6-14-2024

A Framework For Explainable Root Cause Analysis In Manufacturing Systems – Explainable Artificial Intelligence for Shopfloor Workers

Daniel Kiefer
Reutlingen University, daniel_kiefer@outlook.de

Florian Grimm
Reutlingen University, florian.grimm@reutlingen-university.de

Tim Straub
ESB Business School - Reutlingen University, tim.straub@reutlingen-university.de

Günter Bitsch
Reutlingen University, guenter.bitsch@reutlingen-university.de

Clemens van Dinther
IISM, clemens.van_dinther@reutlingen-university.de

A FRAMEWORK FOR EXPLAINABLE ROOT CAUSE ANALYSIS IN MANUFACTURING SYSTEMS – EXPLAINABLE ARTIFICIAL INTELLIGENCE FOR SHOPFLOOR WORKERS

TREO Paper

Daniel Kiefer, ESB Business School, Reutlingen University, Reutlingen, Germany, Daniel.Kiefer@reutlingen-university.de

Florian Grimm, ESB Business School, Reutlingen University, Reutlingen, Germany, Florian.Grimm@reutlingen-university.de

Tim Straub, ESB Business School, Reutlingen University, Reutlingen, Germany, Tim.Straub@reutlingen-university.de

Günter Bitsch, ESB Business School, Reutlingen University, Reutlingen, Germany, Guenter.Bitsch@reutlingen-university.de

Clemens Van Dinther, IISM, KIT, Karlsruhe, Germany, c.v.dinther@gmx.de

Abstract

This paper proposes a novel framework – "Transparent Reasoning in Artificial intelligence Cause Explanation" (TRACE) – that combines root cause analysis, explainable artificial intelligence, and machine learning in a comprehensible manner for the shopfloor worker. The goal is to enhance transparency, interpretability, and explainability in AI-driven decision-making processes as well as to increase the acceptance of AI within an industrial manufacturing area. A human AI collaboration tool in perspective. The paper outlines the need of such a framework, describes the proposed design science approach for the development.

Keywords: Explainable Artificial Intelligence, Root Cause Analysis, Machine Learning.

1 Introduction

Optimizing and understanding industrial processes is an important task for organizations (Ahmed et al., 2022). The trend goes towards the use of Artificial Intelligence (AI) for predictions and the identification of relational dependencies. In many cases, AI is already transforming the operational landscape, for example in automating complex processes and unveiling new avenues for efficiency and productivity (Chakraborty et al., 2017). AI models are often hard to interpret or not interpretable at all, as slight parameter changes can have incomprehensible influence on the model's outcome. For this reason, AI models are often referred to as black boxes, which has led to limited use of AI algorithms because stakeholders are unable to understand, trust, and effectively manage AI-driven processes (Doshi-Velez & Kim, 2017). Workers are not necessarily used to dealing with results of AI algorithms, as the influencing factors often are detached from their operational context.

To achieve effective integration of digital applications in manufacturing, prioritizing human-centric design is paramount for fostering a collaborative digital future. By centering development around the needs and insights of the workforce, we can establish the necessary framework for such integration. The critical question then becomes:

How can advancements in computer science, specifically in Explainable Artificial Intelligence (XAI), be effectively incorporated into the decision-making processes of shopfloor workers in manufacturing environments?

This research aims to pioneer the development of explainable artificial intelligence for shopfloor workers by harnessing and integrating the collective knowledge and methodological strengths from diverse fields. Our goal is to catalyze and refine XAI research within and across various research communities, ensuring that technological advancements are both accessible and beneficial to those at the heart of manufacturing operations.

For this reason, we propose a novel hybrid framework, the "Transparent Reasoning in Artificial intelligence Cause Explanation system" (TRACE). Our hypothesis is that TRACE has high relevance within the context of AI acceptance within the industrial field. In the study of Vössing et al. (2019) for example, one of the obtained results were that "[…] beneficial transparency enhances process performance, as evidenced by a statistically significant reduction in downtime incidents' duration".

2 Proposed Method

For the presentation of our research idea we follow the design science research approach of (Peffers et al., 2006). We currently finished the third step design and development.

Problem Identification & Motivation - The context of the case study lies in a currently running research project. In our specific application, we focus on CNC grinding machines, which are employed to produce high-precision tools, such as turning tools. These tools enable operations like internal grooving and thread turning. Rising setup times are suspected to be due to the production parameters related to batch size one. The increasing complexity of the products, employee changes, and the number of influential parameters further compound these issues. The influential parameters range across machine, human factors, processes, environment, and more.

