

## Doctoral Research Project: Combining Logic-based Approaches and Neural Networks for Effective Defeasible Reasoning over Large-Scale Biomedical Ontologies

**Research fields:** Computer Science and Medical Informatics (Biomedical Ontologies, Knowledge Representation and Reasoning, Neural Networks)

**Laboratory:** : Laboratoire de Recherche en Informatique pour la Santé (Limics), UMRS 1142

**Supervision:**

Chan LE DUC, full professor in Computer Science, Université Sorbonne Paris Nord.

Ivan VARZINCZAK (<https://ijv.ovh>), full professor in Computer Science, Université Sorbonne Paris Nord.

**Contacts:** [chan.leduc@univ-paris13.fr](mailto:chan.leduc@univ-paris13.fr), [ivan.varzinczak@univ-paris13.fr](mailto:ivan.varzinczak@univ-paris13.fr)

**Keywords:** Knowledge representation; formal ontologies; description logics; automated reasoning; neural networks; learning; defeasible reasoning; rationality.

**Description :** Artificial intelligence (AI) methods grounded on explicit symbolic representations of knowledge remain an active area of research. This is the case despite the impressive success recently achieved by AI systems based on statistical methods such as deep learning and alike. Indeed, in numerous AI application scenarios, there is a need to know how declarative knowledge is represented as a knowledge base (KB) and what conclusions follow from it in a principled way, i.e., through the application of logical rules formalising widely accepted reasoning principles. A particular feature of logic-based approaches to AI is the fact that they lend themselves naturally to being explained, making it possible for users to trace back the relevant portion of knowledge in a KB playing a role in drawing a given conclusion. This is particularly useful in medical decision-support systems where healthcare practitioners frequently need to know more about a given conclusion before making a decision.

Description logics (DLs) are among the logical formalisms that are central to many modern AI applications. This is primarily due to the fact they provide the logical foundations of formal ontologies, a pivotal component of semantic technologies meeting a great deal of applications, in particular in the medical domain. For instance, a large number of biomedical ontologies are represented in some sublanguage of the OWL 2 Profile corresponding to a variant of DLs such as  $\mathcal{EL}$ ,  $\mathcal{EL}^{++}$  [1] and others.

Since formal ontologies are meant to capture human knowledge about a given domain, they are prone to containing exceptions, which can lead to inconsistency. Furthermore, unlike (classical) mathematical reasoning, which is always sound and unrevisable, human reasoning is often tentative, incomplete, and, as a result, naturally up for revision. Therefore, endowing DLs with defeasible-reasoning features is a promising endeavour, drawing on a well-established body of research on non-monotonic reasoning in the field of knowledge representation and reasoning. Preferential extensions of DLs turn out to be particularly promising, mostly because they are based on an elegant, comprehensive and well-studied framework for non-monotonic reasoning in the propositional case proposed by Kraus, Lehmann and Magidor in the early 1990s, the so-called KLM approach [7, 8].

In recent work [2], we have provided a comprehensive framework for the formal basis of defeasible reasoning in DLs by (i) defining a general and intuitive semantics; (ii) showing a corresponding representation result linking the semantics with the KLM-style properties; (iii) presenting an appropriate analysis of entailment in the context of ontologies with defeasible information, and (iv) providing an implementation of the underlying theory and an accompanying set of experiments run on sizeable ontologies with defeasible information. Part of the research to be conducted in this PhD proposal is to extend the work mentioned above by introducing defeasibility for various reasoning services, particularly conjunctive queries for DLs. In particular, we intend to (i) extend the formal semantics for defeasibility in DLs to encompass conjunctive queries for DLs, and (ii) investigate efficient algorithms for defeasible conjunctive queries in so-called lightweight DLs, namely the  $\mathcal{EL}$  and DL-Lite families.

Despite the advantages logic provides in terms of modelling intuitiveness, explainability, and verifiability, its indiscriminate use often comes with a high cost: The computational complexity of query answering for logical languages of highly expressive power is intractable in the general case. This means purely logic-based approaches to (classical or non-classical) reasoning over large medical ontologies do not scale well in practice. This forms part of the argument for an exploration of the so-called neuro-symbolic approaches to AI, which are acknowledged by the AI community as a new frontier.

Powerful tools based on Graph Convolutional Networks (GCN) such as GraphSAGE [6] are widely used in recommendation systems [13]. These tools rely on statistical approaches such as deep learning that involve training multi-layered neural networks to optimise an objective function, such as minimising prediction error or reconstructing input data. The learning process seeks convergence, in which model parameters are iteratively updated until the objective function stabilises, i.e., the neural network has found parameter values that minimise prediction errors on the training data. The output of the training process is a trained model that can be used to make predictions on new and unseen data. For instance, one can provide as input a large set of labelled graphs for training a GCN. The trained model can predict properties of a new graph, such as the presence of specific sub-graph patterns, without the need for exhaustive exploration as required by symbolic methods [11]. This principle of deep learning has opened up a new research field, namely neuro-symbolic computing [5], which aims at integrating two of the most fundamental cognitive abilities humans possess: the ability to learn from the environment, and the ability to reason from what has been learned [11, 12, 9].

