FACULTY OF ENGINEERING TECHNOLOGY
GROUP T LEUVEN CAMPUS

Creation and automation of a weighted maturity score based on an analytic hierarchy process and a transitive model

Grigorii Osipov
Jacques Tatossian

Supervisor: Prof. Dr Ing G. Waeyenbergh
Co-supervisor: Prof Dr Ir L. Pintelon
Daily supervisor: Nicole Berx

Master’s Thesis submitted to obtain the degree of Master of Science in Electromechanical Engineering Technology

Academic year 2022-2023

Acknowledgements

We would like to express our great gratitude to Nicole Berx, Professors Liliane Pintelon and Geert Waeyenbergh for the successful completion of our thesis. Their expert knowledge, support and guidance were invaluable throughout our research.
A personal thank you to Nicole, who was instrumental in shaping our research in day-to-day activities, and whose insightful feedback, constructive criticism and encouragement were immensely helpful in improving the quality of our work. Her great introduction and preparation of our topic for us were greatly appreciated during the introductory period, and her flexibility with the direction of the research we took while remaining supportive yet steering us in the right direction cannot be overlooked.
We are equally grateful to our promotors, who with their expert knowledge and direct feedback allowed us to showcase the results in the best way and their motivation, and interest in our work helped keep us focused and driven throughout the research purpose. Additionally, we thank Professor Geert Waeyenbergh for providing us with a management topic.

Personal thank you to those whose input helped us complete our study:

- Dmytro Matsepa – For helping with initial structure and feedback on the automation of the program.
- Prof. Klaus Krippendorff (1932-2022) – For his eponymous and widely used interrater reliability metric, which was a vital part of the paper.
- Creators of SANOL – For having a large forming influence on this paper
- Prof. Lasrado – For input and previous work completed on the topic of non-linearity of maturity.
- Developers of “fast-krippendorff” – For saving us tens of hours of development of Krippendorff’s alpha calculator module.
- Ing. Kaat Van Assche – For the valiant effort in recruiting academic experts.

CONTENTS

1 LIST OF FIGURES

2 LIST OF TABLES

3 LIST OF SYMBOLS AND ABBREVIATIONS

4IR	Fourth Industrial Revolution
AHP	Analytic hierarchy process
AI	Artificial Intelligence
Cobot	Collaborative Robot
CR	Consistency Ratio
CSRAT	Cobot Safety Readiness Assessment Tool
Excel	Microsoft Excel
IT	Information Technology
KU Leuven	Katholieke Universiteit Leuven
MCDM	Multiple Criteria Decision Making
PROMETHEE	Preference Ranking Organization METHod for Enrichment of Evaluation
SMART	Simple Multi-Attribute Ranking Technique
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution

4 ABSTRACT

Maturity models are an important tool in determining the current level of performance of systems based on their maturity score, creating a solid foundation for future growth of industrial implementations. The motivation for this research is to advance the field of robotics and developmental models by enhancing awareness and knowledge towards the safety of collaborative robots.

This is accomplished using a maturity model, which helps identify areas for improvement and reduce the risk of accidents and injuries in the 4^{th} Industrial Revolution industries that use collaborative robots (cobots). By employing a maturity model, one can effectively identify and tackle the safety hazards associated with collaborative robots (cobots). The findings of this research have the potential to inspire industries to evaluate and enhance the safety features of their cobots concerning their surrounding environment. This study addresses the weighted calculation of levels of safety maturity within the implementation of cobots and questions the importance of weighting when calculating representative results. Additionally, this paper aims to offer an understanding of collaborative robots, assess the safety awareness of cobots in industries, explore various weighting models, develop a mathematical model by enhancing the adjusted Cobot Safety Readiness Assessment Tool, and simplify the assessment process by automating calculations with Python code.

The data for this research was collected through a Microsoft Excel survey based on the response of respondents with engineering backgrounds (n=9) and prior experience working with cobots. In addition, to accelerate the data intake, processing and report generation, a flexible Python program was created to implement the accompanying survey structure that can be used for similar research and reduced the time needed to process a single survey from 30 minutes to 0.1 seconds.

Different weighting methods are discussed in this research, with Analytic Hierarchy Process being chosen and the survey broken into multiple groups to decrease the length of the survey. Furthermore, different consistency concepts are discussed and used, such as consistency ratio (CR) and Krippendorff's alpha. Additional topics are considered in how they aid the mathematical model in obtaining a higher consistency and reliability in their results.

After analysing the results obtained from the survey, it was found that 33% of the responses had to be discarded due to a CR higher than 0.1 more than once per person. Nevertheless, the results suggest that an individualized correction coefficient can be calculated based on a person's interpretation during the weighting process, which can be used to adjust the weights and generate an adjusted maturity index score. The mathematical model that uses the coefficient provides more consistent results across different groups of risk factors. Furthermore, the study found that ordinal Krippendorff's alpha exhibits a negligible correlation with a sample size of 6.

5 INTRODUCTION

5.1 Introduction to research

Since the introduction of the first robot almost a century ago, robots have been slowly but surely integrating themselves at a higher rate into modern society compared to the 20th century. With the ever-increasing rise of automation and digitalisation of the industrial sector, also known as the 4th Industrial Revolution (4IR), humans were being replaced by robots steadily. However, around the year 2010, robots have increased the employment rate in the automation industry by increasing and creating new technology jobs (Chung & Lee, 2023). Furthermore, according to Chung and Lee (2022), the relationship between the growth of robotics in the industrial sector and its impact on jobs is complex and multifaceted, involving numerous factors that contribute to a nuanced understanding of the issue.

Over the decades the tasks of the human operator have decreased and have since then been displaced by automation systems in the industrial sector. Numerous advantages arise from this state of change within the industrial sector from robots being able to operate in hazardous environments, to lifting heavier loads and increasing the productivity of enterprises with a higher precision rate. On the contrary, disadvantages occur as well, as robots tend to take up a lot of space and are expensive and time-consuming when it comes to setting them up (Buchert, 2021). In addition, development problems are occurring within companies because of the overwhelming technical demand and incapability of executing strategies due to the technological transition of factories, which is in part due to the 4IR movement currently taking place. To solve these issues, but at the same time hold on to these advantages, similar to that of a robot that exists, collaborative robots (cobots) as shown in Figure 5-1.

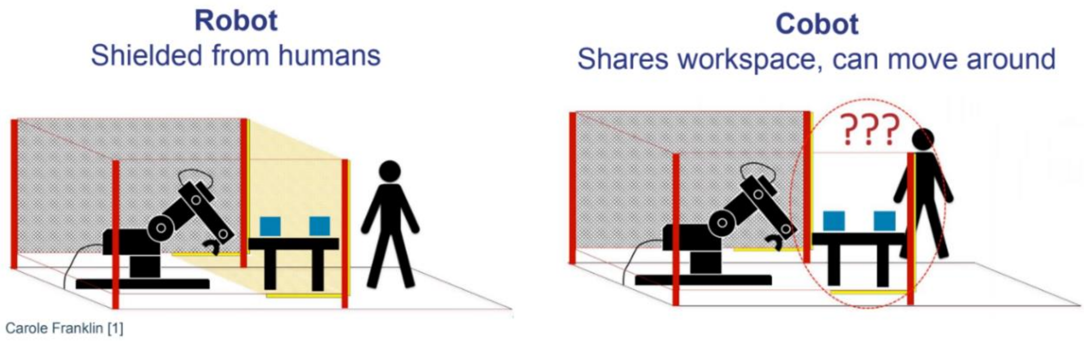

Figure 5-1: *Note*. The main difference between a robot and a cobot is no safety cage is required. From “Cobot. Safety.” “Presentation for The Association for Advancing Automation.” By C. Franklin, 2021. Copyright 2021 of C. Franklin

Collaborative robots can be referred to as socio-technical systems that combine both human and technological interaction, where both the cobot and the human performance depend on one another. Cobots have numerous advantages compared to robots: they are smaller, more flexible, easier to program, and most importantly, they collaborate and work alongside humans (Buchert, 2021). Additionally, they do not require a lot of space and physically share their working environment with humans, requiring no safety cage. Therefore, more emphasis has been applied to the safety of cobots and their surroundings, specifically on the ergonomic, organisational, and psychosocial nature. Thus, increasing the incorporation of cobots within today's enterprises can lead to many positive benefits for both employers and employees. However, at this current point in time, safety, and trust towards cobots has not yet been fully established, hence, limiting the full potential usage and incorporation of cobots within enterprises. The integration of artificial intelligence (AI) can help reduce safety risks associated with human and collaborative robots (cobots) working in the same environment. By improving accuracy and precision, AI can help detect potential hazards and respond appropriately to prevent accidents or injuries. Additionally, AI can help improve the handling of complex tasks with larger workloads, which can reduce the likelihood of human error and related safety risks, as pointed out by (Frąckiewicz, 2023). However, it also gives rise to new risks, including bugs, faulty or insufficient interface communication, and programming-related issues with learning algorithms that can threaten the integrity of the cobot and increase its intelligence (Berx, Decré, Morag, Chemweno, & Pintelon, 2022). A reference to this can be seen when the EU Commission Communication on "Building Trust in Human Centric Artificial Intelligence" declared that "AI systems should integrate safety and security-by-design mechanisms to ensure that they are verifiably safe at every step, taking at heart the physical and mental safety of all concerned " (Berx, Decré, Morag, Chemweno, & Pintelon, 2022).

One of the main focuses of this study is the use of cobots in the 4IR Industry, which aims to digitalize the manufacturing sector through collaboration between humans and machines. The lack of emphasis on practical implementation and safety awareness of cobot risk factors in the 4IR Industry is a concern (Berx, Decré, Morag, Chemweno, & Pintelon, 2022). Cobots must meet industry standards that combine engineering, logistics, and IT, and involve the standardization of products to improve interaction and collaboration between different technologies (Çınar, Zeeshan, & Korhan, 2021). While (Çınar, Zeeshan, & Korhan, 2021) focuses on the technical aspects of 4IR technologies, (Berx, Decré, Morag, Chemweno, & Pintelon, 2022) emphasizes the need to increase safety awareness of cobot risk factors. This thesis aims to address this gap by improving the safety awareness of cobot risk factors when it comes to the 4IR Industry technologies.

This study was conducted to obtain a better understanding of the importance of the current technological challenges to increase the safety awareness of cobots within today's society. Furthermore, robots are a necessary technology in today's world due to the high demand of today's society, where everything must be accomplished with the optimisation of time and accuracy (Javaid, Haleem, Singh, Rab, & Suman, 2022). The completion of this study aims to increase the implementation and trust in using cobots within companies. To achieve this objective, the study establishes an automatic weighted maturity model that can evaluate the level of development in areas where cobot risk factors are present. By using this model, the study intends to identify areas where companies may need to improve their practices to reduce risks and increase the efficiency of cobots in their operations. This approach can help companies become more confident in their use of cobots while ensuring that they safely and effectively. This evaluation is facilitated by a programmed code to simplify the calculation process and to identify critical areas in cobots that require maturity improvement. Enhancing the safety of the cobot and its surrounding environment in these areas can improve the trust in and incorporation of cobots in businesses. This leads to an overall increase in the awareness of risk factors of cobots.

At the start of this thesis, 23 risk factors concerning cobots were already identified and classified into 5 dimensions in the form of a Microsoft Excel (Excel) maturity grid, also known as the Cobot Safety Readiness Assessment Tool (CSRAT) developed by (Berx, Adriaensen, Decré, & Pintelon, 2022). By implementing a maturity model, organizations can assess and manage the level of completeness, development, and readiness in addressing the safety risks associated with cobots. Hence, maturity models are used as a mechanism to assess the maturity of a system implementation and its attached organisation. The CSRAT classified the 23 risk factors and divided them based on their corresponding dimension. In total 5 dimensions were established as shown in Figure 5-2.

- Technology Readiness dimension refers to the technological risk factors associated with cobots involving mechanical safety, physical separation, and energy management to protect the operator. Cybersecurity, additionally, is important to protect the business system from potential electronic vulnerabilities, which may take control of the cobot's controls, potentially endangering the operator. Additionally, insufficient system capacity for data processing can slow down the cobot's response time, creating another potential risk;
- Human Readiness dimensions focus on the importance of the human factor when it comes to managing risks in complex systems. Human-related risk factors are categorized into three subgroups: psychosocial risks, mental strain, and physical workload. Psychosocial risks include trust-related factors and work-related stress which are crucial in increasing trust towards cobots;
- Collaborative Workspace Readiness dimension groups all risk factors related to the workspace in which the cobot and operator will collaborate. The risk factors in the collaborative workspace of cobots include access control, workspace layout, and maintenance. It emphasizes the importance of safe cobot positioning, collision avoidance, and motion planning algorithms to avoid risks to operators in the workspace. The collaborative workspace is influenced by all dimensions, and it cannot exist without considering the human, technological, enterprise, or external dimensions as they all interact with one another;
- Enterprise Readiness dimension groups risk factors related to cobots involving ethical factors, organizational strategy, suboptimal training, and consequences related to the division of labour. It also highlights the importance of corporate vision regarding safety rules adaptation and employee participation in enhancing worker wellbeing and cobot performance;
- External Readiness dimension classifies risk factors in cobots related to safety regulations and environmental factors outside the company's control. Vague legislation can be a potential risk factor, and environmental factors like weather conditions can impact the workspace and cobot components.

These dimensions include various aspects of cobots that need to be addressed to minimize safety risks, such as their design, functionality, programming, training, and maintenance. The goal of the CSRAT and its maturity grid is to determine the current level of readiness of a cobot within an enterprise. This level of readiness will be determined by engineering master students who have prior experience working with cobots. Participants are expected to answer where risk factors reside on a level readiness scale of 1 to 4:

- Level 1: Unaware of the need to identify cobot risk factors;
- Level 2: Aware of the need to identify risk factors for cobots.;
- Level 3: Required Safety Knowledge & Tools are present, but not adapted for cobots;
- Level 4: Required Safety Knowledge & Tools are integrated with Safety Management procedures and are adapted for cobots.

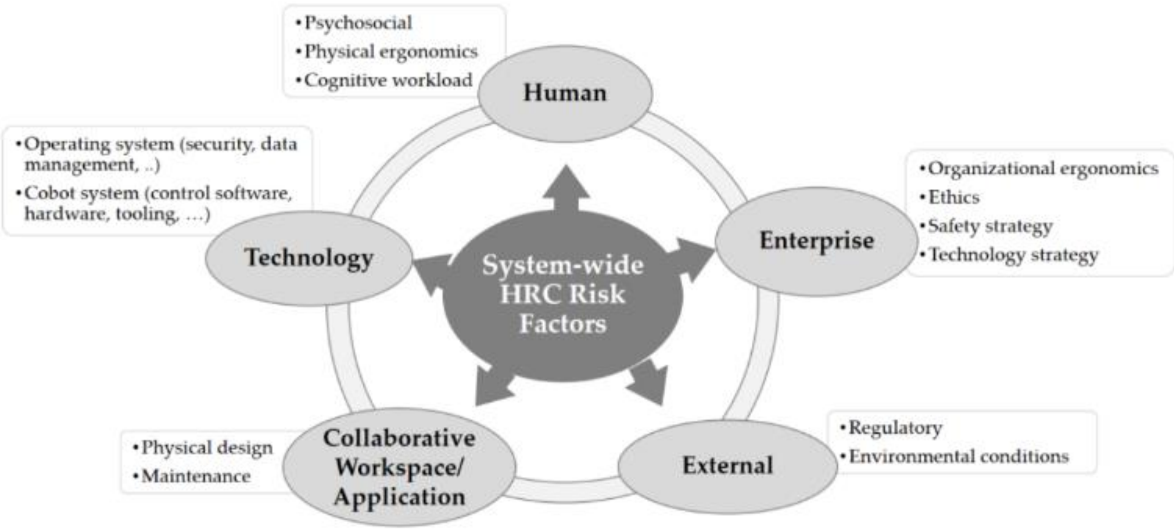

Figure 5-2: *Note.* The 5 dimensions and sub-dimensions of CSRAT. From "Assessing System-Wide Safety Readiness for Successful Human–Robot Collaboration Adoption" by N. Berx, A. Adriaensen, W. Decré, & L. Pintelon, 2022, Safety 8(3), (https://doi.org/10.3390/safety8030048). Copyright 2022 of N. Berx, A. Adriaensen, W. Decré, & L. Pintelon

The Analytical Hierarchy Process (AHP) is used to develop a maturity model for cobots by weighting 23 risk factors. AHP is a decision-making tool that can be used to structure and analyse complex decisions. AHP simplified the complex decision-making processes by comparing two factors against one another one at a time, using a weighting method known as pairwise comparisons (Adams, 2017).

To ensure the consistency of the respondents' pairwise comparisons, the consistency ratio (CR) is used. A CR value below 0.1 indicates a reasonable level of consistency, while a CR value above 0.1 suggests inconsistency (Adams, 2017).

Since inconsistencies are present when applying AHP, the concept of transitivity is used throughout this research to verify whether the pairwise comparisons done by respondents were consistent. Transitivity is an important concept when it comes to multicriteria decision-making. Transitivity states that if A is greater than B and B is greater than C, therefore, A must also be greater than C, as shown in Figure 5-2 . This can be replaced with numerical results and also apply to the weights that are calculated. Transitivity was applied externally in combination with AHP, as it was found AHP lacks transitivity at times when the CR is greater than 0.1, based on the pairwise comparisons done. The consistency ratio uses transitive properties to check whether the pairwise comparisons done by the respondents correspond with consistent results and calculate weights (Ji & Jiang, 2003). Additionally, it is also possible to use transitive properties to advantage and lessen the burden on the respondent during the application phase.

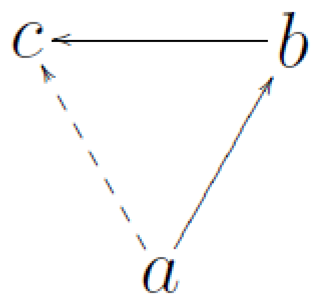

Figure 5-3: *Note.* Concept of transitivity: A>B and B>C then A>C. From “Wiktionary: (Transitive)”. In the public domain.

Krippendorff’s alpha is applied in this study, with the purpose to ensure the reliability of the results and to evaluate the level of consensus among respondents within enterprise. However, as the study progressed, Krippendorff's alpha did not fulfil its initial intended purpose, which was to measure consensus between respondents within an enterprise (Krippendorff, 2011). Therefore, Krippendorff's alpha is only used for theoretical purposes for this thesis.

In this research, a maturity model is developed to assess the level of development of risk factors associated with cobots. The purpose of the maturity model is to identify areas where improvements are needed to align with a predetermined standard. By evaluating the maturity of risk factors, the research aims to provide insights into the current state of cobot risk management and identify specific areas for enhancement. The maturity model will help guide efforts to improve the overall safety and risk management practices related to cobots, ensuring that they meet the desired standard (Helgesson, Höst, & Weyns, 2011).

Furthermore, in this study, the initial stage of the maturity model is being utilized, which involves evaluating the system, in this case, cobots (Helgesson, Höst, & Weyns, 2011). Additionally, for the evaluation process of the maturity model, type 1 and type 2 evaluation methods are applied. In Type 1 the maturity model is tested by the authors of this study to determine whether the maturity model complies with the standards of aims of this study. In type 2, the maturity is sent out to respondents with prior experience in working with cobots which is part of the application process of this study to obtain results (Helgesson, Höst, & Weyns, 2011).

Maturity models and maturity grids slightly differ from one another, as maturity grids are considered as a subset of maturity models (Maier, Moultrie, & Clarkson, 2012). Both are used for this study to for fill the objective of this thesis to obtain a maturity score for each of the 23 risk factors of the cobots. While the maturity grid (CSRAT) was already provided as a basis of this research to allocate the current level of readiness of cobots in the form of a grid structure, the maturity model is developed throughout this research to obtain a level of understanding of the performance of cobots through a mathematical model calculation.

Overall, this study aims to develop a comprehensive understanding of the applications of risks associated with cobots and determine their level of development through a maturity score. The use of AHP and CSRAT in combination with a programmed mathematical model and statistical analysis helps the reliability, accuracy by providing a structured and systematic approach to data analysis, and acceleration in obtaining results.

5.2 Research Questions

This research attempts to answer the following question:

- How can a weighted maturity model be implemented for Cobot Safety Readiness Assessment Tool (CSRAT)?

During the literature review and design, the findings led to further sub-questions, such as:

- How can the automation of the calculation and design of results from the survey reduce the number of manual calculations needed?
- How do individual levels of maturity of a risk factor affect its importance?
- How can the above questions positively affect cobot safety?

5.3 Outline of Paper

The upcoming structure of this paper is organised as follows; In the Research Methodology section, a literature review of the thesis is discussed in detail, providing a more comprehensive understanding of the concepts used and encountered throughout this thesis. In Methods & Materials, the implementation and design processes are discussed, explaining the design cycle of the survey, why these concepts have been chosen, what tools have been used, and the implementation of automation. Results showcase all the data and mathematical model results obtained from the respondents. In the discussion section, observations concerning different types of results and their objective to this research, the patterns in the results, their limitations and other research related to the results are made. Additionally, the validation steps that were created throughout the entirety of this research are also discussed to increase the accuracy of the research, and additional limitations of the research are mentioned. Finally, the Conclusion summarizes the main outcomes and inferences of this research paper.

