Vision of Knowledge Graph Lifecycle Management within Hybrid Artificial Intelligence Solutions

Romana Pernisch[1,2,*], Hennie Huijgens[3], Stefan Schlobach[1], Ruud Mattheij[4], Frank Benders[4], Hubert van Beusekom[4] and Freek Bomhof[4]

[1]Vrije Universiteit Amsterdam, The Netherlands
[2]Discovery Lab, Elsevier, Amsterdam, The Netherlands
[3]Utrecht University of Applied Sciences, Utrecht, The Netherlands
[4]TNO, The Netherlands

Abstract

Knowledge Graphs (KGs) are essential components in AI systems, providing structured and interpretable data representations. However, managing the lifecycle of KGs poses significant challenges due to their dynamic nature, requiring continuous updates, validation, and maintenance. This vision paper addresses the critical need for innovative lifecycle management practices for hybrid AI solutions, KGs being part of them. Given advances in software engineering and software lifecycle, we need to learn from their past and investigate their practices to be applied to hybrid AI. This can be best done in collaboration with industry, such as small to middle-sized companies (SMEs). Our work aims to advance the scientific understanding of KG lifecycle management, offering practical tools and methodologies that benefit various industries, including healthcare, finance, and manufacturing. The implementation of such practices will enhance the overall quality and trustworthiness of AI systems, contributing to broader societal acceptance and integration of AI technologies in the future.

Keywords

Artificial Intelligence, Knowledge Graph, Software Lifecycle, Ontology Evolution,

1. Introduction

The effective lifecycle management of Artificial Intelligence (AI) solutions is crucial for ensuring their utility and reliability when moving from the sandbox into production scenarios. However, today, we are often facing the issue of AI models being developed, published and then abandoned, from a scientific point of view. This hinders the possibility for industry to adopt newest approaches as they are rarely developed with their lifecycle in mind. Unfortunately, this is also hindered by the unavailability of tools to support said lifecycle. As dynamic systems, AI solutions require continuous updates to remain relevant and accurate, which complicates verification, validation, and maintenance efforts.

In this paper, we want to focus on ontologies and knowledge graphs (KGs) as part of AI solutions and how their lifecycle management also remains an unsolved issue [1]. Today, ontologies and KGs have been adopted in many industries and they present a crucial part of hybrid AI systems, combining symbolic knowledge representation with diverse machine learning approaches. Hence, effective lifecycle management of KGs is crucial for ensuring their utility and reliability. This poses a significant challenge for companies, which do not have the funding to invest or develop their own approaches at maintaining such systems. This in turn keeps them from adopting hybrid AI in their production systems.

Therefore, we need to develop innovative tools and methodologies for the lifecycle management of hybrid AI solutions, and within this paper specifically KGs, enabling more responsible and effective use of these technologies in various applications. We propose this challenge to be addressed by borrowing from the domain of software lifecycle management and adapting those approaches. By

SofLiM4KG'24: Software Lifecycle Management for Knowledge Graphs Workshop, co-located with ISWC'24, 12th November 2024, Baltimore, USA

*Corresponding author.
r.pernisch@vu.nl (R. Pernisch); hennie.huijgens@hu.nl (H. Huijgens); k.s.schlobach@vu.nl (S. Schlobach); ruud.mattheij@tno.nl (R. Mattheij); frank.benders@tno.nl (F. Benders); hubert.vanbeusekom@tno.nl (H. v. Beusekom); freek.bomhof@tno.nl (F. Bomhof)

making use of iterative and agile ways of working when developing hybrid AI solutions, we believe that we can find better ways for companies to deal with the lifecycles of AI solutions and in the process develop tools and methodologies. Two aspects play an important role here. An agile way of working ensures that a necessary culture, with the associated technology such as automated pipelines, is created that leads to continuous experimentation, facilitating both fast knowledge graph generation and frequent modification in order to innovate existing graphs [2, 3, 4]. But conversely, an approach based on knowledge graphs can also help to generate or improve the epics and user stories for those experiments [5, 6].

2. Related Work

Current hybrid AI systems utilizing KGs benefit from their ability to encode and reason over complex relationships. However, maintaining and updating KGs is challenging due to their dynamic nature and the need for continuous integration of new information.

