Workflow Provenance in the Computing Continuum for Responsible, Trustworthy, and Energy-Efficient AI

Renan Souza, Silvina Caino-Lores, Mark Coletti, Tyler J Skluzacek, Alexandru Costan, Frédéric Suter, Marta Mattoso, Rafael Ferreira da Silva

Workflow Provenance in the Computing Continuum for Responsible, Trustworthy, and Energy-Efficient AI

Renan Souza[1], Silvina Caino-Lores[2], Mark Coletti[1], Tyler J. Skluzacek[1]
Alexandru Costan[2], Frédéric Suter[1], Marta Mattoso[3], Rafael Ferreira da Silva[1]

[1]Oak Ridge National Laboratory, TN, USA
[2]University of Rennes, Inria, CNRS, IRISA, France
[3]COPPE/Federal University of Rio de Janeiro, Brazil

***Abstract*—As Artificial Intelligence (AI) becomes more pervasive in our society, it is crucial to develop, deploy, and assess Responsible and Trustworthy AI (RTAI) models, *i.e.*, those that consider not only accuracy but also other aspects, such as explainability, fairness, and energy efficiency. Workflow provenance data have historically enabled critical capabilities towards RTAI. Provenance data derivation paths contribute to responsible workflows through transparency in tracking artifacts and resource consumption. Provenance data are well-known for their trustworthiness helping explainability, reproducibility, and accountability. However, there are complex challenges to achieve RTAI, which are further complicated by the heterogeneous infrastructure in the computing continuum (Edge-Cloud-HPC) used to develop and deploy models. As a result, a significant research and development gap remains between workflow provenance data management and RTAI. In this paper, we present a vision of the pivotal role of workflow provenance in supporting RTAI and discuss related challenges. We present a schematic view between RTAI and provenance, and highlight open research directions.**

***Index Terms*—Artificial Intelligence, Provenance, Machine Learning, AI workflows, ML workflows, Responsible AI, Trustworthy AI, Reproducibility, AI Lifecycle, Energy-efficient AI**

I. Introduction

Artificial Intelligence (AI) has experienced remarkable progress in various domains, ranging from healthcare and finance to autonomous vehicles and natural language understanding. Alongside the potential benefits, the development of AI models must be responsible and trusted. This has become a high priority in agencies worldwide, from academia, industry, national laboratories, and government [1]–[4]

The Fairness, Accountability, Transparency, and Ethics (FATE) principles were designed to address algorithmic disparities and enhance AI inclusivity [4]. Furthermore, principles of Sustainability in AI focus on the environmental and societal impact of developing and using AI models that are energy-efficient, and how this impact can be assessed in the short and long terms. This includes analyzing features such as the carbon footprints and computational resources required for training and using models [5]. Moreover, principles such as Explainable AI (XAI) [6] involve techniques that seek to allow humans to comprehend and trust AI models.

Additionally, the Findability, Accessibility, Interoperability, and Reusability (FAIR) principles, which were initially proposed for data, are now adapted to practically define what FAIR means for AI models and datasets [7], aiming at improving reproducibility in AI. For instance, making available model data and metadata (*e.g.*, hyperparameters or initial weights for training) can enhance reproducibility. In this paper, we refer to *Responsible and Trustworthy AI (RTAI)* as practices that prioritize not only model performance (*e.g.*, accuracy, loss) but also principles that play crucial societal roles, such as robustness, explainability, fairness, transparency, auditability, accountability, reproducibility, and energy efficiency.

Workflow provenance data management methods have been evolving towards supporting RTAI. The W3C definition of provenance [8] explicitly mentions the goals of assessments about the quality and trustworthiness of workflow results. Provenance data contain the data derivation paths with information about entities, activities, and people involved in producing results. In the AI context, provenance include information about data sources, data preparation, model design, hyperparameter tuning, model evaluation, and people involved in the lifecycle of an AI model [9], [10]. Queries like *"which sensors were used to generate this data file?"*; *"which data transformations, with their parameters, were used to generate this training dataset?"*; *"which network architecture and hyperparameters were used to train this model?"*; *"which computing infrastructure and how much computing resources were used to obtain this model?"*; *"who trained it?"* are only a few examples that can be answered with provenance data.

However, training and using a model usually involves loosely coupled processes that have data dependencies on one another [9]–[11]. For example, to train a model, data are first obtained in a process, then curated in another, prepared for training in a subsequent process, and finally the training happens in a separate process. These processes are typically not explicitly connected, making it harder to keep track of the dataflows in between, thus hindering RTAI. Capturing

