








Mantik: A Workflow Platform for the Development of Artificial Intelligence on High-Performance Computing Infrastructures

Thomas Seidler^{1,2*}, Fabian Emmerich^{2,3*}, Kristian Ehlert³, Rico Berner¹, Oliver Nagel-Kanzler¹, Norbert Schultz⁵, Markus Quade¹, Martin G. Schultz⁴, and Markus Abel¹

¹ Ambrosys GmbH, Potsdam, Germany ² Institute of Physics and Astronomy, University of Potsdam, Germany ³ 4Cast GmbH & Co. KG, Potsdam, Germany ⁴ Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich GmbH, Germany ⁵ Reactive Core GmbH, Berlin, Germany ¶ Corresponding author * These authors contributed equally.

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Arfon Smith](#)  

Reviewers:

- [@zhaozhang](#)

Submitted: 18 September 2023

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

The use of machine learning (ML) approaches is exponentially increasing, and for many scientific applications, high-performance computing (HPC) infrastructure is used to train large models. However, the tooling for an easy deployment of models for training or inference on HPC infrastructures is not satisfactory, e.g. reproducibility, collaboration and monitoring of ML models are not addressed in existing toolsets. With Mantik, we provide an open-source cloud platform, which simplifies the development of and collaboration on ML models on HPC facilities, and enhances reproducibility by supporting data and code versioning as well as experiment tracking. The users are able to develop their applications in the environment they are most comfortable with – their local machine. Usage of the best-choice IDE and most recent software versions allow to leverage the full potential of the software stack for their research. Using Mantik's remote file service allows for simple management of data in remote storages and keeping track of it. As soon as an application is ready for training or inference, users can immediately submit it to an HPC cluster. During application development, users can train and/or evaluate their models on HPC clusters via CLI on their local machine or our browser-based Mantik cloud platform. The latter only requires an internet browser such that e.g., ML training from your phone becomes feasible. Once training or inference has begun, a user is able to monitor the application in real time on the Mantik cloud platform.

Statement of need

With Mantik, we cover a variety of features that are at the heart of data science and ML. Mantik streamlines fundamental features that form the core of a data scientist's workflow including easy access to external computing resources, data management and collaborative model development. The interest in platforms for ML operations (MLOps) is on the rise, resulting in a rapidly growing market for solutions. Due to this, we will here not provide references to other existing platforms but refer the interested reader to a recent survey by Kreuzberger et.al. ([Kreuzberger et al., 2023](#)).

The Mantik platform provides a secure interface for an easy and unified access to remote computing resources. The interface is extensible to a variety of infrastructures. Currently, the [UNICORE](#) connector is implemented for resources of the Jülich Supercomputing Centre. Further, Mantik supports the connection to commercial clouds by using their respective

41 interfaces. The unified interface for HPC infrastructures simplifies training and deployment of
42 ML models on different sites. The flexible deployment in Mantik is realized by containerization
43 of the developed ML models, which is standard for commercial clouds to avoid dependencies
44 on hardware. This allows to make use of the desired software packages and versions to fully
45 exploit the potential of state-of-the-art ML frameworks.

46 For data handling Mantik provides a remote file service that allows for storing data remotely,
47 e.g. in a cloud, and tracking the data by keeping its location as a reference and supporting
48 versioning. The data is therefore always available to the users and changes can be traced.
49 Mantik assures that the actual data will only be accessed if necessary to optimize network
50 load and wait times. Analogously, code and ML models can be versioned using git and their
51 references can be stored. Whereas training often involves large data sets and big infrastructure,
52 inference may run on small devices which may require optimized, or compiled code. In order to
53 allow for these different use cases, the deployment and versioning of data, code, and models
54 is fully supported by Mantik. For experiment tracking, our platform integrates the existing
55 open-source framework MLflow ([Chen et al., 2020](#)) to provide researchers with the ability
56 to track their modeling results independent of user hardware. Storage of training workflows,
57 experiments, and models, including records of input parameters, metrics, models, artifacts,
58 ensures reproducibility and enables users to share their ML solutions with the community.