With the uptake of semantic technologies to build KGs several approaches to define the KG construction process have arisen as analysed by [7]. Tamašauskaitė and Groth [7] performed a systematic literature analysis to identify a KG construction methodology: (1) identify data, (2) construct the KG ontology, (3) extract knowledge, (4) process knowledge, (5) construct the KG and (6) maintain the KG. They Their focus is on the construction of KGs from unstructured sources, mainly text, but the maintenance of the KG is a fairly neglected step in the overall process. Some other approaches include Radulovic et al. [8], Sequeda et al. [9], and Chessa et al. [10], however, also here the updating of the KG is not really considered in detail. Further, to our knowledge there is no available overall tool support for dealing with the KG lifecycle, as noted by Pernisch et al. [1]. Some first developments have been made, such as [11], but a thorough integration of such developments is still missing and needs to be addressed in the future.

Additionally, when addressing hybrid AI systems, the lifecycle of the KG needs to be integrated with the machine learning parts of the solution. For example, when using KGs in machine learning methods, a vector representation of the KGs is learned. Currently, there are only a few methods that can deal with updates to a KG in order to update such a representation. These are works by Song et al. [12], Cui et al. [13], or Daruna et al. [14]. However, these mostly focus on adding new information and do not regard deletions of information in their systems. Polleres et al. [15] pointed out these challenges in more detail in their survey and vision paper.

Further, existing lifecycle management approaches address artificial intelligence in general [16, 17], methods such as MLOps and AIOps [18, 19], and agile approaches for AI [20] they do not fully address the need for continuous updates and validation of KGs, leading to issues with compliance, accuracy, and usability but mostly hinders easy adaptation for companies. A limited number of studies address the use of KGs for software lifecycle purposes [21, 22, 23]. However, approaches aimed at using KGs to support lifecycle management of AI solutions in practical settings in companies do not seem to exist yet. That is why an investigation of this research gap and its support with practical innovations is necessary.

3. Vision and Approach

Our vision is to establish robust lifecycle management practices for hybrid AI, as well as develop methodologies and tools which support it, which integrate continuous updates, validation, and monitoring.

By leveraging both symbolic and non-symbolic AI approaches, the management of hybrid AI systems can be improved. Symbolic AI (ontologies and KGs) provides structure and reasoning capabilities, while non-symbolic AI handles large-scale data processing and adaptation. These aspects need to be carefully integrated with each other, and at the same time, its advantages should be leveraged. Iterative and agile ways of working will provide a starting point from where onwards we can then adjust and improve the methodology to suit the lifecycle of hybrid AI solutions better.

To develop a lifecycle management framework, we envision working closely with industry by applying iterative and agile approaches. We suggest using an iterative approach in which we would regularly evaluate the project results and adapt to the stakeholders' needs, just as is common for software solutions. By using this approach we will enable agile adaptation to the focus of the project maintaining the interest of stakeholders. In turn, we can revisit the applied methodology and adjust to the needs of the project itself, which will in the longer run result in better lifecycle management. Therefore, we want to create tools for continuous monitoring, analysis, and updating of hybrid AI solutions, ensuring they remain accurate and also, even though challenging, compliant with evolving regulations and standards.

Looking at results that have been achieved on a broad scale in the past decade in the field of a lifecycle-wide scope on information and communications technology (ICT) applications, such as continuous integration and the associated DevOps (or BizDevOps, or DevSecOps) methods and technologies [24, 25, 26, 27, 28, 29], there are still many lessons to be learned in the world of AI systems (MLOps/AIOps). Although AI systems certainly show similarities with ICT systems (they both involve code), there are important differences that necessitate a specifically AI-oriented approach. The differences between AI-based cyber-physical systems and traditional systems impact health prognostics (fit-for-purpose) and management mainly due to the AI-based systems' foundation in information flows, their novel system architectures that become necessary to enable system-internal awareness to ensure that decision-making is based on accurate information, and their unique abilities, as systems that learn and thus change their behaviour invalidates known health and performance indicators, often in unpredictable ways, and lifecycle management actions like updates, upgrades, or maintenance might no longer fit systems that adapted to their operational context [30].

