Wrocław University
of Science and Technology

Department of Computer Science and Systems Engineering

Remigiusz Szczepanowski, Ph.D., D.Sc.
Department of Computer Science and Systems Engineering, Wroclaw University of Science and Technology, Poland
E-mail: remigiusz.szczepanowski@pwr.edu.pl

April 23, 2025

Reviewer's opinion

on Ph.D. dissertation authored by

Klaudia Kantor

entitled:

Neural language models for clinical trial eligibility criteria

1. Problem and its impact

The reviewed doctoral dissertation by Klaudia Kantor, M.Sc. titled" ***Neural language models for clinical trial eligibility criteria*** " addresses the problem of improving the process of recruiting patients for clinical trials using natural language processing (NLP) methods. Clinical trials are meant to introduce new medical therapies, but it is only possible with correctly defined patient inclusion and exclusion criteria. The author of the dissertation points out that eligibility criteria for clinical trials are complex and written in the form of unstructured texts. Because of this, their analysis and comparison with patient data are mostly done manually by medical personnel. According to the author, the efforts of health professionals adjusting fragmentation of eligibility criteria, into logical sections is also one of the reasons why patient data entry and patient categorization by personnel are prone to error. The dissertation shows that the lack of practical tools for automatic analysis of eligibility criteria is a significant problem in both clinical trials and medical informatics applications. Given the above findings, the doctoral dissertation hypothesizes that "***Neural language models can significantly enhance the efficacy of parsing clinical trial eligibility criteria, out- performing traditional methods and consequently enhancing patient recruitment in clinical trials"*** .

From a scientific standpoint, this problem is important because it requires advanced natural language processing in the biomedical domain. Eligibility criteria texts contain specialized medical terminology and have a complex logical structure (e.g., combinations of patient conditions, laboratory results, medical history, etc.). Their correct automatic interpretation requires a combination of NLP methods and medical domain knowledge. The author suggests that previous approaches to parsing eligibility criteria have been inadequate. Traditional methods, such as rule-based systems or regular expressions, have proven

Wrocław University
of Science and Technology

Faculty of Information
and Communication Technology

Department of Computer Science
and Systems Engineering

27 Wybrzeże Wyspiańskiego St
50-370 Wrocław

T: +48 71 320 30 74

www.pwr.edu.pl
www.kinf.pwr.edu.pl
sekretariat.k44.wit@pwr.edu.pl

REGON: 000001614
NIP: 896-000-58-51
Nr konta:
37 1090 2402 0000 0006 1000 0434

inflexible and difficult to transfer to other therapeutic fields. The older NLP techniques also are not fully capable to embrace the complex relationships in these documents. In this context, the problem addressed in the dissertation is part of a broader research challenge of applying modern neural models based on transformer architecture to solve a real-world problem in medical informatics and medicine. It should be noted that the use of large-scale linguistic models to analyze eligibility criteria is a novel aspect of the dissertation, as there are not many research works addressing this topic to date, which makes the results of Klaudia Kantor's dissertation valuable and original. In conclusion, the research problem proposed in the dissertation on automatic linguistic processing of eligibility criteria for clinical trials is of great practical and scientific significance.

2. Contribution

The dissertation provides important scientific contributions at the intersection of computer science and medicine. The author's main achievements include:

a.) Analysis of the clinical trial design features on the recruitment efficiency and overall trial duration: In an early phase of the work (Ch. 3), the author presented a quantitative analysis of how clinical trial design features, particularly the number of eligibility criteria, affect trial operational efficiency. The study used data from 2,051 completed clinical trials (Roche, 2009–2020) and evaluates five operational metrics: screen failure ratio, dropout ratio, pre-enrollment duration, enrollment duration, and overall study duration. Using machine learning models (LightGBM), the analysis identified the number of eligibility criteria and planned patient enrollment as among the important predictors of recruitment inefficiency and prolonged trial timelines. These relationships were non-linear and not detected by traditional linear regression models. High protocol complexity, identified in the number of procedures, planned visits, and endpoints, was consistently associated with poorer recruitment outcomes and higher patient attrition. This finding confirms the practical validity of the dissertation topic and motivates further research on automating criteria analysis.

