Human-agent Explainability: An Experimental Case Study on the Filtering of Explanations

Yazan Mualla[1], Igor H. Tchappi[1,2], Amro Najjar[3], Timotheus Kampik[4], Stéphane Galland[1] and Christophe Nicolle[5]

[1]CIAD, Univ. Bourgogne Franche-Comté, UTBM, 90010 Belfort, France

[2]Faculty of Sciences, University of Ngaoundere, B.P. 454 Ngaoundere, Cameroon

[3]AI-Robolab/ICR, Computer Science and Communications, University of Luxembourg, 4365 Esch-sur-Alzette, Luxembourg

[4]Department of Computing Science, Umeå University, 90187 Umeå, Sweden

[5]CIAD, Univ. Bourgogne Franche-Comté, UB, 21000 Dijon, France

Keywords: Explainable Artificial Intelligence, Human-computer Interaction, Agent-based Simulation, Intelligent Aerial Transport Systems.

Abstract: The communication between robots/agents and humans is a challenge, since humans are typically not capable of understanding the agent's state of mind. To overcome this challenge, this paper relies on recent advances in the domain of eXplainable Artificial Intelligence (XAI) to trace the decisions of the agents, increase the human's understandability of the agents' behavior, and hence improve efficiency and user satisfaction. In particular, we propose a Human-Agent EXplainability Architecture (HAEXA) to model human-agent explainability. HAEXA filters the explanations provided by the agents to the human user to reduce the user's cognitive load. To evaluate HAEXA, a human-computer interaction experiment is conducted, where participants watch an agent-based simulation of aerial package delivery and fill in a questionnaire that collects their responses. The questionnaire is built according to XAI metrics as established in the literature. The significance of the results is verified using *Mann-Whitney U* tests. The results show that the explanations increase the understandability of the simulation by human users. However, too many details in the explanations overwhelm them; hence, in many scenarios, it is preferable to filter the explanations.

1 INTRODUCTION

With the rapid increase of the world's urban population, the infrastructure of the constantly expanding metropolitan areas is subject to immense pressure. To meet the growing demand for sustainable urban environments and improve the quality of life for citizens, municipalities will increasingly rely on novel transport solutions. In particular, Unmanned Aerial Vehicles (UAVs), commonly known as drones, are expected to have a crucial role in future smart cities thanks to relevant features such as autonomy, flexibility, mobility, and adaptivity (Mualla et al., 2019c).

Still, several concerns exist regarding the possible consequences of introducing UAVs in crowded urban areas, especially regarding people's safety. To guarantee it is safe that UAVs fly close to human crowds and to reduce costs, different scenarios must be modeled and tested. Yet, to perform tests with real UAVs, one needs access to expensive hardware. More, field tests usually consume a considerable amount of time and require trained people to pilot and maintain the UAVs. Furthermore, on the field, it is hard to reproduce the same scenario several times (Lorig et al., 2015). In this context, the development of computer simulation frameworks that allow transferring real world scenarios into executable models, *i.e.* simulating UAVs activities in a digital environment, is highly relevant (Azoulay and Reches, 2019).

The use of Agent-Based Simulation (ABS) frameworks for UAV simulations is gaining more interest in complex civilian applications where coordination and cooperation are necessary (Abar et al., 2017). ABS models a set of interacting intelligent entities that reflect, within an artificial environment, the relationships in the real world (Wooldridge and Jennings, 1995). ABS is also used for different simulation applications in different domains (Najjar et al., 2017;

Mualla, Y., Tchappi, I., Najjar, A., Kampik, T., Galland, S. and Nicolle, C.
Human-agent Explainability: An Experimental Case Study on the Filtering of Explanations.
DOI: 10.5220/0009382903780385
In *Proceedings of the 12th International Conference on Agents and Artificial Intelligence (ICAART 2020) - Volume 1*, pages 378-385
ISBN: 978-989-758-395-7; ISSN: 2184-433X

Mualla et al., 2018b, 2019a). Due to operational costs, safety concerns, and legal regulations, ABS is commonly used to implement models and conduct tests. This has resulted in a range of research works addressing ABS in UAVs (Mualla et al., 2019b).

