Leveraging Knowledge Graph to design the Machine-Learning Engineering Body-of-Knowledge

Juliette Mattioli, Dominique Tachet, Fabien Tschirhart, Henri Sohier, Loic Cantat, Boris Robert

Leveraging Knowledge Graph to design the Machine-Learning Engineering Body-of-Knowledge

Juliette MATTIOLI
Thales, France
juliette.mattioli@thalesgroup.com

Dominique TACHET
IRT SystemX, France
dominique.tachet@ext.irt-systemx.fr

Fabien TSCHIRHART
IRT SystemX, France
fabien.tschirhart@irt-systemx.fr

Henri SOHIER
IRT SystemX, France
henri.sohier@irt-systemx.fr

Loic CANTAT
IRT SystemX, France
loic.cantat@irt-systemx.fr

Boris ROBERT
IRT Saint Exupéry, France
boris.robert@irt-saintexupery.com

***Abstract*—A body of knowledge (BoK) is the complete set of concepts, terms, standards and activities promoting abroad awareness of a field or profession to guide practice or work. This paper presents how knowledge-based artificial intelligence (a.k.a. symbolic AI) could be used to build a body of knowledge (BoK) by first identifying relevant documents and data to capture concepts, standards, best practices, and state-of-the-art; then fusing all knowledge items into a knowledge graph, and finally providing query capacities. The overall process of knowledge collection, storage, and retrieval is implemented to support a trustworthy Machine Learning end-to-end engineering methodology, through the Confiance.ai BoK.**

***Index Terms*—Body-of-Knowledge; Knowledge graph; Knowledge extraction; Knowledge fusion; ML Engineering**

I. Rationale for a ML Engineering BoK

Initially used for non-critical tasks with no or very low risk, building an ML-based system was essentially a matter of combining *ad-hoc* engineering practices with the aim of delivering "usable" results as cost-effectively as possible. Several additional constraints need to be considered when dealing with industrially critical systems. Firstly, processes need to be rationalized, justified, made reproducible, optimized, etc. Second, the processes need to ensure that the overarching properties of the system being designed are actually met with the appropriate level of trust [2], including robustness (the ability of a computer system to withstand errors during execution and to cope with erroneous input), cyber-security, dependability (including reliability, availability, maintainability, safety - characteristics).

As with any technology, the ability to demonstrate that the expected service can be delivered in line with stakeholder expectations will be critical to the deployment of Machine Learning (ML) on critical systems [7]. Engineering practices including algorithm engineering, software engineering, system engineering, safety and cyber-security engineering, should address such issues. Thus, the Confiance.ai program [1] proposes a rigorous and interdisciplinary approach to formalize the engineering processes necessary for ML-based critical systems, allowing to guarantee "safe and secure" deployment and maintenance in operational condition, compatible with business usages. This approach is enhanced by a Body-of-Knowledge (BoK):

- To promote the Confiance.ai end-to end methodology [4], [5] which aims to support the engineering of reliable AI systems by mastering the risks related to AI [21]. This methodology encompasses methodological processes that cover the entire lifecycle of an AI system, both the component and the system levels, and is compatible with industrial practices.
- To specify the scope of the methodology and to clarify its place with respect to other related engineering disciplines such as algorithm engineering, data engineering, software and system engineering, safety and security engineering.
- To characterize the contents, and known practices of ML-based system engineering, organizing them in a coherent and comprehensive manner;
- To provide a basis for the development of a curriculum and, where appropriate, a qualification for certification.

Because our data/information and knowledge come from diverse sources and covers various engineering domains[1], the same concept (for example, accuracy) may appear in several sources with complementary or different connotations. For example, the item "accuracy" may exist in data engineering, at the ML model level but also at system level with various meanings. The first step is to define at least a taxonomy of the ML engineering domain [22] and then to link the different occurrences of the same item across a variety of data sources. In addition, engineers become highly specialized and the required competencies (skills, knowledge, experience and attitudes) become dispersed due to the induced complexity of developing a reliable ML-based system. The Confiance.ai BoK captures key information and outlines the core competencies and skills required to specify, design, deploy, maintain and evolve ML-based critical systems to guide and support engineers throughout the lifecycle of an AI-based critical system.

This work has been supported by the French government under the "France 2030" program, as part of the SystemX Technological Research Institute within the *Confiance.ai* Program (www.Confiance.ai).

