A Proposed Risk Categorisation Model for Human-Machine Teaming

Zena Assaad [1]

[1] *The Australian National University, Canberra, Australia*

Abstract

Autonomous systems are becoming more prevalent across a diversity of industries and applications. The development and deployment of these systems is surpassing the promulgation of standards and regulations needed to govern their safety. As the potential applications of autonomous systems continue to broaden, segregating these systems from humans will become increasingly difficult and potentially not feasible in some contexts, such as human-machine teaming (HMT).

A mechanism for categorising risk for HMT operations against levels of autonomy (LOA) and machine functions is proposed. The risk categorisation tool sits within a broader safety framework for HMT. The user centric framework will enable the safe operation of humans alongside machines in a teaming environment in which the machine will not be physically segregated from the human. A key factor to effective safety assurance is proportionality. Autonomous capabilities can vary widely for HMT operations, resulting in varying levels of risk. The proposed risk categorisation tool provides a mechanism for categorising risk for HMT operations.

Keywords

Human-machine teaming, autonomy, safety framework, assurance, risk

1. Introduction

The origin of the word autonomy stems from the Greek words "auto" meaning self and "nomos" meaning governance [2]; reflecting a notion of independence and personal authority [21]. [3] argues that the term autonomy is often conveyed through two interpretations; one denoting self-sufficiency, an ability to take care of oneself, and the other denoting self-directedness, freedom from outside control. The differences in these interpretations have elicited multiple definitions attempting to conceptualise autonomy. These efforts have been married with attempts to define levels of autonomy (LOA) as a mechanism for categorising the varying capabilities of autonomous systems. [18] provides an in depth literature review of the evolution of LOA over the last few decades.

While many LOA taxonomies have been proposed over the years, none of these taxonomies are specific to the application of human-machine teaming (HMT) [18]. [14] presents a framework for adaptive automation processes for human-robot teaming. While the framework presents varying LOA as a method for enhancing human-system performance, a taxonomy for categorising LOA for HMT is not presented.

There currently does not yet exist a globally agreed upon definition of HMT; however, the broader literature defines HMT, often coined the term human-autonomy teaming, around the notion of sharing authority to pursue common goals [12]. In the context of this research, HMT is defined as a combination of human and machine capabilities working together towards an aligned goal [20].

HMT operations have been actualised across a breadth of domains and applications, demonstrating a range of machine capabilities - what the machine is capable of doing - and machine functions - the role or purpose of the machine. LOA are an indication of machine capabilities as they describe the degree to which a system is automated and what level of human intervention is required [17]. How

EICS '22: Engineering Interactive Computing Systems conference, June 21–24, 2022, Sophia Antipolis, France
EMAIL: zena.assaad@anu.edu.au (A. 1)
ORCID: 0000-0002-1529-1088 (A. 1)

risk is measured and how safety is assured for HMT operations will differ depending on the varying capabilities that span across the spectrum that we call autonomy.

Currently, robust mechanisms for assuring the safety of autonomous systems are lacking across most industries. There exists a patchwork of safety standards around robot systems, most prominently in the industrial sector. ISO 15066 [11] specifies safety requirements for collaborative industrial robot systems, as described in ISO 10218-1 [9] and ISO 10218-2 [10], that share the same workspace as humans. ISO 15066 applies a heavy focus on controlling process parameters, such as speed and force, to mitigate potential collisions. Enabling collision mitigation through controlling parameters arises as a common mechanism within the literature around safety assurance of humans operating alongside collaborative robots [7][13].

While standards such as ISO15066 "*Robots and robotic devices — Collaborative robots*" do exist [11], these standards require systems to be physically separated from humans while operating. Given the diversity of potential applications of HMT, segregating machines from humans may not always be feasible. While established and standardised safety frameworks exist across many industries, managing risks associated with autonomous systems introduces unique challenges. The breadth of possible applications of autonomous technologies also introduces challenges for risk management, as the diversity of use cases that come with different LOA and their inherent risks, can be difficult to capture. Understanding levels of risk for different LOA will aid in determining proportionate safety measures required for HMT operations

2. Risk assessment and management

Risk assessment and management is a core pillar of safety assurance for systems, autonomous or otherwise. Established as a scientific field in the 1970s, the practice of risk assessment and management has matured significantly over the proceeding years and is now used across most industries [1].

[8] explore the links between facts and values in risk decision making; demonstrating that risk is often connected with other issues that impact decision making; "*decision making on traffic safety has to be integrated with decision making on traffic planning as a whole, including issues such as travel time, accessibility, environmental impact, costs, etc.*" [8]. When considering risk assessment and management for HMT, the purpose of a system, what it is actually capable of in terms of autonomy, what capacity there is for human intervention and what the human role is within the broader team are fundamental points that need to be considered if risk is going to be assessed and managed *proportionally*.

