A SYSTEMATIC REVIEW OF HUMAN FACTORS AND AI INFLUENCING OPERATOR PERFORMANCE IN MUM-T ENVIRONMENTS

Ivan de Souza Rehder[1], Moacyr Machado Cardoso Junior[1] & Emilia Villani[1]

[1]Aeronautics Institute of Technology

Abstract

This paper conducts a systematic quantitative literature review exploring the interplay between human factors and artificial intelligence (AI) in Manned-Unmanned Teaming (MUM-T) contexts. With AI's rapid advancement and its growing role in military operations, especially in UAV management, a deeper understanding of how human cognitive capabilities intersect with AI is crucial. This review meticulously evaluates the existing body of literature, following a methodical process of gathering information, building a database, and generating a thorough analysis. The results of this review are organized into principal thematic areas, including levels of autonomy, the dynamics of trust in human-machine interactions, cognitive workload management, experimental practices, and analysis of human factors. The findings underscore the intricacies of integrating AI with human operators in MUM-T scenarios, revealing both challenges and opportunities. This comprehensive literature overview aims not only to synthesize current knowledge but also to guide future research and development in the domain, underlining the need for strategies that effectively marry AI capabilities with human expertise in complex military operations.

Keywords: Manned Unmanned Teaming, Human Factors, Artificial Intelligence, Aviation, UAV

1. Introduction

In the contemporary landscape of military operations, the use of Unmanned Aerial Vehicles (UAVs) with manned aircraft has become increasingly prominent. This synergy, leveraging both human judgment and the precision of automated systems, aims to optimize mission efficiency in air operations. The primary focus of this report is to analyze the current state of artificial intelligence use within the Manned-Unmanned Teaming (MUM-T) framework, particularly considering the human factors influencing operators.

MUM-T represents a collaborative approach where pilots control numerous UAVs, enhancing the operational scope. The fusion of artificial intelligence in this setup enables operators to cover larger areas of interest. However, this comes with challenges, such as the potential overload of information, which can significantly impair situational awareness and increase the mental burden on operators. These challenges necessitate a comprehensive study not just of UAV capabilities but also of operator dynamics to develop systems that effectively transmit essential decision-making information.

Research in this domain has explored intelligent operational modes for UAVs and the use of mixed reality for improved situational awareness [1, 2]. Additionally, studies have focused on developing collaborative decision-making structures that integrate human and machine cognition, facilitating a more efficient MUM-T environment [3]. The concept of adaptive autonomy is also under investigation, with research examining the implications of varying autonomous control levels on the operator's cognitive load [4, 5].

Recent insights emphasize the importance of trust and verbal anthropomorphization in human-autonomous collaborations, underscoring the evolving nature of human-machine dynamics [6, 7]. Furthermore, the impact of workload on task-switching dynamics is a crucial aspect of understanding cognitive factors in supervising semi-autonomous systems [8]. This report aims to contribute to this field by identifying patterns in decision-making behavior and assessing the effects of increased autonomous control on human operators, thereby supporting the formulation of pertinent research questions for this project.

2. Method

The method employed in this review is adapted from the systematic quantitative literature review approach developed by [9] and taught at Griffith University[1]. This approach contrasts with traditional narrative reviews and is particularly advantageous for researchers exploring unfamiliar topics. Narrative reviews does not have a set of rules or a protocol to follow, so it no clear how the authors made their decisions about which paper to include or not [10].

Another approach is the meta-analysis, but this necessitates a comprehensive gathering and statistical analysis of articles, and this effort can be compared to an doctoral research.

The systematic method employed here effectively organizes and quantifies research material for the creation of a literature review. It is comprised of three specific stages, outlined below and depicted in Figure 1:

1. Information Acquisition
2. Database Creation
3. Review Generation

2.1 Information Acquisition

This initial phase is the first step towards the definition of the structure on which the review will be based on. It the moment that de research topic and its questions are defined, the keywords and database used and defined and how the articles will be choosen to be part of the review.

2.1.1 Topic Definition and Research Question Formulation

The first step is defining the research topic. A broadly defined topic might result in an extensive and potentially irrelevant collection of literature, whereas a narrow one could lead to only a few papers bringing not enough knowledge to the review. Following the topic definition, it is essential to formulate specific research questions. These questions are fundamental in guiding the scope and direction of the review. Examples of such questions include: What methods are commonly used in this field? What are the primary input and output variables? Which countries are actively involved in researching this topic? What are the main subjects or themes under investigation? This combined step of defining the topic and formulating research questions sets a focused foundation for the subsequent literature review process.

2.1.2 Keyword and Database Selection

This stage involves two interlinked processes: selecting keywords and choosing databases. Initially, the researcher selects keywords for database searches, an iterative step that may need refining to ensure the search yields an optimal number of relevant articles. Then, selecting the databases. Popular database options include Web of Science, Google Scholar, SCOPUS, and SAGE.

