Robust concept development utilising artificial intelligence and machine learning

A case study at GKN Aerospace

Master's thesis in Product Development

Hugo Alfgården & Kevin Karlsson

DEPARTMENT OF INDUSTRIAL AND MATERIAL SCIENCE

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2024
www.chalmers.se

Master's thesis 2024

Robust concept development utilising artificial intelligence and machine learning

An explorative study

HUGO ALFGÅRDEN
KEVIN KARLSSON

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Industrial and Material Science
Division of Product Development
Chalmers University of Technology
Gothenburg, Sweden 2024

Robust concept development utilising artificial intelligence and machine learning
An explorative study
HUGO ALFGÅRDEN
KEVIN KARLSSON

Company Supervisor: Sören Knuts, GKN Aerospace
Academic Supervisor: Alejandro Pradas Gomez, Chalmers University of Technology, Department of Industrial and Material Science
Examiner: Ola Isaksson, Chalmers University of Technology, Department of Industrial and Material Science

Master's Thesis 2024
Department of Industrial and Material Science
Division of Product Development
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: An AI generated picture of the thesis title.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2024

Robust concept development utilising artificial intelligence and machine learning
An explorative study
HUGO ALFGÅRDEN
KEVIN KARLSSON
Department of Industrial and Material Science
Chalmers University of Technology

Abstract

The 80-20 rule suggests that design decisions significantly impact downstream effects, such as product cost, with many of these decisions made during concept generation. This early commitment limits the ability to make changes later in development. Early-stage design requires a variety and quantity of concepts, but designers often fixate on existing designs, limiting innovation. In the aerospace industry, the complexity of concept development and evaluation is particularly challenging. Therefore, this study seeks to explore how AI/ML methods can aid designers in the concept development process.

This thesis was initiated as a results of a 2023 internship at GKN Aerospace, which involved generating a concept for the sectioning and manufacturing of an existing part. Recognizing the intricacies of these phases, the authors explored the potential of AI/ML methods to enhance robustness in concept generation and evaluation.

The aim is to evaluate how GKN Aerospace can effectively integrate AI and ML into their product development workflows. This involves understanding current methodologies and identifying gaps to address before implementation. The focus is on leveraging AI and ML to streamline complex decision-making processes, ultimately providing actionable insights for robust, efficient concept design aligned with the Zero Defect paradigm in aerospace.
Additionally, the thesis identifies gaps in the organization that needs to be address before a possible integration, such as data quality and data secrecy.

The result, building on extensive interview studies and literature studies is that there is potential in incorporating AI/ML in concept development processes. Although, AI methods such as LLMs, still have limitations, including confidently producing incorrect results, a phenomenon known as hallucinations.

The conclusion is that Generative AI, design tools with integrated AI/ML methods, and LLMs still offers opportunities to simplify concept generation. LLMs can assist with ideation, creative reasoning, and cognitive task offloading. Fine-tuned LLMs, trained on internal documentation, provide instant feedback on less complex tasks, helping designers explore a broader design space, mitigate bias, and enhance knowledge, facilitating the development of robust design solutions.

Keywords: Product Development, Concept Generation, Concept Evaluation, Set-Based Design, Artificial Intelligence, Machine Learning

Acknowledgements

We extend our sincere gratitude to Johan Lööf, our manager at GKN Aerospace, for his invaluable support and guidance in setting up our master thesis. His commitment to fostering a conducive research environment greatly shaped our trajectory in the aerospace industry.

We also thank our dedicated supervisors, Sören Knuts and Alejandro Pradas Gomez, whose expertise and mentorship were crucial throughout our journey. Their constructive feedback and collaborative approach enriched the quality of our research.

Our appreciation extends to our examinator, Ola Isaksson, for his thorough examination and insightful suggestions that enhanced the quality of our work.

We would also like to express our heartfelt thanks to all the respondents who took part in the interviews. Their valuable insights and contributions significantly enriched our research, and we are grateful for their willingness to share their expertise.

This endeavor wouldn't have been possible without the collective support of these individuals. We are truly fortunate to have worked with such an exceptional team.

Thank you, Johan Lööf, Sören Knuts, Alejandro Pradas Gomez, and Ola Isaksson, for your indispensable contributions to the successful completion of our master thesis.

Hugo Alfgården, Kevin Karlsson, Gothenburg, May 2024

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

AI	Artificial Intelligence
CAD	Computer Aided Design
CAE	Computer Aided Engineering
DfM	Design for Manufacturing
DSE	Design Space Exploration
FMEA	Failure Mode and Effects Analysis
GAN	Generative adversarial network
GAS	GKN Aerospace Sweden
LLM	Large Language Model
MDO	Multidisciplinary Design Optimization
ML	Machine Learning
PD	Product Development
PDP	Product Development Process
SBD	Set-Based Design
TRL	Technology Readiness Level
VRM	Variations Risk Management
ZD	Zero Defects
ZDM	Zero Defect Manufacturing

Contents

List of Figures

List of Tables

1 Introduction

In this chapter, an introduction to the master thesis is presented.

1.1 Background

Performing engineering analysis is a critical component in the decision-making process, offering a systematic approach to solving complex problems and ensuring optimal solutions. By utilizing simulations, models, and empirical data, engineers can predict the performance and behavior of systems under various conditions. This predictive capability allows for informed decision-making, enabling engineers to evaluate potential outcomes, identify risks, and optimize designs before implementation. As a result, engineering analysis enhances project efficiency and effectiveness while significantly reducing the likelihood of costly errors and failures.

Furthermore, reviewing the reports from such analysis is the foundation for critical decisions, and any inaccuracies or omissions can lead to significant consequences. A thorough review process ensures that all assumptions, methodologies, and results are scrutinized and validated, thus enhancing the credibility of the findings. This process often involves cross-disciplinary collaboration, where experts from various departments within the organization, provide insights and identify potential issues that may have been overlooked. In general, engineering analysis and the meticulous review of its reports are crucial for informed, reliable decision-making and the advancement of engineering practices. This process is complex, thorough and time-consuming, and this process at GKN Aerospace is certainly no exception.

1.1.1 Introduction to GKN Aerospace

GKN Aerospace is a top-tier global provider specializing in airframe and engine structures, landing gear, electrical interconnection systems, transparencies, and aftermarket services. They deliver products and services to a diverse array of commercial and military aircraft and engine manufacturers, as well as other primary suppliers. They are present in 12 countries, employing about 16 000 employees. In total, they have 32 manufacturing locations, where a large part of their workforce is based in Europe (GKN Aerospace, 2024a). This study is focused on GKN Aerospace Sweden (GAS) with head office in Trollhättan and about 2000 employees (GKN Aerospace, 2024c).

GKN maintains strong partnerships with all the leading engine, airframe OEMs and their primary suppliers (GKN Aerospace, 2024d). Since 1930, GKN Aerospace has been consistently delivering advanced engine components to the worlds airplanes and rockets (GKN Aerospace, 2024b). In fact, their engine components exist in over 90% of the worlds passenger planes.

1.1.2 Aerospace Product Development

With their strong knowledge in product development, GKN Aerospace enable the creation of highly durable and complex products with long operational durations under high stresses, while ensuring both safety and performance. As a result, the lead time for developing products in aerospace is rather long, typically 5-10 years (Rabie Jaifer & Bhuiyan, 2021). Therefore it is interesting to see how these processes can be optimized and made more effective by utilizing artificial intelligence (AI) and machine learning (ML) tools.

At GKN Aerospace, an update of working methods aimed at achieving Zero Defect is underway. In this thesis, the goal is to investigate where the bottlenecks and gaps are in order to conduct Set-Based Concept Design studies in our Product Development. Here, concepts, features, and design parameters are varied at an early stage, and the understanding of how technical specifications can be fulfilled in the final product needs to be explored. When making these decisions, AI and ML solutions are sought to simplify the complex process of decision-making in design.

1.1.3 Summer Internship

The authors conducted an internship in the summer of 2023, which inspired the research topic and shaped the problem definition for this thesis. The assignment in question was related to concept generation and evaluation for the sectioning of an existing product that GKN produces. No initial information was given other than finding different ways of dividing the existing product, into smaller parts. A lot was thereby open and much up for own interpretation of how to go about this problem.

Initially, information was gathered from around the company about the considerations necessary for performing such a sectioning and later welding the sectioned pieces together. Without a specification sheet and with little to no prior knowledge of this type of technology, advancing the work was challenging. The information was gathered mostly through expert interviews and internal documentation, which eventually was compiled as several factors to consider, requirements and "no go zones".

Based on all compiled information, a brainstorming session were performed in order to come up with as many theoretical ways of sectioning the product as possible. The thought process were guided by "out of the box thinking" and no concept were initially considered too obscure. The different ways of sectioning were documented and afterwards, an initial non-structured elimination were performed in order to rule out concepts indicating low potential for success.

The summer internship was conducted over an eight week period and during this included information gathering, concept generation and evaluation. The concept generation consumed the majority of the time, yielding over 40 different concepts of sectioning the product. The sectioning of the product, were performed in the CAD-program NX, in order to visualize the different concepts and to understand how certain ways of sectioning, would interfere with different interface on the product. The different concepts were documented in a concept catalogue. This phase was the lengthiest, estimated at five weeks.

1.2 Problem Discussion

During the summer internship, we found that there are in fact a large amount of ways for sectioning the product and thereby concepts to consider. For each new concept, small adjustments could be made to the sectioning, making a another completely new concept. Cuts could be offset 0.001 mm or angled 0.001 degrees different, for example. We found that product development can be very complex and how can we find the best one and verify that it is the best when there are so many different factors to consider. Striving to fulfil one requirement often led to us moving further away from fulfilling another one. In addition, this was only considering the different factors we had found to be important. In real full scale product development projects, there are a substantial amount of additional factors to consider, making it a lot more complex.

The concept evaluation utilized elimination matrices, Pugh matrices, and Kesselring matrices to rank the different concepts by criteria. This criteria-based ranking was a tedious process and took about three weeks. We found that it could sometimes be complex to rank certain concepts as they sometimes were very similar. It was hard to say one was better than the other, with complete certainty.

During this process, we found ourselves wishing for a program capable of generating concepts based on a prompt and a specification sheet, a program fine-tuned on aerospace specific data able to help with ideation, concept generation, and concept evaluation. This, because the task proved time-consuming, requiring considerable consideration. This, because a majority of important considerations was niched to aerospace specific material, components and ”know-how” that was necessary to know in order to successfully.

With all the new AI-support emerging and already available, we thought the area of concept generation and evaluation had the potential to be improved and more efficient by implementing AI-support. With the new surge of interest and improvement in AI/ML methods, and specifically large language models, we thought that a large amount of the work we had done could be automated and more efficiently done.

At present, the problem to be investigated is that the amount of labor intensive work to derive data with sufficient quality in limited time is a limiting factor during product development. Furthermore, The idea to explore is how to make use of recent advancements in AI and ML to aid in the process of robust concept development at GKN Aerospace.

1.3 Aim & Research Questions

The overarching purpose is to evaluate how GKN Aerospace can effectively incorporate AI and ML into their product development workflows. The investigation will focus on comprehending the existing methodologies and best practices employed for finding robust concept solutions. The emphasis lies in exploring the potential of AI and ML to streamline and simplify the complex decision-making processes inherent in concept generation, and evaluation. The ultimate goal is to provide actionable insights and recommendations for leveraging AI and ML effectively in the pursuit of robust and efficient concept design solutions aligned with the Zero Defect paradigm in aerospace. In specific, the thesis aims to answer the following questions:

- **RQ1:** What are the main challenges when performing concept generation and evaluation?
- **RQ2:** What are the gaps that are needed to be filled for a successful implementation of AI/ML methods to improve robust conceptual design work at GKN?
- **RQ3:** What are GKN's current AI/ML capabilities?
- **RQ4:** How can tools from AI and ML be used to simplify concept generation, evaluation and to propose a robust design solution?
- **RQ5:** What AI based methods have a potential to improve robust conceptual design work at GKN and what are their limitations?

1.3.1 Objectives

The report wishes to perform a benchmarking study to investigate the following main points:

- Gain access to current state of the art of AI implementation in industry, and compare this to where GKN currently stands
- Learn from other organizations motivations for implementing AI.
- Learn from others lessons learned from companies being successful/unsuccessful implementing AI.

1.4 Delimitations

- Any AI/ ML model presented in this report will not be trained on company specific data from GKN.
- The thesis will not construct or design new AI/ML models.
- GKN Aerospace has a substantial amount of sensitive and confidential data. Consequently, the recording of interviews may not be feasible due to the critical importance of avoiding spreading of sensitive information beyond the company.

2

Theory

In this chapter the underlying theory is presented.

2.1 Product development process

(Ulrich & Eppinger, 2016) presents a widely recognized approach to a general product development process (PDP). According to their definition, a product is an item that an enterprise sells to its customers. Before products can be sold and delivered, they must first be designed and manufactured. Thus, product development encompasses all activities from identifying a market opportunity to the production, sale, and delivery of a product. At GKN, the view of Ulrich and Eppinger (2016) process exist on a high level indeed.

An Engineering Design (ED) process proposed by (Beitz et al., 1996) (see Figure 2.1), which is a process with an foundation in specifically mechanical engineering tasks and how to solve problems.

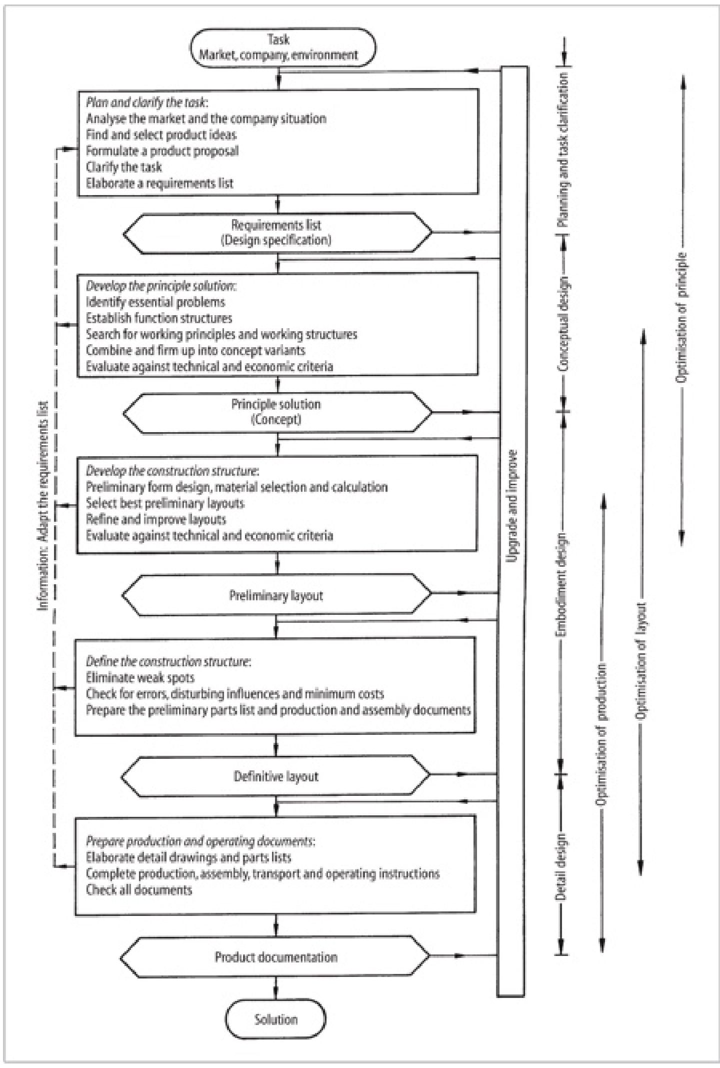

Figure 2.1: The ED process according to (Beitz et al., 1996)

Organizationally, GKN's model is much driven by the overarching phases common in aerospace PD. The aerospace PDP follows the stage-gate review model introduced by (Cooper, 1990). This model treats product innovation as a structured process that can be managed effectively. Like a manufacturing production line, the process is divided into distinct stages, each concluding with a gate. Each gate serves as a quality control checkpoint. To move to the next stage, the product must meet a specified set of deliverables and quality criteria at each gate. This structured approach helps focus on improving the process itself to enhance the quality of the output. The stages are dedicated to specific work, and the gates ensure that the quality is sufficient before moving forward.

2.1.1 Concept Development

The design process constitutes a series of technical activities within a product development process aimed at fulfilling the vision outlined by marketing and business strategies. It involves refining the product vision into technical specifications, developing new concepts, and conducting embodiment engineering of the new product. Companies can innovate by either acquiring existing brands or initiating new product development within their own research and development departments. (Kotler et al., 2005) defines new product development as "the creation of original products, improvements, modifications, and new brands resulting from the firm's own R&D activities."

GKN adheres to the conceptual design approach depicted in (Beitz et al., 1996) and incorporates Set-Based Engineering. This phase focuses on generating a range of broad solutions without committing to specific details initially. It includes breaking down the product into its fundamental functions, exploring various ways to fulfill these functions, and integrating these functions into working principles that dictate how the product will operate. Subsequently, the most feasible, innovative, and effective conceptual solutions are evaluated and selected.
In this report, the viewpoint that concept development entails generating multiple solutions and alternatives to solve a given problem, with the objective of reducing risk to invest in the effort to detail and define a product (Set-Based vs Point based) is taken.

2.1.2 Robust design

Products may deviate from intended designs due to variations in engineering processes. Robust design in industrial engineering aims to enhance product quality and reliability by minimizing such deviations (Park et al., 2006). GKN Aerospace aims for Zero Defects (ZD), the main idea behind the ZD manufacturing concept is not defects and faults detection, but rather faults and defects prediction and provision of suggestions on how those can be avoided. This can be achieved through a combination of Smart Inspection Tools, CPS, Data Analysis, and Knowledge Management tools, as well as Digital Twins (DT) (Nazarenko et al., 2021).

