Université de Rennes

Direction de la Recherche et de l'Innovation

Appel à projets - Année 2025

Contrats doctoraux «Ordinaires»

ECOLE DOCTORALE : MATISSE **POLE DE RECHERCHE : Mathématiques-Numérique**

Unité de recherche d'accueil

Unité de Recherche : IRISA - UMR 6074

Equipe : KERDATA

Encadrement

Directeur de thèse : Gabriel Antoniu **Unité de recherche** : IRISA - UMR 6074

Employeur/Établissement d'origine : Inria

HDR : oui **Nb de thèses en cours** : 3

Mail : gabriel.antoniu@inria.fr **Tél.** : +33 2 99 84 72 44 (87244)

Co-directeur de thèse : Alexandru Costan **Unité de recherche** : IRISA - UMR 6074

Employeur/Établissement d'origine : INSA Rennes

HDR : oui **Nb de thèses en cours** : 2

Mail : alexandru.costan@irisa.fr **Tél.** : +33 2 99 84 75 66 (87566)

Co-directeur de thèse : Silvina Caino-Lores **Unité de recherche** : IRISA - UMR 6074

Employeur/Établissement d'origine : Inria

HDR : non **Nb de thèses en cours** : 1

Mail : silvina.caino-lores@inria.fr **Tél.** : +33 2 99 84 72 31 (87231)

Co-directeur de thèse : Loic Cudennec **Unité de recherche** : AMIAD - Agence Ministérielle pour l'IA de Défense

Employeur/Établissement d'origine : DGA

HDR : oui **Nb de thèses en cours** : 3

Mail : loic.cudennec@def.gouv.fr **Tél.** :

Sujet

Titre : Workflow Provenance and Its Application to Responsible and Trustworthy Artificial Intelligence

Mots clés : artificial intelligence, explainability, reproducibility, energy efficiency, workflow provenance

Cotutelle internationale (oui/non) : non **Si oui, préciser l'établissement et le pays de rattachement** :

Descriptif (15-30 lignes) :
Contexte socioéconomique et scientifique ; Les hypothèses et questions posées ; Les grandes étapes de la thèse et démarche ; Approches méthodologiques et techniques envisagées

Artificial intelligence (AI) has raised concerns in governments, research communities, industry and the public opinion regarding its ethical and practical implications. Consequently, the ideal of Responsible and Trustworthy AI (RTAI) garnered attention worldwide [LWM24, Oak24, IBM24, Mic24, SLD+22]. The path towards RTAI involves developing practices to prioritise not only model performance (e.g., accuracy and loss), but also principles that play crucial societal roles, such as *explainability*, or making AI's decision-making process understandable; *transparency*, ensuring clarity in AI's design, data and operation; and *environmental impact* of the machine learning life-cycle (e.g., large-scale model training) [GMRRG19]. Recent works suggest that workflow provenance (i.e., the documentation and tracking of all processes within AI development) might be the key to support RTAI [SAL+22, KNHJ+23]. Workflow provenance refers to capturing detailed information about all

activities, processes, and transformations applied to data and code during AI development and operations. It includes information about data sources, data preprocessing, model selection, hyperparameter tuning, and evaluation metrics, among others. Capturing this provenance could provide a holistic view of the AI workflow, making it transparent and reproducible, and allowing further analysis of features such as the carbon footprint and computational power required for training a given model [VW21]. However, today there are no comprehensive formalisms able to capture the complex relationships in workflow and model provenance data [BBFM23]. Furthermore, multiple technical challenges arise when attempting to capture, store and manage the full provenance of AI workflows [MCSAGBS21, SS23], and it is still not well understood what information is valuable and how it can be exploited [JRO+20].

