# The Limits of Meaningful Human Control of AI in the Maritime Domain

Lukas Albrecht, Hagen Braun, Tim Robin Kosack, Thomas Krüger

***Abstract*—This paper analyses the viability of meaningful human control as a mechanism to ensure that ethical and safety aims are met, focusing on the maritime context. It concludes that there are technical and conceptual barriers that make meaningful human control non-viable in some applications.**

***Keywords—ethics, artificial intelligence, autonomous vehicles, underwater vehicles, meaningful human control***

## I. INTRODUCTION

Future autonomous systems, in maritime applications and otherwise, will contain Artificial Intelligence (*AI*) components as a main driver for autonomy. This development not only promises increased convenience and efficiency, but also substantial safety advantages. In part, this is because AI may replace the human element in high-risk settings and allows for the delegation of tasks – thereby eliminating human error, deliberate malpractice, and enabling faster response times in case of accidents. However, truly autonomous AI also introduces a variety of characteristic risks. To address these concerns, the concept of Meaningful Human Control (*MHC*) has been introduced. However, reintroducing the human element to a highly autonomous AI system limits its potential. If due to special ethical concerns or safety engineering reasons the human operator needs to be involved in AI decision making, human oversight and human control in a meaningful way are indispensable. But if a truly autonomous AI system is evidentially compliant to safety requirements, must the human operator necessarily be reintroduced to guarantee ethical AI use? What are the applications where MHC needs to be considered even when full autonomy is technologically achievable?

## II. THE BENEFITS OF AUTOMATION

Development in the field of Artificial Intelligence (AI) has brought significant benefits in terms of speed, accuracy, and reliability. This allows for a more efficient accomplishment of tasks and the delegation of more and more tasks from human operators to AI agents, promising to liberate humans from menial work and increase societal well-being [1,2]. Moreover, AI also holds the potential to significantly increase safety for humans in a number of high-risk contexts. From a safety engineering perspective, AI provides extensive possibilities to mitigate or even eliminate risks. Algorithms that identify or assess (undiscovered) risks are already in place in several domains, e.g. healthcare [3], finance [4], and predictive maintenance [5]. Furthermore, in unexpected accident scenarios, AI-based systems may be able to react faster and more accurately, thereby lowering situational risk.

Even more benefits are expected when human operators can be replaced entirely by advanced AI agents, and therefore, are removed from any contextual risk altogether. This kind of AI deployment is particularly useful for safety engineering because it removes the human element both in terms of being the source of, but also subject to harm. In extreme operational environments, such as the high seas and under water, the removal of human operators is especially desirable. Sophisticated machine autonomy promises the complete delegation of tasks to AI agents that can develop the skills required to handle these tasks in unknown environments without significant human input [6]. At present, AI systems that strive for such levels of autonomy within their domains include self-driving cars [7], manned and unmanned aircraft [8], industrial robots [9], and Lethal Autonomous Weapons Systems (*LAWS*) [10].

Further developments in the field of AI and the design of more advanced AI systems not only hold the prospect of the delegation of more tasks in general, but also of more complex tasks in particular. And while current products and demonstrators still need to rely on the human operator as a safeguard, the operational performance of AI is likely to surpass human performance at some point in the future. With increased technological maturity, especially regarding robustness, explainability, as well as safety engineering, the instrumental value of AI as a tool, able to replace human operators and to provide enormous potential in terms of safety and efficiency, can be found in virtually all settings where the use of autonomous systems is deemed viable.

In the transportation sector, the case of self-driving cars serves as a good example of one of the main domains in which the use of AI is headed towards replacing the human element. The final level of driving automation according to the Society of Automotive Engineers – level five "full driving automation" – calls for sustained and unconditional performance of the driving task by the automated driving system [7]. Next to a range of related socio-economic benefits and increased convenience for users, completely autonomous traffic also provides the potential for a substantial increase in safety [11]. Future systems are expected to reduce the number of accidents by eliminating human error and deliberate malpractice. Additionally, in cases where accidents do still occur, AI will be far more capable to take swift action – e.g. to instantaneously steer out of harm's way in unexpected situations where there is little time to react. The delegation to AI presents a particularly promising prospect in this domain, as traffic poses significant risk to road users [12, 13]. Considering the transport sector as a whole, an AI takeover of personal and public transport could reduce the amount of risk passengers are exposed to. In the case of transportation of cargo, it might even remove human involvement, and thus human exposure to risk, completely.

