Human Factors, Business Management and Society, Vol. 56, 2022, 137–144
https://doi.org/10.54941/ahfe1002261

Multidisciplinary Teamwork in Machine Learning Operations (MLOps)

Tapani Honkanen[1], Jonny O'Dwyer[2], and Vesa Salminen[1]

[1]Häme University of Applied Sciences, Vankanlähde 9,13100 Hämeenlinna, Finland
[2]Technological University of the Shannon: Midlands Midwest, University Road, Athlone, Co. Westmeath, N37HD68, Ireland

ABSTRACT

Machine learning operations (MLOps) is an emerging and complex subject area involving experts from several fields and backgrounds. Its main purpose is to enable a more standardized and effective approach to building and maintaining machine learning systems. Machine learning projects have an extremely high failure rate. One of the reasons behind this is the lack of teams designed for building these systems. At the same time, machine learning projects can carry great business risks. This paper takes a scoping review approach in assessing the state of the current literature about multidisciplinary teamwork within the context of MLOps. Most of the literature reviewed on collaboration and teamwork focuses on the intimately related field of data science. These articles are analyzed, and a synthesis is presented of the gaps in the current literature for collaboration within data science. Recommendations for further research directions are given for MLOps.

Keywords: Machine learning operations, Multidisciplinary teamwork, Software engineering, Collaboration, Data science

INTRODUCTION

Development, deployment, and maintenance of different types of machine learning (ML) systems is a challenging task that typically requires expertise from several different fields. Gartner estimate (Van Der Meulen and McCall, 2018) that nearly 85% of artificial intelligence (AI) projects will fail due to delivering erroneous outcomes as a result of bias in data, algorithms or the teams responsible for the development work being done. More recently, Gartner (Rimol, 2021) found that only half of AI related projects make it to production. They also estimate that issues such as transparency, auditability and building of trust will become increasingly important in the field.

Machine learning operations (MLOps) as a field has developed because of the challenges in ML systems development and productionalization. One could argue, that in essence, MLOps is about standardization and simplification of the ML lifecycle in such a way that allows reproducibility between the different development and deployment phases, reduction of business-related risks, and an increase in the success rate of developing, deploying, and monitoring production level ML systems. The field is complementary to data science and aims to support the work of data scientists and other professionals working with these systems. Data science has been defined as the study of data which combines several different fields such as statistics, informatics, communication, and sociology which aims to produce data products as it's outputs (Cao, 2018).

Haakman et al. (2021) have recently proposed that typical ML lifecycle models, such as the Cross-Industry Standard Process for Data Mining (CRISP-DM) or the Team Data Science Process (TDSP), need to be revised with additional steps to further accommodate for feasibility assessment, documentation, evaluation of models, and monitoring of live production models. It is important to note that in many cases, most of these proposed additional steps might require a multidisciplinary approach to be effective. Studer et al. (2021) proposed CRISP-ML(Q) which extends the CRISP-DM model to accommodate for the requirements of ML development, especially in terms of quality assurance. The latter lifecycle model is also endorsed by practitioners within the field of MLOps (Visengeriyeva et al., nd).

Giray (2021) reviewed the literature on the challenges of building ML systems from the perspective of software engineering and suggested that it would be beneficial to have more research on a set of harmonized practices to address both ML and software engineering perspectives, practices for ML project management, and on how to form coherent teams. Motivated by the need to advance current knowledge in the important and nascent field of MLOps, the research questions (RQ) for the present study are:

RQ 1: What is the current state of the literature on multidisciplinary teamwork in MLOps?

RQ 2: Could some of the current findings on collaboration and teamwork in the field of data science teams be extended to MLOps?

DATA COLLECTION AND SELECTION CRITERIA

The search for articles from databases was carried out during the November and December of 2021. Searched databases include Goodle Scholar, Arxiv, Elsevier, IEEE Xplore, SpingerLink, and PapersWithCode. For IEEE Xplore, only free articles were collected. Articles released before 2016 were excluded from this review. Searches were carried out using combinations of the following terms: multidisciplinary, collaboration, team, machine learning, data science, machine learning operations, machine learning systems and mlops. A total of 92 articles were downloaded based on the article title and abstract. After this, articles were excluded in two stages from the collected set of articles. First, article contents were surveyed on a general level, and if article contents did not appear to relate to a research question, the article was excluded from the study. After this, the contents of the included articles were analyzed and, in some cases, exclusion was still done at this point. From all the collected 92 articles, only 23 met inclusion criteria for the present study. The main contents in terms of results of the included articles were summarized into an Excel file for further analysis. These scoping review stages were performed by the primary author of this article.

