Assured Autonomy Survey

Other titles in Foundations and Trends® in Privacy and Security

Hardware Platform Security for Mobile Devices
Lachlan J. Gunn, N. Asokan, Jan-Erik Ekberg, Hans Liljestrand, Vijayanand Nayani and Thomas Nyman
ISBN: 978-1-68083-976-0

Cloud Computing Security: Foundations and Research Directions
Anrin Chakraborti, Reza Curtmola, Jonathan Katz, Jason Nieh, Ahmad-Reza Sadeghi, Radu Sion and Yinqian Zhang
ISBN: 978-1-68083-958-6

Expressing Information Flow Properties
Elisavet Kozyri, Stephen Chong and Andrew C. Myers
ISBN: 978-1-68083-936-4

Accountability in Computing: Concepts and Mechanisms
Joan Feigenbaum, Aaron D. Jaggard and Rebecca N. Wright
ISBN: 978-1-68083-784-1

A Pragmatic Introduction to Secure Multi-Party Computation
David Evans, Vladimir Kolesnikov and Mike Rosulek
ISBN: 978-1-68083-508-3

Assured Autonomy Survey

Christopher Rouff
Johns Hopkins University Applied Physics Laboratory
christopher.rouff@jhuapl.edu

Lanier Watkins
Johns Hopkins University Applied Physics Laboratory
lanier.watkins@jhuapl.edu

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Privacy and Security

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

C. Rouff and L. Watkins. *Assured Autonomy Survey*. Foundations and Trends® in Privacy and Security, vol. 4, no. 1, pp. 1–116, 2022.

ISBN: 978-1-63828-039-2

Foundations and Trends® in Privacy and Security

Volume 4, Issue 1, 2022

Editorial Scope

Topics

Foundations and Trends® in Privacy and Security publishes survey and tutorial articles in the following topics:

- Access control
- Accountability
- Anonymity
- Application security
- Artifical intelligence methods in security and privacy
- Authentication
- Big data analytics and privacy
- Cloud security
- Cyber-physical systems security and privacy
- Distributed systems security and privacy
- Embedded systems security and privacy
- Forensics
- Hardware security
- Human factors in security and privacy
- Information flow
- Intrusion detection
- Malware
- Metrics
- Mobile security and privacy
- Language-based security and privacy
- Network security
- Privacy-preserving systems
- Protocol security
- Security and privacy policies
- Security architectures
- System security
- Web security and privacy

Information for Librarians

Foundations and Trends® in Privacy and Security, 2022, Volume 4, 4 issues. ISSN paper version 2474-1558. ISSN online version 2474-1566. Also available as a combined paper and online subscription.

Contents

Assured Autonomy Survey

Christopher Rouff and Lanier Watkins

Johns Hopkins University Applied Physics Laboratory, USA; christopher.rouff@jhuapl.edu, lanier.watkins@jhuapl.edu

ABSTRACT

Autonomous robots and other systems are no longer just subjects of science fiction, but are becoming common occurrences in our everyday lives. Autonomous vacuum cleaners, lawnmowers, and other household helpers are starting to be common place, with autonomous cars now being tested around the world and autonomous drones starting to be used to deliver packages and groceries. Though they will soon be common occurrences in everyday life, assuring their safety, privacy and security is still a huge challenge. A number of autonomous car accidents have occurred after millions of miles of testing, as well as other injuries from other types of autonomous systems. Assuring the proper behavior and safety of autonomous systems is an important endeavor to reduce risks in using them. This monograph discusses assurance for autonomous systems, the different approaches to assuring autonomy, formal analysis, cybersecurity, certification and research challenges.

Christopher Rouff and Lanier Watkins (2022), "Assured Autonomy Survey", Foundations and Trends® in Privacy and Security: Vol. 4, No. 1, pp 1–116. DOI: 10.1561/3300000027.

1

Introduction

Autonomous systems will soon be ubiquitous in our society, saving us time, performing tasks we do not want to do, caring for us and keeping us safe, often referred to as dull, dirty and dangerous tasks (Connelly *et al.*, 2006). Autonomous robots in homes and businesses are already cleaning floors, mowing lawns, delivering meals and packages, and the technology is now driving cars and trucks. Assuring autonomous systems performance and safety is still a huge challenge. A number of autonomous car accidents have occurred after millions of miles of testing and injuries are also occurring from other types of autonomous systems (Banks *et al.*, 2018; Favarò *et al.*, 2017). Though some autonomous system accidents are minor, others have resulted in deaths to occupants or users, and there is the potential of other damage and injuries from the increasing number and types of autonomous systems that are being proposed.

With the increase of autonomy being used for a wide range of applications, assuring their behavior, trustworthiness, safety and security is still a huge challenge. Providing proper assurance can help prevent injuries, deaths and financial loss. The following subsections give a brief introduction to assured autonomy, providing definitions and terms that will be used in the remainder of this monograph.

1.1 Autonomy

Autonomous systems, also referred to highly automated systems (Falco *et al.*, 2021), have been defined by a number of authors, including Connelly *et al.* (2006), Huang (2007), Huang *et al.* (2007), and Truszkowski *et al.* (2009). Merriam-Webster dictionary defines autonomy as "the quality or state of being self-governing."[1] For software systems, this means they are not dependent on an outside entity for control or decision making. Connelly *et al.* (2006) define an autonomous system as "one that makes and executes a decision to achieve a goal without full, direct human control." Hutchison *et al.* (2018) describe autonomous systems as having the following properties:

Stateful - autonomous systems may need to use a large amount of internal memory to represent the environment in which they are operating, keep track of interactions with people and other entities, making models of the physical world around them, developing plans of actions, and reading and storing sensor data that is constantly being received and that needs to be analyzed. Much of the data autonomous systems receive is interrelated and needs to be retained for differing periods of time for future reference and reasoning purposes.

