Environment-Centric Safety Requirements forAutonomous Unmanned Systems

Conference or Workshop Item

Environment-Centric Safety Requirements for Autonomous Unmanned Systems

Yixing Luo[1,2], Yijun Yu[3], Zhi Jin[1,2], Haiyan Zhao[1,2]
[1]Key Lab. of High-Confidence Software Technologies (MoU), Peking University, Beijing, China
[2]Department of Computer Science and Technology, School of EECS, Peking University, Beijing, China
[3]The Open University, Milton Keynes, United Kingdom

***Abstract*—Autonomous unmanned systems (AUS) emerge to take place of human operators in harsh or dangerous environments. However, such environments are typically dynamic and uncertain, causing unanticipated accidents when autonomous behaviours are no longer safe. Even though safe autonomy has been considered in the literature, little has been done to address the environmental safety requirements of AUS systematically. In this paper, we conduct a systematical literature review and set up a taxonomy of environment-centric safety requirements for AUS. We then analyse the neglected issues to suggest several new research directions towards the vision of environmental-centric safe autonomy.**

***Index Terms*—Unmanned Systems, Autonomy, Environmental Safety.**

I. Introduction

Unmanned systems have been exploited in the missions which are dull, dirty, difficult, and dangerous for humans [1]. Without a human pilot onboard, however, unmanned systems must be sufficiently autonomous to cope with dynamics and uncertainties in the environment. *Autonomous unmanned systems* (AUSs) emerge to avoid harmful situations for an extended period without the need for any direct assistance or intervention from humans. Furthermore, the advent of AUS is transforming traditional industries of transportation, manufacturing and logistics [2].

Historically, numerous crashes have occurred during the testing or operation of AUSs. There have been hundreds of "CLASS A" (destroyed the aircraft or caused at least $2 million damage) crashes of military drones [3]. This year, all Boeing 737 Max 8 aircraft have been grounded after its second crash triggered by an automated system known as the Manoeuvring Characteristics Augmentation System[1]. This presents a dilemma to AUSs between potential benefits and risks. In general, for AUSs applied to safety-critical scenarios (i.e. transportation, search and rescue, surgery, etc.), safety requirements for AUSs should involve all stakeholders, as a tiny mistake could result in human injuries, significant property loss, or damage to the environment.

Safety concerns of AUSs have been a recognised problem on the agenda of stakeholders. Responsible authorities or organisations have formulated safety regulations and standards (e.g., CAP 722 for unmanned aircraft systems by Civil Aviation Authority [4]). The *York Global Initiative for Safe Autonomy*[2] aims at safe autonomy techniques for an enriched, healthier and more sustainable society.

Currently, many researchers concentrate on AUS safety, e.g., faults detection and diagnosis in robotic systems [5] and potential safety and security threats assessment for drones [6]. Here, safety is regarded as an inherent quality and the system is required to be free from losses and accidents caused by failures and errors. However, for an AUS operating in an *actual* environment, merely running the system correctly to ensure the safety of the system is not enough [7]. With an incomplete assumption of system states and environmental conditions, unanticipated accidents may occur in the interaction between the AUS and environment. Therefore, there is a quest for a taxonomy of environmental safety requirements to further constrain the system's autonomous behaviours.

To have a clear view of the fundamental environment-centric safety requirements for AUSs, we carry out a systematic literature review on this topic and figure out countermeasures to provide guarantees for its safe operations. We investigated papers from 2009 to 2019 about safety concerns for AUSs including drones and other robotic systems. All the articles are reviewed with respect to two research questions: (a) what environmental-centric safety requirements are put forward? and (b) how to satisfy these requirements autonomously?

Specifically, we set up a model for conceptual architecture to describe the interaction of autonomous unmanned systems and environment. Entities in the environment are classified as other systems, humans, and constraints for AUS to obey. Environment-centric safety requirements are elicited from these entities, i.e., the AUS is responsible for the insistence and correctness of its behaviours by eliminating internal risks and potential conflicts to protect the environment from harm. We examine the existing countermeasures from the perspective of processes in the AUS's working flow to see whether those requirements can be well implemented in AUS.