The term "robust design" has various definitions. (Taguchi et al., 2000) defined it as the state where the performance of technology, product, or process is minimally sensitive to factors causing variability, such as manufacturing or environmental factors, and aging, while maintaining the lowest possible manufacturing cost. (Suh & Suh, 2001) defines it as meeting functional requirements despite design and process variable tolerances. These definitions emphasize designing for insensitivity to variations.

Robust design theories, including the statistical Taguchi method, robust optimization, and axiomatic design, have evolved from existing design theories. They aim to make product performance insensitive to manufacturing process noises and uncertainties, especially in conceptual design (see Figure 2.2) (Park et al., 2006).

The significance of reducing variation in product characteristics was recognized early in Japan. Japanese engineer Genichi Taguchi's ideas were known in the 1940s (Arvidsson & Gremyr, 2008). He proposed a three-step strategy for product development: system design, parameter design, and tolerance design, emphasizing

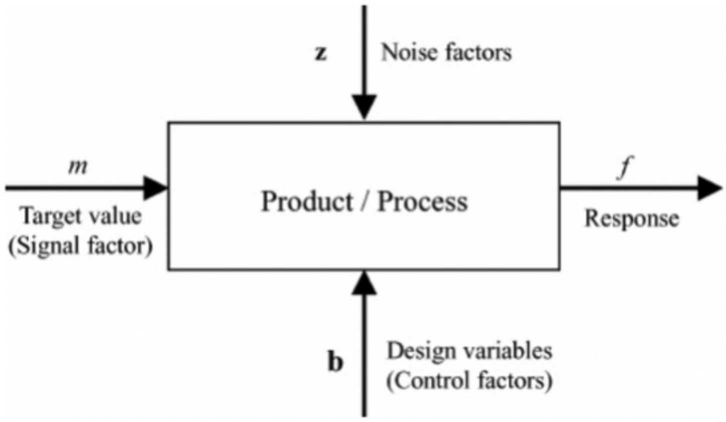

Figure 2.2: Product and process with inputs: Target value (m), Noise factors (z), Design variables (b) resulting in Response (f). (Park et al., 2006)

experimental methods in the latter two steps.

(Arvidsson & Gremyr, 2008) expands on Taguchi's framework, introducing terms like "robustness" and "robust design" and highlighting various methods to achieve robustness in engineering. It explains the fundamental challenge of variation in quality production and cites (Goh, 1993) emphasize the detrimental effects of variability and how important it is to reduce.

Noise factors, often labeled as external, internal, or unit-to-unit noise, are difficult and expensive to control (Taguchi et al., 2000). According to (Phadke, 1995), since noise factors cannot be easily controlled, designing for insensitivity to these factors is preferred over their elimination. (Taguchi et al., 2000) states that a design is robust when it is minimally sensitive to variability and aging at the lowest manufacturing cost. (Arvidsson & Gremyr, 2008) suggests this can be achieved by identifying factors that result in a more robust product or by redesigning the product. (Andersson, 1996) emphasizes that achieving robust design requires a wise choice of concept, using methods like adapted failure mode and effect analysis to account for noise factors, applying error transmission formulas to evaluate different designs, and utilizing design rules that contribute to robustness.

Various methods have been developed to support robust design. Variations Risk Management (VRM), as described by (A. C. Thornton, 2004), serves as an overall framework for reducing variation from system design to production. Additionally, Variation Mode and Effects Analysis, proposed by (Johansson et al., 2006), is useful for variation reduction in concept evaluation and selection phases, and for improving existing designs.

The robust concept design method developed by (Ford, 1996) involves defining the robustness problem, deriving guiding principles, synthesizing a new concept, and evaluating alternative concepts. Other general concepts like robustness and robust design, referred to in (Goh, 1993; Gremyr et al., 2003; A. C. Thornton, 2004; A. Thornton et al., 2000), do not prescribe specific methods for reducing variation like those related to Taguchi's work but view robustness as an engineering problem that can be solved in various ways. These concepts emphasize a product's characteristic of being robust, which relates to its insensitivity to variation.

From the robust design literature, the notion that robustness as an engineering problem is something that can be solved in various ways, together with defining the robustness problem, deriving guiding principles, synthesizing a new concept, and evaluating alternative concepts is important to account for in this work. In this thesis the viewpoint that having more alternatives and solutions reduces the risk of producing a non-robust design.

2.1.3 Set-Based Design

Set-Based Design (SBD), or Set-Based Concurrent Engineering (SBCE) is part of GKNs product development. The distinction between SBD and SBCE is that the term Concurrent Engineering represents a broad industrial development setting involving stakeholders from several organizational functions. SBD is the practical approach to engineering design following the principles of SBCE (Raudberget, 2015). The most important idea of Set Based approaches is to work with sets of plausible solutions, find limitations and constraints until a more narrow feasible design space can be investigated. For this, the following approach is applied (D. Sobek et al., 1999):

1. **Map the Design Space:**
   - Define feasible regions.
   - Explore trade-offs by designing multiple alternatives.
   - Communicate sets of possibilities.
2. **Integrate by Intersection:**
   - Look for intersections of feasible sets.
   - Impose minimum constraint.
   - Seek conceptual robustness.
3. **Establish Feasibility Before Commitment:**
   - Narrow sets gradually while increasing detail.
   - Stay within sets once committed.
   - Control by managing uncertainty at process gates.

In PD, the design phase and mapping the design space can be difficult in early project stages. As knowledge increases though a project and production approaches, design changes can become expansive. Research points to that the earliest decisions in PD have the largest impact on the overall quality of the product (effectiveness) and the overall cost of the project (efficiency). This was first documented by the authors (Clark, 1991).

Before SBD, the design teams used to iterate on one solution. The term "Set-Based" is contrasted with the term "Point-based" (A. Ward et al., 1995), which describes the traditional development methodology. This methodology is called "point-based concurrent engineering" (D. Sobek et al., 1999). Point-based engineering follows the ”do it right the first time” paradigm. Problems with this approach arise when engineers work concurrently with other team members. As the design is critiqued by different groups, each change triggers further changes and analysis, leading to rework and increased communication demands. There is no guarantee the process will converge, and many engineers report it often does not: the team stops designing when time runs out. Without a clear picture of the possibilities, the resulting design can be far from optimal (D. Sobek et al., 1999). Which in turn only is a waste of effort. In reality, the quality of information available in conceptual design phases does not allow the optimal design to be defined.

The picture in Figure 2.3 illustrates the process of point-based engineering. Where every hexagon is a separate activity in the design space.

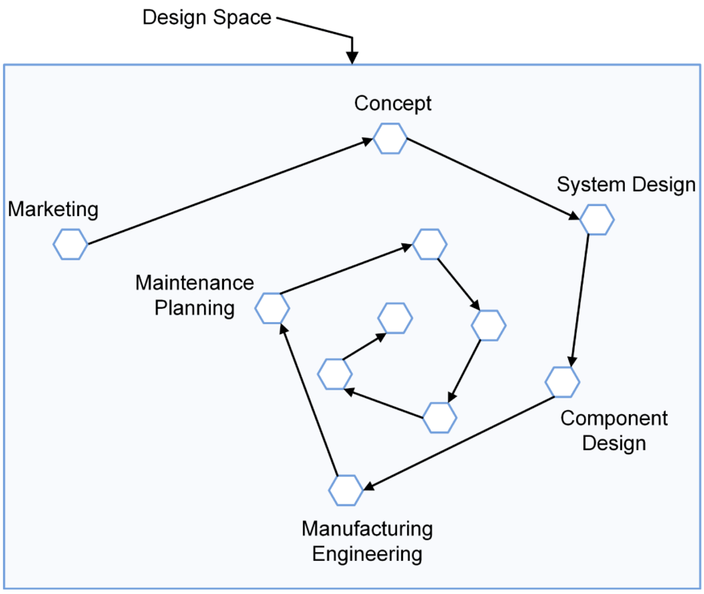

Figure 2.3: Point-based engineering explained (Toche et al., 2020).

In contrast, when engineering teams use SBD, engineers can agree on a range of parametric values instead of solidifying a single value at a time. SBD allows design problems and intricacies to be aligned gradually, offering the best projection (D. K. Sobek & Ward, 1996). The name is based on the fact that team members bring sets of possible solutions and compare them to find a practical intersection, rather than incrementally modifying a single option (Liker et al., 1996). The SBD stages are illustrated in Figure 2.4.
In short, SBD is characterized by developing multiple solutions to design problems in parallel, considering sets of design alternatives rather than a specific design. These sets are gradually narrowed down based on information from customers, manufacturing, tests, and other sources, leaving one solution in the end. A “set” represents different possibilities within the design space, holding multiple versions of elements with a common denominator, such as a family of design solutions, design variations, or manufacturing options (Raudberget, 2015).

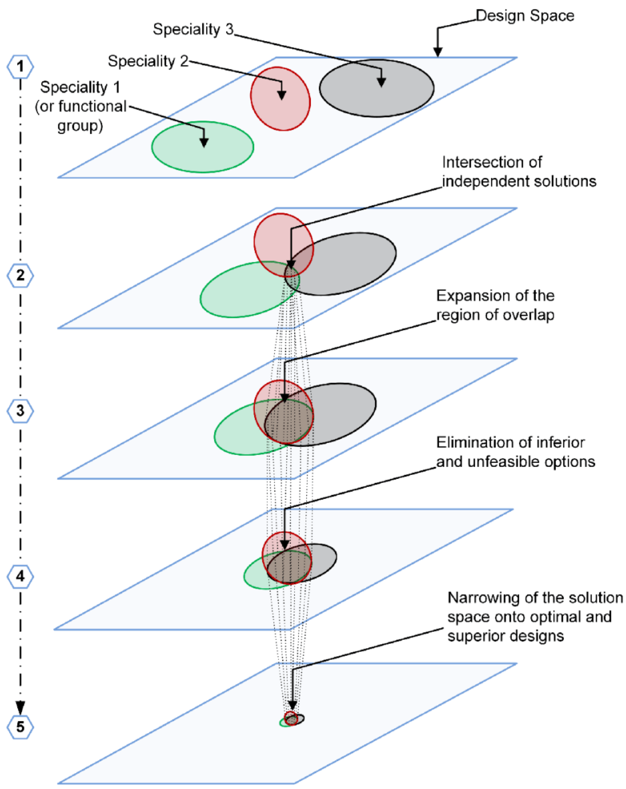

Figure 2.4: SBD engineering explained (Toche et al., 2020).

A benefit that comes with performing SBD is that it avoids narrowing down too quickly on areas of the design space, which, in return, ensures that potential innovation and discovery will not be lost (A. C. Ward & Sobek II, 2014).

2.1.4 Multidisciplinary design optimization

In the figure referenced in 2.4, the rings represent different areas of expertise. At GKN, Multidisciplinary Design Optimization (MDO) is employed to converge these diverse fields into a few optimal design solutions. GKN utilizes the Engineering Work Bench (EWB) program to conduct MDO, where it is possible to automate the MDO process (Madrid et al., 2021).

Originally, MDO is a mathematically rigorous way to formulate and solve optimisation problems dependent on multiple disciplines. However, today MDO has taken a lot of different meanings, for example (Meng, 2022) view MDO as a robust methodology for solving complex design problems involving coupled engineering systems and has garnered significant attention from both industry and academia.

Originally successful in aeronautics and astronautics engineering, the application of MDO has broadened. Typically, design variables and parameters in MDO are treated as deterministic inputs. However, in practical engineering scenarios, uncertainties are prevalent and unavoidable. For instance, the properties of materials, geometric forms, and external working loads on equipment are all subject to variability. In multidisciplinary systems, this uncertain information can propagate through the coupling relationships between different disciplines, leading to cumulative effects that significantly affect system performance and reduce both the reliability and safety of the system (as depicted in Figure 2.5 & 2.6.

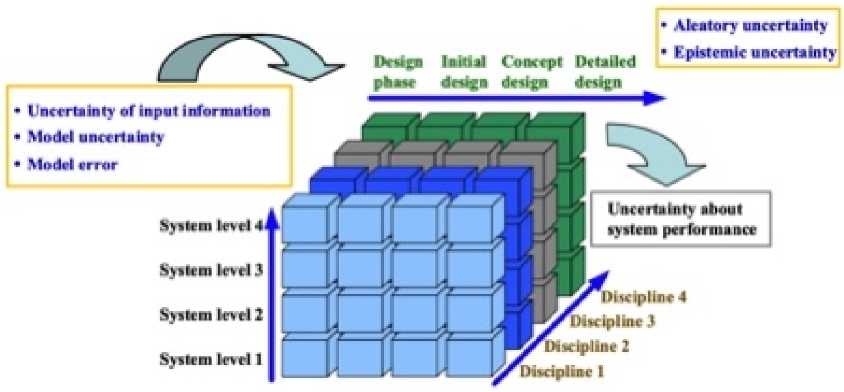

Figure 2.5: A complex system with multiple system levels, multiple disciplines, and multiple sources and types of uncertainty (Meng, 2022).

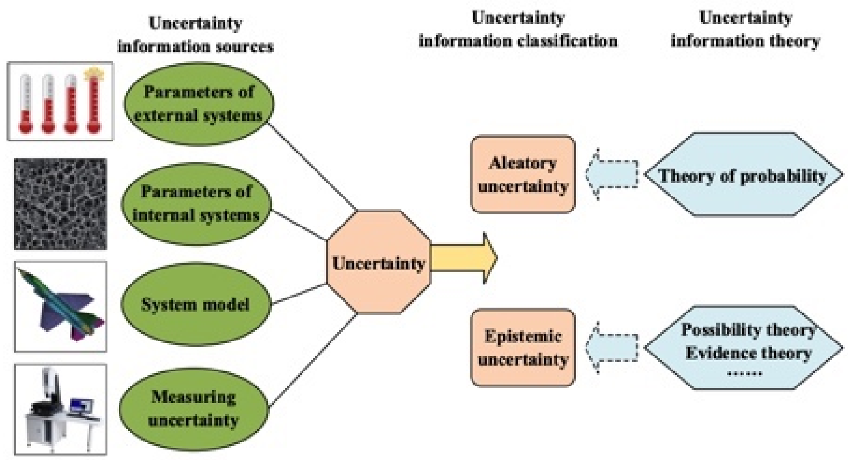

Figure 2.6: Uncertainty information sources and their classification (Meng, 2022).

2.1.5 Concept Evaluation

In the early stages of new product development, particularly after product concepts have been formulated, a critical step involves the evaluation of these design concepts. Proper evaluation is essential as it ensures that customer needs are met and contributes to the creation of products that genuinely satisfy customers (X. Wei et al., 2010). However, the concept evaluation phase comes with its challenges. It is a complex process where sometimes sub-optimal concepts may be initially selected (Dasari, 2021). This is particularly evident in the design of aircraft components, where evaluating a vast array of potential design alternatives involves procedures that are not only costly but also time-consuming, thereby hampering the swift identification of optimal design solutions. It is difficult, and expensive, to derive sufficient quality of data in conceptual phases, especially if different alternative solutions are investigated. Thus, finding effective strategies to streamline and enhance the accuracy of this evaluation phase is paramount in speeding up the development process while ensuring the selection of viable product concepts.

To properly evaluate a concept, gathering as much information as possible is desirable. Therefore, analysis and calculations are essential parts of the process. At GKN, a concept evaluation matrix is used to rank the concept based on its performance in certain areas such as risk and cost.

2.2 OMS

The Operational Management System, or (OMS), is a system located on GKN's intranet. Here, employees can check what is expected and required for specific tasks. It is also possible to see which role is supposed to handle each task. In the context of this report, particular attention is given to the product development workflow, with a focus on concept development and evaluation. The product development process is divided into four major phases in the following order:

1. Pre Study
2. Conceptual Design
3. Preliminary Design
4. Detailed Design
5. Final Design

Furthermore, for the purpose of this thesis, the greatest emphasis is placed on the flow within Conceptual Design. The workflow within Conceptual Design is depicted in Figure 2.7:

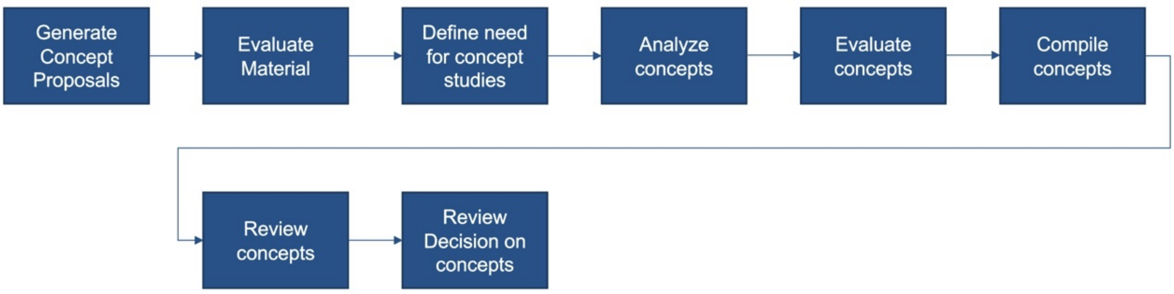

Figure 2.7: OMS process for concept development at GKN.

2.3 Product development methods

In this section, the methods used at GKN to generate robust concepts and evaluate them are presented.

2.3.1 Design for X

The Design for Manufacturing (DfM) and Design for Manufacturing and Assembly (DfMA) methodologies are part of the broader Design for X category. Their primary objective is to optimize the manufacturing and assembly phases of products (Formentini et al., 2022). DfX methodologies are employed to enhance particular aspects of the product in development. The variable ”X” is typically replaced with the specific aspect that need to be taken into account during design, (e.g. Manufacturing, Maintenance, Additive Manufacturing, etc.), and these methodologies are applied to support Engineering Designers during PD.

2.3.1.1 Design for Manufacturing

Presently, each aerospace design organization has developed their own design rules (Rajamani & Punna, 2020). The focus lies on improving product design capability by continuously optimizing. The use of Design for Manufacturing (DfM) or Design for Manufacturing and Assembly (DFMA) has led to simplifying designs at a lower cost and higher efficiency.

