How much is too much?
Levels of AI Explainability within Decision Support Systems' User Interfaces for Improved decision-Making Performance

Ghanshaam Sewnath & Jur Crijnen
Royal Netherlands Aerospace Centre NLR
THE NETHERLANDS

Ghanshaam.Sewnath@nlr.nl Jur.Crijnen@nlr.nl

ABSTRACT

Due to an increasingly complex military battlefield, defences are looking for state-of-the-art solutions that provide operators with the tools to enable a faster and more effective decision-making process than the adversary. These tools are often referred to as Decision Support Systems (DSS) which have persistently been used over the past decades. AI technology is often implemented in DSSs to ensure a lower error rate and faster decision-making when compared to individual human performance. The effectiveness of such an implementation in a DSS, largely depends on the operator's ability to understand, and as a result trust, the advice provided by the AI. Explainable AI (XAI) allows a user to understand how the system came to its advice regarding a decision by visualizing the process in the User Interface (UI) of the DSS. However, this also comes with it an inherent issue, namely: how much of the process should be presented to the user before the user becomes overloaded, decreasing the operator's decision-making performance?

Within this research, an AI-driven application has been developed which assists the operator in planning a military helicopter mission. In this scenario, the operator needs to find two appropriate Landing Zones (LZs) for the soldiers on board the helicopters to approach a terrorist compound in a small city area. The DSS supports the process of selecting appropriate LZs to land the helicopters, taking into account various aspects, such as distance to target area, spot size, surface type, and slope. In order to evaluate how much transparency is needed to achieve optimal levels of trust and task performance, four levels of explainability were defined, each with increased levels of information transparency and control. For each of the four levels, unique UIs were designed, developed and evaluated in testing sessions. The results indicate an increased performance (less time for decision making, high percentage of correct LZ decisions, and a low deviation between the perceived and actual score for the submitted LZs reflecting a decent human machine interaction) for the third and fourth UI design, which offer much more information and more interaction possibilities than the first two levels. Results also suggest that users prefer to personalize their UI to meet their role, experience level, and personal preferences.

1.0 INTRODUCTION

Due to an increasingly complex military battlefield, defences seek for state-of-the-art solutions which provide the operators with the tools to enable a faster and more effective decision making process than the adversary [1]. These tools are often referred to as Decision Support Systems (DSS) which have been used for the past decades as an essential part of military operations [2]. Nowadays often AI methods are implemented in DSSs to allow for low error rate and high speed compared to humans [3]. The effectiveness depends on the information supply to the DSS which often tends to be incomplete, unreliable or not in time. This is the starting point of the NLR Information Governed Operation (IGO) project which runs since 2019.

Previous IGO research focused on the human factors which are associated with operating according to the available information [4] and the usage of attentive user interfaces (AUI) for imagery analysts [5]. Parallel to this project, the Artificial Intelligence (AI) in Task Performance project researched the usage of DSSs in warfare and possibilities for Explainable AI (XAI) [6]. In that report applications for DSSs and the properties of a successful DSS are stated. However, concrete suggestions for technologies or user interface (UI) elements which provide explainability and trust in AI systems are to be researched. This paper continues on those topics.

Currently often a lack of explainability exist in military DSSs. Although this explainability can be interpreted on several levels of the DSS, we focus on the front end (user interface) of the application. Understanding of the algorithm by the operator is ignored because of the complexity of many DSSs and the speed at which decisions are to be made in warfare operations. The characteristics of the DSSs which enable decision making by the operator are researched from a human performance perspective. There will be emphasised on the level of certainty of applications and the UI elements which contribute to this.

1.2 Taxonomy of Explainability

Explainability of AI solutions can be divided in four taxonomies according to Liao et al. [7]. They emphasis on the UI and similar to our research ignore the technology behind the AI solutions. In Table 1, four XAI categories are presented in increasingly interactive and explainable order.

Table 1 Taxonomy of XAI methods [7]

Category of Methods	Explanation Method	Definition
Explain the model (Global)	Global feature importance.	Describe the weights of features used by the model (including visualization that shows the weights of features).
	Decision tree approximation.	Approximate the model to an interpretable decision-tree.
	Rule extraction.	Approximate the model to a set of rules, e.g., if-then rules.
Explain a prediction (Local)	Local feature importance and saliency method.	Show how features of the instance contribute to the model's prediction (including causes in parts of an image or text).
	Local rules or trees.	Describe the rules or a decision-tree path that the instance fits to guarantee the prediction.
Inspect counterfactual	Feature influence or relevance method.	Show how the prediction changes corresponding to changes of a feature (often in a visualization format).
	Contrastive or counterfactual features.	Describe the feature(s) that will change the prediction if perturbed, absent or present.
Example based	Prototypical or representative examples.	Provide example(s) similar to the instance and with the same record as the prediction.
	Counterfactual example.	Provide example(s) with small differences from the instance but with a different record from the prediction.

1.3 Research questions

Based on Liao et al.'s Taxonomy of XAI methods, a set of four XAI User-Interface Levels (UI L1 – UI L4) is defined with each level presenting an increasing level of AI transparency and control [7]. These four UI Levels are further described in Chapter 2.2.1.

This report will answer the following research question, based on the proposed XAI UI Levels:

- Which XAI UI elements should be applied to achieve optimal operator understanding and support when designing and developing an XAI for a DSS?

In order to answer the main research question, the following sub-questions have been defined:

- Which XAI UI elements contribute to the acceptance of the AI's advice?
- Which XAI UI elements increase the operators' trust in the AI's advice?
- Which XAI UI elements contribute to the operators' performance in using the DSS system?
- Which XAI UI elements contribute to the operators' situational awareness when using the DSS system?