Moreover, recent studies find that AI-based systems consequently pose a challenge within main Systems Engineering phases: systems that learn to change their behaviour during operations invalidate verification and validation efforts; updates or upgrades during system lifecycle management might no longer fit systems that adapted to changes in their operational context. That is why we suggest a focus on researching and developing innovation tools that support both AI makers and AI users in both the THINK phase prior to creating AI applications and the USE phase. The focus is on both Innovative AI solutions that support that lifecycle ('what': key technologies) and innovative methods and ways of working ('how': key methodologies).

First Steps We propose to tackle the presented problem and vision by establishing a collaboration, which call a lab, between industry, especially small to middles-sized businesses, and academic partners. We envision to start this work along four tracks, two focusing on the THINK phase and two on the USE phase of the three-phase model depicted in Figure 1. In Track 1, we will start with a project 'AI application mapping', which will conduct research into the development of an innovative toolset to generate data-driven personalised advice for companies for the most suitable AI application, based on added value, innovation capacity, and return on investment [31]. Using this advisory tool, a company can take a well-founded and evidence-based first step in deploying AI in daily operations. Track 2 starts with an exploratory study into the aspects of reuse of existing AI applications as developed in more than fifty labs associated with the Innovation Center for Artificial Intelligence (ICAI)[1] and the experiences on this subject of members of a industry branch association, e.g. International Council on Systems Engineering (INCOSE)[2], a not-for-profit membership organization for system engineers. This exploratory study thus lays the foundation for future research projects in Track 2.

The initial study with which Track 3 starts will investigate innovative ways to monitor and analyse AI systems in an operational setting to give companies better insight into the degree of fit for those systems. It is expected that, based on these innovative solutions, companies in the technical industry sector will have tools to carry out responsible and effective maintenance of existing AI systems by the end of the duration of the lab (planned for five years). In the fourth track, the lab starts with a study

[1] https://www.icai.ai/
[2] https://www.incose.org/

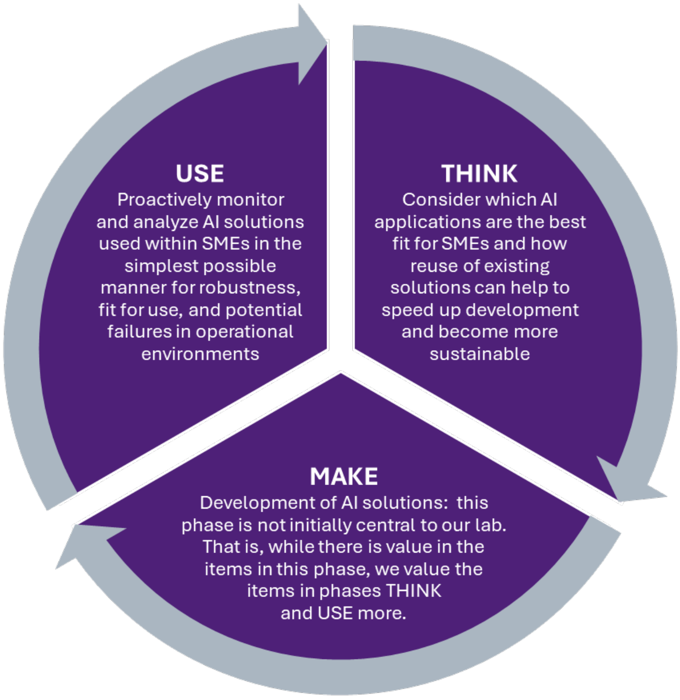

Figure 1: Within that lifecycle we look at two phases for potential innovations: 1) THINK; the phase in which one thinks about which AI could be important to them, and 2) USE; the phase in which one use AI in their operations. Initially, the lab will not delve deeply into the MAKE phase, because we see that many innovative solutions are already being developed.

into developing an innovative tool with which customers can design continuous experimentation in an operational setting through A/B testing of their AI system. This innovation helps companies to organize continuous innovation and renewal of systems responsibly and effectively.