While the use of autonomous AI provides significant potential to eliminate risk caused by humans, and mitigate safety risks that arise in unexpected situations, AI agents, in turn, introduce their own risks. A considerable body of literature addresses the issues that revolve around the technical means employed to harness AI's benefits and the deployment of autonomous agents – including the occurrence of Black Boxes [14], the possibility of Responsibility Gaps [15], and a host of issues concerning algorithmic decision making [16]. More importantly, the prospect of replacing humans with autonomous AI agents comes with a catch. Making use of AI agents that find their own creative solutions to unexpected problems and that are able to operate independently in unknown environments entails the inability to fully predict their behavior and accepting the possibility of undesired outputs. Simply put, valuing AI's autonomy necessitates facing the risk of that autonomy [17].

These risks concerning the use of AI are neither novel in the discussion on the governance of AI nor are they unique to the use cases of risk management mentioned above. However, they are of particular relevance in this regard, as the introduction of new risks goes against the endeavor of minimizing them. More importantly, the types of risk that the use of AI introduces are unlike those that are managed with the help of AI. Whereas safety related risk is expected to be managed more effectively with increased robustness and precision of AI systems, this does not apply to non-technical risk caused by their employment. For example, in the frequently discussed use case of LAWS, various stakeholders have endorsed the paramount importance of responsibility and dignity [18-22]. In this case, the deployment of truly autonomous AI agents that are susceptible to gaps in responsibility is not feasible, regardless of how safe and effective they may be from a technical point of view.

The issues at hand, especially the risk of autonomous AI agents, are not properly addressed by simply improving AI systems deployed in the context of risk management, if these improvements do not also manage the non-technical risk that they themselves cause. Hence, in contexts where aspects pertaining to these risks of AI play a key role, the use of autonomous AI might not be viable, and thus, prevent sophisticated AI systems from successful implementation. Unsurprisingly, addressing these challenges of AI has become a pressing subject in the debate on AI governance, especially with regard to the above-mentioned aspects such as fairness, explainability, and responsibility [23, 24]

## III. THE CALL FOR MEANINGFUL HUMAN CONTROL

One of the most prominent approaches to manage some of the characteristic risks of AI is the concept of Meaningful Human Control (*MHC*). Originating in the discussion on LAWS and the concern of delegating the use of force and decisions over life and death to machines [25-29], MHC has since become a popular instrument to manage various risks that the implementation of AI systems has introduced. The concept has spread to other domains like automated or autonomous driving systems [30, 31] and automated decision-making systems [32, 33]. Yet, despite its apparently ubiquitous endorsement, there is no agreement on what exactly constitutes MHC [34]. Furthermore, discussions on the subject frequently fail to clearly specify the purpose of implementing MHC – i.e. whether it shall increase safety, ensure responsibility or serve a completely different purpose [35].

Nonetheless, it should be stressed that MHC has one very specific advantage: it is the only answer to the problem of responsibility gaps that has been developed so far. MHC closes responsibility gaps by keeping a human operator close enough to the individual decisions made by an AI system that they can genuinely be held responsible for the harms caused by such a system.

MHC is not the only conceptual tool that has been explored in the context of AI ethics. For example, the most influential AI ethics and governance document of the recent past, the *Ethics Guidelines for Trustworthy A*I by the European Commission [36] does not use the terminology of "meaningful human control" at all, and instead opts to engage with the matter in terms of "human agency and oversight" (pp. 15-16). In this context, three methods of oversight are specifically described: Human-in-the-loop (*HITL*), human-on-the-loop (*HOTL*) and human-in-command (*HIC*). These approaches to human oversight can be distinguished by the degree of direct control that a human exerts over individual decisions: whereas with HITL, a human has the capacity to intervene in every individual decision cycle of the AI, with HOTL, human input is limited to intervention during the design cycle and monitoring roles. HIC can be viewed as the minimum that is potentially compatible with the ethical AI use, as the degree of human input in this governance method is limited to decisions of when and how to use AI systems.