[1]data engineering, algorithm engineering, software engineering, system engineering; cyber security, safety and cognitive engineering

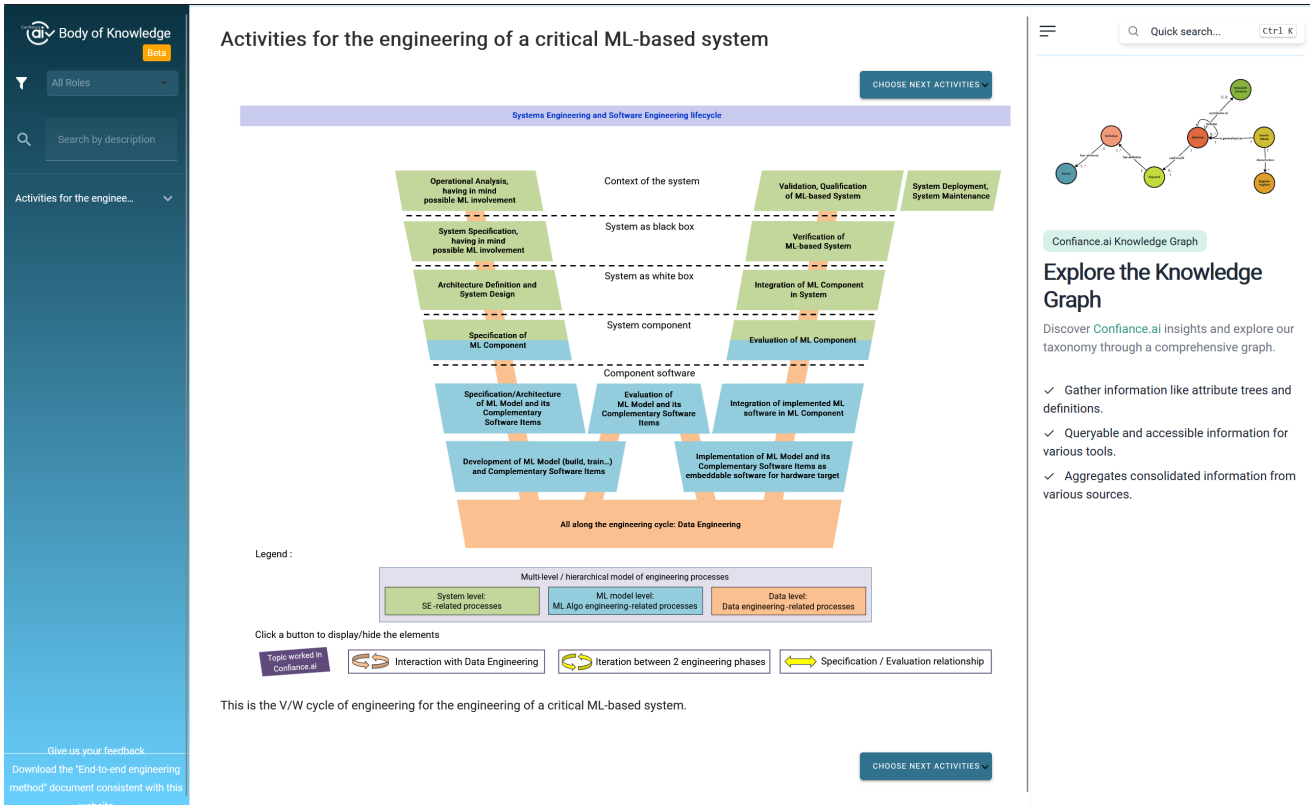

Fig. 1. Trustworthy ML Engineering Body-of-Knowledge - https://bok.confiance.ai/

The key added value is the systematic and comprehensive end-to-end engineering approach it provides for developing, maintaining and evolving ML-based systems. It also provides concrete deliverables, more specifically:

- A terminology that defines the set of main concepts of trustworthy ML engineering, as used by their practitioners; this constitutes the accepted ontology for the trustworthy ML domain.
- An outline of key knowledge, skills and accepted practices, which covers all basic skills required by any professional.

A **Body-of-Knowledge** [6], [25] is a complete set of concepts, terms and activities that make up a application domain. In addition, a BoK may also include technical terms and theoretical concepts as well as recommended practices. Notable examples are SWEBOK (Software Engineering Body-of-Knowledge) by IEEE Computer Society and ACM [6], and PMBOK (Project Management Body-of-Knowledge) [12]. In our context, the Confiance.ai ML Engineering BoK includes a comprehensive taxonomy, a set of fundamental principles, best practices and standards. It provides (see fig. 1) a foundation for engineering and material to support the specification, design, qualification and deployment of ML-based systems.

As part of the Confiance.ai program, an end-to-end methodology has been developed by a large and diverse group of experts, including industry stakeholders, to provide a set of methodological processes to support the development of reliable and trustworthy ML-based systems. These processes cover the entire life cycle of ML-based systems.