2.1.3 inclusion Criteria

Researchers develop criteria to evaluate the articles they retrieve. This step is vital for ensuring that the research results are reproducible. The evaluation of the articles can be by analysing only the title, the keywords, the abstract or the whole text if needed.

2.2 Database creation

After the information acquisition stage, it is probably that it resulted in a good number of papers (recommended is >15 and <300 [9]), but still not read. This moment is when the researcher structurize these papers, define the categories to classify them, and create summary table. At this moment, the reading of at least the abstract of the articles starts.

[1] https://www.griffith.edu.au/griffith-sciences/school-environment-science/research/systematic-quantitative-literature-review

2.2.1 Structure database

With the information of the abstract at least, it is possible to create a database with the are important for the review. Of course, these "important data" can increase or decrease a long the next steps, but is expected to do so. Some useful data can be extracted from the experiment, like the existence of an experiment, the scenario, the evaluation method, the number and the profile of the participants.

2.2.2 First articles

With this first version of an database ready, the first entries should be added to see how the categories and subcategories fit in.

2.2.3 Revise categories and Read all papers

Creating new categories, removing or modifying them are common after these first entries are analysed. This process can be repeated every time a new couple of articles are added to the database until all the articles are read and the categories fit in the scope of the research.

2.2.4 Summary tables

The last step before writing the article should be gather all the important data in summary tables

2.3 Results

After all these steps, the author(s) should now have a great knowledge about the researched topic. They should be able to see how the research is done, where is it applied, what are the gaps, who are the main researchers and where do they come from. The final step should be writing an review paper about the topic.

3. Applying the method

3.1 Information Acquisition

As already made clear, the chosen topic is **"Human Factors and Artificial Intelligence in Manned Unmanned Teaming"**.
The first step is to define a research topic. As the project will study the "Effectiveness of Artificial Intelligence in 'Manned Unmanned Teaming (MUM-T)' configurations evaluated considering Human Factors", the topic of this literature review should be one that analyzes the intersection of the project's main subjects. You should then review the:
"Use of artificial intelligence for pilot assistance in a MUM-T scenario"

- Can we group the articles in groups?;
- Does the article report any experiment? Which scenario/condition?;
- Which human factor(s) does it focus on? How they are evaluated.

To reach articles and papers that can answer our questions, the following keywords were selected.

- Manned Unmanned Teaming;
  Alternatively:
  - MUM-T;
  - Loyal Wingman.
- Artificial Intelligence;
  Alternatively:
  - AI;
  - Cognitive Systems.

