

A 10-Minute Questionnaire on Tool Support for Machine Learning Experiments

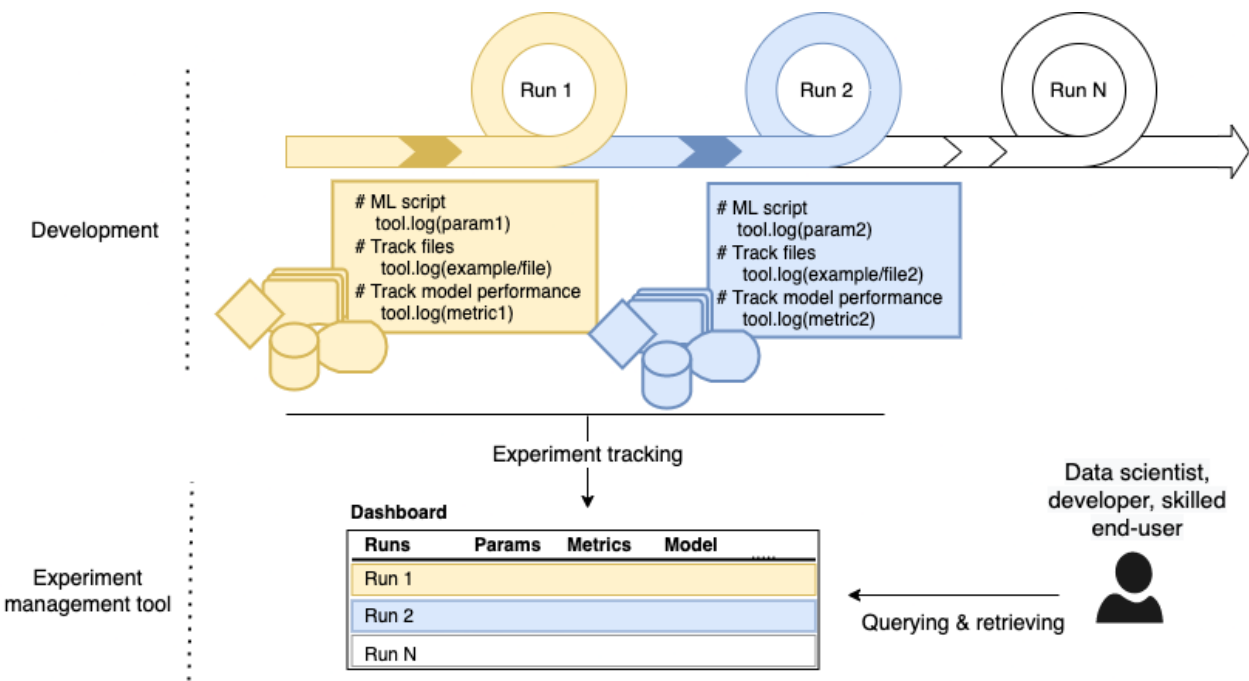
Dear participant,

Since you are an experienced machine learning practitioner, we XXXXXXXXXX
XX would like to hear your opinion on machine learning experiment management tools. We kindly invite you to participate and forward this invitation to further colleagues who might also be interested in this survey. Completing the survey takes approximately 10 minutes.

Experiment management tools support practitioners performing machine learning (ML) or deep learning (DL) experiments to manage all involved artefacts and metadata (datasets, features, scripts, hyperparameters, evaluation metrics, models, ...). Such tools are used to reproduce or trace experiments, analyse experiment results, and collaborate with other practitioners. Popular tools are, for instance, Neptune.ai, DVC, or MLFlow. These tools allow users to track or log artefacts when performing experiment runs. For instance, some tools provide APIs for logging hyperparameters, script versions, metrics, and other artefacts/metadata used during experiments. Some also provide visual user interfaces to later query and visualise experiments, as illustrated below.

***Required**

Illustration of artefact tracking with experiment management tools



This survey aims to elicit information from practitioners on performing ML/DL experiments with and without experiment management tools. If you are not using such tools, you are welcome to report your experiences and challenges with experimentation and give suggestions for improving tooling for ML/DL experiments. If you are using such tools, we aim to investigate the management tools you use, the benefits you perceive, as well as their challenges.

+ What's in it for you?

As a participant, you will learn about experiment management tools, their features and benefits, and how they can be valuable for your own projects. Also, you can receive a state-of-the-art report with the study results to learn more about these tools and their trends. Consequently, you can reflect on your practices and learn about others' practices.

+ What happens to your data?

All collected data will be stored securely and analysed, with information from participants aggregated and reported in an anonymised form for a scientific publication (the state-of-the-art report).

