SZKOŁA GŁÓWNA HANDLOWA W WARSZAWIE
WARSAW SCHOOL OF ECONOMICS

Zakład Wspomagania i Analizy Decyzji

September 4th, 2023

prof. dr hab. Bogumił Kamiński

SGH Warsaw School of Economics

Review of doctoral dissertation of Stephan Wronkowski-Elster MSc

“The design and development of a reference architecture for trustworthy AI

with a focus on corporate planning and decision-making in the process industry”

This document was prepared in response to information number 11.WNEiZ/5301/32/2023 issued on July 4, 2023 referring to resolution of Scientific Council of Economics and Finance of Nicolaus Copernicus University in Toruń asking for a review of doctoral 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” prepared by Stephan Wronkowski-Elster MSc.

The review is prepared in reference to the requirements specified in act “Prawo o szkolnictwie wyższym i nauce” Dz. U. z 2018 r. poz. 1668 with later changes (Act).

The review is prepared in English with the conclusion additionally translated to Polish following the request of Nicolaus Copernicus University in Toruń provided by e-mail on July 17, 2023.

In the first part of the review I evaluate meeting of the formal requirements by the dissertation (art. 187 pkt. 3 and 4 of the Act). Next I present an assessment if, following art. 187 pkt. 1 and 2 of the Act, the dissertation meets the following requirements:

1. Doctorial dissertation presents general theoretical knowledge of the candidate in the area of economics and finance field of studies and the ability to perform the scientific research in an independent way.

2. The dissertation presents an original solution to the research problem.

In accordance with these requirements I perform the evaluation in the following aspects:

1. Whether the solved research problem can be classified within the economics and finance field of studies.
2. Whether the dissertation shows that the candidate has theoretical knowledge in the domain of economics and finance.
3. If the dissertation presents an original solution of a research problem.
4. If the dissertation the ability of the candidate to perform scientific research in an independent way in domains: scientific methodology and written communication.

The review is finished by the conclusion.

Evaluation of formal requirements

The dissertation has a form of a 332 page manuscript and is prepared in the English language. It is accompanied by an abstract in the Polish and the English languages. Therefore it meets the requirements specified in art. 187 pkt. 3 and 4 of the Act.

Assessment of the research problem from the perspective of the economics and finance field of studies

The subject of the dissertation is design and development of a reference architecture for trustworthy AI with a focus on corporate planning and decision-making in the process industry.

First I want to remark that the presented thesis does not present results in the domain of finance. Therefore the assessment is done with respect to the economics domain.

In what follows I use the communicate 7/2010 of Central Commission for Degrees and Titles as a reference scope of economics domain which is:

> *Economics studies the behavior of people and the interactions between them in conditions of limited resources. In particular, the scope of economics research includes:*
>
> - *public sector analysis;*
> - *analysis of growth, development and cyclical fluctuations of the economy and its individual sectors;*
> - *the functioning of markets;*
> - *spatial economy;*
> - *international economics;*

- *theoretical foundations of socio-economic and sectoral policy;*
- *history of economic thought;*
- *economic history.*

> *Economics formulates hypotheses and builds models regarding the relationship between real and monetary variables occurring in the economic process and subjects them to empirical tests. Research in the field of economics is based on fundamental economic categories, such as: economic equilibrium, economic efficiency, rationality of behavior, optimality of decisions in the general economic aspect.*

From this perspective, the dissertation presents a review of the microeconomic and macroeconomic context of use of artificial intelligence solutions in general. However, there is a borderline evidence presented in the text how the product of the thesis, which is Re_fish Reference Architecture, contributes to the development of the economics domain as it is defined in the communicate of the Central Commission for Degrees and Titles.

To justify this judgement let me refer to the hypothesis of the research that is provided by the candidate (this is a single hypothesis that is specified in the dissertation):

> *By developing a reference architecture for an explainable AI system that could combine both subsymbolic and symbolic approaches, confidence in AI models and, thus, decision-making in corporate planning can be improved.*

First, let me highlight that the hypothesis, as stated, is, in my opinion, challenging to be verified. A mere development of a reference architecture is not likely to lead to an improvement in confidence in AI models (and this is the statement of the hypothesis). It can be reasonably assumed that such reference should be first implemented by some organizations, which then could observe this increased confidence. For the remainder of the assessment I assume that this is what the candidate meant (although the hypothesis itself is not stated in this way).

Assuming that the thesis formulates the assumption of implementation of the reference architecture it can be accepted, given the existing results from the economic literature, presented in sections 1.1 and 1.2 of the thesis, that the results present the study of behavior of people in the conditions of limited resources, especially in the domain of functioning of markets.

In conclusion, although the tools and methods used in the thesis are mostly not from the economics discipline, the thesis *can be classified within economics and finance discipline*.

Assessment if the thesis shows that the candidate has theoretical knowledge in the domain of economics and finance

To perform this assessment I analyze the knowledge of the macroeconomic and microeconomic literature presented in the thesis and knowledge of quantitative methods used in economic research.

In the macroeconomic aspect the candidate significantly refers to growth theory. However, there is no direct in-depth reference to a vast literature of endogenous and exogenous growth theory with a relevant critical discussion that would normally be expected.

Let me give one concrete example.