4. Potential Impact and Conclusions

Effective lifecycle management of AI systems is essential for maintaining their reliability and utility. By integrating symbolic and non-symbolic AI approaches, we can develop innovative tools and methodologies to address current challenges. With this work, we see a very high societal/industry-oriented impact. Being able to provide tools and methodologies will especially help small and middle-sized businesses (SMEs) to safely and reliably deploy hybrid AI solutions more effectively. As large corporations already make use of AI solutions, it is hard for SMEs to keep up with the innovation because of the lack of funds. Therefore, we want to encourage the development and adaptation of lifecycle management practices as whole also in the research communities to continue bridging the gap between research and industry.

References

[1] R. Pernisch, M. Poveda-Villalón, D. Conde-Herreros, D. Chaves-Fraga, L. Stork, When ontologies met knowledge graphs: Tale of a methodology, in: The Semantic Web: ESWC 2024 Satellite Events - Hersonissos, Crete, Greece, May 28 - June 1, 2024, Proceedings, volume 13998 of *Lecture Notes in Computer Science*, Springer, 2024.

[2] E. Laurenzi, An agile and ontology-based meta-modelling approach for the design and maintenance of enterprise knowledge graph schemas., Enterprise Modelling & Information Systems Architectures 19 (2024).

[3] M. Cinpoeru, A.-M. Ghiran, A. Harkai, R. A. Buchmann, D. Karagiannis, Model-driven context configuration in business process management systems: an approach based on knowledge graphs, in: Perspectives in Business Informatics Research: 18th International Conference, BIR 2019, Katowice, Poland, September 23–25, 2019, Proceedings 18, Springer, 2019, pp. 189–203.
[4] Y.-S. Ahn, O.-R. Jeong, Time-aware polarisx: Auto-growing knowledge graph., Computers, Materials & Continua 67 (2021).
[5] M. Mancuso, E. Laurenzi, An approach for knowledge graphs-based user stories in agile methodologies, in: International Conference on Business Informatics Research, Springer, 2023, pp. 133–141.
[6] B. Zhang, V. A. Carriero, K. Schreiberhuber, S. Tsaneva, L. S. González, J. Kim, J. de Berardinis, Ontochat: a framework for conversational ontology engineering using language models, arXiv preprint arXiv:2403.05921 (2024).
[7] G. Tamašauskaitė, P. Groth, Defining a knowledge graph development process through a systematic review, ACM Trans. Softw. Eng. Methodol. 32 (2023). doi:10.1145/3522586.
[8] F. Radulovic, M. Poveda-Villalón, D. Vila-Suero, V. Rodríguez-Doncel, R. García-Castro, A. Gómez-Pérez, Guidelines for linked data generation and publication: An example in building energy consumption, Automation in Construction 57 (2015) 178–187. doi:10.1016/j.autcon.2015.04.002.
[9] J. F. Sequeda, W. J. Briggs, D. P. Miranker, W. P. Heideman, A pay-as-you-go methodology to design and build enterprise knowledge graphs from relational databases, in: The Semantic Web – ISWC 2019, Springer International Publishing, Cham, 2019, pp. 526–545. doi:10.1007/978-3-030-30796-7_32.
[10] A. Chessa, G. Fenu, E. Motta, F. Osborne, D. Reforgiato Recupero, A. Salatino, L. Secchi, Data-driven methodology for knowledge graph generation within the tourism domain, IEEE Access 11 (2023) 67567–67599. doi:10.1109/ACCESS.2023.3292153.
[11] D. Conde-Herreros, L. Stork, R. Pernisch, M. Poveda-Villalón, Ó. Corcho, D. Chaves-Fraga, Propagating ontology changes to declarative mappings in construction of knowledge graphs, in: D. Chaves-Fraga, A. Dimou, A. Iglesias-Molina, U. Serles, D. V. Assche (Eds.), Proceedings of the 5th International Workshop on Knowledge Graph Construction co-located with 21th Extended Semantic Web Conference (ESWC 2024), Hersonissos, Greece, May 27, 2024, volume 3718 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024. URL: https://ceur-ws.org/Vol-3718/paper1.pdf.
[12] H.-J. Song, S.-B. Park, Enriching Translation-Based Knowledge Graph Embeddings Through Continual Learning, IEEE Access 6 (2018) 60489–60497. doi:10.1109/ACCESS.2018.2874656, conference Name: IEEE Access.
[13] Y. Cui, Y. Wang, Z. Sun, W. Liu, Y. Jiang, K. Han, W. Hu, Lifelong embedding learning and transfer for growing knowledge graphs, in: B. Williams, Y. Chen, J. Neville (Eds.), Thirty-Seventh Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence, AAAI 2023, Washington, DC, USA, February 7-14, 2023, AAAI Press, 2023, pp. 4217–4224. doi:10.1609/AAAI.V37I4.25539.
[14] A. A. Daruna, M. Gupta, M. Sridharan, S. Chernova, Continual learning of knowledge graph embeddings, IEEE Robotics and Automation Letters 6 (2021) 1128–1135. doi:10.1109/LRA.2021.3056071.
[15] A. Polleres, R. Pernisch, A. Bonifati, D. Dell'Aglio, D. Dobriy, S. Dumbrava, L. Etcheverry, N. Ferranti, K. Hose, E. Jiménez-Ruiz, M. Lissandrini, A. Scherp, R. Tommasini, J. Wachs, How does knowledge evolve in open knowledge graphs?, TGDK 1 (2023) 11:1–11:59. URL: https://doi.org/10.4230/TGDK.1.1.11. doi:10.4230/TGDK.1.1.11.
[16] M. Haakman, L. Cruz, H. Huijgens, A. Van Deursen, Ai lifecycle models need to be revised: An exploratory study in fintech, Empirical Software Engineering 26 (2021) 95.
[17] D. De Silva, D. Alahakoon, An artificial intelligence life cycle: From conception to production, Patterns 3 (2022).
[18] J. Diaz-De-Arcaya, A. I. Torre-Bastida, G. Zárate, R. Miñón, A. Almeida, A joint study of the challenges, opportunities, and roadmap of mlops and aiops: A systematic survey, ACM Computing Surveys 56 (2023) 1–30.