Previously, the approach to designing products started with an analysis of what the product was supposed to do, which gave the form and the materials of the product that are to be made. Thereafter the design was sent to the manufacturing department to be manufactured (Elwakil, 2019). A process that in theory is very simple, but in practice had a lot of drawbacks. Some of them where:

- Aesthetically pleasing design could have complex geometries that was usually considered impossible.
- Preparing the design without considering the tools to be used could result in the need for special, more expensive tools.
- As production volume increased, products had to be specially designed to reduce production time.
- If products shared the same manufacturing process, such as forgings, and the process was ignored in the design phase, the end result of the products could be faulty.

Due to the reasons mentioned above, and in line with a trend of combining manufacturing and modern design activities, a method of incorporating manufacturing considerations into the design phase took shape. The barriers between manufacturing and design departments are fading and will eventually disappear, leading to the emergence of *Design for Manufacturing* (Elwakil, 2019).

To provide context for what DFM does in practice, five specific actions describe DFM. These are: (i) the selection of raw material type, (ii) the selection of raw material geometry, (iii) the definition of dimensional and geometrical tolerances, (iv) the definition of roughness, (v) the characterization of specific shape constraints based on the manufacturing process, and (vi) the selection of secondary processing, such as finishing (Favi et al., 2016).

2.3.1.2 Design for Manufacturing and Assembly

Earliest recorded research of DFMA is back in the 1980's (Formentini et al., 2022). Additionally, there is Design for Assembly (DFA). DFA is a systematic method that aims to systematically reduce assembly time by (i) minimizing the total components in an assembly and (ii) eliminating critical assembly tasks (Boothroyd, 1987).

2.3.2 Zero Defects

Zero Defect Manufacturing (ZDM) endeavors to eliminate defects throughout the entire value stream (Wan & Leirmo, 2023). In many industries, the human factor is often perceived as the weakest link, contributing to variations and defects. Consequently, many industries have opted to replace the human factor with technology as a means of mitigating these issues (Welfare et al., 2019).

The zero defects concept originated as a quality program in the 1960s and adheres to the principles of Six Sigma and Lean Production (Halpin, 1966; Powell et al., 2022). ZDM was founded on the idea that the performance standard should be zero defects, compelling employees to reject any non-conformances (Powell et al., 2022). After its establishment in the U.S., Japan adopted these principles and made some modifications, eventually evolving into methods such as Taguchi methods, Six Sigma, Lean Production, and Total Quality Management (TQM). Although these methods became more prevalent than ZDM, the latter has re-emerged because it goes beyond traditional quality approaches, aiming for the complete elimination of defects. This is achieved not only through the detection and correction of defective products and process parameters but also through defect prediction and prevention, facilitated by its technology-intensive concept. With the introduction of Industry 4.0 in the 2010s, ZDM has thus gained greater traction due to the fact that it presents a digitally enhanced quality management.

The benefits of using ZDM are many, the reduction of defects and improved quality of parts is an competitive advantage, the reduction of scrap and unwanted parts improves the production sustianability. And the assurance of wanted quality in early stages of product development deprecates lead-times (Wan & Leirmo, 2023).

Machine vision in combination with convolutional neural networks (CNNs) enables in-line inspection with unprecedented accuracy (Smith, 2021). The power of this combination was demonstrated by Su et al. (2019) who integrated the technology with augmented reality (AR) to visualize assembly operations using object state and pose estimation. AR can also be utilized as a human-centric approach to assist operators in detecting errors in real-time (Zhao et al., 2022). The utilization of such innovations relies on the knowledge and competency of engineers to support technology integration (Wan & Leirmo, 2023).

Powell et.al has made a list that they consider to be key enabling technologies (KET) of ZDM (Powell et al., 2022). These are:

- Artificial Intelligence (AI): data-driven techniques for automated data analysis and decision making

- Architecture and Standards: integration and communication protocols of industrial software
- Big Data analytics: elaboration, analysis, and visualization of massive amount of industrial data
- Cyber-Physical Systems (CPS): control strategies combining physical and digital resources
- Internet of Things (IoT): multi-source distributed data gathering solutions
- Inspection and monitoring: solutions for the measurement and monitoring of product and process resources
- Simulation and modelling: solutions for the implementation of digital counterparts of product/process/systems (as Digital Twins, etc.)
- Extended Reality (XR): solutions for the integration of virtual and physical representations. (Powell et al., 2022)

Bäst källa hittills - (Fragapane et al., 2023).

2.3.3 ZDM strategies

In a paper published in 2019 by (Foivos Psarommatis & Kiritsis, 2020), a content analysis was conducted on 280 research articles published from 1987 to 2018 in various academic journals and conference proceedings. Based on this extensive review, the authors identified four distinctive strategies within ZDM: detection, prediction, prevention, and repair.

However, in later publications, researchers seem to have reached a consensus that there is a fifth strategy, namely, defect mitigation or compensation (Caiazzo et al., 2022; Fragapane et al., 2023; Powell et al., 2022). It has surfaced during the same study that the detection of defects is the most popular research interest, with 60% of publications related to the topic (Fragapane et al., 2023). The research interest in the other strategies is distributed as follows: Prediction 24%, Prevention 9%, Repair 4%, Defect mitigation or compensation 3%.

Detection in terms of ZDM refers to the activity of identifying non-conformity's, defects, and anomalies by classifying them based on the parameters that led to the undesirable result. This activity can occur at every step of a production process (Caiazzo et al., 2022).

Prediction on the other hand aims to forecast the quality of each part of the product before its production. The prediction is achieved by specific models and historical data analyses (Caiazzo et al., 2022). The methods are composed by applying mathematical modelling with AI technologies (Jagadish et al., 2019).

Prevention refers to the task of monitoring machinery by utilizing inspection tools and quality control. In contrast to detection, prevention involves observing the state of the machine rather than the product itself. This is achieved through machine state analysis so that the process conditions that could lead to defects in the product are identified proactively. Therefore, the expected outcomes are identified, and countermeasures are introduced to prevent the contribution of any defects (Powell et al., 2022).

Repair strategies involve reworking or remanufacturing products. These strategies are time-consuming and expensive for the company; therefore, the products have previously been treated as waste. However, with an increased focus on sustainability, repair has become an important aspect of ZDM since waste is not sustainable (Powell et al., 2022). For this reason, repair methods are optimized with the focus of reduced repair times without obstructing the overall production flow.

Defect mitigation or compensation entails modern strategies for mitigating or compensating defects aim to proactively identify defects or potential defects and seek methods to avoid rework. Defects and deviations are compensated for downstream in the process chain through feedforward control. The ZDM paradigm is based on integrating product and process data from multi-source process chains. Methodologies like stream-of-variation can be used to adjust the downstream process and prevent the propagation of dimensional and geometrical deviations in the measured part (Magnanini et al., 2019). In cases where a model-based solution is not feasible due to line complexity, specific compensation actions can be generated without the need for offline rework (Eger et al., 2018, Eldessouky et al., 2019). Additionally, in assembly systems, components may vary within predefined tolerances, which can result in a defective assembled product due to inherent variability in the parts. Selected assembly methods focus on matching components to minimize the expected deviation in the assembled product (Colledani et al., 2014a, Colledani et al., 2014b). These compensation strategies can reduce the need for rework or even end-of-line inspection, thereby reducing the production of scrap products.

For this thesis, prevention and defect mitigation are of great interest, as they are related to the idea of a robust design. However, it is noted that the literature found regarding ZDM strategies (and especially prevention and defect mitigation) do not adress the fact that design decisions could have a profound impact on downstream effects.

2.4 The fundamentals of Data Science

In the continuously evolving landscape of technology, the terms data science, data mining, artificial intelligence, machine learning and deep learning have emerged as crucial components of the digital era. These fields, while interrelated, possess unique characteristics and applications that distinguish them from one another.

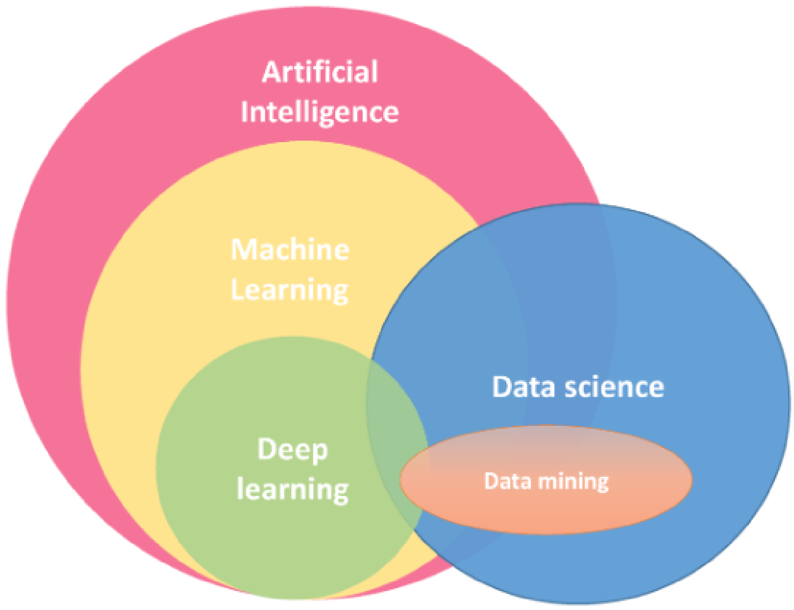

Figure 2.8: Relationship between Data science vs. Artificial intelligence vs. Machine learning vs. Deep learning, according to Kulin et al. (2021).

Figure 2.8, illustrates a venn diagram depicting the relationship between Data science, Data mining, Artificial intelligence, Machine learning and Deep learning. Each subject will be given a more detailed description in the sections below. However, this is just one way to illustrate the relationship between the different concepts as there are many ways to interpret these relations. For example, (Chen, 2022) presents an alternative way.

2.4.1 Data Science

Interest in data science is expanding quickly, with many viewing it as the profession of the future. Similar to how computer science emerged as a distinct field in the 1970s, we are currently seeing the swift establishment of research centers and undergraduate and graduate programs dedicated to data science. The field of computer science arose due to the accessibility of computational resources and the demand for experts in the area. Similarly, data science is now becoming prominent due to the widespread availability of data and the necessity for data scientists who can transform this data into valuable insights. The concept can be seen as a fusion of traditional fields such as statistics, data mining, databases, and distributed systems. As data science has its roots in statistics, some even consider "Data science" as a fancier word for statistics (Van Der Aalst & van der Aalst, 2016).

There has for a long time been confusion around the definition and what data science actually is and some argue a reason for this might be because of how it is intertwined with other data related fields (Provost & Fawcett, 2013a). One general way to describe data science is as the application to solve relevant problems and forecast future problems, utilizing qualitative and quantitative methods (Waller & Fawcett, 2013). Another is to describe data science as a data-driven discipline to use large sets of data in order to explain the complex behavior of a system, that can be hard to understand utilizing more traditional ways as modelling and simulation (Kulin et al., 2016). A third more detailed way to describe data science has been as a multidisciplinary field comprising the entire spectrum of data related activities including activities such as acquiring and storing data, analyzing and cleaning it, visualizing and interpreting the results, making data-driven decisions, with the aim to extract value from data and providing businesses with relevant insights (Kulin et al., 2021). The ability to view business problems from a data perspective is crucial, as data science extends far beyond algorithms for data mining (Provost & Fawcett, 2013a).

2.4.2 Artificial intelligence

Artificial intelligence (AI) is a science focused on developing intelligent machines that mimic human behavior, involving areas such robotics, natural language processing, information retrieval, computer vision and machine learning (Kulin et al., 2021). The idea is that machines created by humans are capable of a lot more than just doing labor intensive work, but also has the potential to develop human-like intelligence (Y. Jiang et al., 2022). Nowadays, artificial intelligence can today be considered being the forefront of industrial innovation, automating the processes of machines through functions like self-monitoring, interpretation, diagnosis and analysis. Manufacturers are today able to use methodologies related to AI, particularly machine learning and deep learning, in order to minimize operational downtime and predict future maintenance needs (Ahmed et al., 2022).

However, the field has been alive for years and there are many subdivisions. According to DARPA (2017), the evolution of AI can be divided into 3 waves.

- The first wave was about "handcrafted knowledge", and involved experts encoding their domain-specific knowledge into rules that a computer could process. This era saw the development of logistics programs, chess playing software and tax preparation programs. For instance, tax experts transformed the complexities of tax laws into commutable rules, enabling the computer to apply logical reasoning to specific scenarios. Logical reasoning like this and programs being able to take a particular fact of a concrete situation and work through is it typical of this first wave. However, these systems lacked the ability to perceive the external world, learn, or abstract knowledge to higher levels. They excelled at solving narrowly defined problems but were deficient in learning capabilities and handling uncertainty (DARPA, 2017).
- The second wave, characterized by "statistical learning," made significant strides in areas like voice and face recognition. These systems excel at perceiving the natural world, such as distinguishing individual faces, and can learn and adapt from specific datasets. However, their logical reasoning abilities are limited compared to first-wave systems. While second-wave technologies can classify data and predict outcomes, they lack contextual understanding, making them

statistically powerful yet often unreliable. This shortcoming sets the stage for the third wave (DARPA, 2017).

- The third way of AI takes future perspective on AI, acknowledging the challenges and limitations found in the previous waves. This wave of AI emphasizes contextual adaptation, where systems gradually develop underlying explanatory models to understand real-world phenomena. These systems will construct and refine these models over time, allowing them to perceive and interpret the world through these frameworks. By leveraging these contextual models, the AI will be able to reason and make informed decisions (DARPA, 2017).

AI can today be found everywhere, and is considered a significant force in transforming socio-economic lives and can be found in industry, healthcare, transportation, education (Y. Jiang et al., 2022). Some compare it to the introduction of internet and social media into everyday life and argues that AI will play just as an important role. Furthermore, AI will also play a role in corporate decision-making and interactions with external stakeholders like customers and employees. Humans coexisting with AI is a reality one must be ready for, and work to ensure this is done in harmony. Identifying which decisions is better of taken entirely by AI, which could benefit from human and AI collaboration and which are still better of completely to human, are important factors firms need to deal with (Haenlein & Kaplan, 2019).

2.4.3 Machine Learning

The popularity for machine learning continues to increase, largely driven by the massive and continuously growing volumes of data and computational capabilities (Badillo et al., 2020). Machine learning, a branch of artificial intelligence, focuses on creating algorithms capable of learning from historical data to enhance a systems performance (Kulin et al., 2021). Involved in machine learning, is identifying and learning from concealed patterns in sets of data and applying that knowledge for classification or predictive purposes (Alloghani et al., 2020). Two major domains of machine learning are supervised and unsupervised learning (Badillo et al., 2020)

Supervised learning is when the machine it taught using pairs of data, including input data and its corresponding output, refereed to as ”input-output pairs”. The algorithms build through supervised learning are thus requiring external assistance. Unsupervised learning is when the machine isn’t provided with a correct answer for each input. Algorithms are thus left to independently identifying structure and patterns in data. Unsupervised learning is mostly used in classification for feature reduction purposes (Mahesh, 2020).

2.4.4 Deep learning

Deep learning, a branch of machine learning, with a primary goal of achieving artificial intelligence (Dong et al., 2021). It works by guiding systems or machines in processing information across multiple layers, enabling them to classify, understand

and forecast results. Key deep learning techniques are Convolutional Neural Networks, Recurrent Neural Networks and Generative Neural Networks (Ahmed et al., 2022). A significant benefit of deep learning, compared to machine learning is its capability to identify complex, high-level features from data, autonomously.

2.4.5 Generative artificial intelligence

As a response to advancements in deep learning technology, "deep generative models" emerged, capable of creating novel content based on existing data. deep generative models that leverage neural networks have led to substantial improvements in the quality of generated content, marking significant advancements in the field of generative artificial intelligence (GAI) (Banh & Strobel, 2023).

Generative AI represents a computational technique and branch of AI capable of crafting new content, spanning texts, images, or audio, that is often difficult to distinguish from human creation. Consequently, GAI the capacity to revolutionize areas dependent on creativity, innovation, and knowledge processing. The widespread adoption of generative AI technology, exemplified by DALL-E 2, GPT-4, and Copilot, revolutionizing everyday work and how we communicate. Other than being used for artistic purposes, generative AI can also assist by taking the role of intelligent question-answering machine (Feuerriegel et al., 2024).

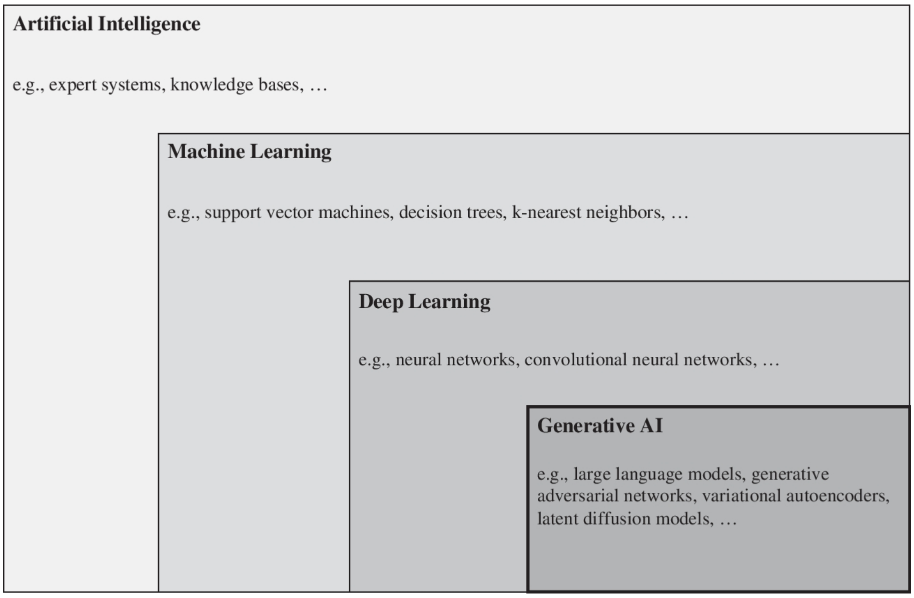

Figure 2.9: Illustration of where GAI exists in relation to AI (Banh & Strobel, 2023)

Worldwide attention were drawn to the field of GAI with the launch of ChatGPT, signifying a major advancement in the field. Additionally, even though GAI has been active for the past decade, ChatGPT ChatGPT's introduction ignited a renewed wave of research and innovation in AI. This momentum has resulted in the creation and release of various advanced tools, including Bard, Stable Diffusion, and

DALL-E. Tools like these demonstrate extraordinary capabilities, performing tasks such as text generation, music composition, image creation, video production, code generation. They are based on a range of cutting-edge models, such as Stable Diffusion, transformer models like GPT-3 and GPT-4, variational autoencoders, and generative adversarial networks. This progress in GAI opens up numerous exciting opportunities across diverse sectors, such as healthcare, business and media (Bengesi et al., 2024).