The poor performance of existing DL reasoners for large-scale ontologies stems from the nondeterminism resulting from logical constructors such as disjunction, negation, and number restrictions [10], which are very useful in the specification of biomedical ontologies [3]. These reasoners are based on tableaux or saturation algorithms that are forced to create several graphs when dealing with disjunction, negation, and number restrictions in the graph labels, since they have to store all of these graphs to be able to backtrack whenever the non-deterministic choice that has been made is a bad one. This leads to generating an exponential number of different graphs in the worst case. By integrating a trained model resulting from a neural network into a tableau-based algorithm, one can consult the model during the proof search and get a suggestion on more promising paths when dealing with nondeterminism. Such a trained model should result from a large dataset of training graphs generated by classical reasoners run offline on a large set of ontologies. Therefore, the main goal of this doctoral research project is the proposal of a neuro-symbolic approach consisting in the integration of a neural network within reasoning algorithms for querying large DL ontologies.

It is strongly expected that the successful candidate will implement the proposed methods, integrate them into inference engines such as OntoRev [4] or ARROW<sup>1</sup> developed by the team, and evaluate the methods with large-scale biomedical ontologies containing nondeterministic constructors such as MONDO.<sup>2</sup>

### Expected qualifications and skills:

- Masters degree in Computer Science.
- Good knowledge of formalisms and techniques for the Semantic Web.
- Good background on Deep Learning.
- Ability to design algorithms and program in both Python, Java or C/C++.
- Communication skills, ability and willingness to interact with peer researchers.
- Proficiency in English.

### Provisional schedule:

- Period 1: Literature review on both defeasible reasoning and deep learning; Consolidation of the student's knowledge on description logics and OWL ontologies; Implementation of Protégé plugins that will serve as modules of the future general system; Presentation of seminars at the local research group; Writing of workshop papers.
- Period 2: Assessment of the most suitable learning model for proof-search optimisation; Design of the learning model and tests in the classical case; Assessment of the challenges in extending it to the defeasible case; Presentation of seminars at the local research group; Writing of papers for international conferences.
- Period 3 : Integration of the learning models into the reasoning algorithms; Extension to more expressive languages equipped with further defeasible constructors; Tests; Presentation of seminars at the local research group; Writing of papers for international conferences.

---

<sup>1</sup><https://gitlab.inria.fr/DLreasoners/arrows>

<sup>2</sup><https://purl.obolibrary.org/obo/mondo>

- Period 4 : Wrapping up of the proposed framework; Evaluation; Writing of papers for presentation at international conferences; Writing of the thesis and PhD defence.

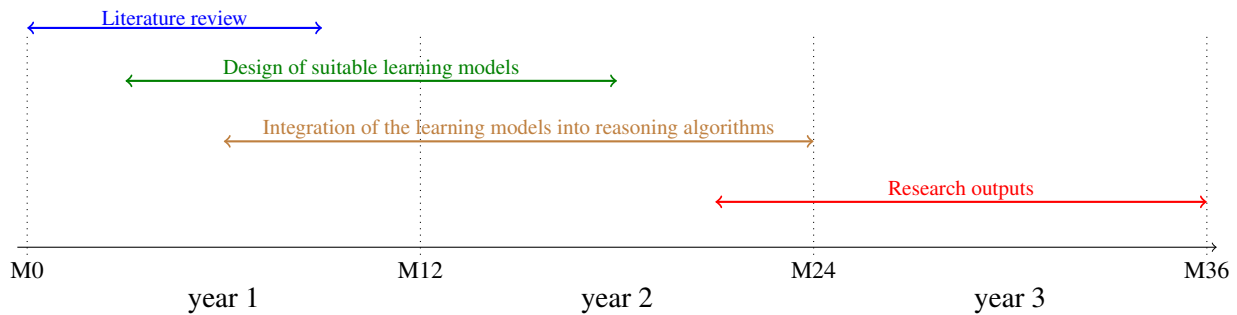


Figure 1: Provisional Schedule for **Period 1**, **Period 2**, **Period 3**, and **Period 4**.

## REFERENCES

- [1] Baader, F., Brandt, S., Lutz, C.: Pushing the  $\mathcal{EL}$  envelope. In: IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK. pp. 364–369 (2005)
- [2] Britz, K., Casini, G., Meyer, T., Moodley, K., Sattler, U., Varzinczak, I.: Principles of KLM-style defeasible description logics. *ACM Transactions on Computational Logic* 22(1) (2021)
- [3] Ceusters, W., Elkin, P., Smith, B.: Negative findings in electronic health records and biomedical ontologies: A realist approach. *International Journal of Medical Informatics* 76, S326–S333 (2007)
- [4] Dong, T., Le Duc, C., Lamolle, M.: Tableau-based revision for expressive description logics with individuals. *J. Web Semant.* 45, 63–79 (2017)
- [5] d’Avila Garcez, A.S., Gori, M., Lamb, L.C., Serafini, L., Spranger, M., Tran, S.N.: Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning. *FLAP* (2019)
- [6] Hamilton, W.L., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems*. pp. 1024–1034 (2017)
- [7] Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence* 44, 167–207 (1990)
- [8] Lehmann, D., Magidor, M.: What does a conditional knowledge base entail? *Artificial Intelligence* 55 (1992)
- [9] Manigrasso, F., Schouten, S.F., Morra, L., Bloem, P.: Probing LLMs for logical reasoning. In: *Neural-Symbolic Learning and Reasoning*. pp. 257–278. Springer (2024)
- [10] Motik, B., Shearer, R., Horrocks, I.: Hypertableau reasoning for description logics. *J. of Artificial Intelligence Research* 36, 165–228 (2009)
- [11] Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., Monfardini, G.: The graph neural network model. *Trans. Neur. Netw.* 20(1), 61–80 (2009)
- [12] Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826* (2018)
- [13] Ying, R., He, R., Chen, K., Eksombatchai, P., Hamilton, W.L., Leskovec, J.: Graph convolutional neural networks for web-scale recommender systems. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*. pp. 974–983 (2018)