Exploring Challenges and Solutions for Non-Functional Requirements for Machine Learning Systems

Khan Mohammad Habibullah

Chalmers | University of Gothenburg

Abstract

Increasing use of Machine Learning (ML) in complex and safety-critical systems has raised concerns about quality requirements and constraints. Non-functional requirements (NFRs) such as fairness, transparency, security, and safety are critical in ensuring the quality of ML systems. However, many NFRs for ML systems are not well understood and the scope of defining and measuring NFRs in ML systems remains a challenging task. Our research project focuses on addressing these issues, using design science as a base of the research method. The objective of the research is to identify challenges related to NFRs and develop solutions to manage NFRs for ML systems. As a part of doctoral research, we have identified important NFRs for ML systems, NFR and NFR measurement-related challenges, preliminary NFR scope and RE-related challenges in different example contexts. We are currently working on the development of a quality framework to manage NFRs in the ML systems development process. In future, we will work more on developing solutions and evaluation of those solutions to manage NFRs for ML systems.

Keywords

Machine Learning (ML), non-functional requirements (NFRs), NFR challenges, quality framework

1. Introduction

Machine Learning (ML) is increasingly and extensively being used in many complex and safety-critical systems (e.g., autonomous vehicles, health care) to perform decision-making and prediction tasks. However, there is growing concern about potential biases and unintended consequences (e.g., non-deterministic behavior of ML and unsafe operations) of ML systems, which makes the development of these systems more complex, expensive, and effort-intensive compared to traditional systems. In addition, ML systems must fulfill certain quality requirements (non-functional requirements, NFRs) such as fairness, transparency, privacy, and safety.

NFRs for traditional software are relatively well understood and established. However, for ML systems, many of these NFRs have different meanings and are not yet well understood [1]. Additionally, new NFRs such as fairness and transparency have become critical in the context of ML, while some NFRs such as compatibility and modularity may have reduced importance. Moreover, new NFRs, such as retrainability, may become relevant. In addition, previously

In: A. Ferrari, B. Penzenstadler, I. Hadar, S. Oyedeji, S. Abualhaija, A. Vogelsang, G. Deshpande, A. Rachmann, J. Gulden, A. Wohlgemuth, A. Hess, S. Fricker, R. Guizzardi, J. Horkoff, A. Perini, A. Susi, O. Karras, A. Moreira, F. Dalpiaz, P. Spoletini, D. Amyot. Joint Proceedings of REFSQ-2023 Workshops, Doctoral Symposium, Posters & Tools Track, and Journal Early Feedback Track. Co-located with REFSQ 2023. Barcelona, Catalunya, Spain, April 17, 2023.

khan.mohammad.habibullah@gu.se (K. M. Habibullah)
0000-0002-0670-1524 (K. M. Habibullah)

CEUR Workshop Proceedings (CEUR-WS.org)

observed quality trade-offs among NFRs (e.g., security vs. performance) in traditional systems have not yet been explored in an ML context [2].

ML forms a part of a larger software system [3], and ML aspects can be decomposed into granular levels, e.g., training data, ML model, and results. Different NFRs may apply to different aspects of the system. Therefore, determining the scope of NFRs for ML systems, including identification, definition, and specification, remains a challenging task. Furthermore, measuring NFRs in an ML context and at these granular levels has not been explored.

Our Ph.D. research focuses on addressing these issues. Using a design science methodology, we aim to identify challenges regarding NFRs for ML, develop artifacts that address these challenges, and evaluate these artifacts in practice. To guide our study, we introduced a main research objective, which is split into sub-objectives: **Objective: Understand challenges in NFRs for ML and create a framework to manage NFRs for ML systems.**

Obj1: Understand NFR-related challenges for ML systems, including definition, scoping, and measurement challenges.

Obj2: Identify importance and criticality of NFRs for ML systems in literature and practice.

Obj3: Identify how to scope the definition and measurement of NFRs for ML systems.

Obj4: Identify and refine existing measurements for NFRs for ML systems.

Obj5: Develop and evaluate a structured framework to identify, define, measure, and document NFRs when developing ML systems.

