Risk-driven Compliance Assurance for Collaborative AI Systems: a Vision Paper

Preprint · February 2021

CITATIONS
0

READS
520

7 authors, including:

Matteo Camilli
Politecnico di Milano
67 PUBLICATIONS **439** CITATIONS
SEE PROFILE

Michael Felderer
German Aerospace Center (DLR)
360 PUBLICATIONS **5,439** CITATIONS
SEE PROFILE

Andrea Giusti
Fraunhofer Italia Research
50 PUBLICATIONS **532** CITATIONS
SEE PROFILE

Anna Perini
Fondazione Bruno Kessler
254 PUBLICATIONS **6,767** CITATIONS
SEE PROFILE

Risk-driven Compliance Assurance for Collaborative AI Systems: a Vision Paper*

Matteo Camilli[1], Michael Felderer[2], Andrea Giusti[3], Dominik Tobias Matt[1,3], Anna Perini[4], Barbara Russo[1], Angelo Susi[4]

[1] Free University of Bozen-Bolzano, Bolzano, Italy
{mcamilli,dmatt,brusso}@unibz.it
[2] University of Innsbruck, Innsbruck, Austria
michael.felderer@uibk.ac.at
[3] Fraunhofer Italia Research, Bolzano, Italy
andrea.giusti@fraunhofer.it
[4] Fondazione Bruno Kessler (FBK), Trento, Italy
{perini,susi}@fbk.eu

Abstract. *Context and motivation.* Collaborative AI systems aim at working together with humans in a shared space. Building these systems, which comply with quality requirements, domain specific standards and regulations is a challenging research direction. This challenge is even more exacerbated for new generation of systems that leverage on machine learning components rather than deductive (top-down programmed) AI. *Question/problem.* How can requirements engineering, together with software and systems engineering, contribute towards the objective of building flexible and compliant collaborative AI with strong assurances? *Principal idea/results.* In this paper, we identify three main research directions: automated specification and management of compliance requirements, and their alignment with assurance cases; risk management; and risk-driven assurance methods. Each one tackles challenges that currently hinder engineering processes in this context. *Contributions.* This vision paper aims at fostering further discussion on the challenges and research directions towards appropriate methods and tools to engineer collaborative AI systems in compliance with existing standards, norms, and regulations.

Keywords: Compliance requirements · Compliance cases · Collaborative AI systems · Machine Learning · Risk management.

1 Introduction

Collaborative AI systems (CAIS) are robotic systems that work with humans in a shared physical space to achieve common goals. To achieve flexibility and accommodate changing needs, the upcoming generation of CAIS heavily rely on Machine Learning (ML) components to mimic human perception skills (e.g., visual perception, speech recognition, or conversing in natural language) as well as

* Pre-print. Accepted for publication at the 27th International Working Conference on Requirement Engineering: Foundation for Software Quality (REFSQ 2021).

learn from humans how to carry out specific tasks by demonstration. Thus, ML-equipped CAIS yield bidirectional human-robot collaboration. For this reason, they must satisfy quality criteria including appropriate behavior with respect to social rules, domain specific standards and laws for certification. Furthermore, such systems often run in dynamic and uncertain environments that make it difficult providing strong assurances of *compliance* [4].

In this paper, we reflect on how research in Requirements Engineering (RE) of software and systems can contribute to define suitable methods for building ML-equipped CAIS (henceforth referred again to as CAIS for the sake of simplicity). We believe that a RE perspective can help reasoning on the trade-off between opportunities and risks deriving from the usage of ML-based solutions in this context. Moreover, existing practices for trust-based human-robot interactions [12] shall be evaluated and eventually revisited through novel RE methods to deal with the assurance of learning agents, in which interactions are driven by ML components. In this setting, we envision risk as a first class concern and propose three research directions to investigate over risk-driven engineering processes that leverage on continuous feedback from empirical evidence collected at run-time, in a closed-loop setting with the surrounding environment, in order to verify semantically meaningful properties through suitable assurance methods.