6 RESEARCH METHODOLOGY

6.1 Literature Review

6.1.1 Introduction

To come up with an answer to the main research question, a literature review was conducted on scientific papers on concepts related to maturity models and the sections of interest which are cobots. To obtain the necessary information regarding the development of a maturity model, a keyword search was done on maturity models and cobots (4IR industry) using IEEE Explore, Science Direct, Research Gate, Springer link, LIMO and search-indexed results. The literature review for this thesis incorporated a selection of 22 papers out of a total of 39. The key points of interest when selecting a research article of importance to guide this research include the following:

- Detailed scientific articles related to the development process of maturity models and how to construct new maturity models;
- Detailed scientific papers related to cobots and their risk factors;
- Scientific papers discussing the different weighting methods, and what their advantages and limitations are;
- Scientific papers discussing concepts that can be applied in correlation with maturity models and how to optimize the automation process of data collection;
- General detailed literature review on maturity models.

As already mentioned, cobots are one of the many current digital technologies that are pushing manufacturing industries to continuously grow and proceed with automation processes. To further increase the incorporation of cobots within enterprises, a maturity model is constructed throughout this study based on already identified risk factors of cobots that have been defined by previous research done by (Berx, Adriaensen, Decré, & Pintelon, 2022)

The upcoming structure of this literature review first discusses what cobots are and why a lot of emphasis is put on the safety aspects of cobots. Since cobots are part of Industry 4.0, a brief discussion is made on Industry 4.0 and the concerns that arise with it. Section 3 focuses on the application process of maturity models and the difference between maturity models and maturity grids. Section 4 of the literature review reviews different weighting methods that are summarised to help develop a mathematical model which is a large part of this study since the combination of both the maturity grid and the mathematical model will develop the maturity model. In addition, the weighting method's function, advantages, and disadvantages towards the objective of this thesis are looked at in-depth. Lastly, different reliability concepts are shown and explained, that increase the reliability of the results of this thesis, such as Krippendorff's alpha and Lasrado's utility curve, with it their advantages and limitations towards the objective of this thesis considered.

6.1.2 Cobot Safety Analysis

In this thesis, safety and trust are critical components of the research, as improving cobot safety awareness can lead to greater trust between humans and cobots. This, in turn, reduces the fear associated with collaborating and sharing workspace with cobots, allowing cobots to be used to their full potential. Various safety standards already exist for cobots to mitigate risks, such as the Occupational Safety Health (OSH) standard mentioned by Berx and colleagues. This standard evaluates the risk of a work activity performed by a cobot that could result in physical harm to the operator or others in the workspace environment. Other emerging safety systems, such as the Safety II approach, which unlike its predecessor (Safety I), assumes that not all risks can be managed due to humans being a crucial part of the system, that there will always be a level of risk when dealing with cobots that cannot be eliminated (Provan, Woods, Dekker, & Rae, 2020). For instance, AI can reduce the risk that occurs in Human-Robot collaboration but can cause other risks in other domains, such as the physical, cognitive, and social realms as mentioned by Berx & al.

Upon further research on the evolution of enterprises within the 4IR, a scientific article by (Çınar, Zeeshan, & Korhan, 2021) was found concerning the 4IR readiness of smart manufacturing enterprises. It was mentioned by (Çınar, Zeeshan, & Korhan, 2021) that the 4IR aims to enhance collaboration sbetween humans and robots through the integration of various technologies. However, the integration of robots in the workplace poses certain risks towards humans, and this thesis seeks to increase awareness of these risks in collaborative robots (cobots), not from a technological perspective but by focusing on human factors through the increase of awareness towards risk factors of cobots. The 4IR Industry is currently taking place and has begun in the mid-2010s. It is the upcoming step in the digitalization of the current manufacturing sector, due to large similar improvements occurring within the industry such as the increase of data and connectivity, an increase of the analysis of data, increasing collaboration between humans and machines, and enhancement in robotics (Çınar, Zeeshan, & Korhan, 2021).

Two major concerns arise due to the transition towards the 4IR Industry; ”The first is a lack of instructions for defining the areas that need to be tackled to implement 4IR technology, and the second is a lack of knowledge about how to practically implement 4IR technology after the areas have been defined” (Çınar, Zeeshan, & Korhan, 2021). A correlation can be seen with another scientific paper written by (Berx, Decré, Morag, Chemweno, & Pintelon, 2022), which discusses that too much focus is put upon the technological factors when it comes to 4IR, with a lack of emphasis put on how the risk factors can influence the safety of humans when working in the collaborative workspace, specifically when it comes to the classification of risk factors of cobots. Both scientific papers refer to the lack of emphasis towards areas that need to be improved to facilitate the incorporation of 4IR technologies, which is related to this paper’s subject of cobots. However, it is unclear whether (Çınar, Zeeshan, & Korhan, 2021) are referring to the lacking emphasis put on areas that need improving in the technological or the safety domain, whereas (Berx, Decré, Morag, Chemweno, & Pintelon, 2022) is specifically referring to the risk factors concerning the safety of cobots towards the operators. Furthermore, the implementation of cobots within today’s industry is part of the 4IR transition and thus must meet 4IR industry standards. Standardisation of products worldwide is one of them, meaning, that a fundamental block of standardised procedures and products will improve the interaction between different parties and as well ease the integration and collaboration between different technologies, one of them being cobots.

(Berx, Decré, Morag, Chemweno, & Pintelon, 2022) and (Çınar, Zeeshan, & Korhan, 2021) concluded 4IR technologies, such as cobots have not reached a sufficient level of maturity and that concerns arise when implementing cobots within enterprises. While (Çınar, Zeeshan, & Korhan, 2021) focused on improving the technical aspects of 4IR technologies to solve these concerns. (Berx, Decré, Morag, Chemweno, & Pintelon, 2022) focused on improving the safety aspects of cobots to increase trust towards cobots, thereby increasing the incorporation of cobots within enterprises.

Since this research is based on previous studies done by (Berx, Decré, Morag, Chemweno, & Pintelon, 2022) thus, to answer the research aims and research question, throughout this research, cobots are referred to as socio-technical systems that combine both human and technological interaction because the performance of one can influence the performance of the other. For that purpose, a lot of emphasis has been applied to increasing the awareness of risk factors of cobots.

6.1.3 Maturity Model and Maturity Grids Analysis

A maturity model is a framework that helps identify weak points and improve the development of businesses, processes, and technological systems through a maturity score that compares it to a standard. It has two stages: assessment of the system and improvement identification to attain a higher maturity score. Maturity models are useful in situations where it is unclear how to change a process or system in the best way. However, it requires a great amount of effort to complete and should be linked to a framework to determine the best evaluation method based on the risks present. Additionally, maturity models are also used as a safety assessment tool for systems (Helgesson, Höst, & Weyns, 2011).

The application process or sending out of maturity models can be done in various ways based on the objective of the maturity model, which is a crucial step in this thesis once the maturity model has been developed to obtain the necessary results. The use of Type 1 evaluations in this study is an important step in verifying the working principle of the maturity model before sending it out to respondents. Type 1 evaluations are conducted by the authors of the model offline, without involving any outside experts. This allows the authors to test the model and ensure that it is functioning as intended and producing accurate results. Additionally, it enables the identification of any potential issues and makes necessary adjustments to ensure the model is functioning as intended before involving respondents in the evaluation process (Helgesson, Höst, & Weyns, 2011).

The use of Type 2 evaluations in this study is essential to ensure the validity of the results obtained from the maturity model. Type 2 evaluations involve practitioners who are experts on the process or system that the maturity model is designed to improve, but who were not directly involved in its development. In the context of this study, this means that respondents with expertise and knowledge on cobots will be involved in the evaluation of the maturity model for assessing the level of maturity of cobots in the workplace. Overall, the use of Type 2 evaluations enhances the credibility and accuracy of the results obtained from the maturity model. By including respondents with prior experience in working with cobots in the evaluation process, this study can obtain a more objective and unbiased assessment of the level of maturity of cobots. This is because the experts can bring their own perspectives and experiences to the evaluation process (Helgesson, Höst, & Weyns, 2011).

Type 3 evaluations are the most costly because they involve carrying out assessments in a practical setting to analyse both the investigated process and the maturity model, but will not be looked at in detail in this study. The order in which these evaluations are carried out is typically type 1, followed by type 2, and finally, type 3. However, the evaluations can also be iterated or carried out in any order that is seen as useful. The purpose of these evaluations is to improve the maturity model by evaluating its understandability, internal consistency, and how well it corresponds with the current state of practice. The feedback from these evaluations can be used to update and evaluate the model (Helgesson, Höst, & Weyns, 2011).

A maturity grid is a grid that shows the current level of maturity based on the assessment of a system at its current state of development and is a subset of a maturity model. Maturity models and maturity grids are often thought of as being the same thing, however, differences between the two exist (Maier, Moultrie, & Clarkson, 2012). The first difference between a maturity grid and a maturity model is that maturity grids do not apply to specific processes like software development and acquisition, but rather to enterprises in any industry (Maier, Moultrie, & Clarkson, 2012). Additionally, maturity grids aim to identify the essential traits that are necessary for any process or organization to create and implement high-performance systems. On the other hand, maturity models concentrate on identifying the optimal practices for a specific process or area and measuring an organization's level of maturity based on how many of these practices they have adopted. Secondly, the assessment methods for maturity models can involve Likert or binary (yes/no scale) questionnaires and checklists, while maturity grids use a grid structure to provide levels of readiness based on the current level of performance. The maturity grid approach provides a descriptive text for the characteristic traits of performance at each level, also known as a "behaviourally anchored scale" (Maier, Moultrie, & Clarkson, 2012).

6.1.4 Weighting Methods

Throughout this thesis, several weighting methods for Multi-Criteria Decision Making (MCDM) were encountered (Odu, 2019), but only a few were discussed in this thesis that aligned best with the aim of this study.

Rating is a weighting method where a total score of 100 (or similar) is used, and where each criterion is assigned a number based on its level of importance. The numbers are integers that range from 1 to 100, where the sum of all criteria must be equal to 100. Therefore, the most important criteria receive a score closest to 100 and the least important criteria receive a score closest to 0. The various criteria must also be independent of each other (Odu, 2019).

Table 6-1: Note. Example of rating weighting. From Weighting Methods for Multi-Criteria Decision Making Technique by G. Odu, 2019. Copyright 2019 of G. Odu.

S/N	Criteria	Weights
1	Cost	10
2	Display Resolution	35
3	Battery Life	15
4	Random Access Memory (RAM)	25
5	Internal Storage	15
	Total	100

The ranking weighting method ranks different risk factors based on their level of importance by assigning a numerical rank to each risk factor, using various mathematical formulas. The higher the rank, the greater the importance of the risk factor in the decision-making process (Odu, 2019). Continuing the example of Table 6-1, this list can be converted to a ranking, however, only the implied weights are kept, with the factors simply being organised into a list such as:

$$[Display\ resolution, RAM, Battery\ Life, Internal\ Storage, Cost]$$

Where the criteria on the leftmost are given the higher weights and the ones on the right are given lower. The weights are allocated based on the numbered position within the list (Ewa, 2013). Below you will find examples of various possible formulae for the ranking method as mentioned by (Odu, 2019).
W_j is the weight of the criterion and is n is the number of criteria within a list.

1. Rank sum normalises weight method.

$$W_j\ (RS) = \frac{n - p_j + 1}{\sum_{k=1}^{n} n - p_k + 1}$$

where p_j is the rank of j criterion, $j = 1, 2, \ldots, n$

2. Rank reciprocal weights method.

$$W_j\ (RR) = \frac{\frac{1}{p_j}}{\sum_{k=1}^{n}(\frac{1}{p_j})}$$

Where p_j is the rank of the j criterion, $j = 1,2, \ldots, n$

3. Rank exponent weight method.

$$W_j\ (RE) = \frac{(n - p_j + 1)^p}{\sum_k^n (n - p_k + 1)^p}$$

Where p_j is the rank of the j criterion, $j = 1,2, \dots, n$

Furthermore, among the most popular weighting methods, the Ranking method is the simplest approach to assign weights to criteria. However, the problem with the ranking weighting method is that all the 3 formulas provide the same pre-set outcome, solely based on the number of criteria. The difference in importance does not depend on the importance of the risk factor itself but instead, it purely depends on positioning in the list. For that reason, the ranking weighting method was not considered as it lacked the discernment needed for highly technical risk factors. Furthermore, it is not good to use this technique when dealing with large numbers of criteria due to the increase in the difficulty of achieving a straight rank (Odu, 2019).

The Analytical Hierarchy Process (AHP) is a decision-making method developed by Thomas Saaty in the 1970s that combines mathematics and psychology to assist in complex decision-making. It utilizes the pairwise comparison method to consider all factors when there is no clear best choice. By comparing risk factors against each other, AHP allows for a comprehensive evaluation of their relative importance. A scale ranging from $1/9\ to\ 9$, developed based on semantic differential scaling research, helps assign the level of importance between alternatives. A value of $\frac{1}{9}$ indicates that one factor is significantly less important than the other, while a value of 9 indicates that one factor is significantly more important than the other as shown in Figure 6-1.

Definition	**Index**	**Definition**	**Index**
Equally important	1	Equally important	1/1
Equally or slightly more important	2	Equally or slightly less important	1/2
Slightly more important	3	Slightly less important	1/3
Slightly to much more important	4	Slightly to way less important	1/4
Much more important	5	Way less important	1/5
Much to far more important	6	Way to far less important	1/6
Far more important	7	Far less important	1/7
Far more important to extremely more important	8	Far less important to extremely less important	1/8
Extremely more important	9	Extremely less important	1/9

Figure 6-1: *Note.* Scale used by pairwise comparison weighting method. From "Weighting by pairwise comparison" by GITTA. Copyright n.d. of GITTA.

After all the risk factors have been compared, a square matrix of size N x N is created, where N is the number of factors being compared. The matrix contains all the comparison results and reciprocals. The final weights are then determined using normalization and dot products (Adams, 2017).

The Consistency Ratio is used to test the consistency of pairwise comparisons, with a CR below 0.1 indicating consistency. AHP can be easily integrated into a digital environment, such as Excel, but it can be time-consuming for respondents to compare all risk factors individually. Inconsistency in weight results is a drawback of AHP, which can arise due to subjective decision-making, the complexity of the decision problem, or errors in input data (Isa, Saharudin, Anuar, & Mahad, 2021).

Preference Ranking Organization Method for Enrichment of Evaluation (PROMETHEE) is another Multiple Criteria Decision Making (MCDM) method that was encountered throughout this study and developed by Brans in 1982 (Abedi, Torabi, Norouzi, Hamzeh, & Elyasi, 2012). The goal of PROMETHEE is to partially rank or fully rank alternatives based on the type of PROMETHEE method used. PROMETHEE I is responsible for partially ranking alternatives, while PROMETHEE II is responsible for fully ranking alternatives. Further research was done on PROMETHEE II since it is an efficient method in dealing with different scales and ranking alternatives based on net outranking flow values. However, a downside of PROMETHEE II is that it doesn't provide a clear path to obtaining the weights and requires assigning values with no clear method, thus requiring the decision makers to come up with and to specify the importance of the weighting scale (Isa, Saharudin, Anuar, & Mahad, 2021).

In the study done by (Isa, Saharudin, Anuar, & Mahad, 2021), AHP and PROMETHEE II were combined to eliminate the weaknesses of both methods and at same highlight their strengths. AHP is a strong method for determining the weights of decision criteria based on an importance scale, which is a drawback of the PROMETHEE II method. On the other hand, PROMETHEE II's strengths lie in identifying the optimal alternative based on outranking flow once the weights have been determined by AHP. This reduces the inconsistencies that occur between the AHP weights, by eliminating the uncertainty of the comparison matrix and to reducing the time required for decision-making (Isa, Saharudin, Anuar, & Mahad, 2021).

In another study done by (Berdie, Osaci, Muscalagiu, & Barz, 2017) a similar approach to that of (Isa, Saharudin, Anuar, & Mahad, 2021) was done, where instead of combining PROMETHEE II to AHP, Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) was instead applied. TOPSIS is an MCDM method which was developed by Hwang and Yoon in 1981 (Roszkowska, 2011) and is applied in combination with AHP to determine the final rank of alternatives. TOPSIS evaluates alternatives based on their proximity to the ideal solution and distance from the worst solution. The ideal solution is determined by the best possible values for each criterion when the result approaches 1, while the worst solution is determined by the worst possible values when the results are closest to 0. The TOPSIS method calculates a score for each alternative based on its distance from the ideal solution and its distance from the worst solution and then ranks the alternatives based on these scores. Similarly to the previous study done by (Isa, Saharudin, Anuar, & Mahad, 2021), AHP will be used to calculate the relative importance of alternatives resulting in weights. However, as PROMETHEE II, TOPSIS also comes with its drawbacks. One is that it lacks the process of determining the relative importance of alternatives. Secondly, it doesn't provide a consistency check of decisions made as AHP (Roszkowska, 2011).

Transitivity is integrated into MCDM, which consists of various types of evaluations concerning conflicting criteria, to determine the ordinal consistency of the pairwise comparisons.

Transitivity states that if A is greater than B and B is greater than C than A should also be greater then C. However, in weighting methods, an occurrence of error exists, both systematic and random. The person can respond by saying that C is greater than A when tasked with comparing 2 criteria directly. Furthermore, if the AHP $CR < 0.1$, then the probability of non-transitivity occurring is significantly low (Wu & Tu, 2021). Additionally, the important difference in comparisons is reciprocal, meaning that each combination is only compared once (as comparisons always take the form of “How much more important is A than B?”) instead of requiring the repetition of B being compared to A.

6.1.5 SANOL

When searching for maturity models and 4IR implementations, a research paper with a similar research question to this study was found, which determined the maturity score of risk factors when it came to the 4IR entrepreneurial implementation in Turkey. The SANOL Maturity Model was created to solve key problems that manufacturing businesses have in different countries and industrial sectors. Furthermore, it focuses on the 4IR industry technologies which is the same focus as this research (Cihan Ünal, 2022), with the sole difference in the main research question being that this research specifically focuses on cobots. As per the authors, the steps taken in the development of SANOL were:

1. The Determination of Industry 4.0 Maturity Model dimensions and sub-dimensions;
2. The Determination of Industry 4.0 Maturity Model dimension and sub-dimension weights;
3. The Determination of the analysis levels of the Industry 4.0 Maturity Model;
4. The Determination of the Industry 4.0 maturity index.

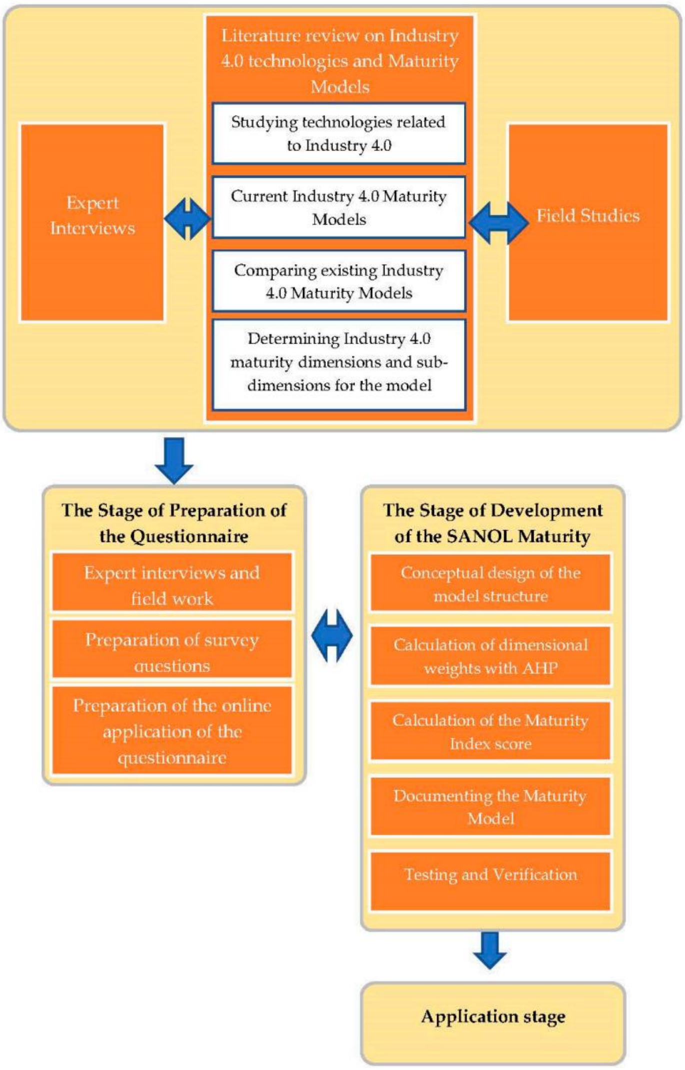

Figure 6-2: *Note.* A multi-methodological research approach for the development of the SANOL Industry 4.0 Maturity Model from "Application of the Maturity Model in Industrial Corporations" by C. Ünal, C. Sungur & H. Yildirim, 2022, Sustainability 14(15), 9478; (https://doi.org/10.3390/su14159478). Copyright 2022 of C. Ünal, C. Sungur & H. Yildirim.