The rest of the paper is organised as follows. Section II describes the interaction architecture between the AUS and its environment. Section III sets up a taxonomy of environment-centric safety requirements for AUSs based on the interaction architecture. Section IV presents the survey results by analysing existing solutions and their maturity from the survey and proposes open research problems. Finally, Section V concludes the paper.

[1]https://en.wikipedia.org/wiki/Boeing_737_MAX_groundings
[2]https://www.york.ac.uk/safe-autonomy

II. Interactions between AUSs and Environment

The robotic systems are AUSs. From systematic literature reviews [8], [9], most robotic systems follow a system architecture of control loops with the monitoring-planning-analysis-execution (MAPE) loops [10]. Following this discovery, we develop a conceptual reference architecture for characterising the interactions between an autonomous unmanned system (AUS) and its environment, as shown in Figure 1.

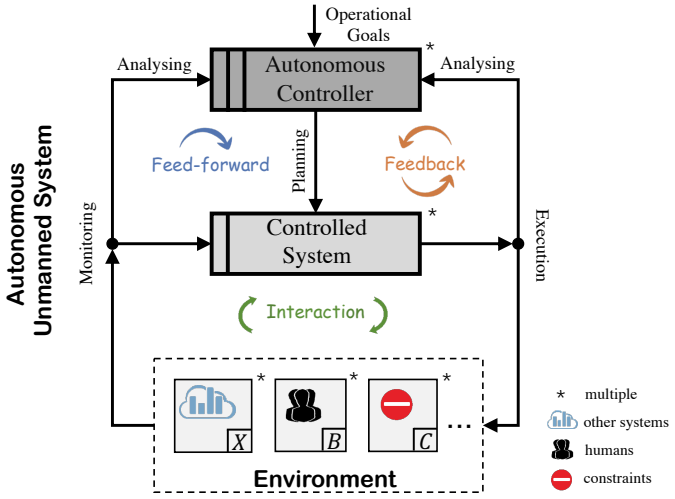

Figure 1. Interaction architecture for AUS and environment

This architecture highlights the structural relationship between an AUS and its operating environment. To highlight the environmental properties in relation to safety requirements, we represent the context diagram of the system by using the notations of the Problem Frames [11]: rectangular nodes denote the domain entities and edges their shared interfaces, where arrows highlight the direction of control.

Based on the control theory, an AUS consists of the controlled system and the controller [12]. It is the autonomous controller that monitors input disturbances, or analyses the execution results, to generate the plan for the controlled system to actuate. These steps form nested feedback and/or feed-forward loop. Here, the *Operational Goals* represents the requirements for the system to achieve. For example, "*more responsive*" is one of the non-functional requirements and "*measuring the distance to the destination*" is one of the functional requirements. Adjustments are taken by the actuators in the controlled system according to the plan computed by the controller to close the gap between the *Operational Goals* and the monitoring inputs in a feed-forward loop, or reduce the error between the *Operational Goals* and the execution outputs in a feedback loop.

The environment interacts with the system through inputs and outputs. Typically, the environment consists of multiple types of domain entities, including other (unmanned) systems sharing the same environment, people around, and physical/logical entities that may pose constraints such as obstacles to avoid, rules/standards to obey, etc. Amongst them, humans are *biddable domains* marked by "B" in the bottom-right corner, the physical entities are *causal domains* marked by "C", and the logical entities are *lexical domains* marked by "X". Differences of these environmental entities in varying degrees of dynamics and uncertainties could have significant impact on the behaviour of the system and its safety requirements.

III. Environment-Centric Safe Autonomy

In the literature, most of the safety concerns about AUS and its runtime environment can be classified as (i) malfunction or incorrect actions of AUS, resulting in failures and errors; (ii) conflicts with entities in the environment. *Environment-centric* requirements are put forward to map these safety concerns into implementable requirements for AUSs, among which, *hazard-elimination* and *conflict-avoidance* describe the ability of AUSs to protect the environment from being threatened.

To address these safety concerns, solutions have been proposed at different stages of the MAPE workflow illustrated in the interaction architecture (Section II). Furthermore, each process can be refined into sub-dimensions with respect to the properties of the safety methods.