Subjective probability is a common approach to managing uncertainty in risk assessments [5]. Note, the reference to uncertainty here is at the operational level rather than at a systems level. Uncertainty at the operational level can result from many factors, a common one being incomplete information [5]. How we understand and conceptualise autonomy within the context of HMT will influence how we analyse risk. Categorising risk levels for HMT operations against LOA and machine functions will facilitate a proportionate approach to risk assessment and management. The risk categorisation matrix presented within this paper sits within a broader HMT safety framework, of which is detailed in section 4, and provides a tool for identifying appropriate levels of risk for HMT operations.

3. Risk categorization model

A method for categorising HMT applications against machine capabilities, expressed through LOA, and machine functions is proposed. Literature around LOA have propositioned taxonomies specifying the degree to which a task is automated. While several taxonomies have been proposed [4][15][16][19], this research builds off the work of [17] which proposes ten LOA. While the proposed levels were designed to be applicable to a *"wide variety of domains and task types"* [6], not all the levels would be applicable to HMT. For the purpose of this research, the following four LOA were identified as being applicable to the context of HMT.

Table 1.
Levels of autonomy - Levels and definitions taken directly from [6]

Levels of Autonomy

Shared Control (SHC)	Both the human and the computer generate possible decision options. The human still retains full control over the selection of which option to implement; however, carrying out the actions is shared between the human and the system.
Blended Decision Making (BDM)	At this level, the computer generates a list of decision options that it selects from and carries out if the human consents. The human may approve of the computer's selected option or select one from among those generated by the computer or the operator. The computer will then carry out the selected action. This level represents a higher level decision support system that is capable of selecting among alternatives as well as implementing the second option.
Automated Decision Making (ADM)	At this level, the system selects the best option to implement and carry out that action, based upon a list of alternatives it generates (augmented by alternatives suggested by the human operator). This system, therefore, automates decision making in addition to the generation of options (as with decision support systems).
Full Automation (FA)	At this level, the system carries out all actions. The human is completely out of the control loop and cannot intervene. This level is representative of a fully automated system where human processing is not deemed to be necessary.

The four LOA detailed in Table 1 were chosen as they reflect a more balanced relationship between human and machine. Each of the levels demonstrates less of a hierarchical structure and more of a collaborative relationship with opportunities for negotiation between the entities. HMT is characterised by a more balanced relationship between human and machine with greater levels of negotiation [12]. This type of relationship requires increased machine capability, which is why the lower LOA identified in [6] were deemed not applicable to the given context.

As machine capabilities cannot be isolated from machine functions, four machine functions were have also been identified, Building on from [6][15] further, the proposed LOA are considered applicable to four machine functions that attempt to identify the role of the machine in a given context. The four machine functions and what they encompass within the context of this framework are detailed.

Table 2.
Machine functions - Machine functions and definitions adapted from [6]

Machine Functions	
Monitoring	Involves sending and registration of input data.
Generating	Involves cognitive functions, such as processing information or input data.
Selecting	Involves decision and action selection.
Implementing	Involves action implementation.

The four machine functions identified represent the possible functions or purpose of a machine within HMT. The functions range from monitoring, which involves lower levels of decision making on part of the machine, through to implementing, which implies implementing decision making with or without human intervention.

To situate HMT operations in the context of machine capability, expressed through LOA, and machine functions, a categorisation matrix has been developed, and is depicted in Figure 1 below. The matrix is a tool for categorising HMT operations against three risk categories to support proportionate risk assessment and management of HMT operations.

