

TRUST4AI.XAI

Explainability provisioning component

The main goal of TRUST4AI.xAI is to enhance transparency and trust in AI systems by providing explainable AI (xAI) services. The component enables model engineers, data scientists, and cybersecurity practitioners to understand how and why AI/ML models produce specific decisions. TRUST4AI.xAI allows users to analyze their models with multiple local and global explanation techniques, improving interpretability, accountability, and compliance in sensitive domains such as cybersecurity and finance.



AI4CYBER
Component



In a context where AI models are often seen as opaque “black boxes,” TRUST4AI.xAI makes AI decisions more transparent and trustworthy by:

- Allowing integration of external AI models (REST, websocket, offline).
- Supporting multiple xAI explainers (LIME, SHAP, Tree-based, DiCE, Anchors, ARIA, HaRIA, GeRIA).
- Providing a user-friendly dashboard for interactive analyses and visualisations.
- Ensuring scalability and maintainability through microservices and container orchestration.
- Supporting decision-making in cybersecurity scenarios (e.g., intrusion detection in AI4CYBER pilots)



Model engineers and data scientists can use TRUST4AI.xAI to perform analyses of their own AI models and visualize explanations in an accessible way.

The frontend is a React-based dashboard that enables intuitive interaction: selecting models, uploading datasets, choosing explainers, and exploring results through interactive visualizations.

The backend (API Gateway with microservices, Kafka message broker, and databases) processes user requests, manages models and explainers, and ensures robust and scalable communication.

The solution supports uploading or connecting to external AI models as well as running local and global explainability analyses.



[YouTube Video Link](#)



Dual licence: open source for restricted version, commercial for full version

itti

Marek Pawlicki (marek.pawlicki@itti.com.pl)
ITTI Sp. z o. o.

www.itti.com.pl



- Pawlicki, M. (2023). Towards Quality Measures for xAI algorithms: Explanation Stability. 2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA), 1–10. <https://doi.org/10.1109/DSAA60987.2023.10302535>
- Pawlicka, A., Pawlicki, M., Kozik, R., & Choraś, M. (2023). The Need for Practical Legal and Ethical Guidelines for Explainable AI-based Network Intrusion Detection Systems. 2023 IEEE International Conference on Data Mining Workshops (ICDMW), 253–261. <https://doi.org/10.1109/ICDMW60847.2023.00038>
- Pawlicki, M. (2023). Towards Quality Measures for xAI algorithms: Explanation Stability. 2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA), 1–10. <https://doi.org/10.1109/DSAA60987.2023.10302535>



Funded by the
European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101070450. Disclaimer: Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the European Commission can be held responsible for them.