The goal of the SANOL maturity model is alike to that of this research, which was to measure the maturity of digital technological transformations in enterprises, to help them reorganize, restructure, and renew their present skill sets when it came to 4IR technologies. Furthermore, SANOL is a maturity model for Industry 4.0 that determines the level of development of dimensions and risk factors of 4IR technologies. It employs the AHP method to compare two risk factors one at a time and forms an N×N matrix. SANOL maturity model outline starts with a literature review to obtain the necessary information to develop the maturity model. The second stage involves preparing the survey and questionnaires to obtain results from respondents, while the third stage involves developing the maturity model itself. This includes designing the model structure, setting the calculation of dimensional weights through AHP, calculating the Maturity Index score, testing and verifying it, and preparing for the application stage (Cihan Ünal, 2022).

6.1.6 Inter-rater Reliability Analysis

Krippendorff's alpha is a statistical measure frequently employed to assess the reliability and consistency of results obtained from multiple raters or observers. Its main objective is to evaluate the level of agreement among respondents or raters when assessing the same dataset. By calculating Krippendorff's alpha, it becomes possible to identify disagreements or inconsistencies in the responses provided by raters, which can arise from both systematic and random disagreements (Krippendorff, 2011).

By identifying areas of disagreement or inconsistency in the dataset, researchers can take steps to improve the consistency and reliability of the data by resolving any discrepancies and establishing clear coding guidelines (Krippendorff, 2011). In this study, Krippendorff's alpha is utilized to evaluate the level of consensus among respondents. However, it can also be employed to examine the disparities or variations that exist between respondents.

The alpha coefficient ranges from -1 to 1, with -1 indicating complete disagreement, 0 indicating no reliability, and 1 indicating complete agreement (Mancar & Gülleroğlu, 2022). A coefficient of at least 0.7 is generally considered acceptable, although this may vary depending on the context and field of study. Krippendorff's alpha is a versatile measure that can work with various data types, such as nominal, ordinal, and ratio data, and can handle incomplete or missing data. It can also work with more than two respondents (Krippendorff, 2011). However, Krippendorff's alpha also comes with its limitations. One limitation of Krippendorff's alpha is the lack of research on its usage in conjunction with other weighting methods. While it is a reliable measure for evaluating inter-rater or observer agreement, its effectiveness in combination with other methods for weighting criteria in multi-criteria decision-making has not been extensively studied. Secondly, the scoring mechanism or coding scheme used by the users must be reliable, or else it can lead to inaccurate alpha coefficient results. Furthermore, the scoring mechanism of Krippendorff's alpha calculates the extent to which the observed agreement among raters is greater than the agreement that would be expected by chance. It does this by comparing the observed agreement to the agreement that would be expected based on the distribution of ratings for each item as shown in the formula below (Krippendorff, 2011).

$$\alpha = 1 - \frac{D_o}{D_e}$$

Where:

$$\alpha \ = \ Krippendorff's\ alpha\ coefficient$$
$$D_o = \ observed\ disagreement$$
$$D_e \ = \ expected\ disagreement$$

Cohen's kappa model compares the observed agreement to the agreement that would be expected by chance to obtain an inter-coder reliability coefficient. Cohen's Kappa is a specialised approach to measure nominal data, which is data that has no hierarchy or order amongst its categories (with the most common example being colours) (Widmann, 2022). Akin to Cohen's kappa model and Krippendorff's alpha, Fleiss' kappa also measured the inter-rater reliability of the data amongst the respondents. Furthermore, Fleiss' kappa is similar to Cohen's kappa model in that it measures the inter-rater reliability specifically for nominal design (Powers, 2012).

Furthermore, unlike Krippendorff's alpha, Fleiss' kappa doesn't allow scenarios where data is missing, thus making it not as sophisticated as Krippendorff's alpha. Additionally, Krippendorff's alpha focuses on the number of agreements over the other outcomes that were expected by chance, while Fleiss' kappa focuses on the number of agreements and disagreements between raters. Thus, Fleiss' kappa can be seen as a similar model to Cohen's kappa, where they are both considered a model that measures the degree of agreement among raters, whereas Krippendorff's alpha is considered a method to measure the reliability or consistency of the coding mechanism itself (Powers, 2012).

6.1.7 Lasrado's Utility Curve

During the design phase of the thesis, it was found and discussed that maturity is not always progressing linearly, and it must be taken into consideration. This was largely prompted by Figure 6- which was found in a publication concerning maturity stages and stage boundary conditions (Lasrado R. K., 2016):

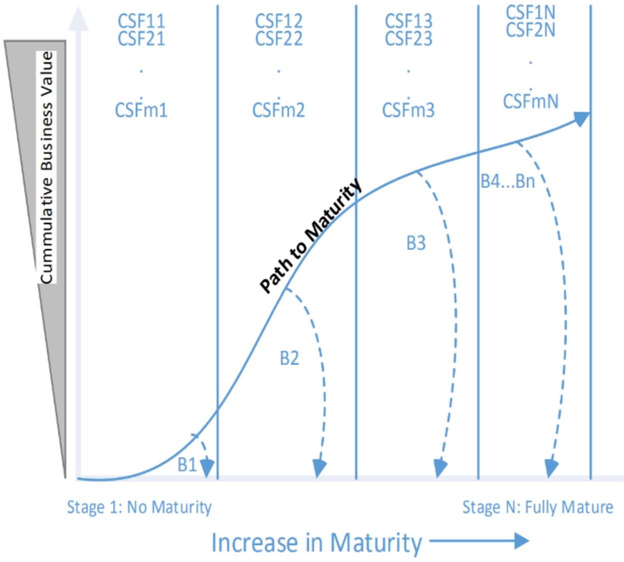

Figure 6-3: *Note.* Critical success factors (CSF) and boundary conditions in maturity models. From "A Methodological Demonstration of Set-Theoretical Approach to Social Media Maturity Models Using Necessary Condition Analysis" by L. Lasrado, R. Vatrapu & K. Andersen, 2016, PACIS 2016 Proceedings. 249. (https://aisel.aisnet.org/pacis2016/249). Copyright 2016 of L. Lasrado, R. Vatrapu & K. Andersen.

However, despite an intensive search for mentions of non-linearity similar to "Path to Maturity", no related studies were found. Due to this, Lasrado was personally reached out to, for further elaboration to establish the basis of the abovementioned figure. He has kindly provided the reference of his previous work of a literature review on maturity models (Lasrado, Vatrapu, & Andersen, 2015) and suggested the reading of more recent publications in the field (Wulf, 2020) and (Bley, Pappas, & Strahringer, 2021). Most importantly, he has brought up the original criticism of the usage of linear maturity in maturity models, while it is acknowledged that is not the best representative of real life that dated back to (King & Kraemer, 1984), which contains Figure 6-4.

As the Path of Maturity curve did not have a name, it was given one of the “Utility Curve” that the authors feel best represents the function representing the connection between the value by the implementation and its maturity level. An example of this would be legal regulations, which provided little, or even negative value if unfamiliar with (Significantly low level of maturity), and quickly grow in value as more of them are understood and implemented to prevent legal and technical risks (Rapid increase post implementation of basics).

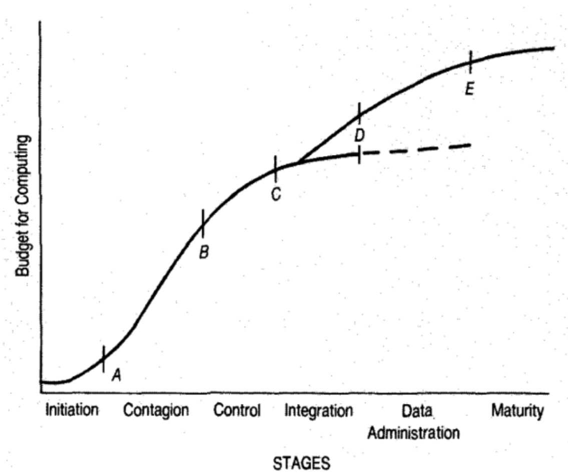

Figure 6-4: *Note.* Budget Curve and Nolan's Stage Theory. From “Evolution and organizational information systems: an assessment of Nolan's stage model” by J. King & K. Kraemer, 1984, Communications of the ACM, Volume 27, Issue 5 (https://doi.org/10.1145/358189.358074). Copyright 1984 of J. King & K. Kraemer.

Upon an in-depth look at the topic and more recent research provided, it was found that it is largely centralised on academic post-doctorate managerial studies, and it is difficult to apply to an engineering master’s thesis, without it rapidly becoming the central topic of the study. As mentioned previously, the topic of non-linearity never takes a larger role during the studies, as the papers shown give a larger priority to the boundaries in-between different stages of maturity and the criteria to find them. Furthermore, it is never implemented in maturity grids and is solely a descriptor for larger maturity concepts, such as maturity models.

However, it is still integral to a full understanding of maturity and its impact on the choices of respondents and thus is not present in the developed maturity model. But it is included in the design of the data collection that is discussed later in the survey for validation of the difference between the importance and utility of risk factors.

7 METHODS & MATERIALS

7.1 Introduction to design of method

In this thesis, a maturity model is developed to enhance safety awareness and trust towards cobots. The model focuses on assessing risk factors related to technology, human interaction, collaborative workspace, enterprise, and external factors. These risk factors have been previously defined by (Berx, Decré, Morag, Chemweno, & Pintelon, 2022) in the Cobot Safety Readiness Assessment Tool (CSRAT). By identifying areas of low maturity, improvements can be made to enhance safety awareness and trust between cobots and humans, thereby promoting the wider implementation of cobots in modern industries.

The research incorporates a maturity grid and a mathematical model. The maturity grid assigns readiness levels to cobots within an enterprise on a scale from 1 to 4, with 4 indicating the highest level of readiness (Berx, Decré, Morag, Chemweno, & Pintelon, 2022). The mathematical model utilizes the AHP to determine the most crucial risk factors. AHP involves pairwise comparisons using a scale of $\frac{1}{9}$ to 9, with weights obtained from the respondents' opinions. This process helps identify the most important risk.

To oversee consensus between respondents, Krippendorff's alpha, an inter-rater reliability measure, is used to assess agreement among the AHP weights. This allows for the possibility of creating a grouped weighted model instead of individual models for each respondent.

To streamline data collection and processing, a Python program hosted on GitHub automates the collection and calculation of results, generating reports in an Excel-compatible format.

7.2 Design of method

7.2.1 Scenario

As cobots are an extremely diverse field, constraints were put in place to ensure consistent results. In the case of the study, participants are tasked with a cobot implementation that they had experience using and maintaining in administrative positions.

In the beginning of a study, a definition of cobots and multiple visual aids are provided, such a Figure 7-1, as respondents were subconsciously pushed towards the industrial implementation and associated values.

Figure 7-1. *Note*. Cobots collaborate with humans and require no safety cage unlike robots. From “Cobots and Collaborative Robots” by “Editorial by Industrial Quick Search”. Copyright n.d. of Editorial Quick Search.

7.2.2 Survey and Synthetic Data

A variety of survey platforms were considered, such as Google Forms and SurveyMonkey, as they are the ones that most respondents are familiar with and have completed below previously. However, as CSRAT has already used Excel and contained a large amount of information for each definition, it was decided that the above is not feasible due to the necessity of all information being present on a single monitor, which the online forms are unable to provide due to their user-experience oriented design.

Microsoft Excel (.xlsx file) was decided to be a medium for the survey, as to continue the expansion of CSRAT. Due to the additional expectation that experts would be completing the survey, it was also desired to provide immediate calculated results, such as weights and consistencies, that are rather complex to ensure their interest and concentration by providing objective values as a reward at the end of the questionnaire.

CSRAT survey was provided with pre-split worksheets, with each one being responsible for each basis. A final worksheet is created and used to provide immediate results post-completion that would be of interest, with more processed results returning sometime after the completion of the study. The worksheets are numbered and expected to be completed from left to right. The participants are further not expected to correct their previous answers based on new information provided for them.

To initialise a survey, a set of instructions was to be created. Initially, as the academic and industry experts (with a few years of experience) were the only ones expected to be completing the survey, they were to be provided with extensive instructions due to a high understanding of the topic. The attachments were expected to be:

1. Detailed explanations of risk factors;
2. Start command to start a stopwatch;
3. Detailed instructions for the survey;
4. Privacy policy (GDPR) of the survey and study.

In the final version of the survey, only privacy was kept within e-mails forwarded to participants and detailed instructions.
The final outline is as follows:

1. Attachments
2. Technology
3. Human
4. Collaborative Workspace
5. Enterprise
6. External
7. Dimension groups
8. Inter-dimensional AHP
9. Utility questions

However, due to the difficulty of finding respondents, all attachments except instructions have been removed to ensure easier understanding and an attractive survey that they would be interesting to take.

Synthetic data was considered for replacement of experts due to initial results showing the difficulty of survey and lack of response from the experts. But due to high complexity of data and transitive properties that are expected during pairwise comparisons (the consistency ratios), there is no previously completed research on the topic. To create the simulation of experts' responses, it is expected to require another in-depth study. Additionally, it was wanted to use real life experts who could provide expertise in the topic of cobots and it was concluded that simplifying the survey and improving recruitment for the topic would provide sufficient results to prove the mathematical model.

7.2.3 Maturity level of risk factors

Given the original CSRAT, the usage is largely kept the same to continue Nicole Berx's research to expand on her model. The change included consist of adding colours to the Excel whenever a level of preparedness is chosen, to ensure increased interaction in the face of a large amount of text. Additionally, drop-down menus were already implemented in CSRAT for user experience. In the final draft, it was chosen to decrease the level of detail and characterisation for the levels of risk factors and remove official definitions due to the difficulty of finding respondents and complaints of high difficulty in combination with a large amount of reading necessary. Each dimension is contained within its worksheet and the order is unchanged from CSRAT.

It is essential to select a level from 1 to 4 to show the preparedness of the correlated risk factor in the application of respondents (Figure 7-2). It shows the preparedness and understanding of the sub-dimension and its effects for a given implementation.

WORKSHEET: COLLABORATIVE WORKSPACE READINESS						
Risk factor class		Level 1. Uninformed	Level 2. Elementary	Level 3. Basic	Level 4. Advanced	Level nr. (from 1 to 4)
Sub-class	***Sub-sub class***	***Unaware of the need to identify cobot risk factors.***	***Aware of the need to identify risk factors for cobots.***	***Required Safety Knowledge & Tools are present, but not adapted for cobots.***	***Required Safety Knowledge & Tools are integrated in Safety Management procedures and are adapted for cobots.***	the level nr for the cell that best describes your situation
Physical design	Access and clearance	Uninformed of safety risks	Elementary awareness of safety risks	High awareness of safety risks Low safety implementation	High awareness of safety risks Complete tool safety implementation	3
	Layout	Uninformed of safety risks	Elementary awareness of safety risks	High awareness of safety risks Low safety implementation	High awareness of safety risks Complete tool safety implementation	3
	Hazardous obstacles/ objects	Uninformed of safety risks	Elementary awareness of safety risks	High awareness of safety risks Low safety implementation	High awareness of safety risks Complete tool safety implementation	3
Maintenance	Hazardous obstacles/ objects	Uninformed of safety risks	Elementary awareness of safety risks	High awareness of safety risks Low safety implementation	High awareness of safety risks Complete tool safety implementation	2

Figure 7-2: *Note.* Implementation of CSRAT in the collaborative workspace dimension after respondent assigned level of readiness of risk factors. Adapted from "Assessing System-Wide Safety Readiness for Successful Human–Robot Collaboration Adoption" by N. Berx, A. Adriaensen, W. Decré, & L. Pintelon, 2022, Safety 8(3), (https://doi.org/10.3390/safety8030048). Copyright 2022 of N. Berx, A. Adriaensen, W. Decré, & L. Pintelon.

7.2.4 Intra-Dimensional & Dimension Groups Comparisons

A variety of weighting methods, as described in 6.1.4 Weighting Methods are considered. The ranking was discontinued from possible selection almost immediately due to its low adaptability for groups where there is a large disparity in the importance of risk factors.
The rating was considered in more detail, however, due to difficulties associated with directly assigning weights to concepts, it is much easier to compare the differences in value between them, providing better results. Additionally, as it was expected that the conceptual usage is the first time the respondents would see it, they would not have associated the importance of each sub-dimension in a quantifiable way in their mind.
Pairwise can be seen as largely the most optimal solution for the associated CSRAT weighting method, however, there was an additional need for simplification due to the formula stipulating the number of necessary comparisons, which is quadratic polynomial, commonly written as $ax^2 + bx + c$.

$$\text{№ } of\ comparisons = \frac{n(n-1)}{2}$$

While the "ideal" situation would compare all the risk factors to each other, according to the formula that would result in 253 comparisons. For that reason, the risk factors are split into 5 groups, as in the CSRAT according to the dimension and then used to calculate the weights within the subgroups (Table 7-1). Whereas it is possible to further decrease the number of comparisons by distributing the risk factors evenly newly designed groups. An example of this would be Enterprise and External groups, where Enterprise contains 3.5x more risk factors, however due to quadratic property requires 21x more comparisons. As such, if there are more groups where the maximum number of sub-dimensions is decreased, it would lead to smaller total number of comparisons needed. An example would be splitting larger dimensions such as Technology and Enterprise into smaller groups that are more specific to their target, however that would require redefining CSRAT which is largely avoided in this paper.

In the case of this study, it was as follows:

Table 7-1: *Note.* Number of comparisons based on the number of sub-dimensions within group.

Dimension	Number of sub-dimensions	Number of comparisons
Technology	6	15
Human	4	6
Collaborative workspace	4	6
Enterprise	7	21
External	2	1
Sum	23	49

Pairwise comparison		
How important is :		**Compared to:**
Access and clearance	1	Layout
Access and clearance	1	Hazardous obstacles/ objects (Physical Design)
Access and clearance	1	Hazardous obstacles/ objects (Maintenance)
Layout	1	Hazardous obstacles/ objects (Physical Design)
Layout	3	Hazardous obstacles/ objects (Maintenance)
Hazardous obstacles/ objects (Physical Design)	2	Hazardous obstacles/ objects (Maintenance)

Figure 7-3: *Note.* Intra-dimensional comparisons in the collaborative workspace dimension through pairwise comparison. From final implementation of survey.

AHP was largely selected due to its previous usage in SANOL (Cihan Ünal, 2022), which had a large influence on this paper due to its high similarity and the higher reliability and cross-reference of methods that have been used before. Additionally, AHP is the best-known and researched MCDM structure, especially with a combination of pairwise comparisons. (Khaira & Dwivedi, 2018)

Intra-dimensional comparisons are completed in the AHP system, directly below the selection of levels within the survey via dropdown menus, to ensure recent thinking about the risk factors (Figure 7-3). An aid is provided for ease of use on every worksheet (Figure 6-1). Additional design choices, such as conditional formatting are also used to increase interactivity and maintain interest in the survey. An example of this would be colours filling the appropriate levels or pairwise comparisons cell turning green when a choice is selected.

A comparison is completed given a comparison grid based on the research (Adams, 2017). $\frac{1}{9}$ to 9 is chosen based on literature review describing possible pairwise comparisons scales (Rosenberg & Navarro, 2018). Furthermore, the same procedure is then used for groups of dimensions (Figure 7-4). This is then followed by application of Krippendorff's alpha further described in 8.2.3 Krippendorff's Alpha.

Dimensions		
Pairwise comparison		
How important is this dimension		**Compared to this dimension**
Technology	1.000	Human
Technology	3.000	Collaborative Work Space
Technology	2.000	Enterprise
Technology	1.000	External
Human	1.000	Collaborative Work Space
Human	3.000	Enterprise
Human	2.000	External
Collaborative Work Space	2.000	Enterprise
Collaborative Work Space	2.000	External
Enterprise	1.000	External

Figure 7-4: *Note.* Group dimensional comparisons of all 5 dimensions via using pairwise comparison. From final implementation of survey

The weights calculated for the risk factors can be seen as:

$$W_{risk\ factor} = W_{dimension} * W_{risk\ factor\ in\ dimension\ group}$$

Where W is the weight generated by the AHP for each factor. The implementation of weights calculation from AHP has used the approximation of eigenvectors based on (Bunruamkaew, 2012). However, as the weights were initially calculated in Excel, which is incapable of measuring the eigenvectors, an approximation was used, which is accurate to 3 significant figures, verified via (Takahagi, s.d.). Due to high accuracy, the substitution was used in development of results. This is shown in more detail in the Calculation of weights & Consistency .

7.2.5 Inter-dimensional comparisons

The inter-dimensional comparisons are created for the purpose of verification of the mathematical model and adjustment of weights. Furthermore, the inter-dimensional comparisons verify whether the system developed maintains the transitive properties. The comparisons use the AHP in the same manner as in Intra-Dimensional & Dimension Groups Comparisons.

The necessary number of comparisons is calculated as shown:

For even total number of risk factors:

$$\text{№}\ of\ \text{AHP}\ comparions = \frac{n}{2}$$

For odd total number of risk factors:

$$\text{№}\ of\ \text{AHP}\ comparisons = \frac{n+1}{2}$$

As there are 23 risk factors, the formula used would be for odd totals:

$$\frac{n+1}{2} = \frac{23+1}{2} = 12\ comparison$$

As there is an odd number, it is additionally necessary to compare a single risk factor twice, where in this case it was “Technology - Application specific hazards”, as shown in Figure 7-5. The comparisons have been assigned randomly to prevent authors’ bias from influencing the combination. Additionally, a system that would prevent difficult combinations of comparisons was considered, however again, bias from authors was avoided, as it was not possible to have additional surveys completed concerning the inter-dimensional comparisons.