As a next example consider the first presented finding presented in the thesis:

> *Finding 1: Impact of AI on economy*
>
> *The impact of AI on economy can be described as the importance of former production factors, labor and capital, become less important, or grow together into a single factor.*

In my opinion, firstly, the presented sentence is highly imprecise, to the level that as a reviewer I have a hard time understanding it. In particular note that the words "importance" and "important" are used in it, and it is not clear what "grow together into a single factor" means. The meaning can be traced back to a literature review presented earlier in the text stating "*AI is a hybrid factor of production that combines capital and labour*". However, this statement is not equivalent to the one presented in the finding.

Here, let me additionally comment on this fragment of the text of the thesis (although this assessment technically falls into "ability of the candidate to perform scientific research in an independent way", where I also discuss it, I put it here as it also directly relates to the scope of review of macroeconomic literature of the candidate). The literature review fragment gives the following citations:

- Gordon (2016), cited several times: such a reference is not present in the bibliography; there is a "Gordon, R. (2017). The rise and fall of American growth: The US standard of living since the civil war. (2017)"entry with a different publishing year; it can be assumed that the candidate meant this book though; still the bibliographic entry does not provide the publisher name which is a standard expectation; also, a standard expectation would be to provide in the text page number in the book the candidate refers to (the book itself has 784 pages and covers a wide array of topics);
- Solow R. (1987), this is indeed a reference to a well-established researcher in the field of economic growth, however, the cited text, is just a 2 page book review; no other works of Solow are cited, in particular results that lead to the award of the Nobel Prize;

- Menzel & Winkler (2018): again there is no such entry in the bibliography; however, there are 2 entries of Menzel and Winkler (2019) – one under letter “C” in the bibliography, other under letter “M”. The hyperlink provided in the entry does not lead to a publication but to the webpage of Christoph Menzel on ResearchGate; it is possible to find the referenced literature position there. Also this is an unreviewed technical report of Bundesministeriums für Wirtschaft und Energie.
- Purdy and Daugherty (2016): this position is missing in the bibliography; the manual search that it is an unreviewed technical report of Accenture company.

I provide this detailed assessment to highlight two facts:

1. The candidate mostly does not refer to the mainstream economic literature in the thesis. It does not mean that non-standard texts (like unreviewed technical reports) should be discarded – indeed they might and do provide valuable insights. However, they should be referred to in the light of the widely accepted economic theory body of knowledge (to prove that the candidate possesses it, which is a requirement that should be verified by this assessment).
2. The quality of the bibliographic references is low. The reader needs to guess what the author refers to and manually search the literature to find a matching source text.

One additional remark regarding this finding (also applicable to other presented findings) is that it is general it does not refer precisely to the scope of the thesis (application of AI in the corporate planning in process industry). It is not clear that this finding applies to this domain of application which is studied in the thesis (even if we agree that it is true in general on the macroeconomic level). Typically, it should be expected that the thesis shows a precise connection of the presented results and its objective.

Additionally, although the implementation of AI tools has become significantly more widespread after the year 2020, the candidate consistently refers to results from before this period, and claims that they are “current”. Let me give one example. In figure 1 the candidate reports the results on total factor productivity for the period up to year 2015 and calls them “*the total factor productivity (TFP) growth rate, which has steadily declined for the countries analysed (USA, UK, EU, and Japan) and is currently at a low level.*”, while for example for USA the data on TFP is available up to year 2022. It is of course valuable to look at historic evidence, but, when the research is devoted to a rapidly developing domain such as AI, it would be expected that the thesis should present the most recent evidence available. It should also be noted that indeed there are recent references in the text, but they are not present consistently in all important areas of the thesis.

Let us now turn to assessment of microeconomic knowledge presented.

Again, I will highlight my point using an example. In formula (f3) on page 71 the candidate presents the way how mean cost per ton of product can be calculated. In this formula there is a $(Mm)/(QU)$ factor presented. First of all, it is clearly incorrect. As M is defined in units per ton while it is later divided by total annual production QU. Additionally, the whole formulation of this factor (by comparison to formulation to other factors) suggests that the candidate considers labor costs as fixed. So even if we assumed that the original formulation $(Mm)/(QU)$ was just an error in typing, it is not clear what the candidate assumes. For a reference, it is standard to assume that labor costs are partially fixed and partially variable (and the ratio between these two components can vary between industries and countries). Also, just below the formula the candidate states that SG&A (selling, general, and administrative) expenses are variable costs, while – typically it is assumed that there is a significant fixed cost components in these costs. What I wanted to highlight here (apart from a clear error) is low level of precision of the candidate in the thesis when referring to economic terms.

Another aspect of general knowledge in domain of economics and finance is application of quantitative methods in the presented research. Let me present an assessment of this knowledge based on the analysis of the survey that the candidate conducted and presented in section 5.3.

Here is a list of comments that can be given to the knowledge side of the presented survey:

1. The candidate does not describe how the process of selection of respondents was performed. This is especially relevant as only 12 responses are analyzed. The candidate states that the surveyed respondents were "experts in the fields of architecture and corporate planning", but the presented data does not even justify this claim. Some of the experts had no experience in planning, some had no experience in architecture; additionally 5 of the experts are from a single company (SAP) that is not a company from process industry (although indeed it is an IT provider for such companies); also other industries of the surveyed respondents do not seem to indicate process industry, like: Inseye, AI/Data Elitmind, self development, Innovation Area, real estate, software consulting. Then the question is did these experts have an adequate experience of working in the corporate planning domain in the process industry to be able to provide educated answers to the questions.
2. The questionnaire included 18 questions measured on a Likert scale from 1 to 5 (see page 271 of the thesis). However, in the analyzed data there are values starting from 0, which is not explained nor justified in the thesis. Given the small sample size such change of encoding might have had a significant impact on the results. Additionally, some respondents gave the same response to all 18 questions (this is technically possible, but at least raises the question about the quality of collected data).
3. In Table 26 the candidate reports that there were no missing data in the responses. However, the fact is that there were missing data. In other place of the thesis it is admitted that missing data was replaced with the median. However, even this statement is not correct. An inspection of the source data revealed that missing data was sometimes replaced with 0 and in other cases with 3 (neither of the values were the median for any of the questions).