What differentiates MHC from mere human oversight is the degree of immediacy with which a human is involved in individual decisions made by the AI system. As perhaps the most important upside of MHC is supposed to be the avoidance of responsibility gaps, it is necessary that a human operator is involved in all ethically weighty decisions made by the AI system. Therefore, with these methods of oversight, as direct human control decreases, so does meaningful human control: HITL exhibits the highest compatibility with the concept, since, by definition, a human is able to intervene in every decision cycle when a HITL approach is employed. The HOTL approach is still compatible with MHC, as human monitoring of AI performance may be sufficient to ensure ethical adequacy and responsible use in less ethically sensitive contexts. The HIC method should be viewed as incompatible with MHC, since following this approach individual decisions made by AI agents after deployment do not fall under the control of humans at all.

From a safety engineering perspective, these methods of oversight have the advantage of interfacing well with the different AI decision making- and risk-levels as, for example, currently discussed in aviation [8] – from assistance to human (Level 1a & b) over human AI teaming (Level 2a Cooperation & 2b Collaboration) to autonomous AI (Level 3a & b). HITL should be expected to be deployed in levels 1 and 2, HOTL in levels 2 and 3, and HIC in level 3.

## IV. BARRIERS TO MEANINGFUL HUMAN CONTROL

The main advantage of MCH is its promise to ensure that *responsibility* is preserved in high automation contexts. However, there are number of conceptual and material barriers to its implementation. For human control to count as meaningful, two requirements must be met: First, the human operator needs to be *significantly involved* in the decision-making process and second, the human operator needs the *necessary expertise* evaluate the decisions of the AI system in an informed way. These two requirements stand in conceptual tension with the benefits of automation outlined above, as they necessitate a re-introduction of the human element into an otherwise highly autonomous system. In any scenario in which the removal of the human operators is an advantage in itself (for reasons of safety, efficiency or otherwise), MHC loses applicability.

Direct involvement in every decision cycle, such as when HITL is applied as the method of human oversight, does not guarantee by itself meaningful human control. The problems of rubber stamping and automation bias illustrate why that might be the case. Rubber stamping refers to a human operator accepting an AI decision without the ability to properly assess it. In the case of rubber stamping, a human operator is nominally in control of the AI system, authorizing or validating AI decisions before they are executed. However, due to a lack of expertise on part of the operator or other contravening factors such as time pressure, the presence of the human operator does not actually result in improvements in the problem areas where MHC is supposed to be a solution. For example, while a human operator can in principle act as a nexus of responsibility when a fully autonomous AI system cannot, this is not the case for obviously unqualified operators or those that feel pressured to quickly authorize AI decision in order to not undermine the performance of the system. The process can be appealing to a person pressing a button every time they see a light turn on. If the light signifies an ethically weighty decision made by an AI, the human operator authorizing it by pressing the button obviously does nothing to improve the adequacy of an AI-enhanced system in terms of the relevant ethical dimensions and does not make the operator meaningfully responsible for any resultant harm.

However, non-meaningful control can also occur in contexts where the human operator is, in theory, able to properly validate the decision of an AI agent. In these cases, the problem of automation bias can potentially undermine meaningful human control. Automation bias occurs when the judgement of an expert operator is undermined by giving undue weight to the output of an automated (in this case, AI) system. It is a documented phenomenon that humans tend to put increased weight on data provided by automated systems even when trusting them goes against their own, best judgement [37, 38]. The ALTAI addresses this problem with the requirement that AI systems should not undermine human agency. Rather, it must be ensured that AI systems enhance human decision-making abilities. Automation bias as a general phenomenon is a barrier to this aim, since it occurs on a subconscious level. However, the existence of the phenomenon shows that implementing MHC is not trivial even in cases where even expert human operators, who are not subject to contravening situational pressures are confronted with highly pre-refined judgements of AI agents.

Ensuring significant involvement in the decision-making process is not trivial in many contexts in which high automation is desirable, either. In the maritime domain, communication with autonomous ships is significantly limited by the low bandwidth available at sea. Due to the high rate of absorption of electromagnetic waves in water radio-based communications are generally not feasible in the underwater environment. Alternative methods of communications come with downsides that make the implementation of MHC difficult, such as the limited range of wired communications or the significant latency inherent to acoustic communications. This means that a significant barrier to fast, reliable, long-range communication exists in one of the prime use-cases of AI in the maritime field, (partially-)autonomous underwater vehicles (*AUVs*) [39]. The more constrained communication between the operator and AI system becomes, the larger the material barrier to the implementation of an MHC paradigm grows: A human operator cannot be expected to be able to intervene in or supervise every decision cycle of the AI system when there is not reliable means of communication between the two parties.

