

### **Author Notes**

- ✓ The type has been replaced in your figures; please proof the figures carefully for any errors.
- ✓ Please confirm all artwork in this article is either original or that permission for use has been obtained and credit given where needed.
- ✓ The last page contains pull quotes that might be used in the article, depending on design and spacing considerations. Are these pull quotes appropriate? If not, please provide several alternatives. Pull quotes should be concise and taken verbatim from article text.

# Trustworthy Autonomous Systems Through Verifiability

**Mohammad Reza Mousavi**, King's College London  
**Ana Cavalcanti**, University of York  
**Michael Fisher and Louise Dennis**, University of Manchester  
**Rob Hierons**, University of Sheffield  
**Bilal Kaddouh**, University of Leeds  
**Effie Lai-Chong Law**, University of Durham  
**Rob Richardson**, University of Leeds  
**Jan Oliver Ringert and Ivan Tyukin**, King's College London  
**Jim Woodcock**, University of York

*Autonomous systems have the promise to address many of our societal challenges in a variety of areas. To realize this potential, these systems need to be trustworthy. We describe research carried out by a U.K. consortium to address a central issue in establishing trustworthiness: verifiability.* **<AU: Please note that the abstract has been trimmed in accordance to magazine style. Please check that the included details are correct.>**

**A**utonomous systems can make decisions, and even take actions, independent of human control or intervention. Such systems promise to improve our lives; driverless trains and robotic cleaners are examples of autonomous systems that are already among us and work well within confined environments. To make the most of their potential and gain justified public acceptance, such systems need to be trustworthy in all scenarios. We must now work to ensure developers can design Trustworthy Autonomous Systems (TAS) **<AU: Kindly check that the expansion of TAS is correct.>** for dynamic and open environments and can provide evidence of the trustworthiness of these systems.

The defining feature of autonomous systems is, unsurprisingly, autonomy: their ability to make decisions. This general description allows for a range of levels of autonomy, depending on who or what retains control. Levels recognized across sectors are often based on the Pilot Authority and Control of Tasks (PACT) categorization developed in aerospace<sup>4</sup> or the subsequent Society of Automotive Engineers levels from the automotive sector.<sup>19</sup> Here, levels range from level 1, essentially capturing human control, all the way up to level 5, wherein the system itself makes all the decisions and can take actions.

Although most deployed systems can be categorized at lower PACT levels, with human operators maintaining a significant level of control (and legal responsibility), the potential applications of fully autonomous systems (level 5) can be of enormous socioeconomic benefit. In producing systems with higher levels of autonomy, developers are likely to start

from systems for specific use cases and operational design domains (such as motorway driving for vehicles) and include more use cases gradually as the technology matures and trust is established.

However, it remains a challenge to proceed with fully autonomous systems in many use cases. While this is partly due to the immaturity of technologies or the unknown added value of autonomy in some use cases, we believe it is more fundamentally concerned with a lack of trust in these systems among their users.

In this article, we discuss how we might improve the trustworthiness of autonomous systems and how verifiability can be a central part of this. We describe a new collaborative research activity in the United Kingdom to tackle the complex, heterogeneous challenges of autonomous systems verification as part of their design and deployment. While we mostly focus on this initiative in the United Kingdom, there are a number of similar initiatives, for example, in Australia (Trusted Autonomous Systems: <https://tasdrc.com.au>); Germany (the Center for Pervasive Computing: <https://www.pervasive-computing.science/>); and the United States [the Stanford Center for AI Safety: <http://aisafety.stanford.edu/>; Assured Autonomy (the Computing Community Consortium): <https://cra.org/ccv/visioning/visioning-activities/2019-activities/assured-autonomy/>; the Institute for Assured Autonomy: <https://iaa.jhu.edu/>; and Good Systems: <https://bridgingbarriers.utexas.edu/good-systems/>]. **<AU: Please note that footnotes are not permitted as per magazine style. Footnotes are incorporated into the text. Please check that the placement is okay.>**

## TAS

Trust in a system is defined as the belief or attitude that the system is helpful and beneficial in achieving the user's goal, particularly in uncertain and risky situations.<sup>20</sup> In traditional cyberphysical systems, trustworthiness often equates to reliability. We are more likely to trust some system if it works reliably. Once we move to autonomous systems, which can make their own decisions and take their own actions, more issues come into play. We also want to know that the system's decisions are for our benefit. This aspect, termed *beneficiality* in Chatila et al.,<sup>8</sup> concerns not just what a system does but why it does it. Is it working for our benefit? Is it trying to help, rather than hinder, us? What does it intend? This aspect of beneficiality might quickly become more important than reliability.

