2022 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER) | 978-1-6654-3786-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/SANER53432.2022.00156

Knowledge Management Challenges for AI Quality

Xiaozhou Li
Faculty of ITC
Tampere University
Tampere, Finland
xiaozhou.li@tuni.fi

Sergio Moreschini
Faculty of ITC
Tampere University
Tampere, Finland
sergio.moreschini@tuni.fi

Aleksandra Filatova
Faculty of ITC
Tampere University
Tampere, Finland
aleksandra.filatova@tuni.fi

Davide Taibi
Faculty of ITC
Tampere University
Tampere, Finland
davide.taibi@tuni.fi

***Abstract*—Developing an AI-based system is uniquely challenging as it requires knowledge across multiple domains. Though the project team is required to be versatile, it is possible that their repertoire cannot cover all of the requirements of the system, which results in damage to the software quality. Therefore, it is critical to have an effective team knowledge management (KM) strategy to detect the valuable "unknown", optimize the "known" task assignment, and enlarge the team knowledge base. Moreover, it is more effective to support the process with data-driven approaches.**

***Keywords*-Software Quality; Artificial Intelligence; Knowledge Management; Challenges**

I. Introduction

The quality assurance of artificial intelligence (AI) based systems has drawn increasing attention in both academia and industry. Among the various issues and comments raised, the developers' skills and training are some of the most critical requirements pertaining to the quality of AI-based systems [1]. Developers' lack of coding skills shall certainly result in low software quality in general, while the lack of AI-related knowledge is a severe issue as well [2]. Especially, for the project team as a whole, their knowledge base regarding software engineering, AI, and the target domain is a critical asset which needs to be well managed.

For AI-based software project teams, regardless of their process and product knowledge, their technical knowledge and their domain knowledge, they need to know "Who Knows What" and "how well" [3]. Moreover, it is critical to elicit the tacit knowledge (i.e., the unknown known and the known unknown [4]) as a team in order to guarantee the shared understanding with all potential requirements tackled by the most competent team members. Hence, an executable strategy for team knowledge management is needed in regards to the quality assurance of the target AI-based system. Therein, data-driven approaches can potentially facilitate addressing the common challenges for tacit knowledge elicitation and management in general [4].

This study presents the identified challenges for AI projects regarding the team KM strategy based on the experience of our on-going AI project. We also propose data-driven solutions that can address such challenges. The vision is a data-driven framework for the optimized AI team knowledge management assuring AI-based system quality.

II. A Case of HealthySit

Herein, we analyze the KM-related issues of an AI-based system project team by investigating the on-going HealthySit project in our research group (shown in Figure 1). The HealthySit project aims to develop an office well-being AI-based platform that monitors the employees' sitting postures and provides gamified features encouraging healthy sitting and standing routines. The system takes photos of each person's sitting posture every minute with a camera connected to a Raspberry Pi and sends them to the server. With the postures identified and rated, it encourages good sitting postures via gamification mechanisms.

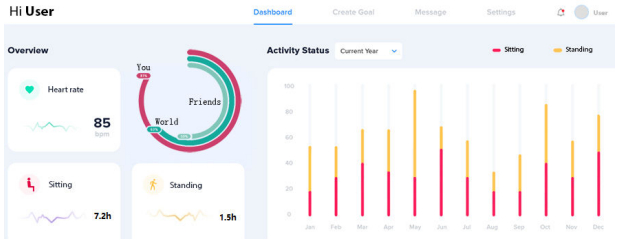

Figure 1. HealthySit System Screenshot

#	Task/Feature	Knowledge	Effort	Assigned
1	Take Photos	HP, PP	3	F.P.,F.L.
2	Setup Network	WT, HP	2	F.L.
3	Send Photos to Server	WT, PP	4	X.L.,F.L.
4	Detect Postures	PP, CV, DL	5	S.M.
5	Classify Postures	ML, DL	4	F.P.
6	Gamification Design	GD, UX	2	X.L.,A.F.
7	User Interface	WT, UX	4	S.M.,A.F.

Table I
Task Assignment based on Required Knowledge

For this AI system, seven main features are planned, which are shown in TABLE I. In addition, for each feature, the required knowledge and effort (in a range of 1 to 5) are estimated respectively by the project group via brainstorming. Shown in TABLE I, the required knowledge includes web techniques (WT), hardware programming (HP), Python

978-1-6654-3786-8/22/$31.00 ©2022 IEEE
DOI 10.1109/SANER53432.2022.00156

programming (PP), computer vision (CV), machine learning (ML), deep learning (DL), user experience (UX), and gamification design (GD). The knowledge level of each member is estimated and verified via brainstorming.