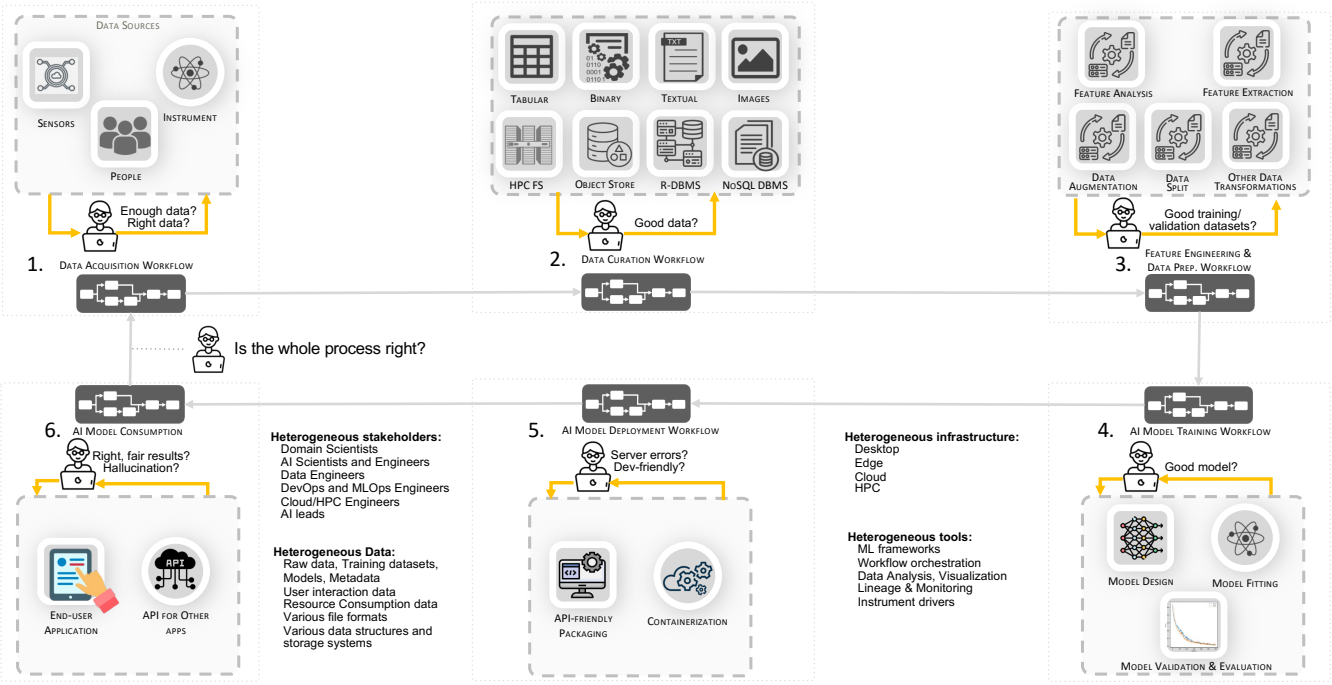

Fig. 1. High-level overview of the AI lifecycle, depicting multiple phases, workflows, stakeholders, data, tools, and infrastructure.

provenance data with their relationships along the AI lifecycle can make these dependencies explicit through data derivation path analyses, enabling a holistic view of the lifecycle, making it more transparent and reproducible [9].

Additionally, the AI lifecycle typically involves people with various expertise working together such as domain experts, DevOps and MLOps engineers, data and Machine Learning (ML) scientists and engineers, each using various specialized tools to perform their respective work [10]. Considering the involvement of different people in multiple processes, leveraging provenance relationships between people, activities, and entities can support accountability in RTAI.

Despite its potential and ongoing community efforts, there are several open challenges in designing and building provenance methods in support of RTAI. For example, it is not trivial to accurately represent provenance data to reflect the complexities of the AI lifecycle. In addition, developing provenance analysis methods to help AI scientists steer experiments and inspect datasets and models for performance, explainability, and fairness is hard. Moreover, AI practitioners must assess the impact of model predictions throughout the model lifecycle, including during inferences in the production stage, worry that models are reproducible, and be concerned with the trade-offs between energy efficiency and performance of their models. Besides, the highly federated, loosely coupled, and hetoregeneous nature of the processes and data in the model lifecycle makes provenance systems more difficult to design. In some domains, especially scientific ones, some processes run at the edge (*e.g.*, close to sensors, collecting data), others in the cloud (*e.g.*, model deployment, exploratory data analysis), while some (*e.g.*, training) may require High-Performance Computing (HPC) environments, in a computing continuum [12].

In this paper, we propose a vision of the role of workflow provenance data management in support of responsible, trustworthy, and sustainable AI in the computing continuum. First, we briefly present the background for this work: the AI lifecycle and a selection of the RTAI principles that we envision that can be supported by provenance data (Sec. II). Then, we present our vision of the main challenges related to the support of RTAI and discuss the potential of provenance methods to address them, highlighting possible research directions (Sec. III). Finally, we conclude in Sec. IV.

II. Responsible and Trustworthy AI: Model Lifecycle and Principles

The AI lifecycle is highly heterogeneous, with various stakeholders, infrastructure, tools, and data involved [9]–[11]. It can be represented as multiple distributed, federated, and not explicitly connected workflows [9]. Fig. 1 gives an overview of this lifecycle and illustrates the connections between these phases. *Data Acquisition, Curation, and Preparation* involve several chained processes. Data may be acquired from sensors, instruments, or people (*e.g.*, via observations, social media, manual inputs). After data acquisition and exploratory data analysis, data need to be curated to improve data quality and data readiness for subsequent phases. Data preparation involves feature engineering, preprocessing, and transforming the collected data to make suitable datasets for model training and evaluation. In order to obtain and configure a suitable model to use in production, stages of *Model Design, Training, Evaluation, and Validation* take place. ML scientists or engineers select suitable algorithms and design

model architectures, specifying model hyperparameters. Model validation assesses performance on validation datasets, aiming at evaluating model generalization. The *Model Deployment* stage deploys the trained and validated model to real-world applications, involving DevOps and MLOps engineers. Finally, in *Model Utilization*, the model is utilized by the end-users.