## An example

Recall the famous 1984 movie *The Terminator* wherein a robot appears to have few qualms about hurting humans. Our trust in such a robot is drastically reduced by its sinister intent; reliability barely comes into it. Indeed, with such a sinister intent, we would prefer the robot to be unreliable. Only once we can be certain about the beneficial nature of an autonomous robot do we want it to be as reliable as possible.<sup>23</sup>

## An aside

We have also investigated real-world examples of autonomous systems where a high level of trust is not justified due to their nontransparent violation of beneficiality, for example, by polluting the environment more than legally allowed.<sup>3</sup> Although trust is itself subjective, being confident

about both reliability and beneficiality is important. And, as we know from decades of research and practice, confidence in software systems is related to the strength of verification we can carry out on the software. In the example mentioned previously, if we can prove that a robot always works both beneficially and reliably, then we are more likely to trust it. Although there are many other issues at play, the verifiability of these key aspects provides important input into trustworthiness.

In the United Kingdom, a £33 million program of interlinked projects is addressing issues related to TAS. The projects comprise large “nodes” tackling key areas, linked together by a coordination, community-building, and engagement hub (<https://www.tas.ac.uk>). **<AU: Please check that the placement of the footnote information is okay.>** While there are many interesting and important nodes, for example, those concerned with resilience (<https://resilience.tas.ac.uk>) or security (<https://security.tas.ac.uk>), in this article, we focus on the work of the Verifiability Node (<https://verifiability.org>) and how it is tackling the verification of reliability and beneficiality in autonomous systems. **<AU: Please check that the placement of the footnote information is okay. Also, please confirm that the <https://security.tas.ac.uk> URL is correct as it gave a security certificate warning.>**

## HETEROGENEOUS VERIFICATION IS ESSENTIAL

Autonomy is not a binary notion and may be introduced in different levels to various systems and application scenarios. However, a key aspect is how (and why) decisions are made within our systems. This can be very different

across automatic systems, where decisions might be precoded; adaptive systems, where decisions might appear from environmental interactions and feedback; or fully autonomous systems, in which decisions may be made in line with internal aims and goals, taking into account the changing context. For each of these levels of autonomy and the mechanisms of its implementation, different verification techniques may be applicable.

Verification techniques range across formal and empirical and across static and dynamic. These comprise logical specification and verification,<sup>21</sup> dynamic testing, including model-based methods,<sup>2,22</sup> simulation-based testing,<sup>5</sup> runtime verification,<sup>3,12</sup> and stochastic methods.<sup>28</sup> While there are many options, it has become clear that we cannot, and should not, rely on one approach and that a heterogeneous (or corroborative) collection of verification approaches is needed.<sup>13,18,25</sup> This is just what the Verifiability Node aims to provide, together with the semantic foundations to design and justify combinations of these heterogeneous concepts and techniques and with applications that highlight the breadth of verification issues across autonomous systems.

## BRINGING IT ALL TOGETHER

In the Verifiability Node, our vision is to carry out foundational research to enable the possibility of having a verified autonomy store. Autonomous systems and the components for autonomy in such a store go through rigorous and extensive verification upon submission and throughout their evolution. Having passed submission checks, components and systems are made available in a package providing the software;

models for design for compatible platforms and environments; properties; and verification evidence. The store also provides automated facilities for the verification of updates to models (to include new algorithms, platforms, and environments) and components (to cater to adaptive and evolving behaviors and for changed or extended functionality) and for incorporating new verification evidence such as deployment test results. Verification covers components and their variability and evolution; their interoperability; and system-level properties for component compositions. Properties can pertain to reactive, real-time, intentional, adaptive, and uncertain aspects of platforms and environments at all levels of abstraction, from planning and decision making all the way to hardware and physical control. In such a setting, users can have widespread access to trustworthy systems, and developers can have widespread access to affordable and trustworthy components. Such a store will enable reuse and reduces the prohibitively high costs for ad hoc verification.