The hypothesis for this research question is that a higher level of information transparency and control over the provided AI advice will also increase acceptance, trust, performance and situational awareness (SA). However, not all information and controls should be available at all times, but only the information and control that serve most users' needs at that time. Providing too much information and control may be overwhelming, which would results in more mistakes being made, lower trust and lower SA.

2.0 APPROACH

The experiment design is based on the literature findings. The four taxonomies of XAI methods are consulted to create four UIs, ranging from a low level of explainability to a high level of explainability [7]. These four interfaces are tested, while backend driven by the AI (rule-based) algorithm, for four assigned scenarios.

2.1 Participants

The participants of the experiment are not familiar with the specified task from a professional level. However, they are all experienced with computer interfacing and usage.

2.2 Experiment sessions

The experiment consist of four equivalent assignments which are randomly assigned to the participants. In all assignments, the operator is asked to designate two areas on a map which are suitable for the landing of a helicopter. This task is similar to the work of a helicopter mission planner. It is assumed that from the helicopter ground troops will be disembarked, which from there will capture a target. These so-called "Landing Zones" (LZs) have several attributes, such as slope, width, area, land type, and distance from target. According to the scenario, planes highlighted on the map appear to be suitable for LZ usages or not. The suitability is determined by the open street map (OSM) database, an AI (rule-based) algorithm, and the input of the operator [8]. The suitability is reflected by an LZ score. The participant will submit two LZ's during each scenario which are the most appropriate in his/her eyes. For submitting the participant will enter the area ID which belongs to the LZ. In addition to the LZ attributes, a red threat circle is placed on the map from which the helicopters should stay clear. An example of one of the assignments is presented in Appendix A.

The experiment is performed on a standard PC configuration (keyboards and mouse) including a Tobii Eye Tracker 4C (software: *Tobii Eye Tracking Core v2.16.8.214*).

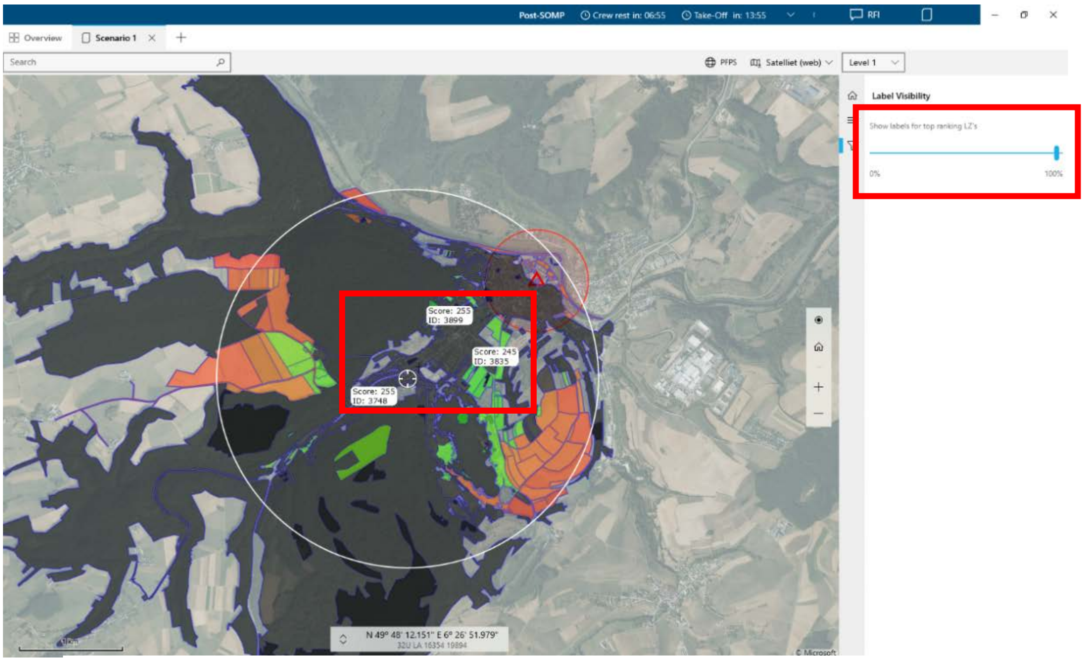


Figure 1 UI L1

2.2.1 Explainability levels

The figures presented in this chapter represent the UIs used during the experiments. For each level of explainability, the specific filters and output are highlighted with red square boxes in the figures. During the experiments these boxes are absent*UI L1*

The first and lowest level of explainable UI only provides LZ information to the operator without any form of interaction in filtering the presented information or insight in features contributed to this recommendation.

Figure 2 UI L2

The presented information is also very limited compared to higher levels. Only through the use of colours does the LZ provide the participant with guidance regarding the suitability of the LZ. The colour is linked to the score which is assigned to each plane. A screenshot of UI L1 is shown in Figure 1. The slider indicated by the right red box is used to (un)display the labels related to the areas. When this slider is set at ninety percent, the top ten percent of LZs according to the AI will show a label.

2.2.1.2 UI L2

The second UI provides the operator with more relevant information compared to L1 as the parameters affecting the AI's recommendation are presented in labels near the LZ. Similar to L1, the planes on the map also show which areas can be ignored as a potential LZ. Nonetheless, interaction possibilities are limited and only the percentage of labels displayed is customizable, as shown in Figure 2.

2.2.1.3 UI L3

Compared to L2, UI L3 provides more filtered interaction possibilities to the operator, as indicated in Figure 3. By using sliders, the minimum and maximum values for several attributes can be adjusted after which the output is updated. A selection of attributes is shown which should enable the operator to change the most crucial parameters. The features which are not indicated in the side-bar are presented in the labels in the map view (such as land type).

