



## Causal Explanations of Property Violations in Distributed Discrete Event Systems

Post-Doc position based in Toulouse, France.

This position is related to the ANITI Chair "Synergistic transformations in model-based and data-based diagnosis" carried out by Louise Travé-Massuyès (<http://aniti.univ-toulouse.fr>), collaboration with INRIA Grenoble - Rhône-Alpes

**Advisor (s):** Yannick Pencolé – [ypencole@laas.fr](mailto:ypencole@laas.fr)  
Gregor Goessler – [gregor.goessler@inria.fr](mailto:gregor.goessler@inria.fr)  
Louise Travé-Massuyès – [louise@laas.fr](mailto:louise@laas.fr) (contact person)

**Net salary:** 2,7K€ per month with some teaching (64 hours per year on average)

**Duration:** 12 months

### SCIENTIFIC CONTEXT

Embedded software is nowadays everywhere (automotive systems, aeronautics, energy, communication, health...), they communicate with each other and also interact with us in our everyday life. They implement increasingly complex and smart functionalities and take over decision making even in critical situations (for instance in autonomous cars). The more complex and smart they are, the more important it becomes that their decisions be explainable to human beings.

In this work, we consider that embedded software behaves as a set of interacting discrete event components (a distributed discrete event system). Their behaviors usually comply with some properties (safety, fairness, ...) but their interaction with their environment might lead to the violation of such properties and these violations must be explained. The ability to automatically construct explanations is then crucial during the whole software life cycle: at design time to quickly localize errors and gain confidence; at runtime to understand the sources of failures and help assigning legal liability (for instance in the case of a car accident). From this topic, we can draw the following scientific challenges:

- 1) how to efficiently build an explanation about the failure of the system globally?
- 2) how to efficiently build local explanations component-wise that are consistent at the system level?
- 3) what are the relationships between a system-level explanation and a set of local explanations?

The Model-Based Diagnosis community has provided a set of techniques that aim at detecting failures and isolating faulty event occurrences based on a discrete event behavioral model of a system and an observable execution log. The strength of a diagnostic process is to determine what happened that is consistent with the observations. However, to explain why the observed outcome occurred, one must borrow techniques from causal analysis. The focus will be put on providing explanations that extract the relevant part of a property violation that can be understood by a human operator. Explanations should indeed retain all information about the system execution that is causally relevant to the failure at hand and make abstraction of irrelevant information. The project will build on a piece of work previously developed [2] that addresses the construction of formal explanations along an approach situated at the crossroads between model-based diagnosis [1; 4; 6] and causal analysis [3; 5].

## Objectives of the project

- Get an overview of the state of the art of explanation in Artificial Intelligence
- Take ownership of existing work and developed algorithms
- Formalize and study how the above could be extended in a distributed discrete event system framework, in which local explanations could be determined for each component while still providing a relevant global explanation when composed
- Revise the algorithms and assess the complexity gain
- Work on providing the explanations with semantic interpretation
- Reflect on a projection of the method to other frameworks than discrete event systems

**Candidate scientific skills:** Good background in at least one of the following domains, and interest in the others is required: symbolic/explainable AI, formal methods, discrete event systems, language theory, automata.

### > Other

- Good level of spoken, read and written English
- Autonomy and initiative
- Good oral communication skills

### > Bibliography

[1] T. Escobet, A. Bregon, B. Pulido, and V. Puig, editors. Fault Diagnosis of Dynamic Systems: Quantitative and Qualitative Approaches. Springer, 2019.

[2] G. Gössler, T. Mari, Y. Pencolé, L. Travé-Massuyès, Towards Explanations of Properties Violations in DES, Submitted to the 30<sup>th</sup> Int. Workshop on Diagnosis DX'19, Klagenfurt, Austria, November 11-13, 2019. Submitted.

[3] G. Gössler and J.-B. Stefani, Fault Ascription in Concurrent Systems. Proc. Trustworthy Global Computing 2015, LNCS 9533, Springer, 2016. <https://hal.inria.fr/hal-01246485>

[4] Walter Hamscher, Luca Console, and Johan de Kleer, editors. Readings in Model-based Diagnosis. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1992.

[5] J. Y. Halpern, A Modification of the Halpern-Pearl Definition of Causality. Proc. International Joint Conference on Artificial Intelligence 2015. <http://ijcai.org/Abstract/15/427>

[6] J. Zaytoon and S. Lafortune, Overview of fault diagnosis methods for Discrete Event Systems. Annual Reviews in Control 37(2), 2013.

## APPLICATION PROCEDURE

Formal applications should include detailed cv, a motivation letter and transcripts of bachelors' degrees.

Samples of published research by the candidate and reference letters will be a plus.

> applications should be sent by email to: advisor email

More information about ANITI: <https://aniti.univ-toulouse.fr/>