Early in the data collection process it was observed that the literature on collaboration and multidisciplinary teamwork is mostly related to data science, and not MLOps. This points to several open future research directions. It is also a limitation for this study and its results. Generalizations can be made based on the results of these articles, but it should not be concluded that all the results can be applied to MLOps directly.

RESULTS

There were no articles found that specifically addressed the topic of multidisciplinary teamwork in the field of MLOps. Many of the collected articles approached the subject of collaboration and teamwork from the perspective of data science.

As shown in Table 1, a categorization of the articles that met the inclusion criteria was performed based on the different elements of multidisciplinary teamwork. These elements were synthesized based on the contents of the collected articles, using the following typical article sections: results (or findings), discussion and conclusions. For those articles, where these sections were not present, the categorization was done based on the primary author's assessment of the contents. For an article to be counted in an element, there had to be a contribution that could benefit other researchers or practitioners. In some cases, categorization was done based on the overall contents of the article that contribute to an element. A single article could be included in several different categories based on its contributions.

As stated above, most of the included articles approached the subject of collaboration and multidisciplinary teamwork through the lens of data science. The above results should be viewed as mostly describing gaps in the literature in this context. In general, it is clearly obvious that less attention has been paid to communication, participant skills, and resourcing, as well as leadership and culture. As there is indication that even today, the role of a data scientist is not always clear (Hukkelberg and Berntzen, 2019), it is probable that even the more often studied aspect of team structures and participant roles will see developments as the field matures. The authors suspect that these roles might still change in the future as a result of other factors, such as tooling which might simplify development efforts in the coming years.

The requirement for building multidisciplinary teams is not new in the field, this was already concluded by Baškarada and Koronios (2017). What is surprising, is the fact that although we have standardized methods and tools, there is some evidence that these are not being used by data science teams (Martinez et al., 2021; Khalajzadeh et al., 2020). Also, it has been concluded that current lifecycle models for data science teams should include project-, team- and also data- and information management elements that should be applied along with data science methodologies (Martinez et al., 2021). As was mentioned earlier, others have proposed additions to these lifecycle methodologies as well (Haakman et al., 2021; Visengeriyeva et al., nd). This would expand the scope of the methodologies to cover the entire ML lifecycle, including MLOps oriented concerns such as memory and compute resource management and optimization, quality assurance, model governance and -retraining, and monitoring of live models.

Table 1. Categorization of the articles which met the inclusion criteria based on elements of multidisciplinary teamwork. The elements were synthesized from the collected articles.

Element of Collaboration	Count	Related Articles
Team structures and roles	13	(Baškarada and Koronios, 2017; Ferrero et al., 2020; Verma et al., 2021; Zhang et al., 2020; Mao et al., 2019; Piorkowski et al., 2021; Hukkelberg and Berntzen, 2019; Hind et al., 2019; Antoniou and Mamdani, 2021; McDavid et al., 2021; Martinez et al., 2021; Martín-Noguerol et al., 2021; Maier et al., 2019)
Collaborative tools and technologies	13	(Ferrero et al., 2020; Verma et al., 2021, Zhang et al., 2020; Mao et al., 2019; Piorkowski et al., 2021; van Stijn, nd., Staron et al., 2021; Park et al., 2021; Hukkelberg and Berntzen, 2019; Hind et al., 2019; McDavid et al., 2021; Khalajzadeh et al., 2020; Wang et al., 2019)
Collaborative practices	12	(Ferrero et al., 2020; Staron et al., 2021; Verma et al., 2021; Zhang et al., 2020; Mao et al., 2019; Piorkowski et al., 2021; Park et al., 2021; Hukkelberg and Berntzen, 2019; McDavid et al., 2021; Martinez et al., 2021; Martín-Noguerol et al., 2021; Wang et al., 2019)
Hindrances to teamwork	10	(Arpteg et al., 2018; Zhang et al., 2020; Mao et al., 2019; Piorkowski et al., 2021; Park et al., 2021; Hukkelberg and Berntzen, 2019; Hind et al., 2019; Martinez et al., 2021; Amershi et al., 2019; Passi and Jackson, 2018)
Workflow	8	(Verma et al., 2021; Zhang et al., 2020; Mao et al., 2019; Piorkowski et al., 2021; Antoniou and Mamdani, 2021; McDavid et al., 2021; Martinez et al., 2021; Khalajzadeh et al., 2020)
Effectiveness of communication	5	(van Stijn, nd.; Staron et al., 2021; Piorkowski et al., 2021; Martín-Noguerol, 2021; Khalajzadeh, 2020)
Participant skills	4	(Baškarada and Koronios, 2017; Ferrero et al., 2020; Amershi et al., 2019; Wang et al., 2019)
Resourcing	2	(Ferrero et al., 2020; McDavid et al., 2021)
Leadership and culture	2	(Baškarada and Koronios, 2017; Ferrero et al., 2020)