As part of achieving these objectives, we developed several research questions and addressed these questions in different phases of the Ph.D. research, described in our published research articles [4, 5, 6, 7]. An overview of the Ph.D. research is presented in Fig. 1. We identified important NFRs, NFR- and RE-related challenges in different ML contexts. We also identified NFRs for ML that have received less attention in literature. This exploration has led to initial solutions. We developed generic definitions for specific NFRs for ML, established initial definition and measurement scopes of NFRs in ML systems, and clustered NFRs based on shared characteristics. We are currently developing a framework to manage NFRs during the ML system development process. In the future, we will evaluate, refine, and improve this framework. This work makes a number of contributions: researchers can use our research results as a guideline to conduct further research on mitigating NFR-related challenges, filling gaps in the literature on important but less researched NFRs for ML, performing research on a specific group of NFRs that share similar characteristics; and developing new methodologies, frameworks, and solutions to manage NFRs for ML systems. Practitioners can use our work as a reference to identify important NFRs for their ML systems, scope and measure important NFRs for their systems, anticipate, address and manage potential NFR-related issues during ML system development.

2. Related Work

NFRs: NFRs are essential for the success of the software, and have been widely researched, but there is still a lack of standard guidelines for eliciting, defining, documenting, and validating NFRs [8]. There is also debate among the RE community about when NFRs should be considered in the RE process [9]. Doerr et al. applied a systematic and experience-based method for eliciting, documenting, and analyzing NFRs, with the aim of creating a comprehensive set of traceable and measurable NFRs [10]. However, the majority of research on NFRs has focused

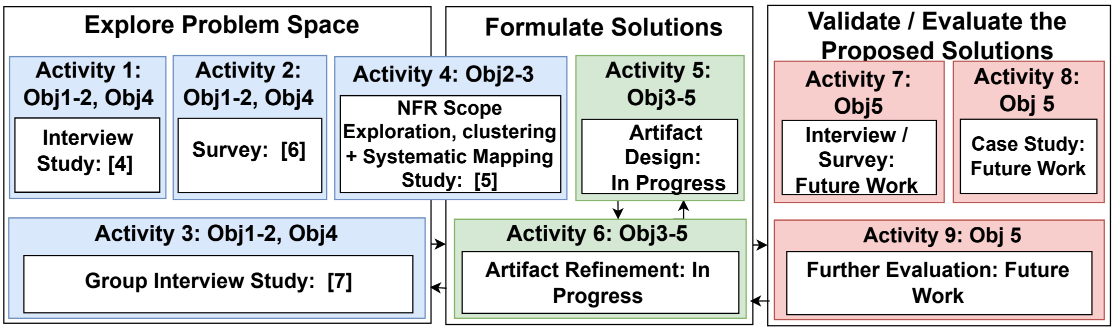

Figure 1: Overview of the Ph.D. thesis. Activities in blue represent completed work, green are work in progress, and red are future work.

on traditional software systems, with relatively little attention given to NFRs in systems using Machine Learning (ML).

NFRs for ML Systems: Horkoff discussed the challenges of NFRs for ML, and research direction, including how the requirements engineering (RE) can be adjusted for solutions to address the challenges related to NFRs for ML systems [11]. Kuwajima et al. illustrated that ML models lack in terms of requirements specification, design specification, interpretability, and robustness [12]. Gruber et al. stated that less research has been done on modeling NFRs, and research tends to focus on functional requirements more [13].

Vogelsang & Borg stated that RE practitioners need to understand ML performance measures to state good functional requirements for ML systems [14]. Khan et al. discussed the importance of documenting NFRs for ML systems and proposed a methodology for documenting and handling NFRs for delivering quality software systems [15]. Villamizar et al. identified quality characteristics relevant to ML systems and NFR related challenges, such as incomplete and fragmented understanding of NFRs for ML and lack of validated RE techniques to manage RE [16]. Martinez et al. performed a systematic mapping study and found that safety and dependability are the most studied properties of AI-based systems [17]. Although previous studies have discussed challenges in addressing NFRs in ML system development, limited research focuses on understanding the current practices and process of defining, allocating, and measuring such NFRs among professionals, and on developing solutions to challenges.

3. Methodology

This Ph.D. thesis follows design science as the main methodology to fulfill the research objectives. Fig. 1 presents the research methods we have used in our thesis so far and a plan of the methods we will use in the future as a part of a broader design science method. The research methods are described below in more detail.