4. The candidate computes Cronbach's alpha, to conclude that there is a high level of agreement between responses. However, in my opinion this is not the case. Even the printout presented in table 27 warns about it as "Item 3" being negatively correlated with the rest. Inspection of the data confirms this observation "Item 3" is negatively correlated with all other items except "Item 18". The candidate ignores this fact.
5. Although analyzed data is discrete, the candidate presents it using histograms and kernel density estimators, see table 28 (which are both typically applied to continuous data). The candidate does not explain the rationale behind this procedure.
6. The formulation and testing of the hypotheses is questionable. Quoting from page 275 (other test is equally problematic)

   H0: The less experience in AI a responder has, the less positive Re_fish will be evaluated

   H1: The more experience in AI a responder has, the more positive Re_fish will be evaluated.

   First, later in table 33 the candidate performs a frequentist test with a point H0 (as opposed to the H0 given above).

   Second, in table 33 the candidate reports p-values for a two-sided H1, while the H1 specified in the thesis is one sided.

   Thirdly, the source data has duplicate values, and also the result of the Shapiro-Wilk normality test presented in table 31 reject normality of the distribution of this variable. So this means that both p-value for Pearson's r correlation and Spearman's rho are inexact. This is especially problematic given that the sample size is very small. The candidate ignores these aspects of the testing procedure.

   Fourthly, the reported p-value is significantly greater than 0.1 (a typically accepted highest value for rejection of H0 in frequentist testing), but the candidate ignores this and states that the tests indicate that H0 should be rejected.

   Finally, if one reads formulation of H0 and H1 they express the same statement (because the candidate applied double negation). Therefore it is even not possible to decide what hypothesis the candidate actually aimed to test.
7. The last comment relates to the conclusion drawn by the candidate from the results. Assuming that indeed (as it is implicitly claimed by the candidate) we would obtain a result that number of years of AI experience is positively associated with positive Re_fish evaluation (although, let me stress that again, the collected data does not provide a statistically significant support for this claim) then the conclusion drawn by the candidate from this result is not justified in my opinion. The conclusion drawn by the candidate is:

   Both results express that the experts value the reference architecture as being "useful" regards to the evaluated aspects.

   This is clearly a different statement than the verified hypotheses. The positive correlation indicates that more experienced in AI experts valued the reference architecture more. This is a significantly different statement, and, moreover, it does not need to mean that it is positive for the Re_fish reference architecture, because it might indicate that less experienced experts cannot really benefit from this architecture, and most likely it is exactly these less experienced experts who need it most (likely someone who has a lot of experience in AI does not need as much guidance as a person with less experience when developing an AI system for the company). Of course this is just one possible interpretation – whether it is correct would need to be verified by gathering some additional evidence. However, I have written it to show what kind of conclusions could be drawn from such a result of statistical analysis (to contrast it with the conclusion given by the candidate which is, simplifying a bit, ”*architecture is useful*”).

Summarizing the evaluation of the fact if the thesis shows that the candidate has theoretical knowledge in the domain of economics and finance I judge that in *its current state the evidence is borderline*. The candidate applies tools from macroeconomics, microeconomics, and quantitative methods domain, however, there are numerous issues with how the candidate uses these results.

Assessment if the dissertation presents an original solution of a research problem

Let me recall the hypothesis of the research that is provided by the candidate:

> *By developing a reference architecture for an explainable AI system that could combine both subsymbolic and symbolic approaches, confidence in AI models and, thus, decision-making in corporate planning can be improved.*

In what follows, I am assuming that the thesis formulates the assumption of implementation of the reference architecture as discussed in the previous section.

First I want to highlight that if the hypothesis were verified in the thesis then it would be an original solution to the problem and would be a valuable research contribution. The subject of XAI is currently one of the most important areas of research of application of AI models in practice. Therefore, in what follows I concentrate on evaluation if indeed it can be assessed that the thesis provides a verification of this hypothesis.

In my opinion to verify the stated hypothesis there should be presented an evidence that implementation of Re_fish indeed has a positive impact on decision-making in corporate planning in process industry. I assess that such evidence is weak in the presented thesis. This link is presented in section 5.3 of the manuscript. The argumentation presented there has two angles. The first one is against the principles of design science. The second is using an expert survey.

The verification against principles of design science is again borderline. Let me quote, as an example, the description of verification against the first guideline (the other guidelines are described in a similar manner):

Guideline 1: *Design as an Artefact: The reference architecture as a purposeful IT artefact, addressing a fundamental organisational problem: the design, construction, and running of a trustworthy AI system.*

The reference architecture Re_fish was designed and created in chapter 5 as an artefact (s. Re_fish business architecture, Re_fish information system architecture and Re_fish technology architecture). The architecture was built by following best practices for design using a combination of ADD and ADM. Therefore, guideline 1 is fulfilled.