To achieve this, we need integrated coverage of everything from models of physical components to low-level control algorithms to higher-level software to services and user interactions. A single universal modeling language, verification tool, or technique is not feasible or desirable, yet we must be able to verify different aspects of these systems and how they operate together to enable trust. Our long-term goal is thus to develop a unifying framework that integrates and coalesces rigorous verification techniques of autonomous systems to quickly and easily verify complex autonomous systems.

The activities in the Verifiability Node can be categorized into the three areas described next:

1. *Foundational Aspects*: These give the formal and practical links between the different notations required; the different semantics used; and the different tools and techniques utilized.
2. *Verification Techniques*: These exist across the different aspects and styles of autonomous systems and autonomous components: verifying cyber-physical systems; verifying subsymbolic artificial intelligence (AI) (for instance, deep learning); and verifying symbolic AI layers via both static and dynamic techniques.
3. *Bridging the Gap*: We must bridge the gap between real-world autonomous systems and human-robot interactions, ranging across unmanned aerial vehicles, service robots, chat-bots, human-robot teams, and so forth to deal with the reality gap.

Figure 1 provides an overview of the structure of the Verifiability Node work plan. Work packages 1-3 are concerned with the foundational aspects; work packages 5-7 address the verification techniques; and work packages 4 and 8, as well as the two crosscutting strands, focus on bridging the gap.

Particularly important for collaboration across activities are common case studies in Strand 1 that allow all the different research avenues to coalesce. We have been developing common case studies across the various work packages of the Verifiability Node, for example, in the areas of

disaster management (a firefighting drone, to be extended with connectivity and interaction mechanisms) and assistive care (a dressing robot). Figure 2 depicts an image of our firefighting drone case study. Figure 3 depicts the robotic arm of our assistive dressing case study. In addition to carrying out fundamental research, the Verifiability Node is engaging with various stakeholders in the crosscutting Strand 2 to build a community through the various organized events and the published policy and popular science articles, all advertised on the Verifiability Node website ([https://](https://verifiability.org)

[verifiability.org](https://verifiability.org)). **<AU: Please check that the placement of the footnote information is okay.>**

## VERIFIABILITY NODE: CURRENT STATUS

The Verifiability Node was established in November 2020 and has already achieved several significant results. These include identifying language and notational abstractions across various domains; studying and identifying the basic building blocks of a semantic framework; and defining algorithmic abstractions, refinements, and translations across various

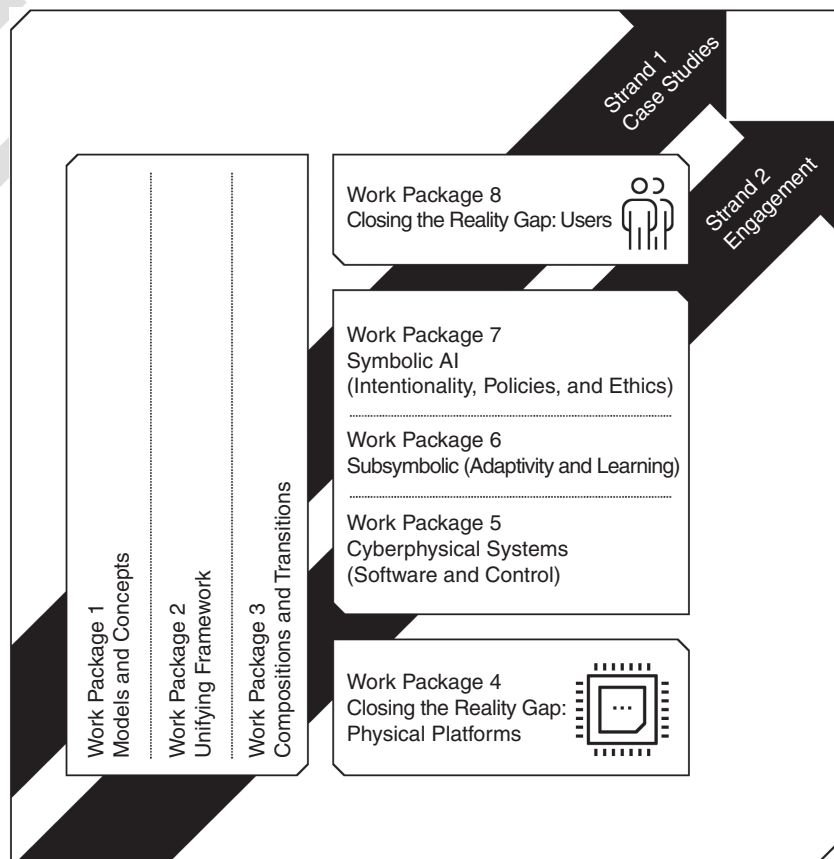


FIGURE 1. A schematic view of the Verifiability Node research program.



