement no 101075332. Any dissemination reflects the authors' views only, and the European Commission is not responsible for any use that may be made of information it contains.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. National Academies of Sciences, Engineering, and Medicine: Human-AI Teaming: State-of-the-Art and Research Needs. The National Academies Press, Washington, DC (2022). https://doi.org/10.17226/26355
2. EASA: EASA Artificial Intelligence (AI) Concept Paper Issue 2: Guidance for Level 1&2 Machine Learning Applications. European Union Aviation Safety Agency (EASA) (2023) https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-roadmap-20. Accessed Mar 2024
3. EASA: Artificial Intelligence Roadmap 2.0. Human-centric Approach to AI in Aviation. European Union Aviation Safety Agency (EASA) (2024) https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-roadmap-20. Accessed Mar 2024
4. Endsley, M.R.: From here to autonomy: lessons learned from human–automation research. Hum. Factors. **59**(1), 5–27 (2017). https://doi.org/10.1177/0018720816681350
5. Lyons, J.B., Sycara, K., Lewis, M., Capiola, A.: Human–autonomy teaming: definitions, debates, and directions. Front. Psychol. **12**, 589585 (2021). https://doi.org/10.3389/fpsyg.2021.589585
6. Chen, J.Y.C., Procci, K., Boyce, M., Wright, J., Garcia, A., Barnes, M.J.: Situation Awareness-Based Agent Transparency. Army Res. Lab., Aberdeen Proving Grounds, MD, Tech. Rep. ARL-TR-6905 (2014)
7. Westin, C., Borst, C., Hilburn, B.: Automation transparency and personalized decision support: air traffic controller interaction with a resolution advisory system. IFAC. **49**(19), 201–206 (2016). https://doi.org/10.1016/j.ifacol.2016.10.520
8. O'Neill, T., McNeese, N., Barron, A., Schelble, B.: Human–autonomy teaming: a review and analysis of the empirical literature. Hum. Factors. **64**(5), 904–938 (2022). https://doi.org/10.1177/0018720820960865
9. Endsley, M.R.: Supporting human-AI teams: transparency, explainability, and situation awareness. Comput. Hum. Behav. **140**, 107574 (2023). https://doi.org/10.1016/j.chb.2022.107574
10. Murdoch, W.J., Singh, C., Kumbier, K., Abbasi-Asl, R., Yu, B.: Definitions, methods, and applications in interpretable machine learning. Proc. Natl. Acad. Sci. **116**(44), 22071–22080 (2019). https://doi.org/10.1073/pnas.1900654116
11. Launchbury, J.: A DARPA Perspective on Artificial Intelligence (2017). https://www.darpa.mil/attachments/AIFull.pdf
12. Bhaskara, A., Skinner, M., Loft, S.: Agent transparency: a review of current theory and evidence. IEEE Trans. Hum. Machine Syst. **50**(3), 215–224 (2020). https://doi.org/10.1109/THMS.2020.2965529
13. Abdul, A., Vermeulen, J., Wang, D., Lim, B.Y., Kankanhalli, M.: Trends and trajectories for explainable, accountable and intelligible systems: an HCI research agenda. In: Proceedings of CHI, Montreal, Canada, pp. 21–26 (2018). https://doi.org/10.1145/3173574.3174156

14. Dudley, J.J., Kristensson, P.O.: A review of user interface design for interactive machine learning. ACM Trans. Interact. Intell. Syst. **8**(2), 1–37 (2018). https://doi.org/10.1145/3185517
15. Liao, Q.V., Gruen, D., Miller, S.: Questioning the AI: Informing design practices for explainable AI user experiences. In: Proceedings of CHI, Honolulu, HI, pp. 1–15 (2020). https://doi.org/10.1145/3313831.3376590
16. Schlatter, T., Levinson, D.: Visual usability: Principles and practices for designing digital applications. Morgan Kaufmann Publishers, Burlington (2013)
17. Airservices Australia and Embraer Business Innovation Center: Urban air traffic management. In: Concept of Operations. https://engage.airservicesaustralia.com/urban-air-traffic-management-concept-of-operations. Accessed Apr 2024
18. Eurocontrol: U-space ConOps and architecture. In: CORUS-XUAM Technical Report. https://www.sesarju.eu/node/4544. Accessed Apr 2024
19. HAIKU: Human-AI Teaming Knowledge and Understanding for Aviation Safety. https://haikuproject.eu. Accessed Mar 2024
20. Burkart, N., Huber, M.F.: A survey on the explainability of supervised machine learning. J. Artif. Intell. Res. **70**, 245–317 (2021). https://doi.org/10.1613/jair.1.12228
21. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. Hum. Factors. **46**(1), 50–80 (2004). https://doi.org/10.1518/hfes.46.1.50_30392
22. Rao, A.S., Georgeff, M.P.: BDI agents: from theory to practice. In: Proceedings of ICMAS, San Francisco, CA, pp. 312–319 (1995)
23. Hoffman, R.R., Mueller, S.T., Klein, G., Litman, J.: Metrics for Explainable AI: Challenges and Prospects. arXiv:1812.040608 (2018). https://doi.org/10.48550/arXiv.1812.04608
24. Sanneman, L., Shah, J.A.: The situation awareness framework for explainable AI (SAFE-AI) and human factors considerations for XAI systems. Int. J. Hum. Comput. Interact. **38**, 1772–1788 (2022). https://doi.org/10.1080/10447318.2022.2081282
25. Cambria, E., Malandri, L., Mercorio, F., Mezzanzanica, M., Nobani, N.: A survey on XAI and natural language explanations. Inf. Process. Manag. **60**(1), 103111 (2023). https://doi.org/10.1016/j.ipm.2022.103111
26. Zimmerman, J., Forlizzi, J., Evenson, S.: Research through design as a method for interaction design research in HCI. In: Proceedings of CHI, San Jose, CA, pp. 493–502 (2007). https://doi.org/10.1145/1240624.1240704
27. Arvola, M.: Interaktionsdesign och UX: om att skapa en god användarupplevelse, 2nd edn. Studentlitteratur, Lund (2021)
28. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model of types and levels of human interaction with automation. IEEE Trans. Syst. Man. Cybern. A Syst. Hum. **30**(3), 286–297 (2000). https://doi.org/10.1109/3468.844354
29. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qual. Res. Psychol. **3**(2), 77–101 (2006). https://doi.org/10.1191/1478088706qp063oa