As UAVs are considered remote robots, communication with humans is a key challenge, since the human user is not capable, by default, of understanding the robot's State-of-Mind (SoM). SoM refers to the non-physical entities such as intentions and goals of a robot (Hellström and Bensch, 2018). This problem is even more accentuated in the case of UAVs since–as confirmed by recent studies in the literature (Hastie et al., 2017)–remote robots tend to instill less trust than robots that are co-located. For this reason, working with remote robots is a more challenging task, specially in high-stakes scenarios such as flying UAVs in urban environments. To overcome this challenge, this paper relies on the recent advances of the domain of eXplainable Artificial Intelligence (XAI) (Preece, 2018; Rosenfeld and Richardson, 2019) to trace the decisions of agents and facilitate human intelligibility of their behaviors when they are applied in a swarm of civilian UAVs that are interacting with other objects in the air or in the smart city.

In existing XAI solutions tackling the explanations of robots/agents behavior to humans, there is a problem with scalability *i.e.* the increasing number of robots/agents providing explanations. The bottle neck of this problem is the *human cognitive load* (Sweller, 2011). Humans have a threshold of how much information they can process at a time. Therefore, in such situations, there should be a way to reduce the cognitive load. This way should be controlled to assure the information with the highest importance is passed, *i.e.* filtering less important information.

In our previous work (Mualla et al., 2019d), we introduced a context model that provides first insights into a possible use case on this topic. In this paper, we define our agent-based model of human-agent explainability. Then, we discuss the filtering of explanations provided by agents to the human user to increase the understandability and instill trust in the remote UAV robots. Three different cases are investigated: "No explanation", "Detailed explanation" and "Filtered explanation". We conduct a human-computer interaction experiment based on ABS of civilian UAVs in a package delivery case study. The rest of this paper is structured as follows: Section 2 discusses related work, whereas Section 3 proposes our model and architecture. In Section 4, an experimental case study is defined and the experimental setup is stated, for which the results are presented, discussed, and analyzed in Section 5. Finally, Section 6 concludes the paper and outlines future works.

2 RELATED WORKS

Recently, work on XAI has gained momentum both in research and industry (Calvaresi et al., 2019; Anjomshoae et al., 2019). Primarily, this surge is explained by the success of black-box machine learning mechanisms whose inner workings are incomprehensible by human users (Gunning, 2017; Samek et al., 2017). Therefore, XAI aims to "open" the black-box and explain the sometimes-intriguing results of its mechanisms, *e.g.* a Deep Neural Network (DNN) mistakenly classifying a tomato as a dog (Szegedy et al., 2013). In contrast to this data-driven explainability, more recently, XAI approaches have been extended to explain the complex behavior of goal-driven systems such as robots and agents (Anjomshoae et al., 2019). The main motivations for this are: *(i)* as has been shown in the literature, humans tend to assume that these robots/agents have their own SoM (Hellström and Bensch, 2018) and that with the absence of a proper explanation, the user will come up with an explanation that might be flawed or erroneous, *(ii)* these robots/agents are expected to be omnipresent in the daily lives of their users (*e.g.* social assisting robots and virtual assistants).

XAI is of particular importance when the AI system makes decisions in multiagent environments (Azaria et al., 2019). For example, an XAI system could enable a delivery UAV modeled as an agent to explain (to its remote human operator) if it is operating normally and the situations in which it will deviate (*e.g.* avoid placing fragile packages on unsafe locations), thus allowing the operator to better manage a set of such UAVs. The example can be extended, in multiagent environment, where UAVs can be organized in swarms (Omiya et al., 2019; Kambayashi et al., 2019) and modeled as cooperative agents to achieve more than what they could do solely, and the XAI system could explain this to the remote operator. Our approach belongs to the goal-driven case and is different than other related works as it relies on a decentralized solution using agents. This choice is supported by the fact that the management of a UAV swarm must consider the physical distance between UAVs and other actors in the system. Additionally, autonomous agents represent, in our opinion, an adequate implementation of the autonomy of UAVs.