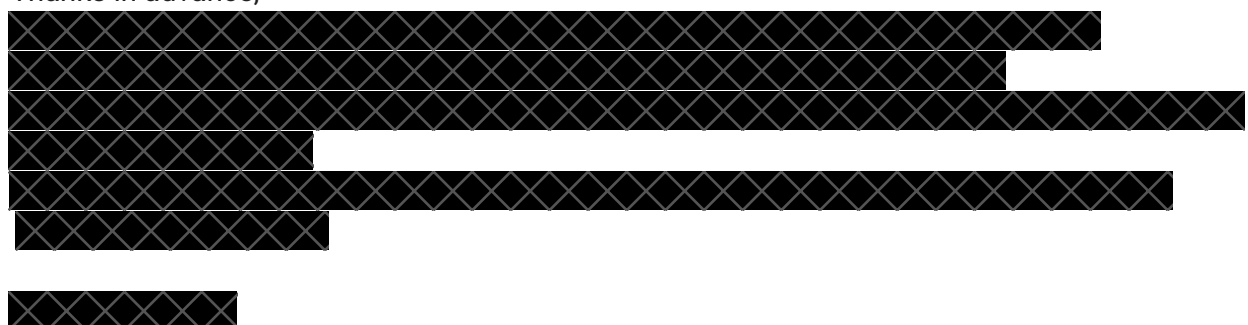
+ What's in it for us?

We will use the collected data to understand the landscape of ML/DL experiment management tools and identify their actual value to practitioners. We will also derive and suggest improvements based on identified challenges.

+ What's in it for everyone else?

Your contribution will benefit both the research and industrial communities to direct future research and eventually lead to better practices and tools. Researchers will obtain information to form systematic knowledge on ML experiment management in ML/DL engineering. Practitioners will receive guidance to design and select better ML experiment management tools for their projects.

Thanks in advance,



1. Informed consent *

Tick all that apply.

- ☐ I hereby consent to participate in this survey. I acknowledge that participation is voluntary.
- ☐ I have read the introduction, and I understand the purpose of this survey.

2. Do you perform ML/DL experiments (a.k.a., model prototyping)? (Further questions will vary based on your answer) *

Mark only one oval.

- ☐ Yes
- ☐ No *Skip to section 10 (Conclusion)*

Participants Performing ML/DL Experiments

3. In terms of ML/DL model training and evaluation, which form of experimentation do you perform? *

Tick all that apply.

- ☐ Automated runs (e.g., I use training loops to perform the iterative task of finding optimal models)
- ☐ Manual runs (e.g., after training and evaluating a model from a run, I inspect the results and decide on changes for the next experiment run)

4. What is the largest number of experiment runs you have ever performed in a project? *

Mark only one oval per row.

	1 - 10	10 - 25	25 - 50	50 - 100	Over 100
Number of runs:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5. If your number of runs is over 100, please estimate:

6. Briefly describe the purpose of your ML/DL experiments and the type of ML you use.

7. Do you use experiment management tools, such as those listed in the next question? (Further questions will vary based on your answer) *

Mark only one oval.

☐ Yes *Skip to question 15*

☐ No *Skip to question 9*

8. If yes, which of the following experiment management tools do you use?

Tick all that apply.

- ☐ Custom (in-house)
- ☐ Weights & Biases (Wand)
- ☐ Neptune
- ☐ Comet.ml
- ☐ Sacred
- ☐ MLFlow
- ☐ TensorBoard
- ☐ TFX
- ☐ Polyaxon
- ☐ DVC
- ☐ ClearML
- ☐ Valohai
- ☐ Pychyderm
- ☐ Google collab
- ☐ Kubeflow
- ☐ Kubeflow_Kale
- ☐ Determine.ai
- ☐ Sumatra
- ☐ StudioML
- ☐ Guild AI
- ☐ Feature Forge
- ☐ Deepkit
- ☐ Dot Science
- ☐ Allegro Trains
- ☐ Verta.ai
- ☐ Datmo
- ☐ Codalab
- ☐ SageMaker Studio
- ☐ Aim
- ☐ Apache Liminal
- ☐ Kedro
- ☐ Metaflow
- ☐ MLAide
- ☐ Lightning-Grid
- ☐ RapidMiner
- ☐ Other: _____

Participants who do not
use experiment
management tools

Now, in the following questions, we are interested in
your personal experiences with ML/DL experiments.

9. Are you aware of experiment management tools, such as those mentioned in the previous section?

Mark only one oval.

☐ Yes

☐ No

10. If yes, why are you not using such tools?

11. How do you manage versions of your experiment artefacts and metadata?

Mark only one oval.