Although the collected literature mostly deals with the work of data scientists, this does not mean that none of it is relevant to MLOps. The above lifecycle models (Martinez et al., 2021), or common workflows, can be considered as critical tools for a discipline that aims to standardize the entire ML lifecycle. These also form a backbone for collaboration. Verma et al. (2021) considered the whole end-to-end ML development process in healthcare and defined how multidisciplinary teams should be organized for the duration of the whole project lifecycle. They elaborated on how ML systems can be integrated into existing clinical workflows to minimize friction, and how trust can be built in the process. Silent testing is mentioned as a form of deploying the model into production iteratively, which refers to shadow deployments in MLOps terminology. In practice, this means that model testing was done in a live production environment with real data but returned prediction results from the model were not communicated to the clinicians. These were used for making comparisons between clinician made predictions and ML model predictions. Antoniou and Mamdani (2021) continued this research stream by studying the validation of ML systems in healthcare. They presented an end-to-end workflow for system validation and explain why this requires a multidisciplinary approach. Their article highlights that system validation should continue even after deployment to make sure that the ML solution remains in working order. In healthcare data, data drift, where system input data distributions change over time, can occur due to alterations in clinical practice or distributions of patient characteristics. Both issues could be argued to be challenges requiring a multidisciplinary approach.

Zhang et al. (2020) mention that it seems that fairness and bias in data science teams is treated mostly as a technical matter. The authors also state that documentation in relation to feature selection and engineering, bias mitigation, and data, are problematic areas in a data science project. During feature selection and engineering, a lot of hidden decisions can be made that affect the later stages of the ML project. In the same manner, improper or lacking methods for documenting and versioning data can cause significant problems, leading to undesired behavior of the model. Documentation and versioning can be seen as collaborative practices (Piorkowski et al., 2021) that help to create and maintain shared mental models between participants in a highly disjointed process, such as is the end-to-end ML lifecycle. In the view of the authors of the current article, these practices also allow for less error prone development processes, and hence reduced business risk. MLOps can support multidisciplinary teams in this respect by providing both tooling and practices to improve the effectiveness of communication and providing consideration for the end-to-end ML lifecycle throughout all project phases.

CONCLUSION

MLOps is a rapidly developing field which aims to improve the success rate of ML projects through standardization and simplification of the end-to-end ML lifecycle. As with ML in general, this field is highly multidisciplinary. One of the main findings of this research is that collaborative aspects of developing and maintaining ML systems has received little attention in the literature. Much of the reviewed work focuses on the collaborative work between data scientists and domain experts. As practitioners are facing challenges in building and maintaining complex systems requiring multidisciplinary expertise, as indicated by Arpteg et al., (2018) and Dey and Lee (2021), the authors would see it beneficial for researchers to generalize based on current best practices within the industry. It is also important to expand the research focus to cover the whole end-to-end ML lifecycle along the development and adoption of newer lifecycle models, such as CRISP-ML(Q) (Visengeriyeva et al., nd). Another possible aspect for further research is to investigate why current methodologies and tools are not being applied by practitioners, as suggested by Martinez et al. (2021) and Khalajzadeh et al. (2020), and how this could be improved. If it holds true, this alone will potentially bring improvements in the success rate of ML projects.

