Wydział Automatyki, Elektroniki i Informatyki
Katedra Inżynierii i Biologii Systemów

dr hab. inż. Profesor Politechniki Śląskiej
Sebastian Student
Dyrektor Centrum Biotechnologii

Gliwice, 18.11.2026

Reviewer's opinion on Ph.D. dissertation

Dissertation title: From literature to models: semi-automatic construction of biomedical Petri net-based models

Author of the dissertation: mgr inż. Adam Aron Rynkiewicz

Dissertation supervisor: prof. dr hab. inż. Piotr Formanowicz

Problem and its impact

The reviewed doctoral dissertation addresses a highly current and important problem from the perspective of modern bioinformatics: the transformation of vast amounts of unstructured knowledge from biomedical literature into structured, formal mathematical models. Modern systems biology faces the barrier of knowledge fragmentation. Although millions of articles are published every year, integrating the phenomena they describe into coherent models requires enormous effort from experts. Adam Aron Rynkiewicz, MSc, Eng., proposed a solution using Petri nets, which combine an intuitive graphical representation with mathematical rigor, along with advances in natural language processing (NLP) and large language models (LLMs).

The author's methodological maturity deserves special recognition, as he rejects the naive pursuit of "full automation" in favor of a semi-automatic, human-in-the-loop approach. The doctoral candidate rightly notes that in scientific applications, precision, interpretability, and traceability of extracted relations to supporting source evidence are crucial, and artificial intelligence should serve as an assistant to the domain expert.

Contribution

I would like to strongly emphasize the exceptional scale and multidimensionality of the independent research, engineering, and analytical work performed by the Doctoral Candidate. The work goes far beyond the standard scope of a computer science dissertation, constituting a rare and highly

Politechnika Śląska
Wydział Automatyki, Elektroniki i Informatyki
Katedra Inżynierii i Biologii Systemów
ul. Akademicka 16, 44-100 Gliwice
Centrum Biotechnologii
ul. Krzywoustego 8, 44-100 Gliwice
+48 32 237 29 06
Sebastian.student@polsl.pl
NIP 631 020 07 36

successful combination of software engineering, advanced AI model training, and systems biology. The main, original contribution includes:

- **Independent creation of datasets:** Due to the lack of ready-made, public corpora linking biomedical literature with Petri net notation, the Author generated massive training and evaluation datasets, as well as an evidence attribution dataset comprising approximately 90,000 relation-evidence pairs.
- **Hands-on fine-tuning of models:** The doctoral candidate did not limit himself to calling APIs of commercial solutions. He independently optimized and fine-tuned an entire family of his own language models (including *cannary-oa-v2*, *cannary-reasoning-v2*, *be-0.6b*), operating on advanced cluster infrastructure (NVIDIA H100) and successfully applying techniques such as Supervised Fine-Tuning and Liger kernel optimizations.
- **Creation of the Watson architecture and platform:** The doctoral candidate designed and implemented an asynchronous backend system (utilizing FastAPI, Celery, RabbitMQ, and PostgreSQL), which binds the relation extraction process into a fully usable research tool.
- **Expert manual validation:** The doctoral candidate demonstrated deep knowledge in the field of biology (HDL metabolism, inflammatory states) by manually assessing hundreds of extracted biological relations for biological plausibility and consistency with the supporting source literature.

Correctness

The strongest methodological aspect of the work is the multi-layered, extremely rigorous process of verifying and evaluating the results. The Doctoral Candidate avoids the common mistake of relying on a single metric and supports the validity of his methodological claims through complementary forms of evaluation:

- **Rigorous data separation preventing data leakage:** The Doctoral Candidate paid particular attention to data separation in the reported experiments. When constructing evaluation sets for the Watson platform, the Author explicitly ensured that the publications used did not overlap with the training corpus, and he evaluated the improved mechanism on previously unseen full-text articles from PubMed. When constructing evaluation sets for the Watson platform, the Author explicitly ensured that the publications used did not overlap with the training corpus, and he evaluated the improved mechanism on entirely new, previously unseen full-text articles from PubMed. This demonstrates a strong research culture.
- **Structural analysis (*in silico knock-out*):** Instead of evaluating the generated Petri nets solely in terms of topology, the Author investigated the dynamics of the system through the