All these issues become even more problematic when a mismatch between the method of oversight and the desired degree of autonomy of the AI system occurs, such as when a human-in-the-loop approach is applied to an otherwise fully autonomous system. At first glance, a more extensive degree of human involvement appears to be desirable in this context, since the loss of human involvement in the decision-making process is the source of the types of issues that MHC is supposed to address (responsibility gaps, ethical inadequacy, etc.). However, we can observe that this kind of approach will either limit the performance of the system as a whole to ensure meaningful control, or undermine the meaningfulness of the control exerted by the human operator in order to maintain performance.

The idea of MHC comes full circle with the reintroduction of expert operators as controllers of AI systems. Whereas humans were originally excluded from specific tasks to reap the benefits of highly autonomous AI systems, the reintroduction of the human operator leads the whole process back to its beginning. The starting point is marked by a task that is being executed without assistance of AI. In order to achieve the benefits of increased automation – better performance in terms of speed, accuracy or safety and the automation of more complex tasks – the involvement of AI reaches a level where the human being is increasingly excluded from the task. Various problems follow from this exclusion of the human operator, both ethical and practical. One of the answers prospective to this development is MHC, where the reintroduction of the human operator is the essential aspect. But with the key aspect being human control, MHC is conceptually incompatible with the highest levels of AI autonomy. The reasons to deploy highly autonomous AI systems are simply in too much tension with the idea of a human operator being meaningfully involved in all relevant decisions made by that AI.

This is not the case at lower levels of autonomy. Human operators are already a necessary element in decision support systems and human-AI teaming set ups. These two approaches differ from the case described above, in which a human operator acts as a vetoing agent to an otherwise autonomous decision-making AI. Instead, the human operator is assumed to be in charge of decision-making ab initio, with the purpose of the AI agent being to either enhance their capabilities (support systems) or to take over ethically non-critical parts of the overall operation (human-AI teaming).

## V. THE LIMITS OF MEANINGFUL HUMAN CONTROL

If we accept that meaningful human control is not a viable solution to problems relating to responsibility and ethical concerns that arise in the context of highly autonomous AI systems, that raises the question: What is the use of MHC? The answer might be that MHC gives us insight regarding the appropriate method of human oversight in use contexts with different levels of ethical relevance. If, for example, guaranteeing ethical standards and avoiding responsibility gaps is non-negotiable in the context of autonomous weapons systems, and MHC is the only way to meet this goal, then we can conclude that the use of fully autonomous AI weapons is unwarranted, and that at most we should be considering decision support systems or human-AI teaming approaches in this context.

Of course, not all domains are characterized by such high degrees of ethical sensitivity. It is not a conceptual necessity that MHC is a requirement for all use cases of AI. For example, if a sufficient level of safety can be clearly demonstrated for highly autonomous vehicles in contexts such as self-driving cars (level 5 automation) or aviation (level 3 EASA guidelines), it becomes difficult to articulate what further ethical barrier to the use of such systems would still remain. The Ethics Guidelines for Trustworthy AI [36] note: "Oversight mechanisms can be required in varying degrees to support other safety and control measures, depending on the AI system's application area and potential risk. All other things being equal, the less oversight a human can exercise over an AI system, the more extensive testing and stricter governance is required" (p. 16). If the area of application is not subject to special ethical concerns and risk is demonstrably low, governing mechanisms other than MHC can take over. This illustrates an upside to the Trustworthy AI approach of the ALTAI guidelines, since it is compatible with MHC in contexts where direct control is necessary, but provides mechanisms to ensure ethical AI development and use in contexts incompatible with MHC, as well.

## VI. CONCLUSION

Meaningful human control is a valuable concept in AI ethics that helps us understand which type of human oversight mechanism is necessary for different use cases for AI systems. It is the only oversight mechanism developed so far that can truly be said to deal with the problem of responsibility gaps. However, the approach is conceptually incompatible with highly autonomous AI systems, as it is impossible to both realize the strict control required for a human operator to truly be responsible for the decisions of the AI system and maintain the benefits that highly autonomous AI provides in terms of speed, accuracy and safety. Domains in which communication between an AI system and human operator are harder to realize, such as the underwater domain in the maritime context, suffer from additional material barriers that could further complicate the implementation of MHC. The existence of certain domains that naturally resist an MHC approach should be recognized as a limiting factor for the feasibility of MHC as a general approach to AI ethics.