REFERENCES

Amershi S. *et al.*, (May 2019) "Software Engineering for Machine Learning: A Case Study," in *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, Montreal, QC, Canada, pp. 291–300. doi: 10.1109/ICSE-SEIP.2019.00042.

Antoniou T. and Mamdani M., (Sep. 2021) "Evaluation of machine learning solutions in medicine," *Can. Med. Assoc. J.*, vol. 193, no. 36, pp. E1425–E1429, oi: 10.1503/cmaj.210036.

Arpteg A., Brinne B., Crnkovic-Friis L., and Bosch J., (Aug. 2018) "Software Engineering Challenges of Deep Learning," *2018 44th Euromicro Conf. Softw. Eng. Adv. Appl. SEAA*, pp. 50–59, doi: 10.1109/SEAA.2018.00018.

Baškarada S. and Koronios A., (Apr. 2017) "Unicorn data scientist: the rarest of breeds," *Program*, vol. 51, no. 1, pp. 65–74, doi: 10.1108/PROG-07-2016-0053.

Cao L., (May 2018) "Data Science: A Comprehensive Overview," *ACM Comput. Surv.*, vol. 50, no. 3, pp. 1–42, doi: 10.1145/3076253.

Dey S. and Lee S.-W., (Jun. 2021) "Multilayered review of safety approaches for machine learning-based systems in the days of AI," *J. Syst. Softw.*, vol. 176, p. 110941, doi: 10.1016/j.jss.2021.110941.

Ferrero E. *et al.*, (Aug. 2020) "Ten simple rules to power drug discovery with data science," *PLOS Comput. Biol.*, vol. 16, no. 8, p. e1008126, doi: 10.1371/journal.pcbi.1008126.

Giray G., (Oct. 2021) "A software engineering perspective on engineering machine learning systems: State of the art and challenges," *J. Syst. Softw.*, vol. 180, p. 111031, doi: 10.1016/j.jss.2021.111031.

Haakman M., Cruz L., Huijgens H., and van Deursen A., (Jun. 2021) "AI Lifecycle Models Need To Be Revised. An Exploratory Study in Fintech," *ArXiv201002716 Cs*, Accessed: Oct. 17, 2021. [Online]. Available: http://arxiv.org/abs/2010.02716

Hind M. *et al.*, (Nov. 2019) "Experiences with Improving the Transparency of AI Models and Services," *ArXiv191108293 Cs*, Accessed: Jan. 14, 2022. [Online]. Available: http://arxiv.org/abs/1911.08293

Hukkelberg I. and Berntzen M., (2019) "Exploring the Challenges of Integrating Data Science Roles in Agile Autonomous Teams," in *Agile Processes in Software Engineering and Extreme Programming – Workshops*, vol. 364, R. Hoda, Ed. Cham: Springer International Publishing, pp. 37–45. doi: 10.1007/978-3-030-30126-25.

Khalajzadeh H., Simmons A. J., Abdelrazek M., Grundy J., Hosking J., and He Q., (Jun. 2020) "An end-to-end model-based approach to support big data analytics development," *J. Comput. Lang.*, vol. 58, p. 100964, doi: 10.1016/j.cola.2020.100964.

Maier T., DeFranco J., and Mccomb C., (Oct. 2019) "An analysis of design process and performance in distributed data science teams," *Team Perform. Manag. Int. J.*, vol. 25, no. 7/8, pp. 419–439, doi: 10.1108/TPM-03-2019-0024.

Mao Y. *et al.*, (Dec. 2019) "How Data Scientists Work Together With Domain Experts in Scientific Collaborations: To Find The Right Answer Or To Ask The Right Question?," *Proc. ACM Hum.-Comput. Interact.*, vol. 3, no. GROUP, pp. 1–23, doi: 10.1145/3361118.

Martinez I., Viles E., and Olaizola I. G., (May 2021) "Data Science Methodologies: Current Challenges and Future Approaches," *Big Data Res.*, vol. 24, p. 100183, doi: 10.1016/j.bdr.2020.100183.