Several works have defined various principles towards RTAI, including FATE [4], FAIR [13], CARE (Collective benefit, Authority control, Responsibility, Ethics) [14], and others [6], [15]–[17]. Rather than proposing new principles or redefining existing ones, here we consolidate, organize, and contextualize the critical principles identified in the literature that can be supported by the workflows and data provenance communities, leaving aside human factors related to economic, political, and philosophical policies. Additionally, our vision does not fully explore all the RTAI principles that could benefit from provenance methods, such as security and privacy, which are not covered in this paper.

Model performance refers to a set of metrics (*e.g.*, accuracy, loss) used to evaluate how effectively a model can make predictions. Although it is the primary focus for optimization by ML scientists, in this context, trade-offs between model performance and other principles must be considered [17].

Robustness refers to the ability of AI models continue to perform well under a wide range of possible conditions, including unexpected or adverse circumstances, requiring validation across various threatening scenarios [17].

Explainability in AI aims to provide a suite of techniques to enable humans to understand and trust models [6]. They can be used to analyze and interpret model inputs and their corresponding predictions. Among other techniques, the ones based on feature importance can evaluate how specific features impact the predictions [18], [19].

Fairness addresses impartial treatment, aiming at mitigating biases and discrimination based on sensitive attributes [16]. By deploying models that treat the individuals according to given policies regulating fair treatment, developers can build inclusive and ethical AI-based systems.

Transparency and auditability emphasize maintaining a clear record of the phases in the AI lifecycle [17], helping to audit (*e.g.*, for compliance, governance) models and associated datasets. While model transparency often refers to "white-box" models [6], [17], in this context, transparency expands this concept and includes a clear documentation of processes, data, models, infrastructure, and humans' interaction in the lifecycle, enabling traceability.

Accountability is related to transparency and auditability, but it highlights the humans interacting with the artifacts (*e.g.*, models, datasets) in the lifecycle. It aims at clarifying the roles and responsibilities of various stakeholders, including AI scientists, engineers, MLOps engineers, and managers [16].

Reproducibility refers to the capability of an independent research team to achieve the same results using the same method, based on the documentation provided by the original research team [20]. It is essential for building trust in AI models [21]. A thorough documentation of the entire lifecycle, including details such as dataset characteristics, the tools and infrastructure used, consumed resources, configurations, parameters, and user interaction information from the humans involved in the process, is necessary for reproducibility.

Energy efficiency in AI balances model performance with the computational resources needed to train, evaluate, deploy, and utilize models. In addition to assessing the computational resources in training only, this includes evaluating processes and data across the entire AI lifecycle, such as data volumes and network transfers, in line with Green AI principles aimed at reducing carbon footprints and promoting frugal and sustainable AI practices [22], [23].

III. A Vision of Workflow Provenance in Support of Responsible and Trustworthy AI

This section discusses the role of workflow provenance data to support the aforementioned RTAI principles. We explore the capabilities typically enabled by workflow provenance data management methods towards supporting RTAI, and discuss key research challenges yet to be addressed.

A. The Role of Workflow Provenance Data in RTAI

Provenance allows retracing the history of artifact creation, supporting trust in processes involving both computational and human actors [24]. It is crucial for reproducibility and accountability in scientific research, particularly in collaborative environments [25]. Provenance frameworks also integrate human factors in fields like business management [26] and policy making [27], aiding in performance monitoring, transparency, stakeholder trust, and decision-making refinement.

In this context, provenance methods emerge as comprehensive frameworks to address scientific, technical, and human challenges in Responsible, Trustworthy, and Sustainable AI [28]. While previous studies [29] and surveys [28] have explored the role of provenance in RTAI principles like explainability [28], [30], auditability [31], and accountability [32], there is no unified vision on its pivotal role to support the RTAI principles. No other provenance-related work has addressed challenges associated with principles not typically linked with trustworthy AI but are still integral to responsible AI, such as energy efficiency in AI. Especially if we consider the various dimensions of heterogeneity in the lifecycle, including data and infrastructure heterogeneity in a computing continuum.

In our vision, workflow provenance data management plays two critical roles:

First, it acts as a ***glue*** connecting the diverse and dispersed artifacts across the lifecycle, including workflows, datasets, models, infrastructures, tools, and people working together [9]. Provenance data have a *data capillarity* ability, like the human circulatory system, where large arteries branch into finer arterioles and capillaries. Similarly, provenance data can transport both coarse and fine-grained information, linking different lifecycle phases and branching into detailed data within each phase as needed. This capability allows for detailed and flexible data capture throughout an artifact's lifecycle.