This project aims to investigate methods to leverage workflow provenance metadata in support of RTAI. It will investigate mechanisms to formalize, capture, store, retrieve, inspect, analyse and visualise metadata in AI-powered workflows, and will explore the relationship between model provenance, metadata, and model and system behavior, aiming to decipher how architectural and algorithmic characteristics impact in the model's outcome and efficiency. First, we will explore the formalisation of ontologies and taxonomies for AI workflow provenance data from multiple angles: system (e.g., hardware, computing infrastructure, storage), platform (e.g., workflow manager, machine learning framework), model (e.g., hyperparameters, performance, architecture), and application (e.g., input and intermediate data, feedback). The resulting theoretical framework will expand our previous work on taxonomies for neural network metadata [RCLJT22] to other machine learning methods. At the technical level, new data structures, algorithms, system architectures and interfaces will be explored to efficiently produce and query a detailed record of data sources, processing steps, and model configurations. We will build upon our previous work on reproducible workflow execution [RCAV21] to rarchitect a proof-of-concept for a large-scale provenance data management system suitable for the analysis AI workflow applications. Finally, we will work towards a methodological framework leveraging provenance matadata taxonomies, the causal model between the studied models and their behavior, and associated statistical findings that support RTAI in practice.

References

[BBFM23] Elisa Bertino, Suparna Bhattacharya, Elena Ferrari, and Dejan Milojicic. Trustworthy AI and data lineage. IEEE Internet Computing, 27(6):5–6, 2023.

[GMRRG19] Eva Garcıa-Martın, Crefeda Faviola Rodrigues, Graham Riley, and Hakan Grahn. Estimation of energy consumption in machine learning. Journal of Parallel and Distributed Computing, 134:75–88, 2019.

[IBM24] IBM Research. Trustworthy AI at IBM. https://research.ibm.com/topics/trustworthy-ai, 2024.

[JRO+20] Fariha Tasmin Jaigirdar, Carsten Rudolph, Gillian Oliver, David Watts, and Chris Bain. What information is required for explainable AI? : A provenance-based research agenda and future challenges. In 2020 IEEE 6th International Conference on Collaboration and Internet Computing (CIC), pages 177–183, 2020.

[KNHJ+23] Amruta Kale, Tin Nguyen, Frederick C Harris Jr, Chenhao Li, Jiyin Zhang, and Xiaogang Ma. Provenance documentation to enable explainable and trustworthy AI: A literature review. Data Intelligence, 5(1):139–162, 2023.

[KRCLJT22] Ariel Keller Rorabaugh, Silvina Ca´ıno-Lores, Travis Johnston, and Michela Taufer. Building high-throughput neural architecture search workflows via a decoupled fitness prediction engine. IEEE Transactions on Parallel and Distributed Systems, 33(11):2913–2926, 2022.

[LWM24] Johann Laux, Sandra Wachter, and Brent Mittelstadt. Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk. Regulation & Governance, 18(1):3–32, 2024.

[MCSAGBS21] Marcal Mora-Cantallops, Salvador Sanchez-Alonso, Elena Garcıa-Barriocanal, and Miguel-Angel Sicilia. Traceability for trustworthy AI: A review of models and tools. Big Data and Cognitive Computing, 5(2), 2021.

[Mic24] Microsoft Research. FATE: Fairness, Accountability, Transparency, and Ethics in AI, 2024.

[Oak24] Oak Ridge National Laboratory. Oak Ridge National Laboratory initiative in secure, trustworthy, and energy-efficient AI, 2024.

[RCAV21] Daniel Rosendo, Alexandru Costan, Gabriel Antoniu, and Patrick Valduriez. E2clab: Reproducible analysis of complex workflows on the edge-to-cloud continuum. In IPDPS 2021-35th IEEE International Parallel and Distributed Processing Symposium, 2021.

[SAL+22] Renan Souza, Leonardo G Azevedo, Vıtor Lourencco, Elton Soares, Raphael Thiago, Rafael Brandao, Daniel Civitarese, Emilio Vital Brazil, Marcio Moreno, Patrick Valduriez, et al. Workflow provenance in the lifecycle of scientific machine learning, 2022.

[SLD+22] Conrad Sanderson, Qinghua Lu, David Douglas, Xiwei Xu, Liming Zhu, and Jon Whittle. Towards implementing responsible AI. In 2022 IEEE International Conference on Big Data (Big Data). IEEE, December 2022.

[SS23] Marius Schlegel and Kai-Uwe Sattler. Mlflow2prov: extracting provenance from machine learning experiments. In Proceedings of the Seventh Workshop on Data Management for End-to-End Machine Learning, pages 1–4, 2023.

[VW21] Aimee Van Wynsberghe. Sustainable AI: AI for sustainability and the sustainability of AI. AI and Ethics, 1(3):213–218, 2021.