FAIR principles for AI models: demonstration of an open-source platform for FAIR reporting of AI models

Johan VAN SOEST[a,b,1], Sander VAN ESSEL[a], Carlos AGUILERA MEDINA[a], Chris VAN DER LANS[a], Shashank CHAKRAVARTHY[a], Yenisel PLASENCIA CALAÑA[a], Judith KAMALSKI[a], Andre DEKKER[a,b]

[a] *Brightlands Institute for Smart Society (BISS), Faculty of Science and Engineering, Maastricht University, Heerlen, The Netherlands*

[b] *Department of Radiation Oncology (Maastro), GROW Research Institute for Oncology and Reproduction, Maastricht University Medical Centre+, Maastricht, The Netherlands*

ORCiD ID: Johan van Soest 0000-0003-2548-0330; Carlos Aguilera Medina 0009-0005-8347-7684; Chris van der Lans 0000-0003-4515-2385; Shashank Chakravarthy 0000-0003-3336-8402; Yenisel Plasencia Calaña 0000-0001-5173-0394; Judith Kamalski 0000-0003-2547-9604; Andre Dekker 0000-0002-0422-7996

Keywords. FAIR, Machine Learning, Artificial Intelligence, Standardization

1. Abstract

The number of Artificial Intelligence (AI) or Machine Learning (ML) models is growing rapidly, and keeping up with developments is hard. In this demonstration we present a platform which can implement FAIR principles for AI models, and make models findable, accessible, interoperable and reusable.

2. Introduction and topic

The number of machine learning models, sometimes referred as artificial intelligence, developed for use in healthcare is rapidly growing. However, implementation and adoption in clinical practice is slow; not only due to regulations such as the Medical Device Regulation (MDR) or Food and Drug Authority (FDA) certifications, but also due to the unknown process of assessing a model before use in clinical practice. Although many reporting guidelines for ML exist in the medical domain [1], these guidelines are primarily geared to reporting in (scientific) publications. These publications are in free-text format, which means models are not directly useable in clinical practice as they are not machine interpretable, and there is no overview of existing models for a specific field or disease.

[1] Corresponding Author: Johan van Soest, email: j.vansoest@maastrichtuniversity.nl

Given the free-text nature of these guidelines, there is no semantic standard to report on aspects such as input and output of the ML model, the type of model (classification, regression, neural network, deep learning etc.) or the model's training and validation performance, or (clinical) acceptance criteria. To overcome these issues, the FAIR principles [2] could help in guiding and capturing descriptions and definitions of AI models using semantically interoperable definitions. After all, ML models are data (a mathematical formula) which can be applied to semantically interoperable, or FAIR data.

For this purpose, we have built the FAIRmodels.org software package. In this software package, users can define metadata descriptions of AI models based on model card descriptions [3], implemented in a CEDAR template [4]. Furthermore, the model itself can be uploaded and described using the Open Neural Network eXchange (ONNX) AI representation standard [5]. Model input and output described in the metadata are afterwards linked to the input and output definitions in the ONNX AI representation. A visual representation of the infrastructure is presented in Figure 1.

All software is available at https://fairmodels.org, where organizations can run their own instance when downloading the software from https://github.com/MaastrichtU-BISS/FAIRmodels-backend.

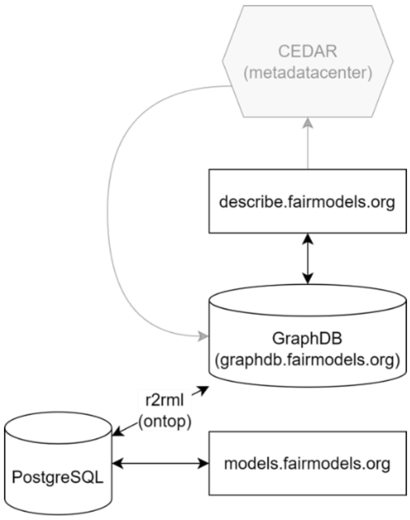

Figure 1 - Infrastructure overview of the fairmodels.org infrastructure, combining a metadata and model repository.

3. Contents of Demonstration

The demonstration will consist of 3 parts. All parts will be demonstrated, whereafter users are able to experiment with the software themselves.

The first part of the demonstration will explain the problem and need for a model repository with structured (meta)data entry and linked to existing clinical terminologies and/or ontologies. During this part, we will show how AI model metadata for a specific prediction model can be stored on the platform.

The second part of the demonstration will explain the concept of standardized AI representations such as ONNX [5] and PMML [6]. We will show and explain what the ONNX representation of an AI model does and does not represent. Hence, this will highlight the gap between AI representations, and the need for clinically useful descriptions of AI models. Afterwards, we will upload an ONNX representation in our platform, and make the link between the AI metadata descriptions and the actual model for the input and output criteria.

Finally, we will show the use of these descriptions. Although this is outside the scope of the platform we will demonstrate, we will show how models described in the platform can be downloaded and used within a hospital, with terminology terms used instead of numerical values which do not represent the semantic meaning of the input / output feature of the model. Finally, we will discuss with the audience the need for, and usefulness of, such tooling to perform commissioning of AI models into clinical practice and reporting of such commissioning activities.

4. Presenters

Johan van Soest is an Assistant Professor at Maastricht University. His research focuses on responsible and translational clinical data science. In his work, he is involved in several (international) research projects regarding federated learning and ethical, legal and societal aspects of data science.

Sander van Essel is a software engineer at Maastricht University. His work focuses on development of (research) software related to data science and AI. His work involves both frontend as well as backend development.

References

[1] C. Garbin, and O. Marques, Assessing Methods and Tools to Improve Reporting, Increase Transparency, and Reduce Failures in Machine Learning Applications in Health Care, *Radiol. Artif. Intell.* **4** (2022) e210127. doi:10.1148/ryai.210127.

[2] M.D. Wilkinson, M. Dumontier, Ij.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. ’t Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data*. **3** (2016) 160018. doi:10.1038/sdata.2016.18.

[3] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I.D. Raji, and T. Gebru, Model Cards for Model Reporting, in: Proc. Conf. Fairness Account. Transpar., Association for Computing Machinery, New York, NY, USA, 2019: pp. 220–229. doi:10.1145/3287560.3287596.

[4] R.S. Gonçalves, M.J. O’Connor, M. Martínez-Romero, A.L. Egyedi, D. Willrett, J. Graybeal, and M.A. Musen, The CEDAR Workbench: An Ontology-Assisted Environment for Authoring Metadata that Describe Scientific Experiments, *Semantic Web--ISWC Int. Semantic Web Conf. Proc. Int. Semantic Web Conf.* **10588** (2017) 103–110. doi:10.1007/978-3-319-68204-4_10.

[5] Open Neural Network eXchange (ONNX), (2024). https://github.com/onnx/onnx (accessed April 11, 2024).

[6] A. Guazzelli, M. Zeller, W.-C. Lin, and G. Williams, PMML: An Open Standard for Sharing Models, *R J*. **1** (2009) 60. doi:10.32614/RJ-2009-010.