3.1. Problem Space Exploration

Interview Study (Obj1-2, Obj4): We conducted an interview study with 10 participants working with ML and requirements engineering in a professional context to explore the perception and current treatment of NFRs in ML systems [4]. Through semi-structured interviews, qualitative data was collected, and we used thematic analysis and coding for data analysis.
Survey (Obj1-2, Obj4): To validate and expand upon the findings from the interview study,

we conducted a survey [5]. Our objectives for this survey matched the interview study, but in addition, we explored whether there is a difference of perspective for participants working in industry, academia or both. The survey participants included practitioners in academic and industrial organizations with experience in ML and RE. 42 individuals responded to at least part of the survey, with 30 responses analyzed based on the demographic information provided and completion of the questions. Most of the data collected was quantitative and analyzed using descriptive statistics, and qualitative data was also collected.

Group Interview Study (Obj1-2, Obj4): We examined NFRs for ML as part of a study on RE topics and challenges in a particular domain—ML-based autonomous perception systems [7]. The identified challenges includes NFR-related challenges. We conducted an interview study with 19 participants from five companies and used thematic analysis to analyze the data.

Preliminary Systematic Mapping Study (Obj2): We performed an exploratory study to establish an initial clustering and scoping of selected NFRs, and an initial estimation of the level of research performed on these NFRs. We performed a preliminary systematic mapping of the selected NFRs for ML systems. We utilized Scopus, a comprehensive meta-database, and we developed search strings by identifying relevant terms and synonyms from related literature and our discussions. To estimate the number of relevant publications for each selected NFR, we screened the titles and abstracts of a sample of 50 papers. Three researchers evaluated the relevance of each paper based on established inclusion and exclusion criteria.

3.2. Artifact Design

Initial Scoping and Clustering (Obj3): Based on the mapping, interview, and survey studies, we clustered ML system NFRs based on shared features and explored the scopes (e.g., data, model, system) that NFRs can be defined over for ML systems. We selected important NFRs for ML from our interview study [4], and defined those NFRs based on our previous experience and a review of literature from research papers, websites, blogs, and forums. To identify the scope of NFRs for ML systems, we identified the key elements of a ML system. We then utilized our prior definitions and experience, along with the titles and abstracts of relevant studies to determine the applicability of each NFR to these elements.

Artifact Framework Design (Obj3-5): Based on the results and recommendations of our previous studies, we are developing a quality framework to specify, allocate, measure, and manage NFRs for ML systems (illustrated in Fig. 2). We will conduct interviews and surveys, then adjust the framework based on recommendations.

3.3. Evaluation of the Proposed Solutions (Obj5)

We will evaluate the artifacts and other solutions we identify to manage NFRs for ML using research methods such as interviews, surveys, and case studies. We will conduct interview studies with participants working with NFRs and ML in a professional context to collect perceptions of domain experts and refine our artifacts and solutions based on the participants' opinions. Then, we will conduct a broader survey to validate the results of the interview data and gain further insights into the artifacts and solutions. Furthermore, we will conduct case studies in industry to evaluate the impacts of our artifacts and solutions in practice. The evaluation process and refinements of our developed artifacts will be done iteratively.

4. Current Results

We summarize results thus far, which have been published [4, 5, 6, 7].
Interview Study [4]: We gained an understanding of the perceptions and challenges related to NFRs in an ML systems context. From the interview data, we identified important NFRs for ML systems, such as accuracy, correctness, reliability, usability, and explainability. We also identified NFR related challenges (e.g., challenging NFRs, and uncertainty) and NFR measurement related challenges (e.g., missing measurement baseline, and complex ecosystems).
Survey [6]: The survey participates offered insight into the importance of NFRs in ML systems, and what differences exist in how NFRs are defined and measured between traditional systems and ML systems. We also compared results for from industrial, academic, or blended contexts. We also gained insight regarding NFR scope, NFR and NFR measurement challenges.
Group Interview Study [7]: In developing autonomous perception systems as part of driving automation systems, practitioners face RE challenges such as difficulties in defining requirements upfront. They often rely on scenarios and operational design domains as RE artifacts. Practitioners identified important NFRs for autonomous perception systems, such as performance, comfort, and integrity. They discussed quality trade-offs, such as accuracy vs. usability.
Preliminary Systematic Mapping Study [5]: We conducted a literature search to estimate the number of relevant publications on each of the NFRs considerd in the interview study [4]. We found that performance, accuracy, and efficiency received the most attention in literature. In contrast, retrainability, justifiability, and testability received the least attention.
Initial Scoping and Clustering [5]: We clarified the scope of NFRs for ML systems by dividing them into clusters based on shared attributes and definitions. For example, NFRs related to functional correctness (e.g., accuracy, consistency, correctness) of ML systems are grouped. We also performed an exploratory scoping of selected NFRs in terms of which elements of the system they can be defined and measured over (e.g., ML algorithm, ML model, or results).