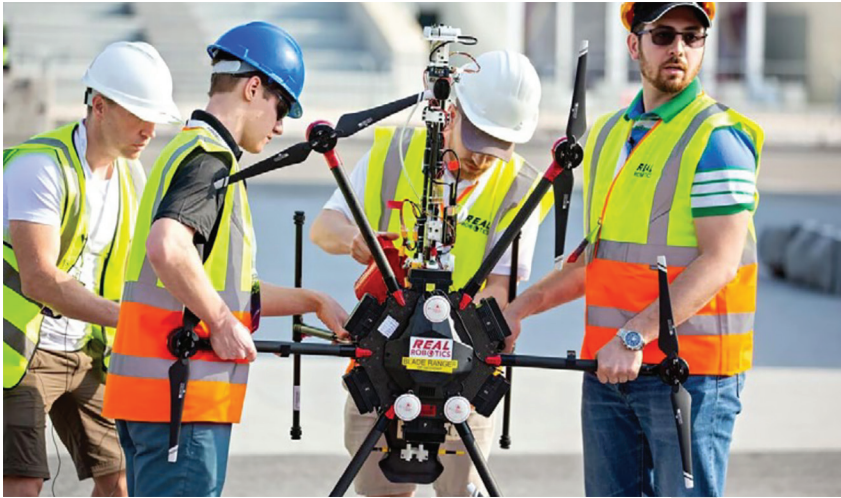


FIGURE 2. The firefighting drone at the Verifiability Node.



FIGURE 3. The assistive dressing robot. (Source: Taken from Chance et al.<sup>7</sup>)

subdomains in the unifying framework. A detailed description of these results can be found in the Verifiability Node Annual Report (<https://verifiability.org/annual-reports/>). **<AU: Please check that the placement of the footnote information is okay.>** We highlight next a few of these results.

- › We designed the first generation of languages to define properties for verification, operational requirements, and mappings between platform-independent and platform models.<sup>6</sup>
- › We formalized heterogeneous semantics, using our Unifying Theories of Programming (UTP)<sup>27</sup> and implementing this in the theorem, proving the framework Isabelle/UTP.<sup>17</sup>
- › We designed a compositional framework for heterogeneous specifications; we took a bottom-up approach by developing a composition of various models for the assistive dressing case study.
- › We accommodated variability in learning and analyzing behavioral models of autonomous systems,<sup>9</sup> using AI (in particular, reinforcement learning) to increase the efficiency of verification strategies.<sup>24</sup>
- › We developed a runtime monitoring algorithm to search for anomalies in the state space of the system<sup>3</sup> as well as a general runtime monitoring framework for autonomous systems.<sup>15,16</sup>
- › We formally verified human-level rules for autonomous systems<sup>1</sup> and ethical concerns in autonomous systems.<sup>10</sup>

## VERIFIABILITY NODE: WHAT NEXT?

The Verifiability Node will continue to work in all the fronts mentioned earlier. For example, the semantics of new notations are being fully formalized and implemented to support automatic generation, and one of our next steps along this line is to mechanize the relevant semantics in Isabelle/UTP. We are also applying these semantic ideas to modeling uncertainty both in case studies and, more widely, in modeling digital twins. In addition, our framework for verifying autonomous decision making,<sup>11</sup> based on verifiable agents, is being developed and expanded to handle the diversity of components.

Within the TAS program, we are collaborating with the Resilience Node on techniques for describing uncertainty in modeling autonomous systems, and we are collaborating with the Security Node on targeting verification to areas identified by security threat analysis. In addition, we aim to expand our collaboration further across other aspects of the program. We will be using formal modeling and verification tools in modeling human behavior and interaction patterns.


More widely, we are keen to collaborate with teams, across academia, industry, and policy, interested in working on common themes. There are existing and emerging standards, such as the ANSI/UL 4600 Standard for Safety for the Evaluation of Autonomous Products; IEEE P7001 Standard for Transparency of Autonomous Systems;<sup>26</sup> and IEEE P7009 Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems.<sup>14</sup> The Verifiability Node has been involved in the design of the latter two standards and is currently engaging in a

number of other standardization initiatives. Details of how to get involved can again be found at the Verifiability Node website (<https://verifiability.org>).

