Trustworthy Autonomous Systems (TAS): The Verifiability Approach

Mohammad Reza Mousavi
Mohammad.mousavi@kcl.ac.uk
https://orcid.org/0000-0002-4869-6794
King's College London, London, UK

Recibido: 11 de septiembre del 2022 / Aceptado: 6 de agosto del 2022
doi: https://doi.org/10.26439/ciis2022.6063

ABSTRACT. Autonomous systems are taking over the decision-making in many crucial aspects of our lives. Trust in them will help users benefit from such systems without harming themselves. Establishing the right level of trust involves a holistic validation and verification process, accounting for aspects such as interactions with the physical world and human users. In this talk, I present our ongoing effort to provide a holistic framework for ensuring the verifiability of autonomous systems.

KEYWORDS: autonomous systems, trust, verifiability, validation and verification, testing

SISTEMAS AUTÓNOMOS CONFIABLES (TAS): EL ENFOQUE DE LA VERIFICABILIDAD

RESUMEN. Los sistemas autónomos se están haciendo cargo de la toma de decisiones en muchos aspectos cruciales de nuestras vidas. Confiar en ellos ayudará a sus usuarios a beneficiarse de dichos sistemas sin dañarse a sí mismos. Establecer el nivel adecuado de confianza implica un proceso holístico de validación y verificación, que tiene en cuenta aspectos como las interacciones con el mundo físico y los usuarios humanos. En esta charla, presento nuestro esfuerzo continuo para proporcionar un marco holístico para garantizar la verificabilidad de los sistemas autónomos.

PALABRAS CLAVE: sistemas autónomos, confianza, verificabilidad, validación y verificación, *testing*

1. INTRODUCTION

Autonomous systems are the result of an integration of software, hardware, and communication systems that enables decision-making with minimal intervention required from their users (Mousavi et al., 2022). Examples of such systems include pacemakers and implantable defibrillators, drones and unmanned aerial vehicles (UAVs), and chatbots. Although decision making in such systems is performed autonomously, they often engage in patterns of interactions with users and hence, their usefulness crucially depends on a smooth orchestration of these interactions.

Trust and trustworthiness are crucial aspects in the development and deployment of autonomous systems: it concerns the users' belief that the system is going to be helpful and safe in challenging scenarios (Araujo et al., 2019). Trusting a system that is not trustworthy can harm users, since users will adapt the systems in challenging scenarios that the system cannot cope with. Likewise, not trusting a system that is trustworthy can lead to avoiding the system in scenarios that the system can cope with and hence, not benefitting from the system. Establishing the right level of trust involves gathering and communicating sufficient evidence for the system's safety and usefulness. A holistic validation and verification process is an essential ingredient for providing such evidence (Mousavi et al., 2022; Araujo et al., 2022).

In this talk, I will go through our verifiability framework for autonomous systems. This involves learning about system and user behavior and capturing their appropriate models (Damasceno et al., 2021); adapting the models by observing and adapting to changes (Damasceno et al., 2019; Tavassoli et al., 2022); generating structured test suites that cover different aspects of system and user behavior (Araujo et al., 2020; Biewer et al., 2022); and an analysis of the test results and explaining the patterns of interaction (Gou et al., 2022).

For each of the above-mentioned four steps, we review our latest results, and point out the challenges before us in establishing a holistic verification framework for autonomous systems.

REFERENCES

Araujo, H. L. S., Damasceno, C. D. N., Dimitrova, R., Kefalidou, G., Mehtarizadeh, M., Mousavi, M. R., Onime, J., Ringert, J. O., Rojas, J. M., Verdezoto, N. X., & Wali, S. (2019, October 21-23). *Trusted autonomous vehicles: An interactive exhibit* [Conference presentation]. 2019 IEEE International Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS), Shenyang, China. https://doi.org/10.1109/IUCC/DSCI/SmartCNS.2019.00091

Araujo, H., Hoenselaar, T., Mousavi, M. R., & Vinel, A. (2020, August 31-September 3). *Connected automated driving: A model-based approach to the analysis of basic awareness*

services [Conference presentation]. 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, London, United Kingdom. https://doi.org/10.1109/PIMRC48278.2020.9217142

Araujo, H., Mousavi, M. R., & Varshosaz, M. (2022). Testing, validation, and verification of robotic and autonomous systems: A systematic review. *ACM Transactions on Software Engineering and Methodology*. https://doi.org/10.1145/3542945

Biewer, S., Dimitrova, R., Fries, M., Gazda, M., Heinze, T., Hermanns, H., & Mousavi, M. R. (2022). Conformance relations and hyperproperties for doping detection in time and space. *Logical Methods in Computer Science*, *18*(1). https://doi.org/10.46298/lmcs-18(1:14)2022

Damasceno, C. D. N., Mousavi, M. R., & da Silva Simao, A. (2019, December 2-6). *Learning to reuse: Adaptive model learning for evolving systems* [Conference presentation]. 15th International Conference, IFM 2019, Bergen, Norway. https://doi.org/10.1007/978-3-030-34968-4_8

Damasceno, C. D. N., Mousavi, M. R., & Simao, A. da S. (2021). Learning by sampling: Learning behavioral family models from software product lines. *Empirical Software Engineering*, *26*, 4. https://doi.org/10.1007/s10664-020-09912-w

Gou, M. S., Lakatos, G., Holthaus, P., Wood, L., Mousavi, M. R., Robins, B., & Amirabdollahian, F. (2022, August 29-September 2). *Towards understanding causality – a retrospective study of using explanations in interactions between a humanoid robot and autistic children* [Conference presentation]. 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Naples, Italy. https://doi.org/10.1109/RO-MAN53752.2022.9900660

Mousavi M. R., Cavalcanti A., Fisher M., Dennis L., Hierons R., Kaddouh B., Law E. L., Richardson R., Ringert J. O., Tyukin I., & Woodcock J (2022). *Trustworthy autonomous systems through verifiability*. IEEE Software.

Tavassoli, S., Damasceno, C. D. N., Khosravi, R., & Mousavi, M. R. (2022, September 12-16). *Adaptive behavioral model learning for software product lines* [Conference presentation]. Proceedings of the 26th ACM International Systems and Software Product Line Conference Graz, Austria. https://doi.org/10.1145/3546932.3546991