Recent works on XAI for agents employ automatically-generated explanations based on folk psychology (Harbers et al., 2010; Broekens et al., 2010). A folk psychology-based explanation commu-

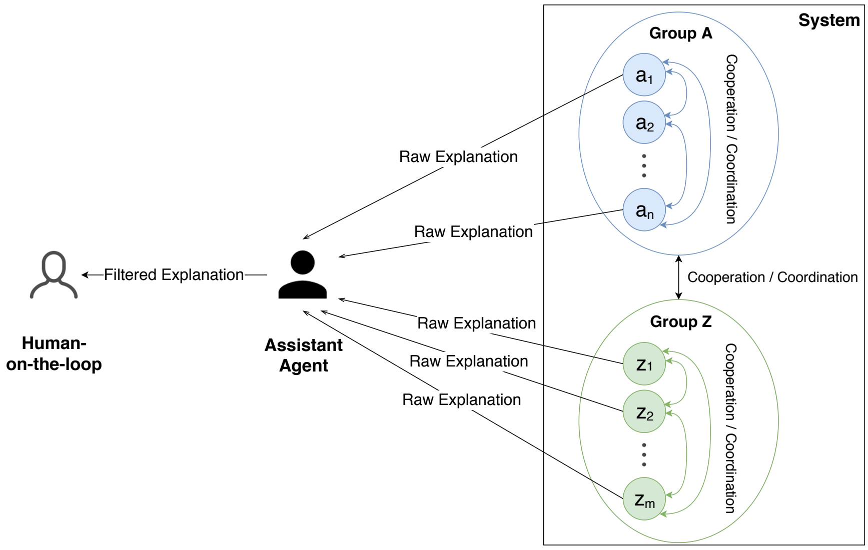

Figure 1: Human-agent EXplainability Architecture (HAEXA).

nicates the beliefs and goals that led to the agent's behavior. An interesting work discussed the generation and granularity (detailed or abstract) of the explanation provided to users in the domain of firefighting (Harbers et al., 2009). However, the paper mentioned that it could not give a general conclusion concerning the preferred granularity. More, and even though it mentioned that detailed explanation is better than abstract explanation in the special case of belief-based explanations, it overlooked the point that too many details could be overwhelming for humans.

3 ARCHITECTURE

Figure 1 shows the architecture of our model named Human-Agent Explainability Architecture (HAEXA). Through cooperation and coordination, agents can be regrouped to form an organized structure. On the right of Figure 1, two groups are shown: Group A with n agents and Group Z with m agents. Inside each group, there are inter-member relations of cooperation and coordination among the agent members of this group. Additionally, inter-group relations can also take place. All agents provide explanations to the *human-on-the-loop*: who benefits from these explanations in after-action decisions, *i.e.* the human does not control the behavior of the UAVs; hence they are fully autonomous. As humans could easily get overwhelmed by information, we introduce the *Assistant Agent* as a personal assistant of the human. In our model, the role of the assistant agent is to filter the raw explanations received by all agents and provide a summary of filtered explanations to the human to reduce the cognitive load.

The following interaction types take place between the HAEXA entities: cooperation, coordination and explanation. The first two are out of the scope of this paper and are widely discussed in the agent and multiagent community (see, *e.g.* Weiss (2013)). In our context, there are two types of explanation: raw explanations provided by the UAV agents, and filtered explanations provided by the assistant agent and based on the raw explanations. The filtering could be realized through different implementations including, learning the behavior of the human, explicit preferences specified by the human, adaptive to the situation in the environment, *etc.* In HAEXA, competition relations among UAV agents are not considered, as the purpose of the architecture is to provide a benefit for the human-on-the-loop which is a mutual goal between all agents. Additionally, it is possible that an agent does not provide any explanation.

4 EXPERIMENTAL CASE STUDY

Two hypotheses are considered in this paper based on the XAI literature (Keil, 2006):

$H1$**:** Explainability increases the understanding of the human-on-the-loop in the context of remote robots like UAVs

$H2$**:** Too many details in the explanations overwhelm the human-on-the-loop, and hence in such situations the filtering of explanations provides less, concise and synthetic explanations leading to higher understanding by her/him.

To prove the mentioned hypotheses, we have used an ABS to simulate an application of UAVs' autonomy and explainability. The case study is performed as a human-computer interaction statistical experiment. The participants will try the simulation and fill out a questionnaire. The results of the questionnaire will be used to investigate the participant understandability of the explanations provided by the UAVs.