2.4.6 Natural Language Models

Natural language processing (NLP) is an application of AI, responsible for pioneering technologies like voice assistants, translation tools, chatbots and a diverse array of everyday tools. NLP focuses on the interaction between computers and human language, specifically on training computers to handle and analyze extensive amounts of text and other forms of natural language data. Artificial intelligence is utilized in NLP to absorb, integrate and interpret real-world data in a manner comprehensible to computers, irrespective of the language used. Similar to how humans having sensory organs like ears and eyes for hearing and seeing, computers possess input mechanisms for reading text and collecting audio. Just as humans rely on their brains to process sensory information, computers rely on programmers to process their inputs (Myers et al., 2023).

2.4.7 Large Language Models & Foundation Models

Part of the domain of NLP is foundation models and large language models (LLMs) (Myers et al., 2023). In recent years, large language models have gained significant traction, garnering attention both in academic circles and within industry, owing to their capacity to support in a multitude of tasks (Chang et al., 2023). LLMs belong to a category of foundational models that undergo extensive training on large datasets, enabling them with the capability to comprehend and generate natural language in order to execute a broad spectrum of tasks (IBM, 2024). Data can be text from articles, books and other information found on the internet, for example (Thirunavukarasu et al., 2023). In essence, LLMs utilizes deep learning technology (Shen et al., 2023) and mark a substantial leap forward in the field of artificial intelligence (Kasneci et al., 2023).

At the basis of LLMs, there lie transformer architecture (Kasneci et al., 2023). The Transformer architecture uses the self-attention mechanism to determine the relevance of different parts of the input when generating predictions. This allows the model to better understand the relationships between words in a sentence, regardless of their position (Vaswani et al., 2017).

Foundation models are the response to a shift in how machine learning conventionally are trained. Instead of constructing task-specific models from the ground up, pre-trained models known as ”foundation models” are customized through fine-tuning and then deployed to cater to a diverse array of domains. These foundation models facilitate the transfer and sharing of knowledge across domains, thereby reducing the necessity for task-specific training data (Mai et al., 2023). Foundation models thereby serve as an initial framework for the development of more specialized tools (Myers et al., 2023).

2.4.8 Model Tuning

Fine-tuning an LLM involves adapting a pre-trained LLM to a specific task or dataset by further training it on relevant data (Hu et al., 2023). Increasing the size of language models does not necessarily enhance their ability to follow a user's intent. For instance, large language models can produce outputs that are inaccurate, harmful, or unhelpful to the user (Ouyang et al., 2022). To tailor a general-purpose LLM for specific tasks, it must be trained on a task-oriented data. Supplementary training like this enables the model to fine-tune its parameters, aligning its capabilities with the particular task or domain of interest (VM et al., 2024). For example, FinGPT is trained and adapted to the domain of finance (X.-Y. Liu et al., 2023), whereas PMC-LLaMA is trained and fine-tuned to suit the medical domain (C. Wu et al., 2024).

While OpenAI models can address many use cases, there is a strong demand for domain-specific LLMs due to concerns about data privacy and pricing. By keeping the stakeholder's dataset and the LLMs on-premise, these models ensure data security. Additionally, fine-tuned LLMs offer high-quality, customized results and exhibit low latency in displaying outputs (VM et al., 2024). This two-stage approach naturally extends to solving tasks in private learning, effectively mitigating data scarcity concerns through the vast scale of the public pre-training dataset. In essence, a private dataset allows for local fine-tuning and alignment of the model (Yu et al., 2021).

2.4.9 ChatGPT

An example of a generative artificial intelligence chatbot is OpenAI's ChatGPT built on the transformer architecture. It was developed through fine-tuning of an LLM, and can thus be considered as an LLM application (Thirunavukarasu et al., 2023). Generative AI refers to a category of machine learning technologies capable of producing new content, including text, images, music, or video, by scrutinizing patterns within existing data (Brynjolfsson et al., 2023). This is categorized as GAI, enabling users to automatically generate content tailored to their individual needs (T. Wu et al., 2023). According to a report by UBS, ChatGPT achieved over 100 million monthly active users by the end of January 2023, a mere two months after its launch (Paris, 2023).

In 2022, leading technology companies worldwide introduced and refined a variety of GAI-products. Notably, OpenAI unveiled DALL-E-2, capable of generating high-quality images based on specific descriptions, while Meta introduced Make-A-

Video, enabling the direct translation of texts into videos. Towards the end of 2022, OpenAI launched the public version of ChatGPT, drawing global attention for its exceptional ability to accurately respond to human requests in natural language (T. Wu et al., 2023).

ChatGPT, an intelligent conversational agent, delivers detailed responses based on given prompts. As part of the GAI landscape, ChatGPT exhibits robust functionality across various language understanding and generation tasks, including multilingual machine translation, code debugging, story creation, error acknowledgment, and even rejection of inappropriate requests, as stated officially. Unlike its predecessors, ChatGPT can retain previous user input within a conversation, facilitating continuous dialogue. In March 2023, following the release of OpenAI's GPT-4, ChatGPT received substantial updates, allowing users to input both textual and visual data concurrently. This enhancement enables the completion of more complex multimodal tasks, such as image captioning, chart reasoning, and paper summarization (T. Wu et al., 2023).

2.5 Decision-making

Decision-making is one of the most important concepts when designing engineering systems (Eres et al., 2014). According to Kahneman (2011), there are two main types of decision-making: system 1 and system 2.

System 1 refers to the mode of decision-making characterized by automatic, effortless responses. These decisions are often shaped by habit and intuition, and they are typically made swiftly. Kahneman (2011) suggests that while intuition can be effective in certain situations, relying solely on System 1 for significant decisions can be precarious. This is because erroneous decisions made through this system may feel right at the moment, despite their potential lack of accuracy (Kahneman, 2011).

In contrast, System 2 is our deliberate, analytical mode of thinking. It requires conscious effort and mental energy to engage, often used for complex problem-solving and logical reasoning. System 2 helps us make careful decisions and override intuitive responses when necessary, ultimately resulting in more reliable decisions (Kahneman, 2011).

2.5.1 Data-driven decision-making

In today's business landscape, there's an abundance of data at our fingertips, driving companies across diverse sectors to harness it for a competitive edge. The sheer volume and diversity of data now surpass what manual analysis can handle, even pushing the limits of traditional databases. Meanwhile, the exponential growth in computing power, coupled with ubiquitous networking, has paved the way for sophisticated algorithms capable of integrating datasets for more extensive and profound insights than ever before. This confluence of factors has led to the widespread adoption of data science in business operations (Provost & Fawcett, 2013b).

Data-driven decision making involves making choices grounded in the analysis of data rather than relying solely on intuition (Provost & Fawcett, 2013b). In recent times, there has been a significant shift in data storage and processing technologies, opening up new avenues for data collection and utilization. This has prompted numerous managers to revise their decision-making approaches, moving away from reliance on intuition and towards a greater reliance on data-driven strategies (Brynjolfsson & McElheran, 2016). For example, data-driven decision-making in maintenance and operations work has spurred the advancement of novel methodologies and algorithms designed to aid engineers in making more optimal choices (Bousdekis et al., 2021).

Statistically, the more data-driven a firm is, the more productive it is (Brynjolfsson et al., 2011). However, even though there is an abundance of data available for decision-making, many decisions often rely on the experiences, opinions, intuitions, and various criteria put forth by product management and stakeholders. These decisions tend to be subjective, marked by inconsistency, and frequently lack transparent explanations or ties to the underlying data and evidence. Furthermore, decisions influenced by opinions, intuitions, and arguments are prone to political manipulation and personal biases, rather than being anchored in business opportunities or customer value. Even when data is incorporated into decision-making processes, an overflow of information can confuse decision-makers rather than providing clarity (Svensson et al., 2019).

3
Method

In this chapter the methodology used for the thesis is presented.

3.1 Methodology

This section describes the method used to achieve the aim and answer the research questions. Initially, an internal interview study focusing on the present state of concept development at GKN will be conducted. In order to learn the current perception, limitations and ongoing initiatives at GKN, an interview study with a range of different stakeholders at GKN was conducted. The interviews were planned as 60 minute recorded sessions, that were transcribed and analyzed using a set of guided questions Thereafter, a literature study projecting into the future state and exploring the opportunities of incorporating AI and ML into the processes was undertaken. Additionally, a second interview study took place with external companies in order to benchmark their progress in incorporating AI/ML tools into their product development processes and to understand how they have achieved it. Thereafter, a benchmarking study was conducted where other companies with similarities to GKN was interviewed to gain insights in industry trends, how far they have come with implementing AI/ML methods in their processes, and lessons learned from such implementations. Finally, a literature study was performed to gain knowledge about the current state of the technologies, its capabilities and limitations. The methodology process flow can be seen in Figure 3.1.
The authors Bell, Bryman, and Harley explain that qualitative studies are well-suited for investigating complex and context-specific problems in-depth, which contributed to the choice of conducting a qualitative study (Bell et al., 2019). Moreover, the authors Patel and Davidsson (2019) describe that a qualitative method provides a greater understanding of a subject, which formed the basis for the decision (Patel & Davidson, 2019).

However, qualitative studies can also involve the interviewer's interpretation and subjective judgments. The authors own notions and experiences can influence data collection, analysis, and interpretation of the gathered results. This can lead to the development of personal opinions and biases in the study. For this study, an abductive approach was chosen, meaning that existing theories and empirical data form the basis for the analysis. By incorporating aspects from relevant theory and conducting interviews with individuals with insights from industry and research, a breadth is created in the study. According to Dubois and Gadde (2002), an abduc-

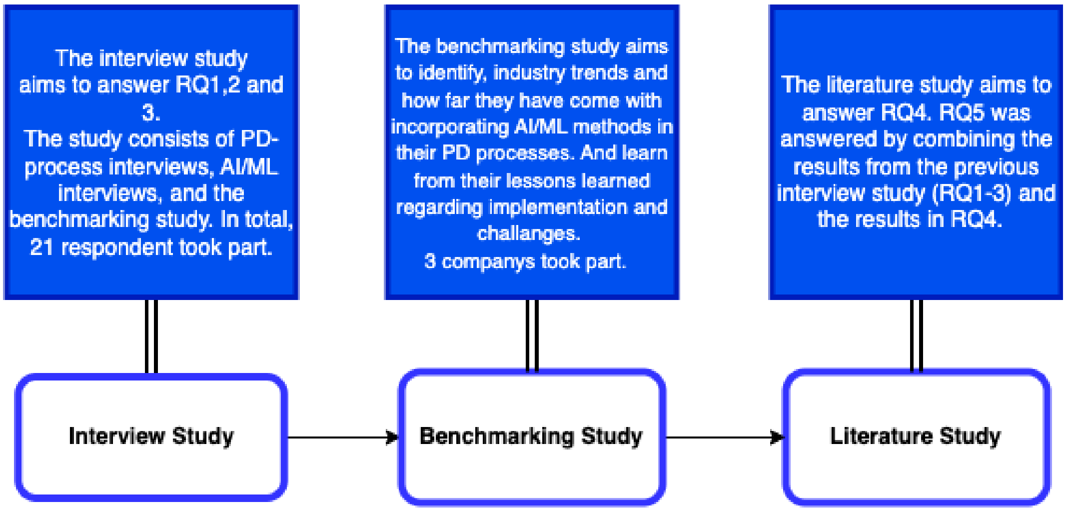

Figure 3.1: The methodology used to answer the research questions.

tive approach allows the purpose of the study and also the theory to evolve based on the empirical evidence (Dubois & Gadde, 2002). Furthermore, the authors explain that it enables the connection between theory, empirics, and interviews.

3.1.1 Study on current concept development processes and AI/ML initiatives at GKN

The very first step, before conducting the qualitative study, is to perform initial research of current product development and concept development processes at GKN. As well as AI/ML initiatives within the company. The material will be in the shape of documents, reports and presentations on the company's current work within AI/ML. The data gathered through the study will then be analysed, structured and categorised so that a view on the company's current state can be presented later on. The results of the study will help partly determine where the company is today in terms of what AI/ML technology they have available and how it is currently utilised. This, in order to develop questions for the first step in the qualitative study which are the interviews that are planned to be held with employees at the company.

3.1.2 Interview study

The initial phase revolves around understanding how concept evaluation and generation is presently conducted at GKN. This includes a detailed examination of the methods employed, the challenges faced, and the opportunities that exist within the current framework. The study aims to identify valuable lessons learned from past experiences in concept evaluation and generation.

The second phase involves interviewing respondents with AI/ML experience to gain insights into the gaps that currently exist at GKN, which need to be addressed before any potential implementation. Additionally, the study seeks to identify AI/ML tools that could be used in concept development at GKN and their limitations. Finally, the second phase aims to understand GKN's current capabilities regarding AI/ML.

The interview study will be conducted in a semi-structured manner and will be transcribed and recorded if accepted by the respondents. The number of respondents is determined by the amount of relevant information gathered. When it is safe to say that enough information has been reached to confidently determine a "best practice" in concept development, the interview study is considered complete.

The respondents to be interviewed are experts in their field at GKN and have extensive experience in concept development during the early stages of product development. Additionally, experts in the fields of AI, data science, and ML at GKN will be interviewed to gain a stronger understanding of the area. Lastly, it is interesting to interview respondents from other industries to learn about their experiences in this field and to benchmark their ways of working against those at GKN.

A crucial step in the interview process will be to as early as possible, establish contact with relevant actors. What actors that are considered relevant is decided by their roles and experience, and by the recommendations from our supervisors. Coordinating schedules in some people's calendars can be challenging, so the sooner we contact the right individuals, the better. We also need to be ready for the possibility of not being able to reach certain actors and have a backup option in this case. A second crucial step is related to interview questions. The better questions, the better the answers will be. Prior to the interviews, through research will be conducted in order to make the most out of the time of the interview. Both to save our time as well as the interviewees.

Respondents will be requested to provide consent for recording the interviews, ensuring that the study retains as much detail as possible for later transcription. Respondents can choose to remain anonymous; however, maintaining maximum credibility is crucial for the study, especially since the respondents are expected to be experts in their respective fields.

Similarly, for interviews conducted offsite at other companies, it will be made clear to the interviewees that they could be anonymous. Additionally, they will be told that the recordings will be deleted after the submission of the report.

In cases where consent for recording interviews is not granted, information will be documented through notes. If any sensitive information is disclosed during the interviews, it will be removed from the notes and will not be included in the final report.

The respondents that took part in the PD-process research are represented in the Table 3.1:
Additionally, two more interviews were conducted in order to gain more knowledge

1:st round interviews: PD-process			
Respondent Code	**Current Position Title**	**Years at GKN**	**Interview Time (min)**
R1	Principal Research Engineer	33	86
R2	Head of Design	26	38
R3	Team Lead Design	19	41
R4	Project Leader	25	41
R5	Design Lead	42	38
R6	Principal Research Engineer	24	46
R7	Project Manager	27	Recording not allowed
R8	Senior Design Engineer	37	66
R9	Design engineer	(7)	32
R10	Senior Design Engineer	26	55
R11	Principal Design Engineer	26	34

Table 3.1: The respondents that participated in the first round of interviews about the PD-process.

about which role or individual actually evaluates the concept and makes the decision to move forward with the actual concept. These interviews are shown in the Table 3.2 below:

Who decides upon the concept evaluation interviews			
Respondent Code	**Current Position Title**	**Years at GKN**	**Interview Time (min)**
R12	SVP	13	30
R13	Head of Industrial Architects	19	45

Table 3.2: Interviews that dowelled deeper into concept evaluation and deciding upon concepts.

The respondents for the second stint of interviews to gain knowledge about current AI/ML initiatives, infrastructure, and IT capability is depicted in Table 3.3:

3.2 Benchmarking study

The overall aim of the benchmark study is to see what is currently considered state of the art within the industry in terms of implementing AI in product development

2:nd round interviews: AI/ML			
Respondent Code	**Current Position Title**	**Years at GKN**	**Interview Time (min)**
R14	Analysis Lead	12	58
R15	CoE Manager	20	50
R16	Research Engineer	1.5	36
R17	CoE Director Engineering IT	5.5	34
R18	CDO	13	30

Table 3.3: The respondents that participated in the second round of interviews about AI/ML.

and compare this with where GKN currently stands. More specifically, it involves examining how it has been implemented in concept generation and evaluation, which could open up opportunities to partake in existing experience. If external actors have been successful/unsuccessful in implementing AI, what did they learn, and what obstacles did they encounter? Much of it is about sharing lessons learned that already exist around the area.

It is also of interest to gain a more strategic perspective and their view on AI and what motivated them to implement it. If they have a more critical perspective towards AI and can provide a motivation for why, this would also be an interesting result. What opportunities did they see/did not see?

The study will also verify or refute statements collected from our previous interview research. If statements arise that can be seen as contradicting previous results, it could be due to differences between industries, which would be interesting. In addition, the study will identify gaps in the implementation of this technology that GKN needs to address. It will also explore whether these gaps are similar across industries or not, which is also of interest to us. The respondents that took part in the benchmarking study can be found in Table 3.4.
To achieve this, an external interview study will be conducted at other companies, preferably aerospace companies, and/or companies that also mainly focus on hardware products with complex geometries.