Such a description is declarative as it has a form self-assessment by the candidate. It does not disqualify it, however, it does not prove the connection between the architecture and the potential economic benefits of its implementation or its impact on decision-making processes. The same reasoning applies to the other guidelines. In particular such self-assessment performed by the candidate does not answer two key questions:

- If the architecture is implementable (i.e. if it is possible to create an instance of the IT system that implements it)?
- If the implementation of the architecture would lead to an increased confidence in the AI models by decision-makers which in consequence would improve decision-making in corporate planning?

An ideal process leading to verification of these statements would be to provide an example implementation of the architecture (to prove its implementability) and then survey the users of such implemented system. I agree that, given the scope of the architecture such task is hard to achieve within the scope of the thesis. However, what I would expect is that the key components of the architecture namely the: symbolic and sub-symbolic components along with their integration and one sample AI model explained by them would possess an implementation. Essentially what I would expect is that slide 18 from the supplementary material attached to the thesis would not be just a mockup of the solution, but it would have some implementation (without requiring it to have a graphical interface or full scope of models covered). This would prove that indeed the key component of reference architecture can be instantiated (its other components are standard) and could also help with verification if indeed it helps the users to build confidence in recommendations produced by an AI model.

The second method used by the candidate is expert survey. I provided detailed comments related to the methodology of the survey in a section of this assessment covering the knowledge of the candidate of the economic theory. Therefore, here, I concentrate solely on the scope of the survey questions. They assumed that respondents had to assess the reference architecture itself (and not the implementation of the architecture in the actual system) in the following dimensions:

- U1: is architecture understandable?
- U2: does the model clearly define all four levels of architecture?

- U3: what is the mapping of processes/components to business actors?
- U4: are business process steps easy to follow?
- U5, U6, U7: is the interface sufficient for the use case?
- P1: is using a Knowledge Graph data bank and subsymbolic explainer a sufficient approach?
- P2: is Data Services component architecture sufficient?
- P3: is Re_fish approach to interaction of symbolic and subsymbolic AI sufficient?
- R1: the question is not a question but a statement by the survey author about the properties of the architecture with regards to model deployment to production. I quote the question verbatim:

  > *R1: One of the results of the thesis is that the explainability of an AI must already be guaranteed in the design and throughout the entire life cycle. In addition to the architecture, this also includes comprehensive lifecycle management. A component for transparency is also the tracker component, which makes it possible to track the status of the non-symbolic machine learning model and to detect deviations if necessary. By separating development, testing and production, it can be ensured that no biased model or its results end up in production.- Slide 15*

- S1: are user roles in the architecture clearly defined?
- S2: the question is not a question but a statement by the survey author about audit module of the architecture. I quote the question verbatim:

  > *S2: One of Re_Fish's key requirements is to address society's growing concerns in AI by ensuring ethical principles and compliance with regulations and standards (GDPR) throughout the lifecycle of the AI model. One of the main components of Re_Fish to ensure this requirement is the use of a separate audit module that contains secure, proprietary (i.e., separation of concerns) access to all information, logging (tracker), metadata etc. of the AI model, but also of Re_Fish itself. - Slide 14,15*

- F1: Are the application and the infrastructure components properly described?
- F2: the question is not a question but a statement by the survey author about the dialogue components. I quote the question verbatim:

  > *F2: The dialogue components of the auditor and the business user are separated. the business users are divided into different groups, knowledge engineer, planner etc. The users can thus be assigned to groups via their roles, which then receive the necessary authorisations. The activities of the users are recorded in the tracker. Slide 15- 18*

- Q1: Can the architecture be easily instantiated?
- Q2: Can it be assessed that the architecture be implemented using the existing technologies?
- Q3: Is it possible to estimate the cost of implementation of the architecture?

Questions Q1 to Q3 are related to the implementability of the architecture, and the experts, in general, assess it as implementable.

No questions are directly related to the question if implementation of the architecture would increase the confidence in AI models. One could consider that questions U5 to U7 are related to this objective, as they ask if the architecture is sufficient for the use case (which can be implicitly understood as providing information about trustworthiness of the AI models; it should be noted though that the questions did not directly ask this question).

Given the above considerations I assess that there is evidence that the presented research provides a verification of the stated research hypothesis, however, this evidence could be significantly improved. This is especially true considering the fact that the process of analysis of the survey results raises significant questions regarding its correctness (as I have explained in the evaluation of candidates knowledge section).

The next aspect of evaluation of the hypothesis is as follows. The title and precise stated objectives of the thesis relate to application of AI for corporate planning in the process industry. However, the research hypothesis stated in the text of the thesis does not relate to the process industry (it is wider). Therefore it is not clear what hypothesis the candidate actually aims to verify. Assuming that the candidate wants to concentrate on corporate planning and process industry it is essential that the study as a whole concentrates on them. This implies two consequences:

- It should be clearly stated how Re_fish is corporate planning for process industry specific (i.e. what makes it specifically designed for this use case; when the Figure 86 is inspected what components and how are specific for the intended use-case). *I assess that this condition is fulfilled positively by the proposed reference architecture*.
- The justifications for the need of Re_fish should be applicable to corporate planning for process industry cases. I find majority of the considerations presented in section 1.1 and 1.2 of the thesis not directly and clearly linked to this use case (and some of them are clearly not applicable as they reference uses of AI that have a significantly different nature). *However, indirectly, also evaluation of this criterion can be assessed positively*.