4.1 Experiment Scenario

The experiment scenario is about investigating the role of XAI in the communication between UAVs and humans in the context of package delivery. In the scenario, one human operator oversees several UAVs that will provide package delivery services to clients. These UAVs will autonomously conduct tasks and take decisions when needed. Additionally, they need to communicate and cooperate with each other to complete specific tasks. The UAVs will explain to the Operator Assistant Agent (OAA) the progress of the mission including the unexpected events and the decisions made by them. Figure 2 shows the interaction between the actors in the proposed scenario. In the following, the steps of the scenario are detailed, with the numbers of steps shown in the figure:

1. When a client puts a request for delivering a package, a notification is sent to the OAA. The OAA will send it to all UAVs, so all UAVs are connected with each other and with the OAA using an assumed reliable network.
2. UAVs that are near, with a specific radius, to the package will coordinate to complete the delivery mission. The decentralized coordination (without the intervention of the operator) can be initiated for several reasons like deciding which UAV should deliver the package, or when the trip is long and it needs more than one UAV to carry the package in sequence.
3. The explanation needed from a UAV is generally about the mission progress, its decisions and its status, *e.g.* which UAV is assigned to the mission after the communication between UAVs, or when a UAV picks up the assigned package and is moving to destination. However, other important kinds of explanation are required regarding the unexpected events, *e.g.* a UAV arrives at the package location and did not find it, or see that it is damaged, or not according to description (maybe heavier). Another example is when a UAV needs charging so it ignores a nearby package.
4. Every UAV will provide explanations to the OAA that will show them to the operator. The UAV will assign a priority to each explanation. The OAA may filter the explanations received from UAVs to give a summary of the most important explanations to avoid overwhelming the operator with a lot of details. The filtering of explanation is based on a *filtering threshold* set by the operator to filter the explanations according to their priorities.

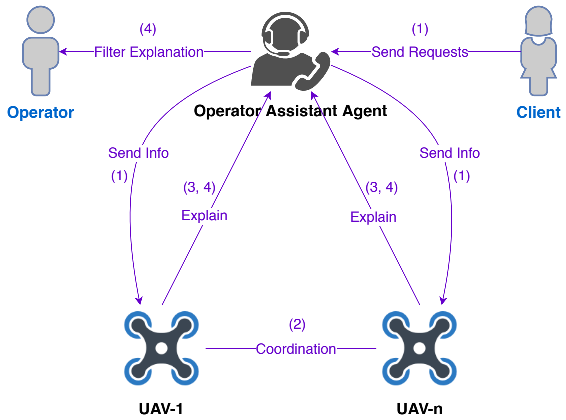

Figure 2: Interaction of actors in the experiment scenario.

To evaluate HAEXA, we focus on the steps 3 and 4, which are related to explainability. Steps 1 and 2 are necessary to build the experiment scenario. As cooperation and coordination are out of the scope of this work, only one group of agents is considered.

4.2 Methodology of the Experiment

The experiment requires the help of some human participants who will watch the simulation and then fill in a questionnaire built to aggregate their responses. It is vital that all participants have the same conditions when watching the execution of the simulation (quality of the video, same place and time, same instructions given, *etc.*). The organizing steps of the experiment are as follows:

1. 27 students (Bachelor, Master, PhD) of the university in the technology domain but in different specialties and different years have participated in this experiment. They were randomly divided into three groups (A, B and C).
2. The simulation is divided into sequences. Each sequence will handle an unexpected event (problem) or more, for example: low battery, damaged package, heavy package. All groups will watch the same simulation sequences but with different explanation capabilities: without textual explanation for Group A, with detailed textual explanation for Group B, and with filtered textual explanation for Group C. The first sequence is a very simple example with no problems to let the users be

familiar with the different actors and their icons. The last sequence is an overwhelming sequence with several UAVs (here 10).

3. After all groups watch the simulation sequences assigned to them, we ask all participants to fill in the questionnaire of the experiment, which is detailed in the next section. All participants have been informed that the experiment follows the EU General Regulation on Data Protection.

4.3 Building the Questionnaire

The questionnaire should include questions so that if we present to a human user the simulation that explains how it works, we could measure whether the user has acquired a useful understanding. Explanation Goodness Checklist can be used by XAI researchers to either design goodness into the explanations of their system or evaluate the explanations goodness of the system. In this checklist, only two responses (Yes/No) are provided. However, this scale does not allow for being neutral, *i.e.* a scale of 3 responses, and for some aspects there is a need for more granularity, *i.e.* the use of more options of the responses.

Studies showed that with the scale of 3 responses, usually the participants tend to choose the middle response because they prefer not to be extremist in their responses. Therefore, in social science the scale of responses is distributed to 5 responses. Explanation Satisfaction and Trust Scales are based on the literature in cognitive psychology and philosophy of science. Therefore, we opt to use these scales where the responses are distributed to a 5-point Likert scale (Hoffman et al., 2018): 1. I disagree strongly; 2. I disagree somewhat; 3. I'm neutral about it; 4. I agree somewhat; 5. I agree strongly.