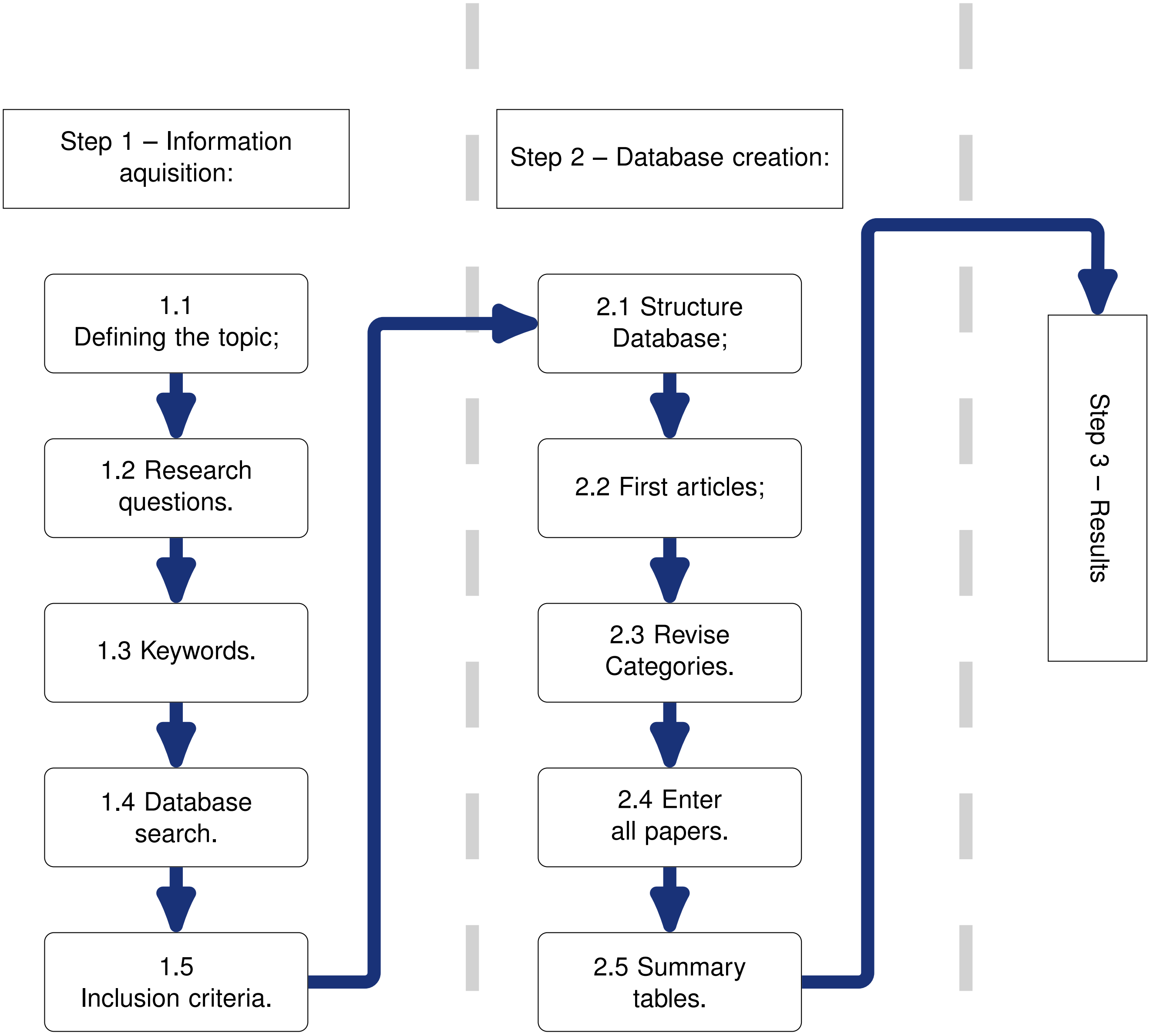

Figure 1 – Method's diagram (adapted from [9])

The database used in this literature review was SCOPUS, the abstract and citation database of the scientific content publisher Elsevier and the search command was:
(
TITLE-ABS-KEY(mum-t: OR (manned OR unmanned) AND "team*") OR TITLE-ABS-KEY ("loyal wingman")
) AND (
TITLE-ABS-KEY("cognitive systems") OR TITLE-ABS-KEY ("artificial intelligence" OR AI) OR TITLE-ABS-KEY("multi-agent" OR "multiagent" OR TITLE-ABS-KEY (multi) AND TITLE-ABS-KEY (agent))
) AND (
LIMIT-TO (DOCTYPE,"cp") OR LIMIT-TO (DOCTYPE,"ar")
) AND (
LIMIT-TO (LANGUAGE,"English")
)

This command found 212 results, a good number to create the starting database.

3.2 Database creation

An Excel file was created, and it's first sheet was a table with the 212 articles containing title, authors, abstract, DOI, URL.
The first articles were added in the sheet and the first connections were made. At this moment was possible to check if the research in the SCOPUS had enough keywords or filters.
To do that, a bibliometric analysis was made using R's Bibliometrix package

3.2.1 Bibliometrix

Bibliometrix [11] is an R package used for bibliometric analysis that is ideal for exploring and analyzing bibliographic data provided to it in a systematic and robust way. It is able to organize a list of authors, journals and countries that published the selected articles and also generate word clouds, thematic maps, a network of co-occurrence of themes.

3.2.2 Word cloud

Figure 2 gives greater prominence to the words with the highest frequency of mentions within the list of keywords proposed by the authors. The larger the source of the word, the more often it was mentioned. We can use this image as a checker to see if all the highlighted terms are consistent with being mentioned in our research. If not, it is ideal to make some changes to the search command in the previous step.

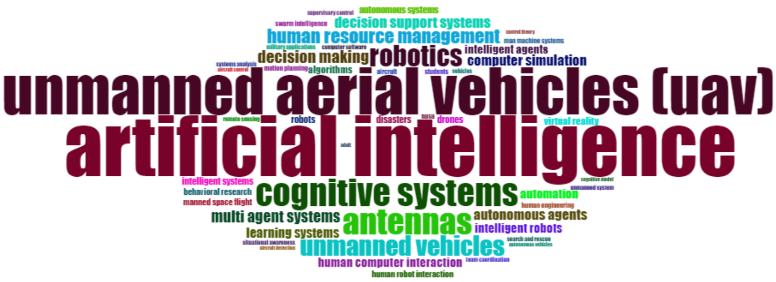

Figure 2 – Cloud of words mentioned in the authors' keyword list

.

3.2.3 Cooccurrence network

Figure 3 lists which topics have the highest co-occurrence within the articles supplied to Bibliometrix. This network is useful for identifying which other keywords orbit the topic being researched.