The above point, however, is related to the line of reasoning behind application of XAI for corporate planning in the process industry presented in the thesis. Let me explain this on an example. On page 37 it is stated:

> *For example, if sales teams are able to understand the decisions of the AI model, they will trust the model and use it even if it makes decisions that are "incomprehensible" at first glance, such as a navigation system that suggests an alternative route based on information that is not (yet) available to the driver.*

This statement, in the context of the thesis, implicitly implies that the "navigation system" is either:

- Not considered to be AI by the author (I would say it is unlikely – these systems involve highly complex algorithms that I think can be reasonably classified as AI solutions).
- It is an example of a system with an explainability layer that makes the users trust it (it could be argued that indeed such a layer is typically provided as such systems visualize routes on maps and typically allow to compare several options).

However, I believe that this example highlights an important aspect of application of AI that is, in my opinion, largely neglected in the thesis, which I think should be addressed as it is crucial for it. In a navigation system users can be considered to use it because they believe that it performs its work well. This belief can be partially built by explainability layer as I have commented, but also there are other sources of such trust. Some of them are: experience (users trust technology because they used it in the past and it worked well and gave significant value added), confidence in the technology used and the quality of the process of its deployment to production. These are example aspects that could be considered substitutes to explainability. In the sales team example, even if the reasons for decisions of the model are not understood by the sales team, they might, for example, by experience learn to trust the model if indeed it consistently produces good predictions. Still, I admit that expainability might be helpful in lowering the barrier of entry for new technologies.

Also explainability, as is discussed in the thesis, is a desirable feature from the regulatory or safety reasons. This aspect is noted in the dissertation. What is lacking in the thesis is a discussion what is indeed the main driver of the need for XAI for corporate planning in process industry: need for trust or regulatory requirements. I believe that an extensive discussion of these two factors is crucial and, if the candidate claims that the regulatory aspect is important, it should be clearly shown and explained in the reference architecture how it is fulfilled by it (in the current presentation of the architecture it was not clear for me how this requirement would be fulfilled).

To summarize this comment let me state that I do not require the candidate to agree with my specific opinions. However, I would expect that the candidate presented a much deeper discussion of such aspects in the thesis in comparison to its current state.

In my opinion the thesis mostly presents a limited critical perspective on the reviewed literature and instead relies on re-statement of the findings presented in source material, while the candidate's contribution is discovering this source material.

To highlight this aspect let me give one more example. The candidate claims that the distinguishing aspect of the thesis is development of the method of combining the symbolic and sub-symbolic approaches. However, a detailed investigation of fragments of the thesis that relate to this aspect reveals the following. The concept of the causal inference engine is taken from the works of Judea Pearl and I do not see a significant value added from the candidate's side. It is mostly a restatement of what Judea Pearl proposed, with a difference that the restatement contains errors. To give precise example of such an error let me mention. In figure 82 the candidate states that a "General cause and effect model" is presented, while in the original article it is considered as just an example of possible causal assumptions about three variables selected in such a way that the assumptions lack testable implications. Similarly the source article gives for this example a formula for the estimand E_S as $E_S = \Sigma z \, P(Y|X,Z)P(Z)$, while the candidate copies it with an error and changes the description to "Pearl formulates the overall problem as a Bayesian equation in that $\exists z = \Sigma_Z P(Y|X,Z)P(Z)$" where not only the $\exists z$ is just hard to interpret, but instead of the original precise notion of giving a formula for the estimand the candidate call this "the overall problem", but does not define it. Finally, the candidate claims that this is a general relationship and at the same time assumes that Z represents gender (see top of page 256).

Similarly the discussion of the explanation process presented in the pages 126-127 is taken virtually verbatim from Chakraborti et al. (2020), but also contains errors in typesetting and imprecisions in the description. Finally, the way how the candidate gives credit to the original authors is in my opinion not fully transparent, as the reference is the following:

> *The planning-based explanations show a complete representation of the respective plan as an explanation. In the following, however, we will first describe typical decision variables that can occur in the context of corporate planning (Chakraborti et al., 2020).*

In my opinion such a reference is not a clear indication that almost two previous pages of the text were based directly on Chakraborti et al. (2020).

Summarizing the assessment if the dissertation presents an original solution of a research problem my conclusion is that *the research problem is currently an important area of study and the candidate provides enough evidence of originally solving it*. However, it should be highlighted that:

1. The way how the hypothesis of the thesis is verified could be significantly improved.
2. The candidate should be more precise in stating which results are his own and new, and which results are just restatements of the research previously published by other authors.

Assessment if the dissertation the ability of the candidate to perform scientific research

Throughout the thesis the candidate has a tendency to state very general statements that are not precisely justified. As an example take a paragraph (I stress that this is a whole contents of the paragraph in the text on page v):

> *The use of AI, especially subsymbolic black-box models, presents the above challenges.*

What is lacking in my opinion is:

- Clear explanation which challenges; the statement "the above" is not precise, the candidate presents diverse challenges in an unstructured way. For example the two paragraphs preceding this paragraphs do not present any challenges that could be directly related to this sentence. They respectively describe process industry and corporate planning in general.
- Clear justification, for each of the challenges, why "the use of AI, especially subsymbolic black-box models" presents these challenges. Now it is just a declarative statement that is a subjective opinion of the candidate.

Such low precision of the presentation is visible in numerous sections in the thesis. As an example "the above" reference is used 20 times in the text and it creates imprecision to what the candidate exactly refers to.