Temporal - autonomous systems often have time and sequence related requirements. They often execute checklists or algorithms that are sequence oriented or where future actions are dependent on past results. An example is going point to point from an initial starting place to a destination. Decisions on the direction from one point may not be known until the autonomous system arrives at that point. In the future these points may be needed to backtrack or return to its starting point.

Distributed - autonomous systems typically contain multiple subsystems that are all communicating with each other over an internal network. Actuators and sensors often have their own processors

[1] See https://www.merriam-webster.com/dictionary/autonomy (date accessed: 28 May 2021).

and commands are sent to them from a central controller. Data may be sent to a modeling agent that keeps track of the system's current position, goals, plans and other high-level information. Race conditions, deadlocks and other distributed system errors can occur and need to be either tested for or checked through formal methods or other techniques.

Cyber-Physical - autonomous systems are often cyber-physical systems (CPSs) that are an integration of hardware and software components. Griffor *et al.* (2017), at the National Institute of Standards and Technology (NIST), describe CPSs as "smart systems that include engineered interacting networks of physical and computational components" (see Figure 1.1). In the NIST description, multiple CPSs can make up a system (such as an autonomous system) and multiple systems can make up a system of systems (such as a smart city). This makes autonomous systems different from traditional software systems in that they need to detect and deal with hardware failures and sensors that may provide faulty or no data since a human may not be available to detect and fix these problems.

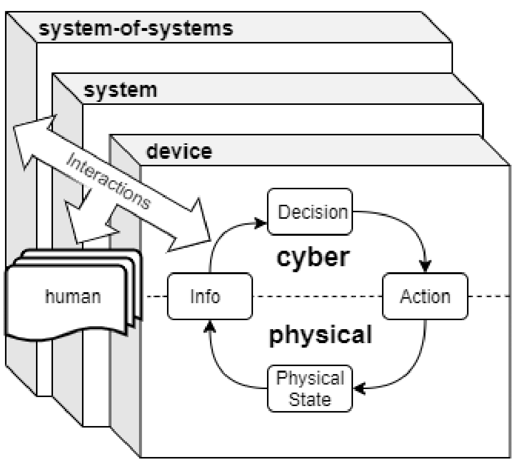

Figure 1.1: NIST conceptual model of a cyber-physical system (based on Griffor *et al.* (2017)).

A question that often comes up when describing autonomous systems is what is the difference between an automated and autonomous system. Both terms refer to processes that may be executed independently from start to finish without any human intervention. Truszkowski *et al.* (2009) describe automated processes as replacing routine manual processes with software and/or hardware. Automation follows a step-by-step sequence of steps that may or may not include human participation. The authors describe an autonomous system as having "self-governance" and "self-direction" and can complete a task independently of a human, and have the goal of emulating human processes rather than simply replacing them. Replacing human processes often requires the use of artificial intelligence (AI). Kunze *et al.* (2018) and Nascimento *et al.* (2019) provide some examples of the types of AI used in autonomous systems.

Autonomy may be applied gradually to systems as the technology is developed, making the system more autonomous over time (Truszkowski *et al.*, 2005). The system may start out with automation, with increasingly sophisticated or intelligent automated steps added until the system is self-governing and emulating human processes. Sheridan and Verplank (1978) describe ten levels of automation, with the final level being able to operate without human supervision, which could be construed as fully autonomous. The ten levels are (Sheridan and Verplank, 1978; MahmoudZadeh *et al.*, 2019):

1. System is controlled by an operator.
2. System helps operator by determining options.
3. System helps operator by determining options and suggesting one option.
4. System selects an action, which the operator may or may not execute.
5. System selects an action and executes it if approved by the operator.
6. System selects an action, informs the operator in plenty of time for the operator to stop the action.

7. System does the whole job and tells the operator what it did.

8. System does the whole job and only tells the operator if the operator asks.

9. System does the whole job and decides whether to tell the operator what it did.

10. System decides if the job should be done, does the whole job and decides if it should tell the operator.

At level 7, one could argue that the system is autonomous, since it is performing a job and only telling the operator about it after it is performed, which would mean that it is emulating a human process at this level (telling someone after a task is performed is often what humans do). Clough (2002) also defined ten levels of autonomy that ranges from a remotely piloted vehicle to a vehicle with human-like performance.

SAE international has defined six levels of automation that a vehicle may have (SAE, 2021). The SAE driving automation levels are:

- Level 0: No driving automation
- Level 1: Driver assistance
- Level 2: Partial driving automation
- Level 3: Conditional driving automation
- Level 4: High driving automation
- Level 5: Full driving automation

Level 0 is no autonomy and may not even have any automation. Level 1 is when one or more functions are automated. Level 1 does not require sensor information from the environment, and could just be a human activity that is automated. An example would be a simple cruise control that maintains a selected speed, but does not maintain distance from other vehicles or have other safety features. Level 2, partial driving automation, requires sensors to sense the environment, but still requires driver assistance. Examples would be lane control maintenance or collision avoidance breaking, which would need to sense the lines on the road or obstacles ahead of the vehicle. Level 3 requires a driver, but the driver is not required to monitor the environment, though the driver must be ready to take control of the vehicle at any time with little notice. Level 4, high driving automation, is where a vehicle is capable of performing all driving functions under certain conditions, with the driver having the option to take control of the vehicle. An example might be that the vehicle can operate autonomously only in good weather or in highway environments with good lane markings. Level 5 is defined as full automation, where the vehicle is able to perform all of the driving under all conditions, with the driver having the option to take control of the vehicle when they want. Level 5 could also be construed as emulating human processes, so could be considered an autonomous system.

At the lower levels of SAE driving automation, such as Levels 2 and 3, vehicles may have several independent systems providing automation, such as adaptive cruise and lane keeping technologies. They are usually different systems and can be operated at the same time, or one without the other. For a Level 4 or 5 vehicle, the automation/autonomy must be one integrated system since all of the components must work together (Figure 1.2). The adaptive cruise control may work with the lane keeping function to pass slower vehicles on a highway, along with the other autonomy components to make decisions about speed limits, directions, obstacle avoidance and other functions.