A. Taxonomy of Safety Requirements for AUSs

1) Hazard-elimination: Conventionally, *hazard-elimination* is required to be designed into an AUS to minimise or eliminate hazards that arise from AUS failures and errors (e.g., design flaws, or poor performance with uncertainties in the environment). *Fail-safe* and *fault-tolerance* are the abilities and requirements of an AUS to protect the environment from being threatened in the event of failures and errors.

A fail-safe system responds in a way that causes little harm to the entities in the environment inherently, in the sense that being fail-safe either anomaly (i.e. deviation from normal behaviours) or system failures that may be posed to the environment can be detected so that risk reduction methods are leveraged to mitigate those potential hazards [13].

Compared with fail-safe, fault-tolerance mechanisms enable the system to continue its intended operations, or at a reduced level, rather than allowing it to fail. That is the reason why fault-tolerance is essential for life-critical systems.

2) Conflict-avoidance: Mitigation of risks is not always sufficient in safety-critical situations in the environment when there is little tolerance to accidents and losses such as human injuries, property damages, ecological harm. In advance, an AUS is required to be aware of avoiding risks, i.e., restriction violations, unintentional decision conflicts or collisions with other systems or human beings, etc.

Typically, AUSs are not working alone. They may share resources with other systems in the environment at different contextual levels. Examples of such *inter-system* conflict-avoidance requirements include UAVs, self-driving cars on the roads, co-robots, etc. Therefore, the relationship of AUSs with related systems should be dealt with carefully to avoid interfering with their normal functions.

During the interaction between human and robotic systems, especially when direct contact occurs between a person and a robot, it is essential to ensure that the human partner is safe [14]. To improve the safety of those humans interacting with an AUS, the system needs to be accountable for its actions. Namely, rational decisions are made to avoid potential physical injuries, privacy invasion, and assets loss.

In real-life scenarios, constraints for AUS safety autonomy always exist, and they are strongly related to environmental conditions. Such constraints, being specifically defined in different scenarios, can be logistic regulations and standards to obey, physical obstacles to avoid in real-world, and spatial restrictions like protected zones, etc. Compared to autonomous systems operating indoor or on the ground, the autonomy of the systems in 3-D open spaces such as the air spaces or the sea areas might be better achieved with fewer physical obstacles. However, natural disturbances and obstacle modelling are harder to handle when planning a collision-free way [6]. It is thus important for AUSs to be aware of the environmental conditions and constraints because the differences in these features could lead to completely distinct safety requirements.

B. Procedures for Safety in AUSs

Workflows in AUSs can be described from monitoring (input), analyzing (feed-forward and feedback), planning and execution (output) procedures, as shown in Figure 1. We further refine them into sub-dimensions at different steps according to the property of proposed solutions.

1) Monitoring: To cope with the complexity of environments, monitoring is the prerequisite for AUSs to detect safety risks and make appropriate decisions. The methods of self-state monitoring and environment awareness fall into proactive and reactive categories.

Equipped with *proactive* sensors, AUSs can collect information about the physical world to achieve internal desires or intentions proactively [15]. This ability refers to being aware of those underlying risks in the system or the uncertainty and dynamics in the environment (e.g., unknown obstacles [16], intruding vehicles [17]) earlier to prevent problems from happening in the first place.

For *reactive* sensing, changes of environmental properties and monitored parameters of systems trigger system's reactions after analysing and planning. Traditional approaches respond to risks by identifying a vulnerability and developing a patch to eliminate it, similar to securing the air vehicle systems [18]. With less time to reason and predict, this cycle repeats itself with each newly found risk.

2) Analysing: With the information collected by the sensors or monitors, the AUS can make reasonable assumptions about its own and environmental states to avoid dangerous situations. The main task is to detect where the failure or risk occurs.

Reasoning and prediction of the external environmental states depend on the prior knowledge acquired from domain experts (i.e., the given environment model) or on a large quantity of experience learnt from both successes and failures in previous explorations (i.e., the learnt environment model). An environment model can be set up before the system is put into use, while a data-efficient learning process with safety-critical constraints [19] is needed for environment modelling if there are insufficient contextual information and domain knowledge at design time.