Similarly it is common in the thesis that the candidate presents own opinions that are not grounded in the presented evidence (either empirical or from the literature). For example on page 24 the candidate states:

> *Often, their [implied: human decision-makers] training only allows for decision-making by going on a so-called "good gut feeling".*

It is not clear on what basis the candidate formulates this statement. The objective of the thesis is analysis of the corporate planning in process industry. Is it indeed the fact that decision makers that are involved in this process lack adequate training for these tasks? Maybe this statement is true, and grounded by the evidence, but it is natural to challenge it. It would be expected to assume that in process industry employees are selected and trained to be able to perform corporate planning decisions using data and facts and not only "good gut feeling".

The candidate presents numerous findings of the thesis. In general this practice should be judged as positive. However, the execution of this practice is in my opinion not precise enough.

Take the fourth finding as an example, I am quoting it in full:

> *Finding 4: Front runners participate most. This could lead to "supercompanies".*

Such a finding is a slogan-like sentence that I judge as not precise enough if it is to evaluated as a result of research.

Similarly, the candidate states above this finding:

> *The findings discussed above from both macro and microeconomic perspectives will be a reference point in the following, not to mention the subdivision of AI into symbolic and subymbolic AI methods.*

I did not find clear references to these findings later in the thesis as is promised in this sentence. Additionally, I would like to highlight that the sentence itself does not seem to be grammatically correct.

Finally the candidate uses the word "finding" in a non-standard way. In scientific context typically a finding is understood as "information that has been found out about something that was previously unseen or unknown". In other words normally it would be expected that findings are candidate's own results. In this thesis it is, in general, not true. Numerous reported findings (like e.g. finding 4 quoted above) are just a re-statement of the results of a single source study. In other words, the candidate often uses the word "finding" in a sense "what was found in the literature" (with a limited creative input from the candidate).

This comment leads to the general remark regarding the thesis. The amount of the material presented in the text is very vast. The total number of pages of the text is 332. The vastness of the text is partially due to the fact that the thesis often deviates significantly into aspects more general and indirectly related to its main objective. It has the following negative consequences:

- Firstly it is often not clear how given considerations or analysis contribute to the achievement of the objective of the thesis (or even if they apply).
- Secondly it makes the thesis hard to follow as it is exceedingly voluminous (thus the thesis does not show that the candidate has appropriately acquired the skill to select the relevant research material that contributes to verification of the formulated hypothesis).
- Thirdly many of the key aspects of the thesis are described without enough precision, likely because already there was a lot of material in the thesis. This is best seen by the use of the word "mentioned" in the thesis. It appears 62 times in it. I would expect that the thesis is more narrow (i.e. would cover only material that is needed for justification of the hypothesis), but at the same time much deeper in these aspects (and not just mention numerous aspects very shallowly). The similar thing can be said in places when the candidate uses the word "briefly" (in a sense that something is briefly described, 16 times).

I recognize the need for putting of the presented research in a wider context of both economic research and AI research. However, those wider considerations should be properly balanced and a focus should be put on a direct objective of the thesis, which itself should be studied in-depth.

The volume of the text contributes to the issues with the quality of references. I have already commented on them but let me give a set of selected examples of issues with the references:

1. On page 314 the candidate gives two references to Pearl J. (2019). However, they are not distinguished (typically they would be by 2019a and 2019b). Therefore when they are cited it is not clear to which of the citations the candidate refers to.
2. On page 314 the candidate gives two references to Pearl J. (2009). However, they are not distinguished (typically they would be by 2009a and 2009b). Therefore when they are cited it is not clear to which of the citations the candidate refers to.
3. I could not find the position (Rogers, 1983) referenced to on page 34.
4. Footnote 6, leads to two short press announcements. They do not allow to verify the claimed thesis of the candidate that “to effectively execute such scenarios” [implied: use of AI in drug development]. They are indeed related to drug development but do not explain how AI is planned to be used. Also I have not found either of the references in the bibliography.
5. On page 35 the reference Kraus et. al (2022) is a report. The link provided in the bibliography is broken (I was able to retrieve the referenced report after fixing the link). The reference itself, although it is not peer reviewed is a valuable material. The candidate, however, uses only it to justify in the following the economic importance of implementation of XAI methods. It is especially limiting as the reference is general and does not directly consider corporate planning applications in process industries.
6. On page 329 the candidate gives a link to the Reference Architecture that does not work. The provided link is: https://sync.luckycloud.de/f/3d519a96aaa940d58f29/.
7. The candidate uses inconsistent style of description of positions in the bibliography section (there are several different styles used)
8. The candidate cites Gordon (2016) but there is not such position in the bibliography (there is other by this author; I have already raised this issue earlier in the assessment)
9. The author cites two works of Garcez, A. d [Artur d'Avila], & Lamb, L. C [Luis C.]., one from 2020 and one from 2023. In my opinion these are identical works. It is not clear why candidate includes them both.
10. Menzel, C., & Winkler, C. (2019) is included twice in the bibliography; one of them is under “C” (I have already raised this issue earlier in the assessment)
11. Chakraborti, T., Sreedharan, S., & Kambhampati, S. (2020) and Chan, F. T. S. (2005) are mixed into one bibliography entry.
12. Chakraborti, T., Sreedharan, S., & Kambhampati, S. (2019) is not cited in the text.
13. Bass, L., Clements, P., & Kazman, R. (2022) has a wrong publication year reported. Also in the text sometimes the candidate refers to it with year 2022 and sometimes with 2021.
14. Bass, L. (2013) is not cited in the text.