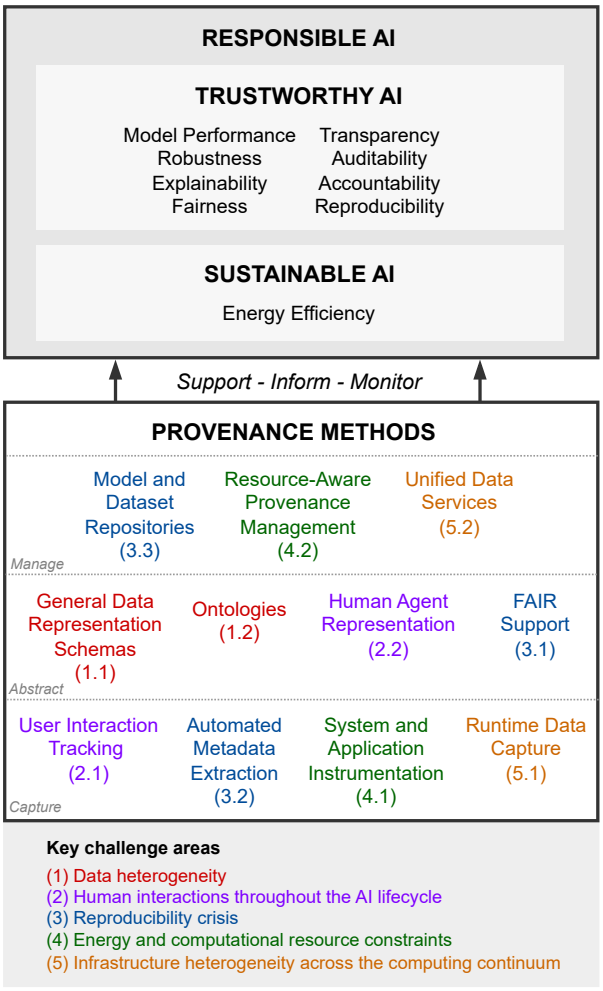

Fig. 2. Relationship of existing provenance methods for supporting, informing and monitoring principles of Responsible AI, including aspects of Trustworthy AI and Sustainable AI.

Second, once captured and stored, the provenance database serves as a ***data-to-insights accelerator*** by systematically organizing data, following community standards, facilitating both comprehensive analyses across workflows and in-depth inspections of individual workflows. This organization supports RTAI principles by enabling seamless navigation, querying, and discovery of insights related to the principles.

B. Supporting RTAI with Provenance Data Management: Key Challenge Areas and Opportunities

Provenance-based methods offer many opportunities to support RTAI principles. Figure 2 shows provenance methods used to capture, abstract and manage provenance data with the purpose of supporting, informing and monitoring the RTAI principles discussed in Sec. II. Nonetheless, multiple challenges arise when targeting full end-to-end workflow provenance data management across the AI lifecycle. Here, we discuss the main open challenges yet to be addressed within five key challenge areas, also depicted in Fig. 2.

(1) Data heterogeneity. Several data types coexist in the AI lifecycle, including raw domain-specific data; curated data; training, test, and validation datasets; models; and telemetry data. Each data type has their own data structure, format, and associated metadata, leading to specialized data manipulation and visualization tools for different data and tasks. Also, in complex scenarios, typically found in scientific domains [9], data may be stored in different storage systems, such as file systems or object stores, or in database systems with varying data models, including relational and NoSQL. Data can come from various sources, with various sizes, and generated at different speeds. Besides, considering the various stakeholders in the lifecycle, human interaction data—which are naturally different from these other data types—might also be generated.

In this context, data integration quickly becomes a challenge to enable data analyses spanning various phases of the lifecycle. Existing works have explored the use of workflow provenance to support evaluation of model performance and robustness, and explainability by recording and linking the execution traces of data producers and consumers [9], [30], [33], [34]. This is useful for establishing relationships between input training datasets and model outcomes. However, combining multiple types of domain-specific data, model performance metrics, human interaction data, and computational resource utilization data (*e.g.*, GPU utilization, temperature, power consumption) is a challenge due to the data heterogeneity.

Furthermore, XAI [18] and fairness AI [35] techniques produce more trustworthy-related metadata that should properly follow the corresponding models. Such metadata can be complemented with provenance data relationships, offering a broader picture, relating AI lifecycle artifacts with explainability and fairness indicators. The work by de Oliveira *et al.* [30] is a step towards this direction by using captured provenance data during the data preparation phase and integrating with explainability data (Shapley values [18]), enhancing explainability that shows attributes' derivation. However, how to capture, represent, store, and visualize explainability and fairness-related data jointly with provenance data to provide meaningful insights is still an open issue, especially considering the broader scope of all other types of data involved.

The W3C PROV framework [8], a W3C recommendation, provides a generic data representation schema (Fig. 2 **(1.1)**) that allows for implementing data integration, making it a suitable abstraction for provenance data in the AI lifecycle. The W3C PROV relationships *used* and *generated* between W3C PROV *Actions* and *Entities* enable gluing the lifecycle phases with their main data sources and data products. W3C PROV *Agents* can represent the humans involved in the lifecycle, like scientists, engineers, and end users. In addition to a coarse-grained glue between phases, these relationships and concepts can be further specialized for richer AI- or domain-specific semantics, in an attempt to capture the multi-granularity aspect within the phases. The works [9], [30] contain examples of specializations of the W3C PROV framework to support data analysis related to RTAI. PROV-ML [9], for instance, is a specialization of the PROV-O ontology (Fig. 2 **(1.2)**), as a way to implement a generic data representation schema, defining how certain phases of the lifecycle can be modeled to be

managed by a provenance system.