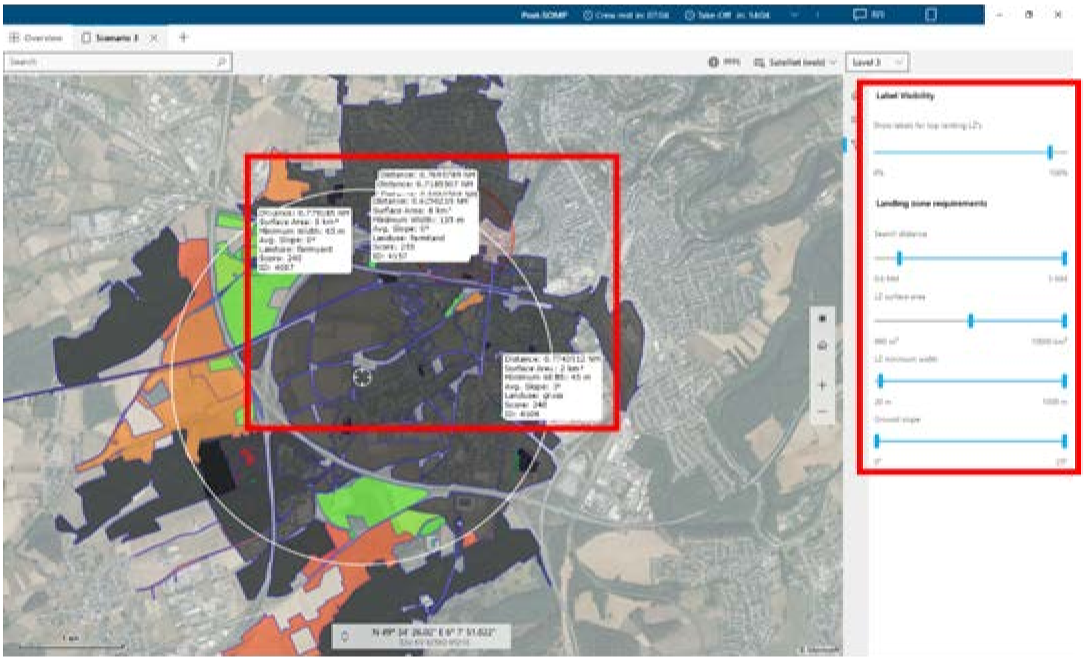


Figure 3 UI L3

2.2.1.4 UI L4

UI L4 has unfiltered interaction possibilities which make all features that the AI uses to calculate its recommendation score adaptable. This creates a transparent interface to enable the operator with all possible information. However, in addition the less interpretable functionalities are displayed as well. For example, the rule-based algorithm can be manipulated to focus more on the ground slope of the area instead of distance from target. This affects the landing zone score and therefore the colour of each plane on the map. All together, these extensive list of features create more a complex UI as can be seen in Figure 4.

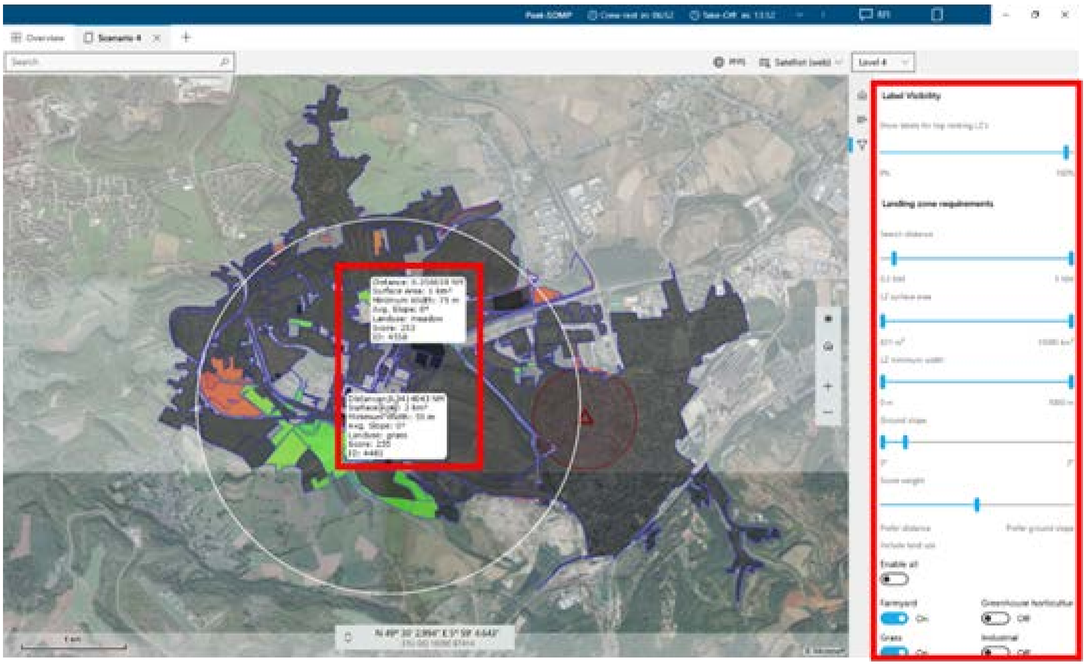


Figure 4 UI L4

2.2.2 Procedure

After the participants are welcomed for the experiment, a short general introduction to the application and research is given. To prevent from measuring a learning curve during the experiment, a familiarization to the interface is provided. This yields the navigation through the map and an explanation of the label visibility function; a feature which is enabled in all four user interfaces.

