



HAL
open science

Challenges for operationalizing XAI in Critical Interactive Systems

Célia Martinie

► **To cite this version:**

Célia Martinie. Challenges for operationalizing XAI in Critical Interactive Systems. ACM CHI Workshop on Operationalizing Human-Centered Perspectives in Explainable AI (HCXAI @ CHI 2021), ACM, May 2021, Online Virtual Conference (originally Yokohama), Japan. hal-03221502

HAL Id: hal-03221502

<https://hal.science/hal-03221502>

Submitted on 12 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Challenges for operationalizing XAI in Critical Interactive Systems

Célia Martinie

ICS, IRIT

University Paul Sabatier Toulouse III

martinie@irit.fr

<https://www.irit.fr/recherches/ICS/people/martinie>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

HCXAI workshop at CHI 2021

Abstract

Large scale critical interactive systems (e.g. aircraft cockpits, satellite command and control applications...) aim to carry out complex missions. Users of such system usually perform predefined tasks for which they are trained and qualified. Nowadays, critical systems tend to embed an increasing number of on-board sensors, which collect large amounts of data to be interpreted and extrapolated, which tend to make more complex user tasks. Artificial Intelligence (AI) could be a powerful option to support the users in managing their tasks and handling this complexity. However, operationalizing AI in critical interactive systems requires proving that the AI behavior is consistent with user tasks, as well as transparent to the users and to the certification stakeholders. Explainable AI (XAI) is key as it could be a significant mean to satisfy this requirement. Nevertheless, at the same time, XAI will also have to comply with needs and common practices for the design and development of critical interactive systems. This position paper discusses the main challenges for operationalizing XAI in critical interactive systems.

Author Keywords

Explainable AI; interactive critical systems; task modelling.

Specificities of operations with critical interactive systems

User goals and tasks: The users (e.g. pilots, air traffic controllers...) mainly interact according to predetermined procedures and predetermined tasks [7]. There are two main types of tasks: mission tasks, which are goal oriented, and platform monitoring and control tasks, which aim to ensure that the current state of the systems provides support for accomplishing the mission tasks. These two types of tasks may interfere with each other. For example, if some elements of a system are not working properly, the user will have to know whether s/he is able to continue her/his mission, or which part of the mission s/he can continue. S/he will also have to perform some recovery tasks to take back the platform in a state that is acceptable for continuing the mission.

Training: Users are qualified and certified by an authority (e.g. JAA, EUROCONTROL...) or by their employer, depending on the application domain.

Introduction

In the field of large scale safety critical interactive systems, command and control user interfaces (e.g. aircraft cockpits, satellite monitoring and control applications...) are defined well in advance, during the development of the systems to be monitored and operated (e.g. engines for an aircraft, batteries for a satellite...) and users of these interfaces are trained before operations. Nowadays, critical interactive systems aim to carry out increasingly complex missions, sometimes outside the scope initially planned. In addition, systems being monitored and operated tend to embed a rising number of on-board sensors, which therefore leads to a considerable increase in the data collected (from the sensors or the data link), which is then interpreted and extrapolated. Users may have to deal with a large number of simultaneous tasks and their complexity requires managing cognitive resources to guarantee mission performance. Artificial Intelligence (AI) could be a powerful option to deal with these issues and to support the users in managing several tasks at the same time. However, operationalizing AI in critical contexts will require proving that the AI behavior is consistent with user tasks, as well as transparent to the users and to the certification stakeholders [1]. Explainable AI (XAI) is key because it could be a significant mean to satisfy this requirement [6]. Nevertheless, at the same time, XAI will also have to comply with needs and common practices for the design and development of critical interactive systems. This position paper highlights the main specificities of operations with critical interactive systems and of engineering and development of such systems. Using these specificities, next section introduces the main challenges for operationalizing XAI in critical interactive

systems. Third section presents an overview on the related work that deal with this topic. Fourth section proposes a possible way for the operationalization of XAI to meet critical interactive systems design and development constraints.

Challenges for the design, engineering and development of XAI in Critical Interactive Systems

Deploying XAI in critical contexts requires taking into account specificities of user tasks in critical context (highlighted in the side bar of this page) as well as applying critical systems engineering practices throughout its development (highlighted in the side bar of next page).