Politechnika Śląska
Wydział Automatyki, Elektroniki i Informatyki
Katedra Inżynierii i Biologii Systemów
ul. Akademicka 16, 44-100 Gliwice
Centrum Biotechnologii
ul. Krzywoustego 8, 44-100 Gliwice
+48 32 237 29 06
Sebastian.student@polsl.pl
NIP 631 020 07 36

simulated disabling (knock-out) of specific transitions. The simulations demonstrated the model's ability to identify potential regulatory bottlenecks within the represented biological network, highlighting transitions whose removal leads to substantial system-wide changes.

- **Expert manual evaluation:** The Author compared the relations generated by the Watson system with manually prepared reference Petri nets for the atherosclerosis process and HDL metabolism. A random 10% sample of the extracted relations was manually evaluated, yielding precision values of 85.46% for atherosclerosis and 82.82% for HDL metabolism.
- **Automated semantic evaluation (LLM-as-a-judge):** The Author prepared a "pseudo-gold standard" using powerful reference models (e.g., DeepSeek R1), and then used independent judge models to evaluate the results. This evaluation assessed semantic meaning, rather than strict character matching.
- **Built-in internal verification (IIE and ECC):** The doctoral candidate embedded verification mechanisms within the process itself. The Iterative Information Extraction (IIE) procedure marks found evidence, preventing hallucinations. In turn, the Extractor-Critic-Consensus (ECC) extension implements a multi-agent solution in which a council of models votes on the correctness of the extracted relation, forcing corrections based on the generated critique. This shifts the paradigm towards higher precision and improved reliability through multi-model verification.

Despite such a rigorous workshop, the work is not free from certain objective methodological limitations, most of which the Doctoral Candidate independently identifies:

- **Lack of handling for negative results:** The extraction schema (substrates, modifiers, products) is oriented towards positive events. The work does not explain how the system handles negative validation (e.g., sentences refuting the existence of an interaction), which risks losing logical negation.
- **Qualitative nature of the models:** The constructed Petri net model is based on the direction of interaction and does not contain precise kinetic rate constants, which prevents the determination of absolute concentrations over time.
- **Problems with segmentation and data from figures:** Segmentation into fragments of about 512 tokens could cause the breaking of semantic links. Furthermore, the text-only approach can lose relations hidden in figures and diagrams.
- **Potential for interdisciplinary validation:** Given that the Watson platform is a *human-in-the-loop* tool, inviting independent molecular biologists or medical professionals for testing would be a valuable future step to verify the target usability of the interface. In the automated evaluation, there is also some sensitivity to the reference set used.

Politechnika Śląska
Wydział Automatyki, Elektroniki i Informatyki
Katedra Inżynierii i Biologii Systemów
ul. Akademicka 16, 44-100 Gliwice
Centrum Biotechnologii
ul. Krzywoustego 8, 44-100 Gliwice
+48 32 237 29 06
Sebastian.student@polsl.pl
NIP 631 020 07 36

Knowledge of the candidate

The dissertation unequivocally confirms the candidate's deep theoretical and general knowledge in the discipline of information and communication technology. Chapters 2 (Theoretical Foundations) and 3 (Literature review) serve as a reliable, comprehensive compendium of the subject. The candidate navigates fluently and with full understanding of advanced artificial neural network architectures (transformer-based architectures, attention mechanisms), vector databases, optimization methods, and the Petri net formalism. Impressively, he also demonstrates remarkable proficiency in biology and medicine (lipoprotein metabolism, inflammatory states), which is crucial in modern bioinformatics. The list of references includes dozens of the latest, leading publications, confirming the Doctoral Candidate's excellent orientation in global research trends.

