Enabling machine learning in software architecture frameworks

Reference:
Moin Armin, Badii Atta, Günnemann Stephan, Challenger Moharram.- Enabling machine learning in software architecture frameworks
2023 IEEE/ACM 2nd International Conference on AI Engineering – Software Engineering for AI (CAIN), 15-16 May 2023, Melbourne, Australia - ISBN 979-83-503-0113-7 - IEEE Explorer, 2023, p. 92-93
Full text (Publisher's DOI): https://doi.org/10.1109/CAIN58948.2023.00021
To cite this reference: https://hdl.handle.net/10067/1977300151162165141

Enabling Machine Learning in Software Architecture Frameworks

Armin Moin
Tech. Univ. of Munich (TUM) &
Univ. of Antwerp
Germany & Belgium
armin.moin@tum.de

Atta Badii
University of Reading
United Kingdom
atta.badii@reading.ac.uk

Stephan Günnemann
*TUM & MDSI**
Germany
s.guennemann@tum.de

Moharram Challenger
Univ. of Antwerp & Flanders Make
Belgium
moharram.challenger@uantwerpen.be

***Abstract*—Several architecture frameworks for software, systems, and enterprises have been proposed in the literature. They have identified various stakeholders and defined architecture viewpoints and views to frame and address stakeholder concerns. However, the Machine Learning (ML) and data science-related concerns of data scientists and data engineers are yet to be included in existing architecture frameworks. We interviewed 65 experts from around 25 organizations in over ten countries to devise and validate the proposed framework that addresses the mentioned shortcoming.**

***Index Terms*—architecture frameworks, viewpoints, views, machine learning, qualitative research**

I. Introduction

Architecture frameworks provide "conventions, principles, and practices for the description of architectures established within a specific domain of application or community of stakeholders" [1]. There exist several well-established examples, including TOGAF [2], [3], DoDAF [4], TEAF [5], MODAF [6], the Zachman Framework [7], the "4+1" View Model of Software Architecture [8], and RM-ODP [9]–[12]. TOGAF, DoDAF, TEAF, and MODAF were primarily concerned with enterprise architectures. MODAF was replaced by NAF [13].

Machine Learning (ML) is increasingly deployed to support a range of self-adaptive and intelligent capabilities in emerging software and information system solutions. Also, ML components assume a prominent role in many systems and organizations. However, prior work in architecture frameworks did not recognize any stakeholders with ML-related concerns. The contribution of this paper is twofold: i) it identifies several stakeholders of modern software, systems, and enterprises that might have ML-related concerns (Section II). ii) It proposes new architecture viewpoints and views to support integrated ML, data science, and software systems development, namely integrative MLOps-DevOps (Section III).

II. ML-enabled architecture frameworks

A. Identified stakeholders

The following stakeholder groups were already considered in prior work on architecture frameworks: 1) end-users, 2) business stakeholders, 3) database designers and engineers, 4) software architects and engineers (i.e., developers), 5) system designers, engineers, and integrators. Additionally, we believe that 6) network engineers and 7) security experts should be distinguished from system engineers and software engineers, respectively, given the sophisticated level of knowledge and skills that is required for designing and managing secured, pervasive technologies of modern software and information systems. Furthermore, we noticed the stakeholder groups below during the initial (one-on-one) expert interviews: 8) safety and regulatory compliance engineers, 9) data protection (privacy) officers, and 10) ethics committees or boards. Also, the online survey participants pointed out the following stakeholder groups: 11) quality assurance (test) engineers, and 12) maintenance managers. Last but not least, we propose counting 13) data scientists (including ML engineers) and 14) data engineers among the stakeholders of modern systems, software, and enterprises, which often contain ML components or ML-enabled (sub-)systems. The proposed stakeholder groups were validated through our online survey.

Data scientists are responsible for analytics modeling. However, there is also a need for technologies regarding the deployment of DA models in products, services, and operational systems. This part is known as analytics operations [14]. Data engineers are typically concerned with this part, which is also called Data Engineering (DE). Together, DA (i.e., data science) and DE are called Data Engineering and Analytics (DEA) [15] or Data Science and Engineering (DSE) [16].

B. Proposed viewpoints and views

We propose two new architecture viewpoint categories to frame the concerns of data scientists and data engineers in architecture frameworks. We call the new viewpoint categories analytics modeling (alternatively DA or data science) and analytics operations (or DE), respectively. Moreover, we propose adopting and adapting existing notations and model kinds to realize corresponding views for the new viewpoints, as well as new views for the viewpoints of other stakeholders communicating and collaborating with data scientists and data engineers. Table I illustrates the viewpoints, as well as notations and model kinds supporting the views corresponding to these viewpoints. The list of notations and model kinds in each row of the table is ordered based on the opinions of the