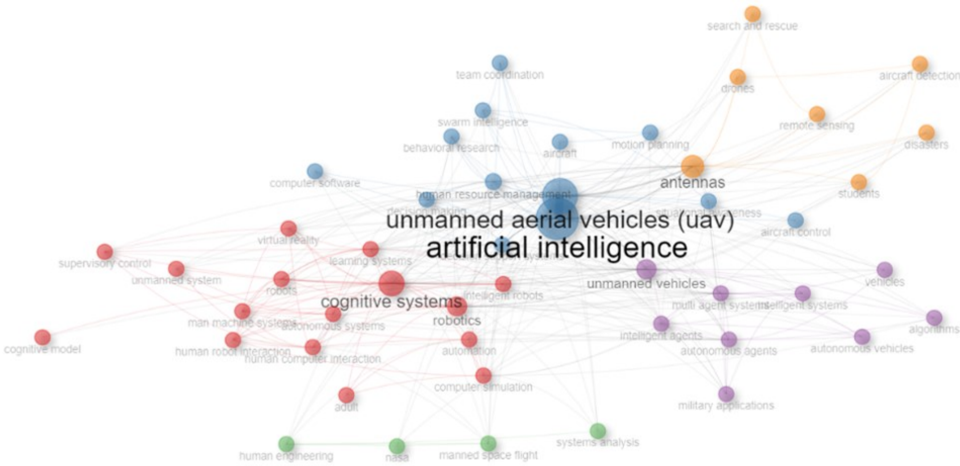

Figure 3 – Map of co-occurring themes.

3.2.4 *Abstract reading and ranking*

At this point we have a list of 212 abstracts which, according to the word cloud, should all be somewhat related, but this number is still too high to read the entire work. This step then proposed creating a metric to classify the abstracts in terms of their relevance to the topic studied. Those 212 abstracts were read and classified, many times, with different categories, until the final classification with scores based on the the power of 2. The categories were:

1 Military context; (Address reconnaissance, mapping, rescue, interception, etc. missions or situations)

2 UAV or USV; (Cite autonomous vehicles, whether air, naval or ground)

4 Decision making; (Addressed solutions to facilitate, or ways to structure, operator decision making)

8 Human Factors; (Take into account mental load, situational awareness, cognitive ability of the operator or other human factor)

16 MUM-T; (Studied the "Manned Unmanned Teaming" paradigm)

The sum possibilities are presented bellow in Table 1
This way, all articles had different scores that were related to a unique combination of relevant areas. This way it not only create a point system that indicated higher relevance papers, but also was ease to check which categories they we in. The next step would be the full reading of the papers in descending order. The goal of this study is to read all the papers until the score of 24 (at least 41 articles).

4. Results

So far, 26 articles have been read. They are presented, along with their respected score, in Table 7 in the end of this document. A summary of their score is presented on Table 2:
Based on the review questions, these are the answer so far:

4.1 Can we categorize the articles into distinct groups?

4 different groups were identified:

- Modes of operation, level of autonomy (LOA);
  These are works that study how the autonomy shifts between low/high autonomy and how this autonomy impacts the pilot. Normally evaluates the level of autonomy with human factors, such as situation awareness, mental workload or performance.

Table 1 – Summary of the possible scores

Key Areas	**Score**	**Key Areas**	**Score**
MUM-T + Military + UAV/USV + Decision making + Human Factors	31	Military + UAV/USV + Decision making + Human Factors	15
MUM-T + UAV/USV + Decision making + Human Factors	30	UAV/USV + Decision making + Human Factors	14
MUM-T + Military + Decision making + Human Factors	29	Military + Decision making + Human Factors	13
MUM-T + Decision making + Human Factors	28	Decision making + Human Factors	12
MUM-T + Military + UAV/USV + Human Factors	27	Military + UAV/USV + Human Factors	11
MUM-T + UAV/USV + Human Factors	26	UAV/USV + Human Factors	10
MUM-T + Military + Human Factors	25	Military + Human Factors	9
MUM-T + Human Factors	24	Human Factors	8
MUM-T + Military + UAV/USV + Decision making	23	Military + UAV/USV + Decision making +	7
MUM-T + UAV/USV + Decision making	22	UAV/USV + Decision making	6
MUM-T + Military + Decision making	21	Military + Decision making	5
MUM-T + + Decision making	20	Decision making	4
MUM-T + Military + UAV/USV	19	Military + UAV/USV	3
MUM-T + UAV/USV	18	UAV/USV	2
MUM-T + Military	17	Military	1
MUM-T	16		

Table 2 – Number of papers read by score

# Papers read	**Score**
4	31
1	29
2	28
5	27
1	25
2	24
11	<23

- Advanced technologies and coordination in autonomous systems;
  How to make the AI algorithms more efficient or how does the UAV decide which action to take or if needs to consult the pilot? These are the questions that this group tries to explore or answer.

- Trust dynamics in human-machine interactions;
  One hypothesis on how to interact with autonomous agents is to anthropomorphize them. This way the team efficiency between the operator and the autonomous agent should increase. At least, this group of works is focused on studying.

- Quality of service and conceptual frameworks;
  These works are focused on defining terms, standards, and protocols.

  It seems that

4.2 Does the article report any experiment? Which scenario/condition?

From the 26 read articles, 15 conducted experiments with 32 participants on average, even though 7 articles did not detail the number of participants. The scenarios were

- 4 ground station simulations;
- 2 numerical simulations;
- 2 driving simulations;
- 2 flight simulations;
- 2 case studies.

The majority perform experiments and it seems that a great part of them perform ground station experiments. Maybe because it is the most common use of the technology so far, or maybe is because the other one are more distant to the reality of the operation, on the get closer it would mean more costs to the study.