4.4 Specific Implementation Elements

The experiment scenario is implemented using RePast Simphony (Collier, 2003), an agent-based simulation framework. The choice of this framework is based on a comparison of agent-based simulation frameworks for unmanned aerial transportation applications showing that RePast Simphony has significant operational and executional features (Mualla et al., 2018a).

The simulation has two panels: the monitoring panel or simulation map, and the explanation panel. The textual explanations have a natural language appearance, with dynamic numbers of the entities (UAVs, packages, charging stations, *etc.*). Some examples of explanations generated by a UAV agent are: "UAV 1 should carry package 3" or "Package 4 is damaged. I can't deliver it". The UAVs will assign priorities to their explanations, and the OAA will filter the explanations allowing to pass only those with a priority higher than the filtering threshold set in the initial parameters of the simulation.

5 RESULTS AND DISCUSSION

All the statistical tests performed in this section were *Mann-Whitney U* tests, as we are evaluating, at a time, one ordinal dependent variable (the 5 responses of the participants to a question) based on one independent variable of two levels (two groups of participants), and the total sample size of all the groups $N < 30$. For all tests, the Confidence Interval CI is 95% so the alpha value $\alpha = 1 - CI = 0.05$, and the $p - value$ will be provided per test below.

5.1 No Explanation vs. Explanation

In this section, we compare the 11 participants of the Group A (No explanation) on one hand with the 16 participants that have received explanation of both the Group B (Detailed explanation) and Group C (Filtered explanation) on the other hand.

Using a *Mann-Whitney U* test ($CI = 95\%, U = 45$, $p = 0.029$), Figure 3 shows the box plot that corresponds to the question: *Do you believe the only one time you watched the simulation tool working was enough to understand it?*, with 5 possible answers (Ref. Section 4.3). The box plot shows that the median response of Groups B and C ($med = 4$) is significantly higher than the median response of Group A ($med = 2$), *i.e.* the participants that received explanations agree more than the participants with no explanation that watching the simulation once is enough.

Using a *Mann-Whitney U* test ($CI = 95\%, U = 43.5$, $p = 0.018$), Figure 4 shows the box plot that corresponds to the question: *How do you rate your understanding of how the simulation tool works?*, with the following possible answers: 5 (Very high), 4 (High), 3 (Neutral), 2 (Low), 1 (Very low). The box plot shows that the median response of Groups B and C ($med = 4$) is higher than the median response of Group A ($med = 3$), *i.e.* the participants that received explanations rate their understanding of the simulation with a higher value than the participants that did not receive any explanation.

According to these two results, the first hypothesis $H1$ is proven. The respectful reader can notice that the questions of Figure 3 and Figure 4 have almost a similar goal. This is explained with the fact that when we have built the questionnaire, we have added

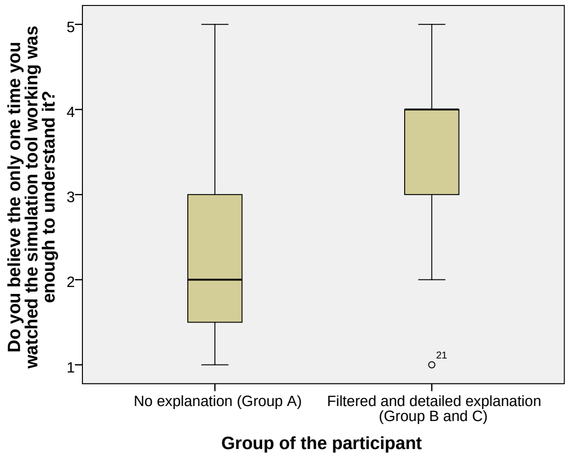

Figure 3: Do you believe the only one time you watched the simulation tool working was enough to understand it? (Explanation vs. No explanation).