Benchmark Interviews				
Respondent Code	**Current Position Title**	**Company Name**	**Years of Experience**	**Interview Time (min)**
R19	Senior Technology Engineer	GKN Fokker	6.5	40
R20	Manager Manufacturing Development	SKF	28	40
R21	Director in Development Quality	C3 (Confidential)	29	44

Table 3.4: The respondents that took part in the benchmarking study.

3.2.1 Literature review

The subsequent phase shifts the focus towards envisioning the future state of concept evaluation. The literature study aims to answer key questions regarding what AI/ML methods are applicable, process improvement, the integration of AI/ML technologies, and the establishment of a robust selection criterion for concepts. Consequently, there is a need to search, organise/describe and analyse AI/ML methods available. Within this context, the study seeks to identify opportunities for improving the concept evaluation and generation process.This, to later identify how AI tools could be applied in a correct way at GKN Aerospace.
To achieve this goal, screening will be conducted on various articles and scientific papers, carefully assessing their relevance to the searched topic through the platforms Scopus, Sciencedirect, Google Scholar, Aerospace Research Central and Chalmers Library. Literature selection will be primarily based on the text's title, and a thorough examination of the abstract and conclusion. If the content is deemed relevant to the research scope, the information from the literature will undergo further analysis and integration into the project.

To ensure the relevance and reliability of the literature, specific criteria were established. Given the rapid advancements in the field, the literature needed to be recent. Peer-reviewed scientific publications were of the highest interest and was favored when possible to ensure high quality, as they are a mark of credibility . However due to the recency bias articles that was very new was also included. Additionally, literature that was frequently cited or referred to was favored, as a high number of citations often indicates central importance to the field. Above all, the relevance of the literature to the study's topic was the primary criterion for inclusion.

To systematically manage the large volume of literature, a folder structure in Google Drive was utilized throughout the project. This folder contained and documented every article read and hold various details about each piece of literature, including the publication title, authors, year, search keywords, notes from reading, the decision of inclusion or exclusion, and the context in which it was used. This approach ensured an organized and efficient handling of information, facilitating a comprehensive and relevant literature review.

4

Results

In this chapter the result and analysis of the research study is presented.

4.1 Outcome of interview study

The following sections show the results from the performed PD-process, AI/ML, decision-making, and benchmarking interviews.

4.1.1 Outcome of PD-process & decision-making interviews

Research question one will be answered through the first round of interviews, based on the answers from respondents connected to the PDP such as design engineers and decision-makers. The respondents can be seen in Table 3.1 and Table 3.2.

As can be seen in Figure 4.1, the result from the interviews is compiled of 42 different challenges which have been categorized into 8 main categories: “Data secrecy“, “Time“, “Organizational synergy“, “Information“, “Human factor“, “Balancing demands“, “Conservative culture“ and “PD-process“. For some of the categories, subcategories has been created to specify variations withing the main category. Indexation/labeling of each challenge has been done in order to efficiently being able to reference to specific challenges in later in the report. Labeling follows the following logic: ”CD1” = ”Challenge Data secrecy 1” and ”CT1” = ”Challenge Time 1”, for example.

Red arrows has been added in order to distinguish between cause and effect.

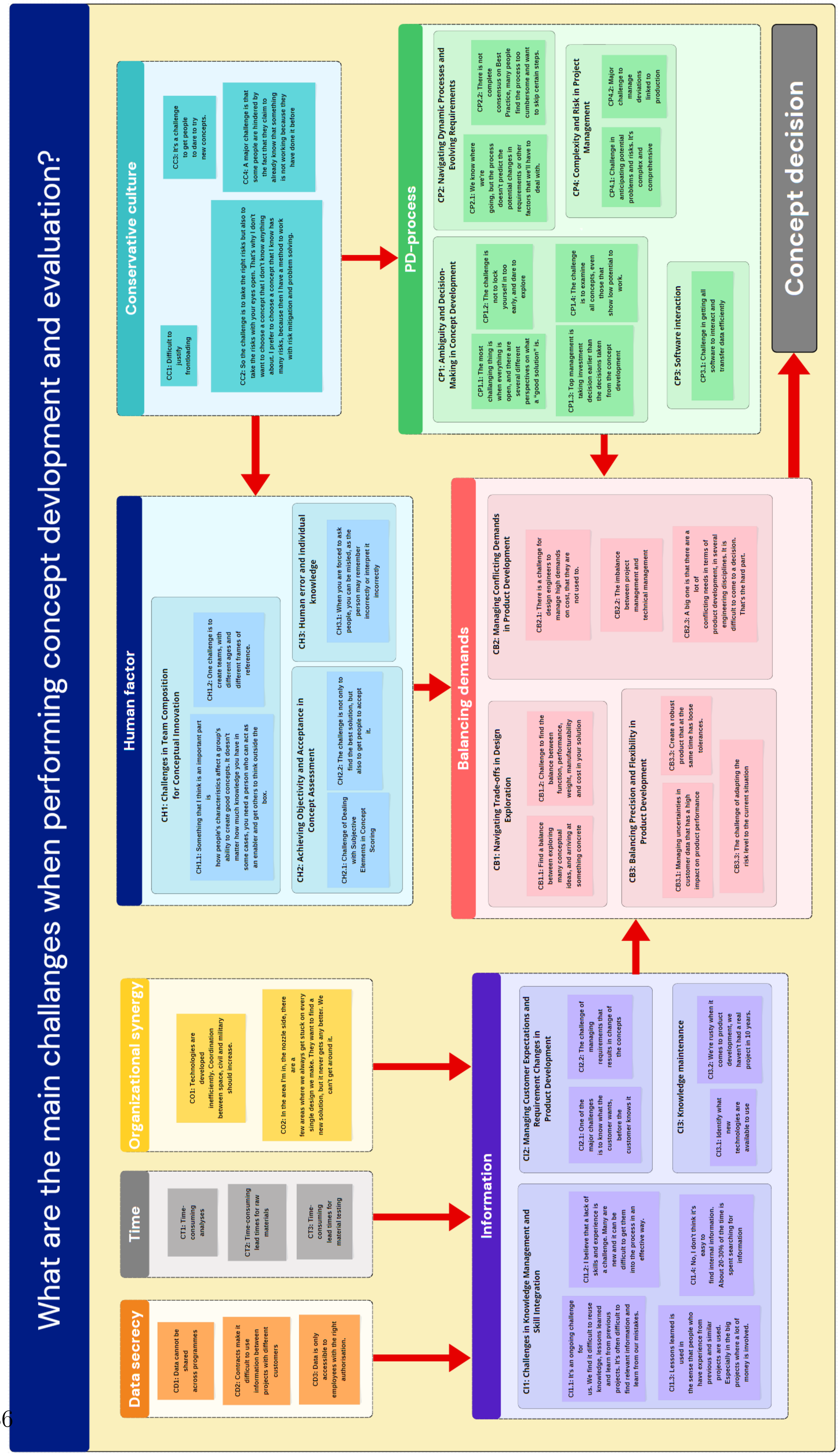

Figure 4.1: Aim diagram PD-process.

Data secrecy:

- CD1: There is a challenge centered around data and that some data cannot be shared across programmes [R1]
- CD2: There is a challenge centered around contracts and the opportunity to use information between different customers. This can have a inhibitory effect, as certain information gained from a project with one customer, cannot be used in projects with another customer [R2].

  > *"Unfortunately, we are sometimes a little slowed down by our agreements. Because what we have learned in collaboration with a customer, we cannot easily use in an assignment with another customer " - R2*

- CD3: There is a challenge centered around access to information, as not all information is available to everyone within the organization [R1]

  > *"This is especially difficult if you have worked at the company for a short time and do not have a wide network of contacts. In addition, there are limitations on information, especially when it comes to projects involving military aspects. Sometimes it is necessary to lock in and limit the availability of information. " - R1*

Time:

- CT1: There is a challenge centred around lengthy, time-consuming analysis of strength of materials and aero-performance. Some analyses can run for 3 months [R5]. On the design side, getting the right loads from the customer is absolutely crucial. If the loads are changed too late, analysis must be redone [R11].

  > *"The analyses are the most time-consuming. Whether it's for aero-performance or strength of materials. These analysis can run for 3 months." - R5*

- CT2: There is a challenge centered around lead-times for materials such as forgings ans castings [R5].

  > *"But another thing that takes a long time is to get materials, forgings, castings. That's where you have the longest lead time. From the time we have done the analyses, finished the model and then send it off to the caster, you need to wait 8-9 months before you get anything back. " - R5*

  This presents issues because despite still being in the concept phase, the material needs to be on-site within 12 months. The purchasing department endeavors to calculate backwards to synchronize this timeline, and it is frequently the case that the concept choice should have been made several weeks prior. Everyone wants enough time, which creates constrains right from the start [R1]. There is also a risk associated with this, as not not everything is decided in PDR when the orders are placed [R3].

- CT3: There is a challenge centered around time-consuming material testing for technology development, to generate required material data [R1].

  "Yes, material testing is often a lead time driver for technology and pre-development. If we don't have a material database, for example, then a material testing process can be very expensive and take a very long time." - R1

Organizational synergy:

- CO1: There is challenge centered around how new technologies are being developed. Today, the process is inefficient and there is a lack of technical management that ensures coordination between different disciplines within the organization [R8].

  "We work inefficiently. There is no technical management to ensure that we develop technologies that ensure that we develop technologies in a good way, and that we coordinate between different projects in a good way" - R8

- CO2: There is a challenge centered around working with the nozzle for example, where several persistent challenges arise with each design iteration. Despite efforts to seek alternative solutions, improvements are hard to achieve. These obstacles tend to persist without resolution [R11].

  "In the area I'm on, the nozzle side, there are a few areas where we always get stuck on every single design we make. We want to find a new solution, but it never gets any better. We can't get around it" - R11

Information:

- CI1: There is a challenge centered around knowledge management and skill integration. There is a lot of information within the organization, but a lot of this knowledge is stored as experience held by the personnel [R10, C1.3]. This might require someone to look for documented lessons learned instead. However, finding information and lessons learned is generally difficult and about 20-30% of time is spent on searching for needed information [R10, CI1.1][R9, CI1.4].

  "It's like a jungle when you have to search for information. Of course, you can search for a few arbitrary words, but then it is important that you know the exact title of the document or who has worked on it. So no, it's not easy." - R9

  This becomes even more challenging for new employees as they are required to find the right people, as opposed to find documented information [R10, CI1.2].

- CI2: There is a challenge centered around managing customer expectations and requirement changes in product development. One of the major challenges is to know what the customer wants, before the customer knows it. Today, the focus is mainly to meet requirements, but in order to possibly gain a larger share of a program, increase revenue and gain advantage in future business negotiation, it would be beneficial to create products that supersede customer expectations [R2, CI2.1].

  > *“I think one of the big challenges is that we manage to understand what it is the customer wants before the customer has figured it out for themselves so that we can offer functionality that adds something extra, an added value. I think we are better at just living up to a set of requirements than figuring out how we can offer something more.“ - R2*

  Furthermore, it is commonplace for customer requirements to evolve throughout the development process, often necessitating the generation of new concepts that were previously deemed finalized [R4, CI2.2].

- CI3: There is challenge centered around knowledge maintenance. Finding lessons learned from previous projects and identify what new technologies are available is challenging, as there is no good source to turn to [R3, CI3.1].

  > *“The big challenges are to find lessons learned from previous projects and identify which new technologies are available to use. There is no good library to turn to, you have to know what has happened in previous projects, know which people to go and talk to“ - R3*

  Additionally, knowledge maintenance also becomes challenging as there can be several years in between product development projects [R13, CI3.2].

Human factor:

- CH1: There is a challenge centered around team composition for conceptual innovation. The quality of the concepts generated is heavily contingent upon the composition of the group. In some instances, the presence of an individual serving as an enabler within the group is crucial for fostering out-of-the-box thinking, regardless of the level of knowledge possessed by other members [R9, CH1.1]. However, creating teams with members of different ages and with different frames of references can be a challenge of its own [R9, CH1.2].

  > *“Another challenge is that you have quite poor spread in the ages because you have much the same thinking. If you are around the same age, you often have quite similar frames of reference. It might help a discussion, but it probably doesn’t help from a creativity perspective, I think“ - R9*

- CH2: There is a challenge centered around achieving objectivity and acceptance in concept assessment. Even though the process of evaluating concepts is standardized and supported by concept choice matrices, subjective elements tend to influence how different concepts are scored and rated. It is common for a project group to get attached to a certain, even though it might not objectively be the best[R1, CH2.1]. The challenge is also not only to find the best solution, but also to get people to accept it[R1, CH2.2].

  > *"But there can still be subjective elements in the scoring itself, so a common pitfall is that a project team falls very much in love with a concept." - R1*

- CH3: There is a challenge centered around human error and individual knowledge. As a lot of information is bound to individuals as experience, a common way to acquire new information is to ask experts in different fields. However, When compelled to rely on inquiries to individuals, there is a risk of being misled, as the individual may recall or interpret the information incorrectly [R3, CH3.1].

  > *"It is more common to seek out a person who has worked with it before and ask questions such as: "what problems did you have?". And then you get what that person remembered at the moment and then you take it to heart. There is always a risk that you miss things that are in lessons learned though by doing this. The person may remember wrong or I interpret it wrong because it is not well enough described " - R3*

Balancing demands:

- CB1: There is a challenge centered around navigating trade-offs in design exploration. Achieving a balance between exploring numerous concepts and arriving at a concrete solution entails navigating a deliberate process of ideation and refinement. A risk arises when one becomes fixed on something without thorough understanding or exploration. Finding a balance is essential [R1, CB1.1].

  > *"However, there is a risk that you lock yourself into something you have not fully understood or explored. There may also be a better solution around the corner, but a faster route has been chosen. On the other hand, one can be deeply engaged in conceptual and abstract ideas, wanting to explore a lot, but never arriving at anything concrete. It's about finding a balance" - R1*

  There is also a trade-off between function, performance, weight, manufacturability and cost and the trick is to find a balance between these factors [R2, CB1.2].

- CB2: There is a challenge centered around managing managing conflicting demands in product development. As economics is becoming a more important factor, design engineers need to manage new demands on costs that they are not used to [R11, CB2.1].

> *“In the project I’m involved in right now, we’ve demanded that the customer decide on the concept selection matrix. So we have included that in the technical spec. Right now they have set that the cost is 50% and then the other is distributed so that the sum is 100%.“ - R11*

Economic requirements are also mostly focused by project management, and there conflict can occur between project management and technical management [R8, CB2.2]. The vast amount of conflicting demands and needs in product development makes it challenging to come to a final decision [R11, CB2.3].

- CB3: There is a challenge centered around balancing precision and flexibility in product development. A major challenge involves securing accurate customer input data to ensure the correct scope is addressed and that the data accurately reflects the customer’s true preferences [R5, CB3.1].

> *“To get the right input data so you know which scope to solve. That is often the biggest problem. I mean, those weight factors I was talking about, getting them so that they really reflect what the company wants“ - R5*

Adapting the risk level to the current situation is also crucial, and it requires the courage to take risks early while endeavoring to minimize risks later in the development process. [R1, CB3.3]. Tolerances also creates challenges, where the goal is to create a product that is robust, while having loose tolerances [R11, CB3.2].

Conservative culture:

- CC1: There is a challenge centered around difficulty justifying front-loading. From a management perspective, more resources should be put into earlier phases of development, as this is where the important decisions are being made. The choice of concept dictates about 80% of the cost [R6].

> *“One obstacle is probably this frontloading that you have a hard time motivating and setting early. And thus may not have time to evaluate all aspects as you should, which means that you will have problems later on“ - R6*

- CC2: There is a challenge centered around risks associated to choosing a concept. The challenge is to take calculated risks while maintaining awareness of the potential outcomes. This is why it is preferable not to choose a concept about which there is little knowledge. It’s better to select a concept with known risks, as it allows for the implementation of risk mitigation strategies [R1].

> *"So the challenge is to take the right risks but also to take the risks with open eyes. That's why I don't want to choose a concept that I don't know anything about. I prefer to choose a concept that I know has many risks, because then I have a method to work with risk mitigation and problem solving" - R1*

- CC3: There is a challenge around getting people to dare dry new concepts. There is conservatism within the airline industry and this makes it challenging to dare look into and create new concepts [R7].

  > *"It's a challenge to deal with conservatism in the industry, and get people to dare something new" - R7*

- CC4: There is a challenge centered around people claiming they already know something is not working because they have done it before. However, this mindset can impede progress, as advancements in technology or alternative approaches may now render the idea viable [R9].

  > *"A big obstacle that I think I see sometimes is that you are hindered by having investigated a certain trail before. Then it becomes easy to conclude that no, we have tested it and it does not work. But you may not see that the processes have changed or that there are new materials or new approaches that make what was previously not possible possible now possible. " - R9*

PD-process:

- CP1: There is a challenge centered around ambiguity and decision-making in concept development. Early phases of development are characterized by large uncertainty and managing this is challenging. A lot is still open and and there are several perspectives on what a good solution is [R1, CP1.1].

  > *"The most challenging thing is usually when everything is open. Each team member has their own idea of what a good solution is, which means there are many different perspectives." - R1*

  Furthermore, when working with design and definition, one are at risk of locking oneself in to early. It is important to balance freedom and design exploration in early stages. It's about engaging people to take responsibility, but at the same time avoid rushing into action without proper planning [R1, CP1.2]. At the end of the day, the challenge is to examine all concepts, even those who show low potential to work [R8, CP1.4]. While the design team is working with concepts, they must handle the fact that top management is taking investments decisions earlier than the decisions taken from concept development [R12, CP1.3].

- CP2: There is a challenge centered around navigating dynamic processes and evolving requirements. There is always conflict between theory and reality. A process embodies an idealized depiction of how something should function, yet in practice, it serves as a framework for actual implementation. While the design team have a clear direction, the process cannot anticipate potential changes in requirements or other unforeseen factors that may arise [R1, CP2.1]. However, even when there are no unforeseen factors, people still still occasionally skip certain steps as the they find the process to cumbersome [R10, CP2.2]

- CP3: There is a challenge centered around getting all software to interact and transfer data efficiently [R1, CP3.1].
- CP4: There is a challenge centered around complexity and risk in project management. Anticipating potential issues is a significant aspect the work, often addressed through methods like P-FMEA. These analyses can be intricate and thorough, focusing on turning abstract ideas into practical plans. However, there are numerous pitfalls, particularly when manufacturing seeks rapid outcomes, risking the creation of hastily developed solutions [R1, CP4.1].