The core principle of task (i.e., feature) allocations based on the knowledge specifications is a feature must be assigned to a member that has the highest level in the according required knowledge. Meanwhile, each feature must be taken by at least one member, while each member shall cover at least two features. The synthesized knowledge level of the members taking charge in each feature and the sum of required effort for each member are both considered.

III. Challenges and Roadmap

Based on the experience with HealthySit project, we summarize the challenges regarding project team KM for AI-based systems and reason with potential data-driven solutions.

Challenge 1. Personal Knowledge Assessment. Currently, the knowledge and competence specification of the project team is self-estimated. Such estimations can be partially automated by crawling and analyzing the Github commit history and StackOverflow post history using their APIs [5]. However, knowledge not related to programming is hard to assess, e.g., gamification design, which still requires personal estimation with clarified and unified standards.

Challenge 2. Balanced Effort Allocation and Competence Coverage. Currently, the team members and the tasks are mapped manually. It can result in issues of unbalanced effort allocation and insufficiently covered tasks. Some features are only taken by one team member with less competence coverage, which might lead to low quality. However, assigning another member to this task might lead to the overloading effort requirements for him/her while others that are more available might be lacking the required knowledge. Measurements of the overall task-knowledge coverage are required as well as an algorithm to optimize the task allocation.

Challenge 3. Tacit Knowledge Transformation. Issues like unbalanced effort allocation and insufficiently covered tasks cannot be simultaneously solved only by proper task allocation. Thus, enriching the team knowledge base is important via discovering the "unknown" and transforming the "unknown" into "known". A potential data-driven solution is to construct the evolving system knowledge graph with the ontology of knowledge considered and an automated learning planning conducted continuously [6] [7] [8].

Challenge 4. "Unknown Unknown" Detection. For innovative projects, understanding and addressing the "unknown" is always an issue in software project management. Using approaches like group workshops, scenarios, prototyping and so on, shall help the process eliciting the unknowns, but can be time-consuming [4]. A potential data-driven approach to elicit the "unknown unknowns" is to expand the team's "shared vision and knowledge" [9] by continuously discovering and learning from others' experiences, e.g., similar projects.

Addressing the challenges, our roadmap towards a data-driven KM strategy for AI-based system quality shall include the following steps: 1) automated personal knowledge assessment with online collaboration platforms, 2) defined measurements for knowledge coverage with an optimization algorithm, 3) evolving project knowledge graph with automated learning plan and 4) similar projects recommendation for expanding shared vision. Provided the feasibility being investigated and validated, following such a roadmap shall gradually construct the data-driven knowledge management framework towards the AI-based system quality assurance.

IV. Conclusion

Effectively managing the shared knowledge of the project team is critical to the success of AI-based software projects and those of other domains as well. We identify some challenges therein based on the experience of our AI-based e-health project and propose the potential data-driven solutions. This article aims to raise the awareness of academia and industrial practitioners towards the knowledge management aspect of AI-based system quality assurance. Herein, we also aim to initiate enlightening discussion towards the data-driven solutions addressing such challenges.

References

[1] V. Lenarduzzi, F. Lomio, S. Moreschini, D. Taibi, and D. A. Tamburri, "Software quality for AI: Where we are now?" in *Int. Conf. on Software Quality*, 2021.

[2] F. Corea, "AI knowledge map: How to classify AI technologies," in *An introduction to data*, 2019.

[3] I. Rus, M. Lindvall, and S. Sinha, "Knowledge management in software engineering," *IEEE software*, vol. 19, no. 3, 2002.

[4] A. Sutcliffe and P. Sawyer, "Requirements elicitation: Towards the unknown unknowns," in *RE*. IEEE, 2013.

[5] Y. Xiong, Z. Meng, B. Shen, and W. Yin, "Mining developer behavior across github and stackoverflow." in *SEKE*, 2017.

[6] D. Fensel, "Ontology-based knowledge management," *Computer*, vol. 35, no. 11, pp. 56–59, 2002.

[7] M. P. Rice, G. C. O'Connor, and R. Pierantozzi, "Implementing a learning plan to counter project uncertainty," *MIT Sloan Management Review*, vol. 49, no. 2, 2008.

[8] S. Moreschini, F. Lomio, D. Häastbacka, and D. Taibi, "MLOps for evolvable AI intensive software system," in *SQA4AI 2022. Co-located with SANER 2022*, 2022.

[9] M. Evans, A. Wensley, and C. W. Choo, "How shared language and shared vision motivate effective knowledge sharing behavior," in *Euro Conf. on Knowledge Management*, vol. 1, 2012.