Auditable Semantic Web Machine Learning Systems

Laura Waltersdorfer[0000−0002−6932−5036]

TU Wien, Vienna, Austria,
laura.waltersdorfer@tuwien.ac.at

Abstract. Research in neurosymbolic Artificial Intelligence (AI) approaches has surged recently: Symbolic and sub-symbolic methods are combined to solve complex tasks. Nevertheless the significance of this field, little systematised knowledge exists yet. To scope our research, we will focus on semantic web machine learning systems (SWeMLS). Furthermore, AI systems have been under scrutiny due to prominent cases of biased or incorrect systems in sensitive domains. Thus, arises the need to make hybrid systems auditable, supporting the examination of their correct functioning. However, also this field has received limited attention. To that end, in this thesis, we want to investigate SWeMLS regarding 1) characteristics, interaction patterns and general system aspects, to provide an overview of this emerging field 2) guiding methodologies, technologies to make them auditable and 3) evaluation purposes by designing a generic end-to-end framework.

Keywords: auditability · semantic web · machine learning.

1 Problem Statement

Traditionally, AI research has been divided into symbolic and sub-symbolic approaches: Sub-symbolic techniques, i.e. machine learning methods and deep learning have successfully been applied to a variety of complex problem contexts, including computer vision, information retrieval and speech recognition [15]. On the other hand symbolic approaches, including logical and semantic web methods, are well suited for reasoning and making latent knowledge explicit.

Both approaches are well established in industry and academia, however also have limitations: Common criticism towards machine learning models is the lack of explainability [10] and the difficulty to be generalisable beyond training data [3]. In contrast, the creation and maintenance of symbolic knowledge is effort-intensive and interoperability between different models is challenging [5]. Thus, scientific interest has grown on how to benefit from the strengths of combining both approaches machine learning and symbolic domain knowledge [5] [21], while overcoming the aforementioned challenges.

Neuro-symbolic AI describing the combination of both approaches [10], is also referred to as the **third wave of AI**. With the emergence and success of innovative approaches, such as artificial neural networks [25] and knowledge

graphs, industrial applications have surged. This development lead to a high divergence in architectures, models and applied techniques and therefore opening a major need for understanding these systems.

While there are significant initial works in this area, such as Van Harmelen and ten Teije proposing a set of design patterns for hybrid systems [26], Seeliger et al. highlighting the semantic aspects of hybrid systems [23] and Sapna et al. exploring these from the machine learning perspective [22], there is not yet a systematic investigation from a general perspective.

Thus, one goal of this research is to systematically research the combination of symbolic and sub-symbolic approaches. Due to the breadth of the field and missing taxonomies to characterise such systems, we want to focus on **semantic web and machine learning systems** (SWeMLS), as a subset of neuro-symbolic AI systems. To manage the scope of this emerging field, our working definition is as follows: In our understanding, SWeMLS need to have a machine learning component interacting with a symbolic knowledge component aiming to achieve a task.

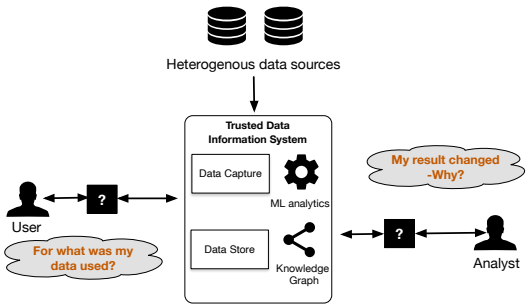

Fig. 1. Use Case for auditable SWeMLS.

The second goal of this research is concerned with the **auditability** of SWeMLS. With the rise of data-intensive analysis and black-box AI systems, and the coverage of prominent cases resulting in biased or incorrect results [7] [19], concerns over the correct functioning of complex systems have grown. As a result, audits aimed at AI have become also more relevant [20, 1] to check for problematic behavior of complex systems. However, with growing complexity, specially in the emerging field of AI, but also in the subset of SWeMLS, a variety of challenges arise. First, the traceability of full system logs becomes unmanageable as systems reach a higher level of complexity [11]. Second, although, audits occur increasingly frequently, a principled methodology or process is also missing covering both lifecycles of machine learning and semantic web components. There are initial efforts, focusing on one subcomponent, such as Model Cards as proposed in [17], or an semantic framework for supporting the AI design lifecycle phase [18].

To illustrate the relevance of auditability in the context of hybrid systems (see Fig. 1), we introduce a real-world use case from the medical domain, in the WellFort project[1]. Personal data from medical devices is shared with the system and also user consent for research purposes. Medical researchers can access this trusted data information system and may analyse anonymised user data for experiments. Both stakeholders, users and analysts have diverging auditing needs for the system: 1) users how and in which context their personal data is used and 2) for the analyst/system operator perspective the evidence of conducted analysis and retrieved results.