(b). Literature review and identification of research gaps: The author conducted the most up-to-date and comprehensive review of research on using NLP in clinical trial eligibility criteria analysis. Chapter 5 of dissertation discusses existing approaches and their limitations. It was shown that most existing approaches are based on outdated technologies (manual rules, simple statistical methods, older language models) and lack applications of the latest transformative or generative models. This critical review has identified a research gap which is the need to apply modern, learning methods to this problem. Note that this comprehensive review of the literature on processing eligibility criteria has been recently published in 2024 on Drug Discovery Today.

unite! University Network for Innovation, Technology and Engineering

HR EXCELLENCE IN RESEARCH

Wrocław University
of Science and Technology

Faculty of Information
and Communication Technology

Department of Computer Science
and Systems Engineering

27 Wybrzeże Wyspiańskiego St
50-370 Wrocław

T: +48 71 320 30 74

www.pwr.edu.pl
www.kinf.pwr.edu.pl
sekretariat.k44.wit@pwr.edu.pl

REGON: 000001614
NIP: 896-000-58-51
Nr konta:
37 1090 2402 0000 0006 1000 0434

(c). Research on semantic similarity of medical texts: The dissertation (Ch. 6) presents experiments evaluating different semantic representations of sentences in the biomedical domain. The author compared the performance of classical and state-of-the-art text embedding models (including transformer-based models) in the task of semantic text similarity for qualifying criteria descriptions. The experiments evaluated how different text preprocessing methods, when combined with sentence representation strategies, perform in the biomedical semantic textual similarity task. The study found that sentence transformers trained on general-domain data can outperform specialized biomedical BERT models in biomedical semantic textual similarity tasks, showing greater versatility and higher overall performance. The CODER model also performed well, winning with most biomedical BERT models, and achieving the best results in binary classification.

(d). Conducting a comparative analysis of the GPT-4 model with BERT models in the task of recognizing named entities (NER): One of the key contributions of the dissertation was the innovative use of the GPT-4 large language model for the task of recognizing named entities (NER) in eligibility criteria (Chapter 7). The author investigated the extent to which the GPT-4 model (used through the so-called GPT-4-turbo API) can recognize relevant medical entities (e.g., names of diseases, drugs, clinical parameters) in the content of criteria, and compared its effectiveness with classical approaches based on BERT models trained specifically for NER. It turned out that BERT-based pretrained models performed better than GPT-4-turbo when fine-tuning conditions were applied. Experimental results revealed also that GPT-4 had potential in situations where fine-tuning data are limited. Thus, the research demonstrates the viability of using a straightforward few-shot prompting approach with GPT-4 for NER tasks in clinical trial eligibility criteria, especially beneficial when no fine-tuning data is available. Despite GPT-4's promising results, pretrained BERT models, considerably smaller in size, consistently showed superior performance. It was also shown that the CODER model pretrained on the UMLS ontology exhibited outstanding capabilities, significantly enhancing the recognition of specialized medical terms. The broader linguistic and grammatical strengths of GPT-4-turbo proved less effective in improving medical NER accuracy.

(e). Organization of a hackathon and practical evaluation of LLM results in the analysis of eligibility criteria: An important part of the work was the holding of a hackathon with the participation of experts from Roche (Ch. 8), which tested the application of LLMs (including GPT-4) for criteria analysis in a near real-world setting. The main objective of the hackathon was to design prompts for an internal ChatGPT-based tool capable of converting clinical trial eligibility criteria into a structured data format. Following an analysis of the criteria and consultations with domain experts, two specific subtasks were defined. Participants were given the flexibility to work on either one of these subtasks or address both. Such an initiative made it possible to gather feedback from end-users (medical and data specialists) on the usefulness of language models. The hackathon revealed important insights for building an

