

14.10.2023



WARSAW SCHOOL OF ECONOMICS
Department of Information Management
Institute of Computer
Science and Digital Economy

dr hab. Andrzej Sobczak, prof. SGH

Warsaw, 15 October 2023

Review of the doctoral dissertation
MA Stephan Wronkowski-Elster

entitled. **"The Design and Development of a Reference Architecture for Trustworthy AI with a Focus on Corporate Planning and Decision-Making in the Process Industry"**.

written under the scientific supervision of
dr hab. Sylwester Bejger prof. UMK
dr Krzysztof Rykaczewski asst. prof. UMK

1. The basis for a review

The review of the dissertation of Stephan Wronkowski-Elster, MA, entitled "The Design and Development of a Reference Architecture for Trustworthy AI with a Focus on Corporate Planning and Decision-Making in the Process Industry" is based on the decision to appoint me as a reviewer by the Council of the Scientific Discipline Economics and Finance of the Nicolaus Copernicus University in Toruń at its meeting on 14/06/2023

2. Choice of dissertation topic and presentation

The last few years, and especially the last few months, have seen an exponential increase in interest in Artificial Intelligence (AI). Colloquially, it can be said that solutions incorporating AI which emerged from universities and research units become part of the operations of many companies. One of the obstacles to massive use of AI

in many industries is the question of how to explain the results obtained from machine learning models. The implementation of the concept of Explainable AI (XAI), which makes it possible to transform opaque and difficult-to-explain machine learning models into transparent decision-making mechanisms that enable the control and interpretability of the results they provide, becomes crucial. By implementing XAI, organizations can help build trust and acceptance of AI among key stakeholders - both internal and external (e.g. customers or suppliers). Explainability becomes the bridge that enables engineers and scientists to explain the principles of complex analytical models to users.

At the same time, it is critical to the scientific rigour of AI work to ensure that machine learning models are continually evaluated and that XAI provides the tools to critically analyse their performance.

XAI implementations promote the construction of more ethical algorithms by enabling the identification and elimination of biases in machine learning. The push for more explanatory AI is also a response to the increasing complexity of models and to the need to regulate them in a rapidly changing legal and ethical environment. In fact, it can be said that the explainability of AI is no longer an option, but an imperative, requiring companies to implement solutions ensuring that the results provided by AI are understandable and credible to end users.

All this, together with the scope and objectives of the work reviewed, which focuses on the creation of a reference architecture for an Explainable Artificial Intelligence system that can combine both sub-symbolic and symbolic approaches, allows the conclusion that the choice of topic is very timely and highly relevant, both from a practical and research point of view.

It should be emphasised that the topic addressed in the reviewed thesis is still relatively unexplored, not only in Poland, but also worldwide (although the last 12-18 months have brought an exponential increase in interest in these issues). This observation results both from the research carried out by the author and presented in the paper but is also confirmed by the analyses carried out independently by the reviewer. It can therefore be concluded that the paper is of innovative nature.

The author focused on the process industry, with particular emphasis on the chemical and pharmaceutical industries. The role of these industries is increasing every year – the best example is the situation during the Covid-19 pandemic, or the current efforts to combat chronic diseases. The reviewer understands the reasons for narrowing the research area to these industries resulting from the implementation conditions. At the same time, the reviewer would like to raise the question of whether, in the author's opinion, the research methods developed are of a more universal nature and can be applied to other industries as well. The author mentions this in the thesis, but the reviewer will ask for a reference to this issue during the defence of the thesis. In particular, what measures should be taken from a methodological point of view in order to be able to apply the results of the work to wider areas of industry from the results of the doctoral candidate's work.

In conclusion it should be said that the doctoral thesis reviewed addresses an important, interesting and very current scientific and research problem. The author has made an appropriate choice of the topic of his dissertation. Both the cognitive and the utilitarian context of the considerations presented in the dissertation deserve to be highlighted.

3. Formal evaluation of the work

The volume of the thesis is 332 pages. The thesis consists of an introduction (defined as Chapter 1), four main chapters, a summary and a conclusion. In addition, the dissertation contains an appendix on the research carried out (questionnaire survey), a presentation on the evaluation of architecture and lists of figures, tables and abbreviations. Due to the relatively large size of the dissertation the prepared indexes are useful in guiding the reader through the content presented. A glossary of terms is the plus of the work. The terminology related to AI is still an area of discussion and the glossary definitely facilitates the reception of the presented content.