[19] N. Nahar, S. Zhou, G. Lewis, C. Kästner, Collaboration challenges in building ml-enabled systems: Communication, documentation, engineering, and process, in: Proceedings of the 44th international conference on software engineering, 2022, pp. 413–425.

[20] K. Vaidhyanathan, A. Chandran, H. Muccini, R. Roy, Agile4mls—leveraging agile practices for developing machine learning-enabled systems: An industrial experience, IEEE Software 39 (2022) 43–50.

[21] M. Shetty, C. Bansal, S. Kumar, N. Rao, N. Nagappan, Softner: Mining knowledge graphs from cloud incidents, Empirical Software Engineering 27 (2022) 93.

[22] I. Mizutani, G. Ramanathan, S. Mayer, Integrating multi-disciplinary offline and online engineering in industrial cyber-physical systems through devops, in: Proceedings of the 11th International Conference on the Internet of Things, 2021, pp. 40–47.

[23] Z. Ge, Technologies and strategies to leverage cloud infrastructure for data integration, Future And Fintech, The: Abcdi And Beyond 311 (2022).

[24] P. Jacobs, B. Schlatmann, D. Mahadevan, Ing's agile transformation, The McKinsey Quarterly (2017).

[25] T. Peeters, K. Van De Voorde, J. Paauwe, The effects of working agile on team performance and engagement, Team Performance Management: An International Journal 28 (2022) 61–78.

[26] K. Balog, The concept and competitiveness of agile organization in the fourth industrial revolution's drift, Strategic Management 25 (2020) 14–27.

[27] M. Calnan, A. Rozen, Ing's agile transformation—teaching an elephant to race, Journal of Creating Value 5 (2019) 190–209.

[28] T. Hahn, M. te Brömmelstroet, Collaboration, experimentation, continuous improvement: Exploring an iterative way of working in the municipality of amsterdam's bicycle program, Transportation Research Interdisciplinary Perspectives 9 (2021) 100289.

[29] H. Huijgens, A. Van Deursen, R. Van Solingen, The effects of perceived value and stakeholder satisfaction on software project impact, Information and Software Technology 89 (2017) 19–36.

[30] 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 Trans. Softw. Eng. Methodol. 31 (2022). doi:10.1145/3487043.

[31] H. Huijgens, D. Eskiyerli, B. Kör, Help! which ai should i choose?-work in progress, The 1st Human-Centered AI Education Practice Conference, Dublin, Ireland (2024).