	Sub-sub dimension comparison		
	How important is :		**Compared to:**
1	Technology-Tooling	1	Human - Work stress
2	Human - Cognitive workload	0.5	Workspace - Hazardous obstacles (Maintenance)
3	Human - Physical workload	1	Enterprise - Tech/digitization strategy
4	Enterprise-Social acceptance	2	Technology - Data management
5	Technology - Application specific hazards	0.33333333	Enterprise - AI decision making
6	Workspace - Access and clearance	1	External - Environmental conditions
7	Enterprise - Privacy	0.33333333	Technology - Control software
8	Enterprise - Training	2	Technology - Hardware
9	Workspace - Hazardous obstacles/ objects (Physical design)	1	Human - Trust
10	Enterprise - Work design	1	Technology - Application specific hazards
11	External - Regulatory	2	Workspace - Layout
12	Enterprise - Safety strategy	1	Technology - Security

Figure 7-5: *Note.* Interdimensional comparisons to verify if comparisons are transitive. From final implementation of survey

The adjustments of weights are necessary to verify and improve the simplified model that we use in this study. It increases proves the validity of the model and adjusts it to be more consistently when combining all different dimensions into a singular final result.

7.2.6 Weights of risk factors

The weights are separately calculated within each group of dimensions, and each dimension has a weight. Based on these factors, it is possible to estimate the weight (importance) of each risk factor (sub-dimension) within the maturity model via:

$$W_{risk\ factor} = W_{dimension} * W_{risk\ factor\ within\ dimension}$$

Where W is representative of weight calculated within AHP and n representative of number of risk factors within a AHP comparison.

$$\sum_{i=1}^{\mathrm{n}} W_i \ = 1$$

This is valid for groups of dimensions, such as “Human”, “Technology” etc. and groups of sub-dimensions, such as “Technology” dimension containing “Control software” and “Hardware”.

Inter-dimensional weights are compared from calculated weights to respondents' inputs, followed by an adjustment in the form of linear scaling, where an in-depth calculation can be found in Calculation of adjusted weights, that takes place later in the study.

7.2.7 Utility Perception of risk factors at varying maturity

Pertaining to the research completed in literature review and the sequential research sub-question, implementation of data collection concerning utility is integrated.

To implement a utility component, it is decided to implement an additional questionnaire, containing questions about the importance of each dimension at different levels of preparedness. Results similar to Figure 7-6 were expected, with usability of risk factors drastically rising in at a specific maturity level.

The theory behind it is shown in and is based on Lasrado's research (Lasrado R. K., 2016). The utility level has high implications for the weights, as the utility can be modelled as non-linear depending on the level of implementation (Figure 7-6). As such, the maturity index can additionally be adjusted depending on the considerations for utility, however, this was not implemented in this study's maturity model and is only used for discussion.

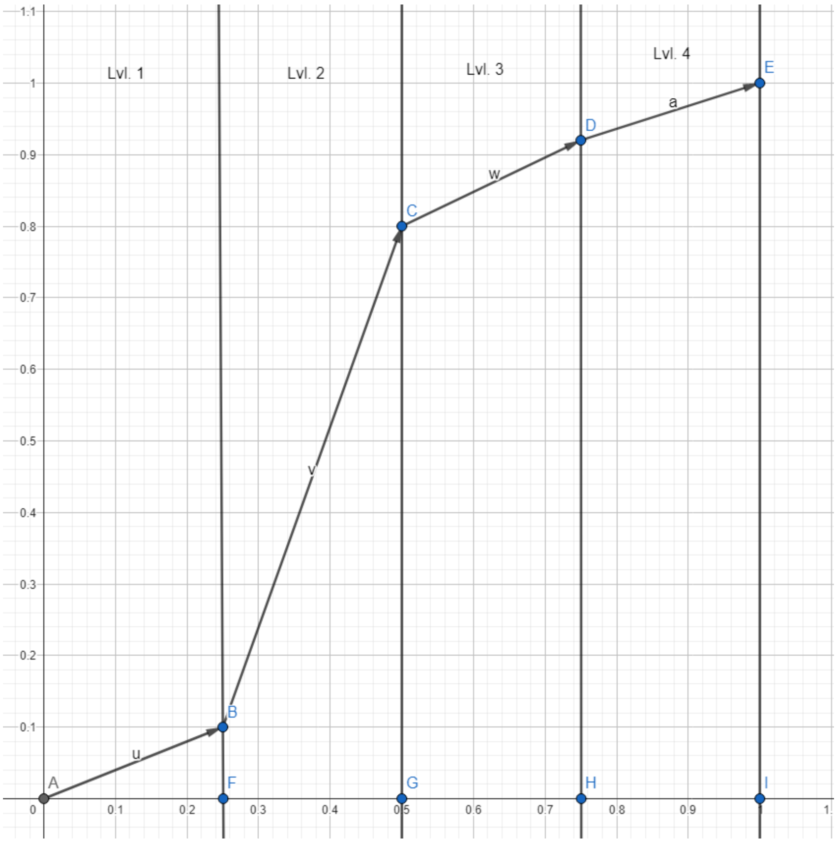

Figure 7-6: *Note.* Value-Maturity chart containing Lasrado's utility curve of a criteria, with value experiencing a large jump due to initial implementation steps.

The implementation consisted of assuming a utility value of 0 at the origin (A) and a value 1 at the rightmost part (E). The values asked of respondents were concerning Early (B), Middle (C) and Late (D) stages.

WORKSHEET: Usefulness progression	
For each dimension below, please reply on a scale of 1 to 10 your perceived usefulness at that stage of implementation of that dimension	
Early stage	
On average, how useful are this dimension's subclasses if you only know their risks?	
Technology	7
Human	8
Workspace	6
Enterprise	8
External	8
Middle stage	
On average, how useful are this dimension's subclasses if you know the risks and the tools, but theyare not yet tailored for cobots?	
Technology	7
Human	8
Workspace	9
Enterprise	7
External	7
Late stage	
On average and for the effort that expended, how useful is complete integration of risk factor identification tools tailored for cobots for each subclass?	
Technology	9
Human	7
Workspace	8
Enterprise	6
External	8

Figure 7-7: *Note.* Utility of the 5 dimensions at different stages to establish a utility curve. From final implementation of survey

7.3 Data collection design iterations

The design of the Excel maturity model has gone through 3 design iterations, with only initially 2 being planned originally. The first iteration was to obtain feedback from the respondents, in order to determine the average time of completion and was used as proof of concept. The second iteration was to be the final Excel maturity model to be sent out to the experts whereas the third was necessary to obtain results for the study upon failure of the second version.

7.3.1 First iteration

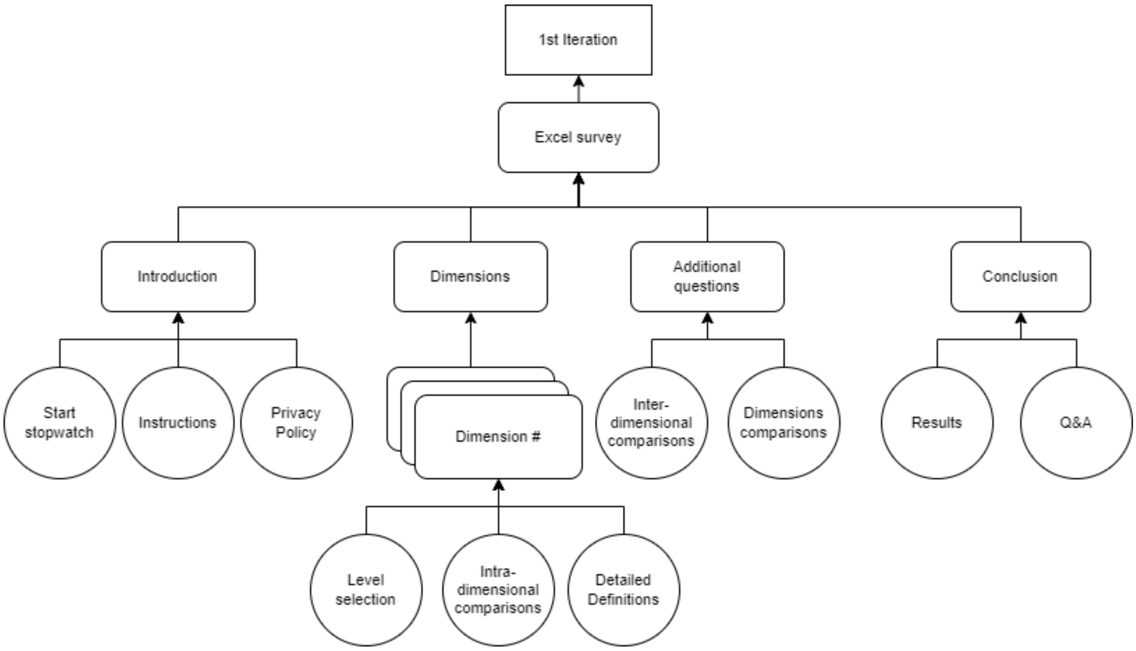

Figure 7-8: *Note.* First design iteration of the Excel survey.

A first draft was designed to test the results of a group of engineering students. This iteration can be considered as proof of concept, with a Q&A section after the results, to obtain feedback on what must be improved within the draft. This design also included stopwatch timing, as it was desired to know how long it would take for students to complete, and as such predict how long it would take the experts for the survey to not take too long for a larger number of responses (Figure 7-8).

7.3.2 Second iteration

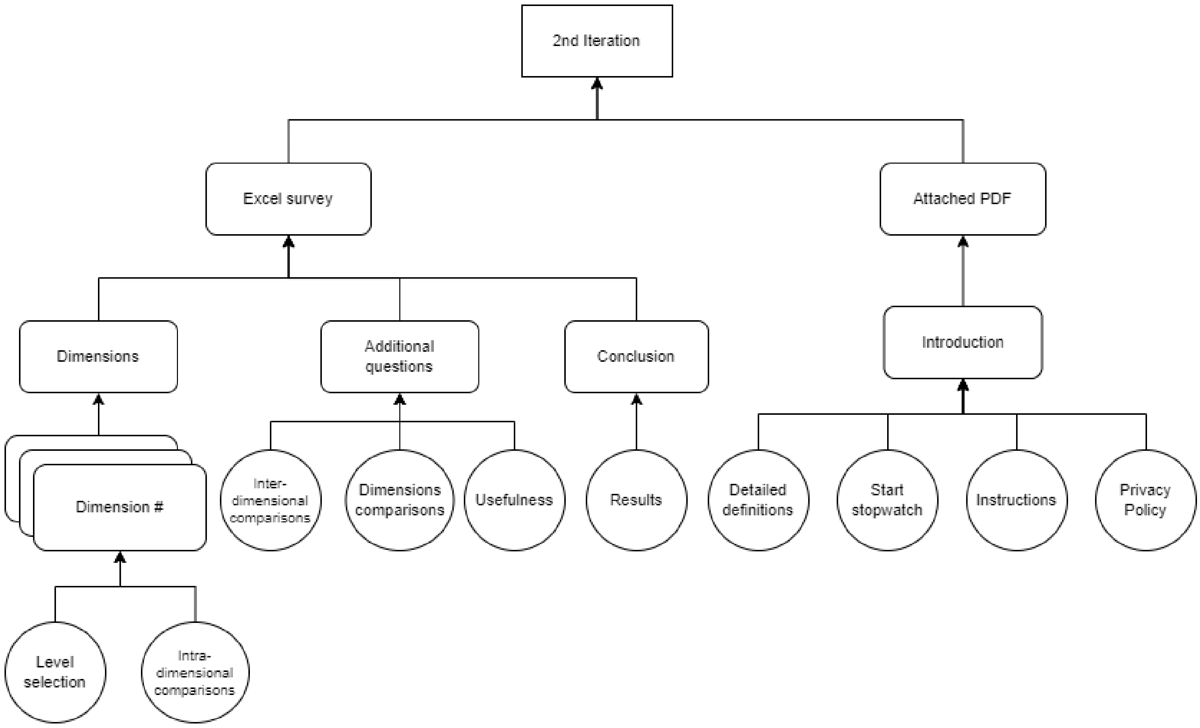

Figure 7-9: *Note.* Second iteration of Excel.

After receiving and digesting all the feedback from the students, a more refined version of the questionnaire was made. A detailed pdf document containing in-depth details such as definitions and instructions was created (largely from the previously used worksheets) to be optional reading, or as a reference guide if the respondent felt it difficult to understand the meaning of a risk factor. Q&A was also removed, and results were made easier to understand (Figure 7-9). At this stage, the additional promotion has been completed, as a month has passed with 0 responses from academic and industrial experts alike. Additionally, at this stage inter-dimensional comparisons were changed from proof of concept to fully functional system, with utility additionally becoming a part of the survey.

7.3.3 Final (Third) iteration

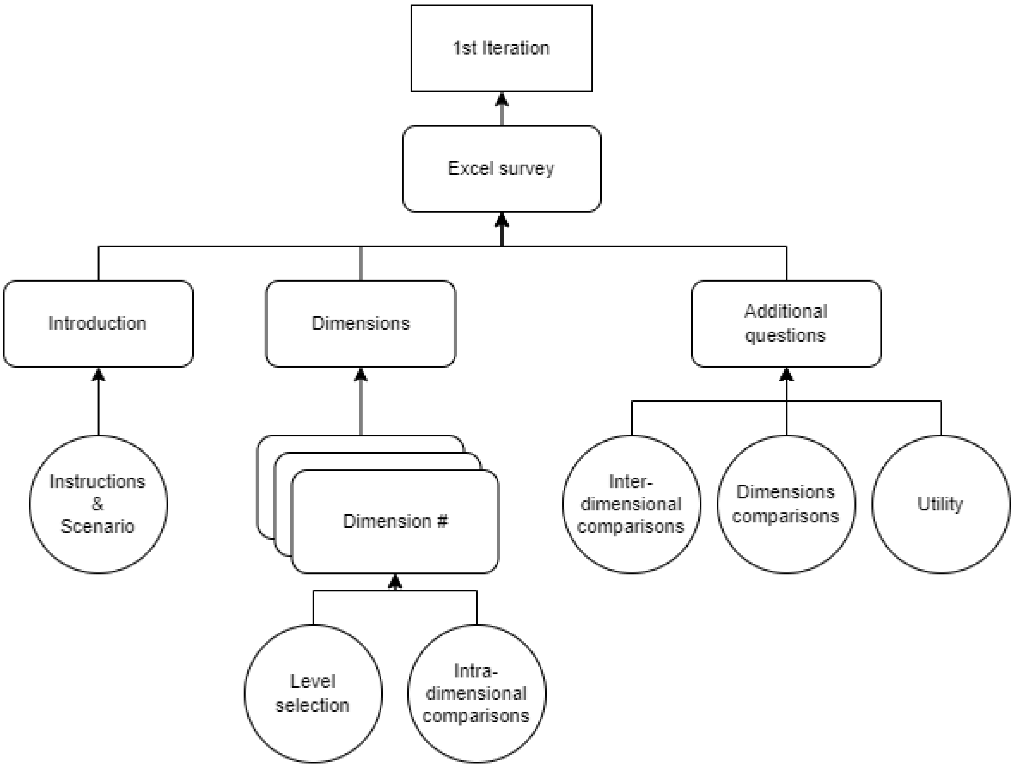

Figure 7-10: *Note.* Third iteration of Excel survey - a simplified version of Excel survey.

A much more simplified and compact version was made towards the end of the study, with the definitions and stopwatch being removed, to decrease the stress and burden on the responder, and to prevent the feeling of being overwhelmed. The respondent profile requirements were additionally reduced to increase the number of participants, as it was decided that usage of synthetic data is highly impractical and would take another master thesis to be done. At this stage, additional considerations went into the design of the survey to improve the user experience (Figure 7-10).

7.3.4 Participant Profile, Recruitment & Synthetic data

At the beginning of the research, after it was known that a survey is necessary, it was needed to decide the target audience of the survey.

For the first draft, used as a proof of concept, a group of engineering students and recent graduates from KU Leuven Group T were reached out to, who then completed the survey which was seen as a validation step in this research, providing an approximate time of completion and leaving qualitative feedback (n=4) in addition to expected results.

For the (second) expected final draft, professors, and academic experts in the engineering faculties of KU Leuven were reached out, those of Group T and the Mechanical Engineering Faculty. Nicole Berx has assisted with participant recruitment via LinkedIn and oral information dissemination about the survey. A recruitment campaign has been launched with the assistance of Ing. Kaat Van Assche, who is a member of the research team in Medical Instrumentation in the abovementioned Mechanical Engineering Faculty. She then reached out to her department and promoted the survey to her colleagues to help obtain more results. The positive encouragement consisted of free pizzas for the department, delivered a month after the initial survey send out. However, no responses were recorded, (n=0).

The third draft expanded the search group to engineering students once again to increase the number of participants. The participant requirements were also lowered complimented by a survey change, lowering completion time and facilitating understanding of maturity model. The participants from the first draft were contacted to complete 2 additional sets of questions due to the high difficulty of finding new respondents, with additional 5 participants found. This brought the total number of people that have successfully completed the survey to 9.

7.4 Automation

Throughout the development of the tool, the results were manually tested multiple times for all components.

7.4.1 Selection of automation tool

Upon initial research and discussion of respondent profiles, an expectation of much higher respondent turnout was expected, with tens of experts participating in the survey.

Initial automation via Excel has been considered and the weight calculation with all its attached attributes was possible in an additional worksheet of Excel, as they are independent of other respondents. However, for Krippendorff's alpha, a complex calculation would have to be made for each new sample of respondents, as it scales varies with the number of respondents as shown in Figure 8-3: Note. Respondents and their table of agreement. Agreement table is used for nominal alpha and is not used in this study. , where each person is represented by a column. Additionally, if results were expected not to be seen by participants and thus cannot be present in an Excel file, it would mean that all the data would have to be compiled by hand every time a new person is added or removed.

The authors approximate that it takes 20 minutes for each survey, with Krippendorff's alpha being of additional concern due to high complexity and necessity for manual adjustment, possibly increasing the time to 30 minutes. In comparison, a program was expected to complete the task and compile all results in seconds, as well as present data in a reportable manner. This has thus spawned a sub-research question concerning possible automation and optimisation of processing time.

	N	O	P	Q	R	S	T	U
1	Weighted Agreement Table							
2								
3		1	2	3	4		r	p
4	1	1	1	1	1		4	0
5	2	0	3	1	0		4	0.470588
6	3	3	2	0	0		5	0.470588
7	4	0	1	1	3		5	0.352941
8	5	0	0	3	1		4	0.470588
9	6	2	1	0	0		3	0.235294
10	7	0	0	3	1		4	0.470588
11	8	0	0	2	3		5	0.470588
12								
13	Weights						n	8
14		1	2	3	4		q	4
15	1	1	0	0	0		r-mean	4.25
16	2	0	1	0	0		epsilon	0.029412
17	3	0	0	1	0		pa	0.386246
18	4	0	0	0	1		pe	0.261246
19							kalpha	0.169204
20	π	0.176471	0.235294	0.323529	0.264706			
21								
22	π*	0.176471	0.235294	0.323529	0.264706			

Figure 7-11: *Note.* Single dimension - nominal Krippendorff's alpha calculation in Excel. From Krippendorff's Alpha Basic Concept by C. Zaiontz, n.d., Real Statistics Using Excel. Copyright n.d. of C. Zaiontz.

As such, a program was decided to be developed to increase the processing speed. As the authors' knowledge of programming is of note when using Python, it was decided to collaborate on it via GitHub and develop it in tandem with the survey. Other variations of programs were not considered due to the high specificity of the requirements, and the authors' desire to provide a presentable and flexible implementation of a complex program.