Issues around trust in technology are not new. Throughout the ages, we have had to find ways to learn to trust new tools that can benefit us. However, the issue of the trustworthiness of autonomous systems brings new challenges. As autonomous systems essentially make their own decisions, independent of us, then our trust in these systems is not solely related to their reliability but to whether they will make the right decisions, even in complex and unpredictable situations. Verifiability has a key role not only in assessing reliability but also in establishing beneficiality: that systems will make decisions beneficial to us.

In this article, we described a large, multidisciplinary project focusing on the issue of trustworthiness in autonomous systems, identifying both its challenges and the results obtained so far. The vision of this Verifiability Node is to enhance trustworthiness through a unifying verification framework allowing for heterogeneous models, techniques, and views to be analyzed in tandem. This leads to holistic and wide-reaching verdicts. Our vision is that such a unified and holistic approach to verifiability will fundamentally change our approach to the verification of autonomous systems and will lead to systems that are by their construction worthy of our trust.

Our framework supports the inherent heterogeneity of autonomous systems and allows domain experts to specify their concerns in

domain-specific models. The Verifiability Node framework takes care of connecting these models and providing holistic verification results, which are also projected back to the respective domains. Distinctive in our long-term vision is the integrated coverage of everything from models of physical components to low-level control algorithms to higher-level software to services and user interactions. To realize this vision, we closely collaborate with some of the other ongoing initiatives around the world (listed earlier) as well as with policymaking and standardization bodies. 

## ACKNOWLEDGMENTS

The work reported here is funded by the UK Research and Innovation <AU: Kindly check that the expansion of UKRI is correct.> TAS Verifiability Node EP/V026801/2, Royal Academy of Engineering, and U.K. Engineering and Physical Sciences Research Council. <AU: Kindly check that the expansion of EPSRC is correct.>

## REFERENCES

1. G. V. Alves, L. A. Dennis, and M. Fisher, "A double-level model checking approach for an agent-based autonomous vehicle and road junction regulations," *J. Sensor Actuator Netw.*, vol. 10, no. 3, p. 41, 2021, doi: 10.3390/jsan10030041.
2. H. L. S. Araujo, T. Hoenselaar, M. R. Mousavi, and A. V. Vinel, "Connected automated driving: A model-based approach to the analysis of basic awareness services," in *Proc. 31st Int. Symp. Personal, Indoor Mobile Radio Commun. (PIMRC)*, 2020, pp. 1-7, doi: 10.1109/PIMRC48278.2020.9217142.
3. S. Biewer *et al.*, "Conformance relations and hyperproperties for doping detection in time and space," *Logical*

## ABOUT THE AUTHORS

**MOHAMMAD REZA MOUSAVI** is a professor of software engineering at King's College London, London, United Kingdom. **<AU: Please provide the postal code.>** His main research interests include model-based testing, particularly applied to software product lines and cyberphysical systems. He has been leading several research initiatives and industrial collaboration projects on health care and automotive systems and their validation, verification, and certification. **<AU: Please provide the author's highest academic degree/field of study and institution. Also, please provide the author's IEEE Member status if applicable.>** Contact him at mohammad.mousavi@kcl.ac.uk.

**ANA CAVALCANTI** is the Royal Academy of Engineering chair in emerging technologies at the University of York, York, YO10 5DD, United Kingdom. Her main research interests include software engineering for robotics: modeling, validation, simulation, testing, and verification. She currently leads the RoboStar Centre Of Excellence in this area and is the chair of the board of the Formal Methods Europe association. **<AU: Please provide the author's highest academic degree/field of study and institution. Also, please provide the author's IEEE Member status if applicable.>** Contact her at ana.cavalcanti@york.ac.uk.

**MICHAEL FISHER** is the Royal Academy of Engineering chair in Emerging Technologies at the University of Manchester, Manchester, United Kingdom. **<AU: Please provide the postal code.>** His main research interests include the verification, responsibility, trustworthiness, and safety of autonomous robotic systems. **<AU: Please provide the author's highest academic degree/field of study and institution. Also, please provide the author's IEEE Member status if applicable.>** He cochairs the IEEE Technical Committee for Verification of Autonomous Systems (<https://www.ieee-ras.org/verification-of-autonomous-systems>). Contact him at michael.fisher@manchester.ac.uk.