> *“Much of the work is about anticipating potential problems, and this often uses something called P-FMEA. These analyses can be complex and extensive, and it is very much about transforming abstract thoughts into concrete plans. There are many pitfalls, especially when manufacturing is striving for quick results and thus risks creating something hasty.“ - R1*

Managing deviations linked to production is also a major challenge [R6, CP4.2]

4.1.2 Outcome of AI/ML interviews

The questions *”What are the gaps that are needed to be filled for a successful implementation of AI/ML methods to improve robust conceptual design work at GKN?”* and *”What are GKN’s current AI/ML capabilities?”* will be answered by the second round of interviews, the AI/ML interviews with respondents within the field. The respondents can be found in Table 3.3. The interview study highlights several limitations and gaps that need to be addressed before implementing AI/ML methods in GKN’s conceptual design work.

The accompanying AIM diagram provides an overview of how the issues relate to one another, based on statements from respondents during the interview. The AIM diagram can be found in Figure 4.2 below.

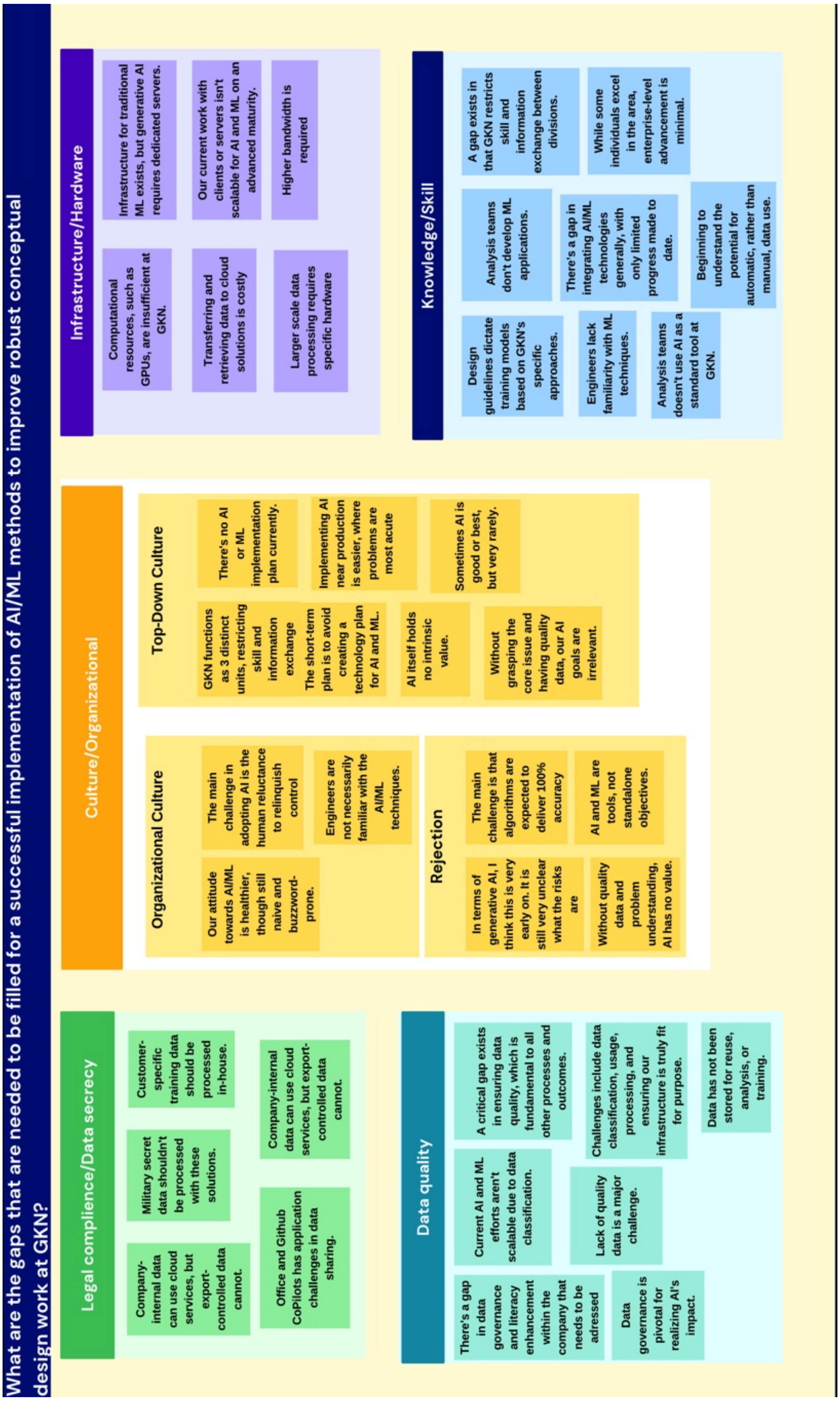

Figure 4.2: AIM diagram two that answers RQ2.

The limitations and gaps are organized into different categories. Citations for each statement will be presented below their respective category as follows. The categories are:

- Culture/Organizational
- Data quality
- Infrastructure
- Knowledge/Skill
- Legal compliance/Data secrecy

4.1.2.1 Culture/Organizational

The sectioning of Culture/Organizational can in turn be divided into the two subgroups, **Top-Down-Culture**, **Organizational Culture** and **Rejection**.The top-down culture exhibits reluctance towards adopting AI/ML methodologies until their value for the organization is proven. It emphasizes that their role is to ensure the company does not rush into adopting AI strategies. Furthermore, it suggests that numerous other improvements could be pursued instead of implementing AI/ML methods, questioning the relevance of AI for their business. Although this cautious approach is sensible, the authors sensed from the interviews a notable skepticism towards AI/ML at the executive level.

The organizational culture indicates that employees are uncertain about the company's direction concerning AI/ML strategies. They are aware of its use in production, and some are involved in its application, but broader awareness is limited. Furthermore, while some employees have significant knowledge and passion for AI/ML, others show little interest. Additionally, discussions about available AI/ML methods are not widespread throughout the company; instead, employees explore these topics independently.

This uncertainty is increased by the fact that there seems to be no clear plan for the implementation of these technologies. One employee explicitly stated:

> *"There is no direct implementation plan of AI or machine learning to my knowledge." - R14*

While employees are cognizant of AI's deployment in production and some are actively engaged in its applications, there is a general lack of widespread knowledge. Moreover, discussions on AI/ML methodologies are not commonly shared across the organization. Instead, individuals tend to explore these topics on their own, and plans for a structured approach to technology integration are intentionally absent, as another employee noted:

> *"The short-term plan from my side is that we should not have a technology plan for AI and ML..." - R18*

This independent exploration of AI/ML is partly due to the company operating as three distinct entities, which limits skills and information sharing. However, there are signs of improvement and potential for better integration [R17].

Despite the presence of enthusiasm and substantial expertise in AI/ML among certain employees, there is a noticeable disparity, with others showing minimal interest. The healthiest attitudes towards AI/ML still contend with naivety and the allure of buzzwords [R17]. One of the significant hurdles mentioned is the reluctance to shift from traditional methods to more automated solutions, as captured in an employee's observation:

> *"The main challenge of implementing AI is the human aspect of having to let go of control... One simply doesn't trust the solution, one would rather trust their Excel model. - R17"*

It has been observed that implementing AI technologies is most effective close to production areas, where issues are felt more acutely and can be addressed directly [R17]. However, organizational challenges persist as engineers are often not well-versed in AI/ML techniques, making integration and application more difficult [R14]. Furthermore, the effective use of AI is contingent on understanding the fundamental issues and having access to quality data. Without these, AI's potential cannot be realized, as emphasized by the following remark:

> *"AI has no intrinsic value. But if we don't understand the basic problem, we don't have the data that is quality assured, then it doesn't matter what we want to do with AI." - R17*

Moreover, while there are recommendations from sources like the Harvard Business Review on the need to introduce and transform using AI/ML, practical guidance on how and where to apply these changes remains elusive [R18]. In some instances, AI proves to be beneficial, but often, simpler solutions such as data visualization are adequate to resolve issues without the need for advanced AI applications [R18]. This highlights a selective utility of AI in the company, emphasizing the need for a more strategic and informed approach towards its adoption and use.

Rejection:
In the discussions surrounding the adoption of AI/ML technologies within the company, interviews revealed a noticeable skepticism, potentially linked to the cautious approach noted in the **Culture/Organizational** section. Several statements indicated a critical perspective toward the use of AI/ML.

One employee emphasized that without quality data and a deep understanding of the underlying problems, AI lacks value [R17]. This point was supported by another respondent who noted that the advantages of AI are often surpassed by simpler methods like data visualization [R18]. Additionally, an important reflection on the role of these technologies was shared:

> *"AI and ML are tools, not standalone objectives." - R18*

Technical challenges were also highlighted, especially concerning AI's capabilities in handling complex geometrical data. One respondent pointed out the difficulty of machine learning models to capture subtle details [R14]. The conversation also touched on the unrealistic expectations placed on AI performance, as one respondent criticized the demand for perfect accuracy:

> *"The main challenge is that algorithms are expected to deliver 100% accuracy, unlike humans." - R16*

Concerns about generative AI were discussed, particularly its early development stage and uncertain risks, including the potential for providing incorrect information confidently:

> *"In terms of generative AI, I think this is very early on. It is still very unclear what the risks are... That they sometimes give the wrong information confidently." - R16*

Furthermore, the context in which AI operates was noted as being full of external variables and caveats, suggesting that the technology's effectiveness is significantly influenced by surrounding conditions [R14].

Finally, the critical role of human judgment was emphasized, especially in ensuring the reliability and applicability of AI solutions. The need to maintain human oversight in decision-making processes to ensure product worthiness was highlighted [R14]. This collection of viewpoints presents a careful and thoughtful approach to integrating AI/ML into company processes, reflecting an organizational ethos of informed and strategic advancement.

The interview study underscores that organizational culture and structure are significant hurdles in adopting AI and ML technologies. Fragmentation within the company, a mixed attitude towards AI/ML, trust issues regarding new technologies, skill gaps among engineers, and the absence of a cohesive strategic plan for AI/ML adoption are key challenges identified. Addressing these cultural and organizational gaps is crucial for successful AI implementation.

4.1.2.2 Data quality

The interviews around AI within the company reveal an important realization: the true challenge lies not in the technology itself but in the foundational aspects of data management. This acknowledgment has come to the forefront, especially as AI pilots have underscored an inadequate understanding and control over the quality of data. One employee noted that data has not been stored for reuse, analysis, or training, which has been a persistent issue in AI implementations:

> *"Data has not been stored for reuse, analysis, or training." - R15*

In response to these challenges, the company has outlined a strategic focus on data quality, governance, and literacy. This approach is rooted in the belief that the success of any AI-driven output is contingent on the quality of the input data. Without high-quality input data, the output will inevitably suffer. As one employee put it:

> *"It's about ensuring data quality because that's what we will base everything else on. This is the most important part." - R18*

The importance of organized and accessible data is also emphasized. Much of the company's data, particularly analysis data, remains unsorted, with some exceptions like sensor data which is neatly categorized and prepped for analysis. This unsorted state presents an obstacle in efficiently leveraging big data for automated processes [R17]. Recognizing this, the company has prioritized the need to better manage and structure their data repositories. One employee succinctly highlighted the challenges associated with data:

> *"The challenges are often around data classification and how we can use the data and how we will process the data and how we should have an infrastructure that really is adapted for the purpose." - R17*

The strategic shift towards a more data-centric approach rather than a purely AI-focused strategy is informed by past experiences. The realization that an AI strategy without a solid data foundation is ineffectual has led to a clearer focus: solving core business problems through better data management rather than relying solely on AI solutions [R17].

However, the interest in AI is not entirely gone. During interviews, it was evident that despite limited knowledge about AI/ML, employees are excited about the possibility of offloading mundane or repetitive tasks to AI systems. This enthusiasm for automating "boring" or "non-stimulative" tasks underscores the potential benefits AI can offer in enhancing workplace efficiency and satisfaction.

Together, these insights indicate that the company needs robust data management practices to fully harness the capabilities of AI. Although significant strides have been made, there remains much work to ensure that data quality, governance, and literacy are up to the mark to fully realize AI's potential. The company's adjusted strategy reflects a mature approach to technological adoption, emphasizing that the foundation of data must be solid for AI applications to be truly effective.

4.1.2.3 Infrastructure

In discussing the challenges faced by the company regarding AI/ML implementation, the topic of infrastructure became an apparent bottleneck. According to the respondent, the company lacks long-term ownership of its code solutions, which poses a significant challenge as they consider how to manage and support their software over many years. The choice of stable and supportable frameworks is crucial for sustainable software development [R15].

The respondent further noted that data downloading speeds are a bottleneck due to limited internet bandwidth. At the company's site, despite having a 10 Gigabit internet connection, practical download speeds during peak office hours are much lower, sometimes around 5MB/s, which significantly slows down data analysis. This

limitation means waiting many hours for just one analysis to complete, which affects productivity [R15].
Additionally, the respondent discussed the infrastructure requirements for modern AI applications, especially when moving beyond traditional machine learning to more complex models like generative AI and LLMs:

> *"Mainly infrastructure and...traditional machine learning there are some...Servers that are able to train for a few hours, but if you go towards more generative AI and LLMs...Then you need dedicated servers" - R15*

Furthermore, the cost of moving and retrieving large datasets to and from the cloud was highlighted as a significant expense, while the actual training of models on this data could be relatively cost-effective:

> *"It's 'transferring' and 'retrieving' the data to the cloud that is expensive, while 'training' the model on the data will probably be quite cheap" - R15*

The conversation also touched upon the limitations in current hardware capabilities within the company, emphasizing that the existing infrastructure, often reliant on individual servers or client systems, is not suitable for scaling up AI and ML operations to a more advanced level:

> *"If we talk about the process of data on a larger scale, you need certain hardware that is no longer a standard laptop. What we do today is often on individual clients or on individual servers and is not scalable up to a higher level of maturity for AI and ML." - R15*

These insights from R15 clearly illustrate the technological and infrastructural challenges the company faces. The need for substantial upgrades and investments in IT infrastructure is critical to enable the company to fully leverage modern AI capabilities and manage data more effectively. These challenges necessitate thoughtful planning and investment to bridge these gaps and prepare the company for future technological demands [R15].

4.1.2.4 Knowledge/Skill

For successful implementation, knowledge and skill about the topic must already be established within the company. The interviews revealed that this knowledge is currently very limited to a few individuals who are far ahead of the greater mass. They also showed that the lack of knowledge about the topic and methods is extremely limited among employees working in product development, for example. However, there is excitement about using AI/ML as an assistant for mundane tasks.

This section could be divided into two distinct parts. The first part focuses on **AI/ML expertise**, highlighting the need for more human capital in the area of AI/ML implementation on a large scale and to advance research.
The second part addresses the **Lack of Knowledge**. Many employees regard AI as a buzzword, yet it is evident they possess limited understanding of what is required,

how it works, and what the methods are capable of. It is crucial that knowledge in this area is expanded before implementation to ensure that everyone shares the same, informed perspective on AI/ML. This will help them be aware of the methods, limitations, capabilities, risks, and benefits.

Reluctance to relinquish control remains a significant barrier to the adoption of AI technologies within the company. One respondent noted that while there is a growing openness to AI and ML, there remains a pervasive sense of naivety and a tendency to get caught up in the allure of buzzwords without a deep understanding of the technology [R17]. This shows a lack of information of AI/ML and their capabilities. In discussions about the progress of integrating AI into more routine and structural areas, such as reporting, the implementation appears to be very preliminary. A specific point made was that the use of AI for structural analysis is not currently standard practice at GKN, indicating a significant area where AI integration could potentially be expanded:

> *"As a standard tool and for structural analysis, it is not considered at GKN." - R14*

Moreover, the adoption of AI and ML within various departments faces additional hurdles due to a lack of familiarity with these technologies among engineers. This organizational challenge compounds the difficulty of integrating advanced AI applications across the company:

> *"Organizationally, I think it's difficult because the engineers are not necessarily familiar with the AI/ML techniques." - R14*

Furthermore, when asked about the overall progress of AI integration within the company, it seems that, while there are individuals with advanced skills and understanding, the broader organizational integration of AI is still nascent. The focus has recently shifted towards a more realistic appreciation of how data can be leveraged effectively and automatically, moving away from manual processes. One respondent articulated this transition, emphasizing the modest progress at the enterprise level:

> *"I would say that we haven't made much progress if you think about it on a level of integration. Then, of course, there are individuals who are far advanced, very competent. But on an Enterprise level? No, I would say we have just sobered up to our view on how data can be used more effectively and above all automatically, and not do it manually." - R17*

4.1.2.5 Legal compliance/Data secrecy

One of the major challenges facing GKN relates to the restrictions on where and how data can be used, particularly due to security and confidentiality concerns. Certain data cannot leave the premises of GKN, necessitating that all training of models must occur in-house. This requirement for local processing demands extensive infrastructure, complicating efforts to leverage cloud computing and other external resources. However, there is an important distinction to be made, as highlighted by one employee:

> *Company-internal data can use cloud services, but export-controlled data cannot. - R15*

In addition, there are strict regulations concerning data derived from military products. Such data is highly sensitive and, as another statement clarifies, must not be processed using external solutions:

> *Military secret data shouldn't be processed with these solutions. - R15*

Nevertheless, opinions within the company about the flexibility of data use for AI training vary. For instance, R17 suggests an alternative approach to circumventing some of these challenges:

> *"Absolutely outsource the training of models...it's really just the data that is sensitive and not the model, we can always anonymize the data, and far from all data is sensitive"*

However, the use of advanced AI technologies such as LLMs faces specific limitations due to export controls, as noted by another team member:

> *Export controls limit the use of local LLMs. - R16*

Furthermore, contractual obligations also restrict the use of information across different projects, especially when these projects involve multiple customers [R2]. This limitation affects the ability to aggregate and utilize customer-specific data efficiently. In terms of processing customer-specific data, there is a clear directive that this should be handled in-house [R14]. Despite this, there is potential to train models with anonymized data off-site, presenting a possible workaround to the stringent in-house requirements.