During the experiment, the four levels of interfaces are presented in random order. The assignment (scenario) is also provided at random by the system. This allows for ruling out the impact of sequencing and scenario difficulty on the user experience and performance of the task. The scenarios are designed to be of similar complexity. Before starting the scenario, the participant will read the assignment (as presented in Appendix A). Then, the scenario is selected and the UI is presented to the operator. The user is not aware of the different levels but only experiences the different (lack of) functionalities. After submitting the LZ ID's, the window is closed and the participant fills in a questionnaire. This is repeated for all four interfaces. Note that for the low level interfaces (especially UI L1) not all information is available to enable the participant to make the right decision. Additionally, it should be stated that an absolute correct answer may not be possible due to the amount of features and interface functions. This is close to reality, as weighing up multiple options is required in situations where ambiguous information is provided, such as designating LZs.

2.3 Measures

Within the research objective performance data and subjective experiences are measured.

2.3.1 Performance

The performance is reflected by three objective measures.

2.2.3.1 Decision duration

The duration for decision making is measured from the moment the scenario is selected until the LZ IDs are submitted. It includes the time participants require to familiarise themselves with the interface as well as information gathering until LZ selection.

2.2.3.2 Correct decisions

The percentage of correct decisions reflect the amount LZs that are submitted which could potentially be appropriate for landing. In each scenario two LZs are selected, therefore this metric could either be 100, 50, or 0 percent. Suitable LZs are areas which do not exceed hard limits (see next performance metric).

2.3.2.1 Landing Zone Score deviation

The Landing Zone Score is a value determined by the AI (rule-based) algorithm according to the attributes belonging to the scenario and ranges from 0 to 255. This score is either *Perceived* or *Real*. The perceived score is logged by the system and indicates the score as seen by the operator while submitting the LZ. The real score is calculated by the system by inserting the scenario parameters in the algorithm. If the operator inserts all parameters correct, then the perceived score is equal to the real score. However, for lower level user interfaces this is not possible as the interaction is limited and for the higher levels it remains challenging. The LZ score is zero when hard limits are exceeded (e.g. too close to target, wrong land type, above maximum ground slope) but holds a value for all remaining areas within limits. The score is determined according to ground slope and distance to target and ranked from nearest by target and least slope, to further away and to ground slope limit. As mentioned before, in UI L4 the weighing of both of these features can be adapted. The absolute difference (deviation) between *Perceived* and *Real* score is used to determine the alignment between the operator and the AI. Deviation close to zero indicates sound alignment between the system and the operator, reflecting decent insight in information for the operator to determine the appropriate LZs.

2.3.2 Subjective Measures

2.3.2.1 Acceptance Scale

The acceptance scale is used to measure the participants' handling with the user interfaces [9]. The items are presented in Appendix B.1. Item 3,6, and 8 are mirrored. Overall usefulness is retrieved from averaging item 1, 3, 5, 7, and 9 and satisfaction from the average of item 2, 4, 6, and 8.

2.3.2.2 Trust between People and Automation

Items from the Checklist for Trust between People and Automation are applied to measure user experiences [10]. These are indicated in Appendix B.2. Only a selection from the original set of items is used because of the (in)applicability of several statements.

2.3.2.3 Situational Awareness

The SA during the experiments is graded by the participants according to the scheme indicated in Appendix B.3. This reflects the perception of to what extend the interface contributes to the information provision for decent decision making.

3.0 RESULTS

The presented results only reflect the data of a limited number (n=8) of participants. The average age of the participants is 25.4 years (SD = 2.9). Due to a relatively low number of participants a statistical analysis of the results, apart from standard error, is left aside.

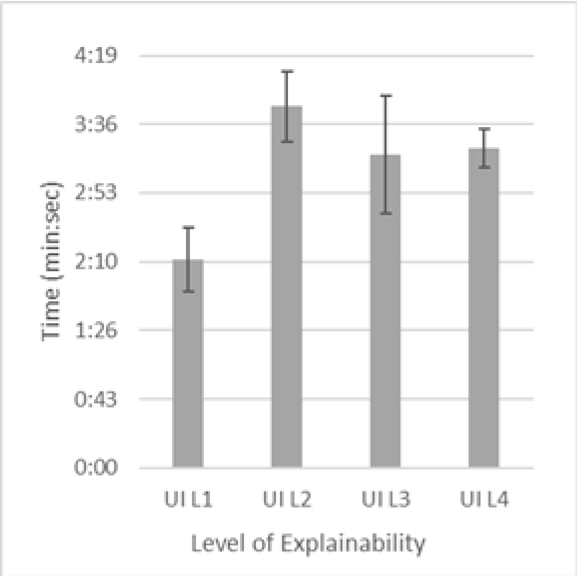


Figure 5 Average Time for Decision making per level of explainability with standard error

3.1 Performance

Figure 5 indicates the average time the participants require for submitting the LZs. Clearly the shortest time is required for submitting the LZs for UI L1. The duration for UI L2 is the highest (on average 3:47 minutes) while for UI L3 and UI L4 the time to assign appropriate LZs is relatively similar.

UI L1 provides barely information and therefore the decision appears to be made relatively quick. It might the case that the operator is just left in the unknown and makes his/her best guess. In UI L2 most of the required information is shown, however, investigation of the data has to be done manually. In UI L3 the participant is assisted by the system and therefore the decision making time is reduced. It is expected that especially for UI L4 the required time is higher due to the large amount of information and interaction possibilities. For UI L3 a high standard error is observed. Supposedly, this is due to the contrast between participants that accept the fact they miss information concerning the land types of the LZ (land types cannot be filtered in L3) and participants that try to identify the land type by observing the labels on the map. The latter is a rather time consuming process.