unite! University Network for Innovation, Technology and Engineering

HR EXCELLENCE IN RESEARCH

Wrocław University
of Science and Technology

Faculty of Information
and Communication Technology

Department of Computer Science
and Systems Engineering

27 Wybrzeże Wyspiańskiego St
50-370 Wrocław

T: +48 71 320 30 74

www.pwr.edu.pl
www.kinf.pwr.edu.pl
sekretariat.k44.wit@pwr.edu.pl

REGON: 000001614
NIP: 896-000-58-51
Nr konta:
37 1090 2402 0000 0006 1000 0434

effective tool to parse clinical trial eligibility criteria. It confirmed that pretrained large language models (LLMs) can extract and structure relevant information from eligibility text into boolean expressions, preserving logical and temporal relationships.

(f). Formulation of requirements and implementation best practices: Based on the research and implementation experience, the author formulated practical guidelines (Chapter 9) for implementing similar NLP tools in the pharmaceutical industry environment for clinical trials. These include data standardization, integration with existing systems, data quality assurance, use of MLOps practices to maintain the model, and medical data security and confidentiality issues.

(g). Development of Proof of Concept tool using GPT-4 model: The most important achievement of the dissertation is the design and implementation of a proof of concept (Proof of Concept) tool for automatic analysis of eligibility criteria (Ch. 10). The author built a prototype system based on the GPT-4o model, enhanced with advanced prompt engineering to improve the precision and consistency of the results obtained. The GPT-4o model was selected due to its optimal balance between high accuracy and processing speed, crucial for practical applications such as analyzing medical texts. Additionally, the favorable performance-to-cost ratio made GPT-4o solution economically acceptable for Roche. Collaboration with domain experts played an important role in developing the tool, ensuring that medical knowledge and practical user needs were considered. The tool was then evaluated in a controlled experiment with users. The developed system correctly recognized the key elements of the criteria and was able to make a preliminary interpretation of them, significantly relieving the workload of a human.

(h). Publications: It should be noted that the results obtained in the dissertation have been partially published and presented in journals and conferences of international scope. The author is a co-author of a publication in The AAPS Journal on clinical trial performance prediction (Ch.3) and first author of a paper presented at the AIME 2022 international conference, where she presented results on semantic textual similarity in medicine. She was also the first author to publish a review article entitled "*Machine learning and natural language processing in clinical trial eligibility criteria parsing: a scoping review*" in the journal Drug Discovery Today (Ch.5). This article is an important contribution to Kantor's dissertation, providing a broad review of the literature on the use of ML/NLP models for eligibility criteria parsing in clinical trials (Ch. 5).

It is worth noting that the research was conducted as part of a so-called implementation PhD in collaboration with the pharmaceutical industry such as Roche. As a result, the proposed solutions were focused on practical applications from the beginning, and their usefulness was consulted with potential end-users. Such an approach increases the chances that the dissertation results will be effectively adapted in real-world clinical trial recruitment

unite! University Network for Innovation, Technology and Engineering

HR EXCELLENCE IN RESEARCH

Wrocław University
of Science and Technology

Faculty of Information
and Communication Technology

Department of Computer Science
and Systems Engineering

27 Wybrzeże Wyspiańskiego St
50-370 Wrocław

T: +48 71 320 30 74

www.pwr.edu.pl
www.kinf.pwr.edu.pl
sekretariat.k44.wit@pwr.edu.pl

REGON: 000001614
NIP: 896-000-58-51
Nr konta:
37 1090 2402 0000 0006 1000 0434

processes in the future. To summarize the contribution of the work: ***the dissertation by Klaudia Kantor contributes valuable new results to the knowledge of automating the analysis of eligibility criteria. It combines a broad literature review, innovative application of state-of-the-art NLP models (transformative and generative), and practical verification of these solutions in a real industrial context.***