15. Dwivedi, R [Rudresh], Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., & Ranjan, R. (2023) and Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R [Ro-hita], Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Ki-zgin, H., Kronemann, B., Lal, B., Lucini, B., . . . Williams, M. D. (2019) are not cited. There is a reference in the text of Diwedi et al. (2022), but it is a different year and a different first author name (but they both might have been typos). Still it is not clear which of the positions were actually cited.
16. Purdy and Daugherty (2016) cited in several places is missing in the bibliography.

There are numerous writing style and precision, typesetting and language issues in the text. For example (I am listing several that I found most confusing):

- The candidate provides only one hypothesis and then a long list of research goals and research questions. I did not find a discussion showing that they are fully logically connected (they are related, but the candidate fails to explain the link between them).
- The candidate provides repetitive definitions of the same terms in different places of the thesis. Often these definitions are not fully consistent (e.g. one is informal, the other is more formal, but not exactly the same). Also many of the definitions, as claimed by the candidate, cannot be classified as definitions, but rather are general descriptions. For example here is what the candidate states to be a definition of *Fair and Ethical Decision-Making* (page 45):

  > *There is an increasing demand by the public for fair and ethical decision-making alongside explainability, e.g., concerns raised by politicians and other stakeholders that AI or algorithmic decision-making is influencing social life more and more, such as the COMPAS system. Pursuant to the GDPR of the European Union, individuals affected by any algorithmic decision have the right to file a claim (Bejger & Elster, 2020; Goodman & Flaxman, 2017; High-Level Expert Group on Artificial Intelligence, 2019; Lipton, G., 2016).*

  After reading this definition I am unable, if I were given some decision process, to decide if it follows fair and ethical decision-makng principles or not.

  Similarly the following explanation is unclear in my opinion:

  > *The difference between interpretability/explainability (the first is more the "understandability" of a model during "runtime" directly when the decision is made) versus an Explanation is as follows: the first is somewhat intrinsic, while the second is more or less explicit (done afterwards).*

  As a last example *transparency* is defined as: "*A model is considered transparent provided that it is understandable*". But just as a next definition we get a definition of *understandability*. This poses a question if thus the candidate states that transparency and understandability are synonyms, or they are different concepts (since they in the text get a separate and different, although similar, description).

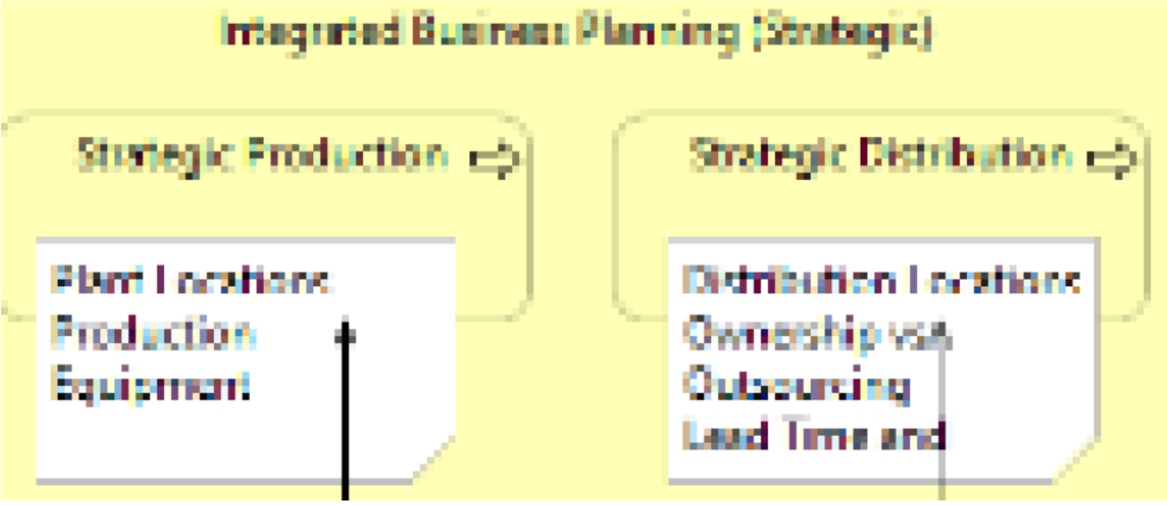
In general, when candidate gives some definitions, it would be also useful to provide examples of the systems that meet these definitions (to prove that these definitions are not just abstract concepts but have concrete instances of systems/objects that implement them).

- The candidate in the text distinguishes “reference architecture” and “reference model”, but later (page 223) writes “The main goal of this work is to develop a reference architecture as a reference model” which contradicts the distinction between these two concepts introduced earlier.
- The candidate often promises to give explanations that in fact are not such. For example let me give explanation of ELI5 from page 166 (the candidate promises to explain how ELI5 works as a AI model explainability framework):

  > *ELI5 (“Explain Like I’m 5”) is a library which is based on the programming language PYTHON. The idea behind ELI5 is that it should be used for AI pipelines, in order to v18isualize and debug various machine learning models by using a unified API. The library provides built-in support for several ML frameworks and a way to explain black box models. The result of using ELI5 could be a table, for instance, where feature importance for a specific model can be seen.*