Analysis is conducted through machine learning, the semantic web component is used to check for checking user consent. Auditability in our context aims to go beyond standard provenance of who, when and what, and links additional contextual data to competency questions to support an auditor in examining the functionality of a system.

2 Importance

Increased use of opaque applications and data mining in sensitive areas, such as health care, HR and education, leads to the following hypothesis: Demand for recurring, controlled examination and verification of hybrid systems, both internally and externally will grow to prevent undesired impacts on stakeholders. Based on this context we identified the following three main challenges in the context of auditable SWeMLS (cf. Fig. 2): P1) a missing systematic understanding of the characteristics and building blocks of SWeMLS, P2) unclear requirements and capabilities for auditing SWeMLS, and P3) missing guidance on evaluation of the auditability level of SWeMLS.

3 Related Work

Semantic Web Machine Learning Systems can be considered as a subset of neuro-symbolic systems, which yet lack a concrete taxonomy. Previous research related to SWeMLS has focused on specific application areas or supersets, such as explainable AI [23] or recommender systems [22]. There have been initial categorisation efforts, such as [26], presenting an initial boxology for hybrid reasoning and learning systems focusing on neuro-symbolic systems. Besold et al. examine neuro-symbolic learning and reasoning from a cognitive perspective and point out several open research directions such as the confluence of knowledge representation and machine learning [4]. Hitzler et al. provide an initial overview of the integration of neuro-symbolic approaches and semantic web in a position paper [13]. Other related surveys include [21], focusing on data mining and knowledge discovery through semantic web technologies, while [8] is a qualitative, non-systematic review concentrating on machine learning techniques with semantic web technologies. Specific research targeting SWeMLS is limited and

[1] https://www.sba-research.org/research/projects/wellfort/

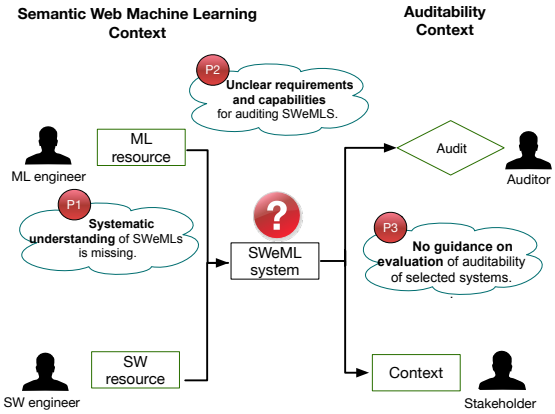

Fig. 2. Problem statement

systematized approaches are missing (cf. P1 in Fig. 2), however the interest and variety of connected topics displays much activity in this field.

Auditability Semantic Web Machine Learning Systems Historical and methodological lessons for auditing can be learned from various domains where audits are common, such as financial, aerospace or medical [20]. Algorithmic and AI audits are still under-researched, however efforts have been made to close this gap: Bandy provides a systematic overview of audits on different public-facing algorithms [1]. Sikos and Philip investigate provenance-aware technologies and data models, showcasing the applicability of semantic web technologies. [24]. In [18], Naja et al. propose an approach to audit ML lifecycle of systems supported by semantic technologies, however it is currently semi-automatic and covering only the design phase of ML systems. Concluding, existing works already investigated auditability from different contexts. However, current solutions are not covering the complexity of SWeMLS, requirements and capabilities for auditing such systems are missing (cf. P2 in Fig. 2) The reliance on (primarily) manual approaches does not scale and makes the evaluation of auditability of SWeMLS challenging (cf.P3 in Fig. 2).

4 Research Questions and Expected Results

Based on the analysis of the research area and concrete gaps, this thesis aims to investigate the following overall research question:

What are general characteristics of semantic web machine learning systems and how to support their auditability? In particular, we will investigate the following four focused research questions:

RQ 1: What are key characteristics and technological elements of semantic web and machine learning systems? To address the gap of sys-

tematic knowledge (P1) due to the recent surge in hybrid systems combining various methods of both approach, we aim to establish a systematically derived taxonomy and characteristic of key technological elements of such systems based on a systematic mapping study and use case analysis.

RQ 2: What are requirements and capabilities to enable auditing SWeML? This question aims to enable audits of these systems (P2). This encompasses the process of auditing as well as the capabilities needed to automate such processes, while building on the findings of RQ1, using the taxonomy to categorize system capabilities and requirements regarding their auditability.

RQ 3: What method enables a semantic-based auditing of SWeML systems? In order to ensure credibility of audits, external parties are invited to conduct the audit. However, in order to enable both internal and external parties to audit SWeML systems, a method is needed to provide the desired information (P2). Traditionally, for this purpose log information is gathered, leading to rich, but often unstructured data. With a semantic-based auditing we want to enrich log-based provenance and provide additional contextual information for audits in a proof of concept.