3) Planning: In response to the changes and safety threats in the environment, decisions for the action policy are made by the AUS. Similar to different types in self-adaptive systems, the planning for AUSs has either *offline* (static) or *online* (dynamic) ways for policy generation and decision-making [20].

Assuming that the safety concern has been revealed before the mission starts, time is taken to optimise safety strategies usually. However, such countermeasures only have predetermined or fixed number of actions and rules based on developers' analysis.

Rule-based motion planning is suitable for handling well-examined errors and failure modes in the system and predetermined constraints outside the system. On the other hand, adaptive actions and policies can be extended at runtime through the online planning process. It seems that the online learning-based and the dynamic planning algorithms have a greater potential than offline ones in more complicated tasks.

4) Execution: To apply policies decided by the planning process, effectors map the commands into actions according to their underlying techniques, getting rid of unsafe states and solving conflicts either *independently* or *cooperatively* [21].

For the independent mechanism, AUSs are fully responsible for their behaviours to ensure environmental safety. In this case, the designers care less about the external systems, but keep raising their own AUSs' self-adaptive ability in face of external risks [22]. However, it calls for a cooperative collision-avoidance and scheduling scheme for AUSs that sharing contexts with other systems and/or human beings. Collaboration and coordination mechanisms can raise the efficiency of problem-solving. Communication protocol needs to be deployed for information sharing [23].

IV. Environmental Safe AUSs

In this section, we clarify our survey process and analyse the results so that it is possible to discern gaps in addressing safety requirements.

A. Systematic Literature Survey

According to the taxonomy illustrated in Section III, our review has ended with a total of 64 papers from 2009 to 2019, after applying inclusion (e.g., considering only journal, conference papers) and exclusion criteria (e.g., removing duplicate entries and considering only the extended version with complete solutions to the question) to the initial search results. We use these selective papers to analyse environment-centric safety concerns for AUSs and evaluate the maturity of solutions at different stages of the workflow in AUSs.

The survey results of the state-of-the-art research on this topic are classified as shown in Table I, in which, we use three kinds of icons to qualitatively indicate the maturity of the solution proposed:

Table I
ENVIRONMENT-CENTRIC SAFETY REQUIREMENTS FOR AUSS FROM DIFFERENT ASPECTS OF AUTONOMY.

Environment-Centric Safety Requirements		Procedures for safety in AUS															
		Monitoring				Analysing				Planning				Execution			
		proactive		reactive		environment model given		environment model learning		offline		online		independent		collaborative	
Hazard-elimination	fail-safe	○	[30]	●	[26]	●	[27]	○		●	[29]	○	[22]	●	[29]	○	
	fault-tolerance	○	[30]	●	[27]	●	[30]	○		●	[23]	◑	[32]	●	[33]	◑	[23]
Conflict-avoidance	inter-system	○	[25]	◑	[35]	◑	[34]	○	[44]	◑	[35]	○	[34]	○	[34]	◑	[49]
	human	●	[37]	○	[8]	◑	[8]	◑	[37]	○	[8]	◑	[38]	◑	[38]	○	[39]
	constraints	●	[40]	◑	[43]	●	[24]	◑	[19]	●	[41]	◑	[19]	●	[19]	○	

● Productivity ◑ Enlightenment ○ Trigger

The icons refer to the maturity of existing solutions for AUSs to satisfy environmental safety requirements in the workflow.

- *Productivity*: methods for this step of the workflow are well-examined or there is a well-established methodology to solve safety concerns, e.g., methods for collision-avoidance threat assessment and decision making in intelligent vehicle systems have been classified in [24].
- *Enlightenment*: There is room for making progress on optimising existing solutions, e.g., safety constrained MDP has been proposed for autonomous robots in a previously unknown environment [19]; however, the technique cannot handle more complex scenarios.
- *Trigger*: More research efforts are needed to narrow the gap between requirements and the realities, e.g., with more unmanned systems sharing the same working space. It is important for the proposed techniques to be aware of and not interfere with other systems [25].