Support for elicitation of (types of) explanations that are relevant to user tasks

In the context of critical interactive systems, explanations have to be consistent with user tasks and have to support to reach the main mission goals. They have to suit the procedures and to integrate with the procedures. Beyond consistency between explanation and user tasks, as far as several methods exist for proposing explanations to the users [3], the design of XAI in critical interactive systems requires to ensure compatibility between the type(s) of explanations provided and user tasks for each user goal. This means that the types of explanations have to be chosen to be the most relevant to mission tasks, as well as to the possible contexts in which they could be provided. For example, in case where the users need to handle an emergency issue in a restricted amount of time, a static explanation may better suit the operational needs than an interactive explanation because users may not have

Specificities of engineering and development of critical interactive systems

Dependable computing:

Engineering critical interactive systems requires identifying and analyzing threats and faults that may impair systems functioning. It also requires being able to develop and integrate several types of mechanisms to prevent, remove, tolerate or forecast these faults [4].

Certification: Standards support for ensuring a certain level of safety for systems, software and their deployment in operational context. For example, the Certification Specification 25 for large aeroplanes [5], in its *section 1302*, states that "... installed equipment must be shown... to be designed so that qualified flight-crew members *trained in its use can safely perform their tasks associated with its intended function...*".

time to drill down or ask for different types of explanations.

Another important point is that to design XAI for critical interactive systems requires assessing the risk that an explanation may trigger a confusion and error for the next steps of the mission. In that case, designers also have to analyze the impact of a possible misunderstanding of the explanation on the success of the mission as well as on safety of operations.

Support for training

Users of critical interactive systems learn to apply specific procedures according to specific contexts, this in order to increase operators' performance and to decrease the number of potential human errors when using the system [2]. Training programs aim to provide operators with a predefined set of skills and knowledge before using the system. Embedding XAI in critical interactive systems will require identifying which user tasks are concerned with understanding AI functions output and their explanations, as well as to identify the most appropriate means to teach XAI analysis to the users in the context of their mission. The training program will thus require to implement these means to train the users to be able recognize, understand explanations and to use them to achieve their goals.

Support for engineering interfaces and interactions

As far as critical interactive systems have to be reliable, embedding XAI in critical interactive systems will require methods and techniques to check that XAI are reliable. For instance, there shall be means to ensure that a given action (or given set of correlated actions) performed by the AI in a given context, each time it is

performed, leads to provide the same explanation to the user. Furthermore, there shall also be means to ensure that there is an explanation for each action (or set of actions) performed by the AI.

Support for providing evidences during the certification process

Operationalizing XAI in critical interactive systems will require being able to show evidences about XAI dependability and about safety of operations performed by users of XAI. Techniques and tools for generating explanations will thus have to provide functions that offer support to show these evidences.

Towards modelling user tasks as a support to engineer XAI

Engineering critical interactive systems requires being able to identify systematically the user goals and tasks, as well as their temporal ordering and interleaving with systems functions. In order to take up the previously highlighted challenges, the systematic identification of user goals and tasks with XAI functions will also be required for their operationalization in critical contexts. Task models aims to represent, in a hierarchical and temporally ordered way, the tasks the user performs to reach a goal [9]. Fig. 2 shows a simple example of a task model that describes a selection of user tasks when playing an interactive computer based game. These user tasks correspond to the selection of a token, and Fig. 1 presents the interface of the game that matches these tasks. Task models support the identification and description of several types of task such as user task (cognitive, perceptive, and motor), abstract tasks, interactive tasks (input, output) and system tasks.

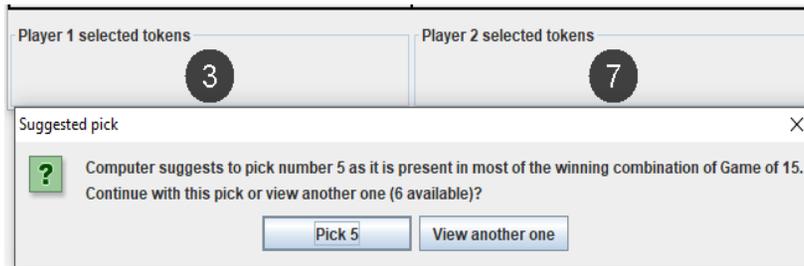


Figure 1. Screenshot of a recommendation for playing the next token (Game of 15 application)