**LOUISE DENNIS** is a senior lecturer at the University of Manchester, Manchester, United Kingdom, **<AU: Please provide the postal code.>** where she leads the Autonomy and Verification Group. Her main research interests include rational agent programming languages and architectures for autonomous systems, particularly ethical machine reasoning, explainability, and creating verifiable systems. **<AU: Please provide the author's highest academic degree/field of study and institution. Also, please provide the author's IEEE Member status if applicable.>** Contact her at louise.dennis@manchester.ac.uk.

**ROB HIERONS** is a professor of testing at the University of Sheffield, Sheffield, United Kingdom. **<AU: Please provide**

**the postal code.>** His main research interests include the automated generation of efficient, systematic test suites on the basis of program code, models, or specifications. **<AU: Please provide the author's highest academic degree/field of study and institution. Also, please provide the author's IEEE Member status if applicable.>** He is the joint editor of the Journal of Software Testing, Verification, and Reliability and a member of the editorial board of The Computer Journal. Contact him at r.hierons@sheffield.ac.uk.

**BILAL KADDOUH** is a lecturer in aerial robotics at the University of Leeds, Leeds, United Kingdom. **<AU: Please provide the postal code.>** His main research interests include robotics and unmanned systems design; control; and multi-robot mission management. He works on the development of unmanned aerial systems technologies for infrastructure inspection and maintenance; beyond the visual line of sight operations; **<AU: Kindly check that the expansions of UAS and BVLOS are correct.>** precision agriculture; and atmospheric studies. **<AU: Please provide the author's highest academic degree/field of study and institution. Also, please provide the author's IEEE Member status if applicable.>** Contact him at b.kaddouh@leeds.ac.uk.

**EFFIE LAI-CHONG LAW** is a professor of computer science, specializing in human-computer interaction, at the University of Durham, Durham, United Kingdom. **<AU: Please provide the postal code.>** Her main research interests include usability and user experience methodologies; multisensory emotion recognition; conversational artificial intelligence (chatbots); and mixed reality. **<AU: Please provide the author's highest academic degree/field of study and institution. Also, please provide the author's IEEE Member status if applicable.>** She has authored more than 200 peer-reviewed articles and played a leading role in a number of research projects on technology-enhanced learning, health, and well-being. Contact her at lai-chong.law@durham.ac.uk.

**ROB RICHARDSON** is a professor of robotics at the University of Leeds, Leeds, United Kingdom. **<AU: Please provide the postal code.>** His main research interests include a broad range of applied robotics including robotics for civil infrastructure inspection and repair; making smart robots; and robotics for 3D printing applications. He has key roles in many large-scale research projects, including Pipebots; Trustworthy Autonomous Systems; and the Advanced Machinery & Productivity Institute (AMPI). His robotic platforms operate in the air, on the ground, in the water, and underground. **<AU: Please provide the author's highest academic degree/field of study and institution. Also, please provide the author's**



**IEEE Member status if applicable.**> Contact him at r.c.richardson@leeds.ac.uk.

**JAN OLIVER RINGERT** is a lecturer in software engineering at King's College London, London, United Kingdom. **<AU: Please provide the postal code.>** His main research interests include model-based software engineering with a focus on applying formal analyses for the verification and synthesis of reactive systems and software system evolution. **<AU: Please provide the author's highest academic degree/field of study and institution. Also, please provide the author's IEEE Member status if applicable.>** Contact him at jan\_oliver.ringert@kcl.ac.uk.

**IVAN TYUKIN** is a professor of mathematical data science and modeling at King's College London, London, United Kingdom. **<AU: Please provide the postal code.>** His main research interests include the mathematical foundations of artificial intelligence (AI) and learning systems; mathematical modeling; adaptive systems; inverse problems with non-convex and nonlinear parameterization; data analytics; and computer vision. **<AU: Please provide the author's highest**

**academic degree/field of study and institution. Also, please provide the author's IEEE Member status if applicable.>** He is a UK Research and Innovation Turing AI Fellow pursuing a research program to develop adaptive, robust, resilient, certifiable, and trustworthy AI systems. Contact him at ivan.tyukin@kcl.ac.uk.