Furthermore, a few intersections between safety and autonomous concerns are not covered by the existing work surveyed.

B. Analysing Results

In the literature, hazard-elimination is regarded as the basic requirement for environmental safety. Failure modes [26] and risk models [27] are identified and established at design time to examine whether controlled variables are out of the safe boundary [28]. On detecting the anomaly, emergency plans such as emergency pause (e.g., stopping system's motion) and emergency stop (e.g., disabling its power module [29]) are carried out according to handmade rules for fail-safe. Online planning is sometimes needed for drones to return autonomously to its starting point after GPS spoofing [22].

At least in some researchers' opinion [30], however, trying to identify and avoid any source of disruptions is in vain and what matters for AUSs is how to control the vulnerabilities and automatically restore its normal state. That partially explains why proactive monitoring is less covered in AUSs. The realisation of fault-tolerance puts forward higher demand on the system to recover from failures or erroneous states. Data-based anomalies detection and analysis are well achieved for fault detection in literature [31]. Recovery policies and measures can be real-time signal reconstruction [32], motion modification, and mission adjustment [33], roles reversal between different agents in the system [23], etc.

Conflict-avoidance is critical for autonomous systems deployed in actual scenarios, as it is the fundamental requirement from the environment. Competitions exist when resources shared by AUSs with other systems are limited, and AUSs should be careful of possible intruders and assure the safety of other systems [34]. In crowded workspaces (e.g., roads, factories, urban airspace), the systems' capability of sensing and avoiding other systems is crucial. Efforts have been done in inter-system communication and coordination (e.g., a UAV can broadcast its location to other vehicles for identifying and making way). Reminder from those scheduling systems like TCAS [35] is widely-used as guidance for aviation navigation. Although those manually specified rules and inter-operable equipment enjoy high credits from the authorities, they cannot solve the problem of bandwidth limit as extra information is transmitted and difficulty in heterogeneous autonomous systems communication [36].

To protect human safety and assets in human-AUS coexistence environment, solutions have been proposed that monitor the amount of energy injected into the system by actuator [8], predict human motions [37], real-time motion plan and control design for human-robot collaborative safety [38], change safe regions of the system dynamically [9]. Realising of high likelihood to fail, robots can also report the risk to its user [39]. Equipped with high-precision cameras for environmental information collection, AUSs may also intrude into human being's private space [6] and be regraded as "spiders" from users' perceptions. Claims for more "friendly" AUSs may help to ease this conflict by increasing the sense of safety and security of human beings.

For various kinds of constraints in the environment, countermeasures for neither physical [40] nor logical [17] or geopolitical [41] restrictions are covered in the literature. Most techniques for conflict avoidance with dynamic or uncertain physical obstacles operate in a timely fashion, which relies on the system's ability to proactively perceiving and planning. Given the behaviour models of the external entity, scene analysing and prediction can be framed as Markov decision

process (MDP) [42] or its variants POMDP [40]. In the absence of models, inverse reinforcement learning (IRL) and its variants Bayesian IRL [19] and deep IRL could infer latent reward function by exploration and exploitation.

For the assurance of safety and privacy of protected areas, constraints can be settled inherently like geo-fencing system used by DJI [43]. However, there are also reasons for temporarily forbidden flights in certain events, e.g., the lack of updating the latest no-fly zones could exert a serious impact on those protected areas.

C. Open Problems and Challenges

Apart from the existing solutions for AUSs working in a mission-critical environment safely, additional directions need further investigations.

Learning-based Environmental Model Structures. With little domain knowledge, it is challenging to acquire a well-established environment model at design time for those uncharted areas. As such, reinforcement learning may help through iterative trial-and-error to accumulate experience from the interactions for the system to be aware of opponent factors in the environment [44]. However, the cost of repairing and the lack of tolerance to execution failures highly restrict its learning ability [19].

Additionally, certifications for learning-based components are suspected given the inherent vulnerability of neural networks [45]. Taking a combined view of software engineering and dependability of artificial intelligence, the safety of learning-enabled components in AUSs should be verified. Initial attempt [46] has been taken at the verification of an artificial neural network-based feed-forward loop controller.