In some domains, where the problem of responsibility gaps is of elevated importance, this implies that the use of highly autonomous AI systems is unjustifiable on ethical grounds. In these contexts, alternative approaches such as AI decision support and human-AI teaming should be pursued instead. However, in less ethically demanding contexts, one may find other human oversight paradigms that do not preclude the use of highly autonomous systems, if a sufficient level of safety can be demonstrated.


[1] Danaher, J. (2019). Automation and utopia: Human flourishing in a world without work. Harvard University Press. http://www.jstor.org/stable/j.ctvn5txpc https://doi.org/10.2307/j.ctvn5txpc

[2] European Commission. (2022). *Annual statistical report.* European Commission. https://road-safety.transport.ec.europa.eu/european-road-safety-observatory/data-and-analysis/annual-statistical-report_en

[3] EPRS. (2022). *Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts.* European Parliament. https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2022)729512

[4] Milana, C., & Ashta, A. (2021). Artificial intelligence techniques in finance and financial markets: A survey of the literature. *Strategic Change*, *30*(3), 189–209. https://doi.org/10.1002/jsc.2403

[5] Keleko, A. T., Kamsu-Foguem, B., Ngouna, R. H., & Tongne, A. (2022). Artificial intelligence and real-time predictive maintenance in industry 4.0: a bibliometric analysis. *AI and Ethics*, *2*(4), 553–577. https://doi.org/10.1007/s43681-021-00132-6

[6] Ezenkwu, C. P., & Starkey, A. (2019). Machine Autonomy: Definition, Approaches, Challenges and Research Gaps. In K. Arai, R. Bhatia, & S. Kapoor (Eds.), *Springer eBooks Intelligent Technologies and Robotics: Vol. 997. Intelligent Computing: Proceedings of the 2019 Computing Conference, Volume 1* (1st ed. 2019, Vol. 997, pp. 335–358). Springer. https://doi.org/10.1007/978-3-030-22871-2_24

[7] SAE. (2021). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles: J3016_202104. SAE International. https://www.sae.org/standards/content/j3016_202104/

[8] EASA. (2023). EASA Concept Paper: First usable guidance for Level 1 & 2 machine learning applications: A deliverable of the EASA AI Roadmap. European Aviation Saftey Agency. https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-concept-paper-proposed-issue-2

[9] Tantawi, K. H., Sokolov, A., & Tantawi, O. (2019). Advances in Industrial Robotics: From Industry 3.0 Automation to Industry 4.0 Collaboration. *2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES-ICON),* 1–4. https://doi.org/10.1109/TIMES-iCON47539.2019.9024658

[10] Scharre, P., & Horowitz, M. C. (2015). *An Introduction to Autonomy in Weapon Systems.* Center for a New American Security. https://s3.us-east-1.amazonaws.com/files.cnas.org/hero/documents/Ethical-Autonomy-Working-Paper_021015_v02.pdf

[11] Ferencz, C., & Máté Zöldy (2024). Exhaustive Investigation of the Promises and Perils of Autonomous Mobility Technology. *Periodica Polytechnica Transportation Engineering*, *52*(1), 86–95. https://doi.org/10.3311/PPtr.22573

[12] European Commission. (2022). *Annual statistical report.* European Commission. https://road-safety.transport.ec.europa.eu/european-road-safety-observatory/data-and-analysis/annual-statistical-report_en

[13] Bureau of Transportation Statistics. (2023). *Motor Vehicle Safety Data.* Bureau of Transportation Statistics. https://www.bts.gov/content/motor-vehicle-safety-data

[14] Eschenbach, W. J. von (2021). Transparency and the Black Box Problem: Why We Do Not Trust AI. Philosophy & Technology, 34(4), 1607–1622. https://doi.org/10.1007/s13347-021-00477-0

[15] Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, *6*(3), 175–183. https://doi.org/10.1007/s10676-004-3422-1

[16] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 205395171667967. https://doi.org/10.1177/2053951716679679