3. Correctness

The dissertation is written in a coherent, clear manner and demonstrates the author's research expertise. The structure of the work is logical and well thought out. It consists of eleven chapters that lead the reader from the theoretical introduction, through the stages of the author's own research, to a summary with conclusions. Chapters 1-4 outline the background and theoretical underpinnings: concepts related to clinical trials and eligibility criteria are introduced, key information resources and biomedical ontologies used in the paper are presented, and a preliminary study on the impact of clinical trial characteristics on the study is described. Chapter 5 provides a detailed literature review, showing that the author has thoroughly reviewed the current state of knowledge. Several sources from recent years are cited, demonstrating that the information is up-to-date. Further chapters (Ch.6-7) present the experiments and solutions proposed by the author. The description of the methodology is detailed and careful. The author precisely defines the objectives of each experiment and selects appropriate methods for their implementation. Each chapter of the work ends with a summary of the most important findings. The final Chapter 11 presents the summary conclusions of the entire dissertation, discusses the limitations encountered and suggests directions for future research.

The results of the experiments are presented clearly, complemented by adequate tables, summaries and graphs. The interpretation of the results is thorough: the author does not limit herself to stating which model is better, but analyzes in depth why this is so, what the potential advantages of one approach over the other are, and in what situations. Importantly, the author maintains a critical approach to her own achievements. The language of the dissertation is precise and clear, despite the specialized subject matter. Medical and IT terms are explained where needed, making the work accessible also to the reader outside the narrow specialization. The work contains an extensive literature list (including both classic items and recent publications from 2021-2023), which shows that the author is familiar with the her research. The formatting, figures and tables were carefully prepared, facilitating understanding of the content.

There were no significant substantive or methodological shortcomings in the work. One of the limitations of the work (incidentally noted by the author herself in the conclusions) is the evaluation experiment of the implementation of the prototype tool (Ch. 10), referring only to criteria from breast cancer research.. While this limits the generalizability of results to some

unite! University Network for Innovation, Technology and Engineering

HR EXCELLENCE IN RESEARCH

Wrocław University
of Science and Technology

Faculty of Information
and Communication Technology

Department of Computer Science
and Systems Engineering

27 Wybrzeże Wyspiańskiego St
50-370 Wrocław

T: +48 71 320 30 74

www.pwr.edu.pl
www.kinf.pwr.edu.pl
sekretariat.k44.wit@pwr.edu.pl

REGON: 000001614
NIP: 896-000-58-51
Nr konta:
37 1090 2402 0000 0006 1000 0434

extent to other medical domains, the author acknowledges this limitation. A well-designed assessment of tool for automatic analysis of eligibility criteria was conducted, demonstrating the feasibility and potential utility of applying large language models (LLMs) to the clinical trial management context. The prototype was developed in collaboration with domain experts and is being considered for integration into Roche's Clinical Trial Distribution Network, further attesting to its practical relevance. Although the user testing group comprised only four participants (including one physician and one clinical data expert), which is insufficient for statistically generalizable conclusions, the study was clearly presented as a Proof of Concept.. The author holds a critical perspective throughout and appropriately treats this evaluation as a preliminary step toward broader validation.. It is worth noting that the designed tool has a practical dimension, since it was created with the participation of clinical experts and is a step towards the implementation of such solutions in the pharmaceutical industry.

In conclusion, the methodological correctness of the dissertation meets high standards. The author has exercised diligence at every stage of the research: from the literature review, to the planning of the experiments, to the analysis of the results. The result is a complete work, comprehensively treating the research problem posed.

4. Knowledge of the candidate

After reading the dissertation the following conclusion can be drown: Klaudia Kantor has extensive knowledge and skills in the domain of Information and Communication Technologies, especially in artificial intelligence and natural language processing. The author is proficient in machine learning, as evidenced by both the use of classical methods (e.g., classification algorithms for prediction of clinical trial indicators) and advanced deep learning techniques (transformers, BERT-type models, GPT-4). It demonstrates knowledge of the architectures of state-of-the-art language models and the ability to apply them in practice.