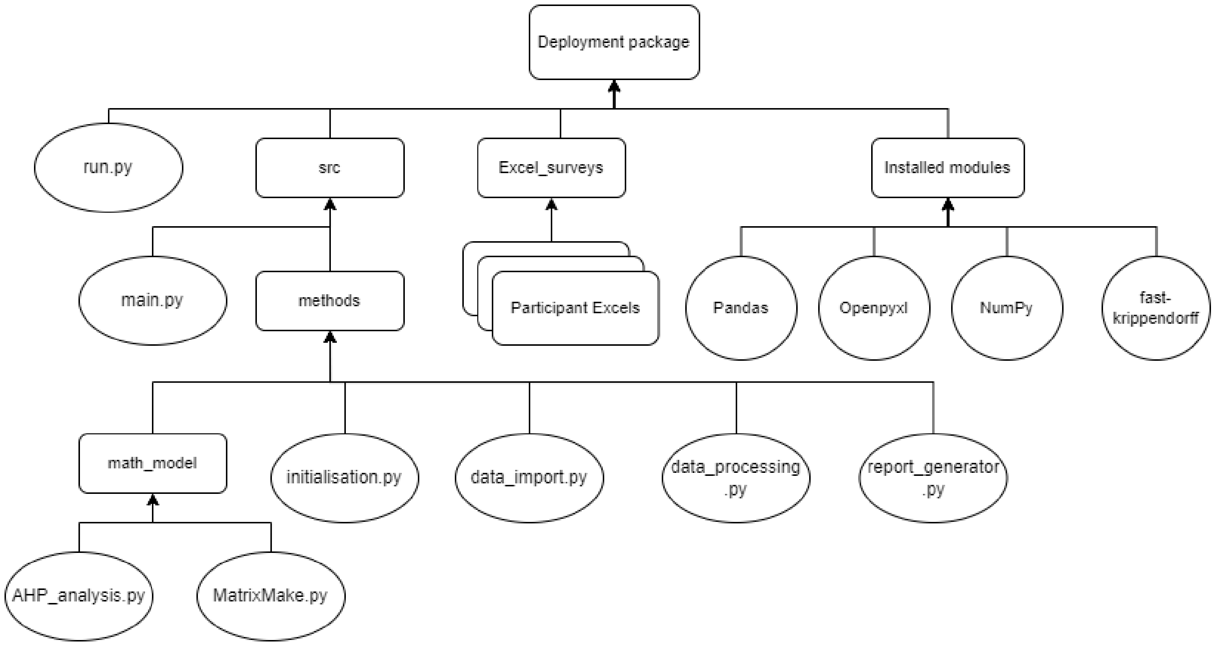

Figure 7-12: *Note.* Deployment package diagram, without miscellaneous files

7.4.2 Initialisation

The initialisation of the program occurs with the retrieval of a configuration file, that is designed to be highly customisable to allow possible integration into other projects. The program further verifies that the input folder exists and that it contains the Excel files inside.
To prevent copies from accidentally appearing in the files scanned, a function exists to check that the files are indeed unique, as it is calculated via SHA256 hash code to prevent duplicates. If there is a file that occurs twice, it is noted in the runtime log file. At the end of initialisation, the program ensures GDPR compliance by changing all names that are not yet in the correct format to the format of XXXX-XXXX, where X can be either a capital letter or a number, which ensures the uniqueness of over files.

7.4.3 Data intake

The process is followed by automatically getting worksheet names based on the configuration file and whether any introduction worksheets need to be skipped. The data is then automatically extracted from all the Excel files located in the input folder. The data is stored in data frame format, which is the primary storage container for the Pandas package. As there are multiple worksheets in each file, they are all stored in unique lists and passed further down the program.

7.4.4 Data processing

The following is done in this section:

- Calculation of selected Krippendorff's alpha;
- Simplification of groups into smaller subgroups and application of Krippendorff's alpha to them;
- Calculation of weights and consistencies for original results for all comparison groups;
- Calculation of final weights.

7.4.5 Report generation

The following reports are generated to be usable with any spreadsheet program, as it is exported in .csv format. The headers for the converted file names are included in .csv(s) for ease of use and cross-referencing during analysis. The specific reports are placed into a newly created folder, which is named with ISO 8601 standard (Wolf, 1997) in the format of "YYYY-MM-DDTHH_MM_SS – n participant(s)".

Inputs:

- Levels selected;
- Interdimensional comparison inputs;
- Dimensional comparisons inputs;
- Inputs used for Krippendorff calculation;
- Simplified inputs used for Simple Krippendorff calculation.

Outputs:

- Consistency ratio calculated for each dimension;
- Weights of each sub-dimension in each dimension;
- Weights of combined sub-dimensions based on the transitivity model;
- Krippendorff's alpha results for all groups.

7.4.6 Continuous Integration via GitHub Actions

A step further is taken to ensure continuous working development. A variety of Pytests were developed for the initial part of the program for further additions to not prevent other parts from working. This is accomplished via GitHub Actions, where additional setup is needed as described in .github/workflows folder. Upon every update (commit) of the project, tests are executed at GitHub servers without any additional steps necessary from the user, where the results of the tests can be found here. As this is not indicative of performance, it was solely used for assistance of development and no testing or results are being used in the study.

7.5 Overview of program characteristics

To see the optimisation of the program, it is necessary to measure its performance. The program is designed with big O notation in mind (Chivers, 2015), to ensure rapid processing and assuredness of short processing time.

The CPU usage as a statistical measure has been considered, but due to optimisation and threading of modern CPUs, CPU usage sampling is not a viable option due to the continuous measurement being impossible, as the process completes tasks in bursts.

The testing is completed with various numbers of Excel files, where copies are created from the participants' files.

7.5.1 Storage sizing

The storage taken is highly variable depending on the aspects considered. It is necessary to define first what size means in the context of the storage taken on the system. The file system uses varying cluster sizes, meaning the smallest amount of disk space that can be allocated to store a file. File system areas systems such as New Technology File System (NTFS), which is most known as Windows's current proprietary file system, have a relatively large average overhead of 4KB in comparison to File Allocation Table (FAT32), which is an older implementation of the same system, which can range from 512 bytes to 32KB. This means that the ideal space taken by the program is different from the space on the disk that it occupies.

The program also comes with a variety of tests, example files and additional files necessary for GitHub test validation, meaning that pure minimum to run the program is different from a standard download of the files.

There are further additional Python Package Index (PyPi) packages, which are collections of code and documentation which can be installed directly within the program without opening any web browsers that are necessary for the program to run. The code repositories will be downloaded and measured directly one by one.

Table 7-2: *Note.* Modules necessary for the program

Module	Implementation
Pytest	Necessary to run continuous integration tests.
Pandas	Necessary for data storage as “Dataframe” class
Openpyxl	Necessary for accessing data within Excel files
NumPy	Most popular python library for completing numerical calculations
Krippendorff	A module necessary for Krippendorff's alpha calculation (Castro & Santiago, 2017)

In Storage sizes, the size of build distributions can be found, which is retrieved from the official module installation website [PyPi](). Further instructions for installation are contained within [this GitHub link](), in the README.md file. Due to the necessity of proficiency in computer usage for installation, the user is expected to be either knowledgeable and/or inquisitive enough to accomplish the installation or to be guided by an author/expert.

7.5.2 Memory usage

A possible test is to ensure that the program complies with the most common computers in the field and gives minimum requirements to run. The program used is “memory_profiler”, which stops the program every 0.1 seconds to measure the usage of memory (RAM) for the entire duration of the program. It requires a PyPi package that is necessary to be installed[1].

The run.py is changed to

[1] Such as with “pip install memory_profiler” in terminal console of a working program.

```
from src import main
from memory_profiler import profile

@profile
def run():
    main.main()

if __name__ == "__main__":
    run()
```

The following commands are then executed in following order.

```
mprof run run.py
mprof plot
```

Where ”mprof” is short for “memory profile”, which is a module installed for this task. The first commands execute the sampling whereas the second one plots the graph.

Additionally, for this test, no renaming is completed, and the number of surveys is expanded to 50 to simulate a higher workload and provide more visual results. The red line in results a is the forced part of memory_profiler that shows the point of highest memory usage.

7.5.3 Time taken to run the program.

The runtime is calculated via an integrated logging function, that generates annotated runtime.log file in which parts of the program accomplished are mentioned. The test is completed 3 times for consistency and then the lowest order polynomial as a trendline is applied to the log-log chart of runtime via the Excel trendline option. The tests are additionally completed with both program and deployment package being located on Solid State Drive instead of slower Hard Disk Drive.

Assumptions:

- The tests are completed with the nominal setting of kappa, as it is computationally complex.
- The test assumes that no renaming is necessary.

7.6 Overview of statistical tests

Here, the research is completed solely with results provided by respondents in comparison to 7.6 Overview of statistical tests, where copies of respondent provided Excels are used to for verification purposes.

7.6.1 Consistency ratios

Calculated for all intra-dimensional comparisons. As one of the dimensions will always be perfectly consistent due to it having only 2 risk factors, the other 4 are looked at, and the surveys are discarded if the 2 or more dimensions are not within the cut-off consistency ratio of 0.1.

7.6.2 Krippendorff's alpha variations

Via a programming module (Castro & Santiago, 2017), (with whom a slight collaboration was completed for bugfixes), the results are calculated via this line, additionally shown below. The "TYPE" must be replaced with either "nominal", "ordinal" or "interval". The developed program provides all results necessary for it to work.

```
return (krippendorff.alpha(reliability_data=newlistconvert, level_of_measurement="TYPE"))
```

8 RESULTS

8.1 Calculation of statistical tests

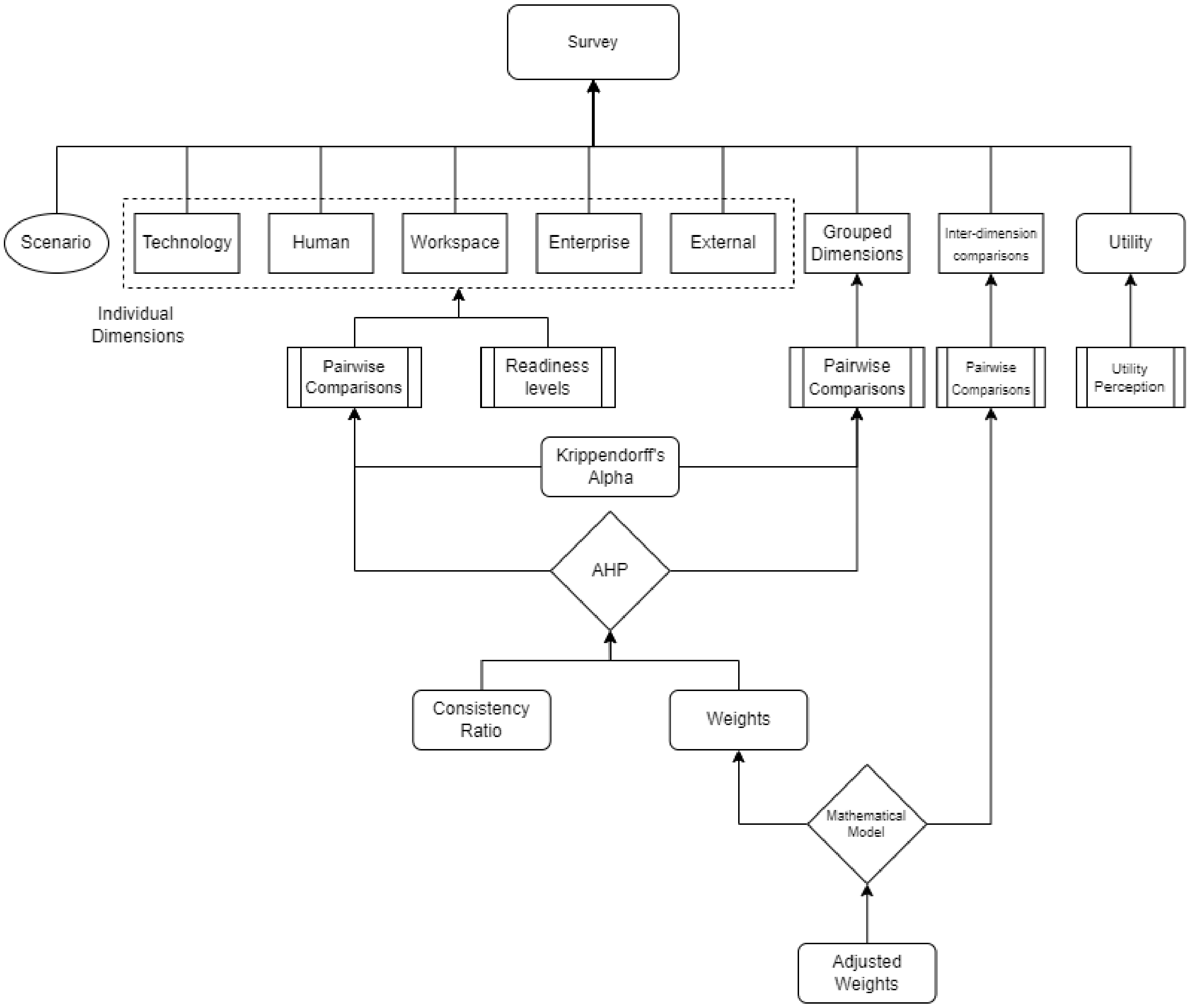

Figure 8-1: *Note.* Survey components and data extracted. From last iteration of survey.

8.1.1 Calculation of weights & Consistency Ratio

The implementation was completed initially in the first iteration in Excel, to provide results after survey completion. The implementation was further expanded to be completed automatically in later design implementation via code. Excel addition was based on (Bunruamkaew, 2012), where it was further expanded and automated in the same manner. The comparisons are completed within the weight portion of the topic, which is done for intra-dimensional (Table 8-1), inter-dimensional and group-dimensional comparison.

Table 8-1: *Note.* Example intra-dimensional pairwise comparisons completed by respondent.

Factor #1	Factor 1's importance relative to Factor 2	Factor #2
C1	1	C2
C1	1/4	C3
C1	1/7	C4
C2	1/8	C3
C2	1/7	C4
C3	1	C4

As referenced before in Intra-Dimensional & Dimension Groups Comparisons, the number of comparisons for a n number of factors is:

$$\text{№ } of\ comparions = \frac{n(n-1)}{2}$$

For this case there are 4 factors, there will be 6 comparisons. The results then populate a top right triangular matrix of a unit matrix of size $n \times n$, starting from top left to top right, followed by the next line until the upper corner is filled out (Table 8-2):

Table 8-2: *Note.* Unit matrix with inputs of respondent.

Factor	C1	C2	C3	C4
C1	1	1	1/4	1/7
C2		1	1/8	1/7
C3			1	1
C4				1

The lower left triangular matrix (Table 8-3), the reciprocal values of the upper diagonal are used. Given a_{ij} is an element of row i and column j, then the formula to calculate a_{ji} would be as below (Table 8-4).

$$a_{ji} = \frac{1}{a_{ij}}$$

Table 8-3: *Note.* Reciprocal matrix formula.

Factor	C1	C2	C3	C4
C1	1	1	1/4	1/7
C2	$\frac{1}{a_{12}}$	1	1/8	1/7
C3	$\frac{1}{a_{13}}$	$\frac{1}{a_{23}}$	1	1
C4	$\frac{1}{a_{14}}$	$\frac{1}{a_{24}}$	$\frac{1}{a_{34}}$	1

Table 8-4: *Note.* Completed reciprocal matrix using Table 4 formula on Table 3 values.

Factor	C1	C2	C3	C4
C1	1	1	1/4	1/7
C2	1	1	1/8	1/7
C3	4	8	1	1
C4	7	7	1	1

Once the pairwise comparisons of all risk factors have been done, a sum of each column is then calculated (Table 8-5), and all values are normalised by dividing by the sum of the column (Table 8-6), such as $a_{11} = \frac{1}{13}$, where a_{11} is a combination of factors C1 and C1.

Table 8-5: *Note.* Reciprocal matrix with totals.

Factor	C1	C2	C3	C4
C1	1	1	1/4	1/7
C2	1	1	1/8	1/7
C3	4	8	1	1
C4	7	7	1	1
Total	13	17	2.375	2.286

Table 8-6: *Note.* Normalised reciprocal matrix.

Factor	C1	C2	C3	C4
C1	0.08	0.06	0.11	0.06
C2	0.08	0.06	0.05	0.06
C3	0.31	0.47	0.42	0.44
C4	0.54	0.41	0.42	0.44

The mean of reach row is then calculated of the normalised matrix by taking the sum of each row and dividing it by the number of risk factors (C1, C2, C3, C4) which in this example is 4:

Table 8-7: *Note.* Normalised reciprocal matrix with averages.

Factor	C1	C2	C3	C4	Average
C1	0.08	0.06	0.11	0.06	0.0775
C2	0.08	0.06	0.05	0.06	0.0625
C3	0.31	0.47	0.42	0.44	0.41
C4	0.54	0.41	0.42	0.44	0.4525

Based on the above, 2 matrices can be extracted.

Matrix A, which is short for "Averages", is extracted from the "Average" column in Table 8-7, Each value is equivalent to the normalised weight of the factor, wherein the example above, C1 has a value of 0.0775 and C3 has a value of 0.41.

$$A = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} 0.0775 \\ 0.0625 \\ 0.41 \\ 0.4525 \end{bmatrix}$$

Matrix B is extracted from Table 8-4 as a complete copy.

$$B = \begin{bmatrix} 1 & \cdots & a_{ij} \\ \vdots & \ddots & \vdots \\ \frac{1}{a_{ij}} & \cdots & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \frac{1}{4} & \frac{1}{7} \\ 1 & 1 & \frac{1}{8} & \frac{1}{7} \\ 4 & 8 & 1 & 1 \\ 7 & 7 & 1 & 1 \end{bmatrix}$$

A 1-dimensional matrix is created by multiplying the B by A matrix, and then scalarly dividing by corresponding column of the final matrix as below:

$$C_{i1} = \frac{B \cdot A}{A_{i1}} = \begin{bmatrix} \frac{0.308}{0.0775} \\ \frac{0.257}{0.0625} \\ \frac{1.68}{0.41} \\ \frac{1.85}{0.453} \end{bmatrix} = \begin{bmatrix} 3.96 \\ 4.11 \\ 4.10 \\ 4.08 \end{bmatrix}$$

Example calculation:

$$C_{11} = \frac{a_1 \cdot a_{11} + a_2 \cdot a_{12} + a_3 \cdot a_{13} + a_4 \cdot a_{14}}{a_1} = \frac{(0.0775 \cdot 1) + (0.0625 \cdot 1) + \left(0.41 \cdot \frac{1}{4}\right) + \left(0.4525 \cdot \frac{1}{7}\right)}{0.0775} = 3.96$$

Consistency index is then calculated as per below, with $\lambda = average(C)$.

$$Consistency\ Index = \frac{\lambda - n}{n - 1} = \frac{4.06 - 4}{4 - 1} = 0.02$$

However, it is necessary to calculate the consistency ratio, as the consistency index varies depending on the size of the matrix. For that, it is necessary to divide the consistency index by a random index, which is a consistency index of a randomly generated reciprocal matrix based on the number of risk factors as per Saaty (Adams, 2017):

Table 8-8: *Note.* The consistency indices of randomly generated reciprocal matrices. From "Application of TOPSIS Method to the Selection of a Production Drilling Rig" by S. Chanda, 2018, Conference: First Zambian

n	1	2	3	4	5	6	7
Random Index	0.00	0.00	0.58	0.9	1.12	1.32	1.41

As our n = 4 in this calculation, the formula is as follows.

$$Consistency\ Ratio = \frac{CI}{RI_n} = \frac{0.0158}{0.9} = 0.175$$

8.1.2 Derivation of adjustment formula

For the purposes of verification and further weights adjustment, an additional formula has been created. It is possible to devise a coefficient formula for a group of risk factors, where their sum will always remain equal to 1 for normalisation as per original weight calculation.

For the formula below, W_d is the weight of dimension before the change, and $W_{d.u.}$ is the updated variant. The μ is the coefficient of the change desired, where if it equals 0, then the weight of dimensions are unchanged, and if it is 1, then the weight is all equal to each other.

Originally given the range of $0 \leq \mu \leq 1$, it was discovered that the only limitation is the for the coefficient is if the weights become negative. For the purposes of this study, the range of $-0.5 \leq \mu \leq 1.5$ was then used.

$$W_d + \left(\frac{1}{n} - W_d\right) * \mu = W_{d.updated}$$

As there are 5 factors within groups of dimensions, the equal weight division would be

$$\frac{1}{n} = \frac{1}{5} = 0.2$$

The weights calculated for the respondent are as follows:

Table 8-9: *Note.* Adjusted weights of dimensions of respondent “0SE1-KM5J”.[2]

	Weight				
Dimension	$\mu = 0$ Original	$\mu = 0.01$	…	$\mu = 0.99$	$\mu = 1$
Human	0.048	0.050	…	0.199	0.2
Technology	0.450	0.448	…	0.201	0.2
Collaborative Workspace	0.212	0.212	…	0.200	0.2
Enterprise	0.185	0.185	…	0.200	0.2
External	0.104	1.06	…	0.199	0.2

[2] For this example, results of respondent “0SE1-KM5J” are used.

Proof:
Given that:

$$\sum_{i=1}^{n} W_{d_i} = 1$$

And

$$W_{d_i} + \left(\frac{1}{n} - W_{d_i}\right) * \mu = W_{d.u._i}$$

Prove that:

$$\sum_{i=1}^{n} W_{d.u._i} = 1$$

Using direct proof:

Write $\sum_{i=1}^{n} W_{d.u._i}$ as a sum of equations including W_{d_i}

$$\sum_{i=1}^{n} W_{d.u._i} = W_{d_1} + \left(\frac{1}{n} - W_{d_1}\right) * \mu + W_{d_2} + \left(\frac{1}{n} - W_{d_2}\right) * \mu + \cdots + W_{d_n} + \left(\frac{1}{n} - W_{d_n}\right) * \mu$$

This can be simplified to

$$\sum_{i=1}^{n} W_{d.u._i} = \left(W_{d_1} + W_{d_2} + \ldots + W_{d_n}\right) + \mu * \left(\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}\right) - \mu * \left(W_{d_1} + W_{d_2} + \ldots + W_{d_n}\right)$$

$$\sum_{i=1}^{n} W_{d.u._i} = \sum_{i=1}^{n} W_{d_i} * (1 - \mu) + \mu$$

As

$$\sum_{i=1}^{n} W_{d_i} = 1$$

It can be concluded that.

$$\sum_{i=1}^{n} W_{d.u._i} = 1 * (1 - \mu) + \mu = 1 - \mu + \mu$$

$$\sum_{i=1}^{n} W_{d.u._i} = 1$$

8.1.3 Calculation of adjusted weights

For coefficient, between $-0.5 \leq \mu \leq 1.5, steps\ of\ 0.1$, a table is completed.

$$W_{risk\ factor\ u.} = W_{risk\ factor} * W_{d.u.}$$

For each version of coefficient, the weights are compared on individual level, where "Input Δ" is the input of AHP inter-dimensional comparison completed by the participant.