**JIM WOODCOCK** is a professor of software engineering at the University of York, York, YO10 5DD, United Kingdom; professor and distinguished researcher at Aarhus University, Denmark; and professor at Southwest University, China. **<AU: Please provide the cities and postal codes.>** His main research interests include the theory and practice of formal methods for software engineering. **<AU: Please provide the author's highest academic degree/field of study and institution. Also, please provide the author's IEEE Member status if applicable.>** He is a Fellow of the U.K. Royal Academy of Engineering and the editor in chief of the Association for Computing Machinery journal *Formal Aspects of Computing*. **<AU: Please supply email address.>**

- Methods Comput. Sci.*, vol. 18, no. 1, pp. 14:1–14:39, 2022, doi: 10.46298/lmcs-18(1:14)2022.
- M. C. Bonner, R. M. Taylor, and C. A. Miller, "Tasking interface manager: Affording pilot control of adaptive automation and aiding," in *Contemporary Ergonomics 2000*, S. Robertson, M. Hanson, and P. T. McCabe, Eds. London, U.K.: CRC Press, 2004, pp. 70–74.
  - A. Cavalcanti *et al.*, "Verified simulation for robotics," *Sci. Comput. Program.*, vol. 174, pp. 1–37, Apr. 2019, doi: 10.1016/j.scico.2019.01.004.
  - A. L. C. Cavalcanti, J. Baxter, and G. Carvalho, "RoboWorld: Where can my robot work?" in *Proc. Softw. Eng. Formal Methods (SEFM)*, Springer-Verlag, 2021, pp. 3–22, doi: 10.1007/978-3-030-92124-8\_1.
  - G. Chance, A. Jevtić, P. Caleb-Solly, and S. Dogramadzi, "A quantitative analysis of dressing dynamics for robotic dressing assistance," *Frontiers Robot. AI*, vol. 4, p. 13, May 2017, doi: 10.3389/frobt.2017.00013.
  - R. Chatila *et al.*, "Trustworthy AI," in *Reflections on Artificial Intelligence for Humanity*, B. Braunschweig and M. Ghallab, Eds. Cham: Springer-Verlag, 2021. **<AU: Please provide the page range.>**
  - C. D. N. Damasceno, M. R. Mousavi, and A. d S. Simão, "Learning by sampling: Learning behavioral family models from software product lines," *Empirical Softw. Eng.*, vol. 26, no. 1, pp. 1–46, 2021, doi: 10.1007/s10664-020-09912-w.
  - L. A. Dennis, M. M. Bentzen, F. Lindner, and M. Fisher, "Verifiable machine ethics in changing contexts," in *Proc. 35th Conf. Artif. Intell. (AAAI)*, AAAI Press, 2021, pp. 11,470–11,478.
  - L. A. Dennis and M. Fisher, *Verifiable Autonomous Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2022.
  - Y. Falcone, S. Krstic, G. Reger, and D. Traytel, "A taxonomy for classifying runtime verification tools," *Int. J. Softw. Tools Technol. Transfer*, vol. 23, no. 2, pp. 255–284, 2021, doi: 10.1007/s10009-021-00609-z.
  - M. Farrell, M. Luckcuck, and M. Fisher, "Robotics and integrated formal methods: Necessity meets opportunity," in *Proc. 14th IFM Conf.*, Springer-Verlag, 2018, vol. 11023, pp. 161–171, doi: 10.1007/978-3-319-98938-9\_10.