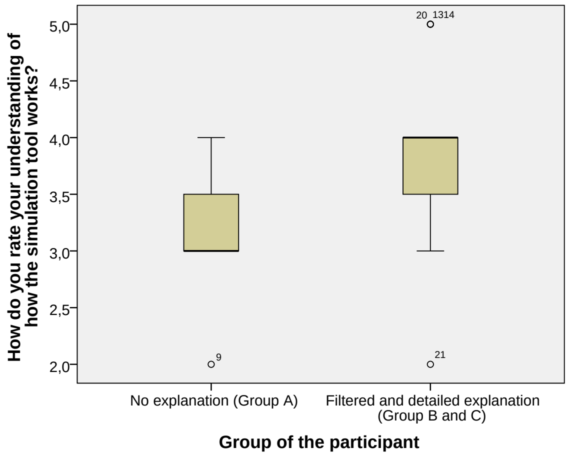

Figure 4: How do you rate your understanding of how the simulation tool works? (Explanation vs. No explanation).

some similar questions to assure the adherence and consistency of the responses of the participants.

5.2 Detailed Explanation vs. Filtered Explanation

In this section, we compare the 8 participants of the Group *B* (Detailed explanation) on one hand with the 8 participants of the Group *C* (Filtered explanation) on the other hand.

Using a *Mann-Whitney U* test ($CI = 95\%, U = 15$, $p = 0.058$), Figure 5 shows the box plot that corresponds to the question: *Do you believe the only one time you watched the simulation tool working was enough to understand it?*, with 5 possible answers (Ref. Section 4.3). The box plot shows that the median response of Group *B* ($med = 4$) is higher than the median response of Group *C* ($med = 3$), *i.e.* the participants that received detailed explanations tend to agree more than those receiving filtered explanations that watching the simulation once is enough to understand it. This result could be explained by the fact that when a participant receives a lot of explanation, she/he tends to feel more confident that watching the simulation once is enough. However, it is worth mentioning here that the $p-value$ was slightly higher than the α value for this test.

The last sequence shown to the participants included 10 UAVs and 16 packages. For this sequence, we asked a specific question related to the second hypothesis $H2$. Using a *Mann-Whitney U* test ($CI = 95\%, U = 13$, $p = 0.044$), Figure 6 shows the box plot that corresponds to the question: *The explanation of how the simulation tool works in the last sequence has too many details*, with 5 possible answers (Ref. Section 4.3). The box plot shows that the median response of Group *B* ($med = 3.5$) is higher than the median response of Group *C* ($med = 2.5$), *i.e.* the participants that received detailed explanations were overwhelmed by the details of the explanations, and think that the explanation was too much detailed compared to the participants that received filtered explanations.

Two findings can be drawn from the results of comparing the Group *B* vs. Group *C*:

1. More details are preferable by the participant and it increases its confidence that watching the simulation once was enough to understand it, but with a questionable significance (Figure 5). This agrees with the findings of (Harbers et al., 2009) where it is mentioned that the participant prefers more details in the explanation.
2. However, with the increase of scalability, the participant is eventually overwhelmed with too many details (Figure 6) and in this case, the filtering of the explanations is essential, and this proves the second hypothesis $H2$. More, filtering of the explanations gives more time to the participant to do other tasks, and this aspect of shared autonomy could be investigated in the future work.

5.3 Limitations

We have tried to normalize the conditions of the experiments by providing the exact experimentation conditions for all participants. However, there may be still some personal factors that make the experience of each participant different. Additionally, when choosing a sample from the population, this sample may have traits that are not representative for the entire population (*e.g.* knowledge and interest to technology, culture, *etc.*), and that influence the responses of

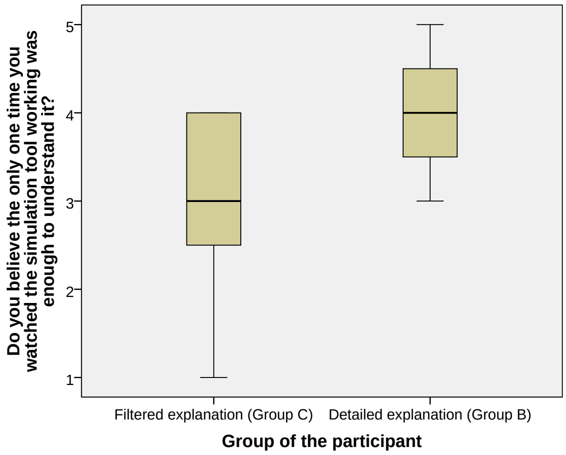

Figure 5: Do you believe the only one time you watched the simulation tool working was enough to understand it? (Detailed explanation vs. Filtered explanation).