Adjustable autonomy is where the level of autonomy can be adjusted based on the task being performed (MahmoudZadeh *et al.*, 2019; Maheswaran *et al.*, 2003). Mostafa *et al.* (2019) define adjustable autonomy as providing "an autonomous system with variable autonomy in which its operators have the options to work in different autonomy states." Zieba *et al.* (2010) define adjustable autonomy as "the property of an autonomous system to change its level of autonomy while the system operates. The human operator, another system or the autonomous system itself, can adjust the autonomy level." Adjusting the level of autonomy in a system can allow a user to take control when the autonomy is no longer necessary, the autonomy is not operating correctly or the user

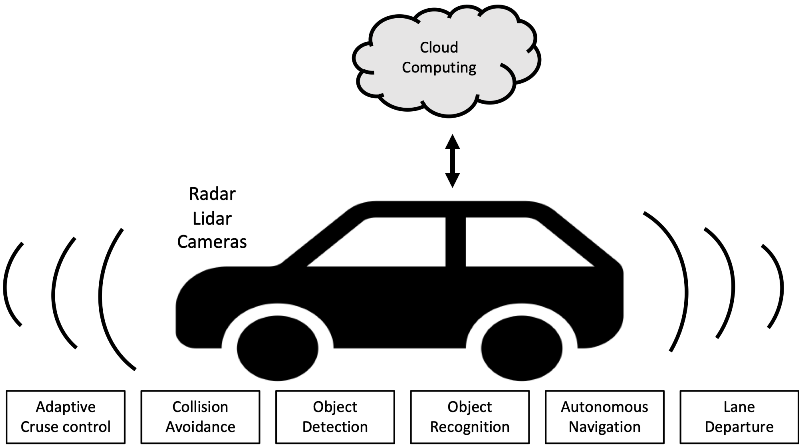

Figure 1.2: Components of an autonomous vehicle.

prefers to have manual control in a given situation. For an autonomous automobile, this could be when the road conditions or the weather makes it difficult for the autonomy to operate, when a sensor fails, or when the driver would just prefer to drive the car themself. Other examples may be when moving a robot through a tight area, where some of the autonomy is still necessary for navigation, or when learning or other intelligence is not operating properly and some manual control is necessary.

Though there is not a clear agreement between practitioners and researchers on when a system is autonomous and not autonomous, the ability to achieve goals given to it by a human with little or no input is important. For this monograph, the authors will use the definition.

1.2 Assurance

The ability of an autonomous system to operate independently of a human adds a large amount of complexity to the system. This added complexity greatly increases the chances of errors in the system, which adds risks since there may not be a human in the loop that could stop it from causing harm. There need to be assurances that autonomous systems will operate as intended, and appropriately, even in unanticipated or emergency situations.

Topcu *et al.* (2020) describe assurance as an interdisciplinary research area. What it means for a system to have assurance can differ depending on the community involved. For software engineers, assurance relates to the correct operation of the software that implements the system (Abrams and Zelkowitz, 1995; Saidi *et al.*, 2020; Smith *et al.*, 2020; Wing, 1990). For software engineers, software assurance can refer to formal verification, testing and other approaches used to ensure the system implementation matches the system requirements.

In cybersecurity, assurance can reflect the confidentiality, integrity and availability of a system, referred to as the CIA triad (Gamundani and Nekare, 2018; Samonas and Coss, 2014). Cherdantseva and Hilton (2013) define information assurance (IA) as providing protection by:

> reducing risks associated with information and information systems by means of a comprehensive and systematic management of security countermeasures, which is driven by risk analysis and cost-effectiveness.

Cooper *et al.* (2010) define IA as a:

> set of technical and managerial controls designed to ensure the confidentiality, possession of control, integrity, authenticity, availability, and utility of information and information systems. IA includes measures that protect and defend information and information systems by ensuring their availability, integrity, authentication, confidentiality, and nonrepudiation.

Similarly, NIST defines IA as (Barker, 2003):

> Measures that protect and defend information and information systems by ensuring their availability, integrity, authentication, confidentiality, and non-repudiation. These measures include providing for restoration of information systems by incorporating protection, detection, and reaction capabilities.

To provide IA for autonomous systems there needs to be a way to detect when they are under cyber attack, protecting themselves from attack, protecting the information they contain from being stolen, and ensuring that those that try to access the system and its information are authorized to do so.

Mueller (2019) describes three features that can serve as a foundation of assured autonomy: accuracy, reduction in bias and the ability to reverse engineer the decision-making processes. Accuracy refers to how the autonomy algorithm "senses and perceives the environment in a manner relatable to humans." This means that the perception of the autonomous system would be similar to how humans perceive the world so that humans can then better relate to how the autonomous system is acting, which reduces the perceived complexity of the system. The reduction in bias refers to the algorithmic and training of the autonomous system, whether the autonomous system is directed toward a result different than what it was originally intended. Algorithmic bias is usually the result of an improper specification of a function, coding errors and other bugs, which causes a degradation in performance of an autonomous system. Training bias is when data that is used by an AI algorithm that is controlling an autonomous system does not represent the environment in which the autonomous system is deployed. This can also be caused by malicious actors feeding the wrong or false data into an AI system, either during training or operations. This means that the autonomous system is taught the wrong information and will not act as intended. The ability to reverse engineer the decision-making processes allows a human to understand why an AI decision was made. When an AI system can explain a decision, it is referred to as Explainable AI (Phillips *et al.*, 2021). AI decision systems often have opaque algorithms, but it is important for humans to understand why a decision was made so that it can be corrected if needed, or just for an operator to understand why an intelligent system operated in a particular manner.