The ML-based systems engineering lifecycle depicted the center of our BoK in figure 1, follows the Concept Paper of EASA [9] and guidelines outlined in the ISO/IEC DIS 5338 standard for AI systems. A preliminary draft of this standard has been made available exclusively to members of the Confiance.ai program, although it is still under development. This illustrates a systems engineering life cycle that integrates machine learning (ML) into the traditional ”V-cycle” of systems development and creates a ”W-cycle” that is tailored to the processes of software engineering. This W-cycle emphases the critical step of evaluating the reliability of the ML model at the algorithm level before implementing it at the software level. It illustrates this lifecycle visually, starting with three key phases: operational analysis with potential ML involvement, system specification considering ML integration, and defining the architecture and designing the system. By using this ML Engineering BoK, organizations can ensure that ML is effectively incorporated into the development process. This will lead to more robust and efficient ML-based systems.

Concepts, knowledge, skills, standards, terminology, guidelines, practices and activities for data scientists, software and systems engineers, safety and cyber engineers, quality engineers and auditors, project managers and customers of ML-based solutions are therefore the core components of this Confiance.ai BoK. All topics within the Body-of-Knowledge con-

tribute to trustworthiness based on the end-to-end Confiance.ai methodology [4], which encompasses all stages of the lifecycle of an ML-based system, from problem specification and the Operational Design Domain (ODD) formulation, through solution definition and design, to qualification, operational use and maintenance. In addition to the system specification including ODD analysis, and trustworthiness assessment processes, a risk analysis process is essential to address and mitigate the risks related to AI/ML technologies, based on add-hoc control-structures specifications. Thus, developing and maintaining the Confiance.ai BoK (i.e. keeping the information up to date and accurate) is a real challenge, but a necessarily condition to guide the engineer toward an industrial delivery of a ML-based system.

This paper promotes the use of a knowledge graph approach, a branch of symbolic AI to capture the dynamics of knowledge flows, to take into account tacit knowledge, and to provide new insights into the interrelationships between stakeholder concerns, from the elicitation of user needs. The paper describes the process of knowledge collection, storage, and retrieval that implements established trustworthy ML end-to-end engineering methodology through the the Confiance.ai BoK. For a more detailed description of the Confiance.ai end-to-end methodology, see [3], [4] or explore the Confiance.ai BoK at https://bok.confiance.ai/.

II. Leveraging Knowledge Graph for BoK design

First proposed by Google in 2012 to improve its search engine, the concept of knowledge graph (KG) is defined as a large knowledge base consisting of many entities and their relationships [8]. Many large-scale knowledge graphs such as Dbpeida [19], are constructed from various structured, semi-structured or unstructured data sources. To improve the quality of information services and provide users with smarter services, a KG is a knowledge representation approach that is capable of extracting, organizing, and effectively managing knowledge from large amounts of data. KG uses semantic retrieval methods for the collection of information from multiple sources for the increansing of search quality.

A graph-based knowledge representation and reasoning formalism derived from conceptual graphs, formalized by [24] as finite bipartite graphs where the set of nodes, is divided into concept and conceptual relation nodes. In such a graph, concept nodes represent classes of individuals, and conceptual relation nodes show how the concept nodes are related to each other [26]. In [10], a knowledge graph acquires information and integrates it into an ontology, and applies a reasoner to derive new knowledge. Moreover, following the definition of [15], Knowledge Graphs are "*structured representations of a fact, consisting of Entities, Relations, and Semantics. Entities can be real-world objects and abstract concepts, relationships represent the relationship between entities, and semantic descriptions of entities and their relationships contain types and properties with defined semantics. Property graphs, where nodes and relations have properties or attributes, or attribute graphs are widely used*".

All of these facets rely via a knowledge inference over knowledge graphs which is therefore one of the core technologies in the design of our Confiance.ai BoK.

With the advantages of unambiguous linkage, efficient querying and dynamic change of data mode, a knowledge graph (KG) is essentially a semantic network that is composed of different entities, concepts and relationships. The Semantic Web community has agreed to use RDF to represent a Knowledge Graph.

Thus, a KG is a directed labeled graph in which domain-specific meanings are associated with nodes (entities or concepts such as people, places, things) connected by edges (relationships or associations). KG is essentially a semantic graph that can easily integrate heterogeneous islands of structured and unstructured information from multiple sources, with the advantages of unambiguous linkage, efficient querying and dynamic data mode change. It is a directed labelled graph associating domain-specific meanings with nodes (entities or concepts like people, places, things) connected by edges (relationships or associations). Entities can be real-world objects or abstract concepts. Relationships represent the relationship between entities and semantic descriptions of entities, and their relationships contain types and properties with well-defined meanings.

As a typical graph structure, knowledge graphs have 1) systematic representation of large amounts of data; 2) easy to retrieve and use; 3) conducive to knowledge discovery and reasoning. A KG is represented in the form of triples $< subject, predicate, object >$, where the nodes of the graph represent entities or concepts, and the edges represent the relationships between entities or concepts. Its purpose is to model, store, and organize complex information in a way that makes it easy for both humans and machines to understand, navigate, and use the knowledge it contains.