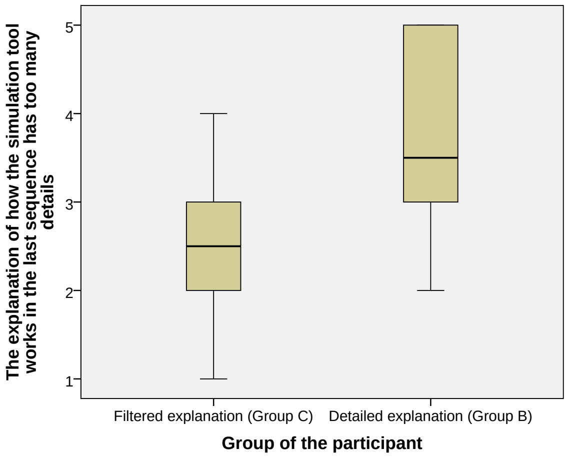

Figure 6: The explanation of how the simulation tool works in the last sequence has too many details. (Detailed explanation vs. Filtered explanation).

the questionnaire. Therefore, the generalization of the results is limited as the experiments were conducted on a sample consisting of students (Bachelor, Master, PhD) in the technology domain which does not necessarily represent the whole population.

6 CONCLUSION

While explainable AI is now gaining widespread visibility, there is a continuous history of work on explanation and can provide a pool of ideas for researchers currently tackling the task of explainability. In this work, we have provided our architecture HAEXA for modelling the human-agent explainability. HAEXA relies on filtering, where the related work is inadequate, the explanations of agents that are provided to the human user considering he/she has a cognitive load threshold of information to handle. To evaluate HAEXA, an experimental case study was designed and conducted, where participants watched a simulation of UAV package delivery and filled in a questionnaire to aggregate their responses. The questionnaire was designed based on the XAI metrics that have been established in the literature. The significance of the results was verified using *Mann-Whitney U* tests. The tests show that the explanation increases the ability of the human users to understand the simulation, but too many details overwhelm them; then, the filtering of explanations is preferable. The generalization of the results is a challenge that needs future research. In the proposed case study, and even though the human user sets the value of the filtering threshold, this value cannot be changed throughout the simulation. This means the filtering does not adapt to the changes of the situation in run time. Therefore, a future work is to implement a dynamic or adaptive filtering of the explanations.

ACKNOWLEDGEMENTS

This work is supported by the Regional Council of Bourgogne Franche-Comté (RBFC, France) within the project UrbanFly 20174-06234/06242. The first author thanks Cedric Paquet for his help in conducting the experiment.

REFERENCES

Abar, S., Theodoropoulos, G. K., Lemarinier, P., and O'Hare, G. M. (2017). Agent based modelling and simulation tools: A review of the state-of-art software. *Computer Science Review*.

Anjomshoae, S., Najjar, A., Calvaresi, D., and Främling, K. (2019). Explainable agents and robots: Results from a systematic literature review. In *Proc. of 18th Int. Conf. on Autonomous Agents and MultiAgent Systems*, pages 1078–1088.

Azaria, A., Fiosina, J., Greve, M., Hazon, N., Kolbe, L., Lembcke, T.-B., Müller, J. P., Schleibaum, S., and Vollrath, M. (2019). Ai for explaining decisions in multi-agent environments. *arXiv preprint arXiv:1910.04404*.

Azoulay, R. and Reches, S. (2019). UAV flocks forming for crowded flight environments. In *Proc. of 11th Int. Conf. on Agents and Artificial Intelligence, ICAART 2019, Volume 2*, pages 154–163.

Broekens, J., Harbers, M., Hindriks, K., Van Den Bosch, K., Jonker, C., and Meyer, J.-J. (2010). Do you get it? user-evaluated explainable bdi agents. In *German*

Conf. on Multiagent System Technologies, pages 28–39. Springer.

Calvaresi, D., Mualla, Y., Najjar, A., Galland, S., and Schumacher, M. (2019). Explainable multi-agent systems through blockchain technology. In *Proc. of 1st Int. Workshop on eXplanable TRansparent Autonomous Agents and Multi-Agent Systems (EXTRAAMAS 2019)*.

Collier, N. (2003). Repast: An extensible framework for agent simulation. *The University of Chicago's Social Science Research*, 36:2003.

Gunning, D. (2017). Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web*.

Harbers, M., van den Bosch, K., and Meyer, J.-J. (2010). Design and evaluation of explainable bdi agents. In *2010 IEEE/WIC/ACM Int. Conf. on Web Intelligence and Intelligent Agent Technology*, volume 2, pages 125–132.