Lastly, ensuring compliance when dealing with regulated data is emphasized as critical [R18]. Adhering to legal and contractual guidelines is paramount when training models, further underscoring the need for meticulous data management practices within the company. These factors collectively influence the strategic decisions around data handling and AI model training at GKN, reflecting the complex interplay between technology capabilities and regulatory requirements.

4.1.2.6 Technology

There is a vision of generating entire CAD models through advanced AI technologies like text-to-speech or directly from lists of requirements specifications. This approach could dramatically transform the design process, providing a streamlined, efficient method for producing detailed CAD models directly from spoken or written inputs.

This section explores where AI technology currently stands according to industry respondents. It examines what is feasible with today's AI/ML programs and applications, and what objectives remain beyond reach. The intention is to discover potentially overlooked new areas and map out the knowledge landscape of these respondents. Furthermore, this discussion delves into the specific kinds of assistance that design engineers seek from AI during critical phases such as concept generation and evaluation.

One respondent, R16, expressed the complexities involved in integrating AI into the traditional design process, particularly when adhering to established design guidelines that reflect the company's unique approach to model building:

> *"When it comes to design, we have our own 'design guidelines' about how we think a model should be built. This means that it's something we really need to train based on our way of thinking. So, that's probably the hardest part. To generally build up a CAD geometry from start to finish is, I believe, a vision or utopia." - R16*

R16 further discussed the potential for LLMs (LLMs) in the design process, particularly in how they can handle subjective elements that do not require absolute precision to be effective:

> *"Generally it's like the more subjective it is, the better a LLM can perform so like if we talk about in a CAD model, you might want to have texted descriptions for different things. And that text description doesn't have to be 100% accurate to be good enough, as long as it describes the system. Or when it comes to like correlation the like this part that has been designed matches these parts from the previous years and if they can find that correlation." - R16*

These insights from R15 and R16 showcase the broad spectrum of applications and the complex challenges associated with integrating AI in design and engineering. While there is excitement about the possibilities AI and ML bring to the table, there is also a recognition of the significant hurdles that need to be overcome, especially in ensuring that these technologies align with existing company standards and practices. The discussions underscore the need for a deeper understanding of how AI tools can be practically and effectively integrated into existing workflows, enhancing and streamlining the design process while adhering to the nuanced needs of engineering disciplines.

4.1.2.7 GAS current AI/ML capabilities

GKN's engagement with AI/ML technologies is evident despite a lack of overwhelming positivity in initial discussions. Researchers and PhD students at GKN are deeply immersed in advanced AI/ML projects, reflecting a commitment to exploring and expanding these capabilities within the company. This effort aims to understand what is currently achievable with AI/ML, identify beneficial methods for PD processes, and determine the potential for broader integration across the enterprise.

Specific AI/ML applications already in use include the NX platform, which utilizes machine learning to enhance surface highlighting [R14]. Additionally, tools such as Office and GitHub CoPilots are deployed to assist in document and code writing, though they present challenges related to data sharing [R15]. The integration of machine learning extends into design optimization as well, with applications like OptiSlang being used to explore design spaces and optimize designs, incorporating machine learning techniques:

> *"OptiSlang is used for design experiments and exploring design spaces, including machine learning for optimization." - R14*

Moreover, there is potential to further utilize AI/ML in areas such as stress reports and design practices, which could serve as valuable data sources for training models that provide guidance, particularly for agent-based models:

> *"GKN could use stress reports and design practices as data for training models for guidance, especially for agent-based models." - R14*

Research projects are another area where AI/ML is actively applied; for instance, there is an ongoing project dedicated to collecting images, storing timestamps, and other process information. This data is integrated and correlated with GKN's main data logging system to enhance operational efficiency [R16]. AI/ML's role in improving production processes is also significant, as seen in its application for automation and efficiency improvements through image recognition in GKN's production lines:

> *"ML is heavily used with GTC for production, automation, and efficiency through image recognition." - R15*

Further explorations into generative AI are underway to improve the quality of data in Health, Safety, and Environmental (HSE) reports, highlighting a proactive approach to leveraging AI for more accurate and reliable data management [R17, R18].

GKN also considers the potential of using public datasets for on-premise model training and fine-tuning, thus enhancing the capabilities of their AI systems without compromising the security of proprietary data [R18]. There's an ongoing discussion about the feasibility and benefits of outsourcing model training using general aerospace data instead of company-specific datasets. This strategy is considered potentially valuable for broad applications, even if it might not provide the finely tuned results that company-specific training would yield [R15].

The level of integration of AI into regular processes, such as report generation, is still in the early stages—described as being at the "thesis level"—with researchers looking into applications of technologies like OpenAI [R15]. While some individuals within the company show advanced skills and competence in AI, the broader organizational adoption is still maturing. The company is gradually moving towards a more data-driven approach, aiming to automate data usage more effectively and reduce manual interventions [R17].

In terms of data privacy and security, GKN is considering outsourcing model training while ensuring sensitive data is anonymized, thus maintaining confidentiality while still advancing their AI capabilities [R17]. This combination of internal advancement and strategic external collaborations underscores GKN's balanced approach to developing and implementing AI/ML technologies.

4.1.3 Benchmarking study

The purpose of the benchmarking study was mainly divided into three parts: 1) Gain access to current state of the art of AI implementation in industry, and compare this to where GKN currently stands, 2) Take part of other organizations motivations for implementing AI, 3) Take part of lessons learned from companies being successful/unsuccessful implementing AI. In total, 3 organizations were interviewed as part of the benchmarking study. The companies interviewed were GKN Fokker, SKF and a large life-science company in a heavily regulated industry who wants to remain anonymous. This company will thus be referred to as C3 onwards.

4.1.3.1 GKN Fokker - R19

AI implementation:
In terms of AI implementation at GKN Fokker, they create software, mainly Python modules to automate various steps in the wiring design process. They aim to automate tasks that are either highly repetitive or complex and extract knowledge from experts and create Python applications to support these tasks. However, looking specifically at the design process, they are working with AI more as a general concept:

> "In the design process, we are not really working with machine learning, but more with AI as a broad definition. We include traditional AI like design of experiment algorithms or optimization algorithms in our work. The more automation modules we create, the more we can explore optimization aspects. We're looking into different design of experiments and optimization algorithms, mainly within the defined project."

One specific example of how they utilize AI is when they are designing and optimizing signal routing for their harnesses:

> "One example is the signal routing module, which is a bit like Google Maps for designing our harnesses. It helps us route signals from A to B, considering constraints and ensuring signals go over harnesses that match their criteria. We also deal with shielding, grounding, and optimizing the selection of components to create the best solution, whether it's the lightest or cheapest."

The key driver for GKN Fokker to implement AI is mainly to make the process quicker or the product more optimal, aiming for cheaper, lighter and safer solutions. In their research and development department, AI and ML is actively considered and actively looked into, although still indeed with caution. They bring up AI-support like ChatGPT and its many interesting applications, but also mention that programs like this are still not necessarily trustworthy even though the results you get might seem like it.

Challenges and lessons learned:
The main challenge GKN Fokker brings up when trying to implement AI has been centered around the abundance of AI-support on the market and state the following:

> “The opportunities and the options out there are so broad. So there’s so many, many different ways in which AI could help you, so many different algorithms out there, and it’s sometimes a bit difficult to choose what is actually the most suitable for me.“

In order to overcome this challenge, they state talking to people has been the most effective:

> “For the first challenge, the easiest way to overcome it is just by talking to people and see what other people and learn from their experience. If you Google about it, there’s always so many things you find you never know what’s trustworthy necessarily. But if you have some good colleagues, whether it’s internal colleagues or external companies we collaborate with, that’s often a bit more of the trusted source.“

GKN Fokker also bring up more specific optimization challenges related to their domain:

> “There are a lot of optimization algorithms out there that deal with a lot of continuous variables, and it’s often a bit more tricky because we deal with a lot of non-continuous variables or categorical variables. Then, often you’re more pointed towards generic algorithms, for example, which can have quite some problems with speed. So, finding the best way to deal with categorical variables is also, I guess, is the challenge we often come across. To deal with this, there is often much trial and error.”

Other challenges encountered when implementing AI are for example, when tools and software become more and more complex, it can be difficult for people to still understand what is going on. Another challenge is related to ChatGPT and generating code. If you want to use ChatGPT, for example, for generating code for you, you never actually know where that code is coming from. It could actually be coming from a licensed source and you’re not actually allowed to use that code. Lastly, getting IT on the same page as you can take a lot of time and and thus being an obstructing factor. They also state that they still have not overcome most of the challenges.

Benefits:
When it comes to metrics measuring the success of AI, they mainly look at whether it saves time or creates more optimal solutions. In specific processes, they have seen time reductions somewhere between 50% and 80%, considering AI as automation.

> “If we consider AI as automation, we’ve seen significant time savings in specific processes, reducing the time required. I would say definitely somewhere from 50 to 80% time saved in some cases.”

However, R19 mentions that optimization potential within design is harder to quantify and consequently, harder to give an honest comparison of.
Required for AI implementation:
When discussing what is necessary to successfully implement AI, GKN Fokker highlight the importance of people and having the right competence in the organization:

> “If you have direct experience with AI, that helps. But other than that, I think it’s often just having people with a good abstraction view. Let’s say people that can look at processes as a whole. That can look at the big picture on how everything is connected because I think that is often required to be able to decide how best to apply different types of AI or different types of algorithms.”

Data:
In general GKN Fokker store data on an internal server. Sometimes, they will have a cloud solution but in that case, it will be an internal cloud. On rare occasions, they make use of external clouds, but then they need to completely make sure that they don’t upload military data.

They are working with LLMs, but are solely developing models in-house.

Future plans:
Looking into the future, GKN Fokker want to investigate more into optimization algorithms and design of experiments. Looking more at machine learning, they have two ongoing research projects:

> “One of the projects we have is looking into the conceptual design phase, where you often have quite a big lack of data. We would also be interested in looking at how machine learning can help us predict some of the missing data early in projects.”

GKN Fokker also want to highlight that AI is not necessarily about replacing people with AI-solutions, but more about ensuring the designs meet safety requirements and are optimal, considering trade-offs between various factors.

4.1.3.2 SKF - R20

AI implementation:

For SKF, the main objective is to develop digital solutions to support their approximately factories and boost them digitally. They are working with the ISA-95 pyramid with its several layers. At the bottom, are their factory floors with lean processes where they connect machines and sensors, and as you move up the pyramid it becomes more abstract. At the top are their ERP systems with all product orders and in between this, are for example connectivity layers and data warehouses. SKF are making solutions throughout the entire ISA-95 pyramid:

> “So we’re making solutions throughout this ISA-95 pyramid, which means we’re looking at: How do you connect machines? How do you create data pipelines? How should the data be processed? And then, you can use it in ML models, for example. One is like predictive maintenance, i.e. you predict where you have deviations. Or you’re trying to optimize parameters, so it’s kind of two different perspectives.”

SKF is also working with Microsoft Copilot and find it especially useful for transcribing and summarizing Teams meetings, especially from a project management perspective. They are also working with GitHub Copilot to make programming more efficient.

Furthermore, SKF is working with AI to capture and create value from handwritten notes in their manufacturing facilites:

> “We have systems where maintenance technicians write notes, but these are unstructured from a database point of view. But they are structured, mentally. With LLMs, we can actually draw a connection, even from these handwritten notes. And that means that if we have thousands of such handwritten notes in Swedish in the factory, now all of a sudden they’re available in a structured way, in all languages, to everyone. And that builds on the fact that then we can start using it in different solutions. For example, for a chatbot for maintenance technicians. And it doesn’t matter if you’re in Mexico, China or Sweden.”

The way SKF implement AI-solutions into the organization is through ”spikes”, where they on a small scale, in specific applications try things out to learn:

> “So it’s not like we’re developing some big solution that we’re going to roll out globally, but we’re still very much in the Exploration and Discovery phases, so to speak.”

SKF is also looking into building their own, internal, Large Language Model where they can upload manuals and documents but this is still in the early phase.
Furthermore, the interviewee gave insights into previous experiences at other organizations:

> “I’ve worked at Ericsson, Volvo, Polestar, SKF and I’ve been to quite

> a lot of AI conferences. And realized that we are quite far ahead when it comes to AI. And the main reason is the values and norms here. As well as the fact that people here adopt technology quite quickly and have a great interest in technology. That is the main thing. We have senior managers who sit and run Home Assistant and connect their entire house and assemble all the sensors. And I'm talking really senior managers. And when top managers have genuine technical interest, that creates a driving force to implement AI."

However, they do not use AI for concept development or product development.

The key driver to implement AI at SKF is primarily to assist "white collar workers" who are less experienced. They mean that really sharp experts don't have much to gain from AI because they are so talented so it's hard for AI to match that knowledge when we look at cutting-edge technology.

Challenges and lessons learned:
One challenge associated with AI and LLMs according to R20, lies in their tendency to generate erroneous information confidently, asserting applicability even when it may not always hold true. This approach is ineffective in a factory setting. Having a model that is 99% accurate is still not enough. Ensuring reliability is crucial, as failure to do so can result in significant costs.

Another challenge highlighted by R20, is related to the pace at which AI is evolving and the preasure this puts on orgnizational structures:

> "One challenge is that AI is evolving extremely fast. But in organizations, you have budget structures, organizational structures that can't keep up with that pace. If you look at when we develop an internal SKF co-pilot, there is someone who does it as a hobby and spends an hour or so on the side. And all of a sudden, you have 100 people who are standing and want support to install this. And then there are no budgets and there are no structures to handle this."

The interviewee occasionally attends AI conferences where the shortage of AI competence is highlighted:

> "Something that is highlighted at AI conferences is the large skill gap when it comes to AI. How do companies ensure that they have enough AI skills? Because this is a huge shortage. Right now, you have a large number of young talents who are interested and want to start working with AI. But there are only a few really good data scientists who can be coaches and mentors for all the young people. And how do you match this up at the company? Because you need to get up to speed in building your AI skills, i.e. machine learning."

Lastly, there seem to be a general challenge centered around data engineering, data science and data management in order to ensure high quality models. According to R20, only a few possess the right AI competence, and if everyone is going to need AI support, there is an enormous amount of data that needs to be processed, be of high enough quality and the competence to handle this is today not sufficient.

Benefits:
According to R20, it's too early to tell whether they can see any general, measurable, improvements from starting to implement AI support. However, in some specific areas, you can get subjective assessment regarding improvements from AI implementation:

> "You can get subjective assessments with the programmers. Some might say they save 50% or 80% of their on certain elements. But then you get to work with other tasks instead. So it doesn't go that much faster just because of that. It may be more efficient, but we can't measure that. It's too early."

Required for AI implementation:
In order to succeed with A implementation, the organization needs to be equipped with a balanced infrastructure:

> "I would say that this is a very boring answer, but we need a balanced infrastructure. If you take every sensor that gives thousands of readings per second and then you push all that data into the cloud. Then you get an incredibly slow solution that will cost a lot because you shovel a lot of data completely unnecessarily, a very long way."

The way SKF work with data is through a three-phase process called ETL, short for "Extract", "Transform", "Load". The ETL process involves extracting data from various sources, transforming it into a standardized format, and loading it into the next step. At SKF, they have ETL processes at every level. At the factory floor, they have edge computers where they have to decide what data they need to process and how to process it before sending it to the next step. They also have to consider what data could and should be processed on the computer and what data should be uploaded to be processed in the cloud. Determining what data is needed at every level, and how it should be processed at every level, is what R20 means by having a balanced infrastructure. The following quote describes an example from the factory floor at SKF:

> "You need to understand what to do with the data at each stage. Currently, we're exploring what we refer to as edge computing. This involves deploying computers capable of capturing data such as vibrations or acoustic emissions, using sensors to detect when, for instance, a grinding wheel makes contact with metal for grinding purposes, enabling real-time signal analysis. We're dealing with a significant volume of signals

> per second, down to milliseconds, to accurately determine when the disc contacts the material. Subsequently, this data needs to be fed into a machine learning model, processed, and then utilized to control the machine once again. It's of course an extreme scenario, often termed high frequency."

With high frequency, everything needs to happen on right away, locally, in real time, on the factory floor. There is no time to upload the data to a cloud and process it. You could also collect data i batches and send it, but this is a different scenario and is not applicable for high frequency scenarios. According to R20, every case is different and there is no template or right or wrong way of how to process data. A lot depends on the use case and if it is a high frequency or low frequency process.

Data

The primary method of storing and securing data at SKF is through the utilization of Azure Cloud and Microsoft's ecosystem. R20 mentions that several other organizations use Multicloud, but this it not something SKF has decided to work with. SKF store a lot of data externally but there is a lot to consider when storing data this way:

> "There's much to consider regarding storing business-sensitive information in clouds. Then, when comparing it to the alternative of either having your own on-premises data center or storing information in the cloud, it's actually more secure when it's in the cloud, considering all the security capabilities that Microsoft and they have there. What if a fire starts? At Volvo, for example, there we had two data centers on each site in case there was a fire on one of them."

Furthermore, R20 brings up ChatGPT and that there is a lot to consider before using its many capabilites, and they can't us ChatGPT however they want, as this might leak company-sensitive information. If they intend to utilize ChatGPT skills or capabilities, they must do so through SKF to ensure that they do not inadvertently disclose internal information

R20 also wants to highlight that is does not matter how much data you have, if it isn't contextualized and lack time stamps and meta data.