The RE community recognizes that appropriate assurance methods for compliance requirements (e.g., defined on human-robot collaboration standards, such as the ISO/TS 15066[5]) need to be defined. More generally, research on RE for AI-based system is considered a relevant and timely topic in recent RE conferences[6] and in dedicated workshops, such as RE4AI at REFSQ[7], as well as in European initiatives (e.g., AI4EU platform[8]). Evidence is also provided by recent surveys reporting the urgent need for effective RE processes [14] and verification methods [11] for "intelligent" components as well as AI systems. An analysis of the RE characteristics for systems that include ML components are reported in [14] whereas the work in [2] discusses challenges and desiderata for increasing their level of assurance. This latter work emphasizes that existing assurance methods for AI systems in general and CAIS in particular are not linked to compliance requirements and possible risks.

The rest of the paper is organized as follows. In Section 2, we introduce an illustrative example of CAIS. We discuss the key challenges in Section 3. Then, we elaborate on our envisioned approach and research directions in Section 4. Finally, Section 5 concludes the paper.

2 Illustrative Example

We introduce our vision of the problem by means of a case from the Industry 4.0 domain taken from [9] that exemplifies the high risks for human safety as well as

[5] https://www.iso.org/standard/62996.html
[6] https://requirements-engineering.org/
[7] https://sites.google.com/view/re4ai/home
[8] http://ai4eu.eu

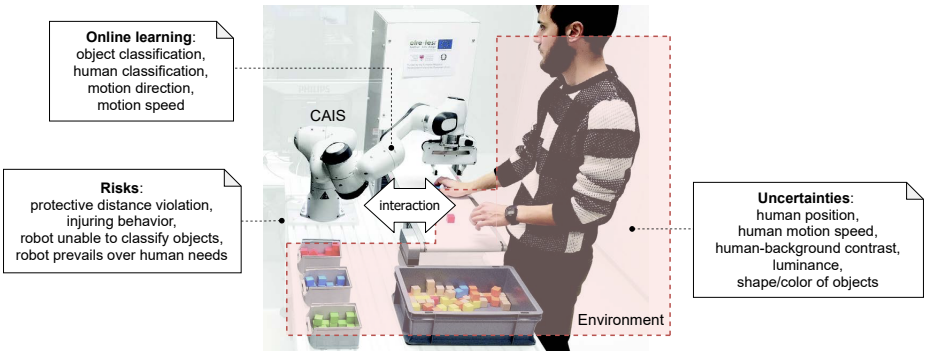

Fig. 1: Illustrative CAIS example in a closed-loop with the surroundings.

relevant dependability concerns. Figure 1 illustrates such example, where an automated controller of a robotic arm attempts to detect and classify objects (e.g., by color and shape) on a conveyor belt and actuates the proper movements to pick and move the object into the right bucket. The system includes a controller, an actuated mechanical system (i.e., robotic arm), and a camera sensor along with a visual perception ML component for classification. This ML component learns on structured heterogeneous data sources associated with *features* (e.g., shape and color of an object) and yields category labels as output. The training and the validation is performed iteratively online by a human operator. The operator collaborates with the robot in order to validate its own actions through gestures.

The operator collaborates with the robot in order to supervise the correct transfer of the desired sorting skill to the robot and can intervene through gestures when corrections are required. The safety control approach of this example is implemented by considering the *speed and separation monitoring* of the standard ISO/TS 15066 for collaborative robotics. Here, a protective separation distance between the human and the robot is checked online using *safety zone*s. The dimension of such zones is dynamically adapted based on the robot motion. Fast motions of the robot can generate safety zones which may be large and therefore negatively affect the realization of collaborative operations. Assuring a safe and successful collaboration in cases in which the robot motions are learnt from humans yields challenges discussed in the next section.

3 Research Challenges

By collaborating with researchers in CAIS engineering, we started to elicit major challenges that are not fully addressed by existing approaches. In this section, we discuss them in light of the key characteristics of CAIS, by focusing on the objective of providing comprehensive, ideally provable, evidence that CAIS exhibit dependable behavior within their viability (due to continuous learning).