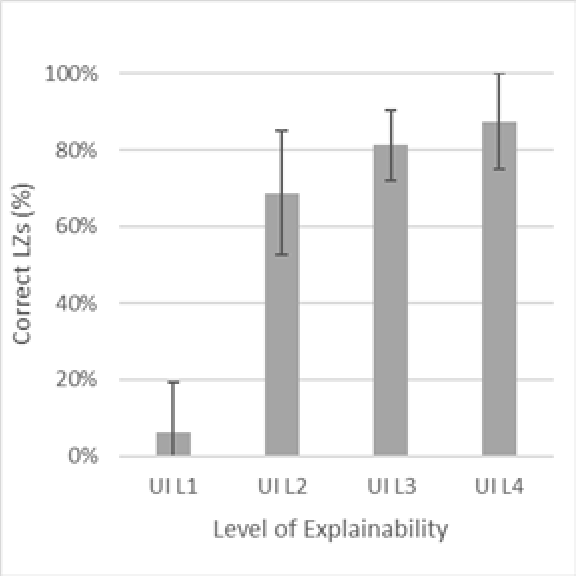


Figure 6 Average percentage of correct LZ Decision making per level of explainability with standard error

In Figure 6 the average percentage of correctly designated LZs is visualized for the different levels of explainability. As can be seen for UI L1 only few LZs are selected which do not exceed hard limits. As interaction possibilities increase the percentage of correct decisions increase. An optimum is observed for UI L4, for which the most interaction possibilities exist. However, the percentage of correct LZs is only 7 percentage points lower for UI L3, which raises the question whether the extensive list of land type options do contribute to the performance substantially.

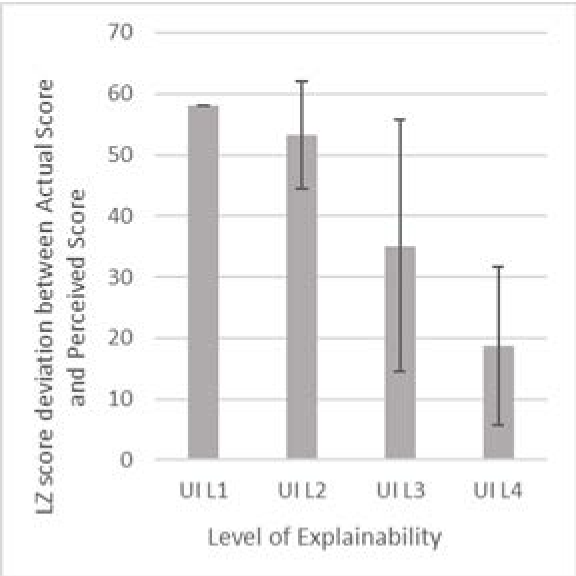


Figure 7 Average absolute LZ score deviation per level of explainability with standard error

Figure 7 shows the LZ score deviation for each UI level. The absolute deviation is retrieved by calculating the difference between the perceived score by the participant and the actual score. This reflects the level of truth of the provided information. The score deviation is only calculated for the submitted LZs which are identified as correct (not exceeding hard limits). UI L1 reveals a clear difference between the observed LZ suitability by the operator compared to the actual appropriateness. Due to a lack of explainability and interaction the operator is not able to adapt the parameters from which he/she has to retrieve the information for decision making. It should be noted that due to the low number of correct designated LZs for UI L1, the score deviation data is retrieved from only a small amount of LZs submitted. As the level of explainability increases the perceived score gradually approaches the actual score. This indicates the ability for the participant to retrieve the required information for their LZ decision. Also it reflects the alignment between the AI and the operator. The standard error for both UI L3 and UI L4 is rather high, resulting from outliers.

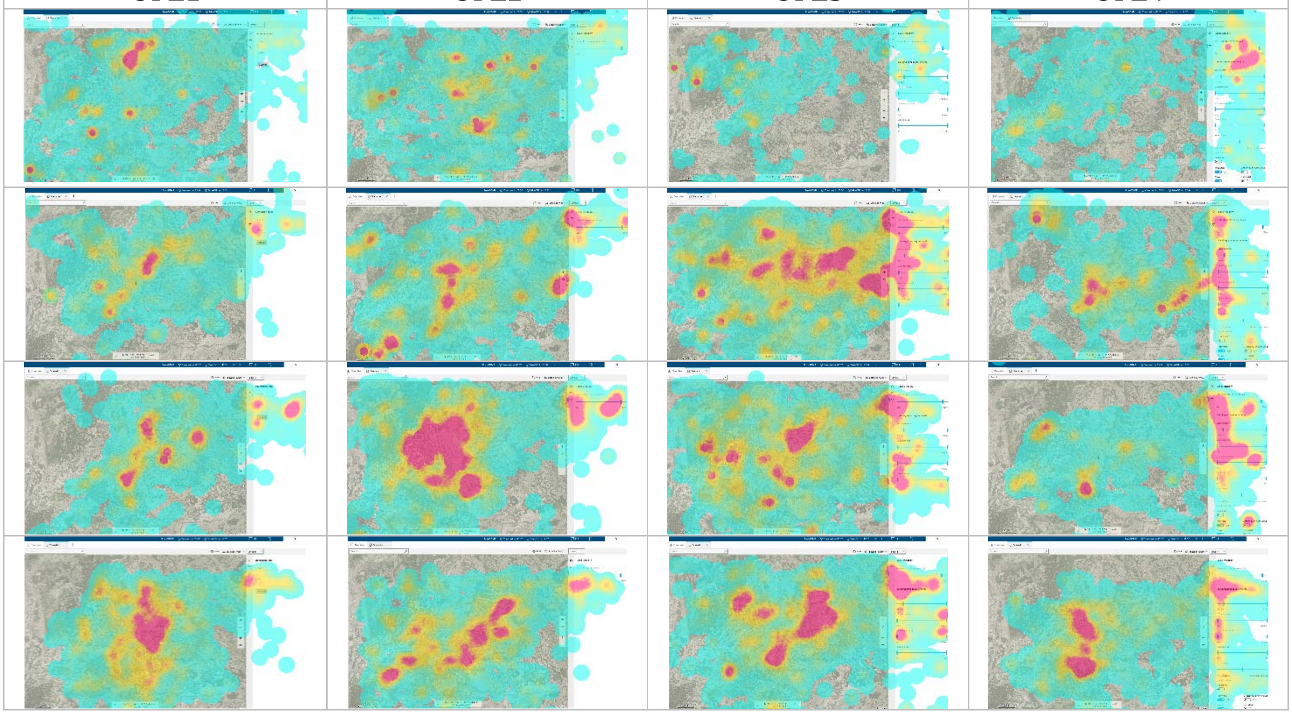
Interestingly, the quality of decision making does not improve much further from UI L3 on as the percentage of correct appointed LZs and the duration to decided remain roughly similar compared to UI L4. This would suggest that up to a certain level the performance does not further increase as explainability and interaction improves. However, alignment between the AI and the human operator increases considerably up to UI L4 which means the insight in the system's truth is better for UI L4 compared to L3, L2 and, L1.