Objectives of a Solution – In requirements workshops with the industrial partner, the requirements and objectives of a potential information systems artefact were derived. During the workshops, one thing became clear. Users want not only a decision system that tells them what to do based on machine learning algorithms, but also why. Explainable components are wanted. In the work cycles with the partners, it became clear that pure SHAPE values and their graphics were not understood by the machine operators. Therefore, the idea came up to combine it with the familiar tools from the Root Cause Analysis family, for example the Ishikawa diagram. The primary aim is to develop a methodology that enhances the transparency, interpretability, and overall trust in AI-assisted decision-making processes within manufacturing systems.

i. It should offer global and local explainability relevant to the manufacturing context based on root cause analysis.
ii. The explanations provided should be understandable for all stakeholders, ranging from machine operators to shopfloor management.
iii. It should make use of a wide and regularly updated data basis encompassing machine data, process data, and other relevant parameters.
iv. The system should utilize existing domain knowledge, exploiting the wealth of experience and insights already present within the organization.
v. Operators should have the possibility to provide feedback and the model should learn from it.

Design & Development – A user is assigned a new work order via the company's SAP System. To execute their task optimally, they input the order number into the user interface to ascertain which features will impact the setup time of the given assignment. This order number is transmitted as a request to a server that houses a pre-trained prediction model, generating an estimated setup time for the specific task. The model is sourced from an AI Cluster in the background, which retrieves historical manufacturing data from the company's operational databases via an intermediary data warehouse for machine learning training. Upon being trained, the model is distributed to the request-handling server, which computes the influencing factors in the form of SHAP values for all features, in addition to the aforementioned setup time estimation. These SHAP values, combined with a root cause analysis of all factors impacting setup time, form the TRACE-diagram visualization combined in a dashboard. The user can now make an informed decision about the tools and machines to be used for their production process, guided by the TRACE-diagram.

Demonstration – Designing an effective dashboard for manufacturing settings presents notable challenges, particularly in tailoring it to meet the specific needs of shopfloor workers. Initial discussions with employees highlighted the unsuitability of conventional SHAPE diagrams for this audience, emphasizing the need for a bespoke solution. To address this, we propose a three-step methodology. The first step involves gathering detailed insights into the current decision-making processes of the workers. The second step entails conducting a survey to ascertain whether employees prefer access to all available data or only to data relevant to their direct influence. The final step is to identify the key attributes that workers value in a dashboard, which may encompass aspects such as functionality, design layout, color schemes, and typography. This approach ensures the development of a user-centric dashboard that enhances decision-making on the shop floor.

Evaluation – Evaluating the effectiveness of design solutions requires real user interaction to identify potential issues. By employing user feedback through surveys, we can pinpoint areas for improvement. Furthermore, the use of technology acceptance models and system usability scale assessments, or similar evaluation tools, allows for a comprehensive assessment of the designs. This multifaceted approach ensures that design decisions are informed by actual user experiences and preferences, leading to more user-friendly and effective solutions.

Communication – This publication marks the beginning of our communication about the TRACE Framework. We will continue to share more information in upcoming publications. We are confident that the TREO Forum will provide valuable feedback on our methodology. This feedback is essential to our ongoing efforts to refine and improve our approach to ensure it meets the needs and expectations of the community.

The TRACE framework has the potential to improve decision-making and production efficiency in manufacturing. It aids shopfloor managers in developing optimal scheduling strategies and transparently reveals reasons for high setup times in meetings. Additionally, it guides workers to select the best machine for upcoming tasks, empowering them with informed decision-making.

References

Ahmed, I., Jeon, G., & Piccialli, F. (2022). From Artificial Intelligence to Explainable Artificial Intelligence in Industry 4.0: A Survey on What, How, and Where. *IEEE Transactions on Industrial Informatics*, *18*(8), 5031–5042.

Chakraborty, S., Tomsett, R., Raghavendra, R., Harborne, D., Alzantot, M., Cerutti, F., Srivastava, M., Preece, A., Julier, S., Rao, R. M., Kelley, T. D., Braines, D., Sensoy, M., Willis, C. J., & Gurram, P. (2017). Interpretability of deep learning models: A survey of results. In *IEEE SmartWorld 2017* (pp. 1–6).

Doshi-Velez, F., & Kim, B. (2017, February 28). *Towards A Rigorous Science of Interpretable Machine Learning*. arxiv.

Peffers, K., Tuunanen, T., Gengler, C., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2006). The design science research process: A model for producing and presenting information systems research. *Proceedings of First International Conference on Design Science Research in Information Systems and Technology DESRIST*.

Vössing, M., Potthoff, F., Kühl, N., & Satzger, G. (2019). Designing useful transparency to improve process performance: Evidence from and automated production line. *In Proceedings of the 27th European Conference on Information Systems (ECIS)*.