Other remarks

Minor editorial flaws and typos are noticeable in the work; however, they do not diminish its scientific value. The most noticeable include:

- Inconsistency in reporting results: On page 58, when describing the model trained using the ORPO method (named *anydef*), the Author rightly states that it surpassed all previous methods. However, unlike the SFT and DPO models, he did not provide the precision score it achieved.
- Unfinished references and inconsistencies: On page 35, an unfilled reference tag was left: "illustrated in Figure X".
- Unreadability of extensive Petri net diagrams: Figures 4.2, 4.3, 4.4, 7.3, and 7.4 in standard print page size are practically unreadable due to the very high density of nodes. In the Author's favor is the fact that they are embedded in the electronic version as vector graphics.
- Errors in generated charts: On the heatmap (Fig. 4.5, p. 45), the axes contain typos, such as "oxidation_*recudtion*" (instead of reduction) and "*absorpted*_FA..." (instead of absorbed).

Conclusion

Taking into account the arguments presented above, I state with full conviction that the reviewed dissertation constitutes an original, deeply thought-out, and comprehensive solution to an extremely difficult bioinformatic problem. The combination of biological knowledge with advanced artificial intelligence algorithms and formal methods of analysis and verification has yielded a highly useful and innovative result. The weaker points I noted are secondary and often stem from the field's inherent limitations.

Politechnika Śląska
Wydział Automatyki, Elektroniki i Informatyki
Katedra Inżynierii i Biologii Systemów
ul. Akademicka 16, 44-100 Gliwice
Centrum Biotechnologii
ul. Krzywoustego 8, 44-100 Gliwice
+48 32 237 29 06
Sebastian.student@polsl.pl
NIP 631 020 07 36

It is also worth strongly emphasizing the outstanding publication record of the Doctoral Candidate, which directly reflects the research conducted in this dissertation. The published scientific articles are organically and directly related to the dissertation. The article entitled "Regulatory Bottlenecks Linking HDL Dysfunction, Inflammation, and Oxidative Stress: A Petri Net-Based Systems Analysis" (Int. J. Mol. Sci., 2026), co-authored by the Doctoral Candidate, forms the core of the dissertation's biological component and is included in the bibliography. In turn, the article entitled "Universal entity linking" (Engineering Applications of Artificial Intelligence, 2025), co-authored by Rynkiewicz as the first author, presents the entity-linking approach developed in Chapter 6 of the dissertation, particularly the generation of contextual entity profiles and vector-based entity disambiguation. Although this work focuses on general-purpose knowledge bases and benchmark datasets rather than biomedical relation extraction, it constitutes an important methodological contribution to the broader research framework presented in the dissertation. It should be noted that this second, published article was not included by the author in the final bibliography of the reviewed dissertation. However, this does not change the fact that it is a resounding confirmation of the excellent quality of his research skills.

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), 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 can conduct scientific work?

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

Politechnika Śląska
Wydział Automatyki, Elektroniki i Informatyki
Katedra Inżynierii i Biologii Systemów
ul. Akademicka 16, 44-100 Gliwice
Centrum Biotechnologii
ul. Krzywoustego 8, 44-100 Gliwice
+48 32 237 29 06
Sebastian.student@polsl.pl
NIP 631 020 07 36

HR EXCELLENCE IN RESEARCH
CENTRUM BIOTECHNOLOGII Politechniki Śląskiej

Taking into account the publications in renowned international journals, the novelty of the implemented artificial intelligence engineering methods, their successful translation into the field of systems biology, and the exceptional amount of independent work put into the development, model training, and rigorous validation of the Watson platform, **I recommend to distinguish** the dissertation for its quality.

Signature

Politechnika Śląska
Wydział Automatyki, Elektroniki i Informatyki
Katedra Inżynierii i Biologii Systemów
ul. Akademicka 16, 44-100 Gliwice
Centrum Biotechnologii
ul. Krzywoustego 8, 44-100 Gliwice
+48 32 237 29 06
Sebastian.student@polsl.pl
NIP 631 020 07 36