1.3 Monograph Outline

The remainder of this monograph expands on the above descriptions of assured autonomy. Section 2 provides an overview of assured autonomy and different aspects of system and software assurances. Section 3 discusses governance, trust, ethics and privacy of autonomous systems. This includes ways the government can be involved in assuring autonomous systems, ways of increasing the trust people have in them, how ethical behavior can be instilled in them, and maintaining the privacy of people who are using or coming into contact with them. Section 4 discusses assuring the correct operation of autonomous systems. This can be done through techniques, such as formal verification, testing and monitoring. Section 5 describes certification of current systems and proposals for certifying autonomous systems. It provides an example of the certification of aircraft software and multiple proposals for how autonomous systems could be certified. Section 6, the conclusion, discusses areas of research in assuring autonomous systems, and some concluding remarks.

References

Abrams, M. D. and M. V. Zelkowitz. (1995). "Striving for correctness". *Computers & Security*. 14(8): 719–738.

Alexander, R., M. Hall-May, and T. Kelly. (2007). "Certification of autonomous systems". In: *Proceedings of the 2nd Systems Engineering for Autonomous Systems (SEAS) Defence Technology Centre (DTC) Annual Technical Conference*.

AlFadhli, M. S., M. S. AlAli, and H. A. AlKulaib. (2021). "The Effect of Suez Canal Blockage on Crude Oil Prices: An Event Study Analysis". *IOSR Journal of Business and Management*. 23(4): 64–66.

Altman, I. (1975). "The Environment and Social Behaviour: Privacy, Personal Space, Territory, Crowding, Monterey".

Anaheed, A., C. Jian, S. Oleg, and L. Insup. (2013). "Assessing the overall sufficiency of safety arguments". In: *21st Safety-critical Systems Symposium (SSS'13), Bristol, United Kingdom*.

Andrews, A. M. (2018). "Concept of Operations for the Tactical use of Autonomous Unmanned Surface Systems". *Tech. rep.* Gravely Naval Research Group, US Naval War College Newport United States.

Awad, E., S. Levine, M. Anderson, S. L. Anderson, V. Conitzer, M. Crockett, J. A. Everett, T. Evgeniou, A. Gopnik, J. C. Jamison, *et al.* (2022). "Computational ethics". *Trends in Cognitive Sciences*.

Banks, V. A., K. L. Plant, and N. A. Stanton. (2018). "Driver error or designer error: Using the Perceptual Cycle Model to explore the circumstances surrounding the fatal Tesla crash on 7th May 2016". *Safety science*. 108: 278–285.

Barker, W. (2003). "Guideline for identifying an information system as a national security system". *Tech. rep.* No. Special Publication 800-59. National Institute of Standards and Technology.

Barrett, M. P. (2018). "Framework for improving critical infrastructure cybersecurity". *Tech. rep.* Version 1.1. Gaithersburg, MD, USA.

Bertot, Y. and P. Castéran. (2013). *Interactive theorem proving and program development: Coq'Art: the calculus of inductive constructions*. Springer Science & Business Media.

Bowen, J. P. and M. G. Hinchey. (1995). "Ten commandments of formal methods". *Computer*. 28(4): 56–63.

Cherdantseva, Y. and J. Hilton. (2013). "A reference model of information assurance & security". In: *2013 International Conference on Availability, Reliability and Security*. IEEE. 546–555.

Clarke, E. M., T. A. Henzinger, and H. Veith. (2018). "Introduction to model checking". In: *Handbook of Model Checking*. Springer. 1–26.

Clarke, E. M., W. Klieber, M. Nováček, and P. Zuliani. (2011). "Model checking and the state explosion problem". In: *LASER Summer School on Software Engineering*. Springer. 1–30.

Clarke, E. M. and J. M. Wing. (1996). "Formal methods: State of the art and future directions". *ACM Computing Surveys (CSUR)*. 28(4): 626–643.

Clough, B. T. (2002). "Metrics, schmetrics! How the heck do you determine a UAV's autonomy anyway". *Tech. rep.* Air Force Research Lab Wright-Patterson AFB OH.

Connelly, J., W. Hong, R. Mahoney Jr, and D. Sparrow. (2006). "Challenges in autonomous system development". In: *Performance Metrics for Intelligent Systems (PerMIS 2006) Workshop*. National Institute of Standards and Technology. Gaithersburg MAryland.

Cooper, S., C. Nickell, V. Piotrowski, B. Oldfield, A. Abdallah, M. Bishop, B. Caelli, M. Dark, E. K. Hawthorne, L. Hoffman, *et al.* (2010). "An exploration of the current state of information assurance education". *ACM SIGCSE Bulletin*. 41(4): 109–125.

Crum, V., D. Homan, and R. Bortner. (2004). "Certification challenges for autonomous flight control systems". In: *AIAA Guidance, Navigation, and Control Conference and Exhibit*. 5257.

Cummings, M. (2019). "Adaptation of human licensing examinations to the certification of autonomous systems". In: *Safe, autonomous and intelligent vehicles*. Springer. 145–162.

Danks, D. and A. J. London. (2017). "Regulating autonomous systems: Beyond standards". *IEEE Intelligent Systems*. 32(1): 88–91.

Dennis, L., M. Fisher, M. Slavkovik, and M. Webster. (2016). "Formal verification of ethical choices in autonomous systems". *Robotics and Autonomous Systems*. 77: 1–14.

Došilović, F. K., M. Brčić, and N. Hlupić. (2018). "Explainable artificial intelligence: A survey". In: *2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)*. IEEE. 0210–0215.

Doyle, T. (2011). "Helen Nissenbaum, privacy in context: technology, policy, and the integrity of social life".

Dubins, L. E. (1957). "On curves of minimal length with a constraint on average curvature, and with prescribed initial and terminal positions and tangents". *American Journal of mathematics*. 79(3): 497–516.

Ducey, D. (2018). "Executive Oder 2018-04: Advancing Autonomous Vehicle Testing annd Operating; Prioritizing Public Safety".

Falco, G., B. Schneiderman, J. Badger, R. Carrier, A. Dabhora, D. Danks, M. Eling, A. Goodloe, J. Gupta, C. Hart, *et al.* (2021). "Governing AI safety through independent audits". *Nature Machine Intelligence*.