5. Research Plan

5.1. Artifact Design: Framework for NFRs for ML

Currently, we are working on developing solutions for managing NFRs for ML systems. We are developing a quality framework for scoping, allocating, measuring and specifying NFRs for ML systems, presented in Fig. 2. The framework consists of four steps. As a first step, practitioners need to identify the important NFRs for ML systems, develop an NFR definition catalogue, and create clusters of important NFRs based on shared characteristics. We will provide a starting list of important NFRs and seed definitions for practitioners to build upon and adapt, an initial version is available in [5]. The second step is to define NFR scope and identify NFR trade-offs, where practitioners need to identify in which part of the system NFRs should be defined (e.g., training data, ML algorithm, ML model), and what are the trade-offs among different NFRs (e.g., safety vs. performance). Thirdly, practitioners need to create a measurement catalogue for the important NFRs for their systems, where they need to specify the techniques to measure specific NFRs. As with the definition catalogue, we will provide an initial catalogue of important NFRs and commonly associated measures as a starting point for practitioners. This can then be extended and adapted as needed for each domain. Finally, practitioners need to fill out a

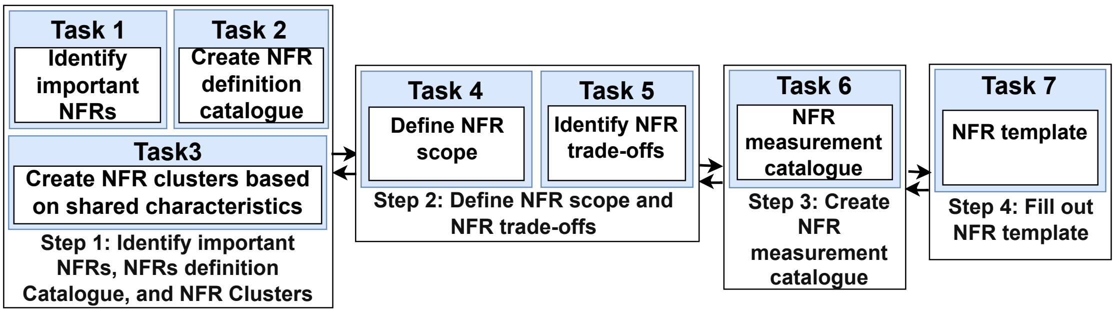
Task 1
Identify important NFRs
Task 2
Create NFR definition catalogue
Task3
Create NFR clusters based on shared characteristics
Step 1: Identify important NFRs, NFRs definition Catalogue, and NFR Clusters
Task 4
Define NFR scope
Task 5
Identify NFR trade-offs
Step 2: Define NFR scope and NFR trade-offs
Task 6
NFR measurement catalogue
Step 3: Create NFR measurement catalogue
Task 7
NFR template
Step 4: Fill out NFR template

Figure 2: Quality framework for managing NFRs in ML systems development.

prescribed requirements template. More details such as example definitions, trade-offs and measurements, will be provided as the framework is gradually developed.

The initial version of this framework is general, across all NFRs and domains. We believe that of our findings and recommendations can be generally applied. However, as part of our evaluation, if we find NFR-specific or domain-specific needs, we may pivot to focus the framework more narrowly on specific NFRs or domains.

5.2. Future Work and Anticipated Challenges

Our future work will start with demonstrating our developed artifact—the quality framework—in practice, and gathering early feedback using interviews and/or surveys with practitioners working with RE and ML. Based on the input from the domain experts, we will refine our proposed quality framework and perform a further evaluation. We also plan to develop a rigorous NFRs definition catalogue and NFRs measurement catalogue specific to ML systems as a part of the framework that will pose features such as NFR measurement techniques, tools, measurement baseline, measurement capturing techniques, measurement challenges, and so on.

In terms of anticipated challenges, it may not be easy to measure the impact of our developed solutions in practice. Finding experts in both RE and ML for the interview and survey purpose could be challenging, according to our previous interview and survey experience. Finding industrial partners for conducting case studies to demonstrate and evaluate our solutions to manage NFRs for ML could be challenging and time-consuming. Furthermore, the fragility of the framework and ensuring the generalizability of our proposed solutions for all ML systems in different contexts could be challenging.