Unlike simple graph or non-relational databases, KGs include an additional embedded layer called a reasoner (or inference engine), which allows them to extract implicit information from existing explicit concepts. Reasoning methods aim to derive new knowledge from existing triplets, which not only provides efficient ways to discover correlations, but also completes the knowledge graph. The consistency and integrity of the KG is ensured by techniques such as consistency inference. Reasoning techniques can perform domain knowledge reasoning by modeling domain knowledge and rules, which can support automatic decisions.

A vital element of a KG is the ontology it uses. To underline this, the authors in [14] define KGs as "*integrating knowledge into an ontology and applying a reasoner to derive new knowledge*". A strong hierarchical structure and low redundancy of the knowledge base are advantages of an ontology. When constructing the Confiance.ai ontology, we first defined Confiance.ai "vocabulary" through a collection of preferred terms that are used to more precisely retrieve content, categorize content, create style guides... Then, we mainly construct the taxonomy, i.e, a set of hierarchically related concepts from the subject words of engineering activities (examples

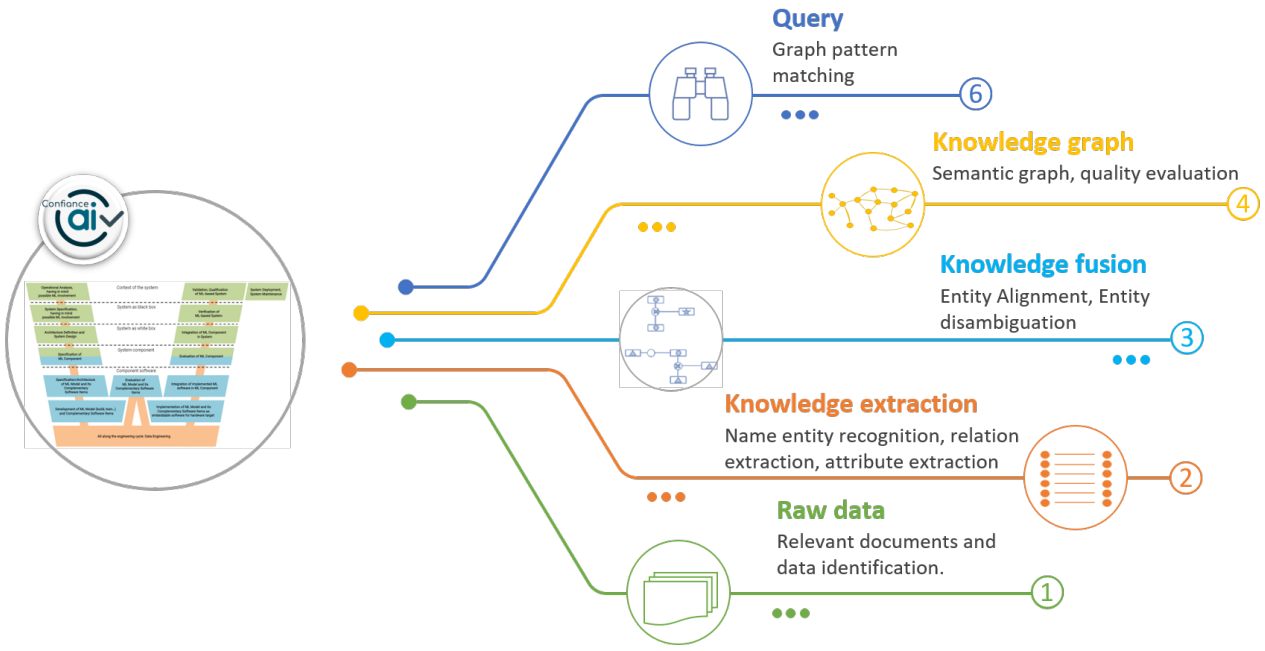

Fig. 2. The BoK design process

of vocabulary and concepts are given in annex of [1]. This classification reflects the "*is-a*" relationships in the vocabulary and captures the hierarchy of levels of abstraction in the ML-engineering domain. Today, human-curated structured or semi-structured data is almost the only source of high quality knowledge bases. To achieve this, the process to populate (and rebuild) a KG with new knowledge items (see fig. 2) is a bottom-up iterative process, which includes fives steps:

