Uniwersytet Ekonomiczny w Katowicach | Katedra Badań Operacyjnych

DR HAB. JERZY MICHNIK, PROF. UE

REVIEW REPORT
on the PhD thesis
"THE DESIGN AND DEVELOPMENT OF A REFERENCE ARCHITECTURE FOR TRUSTWORTHY AI"
by mgr Stephan Wronkowski-Elster

1 Formal Basis

The review has been requested by dr hab. Jerzy Boehlke prof. UMK, Dean of the Faculty of Economic Sciences and Management, Nicolaus Copernicus University in Toruń (contract work 37/2023 (ZPU.2023-01557), dated on June 30th, 2023. The above notification was based on the resolution of Rada Dyscypliny Naukowej Ekonomia i Finanse nr 11.WNEiZ/5301/32/2023, dated on June 14th, 2023. The dissertation was written by mgr Stephan Wronkowski-Elster under the supervision of dr hab. Sylwester Bejger, professor at Nicolaus Copernicus University in Toruń and dr hab. Krzysztof Rykaczewski, professor at Nicolaus Copernicus University in Toruń.

2 General Characteristics of the Dissertation Content

The PhD dissertation is 332 pages long and contains 6 chapters. The main part is preceded by abstracts in English and Polish, list of figures, list of tables, list of abbreviations and followed by references, glossary and 2 appendices. The list of references is comprised of about 460 positions.

The Introduction outlines the broad background of the undertaken research. It begins with the macroeconomic perspective of artificial intelligence (AI) in which impact of AI on economy, potential growth opportunities through AI and impact on labour have been discussed. In the microeconomic perspective the concept of explainable AI (XAI) has been developed and pointed at the supporting role that can be played by XAI. The benefits from

AI can be unevenly distributed among the potential market players, also in connection with the competition for the greatest talents and the best skills.

Chapter 1.3 presents "Motivation and Relevance" of the dissertation. A difference between symbolic AI and sub-symbolic AI methods, as well as between "black box" and "glass-box" (or "transparent") models has been highlighted. The AI is widely used for support in decision making. Spectacular success of sub-symbolic AI (black box) raised a question of its real usability in practice. In the context of decision making the lack of number of features like transparency, understandability, interpretability, etc. has triggered development of XAI. Finally, the key concept of reference architecture is defined as "... as an abstraction on a meta-level, whose idea and goal are to help the design, development, and implementation of systems.".

The main goal of the research "is to establish and create a reference system architecture that promotes explainable artificial intelligence, with the aim of improving decision-making capabilities to facilitate better business planning within the process industry." It is related to the basic research question that is "How can an explainable artificial intelligence system, or agent, be created and integrated into the planning framework of the process industry to increase trust in decision-making AI systems by improving their transparency and decision quality?"

The basic research question is accompanied by the expanded list of auxiliary questions about: specificity of the process industry (6 questions), the XAI and its role in supporting decision making in the corporate planning process (5 questions), construction of XAI (2 questions), guidance on creating reference architecture in the chosen context (9 questions).

Chapter 1.5 introduces the basic foundations of design science and presents a general and detailed plan of research reported in the thesis. The structure and outline together with the brief guide to dissertation content are included in Chapter 1.6.

The dissertation topic is located at the junction of three research planes: corporate management (here planning in the process industry), Explainable Artificial Intelligence (here in corporate planning) and design of a reference architecture (design science). The three chapters that make the second main part of the dissertation, describe them in details based on a wide review of scientific literature. They are: Chapter 2 Planning in the Process Industry, Chapter 3 Explainable Artificial Intelligence in Corporate Planning and Chapter 4 Design of a Reference Architecture for Explainable AI.

The Doctoral Candidate's original contribution to science is the subject of the description in Chapter 5 Development of a Reference Architecture for Explainable AI in Corporate Planning. The process is based on methods described in literature: the Attribute Driven Design (ADD) and the Architecture Development Method (ADM) accompanied by other approaches intended for the creation of a software architecture. The largest part of the chapter is devoted to the detailed description of the development process of reference architecture named Re_fish. Then, In Chapter 5.3 the Author discusses the evaluation of Re_fish architecture. The evaluation has been carried out in two ways. Firstly, using the seven guidelines from the works of Hevner. Secondly, by conducted presentation, discussion and survey in which 12 experts took part. The results of the survey were subjected

to statistical analysis. Finally, the adjustments to Re_fish reference architecture resulting from the comments and recommendations of the participants in the discussions and the survey are presented.

Chapter 6 concludes the work, confirming the achievement of the research objectives of the dissertation and briefly indicating directions for future research.

3 General Remarks

3.1 Important Content-related Remarks

1. p. 48. The form of the hypothesis is questionable. It is difficult to confirm it in the scope of the dissertation. From the reference architecture is a long way to design of concrete system and its practical instantiations and tests. On the other hand, the research goal supported by extensive literature and embedded in design sciences has been achieved. Especially, it has been confirmed by the reference architecture validation against Hevner's research guidelines and results of the survey.

2. p. 49-50. There is overabundance of research questions. Most of them are of the nature of issues that can be clarified by examining the source literature. Only some are actual questions that require original research.

3. p. 226, Fig. 68; p. 248, Fig. 77; p. 260, Fig. 86. The drawings are almost unreadable. As they are essential for the understanding of the Re_fish, this is a major obstacle to the correct evaluation of the results. The verbal description that accompanies the drawings only partly eliminates these difficulties.

4. p. 250, Fig. 78; p. 158, Fig. 85. These drawings are readable only in high magnification in pdf file.