Table 8-10: *Note.* Adjusted intra-dimensional comparisons.

Left column	Expected Importance Δ	Participant Input Δ	Right column
Technology-Tooling	$\frac{W_{Tooling} * W_{Technology\ u.}}{W_{Work\ stress} * W_{Human\ u.}}$	Δ_1	Human - Work stress
Human - Cognitive workload	$\frac{W_{Cog.Work.} * W_{Human\ u.}}{W_{Haz.Obst.} * W_{Workspace\ u.}}$	Δ_2	Workspace - Hazardous obstacles
....	...	...	...
Enterprise - Safety strategy	$\frac{W_{Saf.Strategy} * W_{Enterprise\ u.}}{W_{Security} * W_{Technology\ u.}}$	Δ_{12}	Technology - Security

For each version of coefficient, the difference is calculated in percentage form for each comparison,

$$\%\ discrepancy = \frac{Input\ \Delta}{Expected\ Importrance\ \Delta} * 100 - 100$$

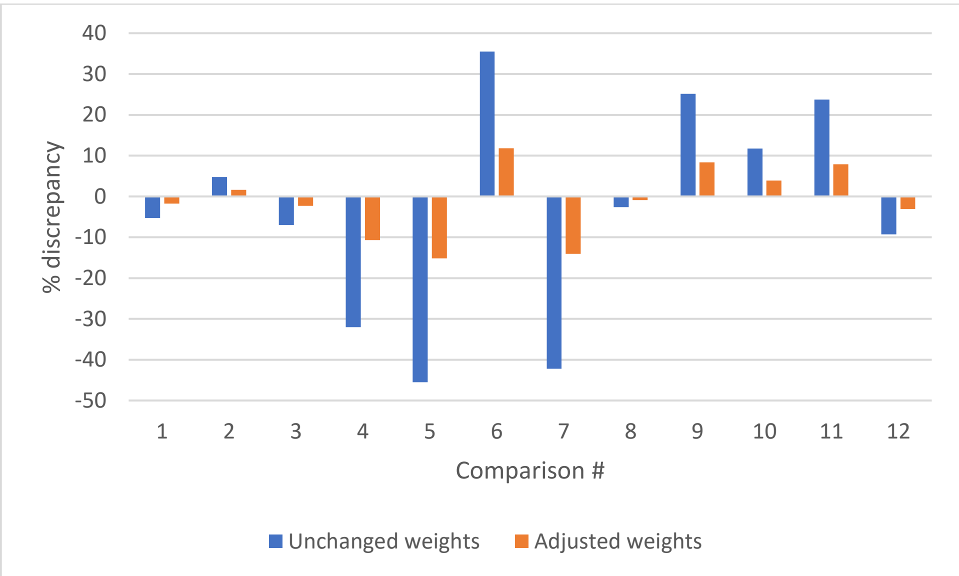

Figure 8-2: *Note.* Simplified desired weight adjustment model. A general decrease in %discrepancy can be seen.

The following are calculated:

$$Mean = \overline{x} = \frac{\sum_{i=1}^{12} \%\ discrepancy_i}{12}$$

$$Standart\ deviation = \sigma = \sqrt{\frac{\sum_{i=1}^{12} (\%\ discrepancy_i - \overline{x})^2}{12 - 1}}$$

Table 8-11: *Note.* Mean and standard deviation of the % discrepancy based on different coefficient of the change desired (μ) values.

μ	**Mean (%)**	**Standard deviation (%)**
-0.5	$\overline{x}_{-0.5}$	$\sigma_{-0.5}$
-0.4	$\overline{x}_{-0.4}$	$\sigma_{-0.4}$
…	…	…
1.5	$\overline{x}_{1.5}$	$\sigma_{1.5}$

In this study it is considered that when the mean is close to 0, the mathematical model has been accurately representative of all dimensions. Additionally, when the standard deviation is close to 0 it means that the mean can be considered non-volatile and consistently representative of mathematical model. Most accurate depiction of μ is then chosen, with all weights changed accordingly.

8.1.4 Krippendorff's Alpha

Due to the excessive sizing and computational necessity of Krippendorff's alpha, an example is not shown. Instead, a Python library is implemented, of hosted on GitHub under name "fast-krippendorff", which was verified to obtain the same result via Excel implementation from (Zaiontz, s.d.). The implementation was further verified by online implementation additionally for accuracy (Freelon, s.d.), where same mathematical result is obtained via an online Krippendorff calculator.

The calculation is completed with pairwise comparisons of each rater being a becoming column, as seen in Figure 8-3, where there are 5 raters, each given a letter as a name. Each cell is their respective response, such as A6 (B9 in spreadsheet) cell being representative of person A, responding with "2" to question № 6. The rows can be seen by the specific identifier of a pairwise comparison, which is identical for all participants. No additional steps are necessary for the calculation of results due to automation. The groups are split into dimension-wide comparisons.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Rating Table							Agreement Table				
2												
3		A	B	C	D	E			1	2	3	4
4	1	1	2		3	4		1	1	1	1	1
5	2	2	2	3	2			2	0	3	1	0
6	3	1	1	1	2	2		3	3	2	0	0
7	4	4	4	3	4	2		4	0	1	1	3
8	5	3	3	4		3		5	0	0	3	1
9	6	2		1	1			6	2	1	0	0
10	7	3	3	3	4			7	0	0	3	1
11	8	4	3	4	3	4		8	0	0	2	3
12	9											
13	10		3									

Figure 8-3: *Note.* Respondents and their table of agreement. Agreement table is used for nominal alpha and is not used in this study. From Krippendorff's Alpha Basic Concept by C. Zaiontz, Real Statistics Using Excel. Copyright n.d. of C. Zaiontz.

A variety of possible implementations are shown in 6.1.6 Inter-rater Reliability Analysis. For this study, AHP was used and as even spacing is expected, ordinal is a best fit as nominal does not consider the difference between the values, relying purely on agreements, whereas the interval rating is not preferrable due to uneven difference between the possible indices, such $0.\dot{1}$ and 0.125, 0.2 and 0.25, representing $\frac{1}{9}, \frac{1}{8}, \frac{1}{5}$ and $\frac{1}{4}$ respectively.

In ordinal implementation, an example of usage would be raters deciding on the performance of individuals on a scale of 1 to 5, with 1 being poor and 5 being excellent. Here the distance between ranks is consistent (distance is always 1) and considered, however, the ranks all have equivalent values.

8.1.5 Grouping of results for Krippendorff's Alpha

Upon initial feedback and calculations from the first iteration, it was decided to further verify the validity of inter-rater reliability for this study. After the first iteration of the survey, Krippendorff's alpha of the first 2 dimensions was calculated in Excel and gave a below 0.05 result, whereas a customary 0.80 was expected for at least some categories, at the end of the study (Adams, 2017). As such pairwise comparisons were simplified into 4 possible groups, with one being original data.

Table 8-12: *Note.* Krippendorff's alpha results to verify coder inter-reliability.

<table>
<tr><td>17</td><td>1/9</td><td>1/8</td><td>1/7</td><td>1/6</td><td>1/5</td><td>1/4</td><td>1/3</td><td>1/2</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td></tr>
<tr><td>9</td><td colspan="2">[1/9 1/8]</td><td colspan="2">[1/7 1/6]</td><td colspan="2">[1/5 1/4]</td><td colspan="2">[1/3 1/2]</td><td>1</td><td colspan="2">[2 3]</td><td colspan="2">[4 5]</td><td colspan="2">[6 7]</td><td colspan="2">[8 9]</td></tr>
<tr><td>5</td><td colspan="4">[1/9 1/8 1/7 1/6]</td><td colspan="4">[1/5 1/4 1/3 1/2]</td><td>1</td><td colspan="4">[2 3 4 5]</td><td colspan="4">[6 7 8 9]</td></tr>
<tr><td>3</td><td colspan="8">[1/9 1/8 1/7 1/6 1/5 1/4 1/3 1/2]</td><td>1</td><td colspan="8">[2 3 4 5 6 7 8 9]</td></tr>
</table>

If the above data fell into the groups, it was then converted into the associated cell of Table 8-13: *Note.* Conversion of first iteration results.

Table 8-13: *Note.* Conversion of first iteration results.

17	1/9	1/8	1/7	1/6	1/5	1/4	1/3	1/2	1	2	3	4	5	6	7	8	9
9	1/5		1/4		1/3		1/2		1	2		3		4		5	
5	1/3				1/2				1	2				3			
3	1/2								1	2							

An example would be a reading of 1/7. It would be converted to:

Table 8-14: *Note.* Conversion of 1/7.

Number of groups	Converted to
17	1/7
9	1/4
5	1/3
3	1/2

All values are iterated through and converted into 4 different tables, from which then only Krippendorff's alpha is calculated.

8.2 Survey Results

8.2.1 Consistency ratios

Table 8-15: *Note.* Consistency ratios - pre-removal

	0SE1-KM5J[3]	57UY-ACHJ	7QGW-6366	8A6M-OHNV	AW0R-DGL7	CEMM-9AS0	LM4B-MEP8	QJD2-FT5B	RF7I-LIDP	Mean
Human	0.080	0.040	0.049	0.085	0.105	0.056	0.138	0.053	0.289	0.099
Technology	0.097	0.024	0.052	0.064	0.000	0.058	0.028	0.452	0.015	0.088
Collaborative Workspace	0.180	0.009	0.014	0.018	0.231	0.035	0.036	0.132	0.258	0.102
Enterprise	0.064	0.037	0.069	0.080	0.529	0.050	0.067	0.154	0.437	0.165
External	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 8-16: *Note.* Consistency ratios - post-removal

	0SE1-KM5J	57UY-ACHJ	7QGW-6366	8A6M-OHNV	CEMM-9AS0	LM4B-MEP8	Mean
Human	0.080	0.040	0.049	0.085	0.056	0.138	0.075
Technology	0.097	0.024	0.052	0.064	0.058	0.028	0.054

[3] As per 7.4.2, these are respondent's names, that were changed to comply with GDPR anonymity compliance in form of XXXX-XXXX

Collaborative Workspace	0.180	0.009	0.014	0.018	0.035	0.036	0.049
Enterprise	0.064	0.037	0.069	0.080	0.050	0.067	0.061
External	0.000	0.000	0.000	0.000	0.000	0.000	0

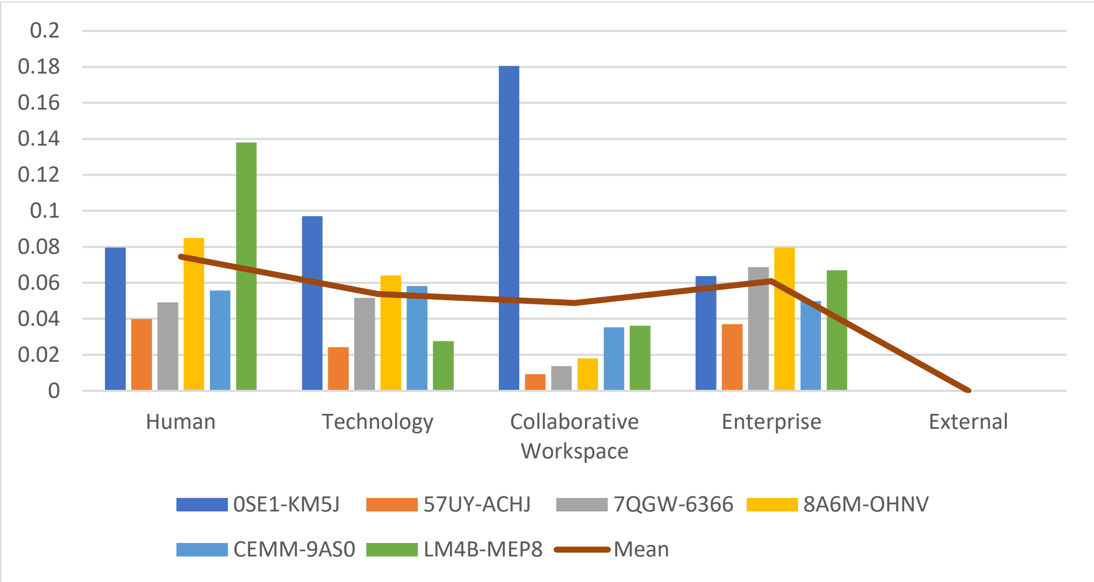

Figure 8-4: Bar chart of consistency ratios and mean- post-removal.

The results show the consistency indices of varying AHP groups, where those highlighted represent a $CR > 0.1$, with an application of filtering to remove inconsistent participants who would negatively impact the study. Figure 8-4 is created to help with the visualisation and understanding of results.

8.2.2 Readiness levels of risk factors chosen by respondents

As this is largely non-impactful for discussion and insight into the study, the inputs used are located in Appendix A.1 Maturity levels

8.2.3 Krippendorff's Alpha

Table 8-17: *Note.* Ordinal Krippendorff's Alpha results

	3	5	9	17
Grouped Dimensions	0.035	0.067	0.106	0.072
Technology	-0.088	-0.087	-0.083	-0.067
Human	-0.019	-0.047	-0.080	-0.108
Collaborative Workspace	-0.031	-0.031	-0.031	-0.075
Enterprise	0.287	0.266	0.252	0.256
External	0.000	0.000	0.000	0.000
Mean	0.037	0.034	0.033	0.016

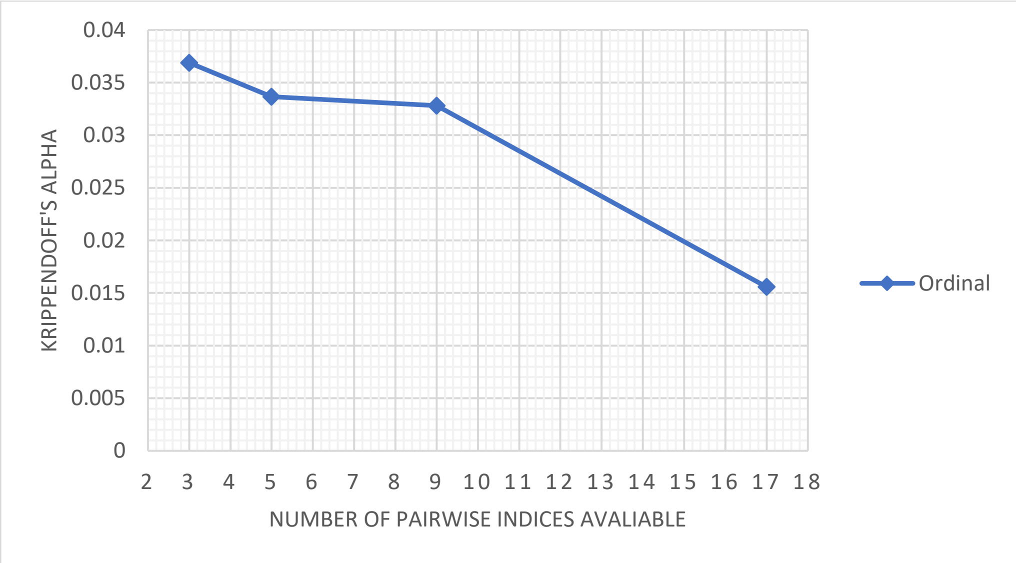

Figure 8-5: *Note.* Krippendorff's alpha – Mean

In this subsection, the results of Krippendorff's Alpha are compiled, and a visualisation of means is shown to provide an overview of original and simplified groups used for calculation. As all calculated alpha results are close to 0.0, with only 0.2 showing indication of agreement, Figure 8-5 shows very low levels of agreement between the raters at all simplified levels.

8.2.4 Initial weights of grouped dimensional and sub-dimensional factors

Table 8-18: *Note.* Weights of grouped dimensions - Initially

	0SE1-KM5J	57UY-ACHJ	7QGW-6366	8A6M-OHNV	CEMM-9AS0	LM4B-MEP8
Technology	0.048	0.375	0.402	0.285	0.277	0.083
Human	0.450	0.253	0.223	0.220	0.261	0.475
Collaborative Workspace	0.212	0.152	0.126	0.220	0.204	0.194
Enterprise	0.185	0.088	0.155	0.196	0.111	0.142
External	0.104	0.132	0.094	0.079	0.146	0.106

The values for individual risk factors can be found in Appendix A.2 Initial weight calculations

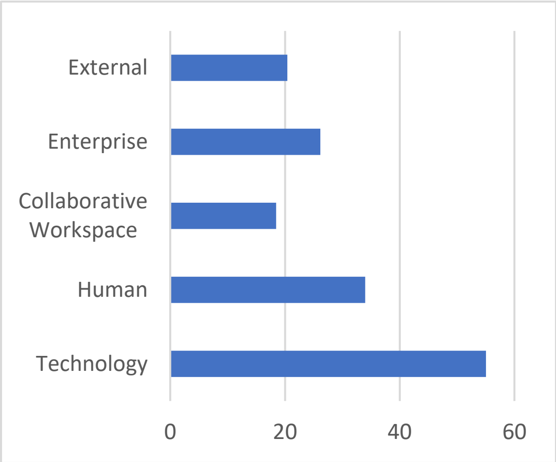

Figure 8-6: *Note.* Standard deviation as % of mean – Grouped dimensions

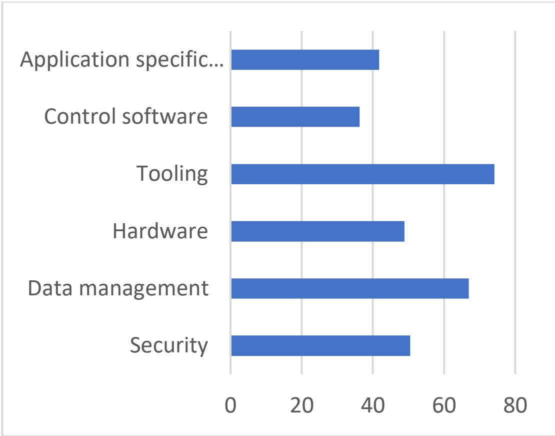

Figure 8 4: *Note.* Standard deviation as % of mean - Technology

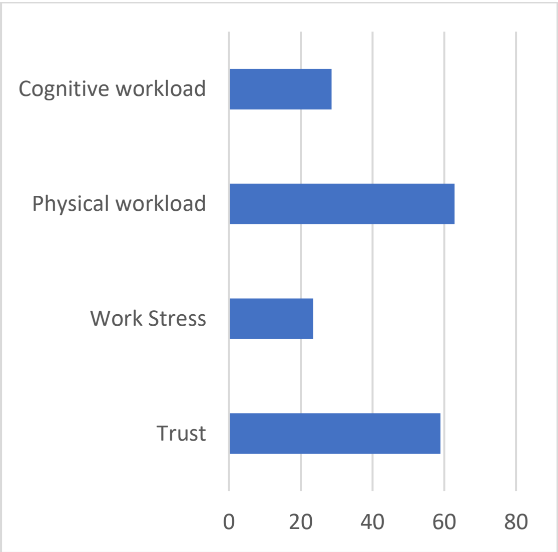

Figure 8-7: *Note.* Standard deviation as % of mean - Human

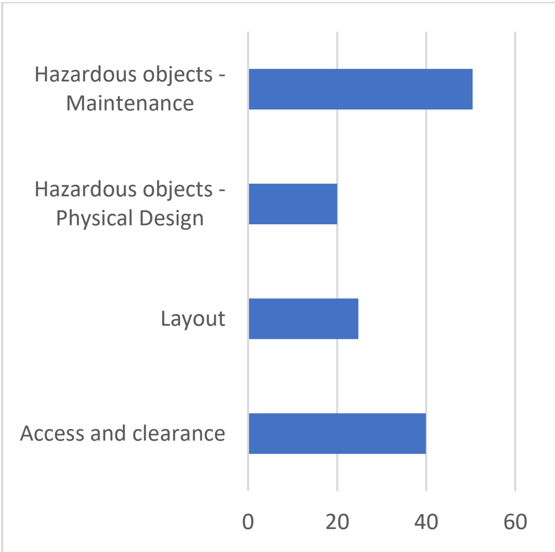

Figure 8-8: *Note.* Standard deviation as % of mean - Collaborative Workspace

Figure 8-9: *Note.* Standard deviation as % of mean - Enterprise

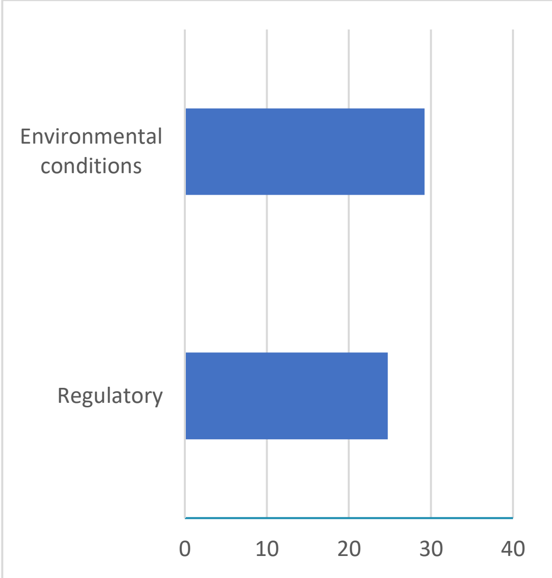

Figure 8-10: *Note.* Standard deviation as % of mean - External

Table 8-19: *Note.* Average Standard deviation as % of mean of sub-dimensions of grouped dimensions.