[17] Sparrow, R. (2007). Killer Robots. *Journal of Applied Philosophy*, *24*(1), 62–77. https://doi.org/10.1111/j.1468-5930.2007.00346.x

[18] Article 36. (2013). *Killer Robots: UK Government Policy on Fully Autonomous Weapons.* Article 36. https://article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf

[19] Human Rights Watch. (2012). *Losing Humanity: The Case against Killer Robots.* Human Rights Watch. https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots

[20] International Committee for Robot Arms Control. (2009). *Mission Statement.* International Committee for Robot Arms Control. http://www.icrac.net/statements/

[21] International Committee of the Red Cross. (2014). *Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects.* International Committee of the Red Cross. https://www.icrc.org/en/document/report-icrc-meeting-autonomous-weapon-systems-26-28-march-2014

[22] UNIDIR. (2014). The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward. United Nations Institute for Disarmament Research. https://unidir.org/files/publication/pdfs/considering-how-meaningful-human-control-might-move-the-discussion-forward-en-615.pdf

[23] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

[24] Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2023). Trustworthy Artificial Intelligence: A Review. *ACM Computing Surveys*, *55*(2), 1–38. https://doi.org/10.1145/3491209

[25] Amoroso, D., & Tamburrini, G. (2020). Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues. *Current Robotics Reports*, *1*(4), 187–194. https://doi.org/10.1007/s43154-020-00024-3

[26] Ekelhof, M. (2019). Moving Beyond Semantics on Autonomous Weapons: Meaningful Human Control in Operation. *Global Policy*, *10*(3), 343–348. https://doi.org/10.1111/1758-5899.12665

[27] Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI*, *5*, Article 15, 15. https://doi.org/10.3389/frobt.2018.00015

[28] Scharre, P., & Horowitz, M. C. (2015). *An Introduction to Autonomy in Weapon Systems.* Center for a New American Security. https://s3.us-east-1.amazonaws.com/files.cnas.org/hero/documents/Ethical-Autonomy-Working-Paper_021015_v02.pdf

[29] Schwarz, E. (2021). Autonomous Weapons Systems, Artificial Intelligence, and the Problem of Meaningful Human Control. *Philosophical Journal of Conflict and Violence*, *5*(1), 53–72. https://doi.org/10.22618/TP.PJCV.20215.1.139004

[30] Mecacci, G., & Santoni de Sio, F. (2020). Meaningful human control as reason-responsiveness: the case of dual-mode vehicles. *Ethics and Information Technology*, *22*(2), 103–115. https://doi.org/10.1007/s10676-019-09519-w

[31] Santoni de Sio, F., Mecacci, G., Calvert, S., Heikoop, D., Hagenzieker, M., & van Arem, B. (2022). Realising Meaningful Human Control Over Automated Driving Systems: A Multidisciplinary Approach. *Minds and Machines*, 1–25. https://doi.org/10.1007/s11023-022-09608-8

[32] Cornelissen, N. A. J., van Eerdt, R. J. M., Schraffenberger, H. K., & Haselager, W. F. G. (2022). Reflection machines: increasing meaningful human control over Decision Support Systems. *Ethics and Information Technology*, *24*(2). https://doi.org/10.1007/s10676-022-09645-y

[33] Wagner, B. (2019). Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems. *Policy & Internet*, *11*(1), 104–122. https://doi.org/10.1002/poi3.198

[34] Robbins, S. (2023). The many meanings of meaningful human control. *AI and Ethics*, 1–12. https://doi.org/10.1007/s43681-023-00320-6

[35] Davidovic, J. (2023). On the purpose of meaningful human control of AI. *Frontiers in Big Data*, *5*, 1017677. https://doi.org/10.3389/fdata.2022.1017677

[36] European Commission. (2019). *Ethics Guidelines for Trustworthy AI.* European Commission. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

[37] Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association : JAMIA*, *19*(1), 121–127. https://doi.org/10.1136/amiajnl-2011-000089

[38] Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, *51*(5), 991–1006. https://doi.org/10.1006/ijhc.1999.0252

[39] Aziz El-Banna, A.A., Wu, K. (2021). Introduction to Underwater Communication and IoUT Networks. In: *Machine Learning Modeling for IoUT Networks*. SpringerBriefs in Computer Science. Springer, Cham. https://doi.org/10.1007/978-3-030-68567-6_1