RQ 4: How to assess the level of auditability of a SWeML system? Based on the identified typologies from RQ1, requirements and capabilities from RQ 2 and the baseline implementation from RQ 3 we will develop an evaluation framework for auditable SWeMLS (addressing P3). The goal will be to provide i) a method for the evaluation of auditability and ii) the support of automatic evaluation. Based on this framework, we will conceptualise approaches for automated evaluation techniques, such as the generation of test cases or graph-based query templates and also test it with suitable users in real world use cases.

5 Research Plan and Preliminary Results

We will apply the design science approach [12] and *engineering cycle* based on Wieringa [28].

In the **problem investigation** phase, we decided to focus on systems that incorporate a semantic web structure and a machine learning component solving a certain task to provide scope for the research. This scoping is necessary to con-

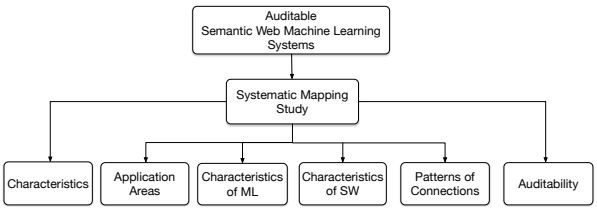

Fig. 3. Focus of Systematic Mapping Study

ceptualise and limit the broad topic of neuro-symbolic systems to a manageable

breadth for a survey. We are conducting a systematic mapping study [14] to identify key characteristics of SWeMLS (cf. Fig. 3), initial results are discussed in [27]. Currently, we are finalising the data extraction which will be concluded by data analysis. Based on these findings we will design a taxonomy for SWeML systems and basic processing flows between components, thus addressing RQ1 and P1.

Furthermore, we will analyse two exemplary use cases with SWeMLS that need to be auditable, which will be analysed for stakeholder, data and processing flows. The first use case is situated in the medical domain, aiming at integrating heterogeneous, sensitive data from multiple data sources. Auditability is added to increase transparency and credibility for conducted analyses via the provided platform. The second use case is in the ecological domain[2], combining semantic web and machine learning components to enrich the provided data [6].

In the **treatment design** phase, we will incorporate the findings of the problem investigation and will derive requirements from the discussed use cases. Based on these requirements and key characteristics from the taxonomy, we will conceptualise building blocks to model the lifecycle of SWeML systems. For the medical use case, we have extended the PROV-DM datamodel [2] for identified requirements for auditability and showed the feasability of our approach [9].

In the **treatment validation**, developed solutions will be evaluated based on the use cases and the coverage of the identified requirements and capabilities. Also the usability of the approach will be assessed in evaluation scenarios.

In the **treatment implementation** the results of the previous phases will be incorporated to demonstrate the feasibility of the approach. Furthermore, suggestions and improvements to extending existing standards and processes will be discussed.

6 Evaluation

The taxonomy of SWeMLS will be based on the results from the systematic mapping study and also bottom-up via the investigated system architectures from the use cases. Requirements and capabilities will be also derived from the use cases to build the auditable SWeMLS framework and methodology. For this purpose, the methodology for developing provenance- aware applications described in [16] will be applied and extended. To validate our approach, we will conduct user studies with the developed auditable SWeMLS framework concerning usability (e.g. execution time, handling) and coverage of the identified requirements and will be compared to existing frameworks and approaches. Specifically, evaluation metrics for provenance-aware applications mentioned in [16] will be also considered, including design-based metrics and implementation metrics.

[2] http://www.obaris.org

7 Reflection and Future Work

With increasing complexity of SWeMLS, the need for auditing hybrid systems rises, to achieve various other goals such as explainability or reproducibility. An essential first step was to scope this research to semantic web and machine learning systems. The focus will be to complete the systematic mapping study to derive characteristics and a taxonomy of SWeMLS. Requirements and capabilities will be analysed through case studies of the discussed projects to identify needs for auditability.

8 Acknowledgement

I would like to thank Dr. Marta Sabou, Dr. Fajar J. Ekaputra and Dr. Tomasz Miksa for their invaluable support and inputs. This work was funded by the Austrian Research Promotion Agency FFG under grant 871267 (WellFort) and 877389 (OBARIS).