3.2 Eye tracking

In Table 2 the eye tracking heatmaps for four participants (rows) are displayed. The columns indicate the four levels of UI explainability. The map view behind the heatmap is standardized as this changed during the experiment. Within this analysis, time spent perceiving the map and the side bar is measured and compared.

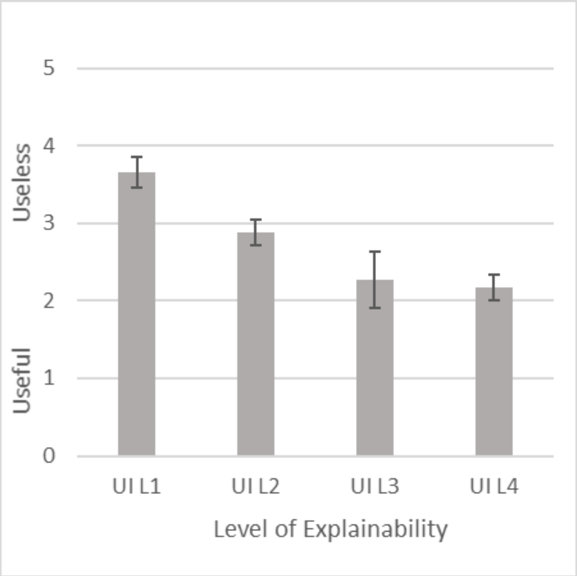
Table 2 Eye tracking heatmaps per level of explainability (columns)

UI L1	UI L2	UI L3	UI L4



In general an increased duration of watching the interface is observed for the higher levels of UI. Thereby, a shift from looking at the map for UI L1 and L2 to looking to the side-bar for controls is identified. This is as expected.

The change in heatmap between UI L3 and UI L4 is interesting. Whereas for UI L3 a fifty-fifty division between map and side-bar viewing is identified the sight moves to merely the side-bar for UI L4. Due to large amount of information and interaction methods for UI L4 the participants hardly watched the map before assigning LZs. this could be undesirable for instance when monitoring a live video feed which requires quick responses from the operator.



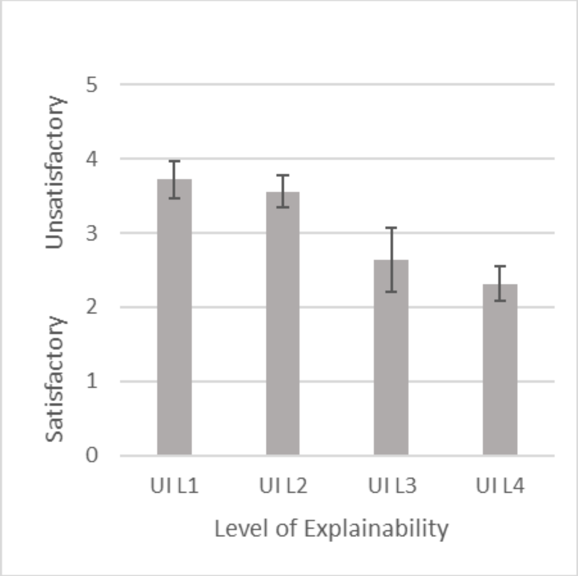


Figure 8 Acceptance level items score with standard error

3.3 Subjective Measures

In Figure 8 the combined items from the Acceptance level questionnaire are visualized. This metric reflects an usefulness and satisfactory rating with values ranging from 1 (useful/satisfactory) to 5 (useless/unsatisfactory). As expected, UI L1 seems the most useless, and unsatisfactory interface as not sufficient information is provided for adequate decision making. For usefulness the participants’ experiences considerably improve up to UI L3 and for satisfaction up to UI L4.