4.3 Which human factor(s) does it focus on? How they are evaluated?

From the 26 read articles, 17 analised human factors somehow. These articles evaluated, singularly or not, these bellow:

7 mental workload:

    2 Using physiological sensors;
    2 Evaluating the operators performance;
    1 Through subjective tests;
    3 Using other methods.

7 performance;

3 trust:

    2 Using eyetracker;
    2 Using HyperNEAT
    1 Through questionnaires;
    3 Using physiological sensors.

3 situational awareness:

    1 Through questionnaires.

Table 3 – Number of papers read by human factors

# Papers read	Human Factors	# Papers read	Method of evaluation
4	Mental workload	2	Physiological sensors
		2	Operator's performance
		1	Subjective tests
		3	Other methods
7	Performance		
3	Trust	2	Eyetracker
		2	HyperNEAT
		1	Questionnaires
		3	Physiological sensors
3	Situational awareness	1	Questionnaires

5. Conclusion

This review was able to summarize the efficiency of artificial intelligence in Manned Unmanned Teaming (MUM-t) by analyzing human factors. So far, the great conclusions, based on the review questions, are:

Can we categorize the articles into distinct groups?;

Yes, it is possible to categorize them in at least 4 different groups. Most of the research is studying the level of autonomy and the operation modes and how those can impact the operators performance/mental workload/situation awareness. A summary of this categorization is shown in Table 4

Does the article report any experiment? Which scenario/condition?

Yes. Most articles relate to an experiment or a case study. Most of them simulate/use a ground station control in their experiments. A summary of the experiment performed are show in Table 5

Which human factor(s) does it focus on? How they are evaluated?

Yes. Most articles are concerned with evaluating the operator's response to the system and most of them do so by evaluating the operator's mental workload. A summary of the experiment performed are show in Table 6

These findings show that while AI can greatly enhance MUM-T operations, it also brings challenges. Balancing AI capabilities with human cognitive limits is crucial for effective missions. Improving trust and reducing workload are key to making AI a robust support tool that enhances human decision-making and performance in complex scenarios.

Table 4 – Articles categorized

Author	Group
Cohen et. al. [6]	Trust dynamics in human-machine interactions
Cohen et. al. (2021) [6]	
Lange et al. (2013) [7]	
Azevedo-Sa et al. (2020) [12]	
Luke Petersen et al. (2019) [13]	
Jayaraman et al. (2018) [14]	
Kiam et al. (2022) [15]	Modes of operation, level of autonomy (LOA)
Andrews et al. (2020) [16]	
Zhao et al. (2020) [4]	
Chen et al. (2018) [2]	
Das et al. (2018) [1]	
Chen et al. (2017) [2]	
Schmitt et al. (2017) [17]	
Brand et al. (2017) [18]	
Agrawal et al. (2020) [19]	Quality of service and conceptual frameworks
Caldwell et al. (2019) [5]	
Schulte et al. (2016) [20]	
Rune et al. (2021) [21]	
Atdelzater et al. (2000) [22]	
Levulis et al. (2018) [23]	Advanced technologies and coordination in autonomous systems
Schmitt et al. (2018) [17]	
İşci et al. (2022) [24]	
Borck et al. (2015) [25]	
Maier et al. (2022) [3]	
Wei et al. (2007) [26]	
Wei et al. (2006) [27]	

Table 5 – Articles grouped by experiment category

Author	Experiment Category
Chen et al. (2018) [2]	Case study
Chen et al. (2017) [2]	
Lange et al. (2013) [7]	
Levulis et al. (2018) [23]	Flight managment
Andrews et al. (2020) [16]	
Atdelzater et al. (2000) [22]	
Schmitt et al. (2018) [17]	
Cohen et. al. (2021) [6]	Ground Station Simulation
Kiam et al. (2022) [15]	
Schmitt et al. (2017) [17]	
Zhao et al. (2020) [4]	
Borck et al. (2015) [25]	Numerical simulation
İşci et al. (2022) [24]	
Rune et al. (2021) [21]	Other
Azevedo-Sa et al. (2020) [12]	Other Simulator
Jayaraman et al. (2018) [14]	
Luke Petersen et al. (2019) [13]	
Das et al. (2018) [1]	UAV flight simulation
Agrawal et al. (2020) [19]	No Experiment
Brand et al. (2017) [18]	
Caldwell et al. (2019) [5]	
Maier et al. (2022) [3]	
Schulte et al. (2016) [20]	
Wei et al. (2006) [27]	
Wei et al. (2007) [26]	