However, we are not aware of any work capable of providing unified data integration across all phases of the lifecycle considering these various types of data that we mentioned. Alleviating the complexity of integrating and managing these heterogeneous data is a key challenge for the provenance community in order to support the development of RTAI, particularly enabling principles of explainability, fairness, accountability, transparency, and energy efficiency.

(2) Human interactions throughout the AI lifecycle. There are several stakeholders involved in the AI lifecycle [9], [10]. Domain experts work towards data acquisition and curation; ML scientists and engineers work in model design, training, and evaluation; MLOps and DataOps engineers work on deployment; and end-users interact with the deployed models in production. Keeping track of different human interactions is essential to understand how each role influences data, processes, and models throughout the cycle.

Examples of human interactions, such as those by an ML scientist, include user-steered hyperparameter fine-tuning for model training, selecting algorithms or model architectures, and preparing training datasets. Such interactions are subject to human error and biases. Tracking them would enable establishing explicit data relationships, for instance, between the action of fine-tuning a model with its performance, or understanding how certain model design decisions impacted the predictions of a model in production.

Methods from experiment steering (also known as computational steering) [36] could be applied to track user interactions (Fig. 2 **(2.1)**) through provenance data within the AI lifecycle. Experiment steering involves both runtime data analysis and support for analysis of user-steered interactions in a scientific experiment [37]. While the use of provenance data for runtime data analysis in AI has been explored particularly for model training [34], these works lack support for other phases of the AI lifecycle and comprehensive integration of human agents in their provenance representations. The work of Rogers and Crisan [38] is a step toward representing human agents (Fig. 2 **(2.2)**) in the AI lifecycle, but it is focused exclusively on model training, leaving room for future research and development that address other lifecycle phases.

Representing the diverse roles and their interactions within the AI lifecycle is extremely challenging due to the wide range of possible interactions, tools used by different individuals, and the scope of knowledge of different human agents (*i.e.*, expected scope of knowledge for an ML scientist differs from that of an MLOps engineer and from the end user interacting with the model in production) [10]. Research efforts are necessary to properly record human interaction data in provenance databases with the goal of supporting RTAI principles that involve aspects of decision making like transparency, accountability, auditability, and fairness.

(3) Reproducibility crisis. AI is facing a reproducibility crisis, which undermines trust in AI models [21]. Given the diversity of data, parameters, tools, infrastructure, and multiple humans involved in the AI lifecycle, scientists and engineers still struggle with inadequate documentation to ensure model reproducibility. Supporting the FAIR principles (Fig. 2 **(3.1)**) in the entire AI lifecycle can enable new ways to achieve reproducibility in AI [7], [13]. For instance, providing machine-readable, structured documentation of the data, processes, and humans involved aids in finding, accessing, and reusing critical artifacts like datasets and models. Interoperability is facilitated when community standards or recommendations, such as W3C PROV [8], are followed [7]. Provenance capture tools [9], [39], [40] have the potential to automatically capture such detailed information, establish the right data relationships, and a provide relevant documentation, hence improving FAIR and reproducibility. When combined with automated metadata extraction [40] (Fig. 2 **(3.2)**), these tools can automatically and systematically register training processes and extract meaningful domain information from datasets.

Model and dataset repositories (Fig. 2 **(3.3)**) [11], [41] have features for model version control with rich associated metadata, facilitating other scientists and engineers to find and reuse AI artifacts. Nevertheless, the metadata attached to the model and accompanying datasets do not capture provenance relationships that allow assessing them in the context of the full AI lifecycle. Unfortunately, even capturing rich details about processes, data, and humans is not enough for reproducibility. For example, one important concern is the non-deterministic behavior inherent to certain ML operators due to arbitrary instruction ordering on GPUs. Provenance tools can store the seed from a random number generator, which helps, but even if the same seed is used to train more than one instance of a model, the resulting model instances may not be identical [42]. Popular packages like PyTorch [43] provide deterministic versions of otherwise non-deterministic functions, although their use may have a negative impact on model performance, thus requiring further trade-off assessment between reproducibility and performance.

(4) Energy and computational resource constraints. Multiple agencies worldwide have identified that AI is too resource-demanding, both for training–particularly when training large models on HPC systems– and for inferences [3]. Government agencies demand sustainable utilization of computational resources to develop AI models and work towards supporting Green AI [23]. However, it is difficult to obtain models that are energy efficient *i.e.*, that optimize model performance within a given amount of resources and usage policy [23]. Making a decision on an optimal model for a given resource-constrained scenario requires combining various data types, such as runtime system data (e.g., telemetry including GPU utilization, memory usage, and temperature variations; CPU metrics; RAM usage; and IO operations), domain-specific key performance indicators, and ML-specific metrics, like loss. The complexity of jointly analyzing these different types of data relates to the discussed data heterogeneity challenge.