1) Relevant documents and data identification.
2) Knowledge extraction: Knowledge elements such as entity, relationship and attribute are extracted from source data and transformed the into understandable knowledge. Named entity recognition, relation extraction and attribute extraction are the three main tasks of knowledge extraction. Key techniques include those based on rules, statistical models and deep learning.
3) Knowledge integration and fusion: The knowledge acquired through knowledge extraction needs to undergo entity disambiguation and co-reference resolution processing through knowledge fusion. This task determines equivalent instances, classes, and attributes based on entity alignment and entity disambiguation. Entity alignment aims to discover entities that represent the same semantics in knowledge elements. Entity disambiguation eliminates the ambiguity of entities in different text, and maps them to actual entities that they refer to.
4) Knowledge graph: In this step, the fully fleshed knowledge graph with a defined relationship is created. After the creation, the knowledge graph should be visualized to see the proper mapping of the relationships between the nodes. This step includes semantic network construction and quality evaluation. Based on the architecture of the KG, semantic network construction refers to the design of conceptual knowledge units in the pattern layer. It standardizes how to describe concepts and how to relate two concepts in a given domain. To improve the quality of knowledge extracted from raw data, quality evaluation is used.
5) Query: Knowledge graph query is graph pattern matching problem where all instances of information according to a specified graph pattern may be found within the KG, through a query function.

Each step and its sub-steps for designing the Confiance.ai BoK are detailed in the following sections.

III. Application to the Confiance.ai BoK

A. Sept 1: Relevant documents and data identification

The first step is dedicated to the identification of data sources, as it has an impact on the entire knowledge graph development process, as well as on the choice of knowledge extraction techniques. For preliminary knowledge extraction activities, all Confiance.ai state of the art reports, the taxonomy, guidelines and methodologies are used and also the relevant standards (such as ISO/IEC 5338, Aerospace Standard 6983, IEEE 7000...) that could help to design ML-based critical system. This step has also been fed by any information needed to characterize and qualify trustworthy AI, to support trustworthiness by design, data engineering for trusted AI, IVVQ Strategy (Integration, Verification, Validation and Qualification) and to address targeted Embedded AI.

B. Step 2: Knowledge extraction

Knowledge extraction, as defined in [11], describes "extracting information from different sources, structuring it, and creating useful knowledge". Thus, the goal of this step is to extract useful information from 'raw' data (unstructured text and other (semi-)structured sources) to build knowledge graphs, to complete existing knowledge graphs, and to discover and recognize entities and associations. This step is the basis of knowledge graph construction, which mainly studies how to automatically extract information from heterogeneous documents to obtain candidate knowledge items. The technologies involved include natural language processing (NLP) and knowledge representation, which can automatically extract structured information such as entities, relationships and entity attributes from unstructured data, semi-structured data and structured data, so as to achieve full and effective use of external data.

The entity is the most basic element of the knowledge graph. It represents a concept. Moreover, the quality of knowledge graph construction is directly affected by the accuracy and integrity of its extraction. Then, relationship extraction involves extracting association relations between entities, i.e. creating semantic relations between entities through association relations, thus forming knowledge with a network structure.

These types of graphs embed a structured representation of facts, consisting of entities, relationships, and semantic descriptions, modeled with RDF (Resource Description Framework[2]) structure. An RDF model is a flexible model for the representation of data in the form of three-element tuples with no fixed schema requirement. It is a graph-based model for describing entities and how they relate to one another on the Web. Many researchers prefer to think of RDF as a set of triples, although it is commonly described as a directed and labelled graph, each consisting of a subject, predicate and object in the form of $< subject, predicate, object >$ the predicate being the relationship between the subject and the object - e.g.:

$$< Data\_Engineering, is\_an\_activity, MLOps >$$

Triples are stored in a triple store and are queried with the SPARQL query language. Compared to both inverted indices and plain text files, triple stores and the SPARQL query language enable users to search for information with expressive queries in order to satisfy complex user needs. Although a model is required for representing data in triples (similar to relational databases), RDF enables the expression of rich semantics and supports knowledge inference [13].

Usual used methods include pattern matching, machine learning, semantic rule extraction... Today, Large Language Models (LLMs) can be instrumental in building knowledge graphs by extracting entities, relationships, and attributes from unstructured text data [23].

[2]https://www.w3.org/RDF/

C. Step 3: Knowledge integration and fusion

Due to the complexity, variety, and volume of data available today, achieving efficient and accurate knowledge graph fusion is a challenging task. Knowledge integration, also called knowledge fusion, is the integration of knowledge from different sources and its cleansing from redundancies, inconsistencies, and ambiguities. This step is useful for both generating and completing knowledge graphs. We take advantage of knowledge graph representation for knowledge fusion introduced by [18] based on the conceptual graph model. This representation is used to represent and store knowledge as well as to perform the fusion. The general approach is to compare and analyze observations (knowledge item) such as $\triangle$, $\circ$ or $\star$, taking into account some *a priori* domain knowledge and to use graph operators as presented in figure 3.