Table 6 – Articles grouped by studied human factors

Author	Workload	Trust	Situational Awareness	Performance
Zhao et al. (2020) [4]	X	–	X	X
Andrews et al. (2020) [16]	X	–	X	X
Levulis et al. (2018) [23]	X	–	X	X
Cohen et. al. (2021) [6]	–	X	X	X
Schmitt et al. (2017) [17]	–	–	X	X
Atdelzater et al. (2000) [22]	–	–	X	X
Kiam et al. (2022) [15]	–	–	X	X
Azevedo-Sa et al. (2020) [12]	–	X	–	–
Luke Petersen et al. (2019) [13]	–	X	–	–
Lange et al. (2013) [7]	–	X	–	–
Wei et al. (2006) [27]	–	X	–	–
Chen et al. (2017) [2]	X	–	–	–
Agrawal et al. (2020) [19]	–	–	–	–
Das et al. (2018) [1]	–	–	–	–
Chen et al. (2018) [2]	–	–	–	–
Schmitt et al. (2018) [17]	–	–	–	–
Brand et al. (2017) [18]	–	–	–	–
Caldwell et al. (2019) [5]	–	–	–	–
Schulte et al. (2016) [20]	–	–	–	–
Borck et al. (2015) [25]	–	–	–	–
İşci et al. (2022) [24]	–	–	–	–
Rune et al. (2021) [21]	–	–	–	–
Jayaraman et al. (2018) [14]	–	–	–	–
Maier et al. (2022) [3]	–	–	–	–
Wei et al. (2007) [26]	–	–	–	–

6. Contact Author Email Address

mailto: ivan@ita.br

7. Copyright Statement

8. Acknowledgments

The authors would like to thank the financial support of the following entities and projects: CAPES-PRINT Project no. 88887.310617/2018-00, FINEP Project 01.20.0195.01, FINEP Project 01.22.0478.00 and FINEP Project 01.22.0313.00."

References

[1] Aditya Das, Patrik Kol, Cody Lundberg, Kris Doelling, Hakki Erhan Sevil, and Frank Lewis. A rapid situational awareness development framework for heterogeneous manned-unmanned teams. In *NAECON 2018-IEEE National Aerospace and Electronics Conference*, pages 417–424. IEEE, 2018.

[2] Jun Chen, Xudong Gao, Xiaowei Chen, and Qiang He. A shifting method for intelligent operational mode of uavs. In *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pages 526–531. IEEE, 2018.

[3] Siegfried Maier and Axel Schulte. A cloud-based approach for synchronous multi-pilot multi-uav mission plan generation in a mum-t environment. In *AIAA SCITECH 2022 Forum*, page 2345, 2022.

[4] ZHAO Zhe, NIU Yifeng, and SHEN Lincheng. Adaptive level of autonomy for human-uavs collaborative surveillance using situated fuzzy cognitive maps. *Chinese Journal of Aeronautics*, 33(11):2835–2850, 2020.

[5] Barrett S Caldwell, Megan Nyre-Yu, and Jordan R Hill. Advances in human-automation collaboration, coordination and dynamic function allocation. In *Transdisciplinary Engineering for Complex Socio-technical Systems*, pages 348–359. IOS Press, 2019.

[6] Myke C Cohen, Mustafa Demir, Erin K Chiou, and Nancy J Cooke. The dynamics of trust and verbal anthropomorphism in human-autonomy teaming. In *2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS)*, pages 1–6. IEEE, 2021.

[7] Douglas S Lange, Phillip Verbancsics, Robert S Gutzwiller, and John Reeder. Trust in sparse supervisory control. In *2013 AAAI Spring Symposium Series*, 2013.

[8] P. N. Squire and R. Parasuraman. Effects of automation and task load on task switching during human supervision of multiple semi-autonomous robots in a dynamic environment. *Ergonomics*, 53:951–961, 8 2010.

[9] Catherine Pickering and Jason Byrne. The benefits of publishing systematic quantitative literature reviews for phd candidates and other early-career researchers. *Higher Education Research & Development*, 33(3):534–548, 2014.

[10] John A Collins and Bart CJM Fauser. Balancing the strengths of systematic and narrative reviews, 2005.

[11] Massimo Aria and Corrado Cuccurullo. bibliometrix: An r-tool for comprehensive science mapping analysis. *Journal of informetrics*, 11(4):959–975, 2017.

[12] Hebert Azevedo-Sa, Suresh Kumaar Jayaraman, Connor T Esterwood, X Jessie Yang, Lionel P Robert Jr, and Dawn M Tilbury. Real-time estimation of drivers' trust in automated driving systems. *International Journal of Social Robotics*, 13(8):1911–1927, 2021.

[13] Luke Petersen, Lionel Robert, X Jessie Yang, and Dawn M Tilbury. Situational awareness, drivers trust in automated driving systems and secondary task performance. *arXiv preprint arXiv:1903.05251*, 2019.

[14] Suresh Kumaar Jayaraman, Chandler Creech, Lionel P Robert Jr, Dawn M Tilbury, X Jessie Yang, Anuj K Pradhan, and Katherine M Tsui. Trust in av: An uncertainty reduction model of av-pedestrian interactions. In *Companion of the 2018 ACM/IEEE international conference on human-robot interaction*, pages 133–134, 2018.