14. M. Farrell *et al.*, "Evolution of the IEEE P7009 standard: Towards fail-safe design of autonomous systems," in *Proc. IEEE Int. Symp. Softw. Reliability Eng. Workshops (ISSREW)*, 2021, pp. 401–406, doi: 10.1109/ISSREW53611.2021.00109.
15. A. Ferrando, L. A. Dennis, R. C. Cardoso, M. Fisher, D. Ancona, and V. Mascardi, "Toward a holistic approach to verification and validation of autonomous cognitive systems," *ACM Trans. Softw. Eng. Methodol.*, vol. 30, no. 4, pp. 1–43, 2021, doi: 10.1145/3447246.
16. M. Fisher, A. Ferrando, and R. C. Cardoso, "Increasing confidence in autonomous systems," in *Proc. 5th ACM Int. Workshop Verification mOnitoring Runtime EXecution (VORTEX)*, ACM, 2021, pp. 1–4, doi: 10.1145/3464974.3468452.
17. S. Foster, F. Zeyda, and J. Woodcock, "Unifying heterogeneous state-spaces with lenses," in *Proc. 13th Int. Colloquium Theoretical Aspects Comput. (ICTAC)*, 2016, vol. 9965, pp. 295–314, doi: 10.1007/978-3-319-46750-4\_17.
18. M. Gleirscher, S. Foster, and J. Woodcock, "New opportunities for integrated formal methods," *ACM Comput. Surv.*, vol. 52, no. 6, pp. 1–36, 2020, doi: 10.1145/3357231.
19. "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," SAE International, Warrendale, PA, USA, Tech. Rep. J3016\_202104, 2021.
20. J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Hum. Factors*, vol. 46, no. 1, pp. 50–80, 2004, doi: 10.1518/hfes.46.1.50\_30392.
21. M. Luckcuck, M. Farrell, L. A. Dennis, C. Dixon, and M. Fisher, "Formal specification and verification of autonomous robotic systems: A survey," *ACM Comput. Surv.*, vol. 52, no. 5, pp. 1–41, 2019, doi: 10.1145/3342355.
22. A. Miyazawa, P. Ribeiro, W. Li, A. Cavalcanti, J. Timmis, and J. Woodcock, "RoboChart: Modelling and verification of the functional behaviour of robotic applications," *Softw. Syst. Model.*, vol. 18, no. 5, pp. 3097–3149, 2019, doi: 10.1007/s10270-018-00710-z.
23. M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, "Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust," in *Proc. ACM/IEEE Int. Conf. 10th HRI*, ACM, 2015, pp. 1–8.
24. U. C. Türker, R. M. Hierons, M. R. Mousavi, and I. Y. Tyukin, "Efficient state synchronisation in model-based testing through reinforcement learning," in *Proc. 36th IEEE/ACM Int. Conf. Autom. Softw. Eng. (ASE)*, 2021, pp. 368–380, doi: 10.1109/ASE51524.2021.9678566.
25. M. P. Webster *et al.*, "A corroborative approach to verification and validation of human-robot teams," *Int. J. Robot. Res.*, vol. 39, no. 1, pp. 73–99, 2020, doi: 10.1177/0278364919883338.
26. A. F. T. Winfield *et al.*, "IEEE P7001: A new standard on transparency," *Frontiers Robot. AI*, vol. 8, p. 665729, Jul. 2021, doi: 10.3389/frobt.2021.665729.
27. J. Woodcock and A. Cavalcanti, "A tutorial introduction to designs in unifying theories of programming," in *Proc. 4th Int. Conf. Integr. Formal Methods (IFM)*, Springer-Verlag, 2004, vol. 2999, pp. 40–66, doi: 10.1007/978-3-540-24756-2\_4.
28. J. M. Zhang, M. Harman, L. Ma, and Y. Liu, "Machine learning testing: Survey, landscapes and horizons," *IEEE Trans. Softw. Eng.*, vol. 48, no. 1, pp. 1–36, 2022, doi: 10.1109/TSE.2019.2962027.

**IN THIS ARTICLE, WE DISCUSS HOW WE MIGHT IMPROVE THE TRUSTWORTHINESS OF AUTONOMOUS SYSTEMS AND HOW VERIFIABILITY CAN BE A CENTRAL PART OF THIS.**

**ALTHOUGH TRUST IS ITSELF SUBJECTIVE, BEING CONFIDENT ABOUT BOTH RELIABILITY AND BENEFICIALITY IS IMPORTANT.**

**IN THE UNITED KINGDOM, A £33 MILLION PROGRAM OF INTERLINKED PROJECTS IS ADDRESSING ISSUES RELATED TO TAS.**

**AUTONOMY IS NOT A BINARY NOTION AND MAY BE INTRODUCED IN DIFFERENT LEVELS TO VARIOUS SYSTEMS AND APPLICATION SCENARIOS.**

XXXXXX

**MORE WIDELY, WE ARE KEEN TO COLLABORATE WITH TEAMS, ACROSS ACADEMIA, INDUSTRY, AND POLICY, INTERESTED IN WORKING ON COMMON THEMES.**

**VERIFIABILITY HAS A KEY ROLE NOT ONLY IN ASSESSING RELIABILITY BUT ALSO IN ESTABLISHING BENEFICIALITY: THAT SYSTEMS WILL MAKE DECISIONS BENEFICIAL TO US.**