Martín-Noguerol T., Paulano-Godino F., López-Ortega R., Górriz J. M., Riascos R. F., and Luna A., (May 2021) "Artificial intelligence in radiology: relevance of collaborative work between radiologists and engineers for building a multidisciplinary team," *Clin. Radiol.*, vol. 76, no. 5, pp. 317–324, doi: 10.1016/j.crad.2020.11.113.

McDavid A. *et al.*, (2021) "Eight practices for data management to enable team data science," *J. Clin. Transl. Sci.*, vol. 5, no. 1, p. e14, doi: 10.1017/cts.2020.501.

Park S., Wang A., Kawas B., Liao Q. V., Piorkowski D., and Danilevsky M., (Jan. 2021) "Facilitating Knowledge Sharing from Domain Experts to Data Scientists for Building NLP Models," *ArXiv210200036 Cs*, Accessed: Dec. 27, 2021. [Online]. Available: http://arxiv.org/abs/2102.00036

Passi S. and Jackson S. J., (Nov. 2018) "Trust in Data Science: Collaboration, Translation, and Accountability in Corporate Data Science Projects," *Proc. ACM Hum.-Comput. Interact.*, vol. 2, no. CSCW, pp. 1–28, doi: 10.1145/3274405.

Piorkowski D., Park S., Wang A. Y., Wang D., Muller M., and Portnoy F., (Apr. 2021) "How AI Developers Overcome Communication Challenges in a Multidisciplinary Team: A Case Study," *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CSCW1, pp. 1–25, doi: 10.1145/3449205.

Rimol, M. (Jul. 09, 2021) "Gartner Identifies Four Trends Driving Near-Term Artificial Intelligence Innovation," *Gartner Identifies Four Trends Driving Near-Term Artificial Intelligence Innovation*. https://www.gartner.com/en/newsroom/press-releases/2021-09-07-gartner-identifies-four-trends-driving-near-term-artificial-intelligence-innovation (accessed Jan. 22, 2022).

Staron M. *et al.*, (Mar. 2021) "Robust Machine Learning in Critical Care -- Software Engineering and Medical Perspectives," *ArXiv210308291 Cs*, Accessed: Dec. 11, 2021. [Online]. Available: http://arxiv.org/abs/2103.08291

Studer S. *et al.*, (Apr. 2021) "Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology," *Mach. Learn. Knowl. Extr.*, vol. 3, no. 2, pp. 392–413, doi: 10.3390/make3020020.

Van Der Meulen R. and McCall T., (2018 Feb. 13) "Gartner Says Nearly Half of CIOs Are Planning to Deploy Artificial Intelligence," *Gartner Says Nearly Half of CIOs Are Planning to Deploy Artificial Intelligence*. https://www.gartner.com/en/newsroom/press-releases/2018-02-13-gartner-says-nearly-half-of-cios-are-planning-to-deploy-artificial-intelligence (accessed Jan. 22, 2022).

van Stijn J. J., (nd) "Team Design Patterns for Moral Decisions in Hybrid Intelligent Systems: A Case Study of Bias Mitigation," p. 13.

Verma A. A. *et al.*, (Aug. 2021) "Implementing machine learning in medicine," *Can. Med. Assoc. J.*, vol. 193, no. 34, pp. E1351–E1357, doi: 10.1503/cmaj.202434.

Visengeriyeva L., Kammer A., Bär I., Kniesz A., and Plöd M., (nd) "CRISP-ML(Q). The ML Lifecycle Process.," *CRISP-ML(Q)*. https://ml-ops.org/content/crisp-ml (accessed Jun. 01, 2022).

Wang A. Y., Mittal A., Brooks C., and Oney S., (Nov. 2019) “How Data Scientists Use Computational Notebooks for Real-Time Collaboration,” *Proc. ACM Hum.-Comput. Interact.*, vol. 3, no. CSCW, pp. 1–30, doi: 10.1145/3359141.

Zhang A. X., Muller M., and Wang D., (Apr. 2020) “How do Data Science Workers Collaborate? Roles, Workflows, and Tools,” *ArXiv200106684 Cs Stat*, Accessed: Dec. 17, 2021. [Online]. Available: http://arxiv.org/abs/2001.06684