CH1 ***Uncertain environment***: The environment in which CAIS operate is often complex with substantial amount of uncertainty even in scenarios where

interacting agents (both robots and humans) are known. For instance, the human operator in the running example (Section 1) is key part of the environment. Thus, assurance methods for CAIS shall deal with inherent variability and uncertainty in human behavior. As an example, the human operator may unpredictably move in a forbidden area and the robotic arm must be able to react in a safe way by enforcing the speed and separation monitoring. In addition to humans, other environment variables can influence the perception capability of the ML components, as shown in Figure 1. For instance, low luminance might lead to decrease the ability of classifying the human operator in specific locations of the shared space. In this case, assurance could leverage on probabilistic approaches by assuming specific distributions of the environment factors. Nonetheless, underlying distributions are often only estimates and do not represent precisely the environment behavior [5].

CH2 ***Adequacy of standards***: Existing standards in the domain of CAIS pose challenges in realizing flexible automation [8]. Difficulties arise from frequently changing production environments and potentially unknown *a-priori* robot motions. These issues become particularly severe when robotic systems swiftly adapt to different task demands by learning from humans using ML components [3]. In fact, existing standards do not specifically refer to CAIS able to learn from humans (e.g., through ML components). For instance, in programming by demonstration [8], fast human motions can induce in turn fast robot motions. Thus, to enforce speed and separation monitoring, the dimension of the safety zones might become large and therefore negatively affect the realization of a successful collaboration. Operational phases can suggest feedback to existing norms and standards that are currently not aligned with practical needs of humans in the context of CAIS. Without clear regulation on what the ML component shall (or shall not) learn, we could put in production CAIS that eventually break trust and prevail over human needs.

CH3 ***Partial and evolving specifications***: Strong compliance assurances usually rely on precise, rigorous, or even formal description of the behavior of the target system. This is anything but trivial in the human-robot collaboration domain and even exacerbated when the robot makes use of ML components. In fact, the data is often the only available "ground truth" of correct behavior for ML components. Available data can only partially represent the correct behavior of CAIS. In our running example (Section 1), the model of the operator's behavior built on available data might not cover an unforeseen movement toward forbidden unsafe areas of the human agent. Furthermore, such behavior constantly changes due to the learning skills of the system. Since CAIS learn from new execution scenarios, design-time specifications must either account for future changes or be incrementally refined online, as the system evolves.

CH4 ***Top-down/bottom-up duality***: There exists a fading boundary between the two approaches in the emerging assurance methods tailored to CAIS. For instance, refinement of top-down decomposition of requirements from standards interleaved with bottom-up analysis of human needs. Another example is top-down partial specification interleaved with bottom-up run-time assurance evidence. Therefore, "traditional" top-down specifications of requirements shall coexist with partial/incomplete or example-based specifications (i.e., examples

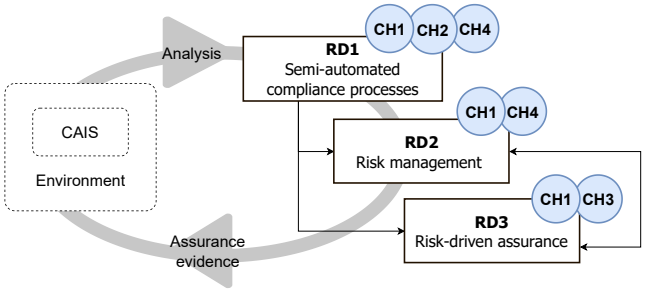

Fig. 2: Research directions RD (Section 4) and related challenges CH (Section 3).

of good/bad behaviors). In our running example, a top-down safety compliance requirement could be "no injuring behaviors with probability 99.99%" whereas the ML visual perception component is built bottom-up on the positive and negative examples (i.e., injured humans), which may not be available. Overall, the top-down/bottom-up duality nature calls for revisited methods built with awareness and endowed with the ability to interleave the two facets.