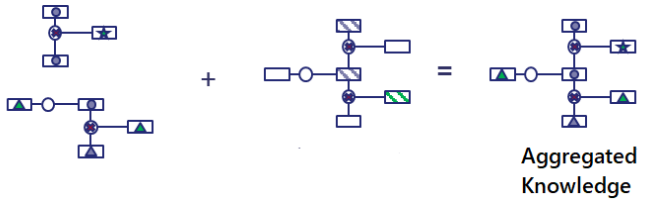

Fig. 3. Aggregation of knowledge by semantic information fusion [20]

Using the same model for both knowledge representation and knowledge aggregation has a major advantage. It allows us to remove the bias due to the translation from one formalism to another when using distinct models. We propose to use it for high-level information fusion approach [16], [17] based on the Maximal Join operator which is an aggregation operator on conceptual graphs [26]. It allows the semantic information fusion of not strictly identical concepts. This join operator is illustrated in figure 3, where the node shape indicates their associated concepts that we can merged.

D. Step 4: Knowledge Graph

Then, RDF model also allows for a more expressive semantics of the modeled data that can be used for knowledge inference. As a result, a KG is a set of interconnected information on a specific set of facts that includes characteristics of many data management paradigms:

- Database: Structured queries can be used to explore data in a database.
- Graph: KGs can be analyzed in the same way that any other network data structure can be.
- Knowledge base: Formal semantics are encoded in KGs, which can be used to understand data and infer new facts.

E. Step 5 : Query

In our context, we consider a BoK to be a (conceptual) graph of knowledge [20]. Finally, the ingested, transformed, integrated and stored knowledge will only become useful, if answers can be efficiently retrieved by our users in an intuitive manner. Today, keyword queries and specialized query languages (e.g., SQL and SPARQL) are the dominant

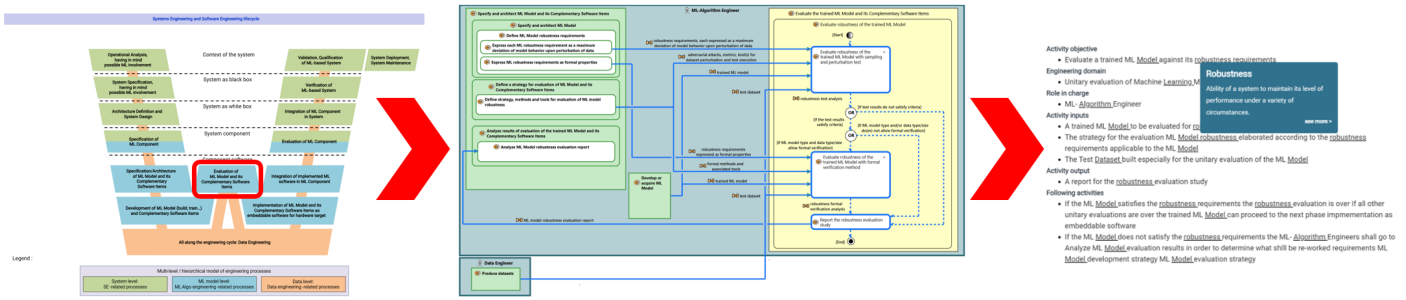

Fig. 4. ML Robustness evaluation query

approaches to information retrieval. But in order to support the search for a specific ML engineering knowledge by querying the KG and selecting the set of relevant engineering views to perform specific ML engineering activities, the identification of similarities between Confiance.ai documents needs to be enabled by searching for isomorphisms between the graphs representing the knowledge extracted from the text. There are several algorithms defined that implement subgraph isomorphism but, subgraph isomorphic problem is an NP-complete problem.

The first component is a generic sub-graph matching mechanism that works with fusion schemes. This will be responsible for the structural consistency of the merged information against the structures of the initial documents throughout the fusion process. The similarity and compatibility functions over the members of the graphs to be fused constitute the fusion approach. By adopting these strategies, the generic fusion algorithm can be adapted to the context in which it is used. Knowledge graph fusion method provides two other alternative operations depending on the fusion strategies used.

1) Information Synthesis: offers the ability to collect and arrange information about a specific topic. All of the collected information items are arranged into a network through information synthesis. The information items' duplicate parts are identified and removed. Fusing techniques are employed in information synthesis to allow the fusion of somewhat diverse information items that describe the same real-life situation. When several sources of information with possibly various precision levels are used to construct a picture of a specific view point, these inconsistencies may appear.
2) Information Query: Through this query function, within a network of information, all occurrences of information according to a specific graph pattern can be recognized. Because of the specialized relation between the query and data graphs, the query graph's structure must be totally contained within the data graph. The search for injective homomorphism between the query graph and the data graph is used by the information query function.

IV. The Confiance.ai BoK

Keyword-based queries have been frequently adopted to allow non-technical users to access large-scale RDF data. Today, the user can click on an engineering activity in the graph that capitalizes the end-to-end methodology and the underlying knowledge will be presented to them.