[15] Jane Jean Kiam, Lukas Fröhlich, and Axel Schulte. Learning decision-making patterns in the context of manned-unmanned teaming. In *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3342–3347. IEEE, 2022.

[16] Jinan M Andrews, Chrstina F Rusnock, Michael E Miller, and Douglas P Meador. Simulation-based evaluation of the effects of varying degrees of control abstraction for manned-unmanned teaming on mental workload of pilots. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2680–2686. IEEE, 2020.

[17] Fabian Schmitt and Axel Schulte. Experimental evaluation of a scalable mixed-initiative planning associate for future military helicopter missions. In *Engineering Psychology and Cognitive Ergonomics: 15th International Conference, EPCE 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings 15*, pages 649–663. Springer, 2018.

[18] Yannick Brand and Axel Schulte. Model-based prediction of workload for adaptive associate systems. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1722–1727. IEEE, 2017.

[19] Ankit Agrawal, Jane Cleland-Huang, and Jan Philipp Steghofer. Model-driven requirements for humans-on-the-loop multi-uav missions. pages 1–10. Institute of Electrical and Electronics Engineers Inc., 8 2020.

[20] Axel Schulte, Diana Donath, and Douglas S Lange. Design patterns for human-cognitive agent teaming. In *Engineering Psychology and Cognitive Ergonomics: 13th International Conference, EPCE 2016, Held as Part of HCI International 2016, Toronto, ON, Canada, July 17-22, 2016, Proceedings 13*, pages 231–243. Springer, 2016.

[21] Stensrud Rune and Sigmund Valaker. Mixed-initiative approaches in the design of a trusted shift of coordination forms in air operations: Supporting collaboration to handle loyal wingmen. Institute of Electrical and Electronics Engineers Inc., 2021.

[22] TF Atdelzater, Ella M Atkins, and Kang G Shin. Qos negotiation in real-time systems and its application to automated flight control. *IEEE Transactions on Computers*, 49(11):1170–1183, 2000.

[23] Samuel J Levulis, Patricia R DeLucia, and So Young Kim. Effects of touch, voice, and multimodal input, and task load on multiple-uav monitoring performance during simulated manned-unmanned teaming in a military helicopter. *Human factors*, 60(8):1117–1129, 2018.

[24] Hasan Isci and Emre Koyuncu. Reinforcement learning based autonomous air combat with energy budgets. In *AIAA SCITECH 2022 Forum*, page 0786, 2022.

[25] Hayley Borck, Justin Karneeb, Ron Alford, and David W Aha. Case-based behavior recognition in beyond visual range air combat. In *FLAIRS Conference*, pages 379–384, 2015.

[26] Wei Ren, Randal W Beard, and Ella M Atkins. Information consensus in multivehicle cooperative control. *IEEE Control systems magazine*, 27(2):71–82, 2007.

[27] Wei Ren and Ella Atkins. Distributed multi-vehicle coordinated control via local information exchange. *International Journal of Robust and Nonlinear Control: IFAC-Affiliated Journal*, 17(10-11):1002–1033, 2007.

[28] Jun Chen, Qilin Zhang, and Bo Hou. An assessment method of pilot workload in manned/unmanned-aerial-vehicles team. pages 1–5. IEEE, 10 2017.

[29] Fabian Schmitt, Gunar Roth, and Axel Schulte. Design and evaluation of a mixed-initiative planner for multi-vehicle missions. In *Engineering Psychology and Cognitive Ergonomics: Cognition and Design: 14th International Conference, EPCE 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part II 14*, pages 375–392. Springer, 2017.

[30] John D Lee and Katrina A See. Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1):50–80, 2004.

9. Appendix

Table 7 – Articles read and their score

Author	Title	Score	Evaluated Human Factors	Experiment	Group
Agrawal, A., Cleland-Huang, J., Steghofer, J.-P.	Model-Driven Requirements for Humans-on-The-Loop Multi-UAV Missions [19]	31	Performance	————	Quality of service and conceptual frameworks
Andrews, J.M., Rusnock, C.F., Miller, M.E., Meador, D.P.	Simulation-Based Evaluation of the Effects of Varying Degrees of Control Abstraction for Manned-Unmanned Teaming on Mental Workload of Pilots [16]	31	Performance and mental workload	Flight simulations	Modes of operation, level of autonomy (LOA);
Cohen, M.C., Demir, M., Chiou, E.K., Cooke, N.J.	The Dynamics of Trust and Verbal Anthropomorphism in Human-Autonomy Teaming [6]	31	Performance and trust	Ground station simulations	Trust dynamics in human-machine interactions
ZHAO, Z., NIU, Y., SHEN, L.	Adaptive level of autonomy for human-UAVs collaborative surveillance using situated fuzzy cognitive maps [4]	31	Performance, mental workload e situational awareness	Ground station simulations	Modes of operation, level of autonomy (LOA);
Caldwell, B.S., Nyre-Yu, M., Hill, J.R.	Advances in human-automation collaboration, coordination and dynamic function allocation [5]	30	————	————	Quality of service and conceptual frameworks
Chen, J., Gao, X., Chen, X., He, Q.	A Shifting Method for Intelligent Operational Mode of UAVs [2]	28	Mental workload	Case study	Modes of operation, level of autonomy (LOA);
Das, A., Kol, P., Lundberg, C., Doelling, K., Sevil, H.E., Lewis, F.	A Rapid Situational Awareness Development Framework for Heterogeneous Manned-Unmanned Teams [1]	28	Situational awareness	Flight with drones	Modos de operação, LOA e consciência situacional
Chen, J., Zhang, Q., Hou, B.	An assessment method of pilot workload in manned/unmanned-aerial-vehicles team [28]	27	Mental workload	Case study	Modes of operation, level of autonomy (LOA);

