

Socio-Technical Anti-Patterns in Building ML-Enabled Software (Supplementary Material)

Alina Mailach
Leipzig University
ScaDS.AI Dresden/Leipzig

Norbert Siegmund
Leipzig University
ScaDS.AI Dresden/Leipzig

Abstract—The artifacts described in this document belong to the paper ‘Socio-technical anti-patterns in building ML-enabled software: Insights from Leaders on the Forefront’. We conducted a large-scale qualitative analysis of existing videos from an online community employing reflexive thematic analysis. This way, we generated 17 anti-patterns in seven contexts and development activities with corresponding causes and recommendation. The artifacts contain initial, interim, and final themes and corresponding codes, showing the evolution of codes and naming of the themes.

I. RESEARCH OVERVIEW

In just a few years, machine learning (ML)-enabled software systems became ubiquitous in our daily lives. Still, there are reports about a large fraction of development efforts failing [1]. While many research work has focused on technical aspects of the development of ML-enabled systems, only lately researchers have turned on investigating team and cross-organization challenges [2], [3]. Such studies provide a rich, yet incomplete picture on the social and organizational aspects of building ML-enabled software. Our paper ‘Socio-technical anti-patterns in building ML-enabled software: Insights from Leaders on the Forefront’ [4] complements this perspective by exploring what experts and leaders report on socio-technical challenges when building ML-enabled software.

We analyzed videos uploaded by the MLOps.community¹ with the goal of identifying and interpreting thematic patterns. We employ a research methodology known as *reflexive thematic analysis* (RTA) [5], [6]. Our two-phased qualitative research design, consisting of an inductive and a deductive phase is depicted in figure 1.

During the inductive phase, 82 videos were taken from the ‘Full Talks’ playlist of the community’s YouTube channel. An initial filtering was conducted on the automatically generated transcripts using the keyword ‘team’ to identify videos that discuss socio-technical challenges. This process resulted in 37 videos being selected for further analysis in the first phase using the guide provided by Braun and Clarke [5]. At this stage, we focused particularly on coding and the development of initial themes. In the next step, the remaining 128 videos were assessed using the initial themes and codes developed earlier to identify relevant videos. The analysis then proceeded through all steps of the guide.

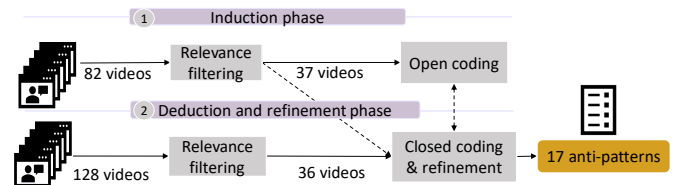


Fig. 1. Two-phased qualitative research design.

In total, 17 anti-patterns were identified across seven different contexts or development activities in three distinct areas: organizational silos, communication within an organization, and organizational leadership vacuum. For each context or activity, we further collected recommendations provided by the speakers.

II. ARTIFACTS

We provide the initial, interim, and final selection of our codes, and links to all analyzed videos in a publicly available repository². The artifacts consist of four files:

- `videos.csv` contains links to all videos that were incorporated in this study, with columns indicating the phase of relevance assessment (first or second) and the final relevance status (included or excluded from later analysis).
- `download_transcripts.py` is a Python script that automatically downloads transcripts from YouTube and saves them as files in the current directory.
- `initial_themes.pdf` contains the initial themes and codes used during the open coding process to establish an initial structure and organize the different codes within their appropriate contexts.
- `interim_and_final_themes.xlsx` depicts more developed themes and codes, and traces the association of codes with their respective themes. The ‘Interim themes’ table represents the stage of research when multiple symptoms, causes, and solutions were identified. The ‘Final themes’ table shows the mapping of codes to final patterns, causes, and recommendations as well as the corresponding final contexts.

¹<https://mlops.community/>

²<https://doi.org/10.5281/zenodo.7520777>

REFERENCES

- [1] “Gartner identifies the top strategic technology trends for 2022,” <https://www.gartner.com/en/newsroom/press-releases/2021-10-18-gartner-identifies-the-top-strategic-technology-trends-for-2022>.
- [2] N. Nahar, S. Zhou, G. Lewis, and C. Kästner, “Collaboration challenges in building ml-enabled systems: Communication, documentation, engineering, and process,” in *Proc. Int. Conf. on Software Engineering (ICSE)*. ACM, 2022, pp. 413–425.
- [3] T. Granlund, A. Kopponen, V. Stirbu, L. Myllyaho, and T. Mikkonen, “MLOps: challenges in multi-organization setup: Experiences from two real-world cases,” in *Workshop on AI Engineering (WAIN)*. IEEE, 2021, pp. 82–88.
- [4] A. Mailach and N. Siegmund, “Socio-technical anti-patterns in building ml-enabled software: Insights from leaders on the forefront,” in *Proc. Int. Conf. on Software Engineering (ICSE)*, 2023.
- [5] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006.
- [6] V. Braun, V. Clarke, N. Hayfield, and G. Terry, *Thematic Analysis*. Springer Singapore, 2019, pp. 843–860.