Existing profiling tools for system and application instrumentation (Fig. 2 **(4.1)**) are able to capture runtime system

information to provide resource consumption data. Via provenance relationships, these resource consumption data can be explicitly connected to the rest of the workflow data, which can serve as proxies for computing energy efficiency [44]. For instance, GPU temperature, a metric typically available in GPU profiling tools [45], can serve as an indicator of energy efficiency, as models that significantly raise GPU temperatures during training require more energy for cooling compared to those that cause less of an increase [44]. Some works have explored the support of computational resource-aware provenance data management in HPC workflows [46] (Fig. 2 **(4.2)**), but we are not aware of works that target RTAI.

Highlighting the correlation between model design choices and resource usage allows AI scientists and engineers to understand the impact of their decisions on resource consumption. However, capturing, transferring, storing, querying, and inspecting potentially large volumes of provenance data for complex analyses may also contribute to the overall energy footprint and resource consumption. Minimizing the impact of provenance management systems, often expressed as system overhead, is still an open challenge [47], [48] to facilitate informed decision making for a more frugal and environmentally friendly AI model development process.

(5) Infrastructure heterogeneity across the computing continuum. Supporting the RTAI principles demands comprehensive queries across the lifecycle. For example, to aid explainability, one needs to combine model evaluation outputs with data transformation parameters for data preparation [30]. For accountability and reproducibility, query detailed historical data recorded in various lifecycle phases. However, to enable these queries, it is necessary to capture and integrate the data that move across multiple infrastructures in the computing continuum (*i.e.*, edge, cloud and HPC) [12].

Even when not all three forms of the continuum are involved, as is often the case in non-scientific domains, some form of a computing continuum still exists, necessitating data movement between environments. For instance, in a simpler scenario, data might initially be downloaded onto a scientist's desktop for preliminary exploratory analysis, then subsets are moved to an environment with more powerful hardware for data preparation and model fine-tuning.

Existing works capture runtime provenance data from devices in edge computing environments [47], in cloud environments [49], and in HPC environments [9] (Fig. 2 **(5.1)**). However, none of these individual works cover all lifecycle phases. Current methods rely on manual modeling of multiple phases, which is error-prone, time-consuming, and complex, making it harder to adopt at scale.

A potential solution for AI lifecycle provenance management in the computing continuum could involve multiple specialized provenance systems for specific infrastructures and requirements, integrated through a unified data service (Fig. 2 **(5.2)**). A proposal in this direction exists [50], but many challenges remain: capturing, organizing, and querying large data volumes across diverse computing infrastructures; dealing with heterogeneous hardware with varying requirements like bandwidth, security, isolation, and resource constraints; developing data capture systems that may compete with user processes, which is particularly critical in HPC environments; and meeting readiness requirements, such as making captured data immediately available for time-sensitive decisions or allowing longer delays, as in the case of retrieving tracked data for documentation purposes.

IV. Conclusion

In this paper, we highlighted the power of provenance relationships to act as a glue that can capture and transfer important information with varying levels of granularity. This capability spans the different phases of the AI lifecycle and extends across highly heterogeneous, federated, and loosely coupled infrastructures within the computing continuum, taking into account the multiple stakeholders interacting with artifacts like data and models. This power stems from the capability to leverage data capillarity, that is, the property of accessing multiple levels of provenance data–from high-level user interaction information to low-level resource consumption telemetry– into comprehensive analyses.

We argue that these properties of provenance data will play a pivotal role in supporting principles of Responsible, Trustworthy, and Sustainable AI (RTAI)—such as accountability, transparency, explainability, and fairness—in complex sociotechnical scenarios. Ongoing research efforts further highlight the potential of provenance data to aid RTAI, yet we note the major challenge of holistically modeling and managing provenance data from multiple distributed infrastructures and dealing with data heterogeneity. The diversity of stakeholders adds another layer of complexity in capturing the provenance of human interactions and determining the appropriate level of detail required to support various principles. Furthermore, fundamental properties of AI workloads like non-determinism and intensive resource consumption bring specific challenges when addressing principles like reproducibility and energy efficiency, forcing provenance system design not only to adapt to these intricate scenarios, but also to do it under strict storage, network, and computing constraints.

With the goal of addressing these challenges, in future work we will explore the composition of an end-to-end provenance data management system tailored to cover the full AI lifecycle. We will support use cases that require analyses that deal with various data types, focusing on assessing trade-offs between model performance and other RTAI principles, such as explainability, reproducibility, and energy efficiency.

Acknowledgment

This research used the Oak Ridge Leadership Computing Facility's resources at the Oak Ridge National Laboratory, supported by the Office of Science of the U.S. Department of Energy, Contract DE-AC05-00OR22725. In addition, as part of the "France 2030" initiative, this work has benefited from a State grant managed by the French National Research Agency (Agence Nationale de la Recherche) attributed to the STEEL project of the CLOUD PEPR program, reference ANR-23-PECL-0007.

REFERENCES

[1] J. Laux, S. Wachter, and B. Mittelstadt, "Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk," *Regulation & Governance*, vol. 18, no. 1, pp. 3–32, 2024.