Falco, G. (2020). "Death by AI: Where Assured Autonomy in Smart Cities Meets the End-to-End Argument". *arXiv preprint arXiv:2002.11625*.

Favarò, F. M., N. Nader, S. O. Eurich, M. Tripp, and N. Varadaraju. (2017). "Examining accident reports involving autonomous vehicles in California". *PLoS one*. 12(9): e0184952.

Federal Aviation Administration. (2017). "The FAA and Industry Guide to Product Certification".

Fisher, M., E. Collins, L. Dennis, M. Luckcuck, M. Webster, M. Jump, V. Page, C. Patchett, F. Dinmohammadi, D. Flynn, *et al.* (2018). “Verifiable self-certifying autonomous systems”. In: *2018 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*. IEEE. 341–348.

Fisher, M., V. Mascardi, K. Y. Rozier, B.-H. Schlingloff, M. Winikoff, and N. Yorke-Smith. (2021). “Towards a framework for certification of reliable autonomous systems”. *Autonomous Agents and Multi-Agent Systems*. 35(1): 1–65.

Fremont, D. J., E. Kim, Y. V. Pant, S. A. Seshia, A. Acharya, X. Bruso, P. Wells, S. Lemke, Q. Lu, and S. Mehta. (2020). “Formal scenario-based testing of autonomous vehicles: From simulation to the real world”. In: *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 1–8.

Gamundani, A. M. and L. M. Nekare. (2018). “A review of new trends in cyber attacks: a zoom into distributed database systems”. In: *2018 IST-Africa Week Conference (IST-Africa)*. IEEE. Page–1.

Gardner, R. W., D. Genin, R. McDowell, C. Rouff, A. Saksena, and A. Schmidt. (2016). “Probabilistic model checking of the next-generation airborne collision avoidance system”. In: *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*. IEEE. 1–10.

Gerdes, J. C. and S. M. Thornton. (2015). “Implementable ethics for autonomous vehicles”. In: *Autonomous fahren*. Springer. 87–102.

Glancy, D. J. (2012). “Privacy in autonomous vehicles”. *Santa Clara L. Rev.* 52: 1171.

Goodall, N. J. (2014). “Machine ethics and automated vehicles”. In: *Road vehicle automation*. Springer. 93–102.

Goodloe, A. E. and L. Pike. (2010). *Monitoring distributed real-time systems: A survey and future directions*. National Aeronautics and Space Administration, Langley Research Center.

Griffor, E. R., C. Greer, D. A. Wollman, M. J. Burns, *et al.* (2017). “Framework for cyber-physical systems”. *Tech. rep.* No. Special Publication 1500-201. National Institute of Standards and Technology.

Grilo, E. S. and B. Lopes. (2018). “Formalization and certification of software for smart cities”. In: *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 1–8.

Guiochet, J., M. Machin, and H. Waeselynck. (2017). “Safety-critical advanced robots: A survey”. *Robotics and Autonomous Systems*. 94: 43–52.

Gurevich, Y., E. Hudis, and J. M. Wing. (2016). “Inverse privacy”. *Communications of the ACM*. 59(7): 38–42.

Hatcliff, J., M. Heimdahl, M. Lawford, T. Maibaum, A. Wassyng, and F. Wurden. (2009). “A software certification consortium and its top 9 hurdles”. *Electronic Notes in Theoretical Computer Science*. 238(4): 11–17.

Helle, P., W. Schamai, and C. Strobel. (2016). “Testing of autonomous systems–Challenges and current state-of-the-art”. In: *INCOSE international symposium*. Vol. 26. No. 1. Wiley Online Library. 571–584.

Herbert-Burns, R. (2009). “The Suez Canal: Strategic & Operational Security Realities-Past, Present, & Future”. *Strategic Insights: Global Maritime Analysis*. (19).

Hinchey, M., C. Rouff, J. Rash, and W. F. Truszkowski. (2005). “Requirements of an integrated formal method for intelligent swarms”. In: *Proceedings of the 10th international workshop on Formal Methods for Industrial Critical Systems*. 125–133.

Hinchey, M., J. P. Bowen, and C. A. Rouff. (2006). “Introduction to formal methods”. In: *Agent Technology from a Formal Perspective*. Springer. 25–64.

Hinchey, M. G. and S. A. Jarvis. (1995). *Concurrent systems: formal development in CSP*. McGraw-Hill, Inc.

Hoare, C. A. R. (1978). “Communicating sequential processes”. *Communications of the ACM*. 21(8): 666–677.

Holzmann, G. J. (1997). “The model checker SPIN”. *IEEE Transactions on software engineering*. 23(5): 279–295.

Huang, H.-M. (2007). “Autonomy levels for unmanned systems (ALFUS) framework: safety and application issues”. In: *Proceedings of the 2007 Workshop on Performance Metrics for Intelligent Systems*. 48–53.

Huang, H.-M., E. R. Messina, and J. Alphus. (2007). "Autonomy levels for unmanned systems (ALFUS) framework". *Tech. rep.* No. Special Publication 1011-II-1.0. National Institute of Standards and Technology.

Hutchison, C., M. Zizyte, P. E. Lanigan, D. Guttendorf, M. Wagner, C. Le Goues, and P. Koopman. (2018). "Robustness testing of autonomy software". In: *2018 IEEE/ACM 40th International Conference on Software Engineering: Software Engineering in Practice Track (ICSE-SEIP)*. IEEE. 276–285.

Inslee, J. (2017). "Executive Oder 17-02: Autonomous Vehicle Testing & Technology in Washigngton State and Autonomous Vehicle Work Group".

Jeannin, J.-B., K. Ghorbal, Y. Kouskoulas, R. Gardner, A. Schmidt, E. Zawadzki, and A. Platzer. (2015a). "A formally verified hybrid system for the next-generation airborne collision avoidance system". In: *International Conference on Tools and Algorithms for the Construction and Analysis of Systems*. Springer. 21–36.

