prof. Anna Gambin
University of Warsaw
a.gambin@uw.edu.pl

16 November 2026

Reviewer's opinion
on Ph.D. dissertation authored by
Adam Aron Rynkiewicz
entitled:
FROM LITERATURE TO MODELS: SEMI-AUTOMATIC CONSTRUCTION OF BIOMEDICAL PETRI NET-BASED MODELS

1. Problem and its impact

The dissertation deals with the semi-automatic extraction of information on interactions between metabolites from scientific articles, and with the use of this information to build computational models of biological processes. The number of biomedical publications is growing rapidly. PubMed alone adds more than a million new articles every year. It is no longer possible for a single expert to follow all the results relevant to a given process, so a lot of useful knowledge remains scattered across individual papers.

Turning this unstructured knowledge into a model, e.g. one based on Petri nets, is difficult. The model has to be suitable for simulation and analysis, but it also has to keep its biomedical meaning. Most of the relevant concepts depend on context. The same metabolite may appear under several names, an interaction may be activating in one setting and inhibiting in another, and a result obtained for one organism or tissue cannot simply be generalised. This is why the author argues that the process should be semi-automatic and should involve experts. I agree with this view. Models built in this way are much better suited to study the mechanisms of diseases and physiological processes. The problem is clearly a scientific one and it has considerable practical value.

It should be noted that hardly anyone, including the author, anticipated how quickly AI tools would develop in the last few years. Today, LLMs can extract interactions between metabolites from text and can even propose a model. Nevertheless, I do not believe that a fully automatic process is the right choice here. LLMs may return interactions that look plausible but are not supported by the text, they may confuse metabolites or the direction and type of an interaction, and they often ignore the biological context. When such a model is simulated, these errors propagate directly to the results. For this reason, I consider the evidence-based approach used by the author particularly valuable. Each extracted relation is linked to the fragment of the source text that supports it, so an expert can quickly check every element of the model against the original paper and correct it if needed. The approach proposed in the dissertation is therefore still justified, even with today's LLMs.

2. Contribution

The contribution of the dissertation consists of two main parts: new computational methods for extracting knowledge from biomedical literature, and new biomedical models built with these methods.

On the methodological side, the author first followed the classical pipeline: recognition of entities such as metabolites, linking them to entries in reference knowledge bases, and extraction of interactions between them. For these steps, he improved existing models and proposed his own, including the FELA method for entity linking. Later, when LLMs made it possible to extract entities and relations in one step, the author built the Watson platform for the semi-automatic construction of interaction networks from scientific texts. Every relation returned by the platform comes with the text fragment that supports it, so the resulting networks can be verified by experts.

The second part of the contribution is the modelling work. The author developed a new Petri net model of processes related to HDL cholesterol under inflammation and oxidative stress. The model was checked by domain experts and analysed further, e.g. with an *in silico* knockout analysis. The dissertation therefore does not stop at proposing tools; it also shows that they can produce a model that is useful in biomedicine.

The main results of the dissertation have been published in peer-reviewed venues. The paper on the lipid model appeared in the *International Journal of Molecular Sciences*. The methodological results were published in two papers: one presented at an international workshop (HybridAIMS and CAI 2025, co-located with CAiSE 2025), and one on entity linking in the journal *Engineering Applications of Artificial Intelligence*.

3. Correctness

In my opinion, the results of the dissertation are reliable. The work is interdisciplinary, so its correctness has to be assessed in two respects: whether the models are biomedically valid and whether the computational methods are sound. In my opinion, both conditions are met.

The Petri net model of lipid metabolism was verified by domain experts, who confirmed that it agrees with the current knowledge of these processes. This is essential, because the model is meant for simulation, and errors in its structure would affect all later analyses.

The computational methods were compared with existing solutions in terms of quality, accuracy and efficiency. For example, FELA was compared with the ReFinED system. The Watson system was evaluated both manually and automatically on the task of finding biologically meaningful interactions in scientific texts. I value this combination. Automatic evaluation gives objective numbers on a larger scale, while manual evaluation shows whether the interactions found really make sense biologically, which standard metrics do not capture well. The author also looked at practical aspects, such as the amount of training data needed and the speed of the system.

I found only minor technical weaknesses, mostly concerning the readability of figures and the description of the experimental setup. They do not change my view on the

correctness of the results. I list them in Sec. 5.

Overall, the software tools are of good quality and were properly validated, and the models built with them and verified by experts are further evidence that they operate correctly.

4. Knowledge of the candidate

The dissertation confirms that the candidate has general knowledge and understanding of the **Information and Communication Technology** discipline.

The dissertation has eight chapters. The first three contain the introduction, the theoretical background and the literature review. The theoretical part covers the basics of Petri net theory: the main concepts, analysis methods such as invariants, and the main classes of Petri nets. It also introduces large language models, including the transformer architecture and the optimisation methods used to train neural networks and LLMs.

The theoretical part also defines the problems of named entity recognition, entity linking and relation extraction. Chapter 3 reviews the literature on these three problems and on the computational methods used to solve them. It also summarises the literature on high-density lipoproteins and inflammation, and describes the current research on the automatic construction of Petri net models.

These chapters show that the candidate is familiar with computational methods and recent advances in AI, and also systems biology, including how systems biology models can be built, also automatically. This combination is relatively rare, and it is needed in this kind of research. To apply AI to biomedical problems, one has to understand what the methods can and cannot do, and also the nature of the biological knowledge being modelled. Therefore, I conclude that the candidate has general knowledge and understanding of the **Information and Communication Technology** discipline.

5. Other remarks[1]

The remarks below are mostly technical and do not lower my high opinion of the dissertation.

- The Petri net figures in Chapter 4 (Figs. 4.2, 4.3 and 4.4) are very hard to read. Places and transitions are identified by tiny numbers, which makes the diagrams difficult to follow. The same applies to Figs. 7.3 and 7.4 in Chapter 7.
- The heatmap with the results of the *in silico* knockout analysis is also hard to read, because the row and column labels are too small.
- In the chapters on computational methods (Chapters 5, 6 and 7), a simple workflow diagram of the experiments would considerably improve the clarity of the presentation. It should show what is compared with what, on which data, and which metrics are used to assess the results.

[1] Optional

6. Conclusion

Taking into account what I have presented above and the requirements imposed by Article 187 of the Act *of 20 July 2018 – The Law on Higher Education and Science* (with amendments)[2], my evaluation of the dissertation according to the three basic criteria is the following:

A. Does the dissertation present an original solution to a scientific problem? (the selected option is marked with **X**)

[X]	[ ]	[ ]	[ ]	[ ]
Definitely YES	*Rather yes*	*Hard to say*	*Rather no*	*Definitely NO*

B. After reading the dissertation, would you agree that the candidate has general theoretical knowledge and understanding of the discipline of **Information and Communication Technology**?

[X]	[ ]	[ ]	[ ]	[ ]
Definitely YES	*Rather yes*	*Hard to say*	*Rather no*	*Definitely NO*

C. Does the dissertation support the claim that the candidate is able to conduct scientific work?

[X]	[ ]	[ ]	[ ]	[ ]
Definitely YES	*Rather yes*	*Hard to say*	*Rather no*	*Definitely NO*

In conclusion, I state that the dissertation meets the requirements of Article 187 of the Act, and I request that the candidate be admitted to the further stages of the doctoral procedure, including the public defence.

_____

Signature

[2] http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20190000276