The work also shows the author has a good orientation in the area of software engineering and MLOps. When designing and implementing the prototype of the tool, the author had to take into account the issues of integration of the NLP model with existing information systems, processing of input data (criterion texts), and presentation of results to the end user. The principles of data standardization and model business continuity (MLOps) mentioned in Chapter 9 confirm her understanding of the full life cycle of an AI-based IT project, namely from the research phase, to implementation, to system maintenance. In particular, the implementation of the prototype tool using the API of the GPT-4o model required the integration of multiple software components (NLP libraries, data processing and storage modules, user interface, etc.), which the author successfully accomplished.

unite! University Network for Innovation, Technology and Engineering

HR EXCELLENCE IN RESEARCH

Wrocław University
of Science and Technology

Faculty of Information
and Communication Technology

Department of Computer Science
and Systems Engineering

27 Wybrzeże Wyspiańskiego St
50-370 Wrocław

T: +48 71 320 30 74

www.pwr.edu.pl
www.kini.pwr.edu.pl
sekretariat.k44.wit@pwr.edu.pl

REGON: 000001614
NIP: 896-000-58-51
Nr konta:
37 1090 2402 0000 0006 1000 0434

The author's familiarity with IT tools and resources specific to the medical domain should also be highlighted. The work used ontologies and medical knowledge bases, which require the ability to use knowledge representation standards and search for information in large data sets. All of the above aspects indicate that Klaudia Kantor, M.Sc. has a solid theoretical and practical foundation in the area of Information and Communication Technologies. She has the ability to critically analyze the scientific literature, design experiments using complex software tools, and implement software solutions. This combination of knowledge and skills fully meets the requirements for doctoral students in **the Information and Communication Technology** discipline.

5. Other remarks[1]

6. Conclusion

The doctoral dissertation of Klaudia Kantor, M.Sc., represents a valuable and original scientific study that significantly combines the latest developments in computer science with the needs of medicine. The dissertation addresses the important problem of automating the analysis of clinical trial eligibility criteria, the solution of which can improve the processes of developing new therapies. The author has demonstrated a thorough understanding of both the nature of the problem under study and the tools needed to solve it. She brought a number of new scientific findings: from a review and diagnosis of the state of research, to experimental comparisons of NLP methods, to the design and evaluation of a prototype system using generative models (GPT-4) in automating the selection of eligibility criteria. The combination of an IT background with an understanding of the needs of medicine indicates her ability to work scientifically in an interdisciplinary manner.

In summary, the ***thesis "Neural language models for clinical trial eligibility criteria" meets all the criteria for doctoral dissertations. It is original, scientifically justified, carefully written and contains results of significant cognitive and practical value. The achievements presented in the dissertation fully justify awarding Klaudia Kantor M.Sc. the doctoral degree***.

Taking into account what I have presented above and the requirements imposed by Article 187 of the A*ct of 20 July 2018 - The Law on Higher Education and Science* (with

[1] Optional

unite! University Network for Innovation, Technology and Engineering

HR EXCELLENCE IN RESEARCH

Wrocław University
of Science and Technology

Faculty of Information
and Communication Technology

Department of Computer Science
and Systems Engineering

27 Wybrzeże Wyspiańskiego St
50-370 Wrocław

T: +48 71 320 30 74

www.pwr.edu.pl
www.kini.pwr.edu.pl
sekretariat.k44.wit@pwr.edu.pl

REGON: 000001614
NIP: 896-000-58-51

Nr konta:
37 1090 2402 0000 0006 1000 0434

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**, and particularly the area of ....?

[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*

Roman Szczerbicki

Signature

unite! University Network for Innovation, Technology and Engineering

HR EXCELLENCE IN RESEARCH

Wrocław University
of Science and Technology

Faculty of Information
and Communication Technology

Department of Computer Science
and Systems Engineering

27 Wybrzeże Wyspiańskiego St
50-370 Wrocław

T: +48 71 320 30 74

www.pwr.edu.pl
www.kinf.pwr.edu.pl
sekretariat.k44.wit@pwr.edu.pl

REGON: 000001614
NIP: 896-000-58-51
Nr konta:
37 1090 2402 0000 0006 1000 0434

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