Jeannin, J.-B., K. Ghorbal, Y. Kouskoulas, R. Gardner, A. Schmidt, E. Zawadzki, and A. Platzer. (2015b). "Formal verification of ACAS X, an industrial airborne collision avoidance system". In: *2015 International Conference on Embedded Software (EMSOFT)*. IEEE. 127–136.

Kane, A., O. Chowdhury, A. Datta, and P. Koopman. (2015). "A case study on runtime monitoring of an autonomous research vehicle (ARV) system". In: *Runtime Verification*. Springer. 102–117.

Kim, K., J. S. Kim, S. Jeong, J.-H. Park, and H. K. Kim. (2021). "Cybersecurity for autonomous vehicles: Review of attacks and defense". *Computers & Security*: 102150.

Klaassen, K. B. and J. C. Van Peppen. (2006). *System reliability*. VSSD.

Kochenderfer, M. J., J. E. Holland, and J. P. Chryssanthacopoulos. (2012). "Next-generation airborne collision avoidance system". *Lincoln Laboratory Journal*. 19(1).

Koopman, P., A. Kane, and J. Black. (2019). "Credible autonomy safety argumentation". In: *27th Safety-Critical System Symposium Safety-Critical Systems Club, Bristol, UK*.

Koopman, P. and M. Wagner. (2017). "Autonomous vehicle safety: An interdisciplinary challenge". *IEEE Intelligent Transportation Systems Magazine*. 9(1): 90–96.

Kornecki, A. and J. Zalewski. (2008). "Software certification for safety-critical systems: A status report". In: *2008 International Multiconference on Computer Science and Information Technology*. IEEE. 665–672.

Kouskoulas, Y., T. J. Machado, and D. Genin. (2020). "Formally Verified Timing Computation for Non-deterministic Horizontal Turns During Aircraft Collision Avoidance Maneuvers". In: *International Conference on Formal Methods for Industrial Critical Systems*. Springer. 113–129.

Kunze, L., N. Hawes, T. Duckett, M. Hanheide, and T. Krajnik. (2018). "Artificial intelligence for long-term robot autonomy: A survey". *IEEE Robotics and Automation Letters*. 3(4): 4023–4030.

Kwiatkowska, M., G. Norman, and D. Parker. (2007). "Stochastic model checking". In: *International School on Formal Methods for the Design of Computer, Communication and Software Systems*. Springer. 220–270.

Laskov, P. and R. Lippmann. (2010). "Machine learning in adversarial environments".

Lawrence, K. (2021). *When "Just-in-Time" Falls Short: Examining the Effects of the Suez Canal Blockage*. SAGE Publications: SAGE Business Cases Originals.

Leveson, N. (2020). "Are you sure your software will not kill anyone?" *Communications of the ACM*. 63(2): 25–28.

Lin, P. (2016). "Why ethics matters for autonomous cars". In: *Autonomous driving*. Springer, Berlin, Heidelberg. 69–85.

Lin, P., G. Bekey, and K. Abney. (2008). "Autonomous military robotics: Risk, ethics, and design". *Tech. rep.* California Polytechnic State Univ San Luis Obispo.

Lindvall, M., A. Porter, G. Magnusson, and C. Schulze. (2017). "Metamorphic model-based testing of autonomous systems". In: *2017 IEEE/ACM 2nd International Workshop on Metamorphic Testing (MET)*. IEEE. 35–41.

Lipsky, M. S. and L. K. Sharp. (2001). “From idea to market: the drug approval process.” *The Journal of the American Board of Family Practice*. 14(5): 362–367.

Luckcuck, M., M. Farrell, L. A. Dennis, C. Dixon, and M. Fisher. (2019). “Formal specification and verification of autonomous robotic systems: A survey”. *ACM Computing Surveys (CSUR)*. 52(5): 1–41.

Maheswaran, R. T., M. Tambe, P. Varakantham, and K. Myers. (2003). “Adjustable autonomy challenges in personal assistant agents: A position paper”. In: *International Workshop on Computational Autonomy*. Springer. 187–194.

MahmoudZadeh, S., D. M. Powers, and R. B. Zadeh. (2019). “Introduction to autonomy and applications”. In: *Autonomy and Unmanned Vehicles*. Springer. 1–15.

Marchant, G. E. and R. A. Lindor. (2012). “The coming collision between autonomous vehicles and the liability system”. *Santa Clara L. Rev.* 52: 1321.

Martin, J., N. Kim, D. Mittal, and M. Chisholm. (2015). “Certification for autonomous vehicles”. *Automative Cyber-physical Systems course paper, University of North Carolina, Chapel Hill, NC, USA*.

Maslow, A. H. (1943). “A theory of human motivation.” *Psychological review*. 50(4): 370.

Maurya, A. and D. Kumar. (2020). “Reliability of safety-critical systems: A state-of-the-art review”. *Quality and Reliability Engineering International*. 36(7): 2547–2568.

Mayer, R. C., J. H. Davis, and F. D. Schoorman. (1995). “An integrative model of organizational trust”. *Academy of management review*. 20(3): 709–734.

McCarthy, R. L. (2021). “Autonomous Vehicle Accident Data Analysis: California OL 316 Reports: 2015–2020”. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*. 8(3): 034502.

McKnight, D. H. (2005). “Trust in information technology”. *The Blackwell encyclopedia of management*. 7: 329–331.

McKnight, D. H. and N. L. Chervany. (2001). “What trust means in e-commerce customer relationships: An interdisciplinary conceptual typology”. *International journal of electronic commerce*. 6(2): 35–59.

Meshka, L. (2020). “Risk Considerations for Autonomy Software”. In: *2020 Annual Reliability and Maintainability Symposium (RAMS)*. IEEE. 1–6.

Mohanty, J. K., P. Dash, and P. Pradhan. (2020). “FMECA analysis and condition monitoring of critical equipments in super thermal power plant”. *International Journal of System Assurance Engineering and Management*: 1–17.

Mostafa, S. A., M. S. Ahmad, and A. Mustapha. (2019). “Adjustable autonomy: a systematic literature review”. *Artificial Intelligence Review*. 51(2): 149–186.