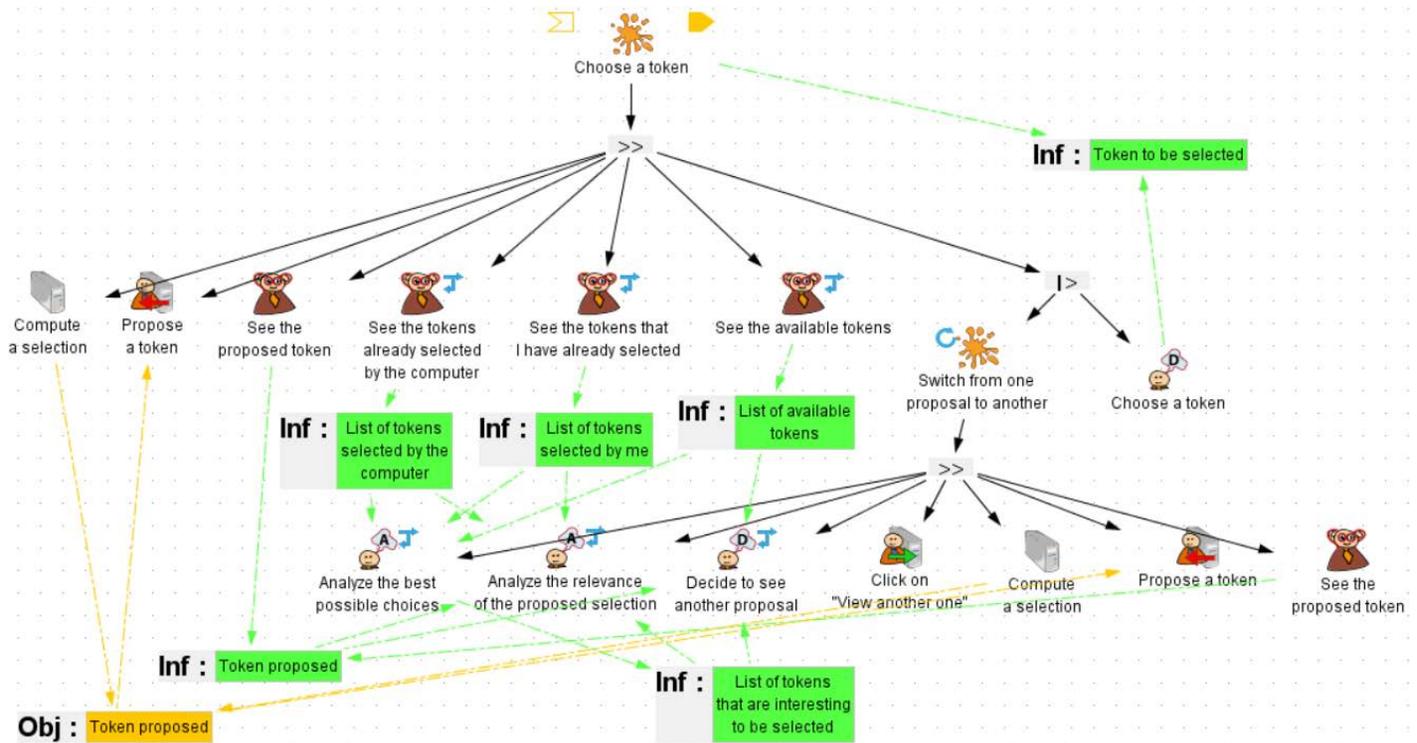


Figure 2. Task model describing main user tasks to select token (with game application recommendations)

Task models support the identification and representation of information and knowledge required for each task. In the same way they support (amongst others) identification and predictive analysis of effectiveness [11] and human errors [8], as well as training program development [12], they could support: analysis of conformance and consistency between XAI and user tasks, identification and analysis of possible human error with XAI, preparation and execution of training program with XAI...

Related work

In the aviation domain, Aditya et al. [1] argue that XAI is an enabler for users' acceptance and certification stakeholders' acceptance. They demonstrate the feasibility of embedding XAI in critical contexts with a prototype of tool that generates explanations for decisions made by a NASA-developed aircraft trajectory anomaly detection AI algorithm. From an engineering perspective, and to show the feasibility of complying with constraints of large commercial aircrafts development processes, Bouzekri et al. [5] identified a list of requirements to integrate a recommender system in the cockpit to support the management of alerts and failures. At last, in the healthcare domain, where people life can be at stake too, Tjoa et Guan [14] highlight similar challenges than the ones identified in this paper. They point out the significance of identifying what they call the "medically related subtleties", and that they exemplify with specificities of medical staff job. They conclude on the importance of identifying which information is required at which moment for a particular task.