[2] U.S.A. Goverment. (2023) Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. [Online]. Available: https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

[3] Oak Ridge National Laboratory. (2024) Oak Ridge National Laboratory initiative in secure, trustworthy, and energy-efficient AI. [Online]. Available: https://www.ornl.gov/ai-initiative

[4] Microsoft Research. (2024) FATE: Fairness, Accountability, Transparency, and Ethics in AI. [Online]. Available: https://www.microsoft.com/en-us/research/theme/fate

[5] A. Van Wynsberghe, "Sustainable AI: AI for Sustainability and the Sustainability of AI," *AI and Ethics*, vol. 1, no. 3, pp. 213–218, 2021.

[6] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, and R. Ranjan, "Explainable AI (XAI): Core Ideas, Techniques, and Solutions," *ACM Computing Surveys*, vol. 55, no. 9, 2023.

[7] S. Samuel, F. Loffler, and B. Konig-Ries, "Machine Learning Pipelines: Provenance, Reproducibility and FAIR Data Principles," in *Provenance and Annotation of Data and Processes*, 2021, pp. 226–230.

[8] P. Groth and L. Moreau. (2020) W3C PROV: an Overview of the PROV Family of Documents. [Online]. Available: https://www.w3.org/TR/prov-overview

[9] R. Souza, L. Azevedo, V. Lourenço, E. Soares, R. Thiago, R. Brandão, D. Civitarese, E. Vital Brazil, M. Moreno, P. Valduriez, M. Mattoso, R. Cerqueira, and M. Netto, "Workflow Provenance in the Lifecycle of Scientific Machine Learning," *Concurrency Computation: Practice and Experience*, pp. 1–21, 2021.

[10] D. De Silva and D. Alahakoon, "An Artificial Intelligence Life Cycle: From Conception to Production," *Patterns*, vol. 3, no. 6, 2022.

[11] H. Miao, A. Li, L. S. Davis, and A. Deshpande, "ModelHub: Deep learning lifecycle management," in *International Conference on Data Engineering*, 2017, pp. 1393–1394.

[12] D. Balouek-Thomert, I. Rodero, and M. Parashar, "Harnessing the Computing Continuum for Urgent Science," *SIGMETRICS Performance Evaluation Review*, vol. 48, no. 2, p. 41–46, 2020.

[13] M. Horsch, B. Schembera, and H. Preisig, "European Standardization Efforts from FAIR toward Explainable-AI-ready Data Documentation in Materials Modelling," in *International Conference on Applied Artificial Intelligence*, 2023.

[14] S. Carroll, E. Herczog, M. Hudson, K. Russell, and S. Stall, "Operationalizing the CARE and FAIR Principles for Indigenous Data Futures," *Scientific data*, vol. 8, no. 1, p. 108, 2021.

[15] V. Dignum, *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer, 2019, vol. 2156.

[16] O. Akinrinola, C. C. Okoye, O. C. Ofodile, and C. E. Ugochukwu, "Navigating and Reviewing Ethical Dilemmas in AI Development: Strategies for Transparency, Fairness, and Accountability," *GSC Advanced Research and Reviews*, vol. 18, no. 3, pp. 050–058, 2024.

[17] C. Wu, Y.-F. Lib, and P. Bouvry, "Survey of Trustworthy AI: A Meta Decision of AI," *arXiv cs.AI*, 2023.

[18] "SHAP Values," https://shap.readthedocs.io/en/latest/, 2024.

[19] M. Ribeiro, S. Singh, and C. Guestrin, "Anchors: High-precision Model-agnostic Explanations," in *AAAI*, vol. 32, no. 1, 2018.

[20] O. Gundersen and S. Kjensmo, "State of the Art: Reproducibility in Artificial Intelligence," in *AAAI*, vol. 32, no. 1, 2018.

[21] M. Hutson, "Artificial Intelligence Faces Reproducibility Crisis," *Science*, vol. 359, no. 6377, pp. 725–726, 2018.

[22] C.-J. Wu *et al.*, "Sustainable AI: Environmental Implications, Challenges and Opportunities," in *ML and Systems Conference*, 2022.

[23] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, "Green AI," *Communications of the ACM*, vol. 63, no. 12, p. 54–63, 2020.

[24] J. Y. Yap and A. Tomlinson, "Provenance-Based Attestation for Trustworthy Computing," in *IEEE Trustcom/BigDataSE/ISPA Conference*, 2015, pp. 630–637.

[25] E. Jandre, B. Diirr, and V. Braganholo, "Provenance in Collaborative in Silico Scientific Research: a Survey," *ACM SIGMOD Record*, vol. 49, no. 2, p. 36–51, 2020.

[26] M. L. Falci, A. Magalhães, A. Paes, V. Braganholo, and D. de Oliveira, "Multimodal Provenance-based Analysis of Collaboration in Business Processes," *JIDM*, vol. 12, no. 5, 2021.

[27] B. Javed, Z. Khan, and R. McClatchey, "An Adaptable System to Support Provenance Management for the Public Policy-Making Process in Smart Cities," *Informatics*, vol. 5, no. 1, 2018.

[28] A. Kale, T. Nguyen, F. Harris, C. Li, J. Zhang, and X. Ma, "Provenance Documentation to Enable Explainable and Trustworthy AI: A Literature Review," *Data Intelligence*, vol. 5, no. 1, pp. 139–162, 2023.