4 Research Roadmap

In this section, we introduce a research roadmap to address the challenges discussed above. Figure 2 shows the high level schema of this roadmap that outlines the Research Directions (RD), and the challenges (CH) that they face. We are following it in our collaboration with Fraunhofer Italia ARENA[9] where we will have the opportunity to validate results in real world CAIS in the industrial manufacturing domain.

RD1 ***Semi-automated compliance processes***: The focus of this RD is on supporting semi-automated derivation of compliance requirements of CAIS from norms and their embedding into proper assurance cases. Requirements elicitation should rest on the continuous analysis of the uncertain environment (CH1) and, as a consequence, the management of possible risks. Multi-paradigm approaches (e.g., goal-oriented techniques [10]) could be exploited to capture different facets of relevant standards and norms regulating CAIS. A promising approach here is the extension of existing modelling techniques by enabling (semi)-automated creation of norm models from domain standards and rules described using natural language (e.g., by exploiting NLP [13]). Furthermore, domain expert knowledge must be considered to support norms evolution based on the feedback and data from the field operation (CH2). Quality criteria for CAIS and a set of compliance requirements (e.g., *coverage* metrics [13]) represents another immature research direction. Namely, the definition of techniques to automate quality assurance processes as well as providing recommendations to engineers for improving the compliance requirements themselves (CH4) require further investigation.

[9] `https://www.fraunhofer.it/en/focus.html`

RD2 ***Risk management***: A successful strategy to cope with uncertainty (CH1), while dealing with safety in particular and dependability concerns in general, is the adoption of a risk management perspective. Uncertainty in CAIS leads to risks that must be identified, analyzed, monitored, and mitigated. Risk models could be adopted to quantify the risk of breaking compliance requirements. The notion of risk is a combination of the likelihood and impact of negative events. In our context, we see the likelihood as the probability that during execution, specific characteristics of the environment (i.e., domain features) cause problems to the ML component (e.g., misclassifications) that might break compliance requirements [7]. Data and domain models will be used to describe constraints, partitions and ranges of domain features (e.g., possible positions of the robotic arm in our running example) and constraints on the underlying data source (e.g., reliability of sensors), respectively. Risk analysis has the potential of quantifying existing risks by integrating probability and impact functions in order to prioritize risks and related compliance requirements that need special attention during the compliance assurance processes. This should inform the top-down and bottom-up compliance requirements management strategies in the learning and operative collaborative activities (CH4). The results of the assurance methods can be used to further bring the overall residual risk below acceptable levels, as defined by proper thresholds associated with assurance cases. Here, there exists the urgent need of novel adequacy criteria prescribing meaningful thresholds for residual risk in the context of CAIS.

RD3 ***Risk-driven assurance***: The risk models developed in RD2 have the potential to drive the automated creation of semantic labels associated with uncertain operating conditions that yield risks (CH1). Prioritized assurance cases and testing scenarios can then be derived from such operating conditions. The comprehensiveness of testing activities, should be then assessed by means of appropriate coverage metrics as discussed in the context of RD1. Further assurance can be provided by using incremental refinement of partial/incomplete specifications (CH3) through verification activities. Partial knowledge captured by risk models can be used to sample execution scenarios associated with high risk. Then run-time data can provide evidence about system compliance that can be used in turn to update the prior knowledge as defined by the risk model. As an example, *falsification* techniques [1,6] (traditionally applied in the context of cyber-physical systems) enhanced with awareness on risks have the potential of driving a CAIS towards compliance issues. In our running example, we could, for instance, falsify a safety compliance property requiring a minimum distance between the robot and the human operator.

5 Conclusion

In this paper, we provided a reflection on how research in the discipline of RE, and software/systems engineering can contribute towards the objective of building effective ML-equipped CAIS with strong, ideally provable compliance assurances. Major challenges that hinder our ultimate goal are discussed to rise awareness and call for contributions from the research community. To deal with

these challenges, we designed a research road-map towards the definition of compliance processes and requirements, risk management for ML-equipped CAIS, and risk-driven assurance.