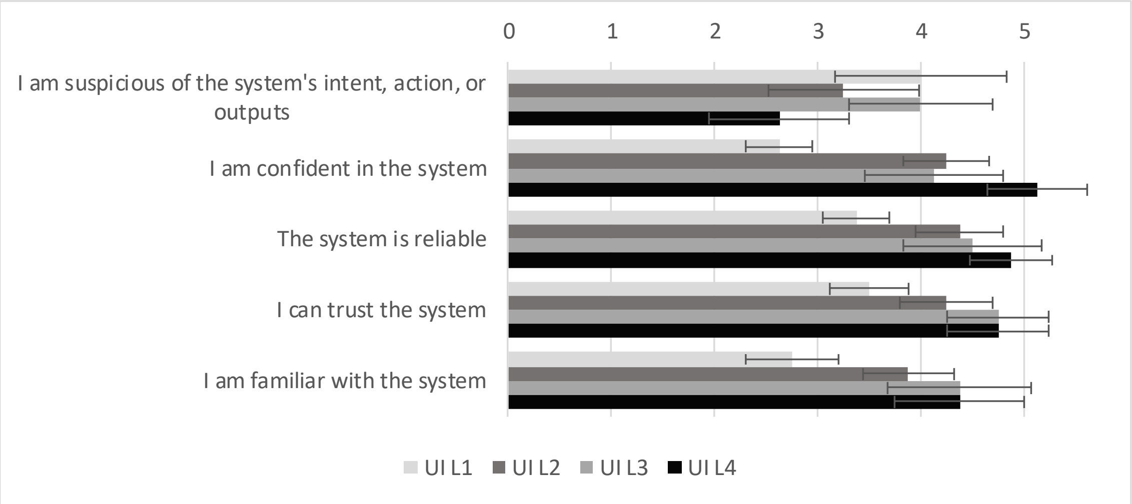


Figure 9 Checklist for Trust between People and Automation items score with standard error

Items from the Checklist for Trust between People and Automation are visualized in Figure 9 for each level of explainability. The highest level of confidence and reliability is experienced for UI L4. UI L3 and L4 have the same average score for trust and feeling familiar with the system. Remarkably, UI L3 indicates a slightly lower confidence level than UI L2. The first item, 'I am suspicious of the system's intent, action, or outputs', is more difficult to interpret. UI L1 and UI L3 score both high, whereas UI L2 and especially UI L4 score lower. For UI L2 this is probably experienced because of the increased insight in information due to the label next to the LZ. The same accounts for UI L4 because of its increased interactivity. The question remains why participants do not experience this to the same extent for UI L3.

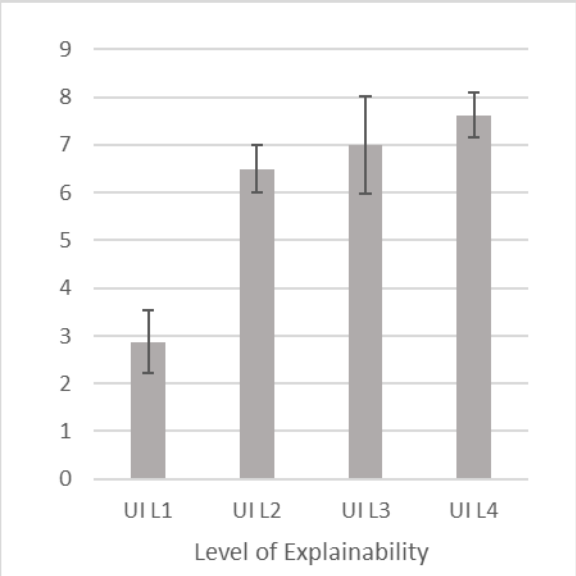


Figure 10 Average Situational Awareness scores with standard error

Situational Awareness (SA) is presented for all levels of explainability in Figure 10. The highest SA is observed for UI L4 but UI L2 and UI L4 are relatively high as well, the SA for UI L1 is noticeably lower and indicates a lack of information to allow for decent consideration of LZs and eventually the decision making. It should be noted that SA scores below seven indicate insufficient SA to perform the task. Only scores from eight and higher reflect "good SA".

3.3.2 Participant feedback

From the answers to the questions of the post-experiment discussion, a summary of the findings is described for each user interface.

3.3.2.1 UI L1

For UI L1, all participants experienced a lack of information and therefore the inability to perform the assigned task. Most comments to the interface concerned the lack of interpretability of the LZ scores and colours. As a result the participants speculate on which LZ would suffice. Few of them mentioned there effort to retrieve information (e.g. distance and land type) from the map view before submitting LZs.

3.3.2.2 UI L2

The most observed comment for UI L2 is the annoyance from information clutter on the map view. Due to the lack of interaction possibilities the participant were urged to retrieve all of the necessary information from the labels near the LZ. Solely the amount of labels is adaptable, however, for the most appropriate LZs participant were required to show numerous labels on the map. Selection of the suitable fields was therefore only possible when zooming in on the field to be able to see the coherent label and its information (to prevent overlap of labels). The absence of sliders resulted in frustration. Additionally, according to one participant hovering over LZs before presenting the labels could have prevented the clutter of labels improving the UX.

3.3.2.3 UI L3

UI L3 demonstrated an improved UX compared to UI L1 and L2. Especially the ability to adapt parameter values was mentioned to be desirable. Several participants commented on the lack of land type selection in the side bar. However, this was mostly overcome by checking the LZ information labels before submitting.

3.3.2.4 UI L4

Mixed opinions are observed for the UI L4. In general most participant preferred this interface over the others, yet, some mention the unintuitive usage of features which do not have 'hard' boundaries. As long as they do not know what is behind the prioritization of LZs by the algorithm, some participant would rather only receive binary advice (suitable / unsuitable). Also the list of land types was mentioned to be too extensive and the large amount of available sliders could induce the participant to set each parameter exactly before taking a look at the map. Most participants spent considerable time on manipulating the side bar options before watching the map.