59 A [browser-based GUI](#) is provided for hosting and maintaining your ML projects. Here, ML
60 tasks can be executed codeless on remote computing resources, including scheduling and
61 interacting with submitted jobs (status, properties, log messages, access to file system). The
62 platform allows the user to securely save credentials required to quickly access remote services.
63 All references to existing data, code, models and experiments that are stored in a particular
64 project are displayed on the Mantik web-platform. Project owners can invite other users to
65 their projects to share their results and ML solutions effortlessly, which fosters collaboration
66 and knowledge transfer.

67 Mantik is designed to be used by researchers, data scientists as well as ML engineers. For
68 example, project KI:STE ([Seidler et al., 2021](#)) aims for building a platform particularly based on
69 Mantik for promoting the application of ML in earth system sciences. In the [project Maelstrom](#)
70 Mantik is already used to develop large-scale ML applications for weather and climate science.
71 Due to the simple structure and the variety of built-in features, Mantik provides a kick-start for
72 all users interested in running ML projects on HPC. Consequently, Mantik saves time allowing
73 the user to focus on research and improving their ML approaches rather than bothering about
74 tooling. Mantik allows for easy bookkeeping of key entities (data, models, experiments) that
75 constitute the ML life cycle, therefore providing reproducibility that forms the basis of scientific
76 work and supports its credibility. Moreover, shareable and open ML projects enable students
77 and junior data scientists to familiarize themselves with state-of-the-art approaches. The ease
78 of use and inherent reproducibility therefore renders Mantik also a great choice for education
79 purposes in the field of ML.

80 The Mantik platform architecture

81 In Mantik the organizational entity is the Mantik project. In [Figure 1](#), we provide our vision
82 for the Mantik platform and how the different core units within the Mantik project relate to
83 the typical development lifecycle of ML solutions.

- 84 ■ Data and AI modeling: In Mantik projects you can store references to all ingredients
85 that you would need to develop your ML solution (data, code, infrastructure).
- 86 ■ Training: A training session of an ML algorithm (code) with some given data on a
87 particular infrastructure, e.g. HPC, is organized in a run.
- 88 ■ AI engineering: Many runs form an experiment that can be used to find the best
89 performing ML solution for your particular use case.

- 90 ■ Deployment: Your evaluation of the experiments results in models that are trained
91 algorithms with fixed sets of parameters, trained on known architectures and meant be
92 used for specific types of data, e.g. video data or texts.