☐ Do not manage versions

☐ Use of dedicated naming convention for folders and files

☐ Use of Git

☐ Use of other version control system

☐ Use of custom database

☐ Other: _____

12. What are the challenges you face in managing artefacts/metadata during or after ML/DL experimentation? Challenges are aspects that make artefact and metadata management difficult.

13. To which extent do you agree with the following statement: "Specialised experiment management tools can improve artefacts and metadata management during ML/DL experiments/prototyping"

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree

14. In terms of tool support, in what other ways do you think ML/DL experiment management can be improved?

Skip to question 25

Participants
who use
experiment
management
tools

Now, in the following questions, we are interested in your personal experiences with respect to the tool(s) you have selected in previous section.

15. In which ML/DL workflow stages do you use your selected experiment management tool(s)?

Tick all that apply.

- ☐ Data preparation
- ☐ Model building
- ☐ Model deployment
- ☐ Other: _____

16. To which extent do you agree with the following statements:

Mark only one oval per row.

	Strongly Disagree (1)	(2)	(3)	(4)	Strongly Agree (5)
Experiment management tools facilitate my ML/DL tasks well	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Experiment management tools are easy to learn and use	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Experiment management tools make me perform experiments more efficiently	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Experiment management tools improve the performance of my models	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall, they provide a benefit to me, compared to not using an experiment management tool	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A simple command-line interface (like Git) is sufficient for querying and making analysis of tracked artefacts and metadata	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A GUI dashboard is essential for querying and analysis of tracked artefacts and metadata	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

I prefer dedicated tools offering strictly experiment management features over multi-purpose tools with additional features

☐☐☐☐☐

17. If applicable, where do you see the benefits/values of the experiment management tool(s) that you use?

Tick all that apply.

- ☐ Time savings
- ☐ Collaboration
- ☐ Reproducibility (i.e. ability to reproduce prior experiment models)
- ☐ Traceability (i.e. ability to map model behaviour to concrete experiment/development assets; for example, fetching specific assets from prior runs based on some conditions)
- ☐ Replicability (re-using prior experiments in a different context. E.g., new datasets)
- ☐ Result analysis
- ☐ Result & model optimization
- ☐ Other: _____

18. Which feature(s) of your experiment management tool(s) do you find important?

Tick all that apply.

- ☐ Versioning
- ☐ Querying
- ☐ Visualisation
- ☐ Computational resources
- ☐ Pipeline support
- ☐ Language agnostic
- ☐ Git integration support
- ☐ Dependency management
- ☐ SaaS (cloud-based)
- ☐ Other: _____

19. Which experiment artefacts do you consider most important to manage?

Tick all that apply.

- ☐ Dataset, data feature & metadata
- ☐ Scripts / source code & metadata
- ☐ Computation & execution data (metrics, logs)
- ☐ Parameters and configuration
- ☐ Models and metadata
- ☐ Pipelines
- ☐ Other: _____

20. Which form/interface of artefact/metadata tracking do you prefer? (e.g., tracking via API in scripts, command-line interface)

Limitations & Challenges

21. The experiment management tool(s) I use has limitations affecting my experiments.

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree

22. What particular limitations of the tool(s) did you experience?

23. When using your selected experiment management tool(s), which challenges did you experience? Challenges are aspects that make using the tool difficult.

24. In what other ways do you think ML/DL experiment management tools can be improved?

Participant
Info

To put your responses into context, we would appreciate if you share your role and experiences with ML/DL experiments with us.

25. What are your current roles?

Tick all that apply.

- ☐ Data scientist
☐ ML engineer
☐ Data engineer
☐ Software engineer
☐ Other: _____

26. How many years of experience do you have working with ML/DL experiments?

27. In which domains do you currently work?

Tick all that apply.

- ☐ Gaming
- ☐ Telecoms
- ☐ Transportation
- ☐ Education
- ☐ Finance
- ☐ Media
- ☐ Consumer retail
- ☐ Health
- ☐ Technology
- ☐ Other: _____

Contact Information

Optional contact information

28. Would you like to receive the state-of-practice report based on the findings from this survey?

Tick all that apply.

- ☐ Yes

29. Would you be available for follow-up questions about your answers?

Tick all that apply.

- ☐ Yes

30. If you answered "yes" to either of the last two questions, please specify your e-mail address.

31. If you want us to address you personally in follow-up communication, please enter your name.

Conclusion

Thanks for your participation! Your input is very valuable to the community.

Conclusion

This survey requires information from participants with ML/DL experiment experience, so we will conclude with you here.

Thanks for your participation! Your input is very valuable to the community.

This content is neither created nor endorsed by Google.

Google Forms