In terms of risk assessment, Confiance.ai specifically states that the probabilistic nature of ML-based systems requires new reliability analysis methods to assess the ability of such systems to meet reliability requirements, as well as novel approaches to assess the reliability of such ML-based systems. ML-based systems are notoriously challenging because of the difficulties in properly defining the environment, context, produced results and internal state, and in proposing a suitable definition of risk for deriving safety objectives and requirements. Furthermore, they raise concerns about the dependency between ML uncertainties and their contribution to overall system-level risk.

Thanks to all the descriptions, engineering knowledge and metrics and key performance indicators capitalized in the BoK, the Confiance.ai BoK is then a Trustworthy ML end-to-end engineering guideline that an engineer should follow throughout his/her engineering activities, assessing functional and non functional properties. Trustworthiness assessment aims to analyze and characterize trust expectations related to the targeted objectives [22]. It contributes, along with the ODD analysis process, to define the system's observable/measurable conditions and properties. Training ML models becomes tractable in terms of optimisation by evaluating ML performance based on quantitative accuracy or loss. In the meantime, predictive accuracy has been widely adopted to indicate the superiority of one AI product over another. However, recent widespread applications of ML have exposed the limitations of accuracy alone to a number of macro-reliability AI characteristics such as:

- Robustness, the system's outcome sensitivity to a change in the input;
- Effectiveness is a measure of its ability to perform the functions necessary to achieve goals or objectives;
- Dependability defined as the ability of a system to deliver a service that can be justifiably trusted;
- Usability describes the extent to which a product or system can be used by specified users to achieve specified goals effectively, efficiently and satisfactorily in a specified context;

- Human agency and oversight means developing and using AI as a tool that serves people, respects human dignity and autonomy, and is under human control and oversight.

For example in Fig. 4, the engineer wants to have information on how "assessing the ML Model Robustness" in the context of the activity of "Evaluation of ML Model". Then, the following activities are described:

- If the ML Model satisfies the robustness requirements the robustness evaluation is over, if all other unitary evaluations are over the trained ML Model can proceed to the next phase implementation as embeddable software.
- If the ML model does not meet the robustness requirements, the ML algorithm engineers go to analyze ML model evaluation results to determine what needs to be reworked ML model development strategy through the ODD definition to design an appropriate ML model evaluation strategy.

Moreover, the end-user can also make a query on the definition of the concept "robustness" provided by the Confiance.ai ontology (fig. 5).

In a future version of our BoK, the user will create questions in natural language, which are mapped in an intermediate language based on logic or by using LLM approaches.

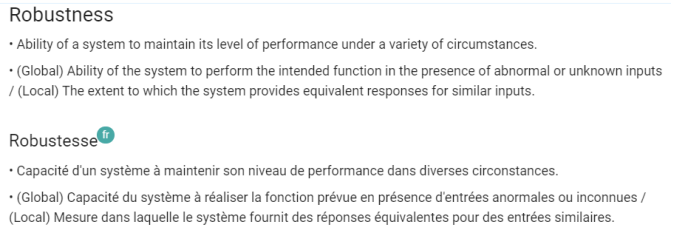

Fig. 5. ML Robustness evaluation query

V. Conclusion

The purpose of building a Trustworthy ML engineering Body-of-Knowledge graph is to assist ML-based system design engineers and ML-based system operations personnel in better specifying, designing understanding, monitoring and maintaining such systems, ultimately improving the safety, cyber security, reliability, availability, and performance. Building a trustworthy ML engineering involves various data sources and artificial intelligence techniques, such as knowledge representation, knowledge graph, semantic network, high level information fusion, graph theory, and more... Confiance.ai's methodological contributions cover the development process of a ML-based system, from initial specification and design to commissioning and supervision of its operation, including in embedded systems. These contributions are manifold:

- A taxonomy used in trustworthy AI;
- A complete documentation of the process, including modeling of activities and roles, with elements enabling corporate engineering departments to implement it;
- A first development of a Trustworthy AI ontology, linking the main concepts of the process and the taxonomy;
- And a "Body-of-Knowledge" which brings together all these elements and makes them accessible on the website of the same name.

While the use of these methodological tools is not in itself a guarantee of the ML-based system's compliance with regulations, it can constitute an element of justification, considered to be part of the state of the art by the notified bodies in charge of verification. This ML Engineering BoK is also a support tool of the following main challenges:

- How to design AI models, so that, by construction, they satisfy trustworthy properties (accuracy, robustness, etc.)?
- How to characterize these AI models, for example, to understand and explain their behavior and their adequacy to the operational domain?
- How to implement and embed those AI models on hardware, by making them fit for the target without losing their trustworthy properties.
- What are the data engineering method to apply in order to manage important volumes of data, account for the evolution of the operational domain, etc.?
- What are the appropriate verification, validation, and certification processes to consider for AI-based systems?

References

[1] J. Adam et al. Towards the engineering of trustworthy AI applications for critical systems - The *Confiance.ai* program, 2022.