The bibliography is relatively extensive. It consists mainly of English and German language publications. The following sources of information were used in the dissertation: books and monograph, scientific journals, research and expert reports of Polish and foreign research institutions. In addition, the doctoral student used a number

of Internet sources. This is not a reproach, as these allow reference to the rapidly changing business reality and technological conditions.

The vast majority of the sources used by the PhD student has been published in recent years, although there are references to articles published 20 years ago and earlier (e.g. *Tushman, M. L., & Nadler, D. A. (1978). Information processing as an integrating concept in organisational design. Academy of Management Review, 3(3), 613-624*). However, they are in the definitive minority and are only supplementary – they show the broader context of the thesis considerations.

The formal aspect of the thesis is unobjectionable. The reasoning is very logical, comprehensible and linguistically correct (there are only minimal punctuation and stylistic errors). The author uses simple language avoiding neologisms and borrowings from other languages. The content of the thesis is presented in manner leaving no doubt as to its interpretation. The clarity of the language, the unambiguousness of the formulations and the conclusions drawn are a great advantage of the thesis.

Some comments can be made on some of the figures and tables. In both printed and electronic form, they are difficult to read (e.g. p. 226 - Figure 68, p. 228 - Table 16).

On the whole, however, the formal side of the study can be considered to be correct and has been prepared with considerable care.

4. Title, scope, thesis and research objectives of the dissertation

The title of the dissertation, 'The Design and Development of a Reference Architecture for Trustworthy AI with a Focus on Corporate Planning and Decision-Making in the Process Industry Focus on Corporate Planning and Decision-Making in the Process Industry', clearly defines the author's intention for the dissertation.

As the scope of the research, the author has chosen manufacturing companies from the chemical and pharmaceutical industry. The subject of the author's research is the reference architecture of the explainable artificial intelligence system, combining two different approaches (subsymbolic and symbolic), which increases the difficulty of the subject matter covered, but at the same time opens up new areas of research.

The content of the work presented is fully in line with its title. The scope of the work, as already mentioned, should be considered important from both a theoretical and practical point of view.

Based on the literature review, the candidate formulated the following research question: How can an explainable artificial intelligence or agent system be developed and integrated into a process industry planning framework in order to increase confidence in decision-making AI systems by improving their transparency and decision quality?

Based on this, he formulated the following thesis: By developing a reference architecture for an explainable artificial intelligence system that can combine both sub-symbolic and symbolic approaches, confidence in artificial intelligence models, and thus quality of the decision-making in business planning, can be increased. (p. ix).

The main objective of the dissertation is to develop a reference architecture that promotes explainable artificial intelligence to improve decision-making and increase the quality of business planning in the process industries (p. x). I believe that such an objective of the thesis is sufficiently relevant for the purpose of formulating the thesis and obtaining the doctorate. It should be emphasised that the structure of the examined thesis and the flow of logical argumentation are purposeful for the achievement the achievement of the main objective of the thesis.

The doctoral candidate has formulated four specific objectives of the thesis, each decomposed to sub-objectives (p. 9). The specific aims are stated in a logical and concise manner:

RQ1: What are the specifics of the process industry?

RQ2: What is Explainable AI and how can it support decision making in the corporate planning process?

RQ 3: How is a Reference Architecture for an explainable AI system being designed and developed?

RQ 4: How to provide guidance on creating a reference architecture for explainable artificial intelligence in the operational planning context?

The decomposition of the main objective of the dissertation into specific objectives is done correctly. The implementation of the specific objectives ensures the achievement of the main objective of the dissertation. The very well planned research process is worth highlighting. It has resulted in an interesting thesis.

The very strong practical orientation of the dissertation should also be emphasised. Therefore, it is interesting and useful not only for the scientific community, but also for specialists in companies responsible for the implementation of systems containing artificial intelligence mechanisms.