[29] K. Werder, B. Ramesh, and R. Zhang, "Establishing Data Provenance for Responsible Artificial Intelligence Systems," *ACM Transactions on Management Information Systems*, vol. 13, no. 2, pp. 1–23, 2022.

[30] R. L. de Oliveira, J. C. Duarte, and K. de Faria Cordeiro, "Machine Learning Model Explainability supported by Data Explainability: a Provenance-Based Approach," *JIDM*, vol. 15, no. 1, 2024.

[31] T. Nakagawa, K. Narita, and K.-S. Kim, "How Provenance helps Quality Assurance Activities in AI/ML Systems," in *International Conference on AI-ML Systems*, 2023.

[32] D. Lange, "Autonomous Decision Provenance as a Requirement for Building Trust," in *Disruptive Technologies in Information Sciences VI*, vol. 12117, 2022, p. 1211706.

[33] P. Missier and R. Torlone, "From Why-Provenance to Why+Provenance: Towards Addressing Deep Data Explanations in Data-Centric AI," in *Symposium on Advanced Database Systems*, 2024.

[34] D. Pina, A. Chapman, L. Kunstmann, D. de Oliveira, and M. Mattoso, "DLProv: A Data-Centric Support for Deep Learning Workflow Analyses," in *Workshop on Data Management for End-to-End Machine Learning*, 2024, p. 77–85.

[35] B. Richardson and J. E. Gilbert, "A Framework for Fairness: A Systematic Review of Existing Fair AI Solutions," *ArXiv cs.AI*, 2021.

[36] M. Mattoso, J. Dias, K. Ocaña, E. Ogasawara, F. Costa, F. Horta, V. Silva, and D. de Oliveira, "Dynamic Steering of HPC Scientific Workflows: a Survey," *FGCS*, vol. 46, pp. 100–113, 2015.

[37] R. Souza, V. Silva, J. Camata, A. Coutinho, P. Valduriez, and M. Mattoso, "Keeping Track of User Steering Actions in Dynamic Workflows," *Futute Generation Computer Systems*, vol. 99, pp. 624–643, 2019.

[38] J. Rogers and A. Crisan, "Tracing and Visualizing Human-ML/AI Collaborative Processes through Artifacts of Data Work," in *CHI Conference on Human Factors in Computing Systems*, 2023.

[39] R. Han, M. Zheng, S. Byna, H. Tang, B. Dong, D. Dai, Y. Chen, D. Kim, J. Hassoun, and D. Thorsley, "PROV-IO+: A Cross-Platform Provenance Framework for Scientific Data on HPC Systems," *IEEE Transactions on Parallel & Distributed Systems*, vol. 35, no. 05, pp. 844–861, 2024.

[40] V. Silva, V. Campos, T. Guedes, J. Camata, D. de Oliveira, A. L. Coutinho, P. Valduriez, and M. Mattoso, "DfAnalyzer: Runtime Dataflow Analysis Tool for Computational Science and Engineering applications," *SoftwareX*, vol. 12, p. 100592, 2020.

[41] Hugging Face – Model Hub. [Online]. Available: https://huggingface.co/docs/hub/en/models-the-hub

[42] M. Srivastava, S. Arora, and D. Boneh, "Optimistic verifiable training by controlling hardware nondeterminism," *arxiv cs.CR*, 2024.

[43] "Reproducibility — PyTorch 2.3 documentation." [Online]. Available: https://pytorch.org/docs/stable/notes/randomness.html

[44] S. Mittal and J. Vetter, "A Survey of Methods for Analyzing and Improving GPU Energy Efficiency," *ACM Computing Surveys*, 2014.

[45] AMD. (2024) Library for AMD GPU Profiling. [Online]. Available: https://rocm.docs.amd.com/projects/amdsmi

[46] J. Kumar, M. C. Crow, R. Devarakonda, M. Giansiracusa, K. Guntupally, J. Olatt, Z. Price, H. Shanafield, and A. Singh, "Provenance–aware Workflow for Data Quality Management and Improvement for Large Continuous Scientific Data Streams," in *IEEE Big Data*, 2019.

[47] D. Rosendo, M. Mattoso, A. Costan, R. Souza, D. Pina, P. Valduriez, and G. Antoniu, "ProvLight: Efficient Workflow Provenance Capture on the Edge-to-Cloud Continuum," in *IEEE CLUSTER*, 2023, pp. 221–233.

[48] R. Souza, L. Azevedo, R. Thiago, E. Soares, M. Nery, M. Netto, E. V. Brazil, R. Cerqueira, P. Valduriez, and M. Mattoso, "Efficient Runtime Capture of Multiworkflow Data Using Provenance," in *IEEE eScience*, 2019.

[49] D. Rosendo, P. Silva, M. Simonin, A. Costan, and G. Antoniu, "E2Clab: Exploring the computing continuum through repeatable, replicable and reproducible edge-to-cloud experiments," in *IEEE CLUSTER*, 2020.

[50] R. Souza, T. J. Skluzacek, S. R. Wilkinson, M. Ziatdinov, and R. F. da Silva, "Towards Lightweight Data Integration using Multi-workflow Provenance and Data Observability," in *IEEE eScience*, 2023.