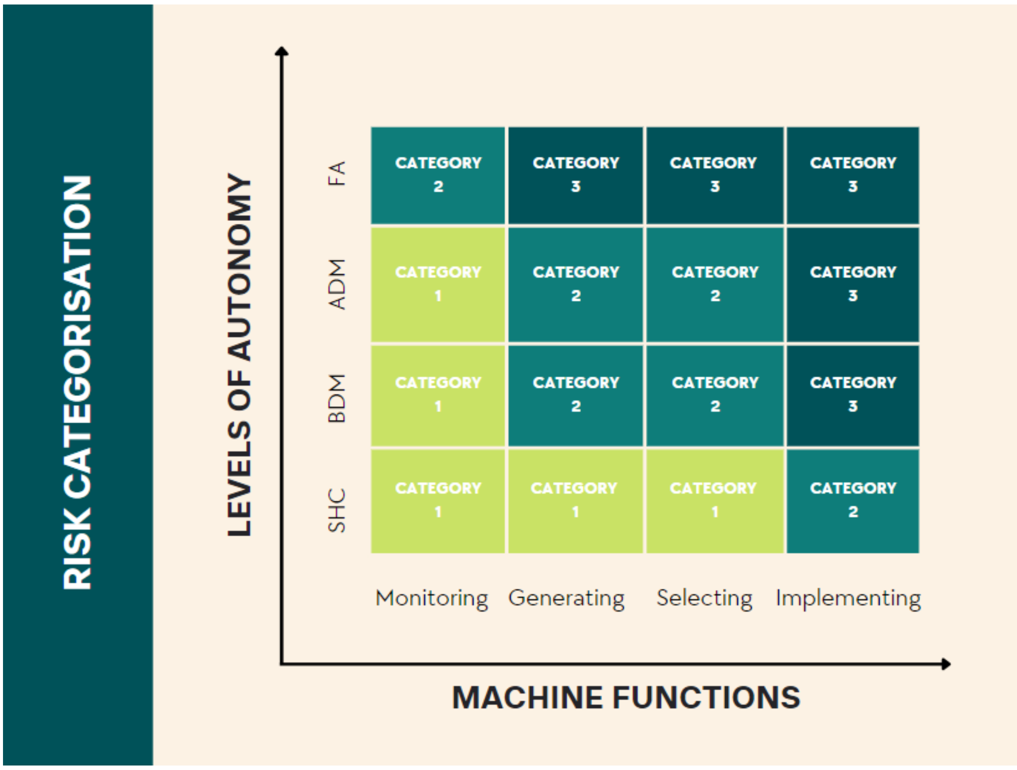

Figure 1. Risk categorisation matrix

The matrix presented in Figure 1 illustrates three risk categories for HMT operations. Risk category 1 encompasses capabilities that demonstrate lower levels of autonomy and greater levels of human supervision. Risk category 2 encompasses capabilities that demonstrate greater levels of autonomy that require less human supervision and risk category 3 encompasses capabilities that demonstrate high levels of autonomy that involve minimal human supervision. Situating HMT operations within these risk categories will ensure proportionate and effective safety assurance can be demonstrated through the broader HMT safety framework.

4. HMT safety framework

The presented risk categorisation matrix sits within a broader HMT safety framework as a mechanism for identifying appropriate levels of risk. The proposed broader framework will demonstrate the safety assurance of both entities -human and machine - within HMT. In a teaming context, the human role is less authoritative and more collaborative, as is demonstrated through increased opportunities for negotiation between the two entities [12].

Capturing all the broader risks that come with HMT can be challenging. As such, guiding principles have been developed to help guide users with identifying the risks of HMT. The guiding principles include:

- Adaptability - understanding the capacity to which the human and the machine can adapt to their environment.
- Goal setting and goal actualisation - as HMT is defined by the pursuit of a shared goal, it is necessary to understand how goals are determined and actualised for both humans and machines.
- Communication - understanding how, what, why and when information is communicated between human and machine.
- Ethics - understanding the ethical implications of humans operating in close proximity to a machine within specific environments.
- Trust - understanding how trust between the two entities influences decision making.

The HMT safety framework will provide assurance of both entities, and in addition to addressing physical safety, the framework will also include psychosocial considerations such as trust. The framework will address how a system or capability operates in a specific environment and, more importantly, how humans operate alongside these capabilities. The HMT safety framework will be targeted at the implementation stage, with specific focus on user experience. It will act as a guiding set of processes for users to follow to ensure the safe operation of humans alongside machines in a teaming environment.

5. Conclusion and next steps

Machine capabilities exist across a spectrum of autonomy. LOA applicable to HMT were identified alongside machine capabilities. These factors are used to categorise HMT operations against three levels of risk. Different machine capabilities and functions will yield different risks. The risks that come with lower capabilities and functions, and thereby lower levels of uncertainty, will differ to the risks that emerge from higher machine capabilities and functions that entail greater levels of uncertainty. It follows that different risk analyses need to be applied to ensure proportionate measures of safety are being implemented.

The next stages of this research will include further development of the three risk analysis categories. Each category will be developed against case study analyses across multiple industries to ensure the outputs are applicable across a diversity of industries. The final output will be a cross-sector safety management framework for HMT.

6. Acknowledgements

The research for this paper received funding from the Australian Government through Trusted Autonomous Systems, a Defence Cooperative Research Centre funded through the Next Generation Technologies Fund.