5. Research tools used

In the thesis, the candidate used a number of interrelated research methods (included in the research plan - page 58). The leading method was a design science based research approach. The author applied guidelines developed by Hevner and his co-workers in order to help understand, conduct and evaluate design science research. Using these, candidate developed a system reference architecture to serve as a model for trustworthy artificial intelligence applied to business planning in the process industry. The evaluation of the results of dissertation took place in two stages. The first step was to evaluate the design process based on design science principles. The second stage involved a survey of experts in the fields of architecture and business planning to further evaluate the project. The survey was developed based on criteria derived from best research and analysis practices, following the methodology of Saunders et al. (2023) and Sekaran and Bougie (2019).

The results of the survey were statistically analysed. In their responses, the experts identified additional concerns, requirements and constraints that were identified as gaps in the PhD project. These gaps will be taken into account in the next iteration of the design cycle. The survey confirmed the hypothesis that the reference architecture developed can practically answer the research questions posed by the author in Chapter I.

In conclusion, thanks to the appropriately selected research tools, the candidate has comprehensively addressed the adopted objectives of the thesis - main and specific - and the formulated research hypothesis.

6. Work organization

The main body of Stephan Wronkowski-Elster's thesis consists of an introduction (referred to as Chapter 1 in the thesis), four substantive chapters and a conclusion. The structure of the thesis is typical and consisting of theoretical part and a research and design part.

Chapter 1 is divided into four sections. Section 1.1 discusses the macroeconomic perspective of Artificial Intelligence (AI), given its importance and relevance to the topic of the thesis. Section 1.2 presents the microeconomic perspective of AI implementation. Section 1.3 describes the motivation for the research and the relevance of the topic, highlighting the need for human trust in the decision-making process of AI systems. Section 1.4 defines the purpose of the thesis and the research questions, and section 1.5 presents the research approach. Section 1.6 describes the structure of the thesis.

In the second chapter, the author focused on the topic of planning in the process industry. After an introduction (Section 2.1), the importance of the process industry is presented (Section 2.2), including the key trends and challenges currently facing the industry (Sections 2.2.1 to 2.2.3). Section 2.2.4 describes AI use cases in the process industry. Planning and decision-making processes are presented in Section 2.3. Sections 2.3.1 to 2.3.4 discuss planning scenarios, integrated business planning, decision types and stakeholders of business planning in the process industry. Section 2.4 describes the support of modern information systems for business planning. Section 2.5 distinguishes between classical decision support systems and AI systems, and discusses business analytics and predictive and prescriptive analytics.

In Chapter 3 the author provides an overview of issues related to Explainable Artificial Intelligence (XAI), starting with a general introduction in Section 3.1, a description of AI in Section 3.2, and an in-depth analysis of issues related to machine learning. Section 3.2.2 and 3.2.3 discuss knowledge-based systems and neuro-symbolic AI methods. Section 3.3 defines the concept of explainable AI. It discusses the importance of explainability throughout the life cycle of an AI system, not just at the development or production stage. Section 3.4 highlights the importance of ethical, legal and regulatory requirements for AI and their impact on explainable AI. It also provides a brief overview of the evolving field of AI ethics and legal and regulatory requirements. Section 3.5 presents a map of AI stakeholders and their requirements for explainable AI.

Chapter 4. focuses on the development of reference architectures for explainable AI systems. It begins with an introduction in Section 4.1, followed by a theoretical foundation for reference architecture development based on knowledge base methods in Section 4.2. Section 4.3 discusses the methodology for creating reference architectures using the TOGAF and Attributive Architecture Design. The following subsections cover in detail the stages of designing the architecture vision, identifying stakeholders and business requirements, defining the scope and objectives, through to the individual steps of creating a reference architecture (according to the TOGAF ADM).

Chapter 5. presents the process of developing a reference architecture for a trusted explanatory AI system called Re_fish (a combination of the words 'Rejewski' and 'Babelfish'). The following sections describe in detail the development of the Re_fish architecture using the ADD and TOGAF ADM methodologies. Sections 5.2.1 to 5.2.7 present the business, application and technology architectures. Section 5.2.8 deals with AI application lifecycle management and section 5.2.9 describes the capabilities and solutions. Section 5.3 evaluates the reference architecture and Section 5.4 summarises the feedback and documents any gaps.

The whole dissertation ends with a Conclusion, which presents a summary of the results obtained, discusses the main conclusions (including a reference to the objectives and research hypothesis adopted at the beginning of the work) and outlines areas for further research. The author also points out the limitations of the research method used.