5. p. 257. "One of the most important questions which has remained unanswered can the symbolic and non-symbolic modules of Re_fish be linked?". **Why it is a difficult problem? Formal, technical or operational issues? What are the most important criteria for establishing this connection?**

6. **To what extent is the proposed architecture universal, and to what extent does it refer to the case of corporate planning in the process industry considered in the dissertation?**[1]

7. Graphical representation of the proposed reference architecture, using dedicated modelling languages, is contained in illegible drawings. This is a significant drawback, due to the important role of this representation in explaining the structure of the project.

[1] Questions in bold may, among other things, be discussed during the defence.

8. Chapter 5.3. Incorrectly performed statistical analyses do not add any value to the results of the work. (I.a.: on p. 275-76, statistical hypotheses are wrongly formulated; Table 33, 34, high values of *p-value* do not allow to reject null hypothesis.) Due to the small size of the sample, an adequate method of interpreting the results of the survey is a qualitative analysis supported by a limited use of descriptive statistics.

3.2 Minor Remarks and Points of Debate

The remarks listed below are ordered according to appearance in the dissertation, not according to importance.

1. p. 30. "The emergence of AI as a new factor of production has led to a hybrid combination of labour and capital." This statement is unclear and not supported by reference to the literature.

2. p. 42, footnote: "bounded rationality or bounded rationality is a reality" . ??

3. p. 101. "Information may be defined as judicial knowledge relevant to a decision (Klein & Scholl, 2011).". Is the term "judicial knowledge" appropriate in the context?

4. p. 109. "innovations were thus managed, using the Stage Gate model.". Why this model? It is not the best approach, nor the only one possible.

5. p. 121. Variables in (f5)-(f6) are not defined.

6. p. 126, "the planning Algorithm A". This symbol seams to be wrong. it is not the same as A that appears in the next sentence (the set of current states). Something wrong in formula (f8). On the left of an arrow there is Cartesian product with missing the second set.

7. p. 127, formula (f13). What is the meaning of the symbol "the question mark over the arrow"?

8. p. 150, Fig. 53. What is a source of the graph?

9. p. 157. "Non-monotonous inference in classic logic it that the set of possible equivalent cases is growing monotone". There is contradiction in this sentence.

3.3 Editorial Notes

There are quite a large number of editorial and typographical errors in the dissertation. Some of them are listed below.

1. Some publications are referred by first name instead of family name. It makes difficult to identify citations. Ex.: Jonathan Gillham. (2017) referred under 'J'; E.J. Russo, & P.J.H. Schoemaker (1992) referred under 'E'.

2. p. viii, part of the multiple citation (Bejger & Elster, 2019; Bejger & Elster, 2021) appears twice.

3. p. 56, " the goal of behavioural science goal is to research"

4. p. 63 "In Chapter 4.5 consists of a short discussion on the requirements. Chapter 4.6 concludes the chapter by summarising the findings.". There are no such chapters.

5. p. 69, last item in the list appears twice.

6. p. 71, "G = unit cost of utility j (e.g., €/MWh)". In formula (f3) it is denoted as "g_j".

7. p. 77, "in den USA die FDA". ??

8. p. 82, Fig. 16/ Legend is missing.

9. p. 85, "In figure 21, it can be seen ...". Perhaps it should be Fig. 22.

10. p. 87, "Figures 22 to 25 show ...". Perhaps it should be Fig. 23 to 25.

11. p. 87, "energy consumption declined from 1990 to 2020 – dropping from 752 tGW to 589 tGW.". What is this unit? Energy is measured in i.a. GWh.

12. p. 87, 88, 89, Fig. 26, 27, 29. The units are not specified.

13. p. 90, "Finding 10: Challenges in the process industry ergeben sich, wie bereits oben beschrieben, aus der hoc", ??

14. p. 116, "Wild categorises a planning or decision-making problem in accordance with the following criteria (Klein & Scholl, 2011; Wild, 1982):". The whole paragraph with the list appears twice.

15. p. 220, "The scenario planning system of the DFKI (AISOP).". This sentences is not finished.

16. p. 238, "... allows for a finite set of actions, "cost" is a non-negative cost function, I is the initial state, and G is the goal.". This is a repetition of the sentence from 3rd line above.

4 Summary and Final Conclusion

As the advantages of the presented dissertation can be indicated:

- extensively researched and presented subject, which serves as the basis for the scientific result of the dissertation,
- in-depth analysis of the main issue – designing a reference architecture for XAI,

- detailed development of a multi-module structure for the XAI reference architecture,
- taking into account many aspects of the potential use of the proposed architecture, including stakeholders, legal regulations, and technical and operational conditions.

A certain number of quite serious defects weaken the positive evaluation of the work. Among them are:

- the hypothesis is not properly formulated,
- too many research questions; most of them are merely auxiliary questions that facilitate the research process,
- Graphical representation of the reference architecture, using dedicated modelling languages, is a vital part of the description of the proposed solution. Its illegibility significantly undermines the communication of the central concept in the dissertation.
- the statistical analysis is not only unsupported by the sample size but also erroneous.

Considering all the above pros and cons, it can be considered that the positive aspects of the presented work outweigh its imperfections and that the dissertation presented by Doctoral Candidate Stephan Wronkowski-Elster meets the requirements of Polish law (Art. 187 Ustawy o szkolnictwie wyższym i nauce z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce /Dz. U. z 2022, poz. 574 z późn. zm./) and may be admitted to further stages of the PhD procedure.

Katowice, September, 4, 2023

dr hab. Jerzy Michnik, prof. UE