DOI: 10.11576/dataninja-1171

Improving Trust in AI Through Sustainable and Trustworthy Reporting

Raphael Fischer RAPHAEL.FISCHER@TU-DORTMUND.DE
Lamarr Institute for Machine Learning and Artificial Intelligence, TU Dortmund University, Germany

Mirko Bunse MIRKO.BUNSE@CS.TU-DORTMUND.DE
Lamarr Institute for Machine Learning and Artificial Intelligence, TU Dortmund University, Germany

Abstract

This extended abstract outlines STREP, our (S)ustainable and (T)rustworthy (REP)orting framework. It communicates performance indicators of systems that build on artificial intelligence and thus makes them more trustworthy.

Keywords: Trustworthy AI, Sustainability, Resource-awareness

1. Introduction

While artificial intelligence (AI) and machine learning (ML) are ubiquitous tools in various domains, their trustworthiness is frequently called into question [1]. One important factor for increasing trust in AI systems resides in communicating novel technological advances and results to the users of such systems. Current ways of reporting the performance of an AI system, however, often produce outcomes that are hard to reproduce, lack information on the computing setup and on the resource usage, and focus on expert users instead of non-experts. To address these issues, we recently proposed the STREP framework as an important step towards more (S)ustainable and (T)rustworthy (REP)orting [2].

This extended abstract and the associated poster summarize the key points of STREP, such as customizable reporting options, interactive controls, and labels for more abstract communication. Our work highlights the importance of resource efficiency, interactivity, comprehensibility, usability, and reproducibility in ML reporting. Through these efforts, we advance the state-of-the-art in ML reporting by promoting sustainability [3], trustworthiness, and user-centric design. We also discuss STREP within the broader context of the triangular research vision—a joint consideration of data, knowledge, and context—that we pursue at the Lamarr Institute.

2. Sustainable and Trustworthy Reporting

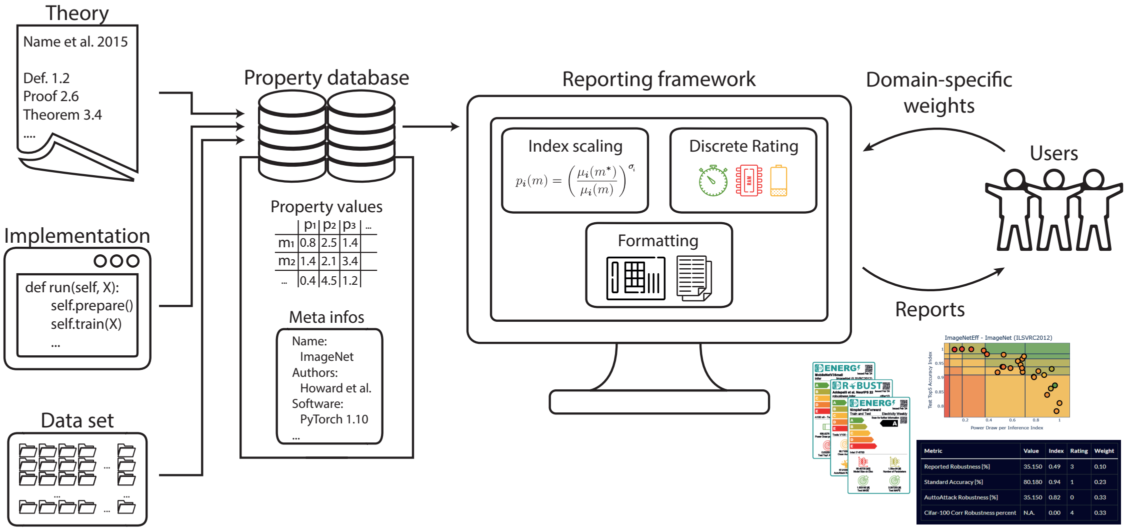
Reporting the performance of a ML system requires a thorough characterization of the corresponding experiments and results. In STREP, schematically displayed in Figure 1, we denote an experimental evaluation setup as a tuple (d, t, m), which corresponds to solving a specific task t on given data d via some method m. An example would be to classify (t) a fixed number of ImageNet images (d) with MobileNetV2 (m) [4]. An evaluation of this kind results in a trained model with specific properties $\boldsymbol{p}_{(d,t,e)}$ which describe the predictive quality (e.g., accuracy, error) of the model and its resource demand (e.g., number of parameters, energy draw).

Unavoidably, these properties are subject to the execution environment e, i.e., to the software and hardware that are used during the evaluation; therefore, they are hard to compare across different execution environments. STREP solves this issue via relative index scaling, a mapping of all real-valued properties onto the unit scale to allow for straightforward comparisons and aggregations.

Since a user might not find all properties of a model to be equally relevant, ML reporting has to offer an interactive investigation of the underlying results. STREP allows uses to control the importance of method properties for their overall performance assessment, hence enhancing user engagement and supporting the understanding of the reported results. To benefit also non-expert audiences, STREP supports the generation of high-level ML labels [5] that can inform users in a more abstract and more easily comprehensible way.

We have used STREP to gain insights into existing benchmarks and open databases from various domains. Our experiments showcase how dramatically

© Fischer & M. Bunse. Licensed under CC BY 4.0.



	p1	p2	p3	...
m1	0.8	2.5	1.4	
m2	1.4	2.1	3.4	
...	0.4	4.5	1.2	

Metric	Value	Index	Rating	Weight
Reported Robustness [%]	35.150	0.49	3	0.10
Standard Accuracy [%]	80.180	0.94	1	0.23
AuttoAttack Robustness [%]	35.150	0.82	0	0.33
Cifar-100 Corr Robustness percent	N.A.	0.00	4	0.33

Figure 1: The proposed framework for sustainable and trustworthy reporting, originally presented in [2]

under-reported resource usage is in public databases like Papers With Code. At the same time, our experiments demonstrate how well a thorough reporting on resource demand can improve the understanding of model performance.

3. Trustworthy and Resource-Aware AI at the Lamarr Institute

STREP is a prime example of the research vision that we pursue at the Lamarr Institute for Machine Learning and Artificial Intelligence. We believe that AI systems need to be designed and implemented along three dimensions: data, knowledge, and context. This understanding can also be seen in STREP - when reporting on empirical performance measurements (data), the context of the experiments (e.g., execution environment) as well as the knowledge of the target audience need to be specifically considered.

In addition to the systematic reporting of AI performance, our institute also investigates trustworthy AI in terms of interpretability, explainability, and ethics, as well as resource-aware AI. We address these topics in diverse application fields and interdisciplinary research areas.

References

[1] Nicole Krämer, Magdalena Wischnewski, and Emmanuel Müller. Interacting with autonomous systems and intelligent algorithms–new theoretical considerations on the relation of understanding and trust. *PsyArXiv*, 2023.

[2] Raphael Fischer, Thomas Liebig, and Katharina Morik. Towards more sustainable and trustworthy reporting in machine learning. *Data Mining and Knowledge Discovery*, 2024.

[3] Aimee van Wynsberghe. Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics*, 2021.

[4] Raphael Fischer, Matthias Jakobs, Sascha Mücke, and Katharina Morik. A unified framework for assessing energy efficiency of machine learning. In *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, 2022.

[5] Katharina J. Morik, Helena Kotthaus, Raphael Fischer, Sascha Mücke, Matthias Jakobs, Nico Piatkowski, Andreas Pauly, Lukas Heppe, and Danny Heinrich. Yes we care!-certification for machine learning methods through the care label framework. *Frontiers in Artificial Intelligence*, 2022.