Grouped dimension	Mean standard deviation as % of mean of sub-dimensions
Technology	53.1
Human	43.5
Collaborative Workspace	33.8
Enterprise	45.4
External	27.0

In above section, standard deviation is used to see the inter-user agreement. The charts show the individual σ of each risk factor. The risk factors are then compiled into groups to observe the means of standard deviations of each dimensions. This is an alternative to Krippendorff's alpha for a different evaluation of inter-rater reliability. It can be seen that the Krippendorff's alpha can be seen as less representative of agreement between the respondents.

8.2.5 Inter-dimensional comparisons

As this is largely non-impactful for discussion and insight into the study, the inputs used are located in Appendix A.3 Inter-dimensional comparisons

8.2.6 Adjustment of Weights of Groups and Risk factors

The following 2 tables are calculated from: A.5 Percentage deviation from ideal inter-dimensional comparison from which mean and standard deviation are calculated

Table 8-20: *Note.* Mean % of discrepancy from inter-dimensional comparison

μ	0SE1-KM5J	57UY-ACHJ	7QGW-6366	8A6M-OHNV	CEMM-9AS0	LM4B-MEP8
-0.5	-1095.1	368.2	78.6	114.4	29.0	105.1
-0.4	-2367.9	264.3	82.6	74.8	17.3	73.2
-0.3	12069.3	204.3	86.4	59.3	8.4	56.9
-0.2	1660.7	165.7	90.0	51.8	1.4	47.3
-0.1	881.2	139.2	93.3	48.0	-3.9	41.3
0	595.3	120.3	96.5	46.3	-8.2	37.6
0.1	447.4	106.5	99.4	45.8	-11.4	35.4
0.2	357.3	96.4	102.3	46.1	-14.0	34.3
0.3	296.9	89.0	105.0	47.0	-15.8	34.2
0.4	253.9	83.7	107.8	48.3	-17.1	34.9
0.5	222.0	80.1	110.4	50.0	-18.0	36.3
0.6	197.6	78.0	113.1	51.9	-18.4	38.4
0.7	178.6	77.0	115.8	54.1	-18.4	41.3
0.8	163.8	77.2	118.6	56.5	-18.0	45.2
0.9	152.2	78.3	121.5	59.1	-17.3	50.1
1	143.5	80.3	124.6	61.9	-16.2	56.6
1.1	137.4	83.2	127.8	64.9	-14.9	65.1
1.2	134.0	86.9	131.4	67.9	-13.1	76.9
1.3	133.8	91.5	135.5	71.2	-11.1	94.1
1.4	138.4	97.0	140.3	74.6	-8.7	121.5
1.5	151.2	103.4	146.1	78.2	-5.9	172.7

Table 8-21: *Note.* Standard deviation of discrepancy from inter-dimensional comparison

μ	0SE1-KM5J	57UY-ACHJ	7QGW-6366	8A6M-OHNV	CEMM-9AS0	LM4B-MEP8
-0.50	3603.6	1004.6	199.3	322.1	225.7	299.2
-0.40	7824.5	712.3	219.8	190.1	193.0	211.8
-0.30	40057.8	542.8	239.9	138.3	167.2	169.8
-0.20	5535.1	432.8	258.7	113.3	146.2	146.4
-0.10	2947.5	356.3	276.0	100.6	128.9	132.7
0.00	1996.7	300.8	291.7	94.4	114.4	124.7
0.10	1502.8	259.5	306.0	92.1	102.0	120.4
0.20	1200.2	228.5	319.0	92.2	91.5	118.6
0.30	995.8	205.7	330.9	94.0	82.6	118.7
0.40	848.5	189.6	341.8	96.9	75.0	120.0
0.50	737.3	179.1	351.7	100.8	68.6	122.2
0.60	650.5	173.7	360.8	105.3	63.5	125.1
0.70	581.0	172.5	369.1	110.5	59.5	128.5
0.80	524.2	175.2	376.8	116.1	56.8	132.3
0.90	477.0	181.0	383.9	122.1	55.3	136.4
1.00	437.4	189.7	390.4	128.5	55.0	141.0
1.10	403.8	200.6	396.4	135.3	55.9	146.4
1.20	375.1	213.7	401.9	142.4	58.0	153.3
1.30	350.6	228.5	407.1	149.9	61.0	163.7
1.40	330.0	245.0	411.9	157.6	65.1	183.7
1.50	314.4	263.1	416.5	165.6	70.0	232.4

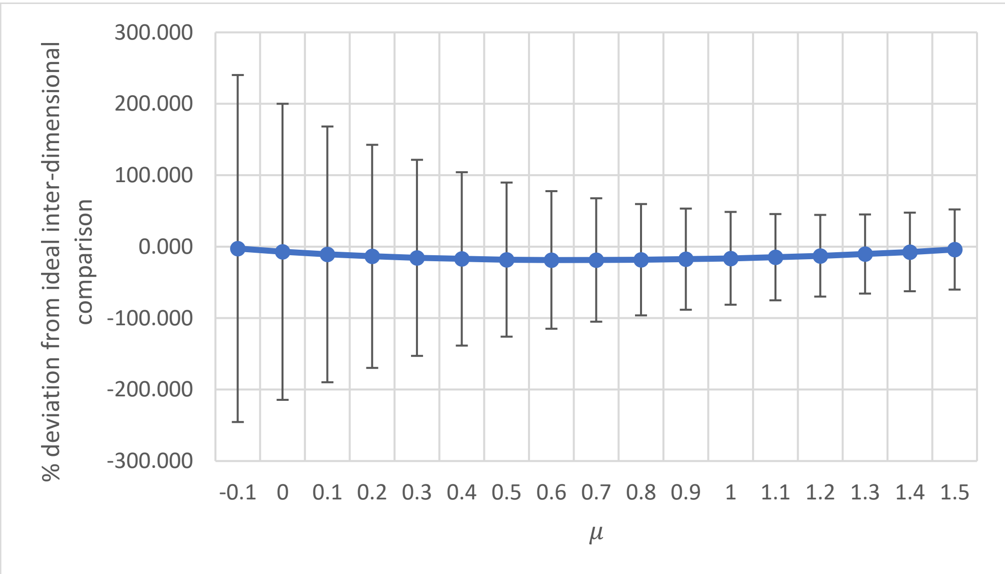

Figure 8-11: *Note.* μ and mean % deviation from ideal inter-dimensional comparison (with standard deviation as error bars)- Example.

For the results above, a chart is plotted with mean as a function of μ. Standard deviation is used as error bars. The coefficient with the mean closest to 0 is chosen, and if there are mean values that are similar to each other, the value with lower standard deviation is chosen.

Table 8-22: *Note.* Optimal individual μ and associated statistical characteristics.

	Optimal μ	Deviation from true value (%)	σ of deviation from true value (%)
0SE1-KM5J	0.6	117.6	314.9
57UY-ACHJ	0.7	70.7	166.0
7QGW-6366	-0.5	87.8	232.8
8A6M-OHNV	0.1	37.3	77.2
CEMM-9AS0	1.5	-3.9	73.6
LM4B-MEP8	0.7	48.5	117.0

8.2.7 Comparison of maturity indices

Table 8-23: *Note.* Non-weighted average dimension maturity indices

	0SE1-KM5J	57UY-ACHJ	7QGW-6366	8A6M-OHNV	CEMM-9AS0	LM4B-MEP8
Technology	2.17	2.67	3.50	2.33	2.33	2.67
Human	2.75	2.75	3.25	2.00	-	1.75
Collaborative Workspace	3.50	2.50	2.75	2.75	2.75	2.25
Enterprise	2.71	2.43	3.00	1.57	2.57	2.29
External	1.50	2.50	3.50	1.50	2.00	2.00

Table 8-24: *Note.* Weighted original average dimension maturity indices

	0SE1-KM5J	57UY-ACHJ	7QGW-6366	8A6M-OHNV	CEMM-9AS0	LM4B-MEP8
Technology	1.96	2.81	3.60	2.36	2.54	2.59
Human	2.84	2.65	3.09	2.16	-	2.35
Collaborative Workspace	3.41	3.04	2.77	2.47	2.84	2.52
Enterprise	2.88	2.87	3.29	1.74	2.71	2.62
External	1.50	2.67	3.75	1.50	2.00	2.33

Table 8-25: *Note.* Weighted optimal average dimension maturity indices

	0SE1-KM5J	57UY-ACHJ	7QGW-6366	8A6M-OHNV	CEMM-9AS0	LM4B-MEP8
Technology	1.90	2.86	3.60	2.35	2.52	2.93
Human	2.84	2.65	3.09	2.16	-	2.35
Collaborative Workspace	3.41	3.04	2.77	2.47	2.84	2.52
Enterprise	2.88	2.87	3.29	1.74	2.71	2.62
External	1.50	2.67	3.75	1.50	2.00	2.33

In these tables, a progression of calculated maturity indices is shown. The first table shows unweighted means of dimensions, whereas in the following 2 implementations weights are used. The third one uses the optimised group dimension coefficients based on final application of maturity model.

8.2.8 Utility

Table 8-26: *Note.* Utility curve inputs

	8A6M-OHNV	CEMM-9AS0	LM4B-MEP8	57UY-ACHJ	0SE1-KM5J	7QGW-6366	Total
	Early stage						
Technology	3	7	4	5	7	3	29
Human	5	8	10	6	10	3	42
Workspace	1	6	6	4	3	2	22
Enterprise	6	8	3	3	3	4	27
External	1	8	4	2	8	5	28
	Middle stage						
Technology	5	7	6	4	9	5	36
Human	5	8	9	7	10	5	44
Workspace	3	9	7	4	6	4	33
Enterprise	6	7	5	4	6	5	33
External	4	7	5	2	8	6	32
	Late stage						
Technology	7	9	9	7	10	8	50
Human	6	7	8	8	10	6	45
Workspace	8	8	8	4	9	5	42
Enterprise	8	6	6	3	7	6	36
External	8	8	6	2	10	7	41

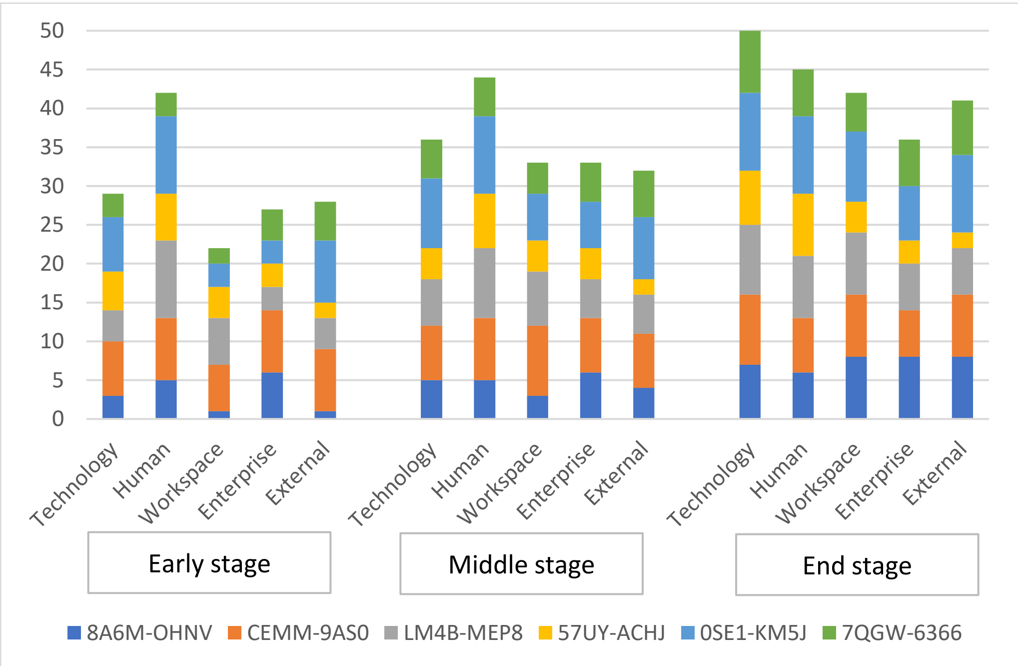

Figure 8-12: *Note.* Stacked columns of perceived utility at different stages of implementation

In this part, the respondents replied to the following questions:

How useful are this dimension's subclasses if you know:

- Only their risks?
- Risks and the tools, but they are not yet tailored for cobots?
- Have complete integration of risk factors tailored to cobots?

As shown in Figure 8-12:, a non-consistent increase in utility is perceived. Technology and Workspace dimensions have a continuous increase in utility from stage to stage, whereas Human dimension is nearly maximal at earlier stages of implementation and only grows slightly. Technology readiness, however, undergoes a massive increase in importance to massively outperforming other dimensions. External dimension, pertaining to the regulations continues to increase in usability, while Enterprise dimension only increases slightly form stage to stage.

8.3 Program characteristics

8.3.1 Time taken to run the program.

Table 8-27: *Note.* Time taken to run the program.

Number of files used	Run #1 (s)[4]	Run #2 (s)	Run #3 (s)	Average (s)
1	1.15	1.15	1.16	1.15
5	1.44	1.42	1.4	1.42
10	1.83	1.86	1.91	1.87
50	4.7	4.7	4.63	4.68
100	8.3	8.33	8.2	8.28
500	41.53	41.27	41.11	41.30
1000	95.64	98.41	97.93	97.3

[4] The run # refers to the program runtime, which is repeated 3 times for consistency.

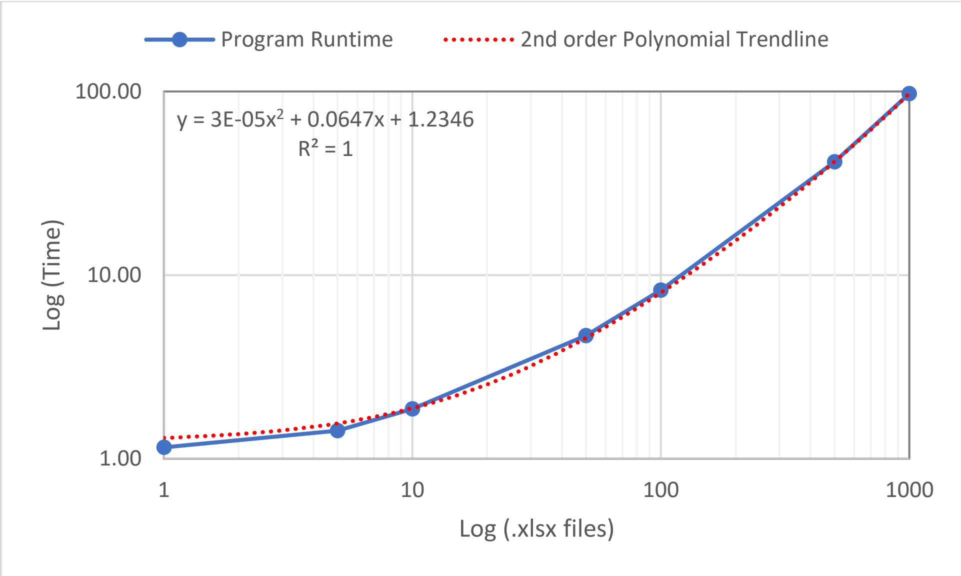

Figure 8-13: *Note.* Time taken to run the program and associated lowest order exponential.

8.3.2 Storage sizes

Table 8-28: *Note.* Size variations of varying parts of deployment package

	Size allocated	True size
Completed survey size	4.39 MB	4.39 MB
Total Outputs size(n=6)	32 kB	18.7 kB
Pytest directory	84 kB	63.2 kB
Necessary files	160 kB	104 kB
Deployment package (n=6 with no reports)	26.7 kB	26.6 kB
Deployment package with no sample surveys	384 kB	300 kB
Deployment package with no sample surveys and modules installed	26.7 MB	26.6 MB

Table 8-29: *Note.* Free space necessary for installation of mandatory modules

Module	Size
Pytest	320 kB
Pandas	10.7 MB
Openpyxl	250 kB
NumPy	14.8 MB
Krippendorff	18.0 kB
Memory_profiler	31.8 kB

8.3.3 Memory usage

Table 8-30: *Note.* RAM usage and important time points.

Task	Time (s)	Memory usage (MB)	Acronym
Importing modules	0	0	IM
Program starts to be run	1.4	85	
Initialisation start	1.4	85	
Initialisation complete	1.7	100	
Setting up Excel import settings	1.7	100	ES
Excel import start	7.7	173	EI
Excel import ended	22.9	144	
Calculations and report generation	22.9	144	C
Program ended	24	80	

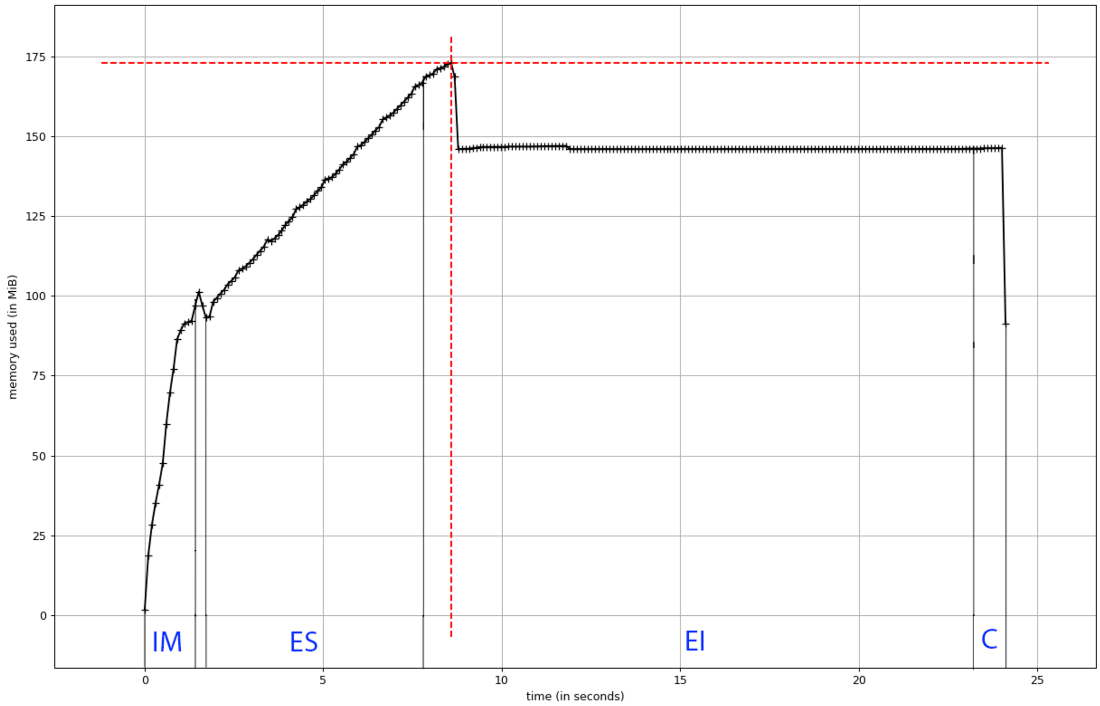

Figure 8-14*: Note.* RAM usage-time chart

The largest memory usage as shown in Figure 8-14*: Note.* RAM usage-time chart, consists of program initialisation, with necessary modules occupying half of memory and the usage of them and necessary variables being initiated taking up another half. The processing of Excels has a constant memory use rate, as a file is always being open. At the end of processing RAM usage drops as only lightweight report files are being generated.

9 DISCUSSION

9.1 Results Discussion

9.1.1 Consistency ratios

The consistency ratios are looked at to determine the diligence and attention the respondent pays to the survey, as well as systematic comprehension and ability to consistently judge criteria. Overall, a third of the results had to be discarded due to standards set at the beginning of the research. This can be attributed to lower respondent standards set in the third iteration of the survey, which compromised engineering students and fresh graduates, instead of expected academic experts in second iteration. This means that they were less likely to take it seriously and skim the text when completing the questions.

The largest failures of consistency can be contributed to extreme weights assigned during some comparisons, which include those that have the lowest calculated weight, such as “AI decision making” and “Social Acceptance”. For the results that were discarded, it was highly likely that they assigned extremely high values, such as 8 or 9 during AHP to those sub-dimensions and did not compensate in other comparisons sufficiently enough. This occurrence can be attributed to consistency ratios of 0.2 or higher, whereas for those between it can be expected that it was due to a systematic error by a participant, where small inconsistencies combined and caused the consistency ratio to rise.

The dimension of “Enterprise” has given the most issue to the respondents, as it contains more abstract risk factors, such as “AI decision making” and “Tech/Digitization strategy”, while those dimensions with more clear-cut constituents, such as “Technology”, which contains “Work Stress”, “Trust” and “Physical workload”, had by far the lowest average, with the exception of one person who prioritised them.

Overall, the implementation can be considered a success, as it provided results as expected, giving a quantifiable way to see the consistency of responses of AHP. The target of answering the research question with high accuracy results has been worked towards, at a small cost to sampled numbers.