  I claim that from such a description the only thing the reader learns are general statements that this is an model explainability framework (which is known before giving this description). The reader does not get a precise (in a scientific sense) understanding how the ELI5 framework works and what it does. The provided description is mostly software developer oriented, and not research oriented.
- Page 230, the candidate writes “*To this end, Bejger and Elster (Bejger & Elster, 2020) call for existing life cycle models for AI and machine learning to be adapted so that no bias or the like can occur from the beginning to the end of the use of an AI model.*”. This is just one example of a style that is not precise. What does the candidate mean by “*or the like*” in this sentence. This is an informal style that is typically not recommended in scientific writing.
- The candidate often provides captions for tables/figures that are highly general without giving precise information what they present. For example Figure 67 has the following caption “High level conceptual diagram”, but does not say of what.
- Page 287, the candidate writes “to proof” (a shortened version of “to proofread”) while most likely “to prove” is meant.
- The candidate consistently uses the “so-called” term, which normally has a negative meaning (used to show that you think a word that is used to describe someone or something is not suitable or not correct); I would not assume that this is the intention of the candidate, however, typically when we use “so-called” in a neutral sense this should be connected with a term that is not widely used, which is often not the case in the thesis.

- Use of confusing and not explained notation like “*N/A -> AR N/A*” on page 280 (I can guess what the candidate meant, but this is not a style typically used in scientific writing)
- Table 37, which is actually a figure. On the figure there is a mixup of “neutral” and “neither agree nor disagree” labels (in the text the candidate states that they mean the same) and also “neutral” is put as worst in the order (which should not be done as the candidate is working with ordered categorical variable).
- The use of typically not applied typographic combinations like “->” and “-Y” on page 270 (I assume the candidate wanted to put an arrow there).
- On page 230 the author gives reference to table 11, but it seems to be an incorrect reference.
- Key tables in the text are using extremely small font, like table 16 or table 22 (but fortunately when zoomed-in they can be read).
- The most important figures in the text, like figure 68 or figure 86 (and many other related) are practically illegible both in print and in PDF. And these figures are crucial since they present the key product of the thesis. Let me show a zoomed-in screenshot of a selected part of figure 86:



- On page 163 there is an enumeration, but it is not clearly marked as such (it is made as a continuous text with consecutive paragraphs).
- The candidate does not use paragraph indentation (it is a standard practice to do so in scientific writing). Also the candidate uses an inconsistent inter-paragraph spacing of text (it was not clear to me if this varying spacing should bear some meaning – e.g. if large space implied a new thread in the reasoning).
- On page 155 the candidate mentions KNN (normally understood as k-nearest neighbors algorithm), but this model is never referenced in the text nor defined.
- There are fragments of German text as parts of English text in the thesis: “*Finding 10: Challenges in the process industry ergeben sich, wie bereits oben beschrieben, aus der hoc The challenges in the process industry arise from various aspects.*”.

To conclude, let me highlight one specific significant aspect that raises questions is approach of the candidate to defining sets of related concepts in the text. For example symbolic AI and subsymbolic AI are defined before the concept of AI is defined. Additionally these two definitions are given by explaining of their differences (table 1 in the text). In what follows the definition of AI is given on page 42, and it is:

> *Thinking of an agent (which in this work is placed equal to model or algorithm) operating autonomously while perceiving the environment as persisting over a prolonged period, adapting to change, and creating and pursuing (the right) goals.*

However, this definition contradicts the information presented in table 1, where the candidate states that symbolic AI is, in particular "rigid and static", which is contrasted to subsymbolic AI, which is described as "flexible and adaptive". Which means that symbolic AI is not AI (per candidate's definition, which requires adaptation to change).

I agree that the candidate captures well on an intuitive level the important aspects of various concepts, but at the same time, in a research document (as opposed to the e.g. a business document that is meant to build an intuition of the reader) providing a precise system of definitions that is internally consistent is of crucial importance.

In summary *the ability of the candidate to perform scientific research is again borderline positive*. The candidate indeed has shown the ability to conduct scientific research. However, the number issues that can be found in the dissertation is significant.

Summary

The subject of XAI that is researched in the assessed thesis is currently one of the most active areas of AI research. Therefore I positively assess the selection of the area of the thesis as interesting and important scientifically.

In the previous sections I presented a detailed assessment of the thesis in the domains of:

1. Whether the solved research problem can be classified within the economics and finance field of studies.
2. Whether the dissertation shows that the candidate has theoretical knowledge in the domain of economics and finance.
3. If the dissertation presents an original solution of a research problem.
4. If the dissertation the ability of the candidate to perform scientific research in an independent way in domains: scientific methodology and written communication.

For all of the areas I have given a summary evaluation in the respective sections of the thesis. In each of them my opinion is positive, however, also in each of them there were significant issues that I have raised.

Conclusion

I judge, despite the issues I have risen in the assessment, that the dissertation meets the requirements of the Act in the discipline of economics and finance.

I would like to stress, however, that in the case that the thesis is allowed for a public defense, the candidate should provide for the defense committee detailed explanations related to the issues I have covered in my assessment.

Following the request of the Nicolaus Copernicus University in Toruń below I present the translation of the conclusion in Polish language.

Konkluzja w języku polskim

W mojej ocenie, pomimo zastrzeżeń, które podniosłem w recenzji, praca doktorska spełnia wymogi ustawowe w dyscyplinie ekonomia i finanse.

Chciałbym podkreślić, że w przypadku dopuszczenia do publicznej obrony, kandydat powinien przedstawić komisji doktorskiej szczegółowe wyjaśnienia związane z kwestiami, które zawarłem w mojej recenzji.

……………………………………………………………………

Bogumił Kamiński

Podpisano przez /Signed by:
BOGUMIŁ KAMIŃSKI
Data /Date: 2023.09.04 17:22
mSzafir