Addressing Complexity in Online Planning. Towards full autonomy, it is critical that the AUSs are capable of responding to anomalous events while minimising underlying risks. Through dynamic planning online, more uncertain and risky situations in the environment can be handled; however, due to unpredictability, less reliability, and trustworthiness, it is not so well accepted by safety regulators and authorities compared with manually specified rules [47].

Besides, online planning occupies intensive computing resources of embedded systems or depends on the high-bandwidth and low-latency to offload computation to the cloud. Therefore, a common way is to combine the strengths of individual methods, i.e., by adjusting the parameters of manually created rules actively at runtime [17] or by allowing online modifications of precomputed offline results [48]. To relieve this burden, it is hoped that computations can be off-loaded to the cloud system with less latency if the 5th Generation Wifi is widely deployed.

Collaboration for Environment Safety. To go beyond the capabilities or knowledge of the individual system, a loosely-coupled network formulated by multiple interacting controlled systems for the mission complement is important in practice (e.g., Amazon Prime Air). With alternative equipment and agents, fault-tolerance can be better achieved intra-system [23]. Additionally, coordination and negotiation between the system and its collaborators help with a more efficient resolution for potential conflicts [36]. Considering the heterogeneous technologies underlying autonomous cyber-physical systems, it is difficult for regulatory authorities to make generally-applicable rules [47].

Without a centralised architecture or central controller, however, a common semantics for information processing is required to achieve consensus agreement amongst intra-system or inter-system [49] parties in a collaborative context. In another way, raising the ability to be self-adaptive and deployment enough for the safety of the entire group, e.g., minimising the gaps with partners or learn from its opponents in the environment.

V. Conclusions

AUSs are likely to flourish in a foreseeable future. However, they need to ensure the safety of the entities in their operating environment.

This paper reports our findings from a literature review of environment-centric concerns for AUSs. Using a control-theoretic reference architecture, we classify the requirements of environmental safety in an MAPE process and examine their somewhat latent relationships. From this analysis, a taxonomy of *environment-centric* safety requirement is proposed and a few research directions are suggested to provide safety to AUSs. We hope that this work will motivate the research community to focus more on addressing the challenges identified, making the future AUSs environmentally safer.

Acknowledgement

The work is supported in part by the National Natural Science Foundation of China under Grant Nos. 61620106007 and 61751210, the EPSRC grants in the UK (SAUSE), ERC Adv. Grant on Adaptive Security and Privacy, EU H2020 EngageKTN grant on Drone Identity. Zhi Jin is corresponding author.

References

[1] S. G. Gupta, M. M. Ghonge, and P. Jawandhiya, "Review of unmanned aircraft system (UAS)," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 2, no. 4, pp. 1646–1658, 2013.

[2] M. Erdelj, E. Natalizio, K. R. Chowdhury, and I. F. Akyildiz, "Help from the sky: Leveraging UAVs for disaster management," *IEEE Pervasive Computing*, vol. 16, no. 1, pp. 24–32, 2017.

[3] E. Chow, A. Cuadra, and C. Whitlock, "Fallen from the skies," *The Washington Post*, vol. 20, 2014.

[4] C. A. Authority, "CAP 722 Unmanned Aircraft System Operations in UK Airspace—Guidance," *Directorate of Airspace Policy*, 2010.

[5] E. Khalastchi and M. Kalech, "On Fault Detection and Diagnosis in Robotic Systems," *ACM Computing Surveys (CSUR)*, vol. 51, no. 1, p. 9, 2018.

[6] R. Altawy and A. M. Youssef, "Security, privacy, and safety aspects of civilian drones: A survey," *ACM Transactions on Cyber-Physical Systems*, vol. 1, no. 2, p. 7, 2017.

[7] N. Leveson, *Engineering a safer world: Systems thinking applied to safety*. MIT press, 2011.

[8] G. A. Folkertsma, S. S. Groothuis, and S. Stramigioli, "Safety and guaranteed stability through embedded energy-aware actuators," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 2902–2908.

[9] M. P. Polverini, A. M. Zanchettin, and P. Rocco, "Real-time collision avoidance in human-robot interaction based on kinetostatic safety field," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 4136–4141.