9.1.2 Brief overview of design iterations

The conclusions that were drawn showed that definitions are hard to understand, and results, consisting of consistency ratios and weights of dimensions are hard to interpret. Furthermore, the feedback received from the students indicated that the initial design was overwhelming and lacked necessary eye-catching behaviour. This was the large reason why the definitions were first moved to a separate PDF file, followed by complete removal. However, by removing the definitions of each risk factor within each dimension, the interpretation and uncertainty of respondents increased as well. This was however the cost of trying to reach an answer to the main research question, leading to a choice of decreasing the complexity of the survey to obtain more completed surveys.

9.1.3 Krippendorff's Alpha

As seen from the results, the alpha is close to 0 and it can be seen that low amount of consensus across the pairwise comparisons can be seen. In academic discussion there is no consensus, however an alpha above 0.2 can be seen as showing low corelation, however here it only occurs within 1 dimension. In general, the mean is very close to 0, and it can be concluded that there is no general agreement between raters.

The fact that the "External" dimension is considered when calculating the mean of the alpha decreases it, as when there is only 1 comparison. Krippendorff's alpha is not built for a single comparison and has a value of 0 when it occurs. However, on a larger scale that does not affect the inter-rater reliability due to average being extremely low already.

9.1.4 Mathematical model

To provide conclusions in the beginning, the creation of weight adjustment has been successful. As based on 2 separate sources, weights have been compiled into a more consistent structure. While not ideal, an improvement has been made in both the mean and standard deviation of difference from inter-dimensional comparisons.

While this study was tasked with answering a creation of a single weighting system, the authors conclude that from the varied responses, it would be unwise to simplify into a singular weight model due to the large standard deviation for weights calculated. The authors however present their findings on an individual level, which shows more promise as their responses are judged solely on their inputs alone, with only mathematical models affecting them.

While initially a range of is proposed due to possible negative weights due to the formula used, it was found that positive effects can be had if the scale is bypassed, as long as the weights do not become negative, with less than half of coefficients going beyond the originally proposed scale.

If the tables within weighted and non-weighted maturity indices are investigated, it can be found that the weighting system developed allows the general mean index to remain the same, however, it largely shifts the importance of all dimensions. The adjustment implemented has unchanged the final maturity indices from the original AHP system or has additionally furthered the difference from the original unweighted maturity indices.

This mathematical model developed can be said to improve the development of risk factors and dimensions as a whole, in return furthering the cobot safety systems. An example would be changing the perception of what is considered to be the current level of dimension, in this study allowing the level to go up in some cases by 0.6, allowing more specific usage of resources for more equal and accurate usage of resources to bring cobots up to 4IR standards.

9.1.5 Utility curve based on maturity level

This part of the research shows that the importance of the risk factors and ability to be usable and a perceived net positive are different things. This is best exemplified by the External dimension, which consistently has the lowest weight out of all respondents by not a small margin, however when asked to judge its importance at various stages it shows a completely different view of the impact it has. As such the level of importance that the person has put on a risk factor can be further impacted not just by the weight of said factor, but also by the usability and utility of the level of selected factor. Moreover, based on the mean weights of dimensions, the middle stage is the best approximation of ranking weights. However, unlike in mean weights, the utility of the technological dimension overtakes the human dimension, despite it being close. In addition, the perceived utility has a much closer cluster of results in comparison to weights, where the least important readiness weights can be up to 3x less important than the most important ones.

It thus can be said that it should be included in the mathematical model in future cases.

9.1.6 Automation summary

While the program is capable of high-speed file reading, the further suggestions would be to implement multi-threading to speed up the processing and use another file format, that has faster read speeds in comparison to user experience oriented .xlsx files. In addition, the design failed to obtain the desired non-quadratic time complexity, however, the authors note that is difficult to bypass as the programs increase in size. While the lowest order function that fits the log-log chart is quadratic, Excel suggests that a cubic function better represents the chart, however, due to a relatively low-level coefficient, it was negated. Additionally, it can be seen that for the approximation function of $y = 3 * 10^{-5}x^2 + 0.0647 * x + 1.2346$ at the lower sample sizes, the quadratic part is negligible, which can be also said for the cubic function. However, for larger sample sizes it becomes more easily visible.

As the program was designed from scratch, it is extremely lightweight. However additional steps can be taken to decrease the program initialisation time, which is more impactful at a lower sample count. It would only consist of importing the necessary portions of modules. An additional external review would be best due to the highly specific implementation.

The space taken by programs is additionally dependent on the number of samples used. The absolute minimum space used by the deployment package alone is 26.7 MB and each sample is 4.39MB. At a low number of files (n<10), that deployment package affects the space needed, however at larger sample sizes, the size necessary can be largely modelled as a linear function of samples alone. Due to the low number of files used in the program, the allocated size is less having a 1% variation from size and can be concluded that is not impactful.

It can be concluded that the program is capable of being completed on any computer that contains the necessary packages and Python, 0.25 gigabytes of RAM and insignificantly more storage space to contain the deployment package, which can vary depending on the number of samples. While it needed large amounts of time to be able to process the inputs, the current automation is largely flexible and allows for rapid development of any further adjustment of the mathematical model used, which can be seen as a response to the research sub-question of possible rapid processing of results.

9.2 Limitations of the research

9.2.1 Insufficient sample size and respondent profile

During the application stage of the initial design of the Excel maturity model, results were obsolete due to multiple reasons. The first reason is that cobots are a specific niche within the vast technological sector, thus finding industrial experts with prior experience in working with cobots was hard to come by. Additionally, most engineers that came across this research have more experience IN working with robots instead of cobots, which increased the difficulty of obtaining trustworthy and accurate results. Furthermore, the Excel survey in itself was long and overwhelming when receiving feedback from respondents, with an average completion time of 35 minutes. Although online surveys are a norm in today's research community and that can collect large amounts of data, the main focus will always be maintaining a high response rate. However, since this survey's completion time is too long and is quite demanding on respondents, it reduces the response rate resulting in different scenarios as mentioned by Hoang (Tran, s.d.) :

- Overwhelming surveys will push respondents to stop midway or click through carelessly;
- Long surveys are less engaging, whilst completing short surveys will feel rewarding for the respondents;
- If the survey is long and overwhelming, respondents might become frightened by the survey.

9.2.2 CSRAT

CSRAT is a tool that assesses the readiness of a cobot based on a set of predefined criteria and does not consider changes that occur during operation. This means that if any modifications are made to the cobot or unexpected events occur, CSRAT needs to be manually updated to ensure that it accurately reflects the current state of the cobot. This limitation was highlighted by (Berx, Adriaensen, Decré, & Pintelon, 2022).

Additionally, the authors conclude that it is difficult to comprehend the risk factors on the conceptual level, followed by comparing them all with each other. It is likely that the respondents have never directly interacted with the larger definitions and possibly never heard of them before, such as with "AI decision making", which is currently on the frontier of possible future implementations in comparison to normally used safety systems.

9.2.3 Time constraint

Although the methods used throughout this research answer the research question and research aims, a level of uncertainty still exists whether this opted methodology was the best-chosen procedure. Since a vast amount of research has been done on maturity models, weighting methods, and mathematical models, resulting also in many hybrid models, it is hard to determine which model fits best in this research. Furthermore, this study was accomplished in 9 months, which is not enough time to determine and test all the best possible solutions, to see which outcome is most beneficial for this research.

9.3 Future research

This study can be said to have furthered the understanding of safety and trust towards cobots in the field of maturity models. However, for future research, it is recommended that an alternative weighting method is implemented, due to the lack of results obtained for this study and based on the feedback received from respondents. Additionally, an Excel survey was used for this study and can still be recommended for future research; however, future research needs to shorten the time of completion to increase response rates. An alternative weighting model, that does not rely on the usage of Excel, will allow for easier automation, and interest of respondents during completion.

Furthermore, the study recommends having a strong knowledge of the sector of interest before using CSRAT or creating a similar assessment tool regarding other topics. Since industries are continuously developing and improving it is important to constantly update CSRAT or other assessment tools manually so that it's in sync with the latest technology norm within that sector of interest. It would also be of interest to work with a company, where completion of CSRAT or similar is paired with a product purchase, which would largely alleviate the respondent requirements and increase the sample size. In addition to that, it would allow industry experts and the ability to review similar/same cases of applications of cobots.

Implementation of ordinal Krippendorff's alpha for individual weights is recommended for the future, as pairwise comparisons have been proven lacklustre in providing significant results, which is in part due to a large number of comparisons. Individual weights do not have such a problem and would be the next stage to explore if continued with this research. Interval implementation would be most interesting due to scalar difference instead of regular spacing of ordinal Krippendorff's alpha.

As the simplification model for Krippendorff's alpha was implemented later in the study, possible variations of simplification were not considered, however, authors believe that alpha values would largely remain unchanged, maintaining that Krippendorff's alpha is not best suited for use on pairwise comparisons. It would be of interest to complete further research into the possible simplification of groupings of pairwise comparison scales for similar applications where a smaller scale is necessary.

While the optimisation was only completed for the grouped dimensions weights, it would be of further interest to complete similar optimisation of individual risk factors. A possible sampling could be implemented, where each dimension is used as a variable, which largely increases the complexity due to up to 7 degrees of freedom, however with automation, it is expected to not take longer than a few seconds per individual due to simplicity of linear algebra being used.

A further implementation of Lasrado's utility curve can be of further consideration, as this study has shown that a it has a perceivable effect depending on maturity level of implementation. In this study it was briefly consulted, yet never implemented. In the future research, authors suggest using:

- Maturity levels
- Intra-dimensional Weights
- Verification system (such as Inter-dimensional comparisons)

- Utility perception coefficient

10 CONCLUSION

This section will summarise the key research findings concerning the research aim and research question, as well as the value and contribution thereof. This research aimed to answer the main research question of whether it is possible to obtain a general weighted maturity model, where in future the participants would only have to complete the maturity grid and the weights are already pre-calculated, based on the single weighted model.

In order to gain a deeper comprehension of the areas that require improvement when it comes to ensuring the safe implementation of cobots in modern enterprises, the level of maturity of different risk factors of cobots must be determined to find risk factors of low maturity, such that the safety and trust towards cobots are increased. The weighted model provides an improvement by further showing which factors are at additional risk due to their importance, based on the opinions of engineers.

Although a single working model that would be applicable to all respondents was not found due to low agreement between respondents, this research has developed a method for calculating the weights of risk factors and implementing a verification model to enhance the consistency of responses on an individual level, thus creating weights on personal views.

This paper further explored the automation of survey results, identifying the optimisation and exploring the usage of Excel as survey media. While it was largely successful and decreased the processing time from 30 minutes to 0.1 seconds, the fact that it requires specific knowledge of programming makes it less flexible for less IT inclined, despite the program being a useful blueprint of future references.

Additionally, Krippendorff's Alpha has seen wide usage throughout the paper, however, the authors conclude that it is not suggested for pairwise comparisons, also given the fact that there is a wide range of opinions. It is suggested that the research is continued with the application of Alpha once again, but with the interval setting and on weights alone. The creators further suggest a collaboration with a cobot production company, which would allow them to contact engineers and managers directly, where survey completion can be worked into contracts for further alleviating the error of low sample sizes and lacklustre respondent profiles.

The topic of Lasrado's utility curve is briefly explored and concluded that in future research it can be a part of a larger implementation of a mathematical model to further cobot safety. Despite being a small part of the study, the results provide apparent proof that the importance and perceived utility of a risk factor is not the same.

This study can be said to have furthered research by modifying CSRAT and reviewing the risk factors used in cobot safety, as well as providing an optimisation technique that would allow for possible strategy and order of things of specific sub-dimensions that need work to be in line with 4IR integration and high standards of safety of cobots, providing further benefits to corporations and works that use them.

11 BIBLIOGRAPHY

Abedi, M., Torabi, S., Norouzi, G.-H., Hamzeh, M., & Elyasi, G.-R. (2012). PROMETHEE II: A knowledge-driven method for copper exploration. *Computers & Geosciences*, 255-263. doi:10.1016/j.cageo.2011.12.012

Adams, B. (2017, November 17). *The AHP Pairwise Process*. Retrieved from Medium: https://medium.com/dlprodteam/the-ahp-pairwise-process-c639eadcbd0e

Berdie, A. D., Osaci, M., Muscalagiu, I., & Barz, C. (2017). A combined approach of AHP and TOPSIS methods applied in the field of integrated software systems. *IOP Conference Series Materials Science and Engineering*, *200* , p. 012041. doi:10.1088/1757-899X/200/1/012041

Berx, N., Adriaensen, A., Decré, W., & Pintelon, L. (2022, July 1). Assesing System-Wide Safety Readiness for Successful Human-Robot Collaboration Adoption. *Safety, 8*(3). doi:10.3390/safety8030048

Berx, N., Decré, W., Morag, I., Chemweno, P., & Pintelon, L. (2022). Identification and classification of risk factors for human-robot collaboration from a system-wide perspective. *Computers & Industrial Engineering, 163*, 107827. doi:10.1016/j.cie.2021.107827

Bley, K., Pappas, I., & Strahringer, S. (2021). INNOVATION CAPABILITY IN SMALL INDUSTRIAL COMPANIES -A SET THEORETIC APPROACH TO MATURITY MODELS. *European Conference on Information Systems.* Online: ECIS.

Buchert, J.-m. (2021, October 6). *Cobot vs Industrial Robot: Differences and Comparison*. Retrieved from Man+Machines: https://manplusmachines.com/differences-robots-cobots/

Bunruamkaew, K. (2012). *How to do AHP analysis in Excel*. Retrieved from University of Tsukuba - Division of Spatial Information Science: http://giswin.geo.tsukuba.ac.jp/sis/gis_seminar/How%20to%20do%20AHP%20analysis%20in%20Excel.pdf

Castro, & Santiago. (2017). Fast Krippendorff: Fast computation of Krippendorff's alpha agreement measure. *GitHub repository*. Retrieved from https://github.com/pln-fing-udelar/fast-krippendorff

Chivers, I. (2015). An introduction to Algorithms and the Big O Notation. In I. Chevers, *Introduction to Programming with Fortran: With Coverage of Fortran 90, 95, 2003, 2008 and 77* (pp. 359-364). Springer Science & Business Media.

Chung, J., & Lee, Y. S. (2023). The Evolving Impact of Robots on Jobs. *ILR Review,, 76*(2), 290–319. doi:https://doi.org/10.1177/00197939221137822

Cihan Ünal, C. S. (2022, August 2). Application of the Maturity Model in Industrial Corporations. *Sustainability*. doi:10.3390/su14159478

Çınar, Z. M., Zeeshan, Q., & Korhan, O. (2021, June 11). A Framework for Industry 4.0 Readiness and Maturity of Smart Manufacturing Enterprises: A Case Study. *Sustainability, 13*(12), 6659. doi:10.3390/su13126659

Ewa, R. (2013, January). Rank Ordering Criteria Weighting Methods – A Comparative Overview. *Optimum.Studia Ekonomiczne*, 14-33. doi:10.15290/ose.2013.05.65.02

Frąckiewicz, M. (2023, April 1). *The Role of Artificial Intelligence in Collaborative Robots (Cobots)*. Retrieved from TS2: https://ts2.space/en/the-role-of-artificial-intelligence-in-collaborative-robots-cobots/

Franklin, C. (2021). Cobot Safety. *Presentation for The Association for Advancing Automation.*

Freelon, D. (s.d.). *ReCal for Ordinal, Interval, and Ratio Data (OIR)*. Retrieved from Deen Freelon, Ph.D.: http://dfreelon.org/utils/recalfront/recal-oir/

Helgesson, Y. Y., Höst, M., & Weyns, K. (2011, August 25). A review of methods for evaluation of maturity models for process improvement. *Journal of Software Maintenance and Evolution: Research and Practice, 24*. doi:10.1002/smr.560

Isa, M. A., Saharudin, N. S., Anuar, N. B., & Mahad, N. F. (2021, July). The application of AHP-PROMETHEE II for supplier selection. *Journal of Physics: Conference Series, 1988*(1), 012062. doi:10.1088/1742-6596/1988/1/012062

Javaid, M., Haleem, A., Singh, R. P., Rab, S., & Suman, R. (2022). Significant applications of Cobots in the field of manufacturing. *Cognitive Robotics*, 222-233. doi:10.1016/j.cogr.2022.10.001

Ji, P., & Jiang, R. (2003, July 24). Scale transitivity in the AHP. *Journal of the Operational Research Society, 54*(8), 896-905. doi:10.1057/palgrave.jors.2601557

Khaira, A., & Dwivedi, R. (2018). A State of the Art Review of Analytical Hierarchy Process. *7th International Conference of Materials Processing and Characterization, March 17-19, 2017. 5*, pp. 4029-4035. Materials Today: Proceedings. doi:10.1016/j.matpr.2017.11.663

King, J., & Kraemer, K. (1984, May). Evolution and organizational information systems: an assessment of Nolan's stage model. *Communications of the ACM, 27*(5), 466-475. doi:10.1145/358189.358074

Krippendorff, K. (2011, January 25). Computing Krippendorff's Alpha-Reliability. *s.l.* Retrieved April 14, 2023

Lasrado, L., Vatrapu, R., & Andersen, K. N. (2015). Maturity Models Development In Is Research: A Literature Review. *System design for, with and by users.* Oulu, Finland: IRIS38. doi:10.13140/RG.2.1.3046.3209

Lasrado, R. K. (2016). A Methodological Demonstration of Set-Theoretical Approach to Social Media Maturity Models Using Necessary Condition Analysis. *PACIS 2016 Proceedings*, (p. 249). Retrieved from https://aisel.aisnet.org/pacis2016/249

Maier, A., Moultrie, J., & Clarkson, P. (2012, March). Assessing Organizational Capabilities: Reviewing and Guiding the Development of Maturity Grids. *Engineering Management, IEEE Transactions on, 59*, 138-159. doi:10.1109/TEM.2010.2077289

Mancar, S. A., & Gülleroğlu, H. D. (2022). Comparison of Inter-Rater Reliability Techniques in Performance-Based Assessment. *International Journal of Assessment Tools in Education, 9*(2), 515-533. doi:10.21449/ijate.993805

Odu, G. (2019, May 20). Weighting Methods for Multi-Criteria Decision Making Technique. *Journal of Applied Sciences and Environmental Management, 23*(8). doi:10.4314/jasem.v23i8.7

Powers, D. M. (2012). The Problem with Kappa. *Conference of the European Chapter of the Association for Computational Linguistics.* doi:10.13140/RG.2.1.3492.0487

Provan, D. J., Woods, D. D., Dekker, S. W., & Rae, A. J. (2020). Safety II professionals: How resilience engineering can transform safety practice. *Reliability Engineering & System Safety, 195*, 106740. doi:10.1016/j.ress.2019.106740

Rosenberg, B. D., & Navarro, M. A. (2018). Semantic Differential Scaling. In *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation* (pp. 1504-1507). Thousand Oaks: SAGE Publications, Inc.

Roszkowska, E. (2011). Multi-criteria Decision Making Models by Applying the Topsis Method to Crisp and Interval Data. *s.l.*

s.n. (s.d.). *Cobots and Collaborative Robots*. Retrieved from Editorial by Industrial Quick Search: https://www.iqsdirectory.com/articles/automation-equipment/collaborative-robots.html

s.n. (s.d.). *Weighting by pairwise comparison*. Retrieved from Geogrpahic Information Technology Training Alliance: http://www.gitta.info/Suitability/en/html/Normalisatio_learningObject3.html

Takahagi. (s.d.). *AHP (Analytic Hierarchy Process) Calculation software by CGI*. Retrieved from http://www.isc.senshu-u.ac.jp/~thc0456/EAHP/AHPweb.html

Thuluviel. (2010). *File:Transitive.PNG*. Retrieved from Wiktionary: https://commons.wikimedia.org/wiki/File:Transitive.PNG

Tran, N. H. (s.d.). *Why Response Rate is Crucial to online research - and how to keep it high*. Retrieved from Decsion Lab: https://www.decisionlab.co/blog/why-response-rate-is-crucial-to-successful-online-research-and-how-to-keep-it-high

Widmann, M. (2022, August 4). *Cohen's Kappa: What It Is, When to Use It, and How to Avoid Its Pitfalls*. Retrieved from THE NEW STACK: https://thenewstack.io/cohens-kappa-what-it-is-when-to-use-it-and-how-to-avoid-its-pitfalls/

Wolf, M. a. (1997, September 15). *Date and Time Formats*. Retrieved from W3: https://www.immagic.com/eLibrary/ARCHIVES/TECH/W3C/W3C_DATE.pdf

Wu, Z., & Tu, J. (2021, March). Managing transitivity and consistency of preferences in AHP group decision making based on minimum modifications. *Information Fusion*. doi:10.1016/j.inffus.2020.10.012

Wulf, J. a. (2020). Evolutional and Transformational Configuration Strategies: A Rasch Analysis of IT Providers' Service Management Capability,. *Journal of the Association for Information Systems, 21*(3). doi:10.17705/1jais.00613

Zaiontz, C. (s.d.). *Krippendorff's Alpha Basic Concepts*. Retrieved from Real statistics using Excel: https://real-statistics.com/reliability/interrater-reliability/krippendorffs-alpha/krippendorffs-alpha-basic-concepts/

FACULTY OF ENGINEERING TECHNOLOGY
GROUP T LEUVEN CAMPUS
Andreas Vesaliusstraat 13
3000 LEUVEN, België
tel. + 32 16 30 10 30
fet.groupt@kuleuven.be
www.fet.kuleuven.be