[2] M. Adedjouma, C. Alix, et al. Engineering dependable AI systems. In *17th Annual System of Systems Engineering Conference (SOSE)*, pages 458–463. IEEE, 2022.

[3] A. Awadid, K. Amokrane-Ferka, H. Sohier, J. Mattioli, et al. AI Systems Trustworthiness Assessment: State of the Art. In *Workshop on Model-based System Engineering and Artificial Intelligence-MBSE-AI Integration 2024*, 2024.

[4] A. Awadid, X. Le Roux, B. Robert, M. Adedjouma, and E. Jenn. Ensuring the Reliability of AI Systems through Methodological Processes. In *The 24th IEEE International Conference on Software Quality, Reliability, and Security*, 2024.

[5] A. Awadid, B. Robert, and B. Langlois. Mbse to support engineering of trustworthy ai-based critical systems. In *12th International Conference on Model-Based Software and Systems Engineering*, 2024.

[6] P. Bourque, R. Dupuis, A. Abran, JW. Moore, and L. Tripp. The guide to the software engineering body of knowledge. *IEEE software*, 16(6):35–44, 1999.

[7] B. Braunschweig, R. Gelin, and F. Terrier. The wall of safety for AI: approaches in the *Confiance.ai* program. In *Proceedings of the Workshop on AI Safety 2022 (SafeAI 2022)*, volume 3087 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2022.

[8] Z. Chen, Y. Wang, B. Zhao, et al. Knowledge graph completion: A review. *IEEE Access*, 8:192435–192456, 2020.

[9] EASA. Concept paper first usable guidance for level 1 machine learning applications, 2021.

[10] L. Ehrlinger and W. Wöß. Towards a definition of knowledge graphs. *SEMANTiCS (Posters, Demos, SuCCESS)*, 48(1-4):2, 2016.

[11] D. Fensel, U. Simsek, K. Angele, , et al. *Knowledge graphs*. Springer, 2020.

[12] A Guide. Project management body of knowledge (pmbok® guide). In *Project Management Institute*, volume 11, pages 7–8, 2001.

[13] Shai Hertz, Mans Olof-Ors, Enav Weinreb, Oren Hazai, Geoff Horrell, Yael Lindman, Yehonatan Mataraso, and Phani Nivarthi. Machine learning-based relationship association and related discovery and search engines, May 28 2019. US Patent 10,303,999.

[14] A. Hogan, E. Blomqvist, M. Cochez, , et al. Knowledge graphs. *arXiv preprint arXiv:2003.02320*, 2020.

[15] S. Ji, S. Pan, et al. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems*, 33(2):494–514, 2021.

[16] C. Laudy. Semantic knowledge representations for soft data fusion. *Efficient Decision Support Systems-Practice and Challenges from Current to Future*, 2011.

[17] C. Laudy and JG. Ganascia. Introducing semantic knowledge in high-level fusion. In *MILCOM 2009-2009 IEEE military communications conference*, pages 1–7. IEEE, 2009.

[18] C. Laudy, JG. Ganascia, and C. Sedogbo. High-level fusion based on conceptual graphs. In *2007 10th international conference on information fusion*, pages 1–8. IEEE, 2007.

[19] J. Lehmann, R. Isele, et al. Dbpedia–a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic web*, 6(2):167–195, 2015.

[20] J. Mattioli, C. Laudy, PO. Robic, and HG. Chalé-Góngora. Body-of-knowledge development by using artificial intelligence. In *2022 17th Annual System of Systems Engineering Conference (SOSE)*, pages 14–19. IEEE, 2022.

[21] J. Mattioli, X. Le Roux, B. Braunschweig, Cantat, et al. AI engineering to deploy reliable AI in industry. In *2023 Fifth International Conference on Transdisciplinary AI (TransAI)*, pages 228–231. IEEE, 2023.

[22] J. Mattioli, H. Sohier, A. Delaborde, et al. An overview of key trustworthiness attributes and KPIs for trusted ML-based systems engineering. *AI and Ethics*, pages 1–11, 2024.

[23] LP. Meyer, C. Stadler, et al. Llm-assisted knowledge graph engineering: Experiments with chatgpt. In *Working conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow*, pages 103–115. Springer Fachmedien Wiesbaden Wiesbaden, 2023.

[24] ML. Mugnier and M. Chein. Conceptual graphs: Fundamental notions. *Revue d'intelligence artificielle*, 6(4):365–406, 1992.

[25] PA. Quezada-Sarmiento, LE. Enciso-Quispe, J. Garbajosa, and H. Washizaki. Curricular design based in bodies of knowledge: Engineering education for the innovation and the industry. In *2016 SAI Computing Conference (SAI)*, pages 843–849. IEEE, 2016.

[26] JF. Sowa. Conceptual graphs for a data base interface. *IBM Journal of Research and Development*, 20(4):336–357, 1976.