References

1. Abbas, H., Fainekos, G., Sankaranarayanan, S., Ivančić, F., Gupta, A.: Probabilistic temporal logic falsification of cyber-physical systems. ACM Trans. Embed. Comput. Syst. **12**(2s), 1–30 (2013)
2. Ashmore, R., Calinescu, R., Paterson, C.: Assuring the machine learning lifecycle: Desiderata, methods, and challenges (2019), `https://arxiv.org/abs/1905.04223`, late accessed: November 2020
3. Billard, A.G., Calinon, S., Dillmann, R.: Learning from humans. In: Springer handbook of robotics, pp. 1995–2014. Springer (2016)
4. Breaux, T.D., Vail, M.W., Anton, A.I.: Towards regulatory compliance: Extracting rights and obligations to align requirements with regulations. In: 14th IEEE International Requirements Engineering Conference (RE). pp. 49–58. IEEE (2006)
5. Camilli, M., Russo, B.: Model-based testing under parametric variability of uncertain beliefs. In: de Boer, F., Cerone, A. (eds.) Software Engineering and Formal Methods. pp. 175–192. Springer International Publishing, Cham (2020)
6. Dreossi, T., Fremont, D.J., Ghosh, S., Kim, E., Ravanbakhsh, H., Vazquez-Chanlatte, M., Seshia, S.A.: Verifai: A toolkit for the formal design and analysis of artificial intelligence-based systems. In: Dillig, I., Tasiran, S. (eds.) Computer Aided Verification. pp. 432–442. Springer International Publishing, Cham (2019)
7. Foidl, H., Felderer, M.: Risk-based data validation in machine learning-based software systems. In: proceedings of the 3rd ACM SIGSOFT international workshop on machine learning techniques for software quality evaluation. pp. 13–18 (2019)
8. Giusti, A., Zeestraten, M.J.A., Icer, E., Pereira, A., Caldwell, D.G., Calinon, S., Althoff, M.: Flexible automation driven by demonstration: Leveraging strategies that simplify robotics. IEEE Robotics Automation Magazine **25**(2), 18–27 (2018)
9. Giusti, A., Steiner, D., Gasparetto, W., Bertoli, S., Terzer, M., Riedl, M., Matt, D.T.: Kollaborative robotik - maschinelles lernen durch imitation. Industrie 4.0 Management pp. 43–46 (2019)
10. Ishikawa, F., Matsuno, Y.: Evidence-driven requirements engineering for uncertainty of machine learning-based systems. In: 2020 IEEE 28th International Requirements Engineering Conference (RE). pp. 346–351. IEEE (2020)
11. Ishikawa, F., Yoshioka, N.: How do engineers perceive difficulties in engineering of machine-learning systems? - questionnaire survey. In: 2019 IEEE/ACM Joint 7th International Workshop on Conducting Empirical Studies in Industry (CESI) and 6th International Workshop on Software Engineering Research and Industrial Practice (SER&IP). pp. 2–9. IEEE (2019)
12. Rahman, S.M., Wang, Y., Walker, I.D., Mears, L., Pak, R., Remy, S.: Trust-based compliant robot-human handovers of payloads in collaborative assembly in flexible manufacturing. In: 2016 IEEE International Conference on Automation Science and Engineering (CASE). pp. 355–360. IEEE (2016)
13. Torrea, D., Abualhaijaa, S., Sabetzadehb, M., Briand, L., Baetens, K., Goes, P., Forastier, S.: An ai-assisted approach for checking the completeness of privacy policies against gdpr. In: 28th IEEE Intl. Requirements Engineering Conference, RE 2020, Zurich, Swiss, August 31 - September 4, 2020. pp. 136 – 146 (2020)

14. Vogelsang, A., Borg, M.: Requirements engineering for machine learning: Perspectives from data scientists. In: 2019 IEEE 27th International Requirements Engineering Conference Workshops (REW). pp. 245–251. IEEE (2019)