The titles of the chapters are correct and correspond to the content presented. The structure of the thesis is correct and clear. The logic of the argumentation is maintained in its different parts. This is a strong point of the assessed thesis.

7. Substantive evaluation of the dissertation

Before mentioning the main achievements of the author, I would like to emphasise the innovation and topicality of the subject matter. The issues addressed in the thesis are interdisciplinary by nature, which increases difficulty of the research process.

Among the most important achievements of the author of the doctoral thesis I would mention

- Conducting an in-depth literature analysis of the role of artificial intelligence in the economy (micro and macro), with a particular focus on the explainable artificial intelligence, including regulatory issues (e.g. EU GDPR, PE-6-2023-INIT).
- Conduct an in-depth analysis of the impact of artificial intelligence on the process sector, with a particular focus on the pharmaceutical industry.
- Develop a research process leading to the development of a reference architecture for an explicable artificial intelligence system that can combine both subsymbolic and symbolic approaches.
- Creation of a proprietary reference architecture called 'Re_fish' that can be used in the context of business planning in the process industry, using a hybrid knowledge-based approach.
- Use of recognised market standards to develop the reference architecture - such as the TOGAF architectural framework and the ArchiMate modelling language.
- A strong practical focus. As a result, it contains content that is useful to a wide audience – both companies implementing AI solutions and vendors of these solutions.

Weaker elements of the thesis would include:

- No real verification of the developed reference architecture – even in a simplified form i.e. a prototype.
- Only very briefly addressed directions for further research.

Finally, I would like to formulate a few questions for candidate:

- How does he see the possibility of using the developed reference architecture in other industries and sectors?
- Is it possible to use the developed reference architecture in the area of large language models (LLM), and to what extent?
- How should it look like to try to verify the use of the developed reference architecture in practice?

8.

9. Conclusions (Eng.)

Despite the minor remarks made in my review the dissertation under review is an example of an extremely interesting study which (it is hoped) will also inspire others to take up the subject. It demonstrates the freedom of movement of the candidate in the world of very complex issues. The very deep connection of the content presented to business practice should also be highlighted.

In fact all the research questions posed in the study have been answered and the research hypothesis can be considered verified.

Taking all this into account, I conclude that the doctoral thesis of Stephan Wronkowski-Elster – – entitled "The design and development of a reference architecture for trustworthy AI with a focus on corporate planning and decision-making in the process industry" and written under the scientific supervision of Dr Sylwester Bejger, Prof. UMK and Dr Krzysztof Rykaczewski constitutes original scientific research and meets all the requirements of the Act on Academic Degrees and Titles. I request that it be approved for public defence.

Finally, I recommend and strongly encourage the candidate to consider making an extra effort to disseminate the results obtained, e.g. in the form of a monograph or at least a white paper.

10. Konkluzje (PL.)

Pomimo zgłoszonych – w trakcie całej recenzji – drobnych uwaga polemicznych, recenzowana rozprawa stanowi przykład niezwykle interesującego badania, które – można mieć nadzieję – zainspiruje także innych do podjęcia tej problematyki. Jest ona dowodem swobody poruszania się Doktoranta w świecie bardzo złożonych zagadnień. Na podkreślenie zasługuje także bardzo głęboki związek prezentowanych treści z praktyką gospodarczą.

Praktycznie na wszystkie pytania badawcze postawione w pracy udało się uzyskać odpowiedź, a postawioną hipotezę badawczą można uznać za zweryfikowaną.

Biorąc to wszystko pod uwagę stwierdzam, że rozprawa doktorska Stephana Wronkowskiego - Elstera napisana pod kierunkiem naukowym dr hab. Sylwestra

Bejgera, prof. UMK oraz dr. Krzysztofa Rykaczewskiego pt.: „The design and development of a reference architecture for trustworthy AI with a Focus on Corporate Planning and Decision-Making in the Process Industry” jest oryginalnym badaniem naukowym i spełnia wszystkie wymogi ustawy o stopniach i tytułach naukowych. Wnoszę o dopuszczenie jej do publicznej obrony.

Na koniec rekomenduję i gorąco zachęcam, aby rozważyć podjęcie dodatkowego wysiłku i rozpropagować uzyskane efekty – np. w formie monografii lub przynajmniej white paper.

A handwritten signature in black ink, appearing to be 'A. L. M.', located in the lower right quadrant of the page.