[10] S. Gerasimou, R. Calinescu, S. Shevtsov, and D. Weyns, "Undersea: an exemplar for engineering self-adaptive unmanned underwater vehicles," in *2017 IEEE/ACM 12th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS)*. IEEE, 2017, pp. 83–89.

[11] M. Jackson, *Problem frames: analysing and structuring software development problems*. Addison-Wesley, 2001.

[12] K.-Y. Cai, J. W. Cangussu, R. A. DeCarlo, and A. P. Mathur, "An overview of software cybernetics," in *Eleventh Annual International Workshop on Software Technology and Engineering Practice*. IEEE, 2003, pp. 77–86.

[13] N. G. Leveson and K. A. Weiss, "Software system safety," in *Safety Design for Space Systems*. Elsevier, 2009, pp. 475–505.

[14] M. Vasic and A. Billard, "Safety issues in human-robot interactions," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 197–204.

[15] R. McAllister, Y. Gal, A. Kendall, M. Van Der Wilk, A. Shah, R. Cipolla, and A. V. Weller, "Concrete problems for autonomous vehicle safety: Advantages of bayesian deep learning." International Joint Conferences on Artificial Intelligence, Inc., 2017.

[16] O. Adiyatov and H. A. Varol, "A novel RRT*-based algorithm for motion planning in dynamic environments," in *2017 IEEE International Conference on Mechatronics and Automation (ICMA)*. IEEE, 2017, pp. 1416–1421.

[17] Z. N. Sunberg, M. J. Kochenderfer, and M. Pavone, "Optimized and trusted collision avoidance for unmanned aerial vehicles using approximate dynamic programming," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 1455–1461.

[18] D. Cofer, A. Gacek, J. Backes, M. W. Whalen, L. Pike, A. Foltzer, M. Podhradsky, G. Klein, I. Kuz, J. Andronick *et al.*, "A Formal Approach to Constructing Secure Air Vehicle Software," *Computer*, vol. 51, no. 11, pp. 14–23, 2018.

[19] A. Wachi, Y. Sui, Y. Yue, and M. Ono, "Safe exploration and optimization of constrained mdps using gaussian processes," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018, pp. 6548–6556.

[20] M. Salehie and L. Tahvildari, "Self-adaptive software: Landscape and research challenges," *ACM transactions on autonomous and adaptive systems (TAAS)*, vol. 4, no. 2, p. 14, 2009.

[21] T. Taleb, A. Benslimane, and K. B. Letaief, "Toward an effective risk-conscious and collaborative vehicular collision avoidance system," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 3, pp. 1474–1486, 2010.

[22] D. He, Y. Qiao, S. Chan, and N. Guizani, "Flight security and safety of drones in airborne fog computing systems," *IEEE Communications Magazine*, vol. 56, no. 5, pp. 66–71, 2018.

[23] I. Vistbakka, A. Majd, and E. Troubitsyna, "Multi-layered Approach to Safe Navigation of Swarms of Drones," in *International Conference on Computer Safety, Reliability, and Security*. Springer, 2018, pp. 112–125.

[24] J. Dahl, G. R. de Campos, C. Olsson, and J. Fredriksson, "Collision Avoidance: A Literature Review on Threat-Assessment Techniques," *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 1, pp. 101–113, 2019.

[25] M. Vierhauser, S. Bayley, J. Wyngaard, W. Xiong, J. Cheng, J. Huseman, R. R. Lutz, and J. Cleland-Huang, "Interlocking Safety Cases for Unmanned Autonomous Systems in Shared Airspaces," *IEEE Transactions on Software Engineering*, 2019.

[26] D. H. Stamatis, *Failure mode and effect analysis: FMEA from theory to execution*. ASQ Quality press, 2003.

[27] Y. Y. Haimes, *Risk modeling, assessment, and management*. John Wiley & Sons, 2015.

[28] R. Melnyk, D. Schrage, V. Volovoi, and H. Jimenez, "A third-party casualty risk model for unmanned aircraft system operations," *Reliability Engineering & System Safety*, vol. 124, pp. 105–116, 2014.