References

1. Bandy, J.: Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits. Proceedings of the ACM on Human-Computer Interaction **5**(CSCW1), 1–34 (2021)
2. Belhajjame, K., B'Far, R., Cheney, J., Coppens, S., Cresswell, S., Gil, Y., Groth, P., Klyne, G., Lebo, T., McCusker, J., et al.: Prov-DM: The prov data model. W3C Recommendation (2013)
3. Bengio, Y., Deleu, T., Rahaman, N., Ke, R., Lachapelle, S., Bilaniuk, O., Goyal, A., Pal, C.: A meta-transfer objective for learning to disentangle causal mechanisms. arXiv preprint arXiv:1901.10912 (2019)
4. Besold, T.R., Garcez, A.d., Bader, S., Bowman, H., Domingos, P., Hitzler, P., Kühnberger, K.U., Lamb, L.C., Lowd, D., Lima, P.M.V., et al.: Neural-symbolic learning and reasoning: A survey and interpretation. arXiv preprint arXiv:1711.03902 (2017)
5. Bonatti, P.A., Decker, S., Polleres, A., Presutti, V.: Knowledge graphs: New directions for knowledge representation on the semantic web (dagstuhl seminar 18371). In: Dagstuhl Reports. vol. 8. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik (2019)
6. Breit, A., Waltersdorfer, L., Ekaputra, F.J., Sabou, M.: An Architecture for Extracting Key Elements from Legal Permits. In: 2020 IEEE International Conference on Big Data (Big Data). pp. 2105–2110. IEEE (2020)
7. Buolamwini, J., Gebru, T.: Gender shades: Intersectional accuracy disparities in commercial gender classification. In: Conference on fairness, accountability and transparency. pp. 77–91. PMLR (2018)
8. D'Amato, C.: Machine Learning for the Semantic Web: Lessons learnt and next research directions. Semantic Web **11**(1), 195–203 (Jan 2020)
9. Ekaputra, F.J., Ekelhart, A., Mayer, R., Miksa, T., Šarčević, T., Tsepelakis, S., Waltersdorfer, L.: Semantic-enabled Architecture for Auditable Privacy-Preserving Data Analysis. Semantic Web Journal (under-review) (2021)

10. Garcez, A.d., Lamb, L.C.: Neurosymbolic AI: The 3rd Wave. arXiv preprint arXiv:2012.05876 (2020)
11. Herschel, M., Diestelkämper, R., Lahmar, H.B.: A survey on provenance: What for? What form? What from? The VLDB Journal **26**(6), 881–906 (2017)
12. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. Design Science in IS Research MIS Quarterly **28**(1), 75–105 (2004)
13. Hitzler, P., Bianchi, F., Ebrahimi, M., Sarker, M.K.: Neural-symbolic integration and the semantic web. Semantic Web **11**(1), 3–11 (2020)
14. Kitchenham, B., Charters, S., et al.: Guidelines for performing systematic literature reviews in software engineering version 2.3. Engineering **45**(4ve), 1051 (2007)
15. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
16. Miles, S., Groth, P., Munroe, S., Moreau, L.: Prime: A methodology for developing provenance-aware applications. ACM Transactions on Software Engineering and Methodology (TOSEM) **20**(3), 1–42 (2011)
17. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D., Gebru, T.: Model cards for model reporting. In: Proceedings of the conference on fairness, accountability, and transparency. pp. 220–229 (2019)
18. Naja, I., Markovic, M., Edwards, P., Cottrill, C.: A Semantic Framework to Support AI System Accountability and Audit. In: Proceedings of the 2021 Extended Semantic Web Conference (2021)
19. Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S.: Dissecting racial bias in an algorithm used to manage the health of populations. Science **366**, 447–453 (2019)
20. Raji, I.D., Smart, A., White, R.N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., Barnes, P.: Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. pp. 33–44 (2020)
21. Ristoski, P., Paulheim, H.: Semantic web in data mining and knowledge discovery: A comprehensive survey. Journal of Web Semantics **36**, 1–22 (2016)
22. Sapna, R., Monikarani, H., Mishra, S.: Linked data through the lens of machine learning: an enterprise view. In: 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT). pp. 1–6. IEEE (2019)
23. Seeliger, A., Pfaff, M., Krcmar, H.: Semantic web technologies for explainable machine learning models: A literature review. Proceedings of the 1st Workshop on Semantic Explainability co-located with the 18th International Semantic Web Conference (ISWC 2019) **2465**, 30–45 (2019)
24. Sikos, L.F., Philp, D.: Provenance-aware knowledge representation: A survey of data models and contextualized knowledge graphs. Data Science and Engineering **5**, 293–316 (2020)
25. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in neural information processing systems. pp. 3104–3112 (2014)
26. Van Harmelen, F., Ten Teije, A.: A boxology of design patterns for hybrid learning and reasoning systems. Journal of Web Engineering **18**(1-3), 97–124 (2019)
27. Waltersdorfer, L., Breit, A., Ekaputra, F., Sabou, M.: Bridging Semantic Web and Machine Learning: First Results of a Systematic Mapping Study (accepted for publications). In: Proceedings of the International Conference on Database and Expert Systems Applications (2021)
28. Wieringa, R.: Design Science Methodology for Information Systems and Software Engineering. Springer Berlin Heidelberg (2014)