Author's background and interest

The author has a long-term experience in the area of interactive critical systems, including design and implementation issues of command and control systems. Her main research skills and knowledge target methods, processes, techniques and tools to engineer critical interactive systems and for their integration within an organizational context for safe operations. She has been working closely with main actors in the area of safety critical interactive systems including Airbus for aircraft cockpits, CNES and ESA for satellite ground segments, and EUROCONTROL for air traffic management. She could share that perspective during the workshop. Participation to the workshop would also be beneficial to her work by exchanging with knowledgeable researchers in the area of XAI.

References

- [1] Aditya, P. Saraf, Kennis Chan, Martin Popish, Jeff Browder and John Schade. Explainable Artificial Intelligence for Aviation Safety Applications. AIAA AVIATION 2020 FORUM, June 15-19, 2020, virtual event, ARC, <https://doi.org/10.2514/6.2020-2881>
- [2] Aguinis, H., & Kraiger, K. (2009). Benefits of Training and Development for Individuals and Teams, Organizations, and Society. *Annual Review of Psychology* 2009 60:1, 451-475.
- [3] Vijay Arya, Rachel KE Bellamy, Pin-Yu Chen, Amit Dhurandhar, Michael Hind, Samuel C Hoffman, Stephanie Houde, Q Vera Liao, Ronny Luss, Aleksandra Mojsilovic, and others. 2019. One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques. arXiv preprint arXiv:1909.03012 (2019).
- [4] Avizienis, A., Laprie, J.-C., Randell, B., Landwehr, C., 2004. Basic concepts and taxonomy of dependable and secure computing. *IEEE Trans. Dependable Secure Comput.* 1 (1), 11–33.

- [5] E. Bouzekri, A. Canny, C. Fayollas, C. Martinie, P. Palanque, E. Barboni, Y. Deleris, C. Gris. Engineering issues related to the development of a recommender system in a critical context: Application to interactive cockpits, *International Journal of Human-Computer Studies*, Volume 121, 2019, pages 122-141, <https://doi.org/10.1016/j.ijhcs.2018.05.001>
- [6] EASA. (2020). Artificial Intelligence Roadmap: A human-centric approach to AI in aviation. Version 1.0, February 2020. EASA.
- [7] EASA. (2017). CS 25 Certification Specifications and Acceptable Means of Compliance for Large Aeroplanes, Amendment 19. EASA.
- [8] Racim Fahssi, Celia Martinie, Philippe Palanque. 2015. Enhanced Task Modelling for Systematic Identification and Explicit Representation of Human Errors. In: Abascal J., Barbosa S., Fetter M., Gross T., Palanque P., Winckler M. (eds) *IFIP TC13 Conference on Human-Computer Interaction – INTERACT 2015*. Lecture Notes in Computer Science, vol 9299.
- [9] Longo L., Goebel R., Lecue F., Kieseberg P., Holzinger A. (2020) Explainable Artificial Intelligence: Concepts, Applications, Research Challenges and Visions. In: Holzinger A., Kieseberg P., Tjoa A., Weippl E. (eds) *Machine Learning and Knowledge Extraction*. CD-MAKE 2020. Lecture Notes in Computer Science, vol 12279. Springer, Cham.
- [10] Célia Martinie, Philippe Palanque, Elodie Bouzekri, Andy Cockburn, Alexandre Canny, and Eric Barboni. 2019. Analysing and Demonstrating Tool-Supported Customizable Task Notations. *Proc. ACM Hum.-Comput. Interact.* 3, EICS, Article 12 (June 2019), 26 pages. DOI: <https://doi.org/10.1145/3331154>
- [11] Célia Martinie, David Navarre, Philippe Palanque, and Camille Fayollas. 2015. A generic tool-supported framework for coupling task models and interactive applications. In *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '15)*. ACM, New York, NY, USA, 244-253.
- [12] Célia Martinie, Philippe Palanque, David Navarre, Marco Winckler, and Erwann Poupart. 2011. Model-based training: an approach supporting operability of critical interactive systems. In *Proceedings of the 3rd ACM SIGCHI symposium on Engineering interactive computing systems (EICS '11)*. ACM, New York, NY, USA, 53-62.
- [13] Stanton, N. Baber, C., Walker, G., Houghton, J., McMaster, R. Stewart, R., Donald, H., Jenkins, D. Young, M., Salmon, P. 2008. Development of a generic activities model of command and control. *Cog., Tech. & Work*.
- [14] E. Tjoa and C. Guan, A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI, in *IEEE Transactions on Neural Networks and Learning Systems*, doi: 10.1109/TNNLS.2020.3027314