4.0 CONCLUSION & RECOMMENDATION

From the results it can be concluded that UI L4 is the most preferred option, followed closely by UI L3. The number of mistakes made in selecting appropriate LZs and the duration to come to this decision is smallest for UI L4 and UI L3. It appears that more information and interactivity results on the user having a better understanding of how the AI calculates its score and therefore a stronger collaboration between both entities. Overall usefulness and satisfactory is also higher for UI L4 compared to the other levels. Similarly, Operator SA, confidence and reliability was the highest for UI L4, closely followed by UI L3. This is probably due to UI L3 generating a level of suspicion as a result of lacking a land type selection filter. Concerning the trust in the system, both UI L3 and UI L4 scored equally high.

For the measured results as well as questionnaire feedback, we can conclude that our hypothesis that more information and control leads to higher levels of acceptance, trust, performance and situational awareness is supported. However, presenting all information and options at once is not optimal for all users. Instead XAI UI designers should attempt to only provide information that is relevant to the users' need at the moment. This will reduce clutter, create focus in performing the task, and minimize information overload. For instance, a user may only want to see the labels of LZs over which they hover with their mouse and not all labels at all times.

Furthermore, the operator should be able to choose whether they have access to information controls and filters when required. This is also related to the recommendation not to provide information all at once, but instead provide the user with the option to get more information through interaction. For instance, in UI L4 all land use types are shown in a large list, but categorizing this list and making it collapsible per category can make in significantly less cluttered and more manageable.

Finally, operators should be able to adjust the UI to their own personal needs and preferences. Providing a settings option to increase font size, audio sounds, and even which UI buttons and visuals are presented should all be adjustable to the operator in order to provide an optimal user-experience.

REFERENCES

[1] Clark, B., Patt, D., & Schramm, H. (2020). Mosaic Warfare: Exploiting Artificial Intelligence and Autonomous Systems to Implement Decision-Centric Operations. CSBA. https://csbaonline.org/research/publications/mosaic-warfare-exploiting-artificial-intelligence-and-autonomous-systems-to-implement-decision-centric-operations

[2] Davis, P. K., Kulick, J., & Egner, M. (2005). Implications of modern decision science for military decision-support systems. Rand, Project Air Force.

[3] Gupta, S., Modgil, S., Bhattacharyya, S., & Bose, I. (2021). Artificial intelligence for decision support systems in the field of operations research: Review and future scope of research. Annals of Operations Research. https://doi.org/10.1007/s10479-020-03856-6

[4] Van Miltenburg, M. P. G., Bos, T. J. J., De Reus, A. J. C., & Den Ridder, S. V. (2019). Human Factors ontwikkeling in IGO (NLR-TR-2019-265).

[5] Van Miltenburg, M., & Sewnath, G. (2021). Application of Gaze-Based Adaptive Notifications to Support Image Analysts.

[6] Claessen, O., Crijnen, J. A., & Van den Oever, F. H. J. (2020). The Use of Decision Support Systems in Warfare (NLR-CR-2020-207).

[7] Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: Informing Design Practices for Explainable AI User Experiences. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–15. https://doi.org/10.1145/3313831.3376590

[8] OpenStreetMap contributors (2021). Plant dump retrieved from https://planet.osm.org, https://www.openstreetmap.org.

[9] Van der Laan, J.D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. Transportation Research - Part C: Emerging Technologies, 5, 1-10.

[10] Jian, J.-Y., Bisantz, A. M., Drury, C. G., & Llinas, J. (1998). Foundations for an Empirically Determined Scale of Trust in Automated Systems, United States Air Force Research Laboratory

APPENDIX A ASSIGNMENT

Assignment 1: Eliminate a high-value target in an urban area.

Designate two Chinook LZs at least 0.6NM from the target. The soldiers will approach the target and clear the area on foot. Hostile anti-aircraft guns are spotted near the target and therefore LZs should be at least 0.3 NM from the threat. The following LZs qualities should be kept in mind:

- • Ground slope: less than 4 degree
- • LZ surface area: larger than 1.100m2
- • Appropriate land types: grass, vineyard, farmyard, meadow
- • Minimum distance from target: 0.6NM
- • LZ minimum width: 20m

APPENDIX B QUESTIONNAIRES

B.1 Acceptance Scale [8]

1. Useful |__|__|__|__|__| Useless
2. Pleasant |__|__|__|__|__| Unpleasant
3. Bad |__|__|__|__|__| Good
4. Nice |__|__|__|__|__| Annoying
5. Effective |__|__|__|__|__| Superfluous
6. Irritating |__|__|__|__|__| Likeable
7. Assisting |__|__|__|__|__| Worthless
8. Undesirable |__|__|__|__|__| Desirable
9. Raising Alertness |__|__|__|__|__| Sleep-inducing

B.2 Items from the Checklist for Trust between People and Automation [9]

Below is a list of statement for evaluating trust between people and automation. There are several scales for you to rate intensity of your feeling of trust, or your impression of the system while operating a machine. Please mark on each line the point which best describes your feeling or your impression.

(Note: not at all=1: extremely=7)

1. I am suspicious of the system's intent, action, or outputs |__|__|__|__|__|__|__|
2. I am confident in the system |__|__|__|__|__|__|__|
3. The system is reliable |__|__|__|__|__|__|__|
4. I can trust the system |__|__|__|__|__|__|__|
5. I am familiar with the system |__|__|__|__|__|__|__|

B.3 Situational Awareness Scale