Reproducibility

Collaboration

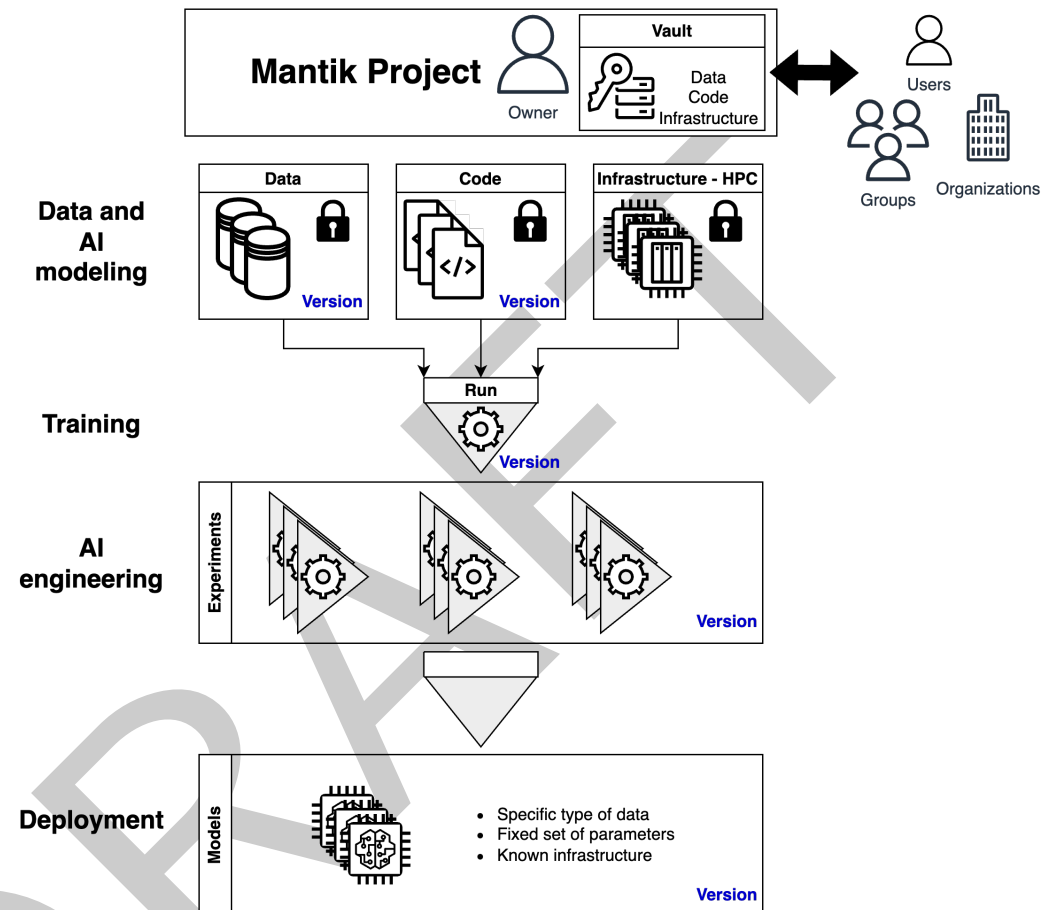


Figure 1: Overview of the organization of the Mantik platform.

Acknowledgements

93 T.S. and M.A. acknowledge financial support from the project KI:STE funded by Bundesmin-
94 isterium für Bildung und Forschung (BMBF). R.B., M.Q. and M.A. acknowledge financial
95 support EMDRIVE and CECAS funded by BMBF. F.E., K.E., M.S. and M.A. gratefully ac-
96 knowledge funding from the MAELSTROM EuroHPC-JU project (JU) under 955513. The JU
97 receives support from the European Union's Horizon Research and Innovation Programme and
98 United Kingdom, Germany, Italy, Luxembourg, Switzerland, and Norway. We are grateful for
99 all contributions to the Mantik platform and in particular to the outstanding people involved
100 in its development: the [Mantik contributors](#).
101

References

- Chen, A., Chow, A., Davidson, A., DCunha, A., Ghodsi, A., Hong, S. A., Konwinski, A., Mewald, C., Murching, S., Nykodym, T., Ogilvie, P., Parkhe, M., Singh, A., Xie, F., Zaharia, M., Zang, R., Zheng, J., & Zumar, C. (2020). Developments in MLflow: A system to accelerate the machine learning lifecycle. *Proceedings of the Fourth International Workshop on Data Management for End-to-End Machine Learning*. <https://doi.org/10.1145/3399579.3399867>
- Kreuzberger, D., Kühl, N., & Hirschl, S. (2023). Machine learning operations (mlops): Overview, definition, and architecture. *IEEE Access*, 11, 31866–31879. <https://doi.org/10.1109/ACCESS.2023.3262138>
- Seidler, T., Schultz, N., Quade, Dr. M., Autermann, C., Gräler, Dr. B., & Abel, P. Dr. M. (2021). *Easing and promoting the application of ML and AI in earth system sciences - introducing the KI:STE platform*. <https://doi.org/10.5194/egusphere-egu21-9632>

DRAFT