Harbers, M., van den Bosch, K., and Meyer, J.-J. C. (2009). A study into preferred explanations of virtual agent behavior. In *Int. Workshop on Intelligent Virtual Agents*, pages 132–145. Springer.

Hastie, H., Liu, X., and Patron, P. (2017). Trust triggers for multimodal command and control interfaces. In *Proc. of 19th ACM Int. Conf. on Multimodal Interaction*, pages 261–268. ACM.

Hellström, T. and Bensch, S. (2018). Understandable robots-what, why, and how. *Paladyn, Journal of Behavioral Robotics*, 9(1):110–123.

Hoffman, R. R., Mueller, S. T., Klein, G., and Litman, J. (2018). Metrics for explainable ai: Challenges and prospects. *arXiv preprint arXiv:1812.04608*.

Kambayashi, Y., Yajima, H., Shyoji, T., Oikawa, R., and Takimoto, M. (2019). Formation control of swarm robots using mobile agents. *Vietnam J. Com. Sci.*, 6:193–222.

Keil, F. C. (2006). Explanation and understanding. *Annu. Rev. Psychol.*, 57:227–254.

Lorig, F., Dammenhayn, N., Müller, D.-J., and Timm, I. J. (2015). Measuring and comparing scalability of agent-based simulation frameworks. In *German Conf. on Multiagent System Technologies*, pages 42–60. Springer.

Mualla, Y., Bai, W., Galland, S., and Nicolle, C. (2018a). Comparison of agent-based simulation frameworks for unmanned aerial transportation applications. *Procedia computer science*, 130(C):791–796.

Mualla, Y., Najjar, A., Boissier, O., Galland, S., Haman, I. T., and Vanet, R. (2019a). A cyber-physical system for semi-autonomous oil&gas drilling operations. In *2019 Third IEEE Int. Conf. on Robotic Computing (IRC)*, pages 514–519. IEEE.

Mualla, Y., Najjar, A., Daoud, A., Galland, S., Nicolle, C., Yasar, A.-U.-H., and Shakshuki, E. (2019b). Agent-based simulation of unmanned aerial vehicles in civilian applications: A systematic literature review and research directions. *Future Generation Computer Systems*, 100:344–364.

Mualla, Y., Najjar, A., Galland, S., Nicolle, C., Haman Tchappi, I., Yasar, A.-U.-H., and Främling, K. (2019c). Between the megalopolis and the deep blue sky: Challenges of transport with UAVs in future smart cities. In *Proc. of 18th Int. Conf. on Autonomous Agents and MultiAgent Systems*, pages 1649–1653.

Mualla, Y., Najjar, A., Kampik, T., Tchappi, I., Galland, S., and Nicolle, C. (2019d). Towards explainability for a civilian uav fleet management using an agent-based approach. *arXiv preprint arXiv:1909.10090*.

Mualla, Y., Vanet, R., Najjar, A., Boissier, O., and Galland, S. (2018b). Agentoil: a multiagent-based simulation of the drilling process in oilfields. In *Int. Conf. on Practical Applications of Agents and Multi-Agent Systems*, pages 339–343. Springer.

Najjar, A., Mualla, Y., Boissier, O., and Picard, G. (2017). Aquaman: Qoe-driven cost-aware mechanism for saas acceptability rate adaptation. In *Int. Conf. on Web Intelligence*, pages 331–339. ACM.

Omiya, M., Takimoto, M., and Kambayashi, Y. (2019). Development of agent system for multi-robot search. In *Proc. of 11th Int. Conf. on Agents and Artificial Intelligence, ICAART 2019, Volume 1*, pages 315–320.

Preece, A. (2018). Asking 'why'in ai: Explainability of intelligent systems–perspectives and challenges. *Intelligent Systems in Accounting, Finance and Management*, 25(2):63–72.

Rosenfeld, A. and Richardson, A. (2019). Explainability in human–agent systems. *Autonomous Agents and Multi-Agent Systems*, pages 1–33.

Samek, W., Wiegand, T., and Müller, K.-R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296*.

Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation*, volume 55, pages 37–76.

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R. (2013). Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*.

Weiss, G. (2013). *Multiagent Systems*. Intelligent Robotics and Autonomous Agents. The MIT Press, Boston, USA.

Wooldridge, M. and Jennings, N. R. (1995). Intelligent agents: Theory and practice. *The knowledge engineering review*, 10(2):115–152.