[29] M. Y. Jung, R. H. Taylor, and P. Kazanzides, "Safety design view: a conceptual framework for systematic understanding of safety features of medical robot systems," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 1883–1888.

[30] W. Young and N. G. Leveson, "An integrated approach to safety and security based on systems theory." *Commun. ACM*, vol. 57, no. 2, pp. 31–35, 2014.

[31] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.

[32] R. G. Dutta, X. Guo, T. Zhang, K. Kwiat, C. Kamhoua, L. Njilla, and Y. Jin, "Estimation of safe sensor measurements of autonomous system under attack," in *Proceedings of the 54th Annual Design Automation Conference 2017*. ACM, 2017, p. 46.

[33] J. D. Andrews, J. Poole, and W.-H. Chen, "Fast mission reliability prediction for Unmanned Aerial Vehicles," *Reliability Engineering & System Safety*, vol. 120, pp. 3–9, 2013.

[34] J. Godoy, I. Karamouzas, S. J. Guy, and M. L. Gini, "Moving in a Crowd: Safe and Efficient Navigation among Heterogeneous Agents," in *IJCAI*, 2016, pp. 294–300.

[35] J. E. Holland, M. J. Kochenderfer, and W. A. Olson, "Optimizing the next generation collision avoidance system for safe, suitable, and acceptable operational performance," *Air Traffic Control Quarterly*, vol. 21, no. 3, pp. 275–297, 2013.

[36] S. Niemczyk and K. Geihs, "Adaptive run-time models for groups of autonomous robots," in *Proceedings of the 10th International Symposium on Software Engineering for Adaptive and Self-Managing Systems*. IEEE Press, 2015, pp. 127–133.

[37] J. S. Park, C. Park, and D. Manocha, "Intention-Aware Motion Planning Using Learning Based Human Motion Prediction," in *Proceedings of Robotics: Science and Systems*, Cambridge, Massachusetts, July 2017.

[38] J.-H. Chen and K.-T. Song, "Collision-Free Motion Planning for Human-Robot Collaborative Safety Under Cartesian Constraint," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1–7.

[39] J. Sattar and J. J. Little, "Ensuring safety in human-robot dialog—A cost-directed approach," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 6660–6666.

[40] Y. Wang, S. Chaudhuri, and L. E. Kavraki, "Bounded policy synthesis for POMDPs with safe-reachability objectives," in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2018, pp. 238–246.

[41] T. R. Jorris and R. G. Cobb, "Three-dimensional trajectory optimization satisfying waypoint and no-fly zone constraints," *Journal of Guidance, Control, and Dynamics*, vol. 32, no. 2, pp. 551–572, 2009.

[42] L. Liu and N. Michael, "An MDP-based approximation method for goal constrained multi-MAV planning under action uncertainty," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 56–62.

[43] DJI, "DJI Introduces New Geofencing System for its Drones," News, Nov. 2015. [Online]. Available: https://www.dji.com/newsroom/news/dji-fly-safe-system.

[44] J. Foerster, R. Y. Chen, M. Al-Shedivat, S. Whiteson, P. Abbeel, and I. Mordatch, "Learning with opponent-learning awareness," in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2018, pp. 122–130.

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

[46] C. E. Tuncali, H. Ito, J. Kapinski, and J. V. Deshmukh, "Reasoning about safety of learning-enabled components in autonomous cyber-physical systems," in *2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC)*. IEEE, 2018, pp. 1–6.

[47] C. Torens, J. C. Dauer, and F. Adolf, "Towards Autonomy and Safety for Unmanned Aircraft Systems," in *Advances in Aeronautical Informatics*. Springer, 2018, pp. 105–120.

[48] D. Fridovich-Keil, S. L. Herbert, J. F. Fisac, S. Deglurkar, and C. J. Tomlin, "Planning, fast and slow: A framework for adaptive real-time safe trajectory planning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 387–394.

[49] F. L. Leite, R. Adler, and P. Feth, "Safety assurance for autonomous and collaborative medical cyber-physical systems," in *International Conference on Computer Safety, Reliability, and Security*. Springer, 2017, pp. 237–248.