Mueller, J. M. (2019). “The ABCs of assured autonomy”. In: *2019 IEEE International Symposium on Technology and Society (ISTAS)*. IEEE. 1–5.

Mullins, G. E., P. G. Stankiewicz, R. C. Hawthorne, and S. K. Gupta. (2018). “Adaptive generation of challenging scenarios for testing and evaluation of autonomous vehicles”. *Journal of Systems and Software*. 137: 197–215.

NAC-482A. (2020). “Nevada Administrative Code Chapter 482A - Autonomous Vehicles”. URL: http://www.leg.state.nv.us/NAC/NAC-482A.html.

Nascimento, A. M., L. F. Vismari, C. B. S. T. Molina, P. S. Cugnasca, J. B. Camargo, J. R. de Almeida, R. Inam, E. Fersman, M. V. Marquezini, and A. Y. Hata. (2019). “A systematic literature review about the impact of artificial intelligence on autonomous vehicle safety”. *IEEE Transactions on Intelligent Transportation Systems*. 21(12): 4928–4946.

Nejati, F., A. A. A. Ghani, N. K. Yap, and A. Jaafar. (2017). “Handling state space explosion in verification of component-based systems: A review”. *arXiv preprint arXiv:1709.10379*.

NHTSA. (2016). “Federal Automated Vehicles Policy: Accelerating the next revolution in roadway safety”. National Highway Traffic Safety Administration, Departmenht of Transportation, USA.

Nissenbaum, H. (2020). *Privacy in context*. Stanford University Press.

Nordenström, M. (2020). “Future certification of autonomous vehicles and the use of virtual testing methods”. *MA thesis*. Sweden: KTH Royal Institute of Technology, School of Engineering Sciences.

NSTC and USDOT. (2020). "Ensuring American Leadership in Automated Vehicle Technologies: Automated Vehicles 4.0". National Science & Technology Council and the United States Department of Transportation.

Owen, D., B. Cukic, and T. Menzies. (2002). "An alternative to model checking: Verification by random search of AND-OR graphs representing finite-state models". In: *7th IEEE International Symposium on High Assurance Systems Engineering, 2002. Proceedings.* IEEE. 119–126.

Owen, D., T. Menzies, M. Heimdahl, and J. Gao. (2003). "On the advantages of approximate vs. complete verification: Bigger models, faster, less memory, usually accurate". In: *28th Annual NASA Goddard Software Engineering Workshop, 2003. Proceedings.* IEEE. 75–81.

Phillips, P. J., C. A. Hahn, P. C. Fontana, A. N. Yates, K. Greene, D. A. Broniatowski, and M. A. Przybocki. (2021). "Four Principles of Explainable Artificial Intelligence (Draft)". *Tech. rep.* No. NISTIR 8312.

Poulsen, K. (2010). "Hacker disables more than 100 cars remotely". *Internet]. Available: www. wired. com/threatlevel/2010/03/hacker-bricks-cars.*

Powell, J. D., D. Owens, and T. Menzies. (2004). "SPIN or LURCH: a comparative assessment of model cheching and stochastic search for temporal properties in procedural code."

Pyrgies, J. (2020). "Towards DO-178C certification of adaptive learning UAV agents designed with a cognitive architecture". In: *2020 IEEE/ACM 42nd International Conference on Software Engineering: Companion Proceedings (ICSE-Companion).* IEEE. 174–177.

Radio Technical Commissionn for Aeronautics (RTCA). (2011). "Software Considerations in Airborne Systems and Equipment Certification". *Tech. rep.* No. DO-178C.

Ramos, K. G., I. C. N. Rocha, T. D. D. Cedeño, A. C. dos Santos Costa, S. Ahmad, M. Y. Essar, and C. Tsagkaris. (2021). "Suez Canal blockage and its global impact on healthcare amidst the COVID-19 pandemic". *International Maritime Health.* 72(2): 145–146.

Rouff, C., J. Rash, and M. G. Hinchey. (2000). "Experience using formal methods for specifying a multi-agent system". In: *Proceedings Sixth IEEE International Conference on Engineering of Complex Computer Systems. ICECCS 2000*. IEEE. 72–80.

Rouff, C., A. Vanderbilt, W. F. Truszkowski, J. L. Rash, and M. G. Hinchey. (2004). "Formal Methods for Autonomic and Swarm-based Systems". In: *Proceedings of the International Symposium on Leveraging Applications of Formal Methods, ISoLA 2004*. Paphos, Cyprus.

Rousseau, D. M., S. B. Sitkin, R. S. Burt, and C. Camerer. (1998). "Not so different after all: A cross-discipline view of trust". *Academy of management review*. 23(3): 393–404.

Sadigh, D., K. Driggs-Campbell, A. Puggelli, W. Li, V. Shia, R. Bajcsy, A. Sangiovanni-Vincentelli, S. S. Sastry, and S. Seshia. (2014). "Data-driven probabilistic modeling and verification of human driver behavior". In: *2014 AAAI Spring Symposium Series*.

SAE. (2016). "j3061, cybersecurity guidebook for cyber-physical vehicle systems". *Society for automotive engineers*. 1: 52.

SAE. (2021). "Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems". *Tech. rep.* No. Technical Report J3016 202104. On-Road Automated Vehicle Standards Committee, SAE International.

Saidi, S., D. Ziegenbein, J. V. Deshmukh, and R. Ernst. (2020). "EDA for autonomous behavior assurance". In: *2020 IEEE/ACM International Conference On Computer Aided Design (ICCAD)*. IEEE. 1–3.

Samonas, S. and D. Coss. (2014). "The CIA Strikes Back: Redefining Confidentiality, Integrity and Availability in Security". *Journal of Information System Security*. 10(3).

Scheidt, D., W. D'Amico, and R. Lutz. (2014). "Safe Testing of Autonomy in Complex, Interactive Environments (TACE)". *ITEA J.* 35: 323–331.