*Munich Data Science Institute

TABLE I
NOTATIONS AND MODEL KINDS SUPPORTING THE VIEWPOINTS AND VIEWS OF ML-ENABLED ARCHITECTURE FRAMEWORKS

Stakeholders communicating or collaborating	**Notations and model kinds supporting the viewpoints and views**
Data scientists with their peers	i) mathematical notations, ii) charts, diagrams, or plots, iii) DFGs or CGs [17], iv) PGMs [18], v) data flow diagrams (or UML activity diagrams showing the flow of data rather than the flow of control), for example, for the data analytics pipeline
Data engineers with their peers	i) data flow diagrams, ii) UML class diagrams, iii) DFGs or CGs [17], iv) Entity-Relationship (ER) diagrams, v) mathematical notations
End-users with data scientists and engineers	i) text documents, ii) charts, diagrams, or plots, iii) tables or matrices, iv) data flow diagrams, v) UML use case diagrams
Business stakeholders with data scientists and engineers	i) charts, diagrams, or plots, ii) text documents, iii) tables or matrices, iv) UML use case diagrams
Database designers and engineers with data scientists and engineers	i) ER diagrams, ii) UML class diagrams, iii) data flow diagrams, iv) UML use case diagrams, v) tables or matrices
Software architects and engineers with data scientists and engineers	i) UML class diagrams, ii) data flow diagrams, iii) UML use case diagrams, iv) ER diagrams
System designers, engineers, integrators, and network engineers with data scientists and engineers	i) UML deployment diagrams, ii) data flow diagrams, iii) DFGs or CGs [17] augmented with physical (i.e., deployment) information, iv) UML class diagrams
Security experts with data scientists and engineers	i) data flow diagrams, ii) UML deployment diagrams, iii) DFGs or CGs [17] augmented with physical (i.e., deployment) information, iv) ER diagrams, v) mathematical notations, vi) UML class diagrams
Safety and regulatory compliance engineers, data protection (privacy) officers, and ethics committees or boards with data scientists and engineers	i) text documents, ii) data flow diagrams, iii) ER diagrams, iv) DFGs or CGs [17] augmented with physical (i.e., deployment) information, v) tables or matrices, vi) UML deployment diagrams

survey participants concerning the suitability of each option for the specific purpose.

III. CONCLUSION AND FUTURE WORK

In this paper, we have enhanced architecture frameworks to address ML-enabled software and information systems. We have identified the stakeholders who might have concerns with respect to the ML aspects, namely data scientists and data engineers. Moreover, we have proposed new architecture viewpoint categories (i.e., analytics modeling and analytics operations) as well as architecture views to frame and address the stakeholder concerns. In the future, the enhancement of architecture frameworks for other sub-disciplines of Artificial Intelligence (AI) beyond ML will be on the agenda. Last but not least, more in-depth studies, including case studies, will be required to show the effectiveness and efficiency of the proposed framework for real-world systems.

REFERENCES

[1] "ISO/IEC/IEEE 42010:2011 Systems and software engineering — Architecture description," ISO / IEC / IEEE, Standard, 2011. [Online]. Available: https://www.iso.org/standard/50508.html

[2] "The Open Group Architecture Framework (TOGAF) version 8.1.1," The Open Group, Standard, 2006. [Online]. Available: https://pubs.opengroup.org/architecture/togaf811-doc/arch/toc.html

[3] "The Open Group Architecture Framework (TOGAF) version 9.2," The Open Group, Standard, 2018. [Online]. Available: https://pubs.opengroup.org/architecture/togaf9-doc/arch/index.html

[4] "The US Department of Defense Architecture Framework (DoDAF)," Department of the Defense, Standard, 2010. [Online]. Available: https://dodcio.defense.gov/Library/DoD-Architecture-Framework/

[5] "The US Treasury Enterprise Architecture Framework (TEAF)," Department of the Treasury, Standard, 2000. [Online]. Available: www.treas.gov/cio

[6] "The British Ministry of Defence Architecture Framework (MODAF)," Ministry of Defence, Standard, 2012. [Online]. Available: https://www.gov.uk/guidance/mod-architecture-framework

[7] J. A. Zachman, "A framework for information systems architecture," *IBM Systems Journal*, vol. 26, no. 3, 1987.

[8] P. Kruchten, "Architectural blueprints — the "4+1" view model of software architecture," *IEEE Software*, vol. 12, no. 6, pp. 42–50, November 1995.

[9] "ISO/IEC 10746-1 - Information technology — Open Distributed Processing — Reference model: Overview," ISO/IEC, Standard, 1998. [Online]. Available: https://www.iso.org/standard/20696.html

[10] "ISO/IEC 10746-4:1998 Information technology — Open Distributed Processing — Reference Model: Architectural semantics — Part 4," ISO/IEC, Standard, 1998. [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso-iec:10746:-4:ed-1:v1:en

[11] "ISO/IEC 10746-2:2009 Information technology — Open distributed processing — Reference model: Foundations — Part 2," ISO/IEC, Standard, 2009. [Online]. Available: https://www.iso.org/obp/ui#iso:std:iso-iec:10746:-2:ed-2:v1:en

[12] "ISO/IEC 10746-3:2009 Information technology — Open distributed processing — Reference model: Architecture — Part 3," ISO/IEC, Standard, 2009. [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso-iec:10746:-3:ed-2:v1:en

[13] "NATO ARCHITECTURE FRAMEWORK Version 4," The North Atlantic Treaty Organization (NATO), Standard, 2018, https://www.nato.int/cps/en/natohq/topics_157575.htm?

[14] J. Pivarski, C. Bennett, and R. L. Grossman, "Deploying analytics with the portable format for analytics (pfa)," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 579–588.

[15] "TUM Data Engineering and Analytics Master's Program," https://www.tum.de/en/studies/degree-programs/detail/data-engineering-and-analytics-master-of-science-msc, n/a, accessed on 2022-03-21.

[16] R. Raj and et al., "An empirical approach to understanding data science and engineering education," in *Proceedings of the Working Group Reports on Innovation and Technology in Computer Science Education*, ser. ITiCSE-WGR '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 73–87. [Online]. Available: https://doi.org/10.1145/3344429.3372503

[17] M. Abadi and et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: http://tensorflow.org/

[18] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.