Acknowledgements

This thesis is funded by the Swedish Research Council (VR project iNForM).

References

[1] R. Binns, Fairness in machine learning: Lessons from political philosophy, in: Conf. on Fairness, Accountability and Transparency, PMLR, 2018, pp. 149–159.

[2] T. Kamishima, S. Akakamishima, J. Sakuma, Fairness-aware learning through regularization approach, in: 2011 IEEE 11th Int. Conf. on Data Mining Workshops, IEEE, 2011, pp. 643–650.

[3] D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young,

J.-F. Crespo, D. Dennison, Hidden technical debt in machine learning systems, Advances in neural information processing systems 28 (2015) 2503–2511.

[4] K. M. Habibullah, J. Horkoff, Non-functional requirements for machine learning: understanding current use and challenges in industry, in: 2021 IEEE 29th Int. Requirements Engineering Conf. (RE), IEEE, 2021, pp. 13–23.

[5] K. M. Habibullah, G. Gay, J. Horkoff, Non-functional requirements for machine learning: An exploration of system scope and interest, in: 2022 IEEE/ACM 1st Int. Workshop on Software Engineering for Responsible Artificial Intelligence (SE4RAI), IEEE, 2022, pp. 29–36.

[6] K. M. Habibullah, G. Gay, J. Horkoff, Non-functional requirements for machine learning: Understanding current use and challenges among practitioners, Requirements Engineering (2023) 1–34.

[7] K. M. Habibullah, H.-M. Heyn, G. Gay, J. Horkoff, E. Knauss, M. Borg, A. Knauss, H. Sivencrona, P. J. Li, Requirements engineering for automotive perception systems, in: 29th Int. Working Conf. on Requirement Engineering: Foundation for Software Quality, Springer, 2023.

[8] M. Glinz, On non-functional requirements, in: 15th IEEE Int. Requirements Engineering Conf. (RE 2007), IEEE, 2007, pp. 21–26.

[9] L. Chung, B. A. Nixon, E. Yu, J. Mylopoulos, Non-functional requirements in software engineering, volume 5, Springer Science & Business Media, 2012.

[10] J. Doerr, D. Kerkow, T. Koenig, T. Olsson, T. Suzuki, Non-functional requirements in industry-three case studies adopting an experience-based NFR method, in: 13th IEEE Int. Conf. on Requirements Engineering (RE'05), IEEE, 2005, pp. 373–382.

[11] J. Horkoff, Non-functional requirements for machine learning: Challenges and new directions, in: 2019 IEEE 27th Int. Requirements Engineering Conf. (RE), IEEE, 2019, pp. 386–391.

[12] H. Kuwajima, H. Yasuoka, T. Nakae, Engineering problems in machine learning systems, Machine Learning 109 (2020) 1103–1126.

[13] K. Gruber, J. Huemer, A. Zimmermann, R. Maschotta, Integrated description of functional and non-functional requirements for automotive systems design using SysML, in: 2017 7th IEEE Int. Conf. on System Engineering and Technology (ICSET), IEEE, 2017, pp. 27–31.

[14] A. Vogelsang, M. Borg, Requirements engineering for machine learning: Perspectives from data scientists, in: 2019 IEEE 27th Int. Requirements Engineering Conf. Workshops (REW), IEEE, 2019, pp. 245–251.

[15] A. Khan, I. F. Siddiqui, M. Shaikh, S. Anwar, M. Shaikh, Handling non-fuctional requirements in IoT-based machine learning systems, in: 2022 Joint Int. Conf. on Digital Arts, Media and Technology with ECTI Northern Section Conf. on Electrical, Electronics, Computer and Telecommunications Engineering, IEEE, 2022, pp. 477–479.

[16] H. Villamizar, T. Escovedo, M. Kalinowski, Requirements engineering for machine learning: A systematic mapping study, in: 2021 47th Euromicro Conf. on Software Engineering and Advanced Applications (SEAA), IEEE, 2021, pp. 29–36.

[17] S. Martínez-Fernández, J. Bogner, X. Franch, M. Oriol, J. Siebert, A. Trendowicz, A. M. Vollmer, S. Wagner, Software engineering for AI-based systems: a survey, ACM Transactions on Software Engineering and Methodology (TOSEM) 31 (2022) 1–59.