Scheidt, D., R. Lutz, W. D'Amico, D. Kleissas, R. Chalmers, and R. Bamberger. (2015). "Safe Testing of Autonomous System Performance". In: *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*.

Schmittner, C., Z. Ma, C. Reyes, O. Dillinger, and P. Puschner. (2016). "Using SAE J3061 for automotive security requirement engineering". In: *International Conference on Computer Safety, Reliability, and Security*. Springer. 157–170.

Seshia, S. A., D. Sadigh, and S. S. Sastry. (2015). "Formal methods for semi-autonomous driving". In: *2015 52nd ACM/EDAC/IEEE Design Automation Conference (DAC)*. IEEE. 1–5.

Sheridan, T. B. and W. L. Verplank. (1978). "Human and computer control of undersea teleoperators". *Tech. rep.* Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.

Siil, K., A. Rubin, M. Elder, A. Dahbura, M. Green, and L. Watkins. (2020). "Mission Assurance for Autonomous Undersea Vehicles". In: *2020 IEEE Security and Privacy Workshops (SPW)*. IEEE. 244–249.

Skrickij, V., E. Šabanovič, and V. Žuraulis. (2020). "Autonomous road vehicles: recent issues and expectations". *IET Intelligent Transport Systems*. 14(6): 471–479.

Smith, B., M. S. Feather, T. Huntsberger, and R. Bocchino. (2020). "Software Assurance of Autonomous Spacecraft Control". In: *2020 Annual Reliability and Maintainability Symposium (RAMS)*. IEEE. 1–7.

Stamp, M. (2011). *Information security: principles and practice*. John Wiley & Sons.

Such, J. M. (2017). "Privacy and Autonomous Systems." In: *IJCAI*. 4761–4767.

Tanzi, T., L. Apvrille, J.-L. Dugelay, and Y. Roudier. (2014). "UAVs for humanitarian missions: Autonomy and reliability". In: *IEEE Global Humanitarian Technology Conference (GHTC 2014)*. IEEE. 271–278.

Topcu, U., N. Bliss, N. Cooke, M. Cummings, A. Llorens, H. Shrobe, and L. Zuck. (2020). "Assured Autonomy: Path Toward Living With Autonomous Systems We Can Trust". *arXiv preprint arXiv:2010.14443*.

Torens, C., F.-M. Adolf, and L. Goormann. (2014). "Certification and software verification considerations for autonomous unmanned aircraft". *Journal of aerospace information systems*. 11(10): 649–664.

Truszkowski, W., H. Hallock, C. Rouff, J. Karlin, J. Rash, M. Hinchey, and R. Sterritt. (2009). *Autonomous and autonomic systems: with applications to NASA intelligent spacecraft operations and exploration systems*. Springer Science & Business Media.

Truszkowski, W., M. Hinchey, J. Rash, and C. Rouff. (2004a). "NASA's swarm missions: The challenge of building autonomous software". *IEEE IT professional*. 6(5): 47–52.

Truszkowski, W., J. Rash, C. Rouff, and M. Hinchey. (2004b). "Some autonomic properties of two legacy multi-agent systems-LOGOS and ACT". In: *11th IEEE International Conference and Workshop on the Engineering of Computer-Based Systems*. IEEE. 490–498. ISBN: 0-7695-2125-8.

Truszkowski, W., C. Rouff, S. Bailin, and M. Riley. (2005). "Progressive autonomy: a method for gradually introducing autonomy into space missions". *Innovations in Systems and Software Engineering*. 1(2): 89–99.

U.S. DOT. (2021). "Automated Vehicles: Comprehensive Plan". United States Department of Transportation.

UK Ministry of Defence. (2017). "Defence Standard 00-56 Issue 7 (Part 1): Safety management requirements for defence systems". Feb.

Van Dijke, J., M. Van Schijndel, F. Nashashibi, and A. de La Fortelle. (2012). "Certification of automated transport systems". *Procedia-Social and Behavioral Sciences*. 48: 3461–3470.

Vinnem, J. E. and I. B. Utne. (2018). "Risk from cyberattacks on autonomous ships". *Safety and Reliability-Safe Societies in a Changing World*.

Wachenfeld, W. and H. Winner. (2016). "The release of autonomous vehicles". In: *Autonomous driving*. Springer. 425–449.

Watkins, L., D. Hamilton, K. Kornegay, and A. Rubin. (2021). "Triaging Autonomous Drone Faults By Simultaneously Assuring Autonomy and Security". In: *2021 55th Annual Conference on Information Sciences and Systems (CISS)*. IEEE. 1–6.

Webster, M., C. Dixon, M. Fisher, M. Salem, J. Saunders, K. L. Koay, and K. Dautenhahn. (2014). "Formal verification of an autonomous personal robotic assistant". In: *2014 AAAI Spring Symposium Series*.

Westin, A. F. (1968). “Privacy and freedom”. *Washington and Lee Law Review*. 25(1): 166.

Wing, J. M. (1990). “A specifier’s introduction to formal methods”. *Computer*. 23(9): 8–22.

Wing, J. M. (2020). “Trustworthy AI”. *arXiv preprint arXiv:2002.06276*.

Wright, R. G. (2020). *Unmanned and Autonomous Ships: An Overview of MASS*. Routledge.

Yan, C., W. Xu, and J. Liu. (2016). “Can you trust autonomous vehicles: Contactless attacks against sensors of self-driving vehicle”. *Def Con*. 24(8): 109.

Zhao, X., V. Robu, D. Flynn, K. Salako, and L. Strigini. (2019). “Assessing the safety and reliability of autonomous vehicles from road testing”. In: *2019 IEEE 30th International Symposium on Software Reliability Engineering (ISSRE)*. IEEE. 13–23.

Zieba, S., P. Polet, F. Vanderhaegen, and S. Debernard. (2010). “Principles of adjustable autonomy: a framework for resilient human–machine cooperation”. *Cognition, Technology & Work*. 12(3): 193–203.