7. References

1. Aven, T. (2016). Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research*, *253*(1), 1–13. https://doi.org/10.1016/j.ejor.2015.12.023
2. Bradshaw, J. M., Feltovich, P. J., Jung, H., Kulkarni, S., Taysom, W., & Uszok, A. (2004). Dimensions of Adjustable Autonomy and Mixed-Initiative Interaction. In *Agents and Computational Autonomy: Potential, Risks, and Solutions* (Vol. 2969, pp. 17–39). Springer-Verlag.
3. Bradshaw, J. M., Hoffman, R. R., Johnson, M., & Woods, D. D. (2013). The Seven Deadly Myths of "Autonomous Systems." *IEEE Computer Society*, 1541–1672.
4. Clothier, R., Williams, Brendan. P., & Perez, T. (2019, February). *Autonomy from a Safety Certification Perspective*. 18th Australian International Aerospace Congress, Melbourne, Australia. https://www.researchgate.net/profile/Reece-Clothier/publication/331587067_Autonomy_from_a_Safety_Certification_Perspective/links/5c81c6ce458515831f8f3571/Autonomy-from-a-Safety-Certification-Perspective.pdf
5. Dubois, D. (2010). Representation, Propagation, and Decision Issues in RiskAnalysis Under Incomplete Probabilistic Information. *Risk Analysis*, *30*(3), 361–368. https://doi.org/10.1111/j.1539-6924.2010.01359.x
6. Endsley, M., R., & Kaber, D., B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, *42*(3), 462–492. https://doi.org/10.1080/001401399185595
7. Falconi, R., Sabattini, L., Secchi, C., Fantuzzi, C., & Melchiorri, C. (2014). Edge-weighted consensus-based formation control strategy with collision avoidance. *Robotica*, *33*(2), 332–347. https://doi.org/10.1017/S0263574714000368
8. Hansson, S. O., & Aven, T. (2014). Is risk analysis scientific? *Risk Analysis*, *34*(7), 1173–1183.
9. ISO, I. (2011). *Robots and robotic devices—Safety requirements for industrial robots—Part 1: Robots*. International Organisation for Standards. https://www.iso.org/standard/51330.html
10. ISO, (ISO). (2011). *Robots and robotic devices—Safety requirements for industrial robots—Part 2: Robot systems and integration*. International Organisation for Standards. https://www.iso.org/standard/41571.html
11. ISO, (ISO). (2016). *Robots and robotic devices—Collaborative robots*. International Organisation for Standards. https://www.iso.org/standard/62996.html

12. Lyons, J. B., Sycara, K., Lewis, M., & Capiola, A. (2021). Human–Autonomy Teaming: Definitions, Debates, and Directions. *Frontiers in Psychology*, *12*, 19–32. https://doi.org/10.3389/fpsyg.2021.589585
13. Matthews, M., Chowdhary, G., & Kieson, E. (2017). *Intent Communication between Autonomous Vehicles and Pedestrians*. https://arxiv.org/abs/1708.07123v1
14. Parasuraman, R., Barnes, M., & Cosenzo, K. (2007). Adaptive Automation for Human-Robot Teaming in Future Command and Control Systems. *The International C2 Journal*, *1*(2), 43–68.
15. Parasuraman, R., Sheridan, T., B., & Wickens, C., D. (2000). A Model for Types and Levels of Human Interaction with Automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *30*(3), 286–297.
16. Parker, J. (2021). The Challenges Posed by the Advent of Maritime Autonomous Surface Ships for International Maritime Law. *Australian and New Zealand Maritime Law Journal*, *35*(1), 31–42.
17. Sheridan, T., B., & Verplank, W., L. (1978). *Human and Computer Control of Undersea Teleoperators* (Mechanical Engineering, Massachusetts Institute of Technology) [Technical Report]. https://apps.dtic.mil/sti/pdfs/ADA057655.pdf
18. Vagia, M., Transeth, A. A., & Fjerdingen, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied Ergonomics*, *53*, 190–202. http://dx.doi.org/10.1016/j.apergo.2015.09.013
19. Vine, R., & Kohn, E. (2020). *Concept for Robotic and Autonomous Systems V1.0*. Joint Warfare Council. https://tasdcrc.com.au/wp-content/uploads/2020/12/ADF-Concept-Robotics.pdf
20. Walliser, James. C., de Visser, Ewart. J., & Shaw, Tyler. H. (2019). Team Structure and Team Building Improve HumanMachine Teaming With Autonomous Agents. *Journal of Cognitive Engineering and Decision Making*, *13*(4), 258–278. https://doi.org/10.1177/1555343419867563
21. Weinstein, N., Przybylski, Andrew. K., & Ryan, Richard. M. (2012). The index of autonomous functioning: Development of a scale of human autonomy. *Journal of Research in Personality*, *46*, 397–413. http://dx.doi.org/10.1016/j.jrp.2012.03.007