Continued on next page

Table 7 Continued from previous page

Author	Title	Score	Evaluated Human Factors	Experiment	Groups
Schmitt, F., Schulte, A.	Experimental evaluation of a scalable mixed-initiative planning associate for future military helicopter missions [17]	27	Mental workload	Flight simulations	Advanced technologies and coordination in autonomous systems
Lange, D.S., Verbancsics, P., Gutzwiller, R.S., Reeder, J.	Trust in sparse supervisory control [7]	27	Trust	————	Trust dynamics in human-machine interactions
Schulte, A., Donath, D., Lange, D.S.	Design patterns for human-cognitive agent teaming [20]	27	————	————	Quality of service and conceptual frameworks
Schmitt, F., Roth, G., Schulte, A.	Design and evaluation of a mixed-initiative planner for multi-vehicle missions [29]	27	Performance and mental workload	Ground station simulations	Modes of operation, level of autonomy (LOA);
Kiam, J. J., Frohlich, L.,. Schulte, A.	Learning Decision-Making Patterns in the Context of Manned-Unmanned Teaming. [15]	25	Performance	Ground station simulations	
Brand, Yannick and Schulte, Axel	Model-based prediction of workload for adaptive associate systems [18]	24	Mental workload	————	Modes of operation, level of autonomy (LOA);
Levulis, Samuel J and DeLucia, Patricia R and Kim, So Young	Effects of Touch, Voice, and Multimodal Input, and Task Load on Multiple-UAV Monitoring Performance During Simulated Manned-Unmanned Teaming in a Military Helicopter [23]	24	————	————	
İşci, H., Koyuncu, E.	Reinforcement Learning Based Autonomous Air Combat with Energy Budgets [24]	23	————	Numerical simulations;	Advanced technologies and coordination in autonomous systems
Borck, H., Karneeb, J., Alford, R., Aha, D.W.	Case-based behavior recognition in beyond visual range air combat [25]	23	————	Numerical simulations;	Advanced technologies and coordination in autonomous systems

Continued on next page

Table 7 Continued from previous page

Author	Title	Score	Evaluated Human Factors	Experiment	Groups
Maier, S., Schulte, A.	A cloud-based approach for synchronous multi-pilot multi-UAV mission plan generation in a MUM-T environment [3]	23	————	————	Advanced technologies and coordination in autonomous systems
Rune, S., Valaker, S.	Mixed-initiative approaches in the Design of a trusted shift of Coordination Forms in Air Operations: Supporting Collaboration to handle Loyal Wingmen [21]	19	————	————	Quality of service and conceptual frameworks
Azevedo-Sa, H., Jayaraman, S.K., Esterwood, C.T. et al	Real-Time Estimation of Drivers' Trust in Automated Driving Systems [12]	14	Trust	Driving simulations	Trust dynamics in human-machine interactions
Luke Petersen, Lionel Robert, X. Jessie Yang, Dawn M. Tilbury	Situational Awareness, Drivers Trust in Automated Driving Systems and Secondary Task Performance [13]	14	Situational awareness. Trust.	Driving simulations	Trust dynamics in human-machine interactions
Jayaraman, S. K., Creech, C., Robert Jr, L. P., Tilbury, D. M., Yang, X. J., Pradhan, A. K., & Tsui, K. M.	Trust in AV: An Uncertainty Reduction Model of AV-Pedestrian Interactions [14]	14	————	————	Trust dynamics in human-machine interactions
T.F. Atdelzater; E.M. Atkins; K.G. Shin	QoS negotiation in real-time systems and its application to automated flight control [22]	13	Performance	Flight simulations	Quality of service and conceptual frameworks
Wei Ren; Randal W. Beard; Ella M. Atkins	Information consensus in multivehicle cooperative control [26]	6	————	————	Advanced technologies and coordination in autonomous systems
Wei Ren, Ella Atkins	Distributed multi-vehicle coordinated control via local information exchange [27]	6	Trust	————	Advanced technologies and coordination in autonomous systems

Continued on next page

Table 7 Continued from previous page

Author	Title	Score	Evaluated Human Factors	Experiment	Groups
John D. Lee and Katrina A. See	Trust in Automation: Designing for Appropriate Reliance [30]	4	Trust	———	Trust dynamics in human-machine interactions