Lastly, a final word by R20 at SKF describing a more general view on AI:

> "I probably won't be replaced by AI, but I may well be replaced by someone who is skilled at using AI, so it's essential for me to understand that tool. It's like today. Would you hire a person who brings the phone books when looking for a number, or hire a person who goes on the internet to look up the phone number? So if you're not part of the development, then you'll be left behind; that's how it is." - R20

4.1.3.3 Lifescience company

AI implementation:
The interviewee, R21, holds the position of quality manager at C3. He approaches AI implementation from a quality perspective and describes its relevance to his work, rather than focusing on how AI is implemented across the entire organization. R21 use AI support in two main areas:

> “I utilize AI in two main areas. The first is predictable quality, which involves predicting quality problems before they occur or escalate. This entails correlating data using machine learning techniques. Additionally, it’s about accessing not only numerical sources but also non-numerical sources through NLP, which stands for Natural Language Processing. Currently, there’s a project underway focusing on this.“

The second area is related their management system:

> “The second area is related to our management system, which consists of about 1000 documents that have grown over 30-40 years. As a human being, you have a lot of bias. With ChatGPT, we rewrite these in a better way.“

R21 continues to state that C3 has more opportunities than they have money, and that it’s always a competition about which product to choose or which projects to run. With growth companies, this is always the case, which puts high demands on productivity and this is where AI is assisting, according to R21.

R21 is also involved in innovation projects around ChatGPT where C3 has purchased a base model and a knowledge base, placed it inside the walls of the organization, and integrated it to their infrastructure. They have also disabled the learning of the model.

The key driver to implement AI is mainly to increase productivity and find information more efficiently. However, R21 want to make it clear that they have still not implemented AI fully into the organization and there is still much business to do before. How AI is supporting in specifically the PD-process, R21 can’t disclose, but only that it is related to evaluation of concepts.

Challenges and lessons learned:
There is a lot of discussion regarding challenges related to AI today at C3. One challenge brought up by R21 is related to LLMs and how complex it is to understand how the data is being used:

> “There are numerous discussions taking place in that regard, which can be challenging. With AI tools such as ChatGPT or their equivalents, once you train them and feed them data, you lose control. The information gets absorbed into the model, becoming fragmented and un-

> traceable. Subsequently, you relinquish control over its usage. From a commercial standpoint, there are risks regarding business perspectives that must be considered."

Additionally, there are legal obligations concerning data. Given that C3 manage a substantial amount of personal data, stringent laws and ethical considerations come into play. Moreover, compliance with EU legislation is imperative. These factors impose significant constraints on their operations. The largest danger is that personal and company-sensitive data is leaked.

Benefits:

C3 have still not implemented AI fully into the organization, which makes it difficult to give exact figures or measurable benefits. However, strictly looking at productivity in PQS, R21 can see a 80% decrease in time.

R21 can also see huge improvements with their verification testing. They can go from not understand anything, to understand in something in three minutes. Looking at predictive quality and ability to predict failures, they are able to avoid large costs.

Data:

At C3, cyber security is very strict who has access to certain data is extremely regulated:

> "It is extremely controlled, who has access to particular data. My organization has full access. But those who do those studies do not get access to the data, until the study has been done to avoid human error and influence."

R21 goes on to highlight the role of data quality and its importance to succeed with AI:

> "If you don't have the right data quality, you will never succeed with AI. There has to be incredible discipline when it comes to data quality and data integrity."

4.2 Outcome of literature study

This section will explore the outcomes of the literature study, delving deeper into the possibilities and opportunities of integrating AI/ML into concept development, specifically in the areas of concept generation and concept evaluation.

4.2.1 Large Language Models in Concept Development and Generation

Large language models (LLMs) have revolutionized the field of nautral language procesessing (NLP) and increasingly influence various sectors beyond academic circles. These models lead the performance across numerous benchmarks in natural language understanding (A. Wang et al., 2019).

Engineering is a field rich in knowledge, poised for significant advancements through the adoption of cutting-edge techniques from the NLP sector. (Göpfert et al., 2023) suggest that foundation models like LLMs are capable of supporting creative reasoning tasks within the engineering design process, thereby augmenting and meshing with traditional computational approaches such as topology optimization.

Furthermore, (Gomez et al., 2024) work show that there is promise in using LLMs when generating CAD-models by prompts, which is an important aspect in concept generation and development. The research use text-to-CAD to illustrate this and comes to the conclusion that the use cases in their research show that LLMs can be applied at the System Level and Detail phases of the product design process. The authors also state that LLMs used to design complex systems and geometries, is an area where there is limited research currently.

This section aims to research what other applications of LLMs in Concept Generation and Development can be found in the current literature.

4.2.1.1 AutoTRIZ - Ideation with LLMs

Intuitive and structured ideation methods such as brainstorming, morphological analysis, and mind-mapping (Camburn, Arlitt, et al., 2020) play crucial roles in enhancing the creative ideation processes among human designers for concept generation. Among these methodologies, the Theory of Inventive Problem Solving (TRIZ) (Altshuller, 1999) stands out as a prominent approach extensively utilized for systematic innovation. TRIZ offers a knowledge-based framework designed to solve engineering problems by addressing technical contradictions through inventive principles derived from a vast patent database.

Despite its benefits, the usage of TRIZ at GKN has been intermittent. The directives of ideation given in OMS, is primarily used for quality control instructions, traditionally and does not support creative or synthetic directives. Moreover, the intricacies of TRIZ resources pose significant cognitive challenges, affecting how effectively individuals can learn and apply the methodology. Therefore, the effectiveness of TRIZ is substantially dependent on the users reasoning abilities and their familiarity with its principles, as demonstrated by studies integrating NLP (Guarino et al., 2022; Hall et al., 2022).

The use of Set-Based design methods at GKN highlights the importance of thoroughly exploring the design space and progressively refining it as additional information becomes available. This method underscores the significance of expansive initial ideation before converging on feasible design solutions.

Recent advancements in technology, particularly the development of LLMs, are reshaping the landscape of ideation and innovation. LLMs are increasingly employed to process extensive design documentation, represent designs in specialized formats, and identify user needs for product development (Qiu & Jin, 2023; B. Wang et al., 2023). These models have also been utilized to distill design-related knowledge from extensive reports and documents (Qiu & Jin, 2023) and to decompose conceptual design tasks into Function-Behavior-Structure (FBS) formats, facilitating ideation across different aspects (B. Wang et al., 2023). Furthermore, (Han et al., 2023) introduced an LLM-based attribute-sentiment-guided summarization model to extract user needs from online product reviews, demonstrating the versatility of LLMs in capturing and utilizing user-generated content.

One significant innovation is the development of AutoTRIZ (see Figure 4.3), an LLM-based intelligent tool that automates the TRIZ methodology (S. Jiang & Luo, 2024). AutoTRIZ begins with a problem statement from the user and automatically generates a detailed solution report, which explains the reasoning process based on TRIZ and the solutions derived. This tool reduces the entry barrier to TRIZ by eliminating the need for extensive training, thereby lowering the cognitive load and accelerating the ideation process (Ilevbare et al., 2013).

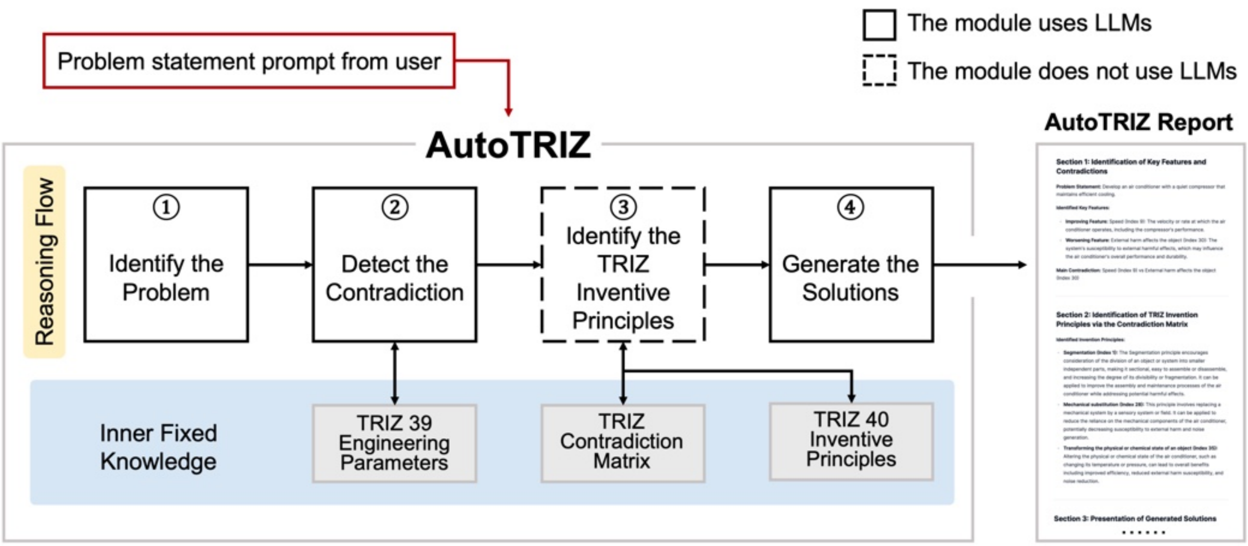

Figure 4.3: The process flow in AutoTRIZ (S. Jiang & Luo, 2024).

LLMs have shown extensive potential in various engineering fields, including microfluidic devices (Nelson et al., 2023), robotics (Stella et al., 2023), and web user interfaces (A. Li et al., 2023). These models, especially those in the GPT series, possess a broad knowledge base that can significantly aid in ideation and innovative problem-solving across multiple domains.

Looking forward, the framework implemented in AutoTRIZ could be extended to automate other knowledge-based innovation methods. For example, the application of LLMs to utilize design heuristics identified from thousands of design outcomes could further enhance ideation quality and innovation (Yilmaz et al., 2016). By integrating various ideation methodologies into LLM reasoning modules, a more robust and versatile tool could be developed, potentially transforming the field of design and engineering innovation.

4.2.1.2 Generative Design with AI

Generative product design tools, often integrating AI, are transforming early-stage design by allowing for the exploration of complex shapes challenging for human designers to conceive or perfect alone (Chandrasegaran et al., 2013). In this thesis, when generative design are mentioned, it is the generative design of a product using a commercially available design tool with algorithmic computation to achieve complex goals within the aspects of product design.

(Saadi & Yang, 2023b) defines generative design as something that involves the collaboration of human designers and algorithmic computation to achieve complex goals with superior results than that of each entity when creating independently. Furthermore, the work by Saadi and Yang (2023b) focuses on generative design where the optimization tool takes on a larger function in the design process.

Generative design tools in the design process can take on many forms with varying levels of involvement from the tool. The design process can be driven by the designer, with minor involvement from computational tools in tasks such as ideation or analysis. For example, Autodesk DreamSketch uses a generative design algorithm to produce multiple 3D sketches based on a designer's initial problem definition. On the other hand, the generative design process can have more substantial tool involvement, as is the case with many commercially available generative design tools. The commercially available design tools mentioned in the work by (Saadi & Yang, 2023b) is primarly Fusion 360, CATIA, NTopology and a unnamed Design Space Exploration tool.

Designers input design goals and specifications into the tool. The tool will explore possible solutions and generate several valid designs that meet the requirements. In this process, generative design tools can be used to take on many tasks in the design process, including idea generation and product optimization. Generative design in which the design tool takes on a more active role has the potential to drastically change the design process while leading to more creative geometries.
Traditionally used in later design stages, these tools are now highly beneficial from the beginning, significantly impacting design processes and outcomes. For example, General Motors utilized Autodesk's **Fusion 360** Generative design to create a seat bracket that was both 40% lighter and 20% stronger than its predecessor, merging eight components into a single 3D printed part (Briard et al., 2020). This

highlights how AI can enhance product performance beyond original capabilities, merging human creativity with machine efficiency (Saadi & Yang, 2023b). The role of generative design extends to redefining tasks traditionally handled by designers, such as concept generation and product optimization. By setting early specifications, manufacturing methods, and product architecture, designers enable these tools to generate multiple viable outputs, altering traditional design approaches significantly.

The integration of these tools also influences designer behaviors, such as their communication and confidence, and can lead to alterations in generated designs to enhance aesthetics (Krish, 2011; Zhang et al., 2021). The design process may range from minor tool involvement in ideation to substantial roles in systems like **Autodesk DreamSketch**, which generates multiple 3D sketches from a designer's initial inputs (Kazi et al., 2017). To further exemplify, **Fusion 360** has a generative design option where, users can define design goals and constraints, such as material usage, manufacturing limitations, and performance requirements. The software then uses AI algorithms to generate multiple design options that meet those criteria (Autodesk, 2024).

Drawbacks of AI in the design process: While AI tools like generative design can significantly aid in ideation and early-stage design tasks, studies by Lopez et al. (2018) highlight some challenges. For example, while AI can produce numerous practical ideas, it occasionally fails to completely match the subtle details of human-set design briefs. Vlah et al. (2020) explored how topology optimization and generative design are implemented in industrial settings, finding that these tools require engineers to substantially rethink their approach to setting up the design space. Computational tools influence early-stage design significantly, particularly in terms of aesthetics, enabling designers to explore specific shape grammars through parametric models (Alcaide-Marzal et al., 2020).

Designers in Generative-Driven Design: The cognitive processes and collaborative dynamics of designers are critically shaped by their interaction with generative design tools. Song, Soria Zurita, et al. (2020) emphasize that computational tools can profoundly impact designers' exploration strategies and the final designs. This interaction necessitates an adaptive approach from designers as they navigate the new landscape of AI-driven design tools, requiring them to adjust parameters and design approaches as projects evolve (Vlah et al., 2020).

Changes in collaboration styles and communication strategies significantly influence design processes and outcomes. For example, studies have shown that AI tools can either enhance or disrupt team dynamics depending on how these tools are integrated into the design process (Phadnis et al., 2021; J. (Zhou et al., 2020). The design of AI tools themselves can also significantly impact how designers interact with these systems and the effectiveness of the resulting designs (Chaudhari & Selva, 2023; Pillai et al., 2020). Therefore, understanding how to effectively utilize AI within the design process is essential for capitalizing on its potential benefits while mitigating its challenges.

4.2.1.3 Computational engineering design

Today, the digitization of engineering design is well advanced. In the past, technical drawing was performed on drawing boards until software for computer-aided design (CAD) was developed in the second half of the twentieth century, which is generally adopted today (Göpfert et al., 2023). Currently, finite element analysis, topology optimization, design-support tools for additive manufacturing, and more are used. With model-based systems engineering and digital twins, the PDP became centered around digital models. Virtual and augmented reality enables visualization and interaction with designs. Due to the breadth of the field, only a brief overview of advances in computational engineering design can be given here, focusing on design generation, design strategy learning, and NLP (in particular LLMs) for engineering design.

In recent years, the intersection of deep learning and mechanical design has seen remarkable progress, transforming the conceptualization and realization of 3D structures (Jadhav & Farimani, 2024). The capability to generate 3D structures from diverse input modalities, including text (Nichol et al., 2022; Sanghi et al., 2022), images, and sketches (C. Li et al., 2022), has expanded the possibilities and made the field of mechanical design more accessible and versatile. However, a notable limitation is that these models are not inherently designed to account for mechanical specifications or functional constraints.

NLP tasks include text generation, compliance with specific task directives (J. Zhou et al., 2023), and the demonstration of emergent reasoning capabilities (Huang & Chang, 2023). Given the significant computational resources required for training and fine-tuning LLMs for specific downstream tasks, these models have demonstrated an exceptional ability to generalize to new tasks and domains. This adaptability is achieved through the in-context learn ing (ICL) paradigm, which has significantly deepened our understanding of LLMs' capabilities. By utilizing minimal natural language templates and requiring no extensive fine-tuning, LLMs have established themselves as efficient "few-shot learners" (Perez et al., 2021).

In chemistry, LLMs independently design, plan, and execute complex experiments (Boiko et al., 2023). In mathematics and computer science, they have solved long-standing issues such as the cap set problem and improved algorithms for the bin-packing challenge (Romera-Paredes et al., 2024). LLMs also contribute significantly to biomedical research (Thapa & Adhikari, 2023) and materials science (Xie et al., 2023), enhancing scientific understanding and capabilities across various fields (Göpfert et al., 2023). Furthermore, they effectively optimize foundational problems like linear regression and the traveling salesman problem, often outperforming custom heuristics with simple prompts (Yang et al., 2024).
In mechanical engineering, the fine-tuned LLM has showcased proficiency in retrieving knowledge, generating hypotheses, conducting agent-based modeling, and linking various fields through ontological knowledge graphs [91]. Moreover, LLMs have successfully automated the creation of initial design concepts by integrating domain knowledge (Makatura et al., 2023; Q. Zhu & Luo, 2023). Additionally,

LLMs display robust skills in design-related tasks, including sketch similarity analysis, material selection, analysis of engineering drawings, CAD creation, and tackling spatial reasoning problems (Picard et al., 2023).
The authors (Jadhav & Farimani, 2024) present a framework that leverages in-context and few-shot learning, along with LLMs' inherent reasoning and optimization capabilities, for structural optimization, particularly in truss design. This approach allows LLMs to generate and iteratively refine design concepts with minimal input, effectively optimizing structural outcomes. Additionally, the versatility of LLM-based optimization in processing categorical data marks a significant shift from traditional optimizers that focus on numerical data.

Generative adversarial networks (GANs), feedforward neural networks, variational autoencoders, as well as reinforcement learning systems have been used in design-related generation tasks such as topology optimization or shape synthesis based on visual modalities such as images, voxels, and point clouds (Regenwetter et al., 2022). Other work focuses on learning design strategies. Given a state in solving a truss design problem, Raina et al. (2021) predicts what actions humans perform next. Gyory et al. (2021, 2022) analyze real time data of design teams to suggest measures from a predefined list if the communication or action frequency appears to be too low.

In design research, NLP has been applied to requirements extraction, ontology construction, patent analysis, and more (Siddharth et al., 2022). Using foundation models such as LLMs or pre-trained multi-modal models in the engineering design process is a recent and unexplored topic (Göpfert et al., 2023).

Several studies have experimented with using LLMs to provide designers with inspirational stimuli for ideation. In three explorative studies, S. Jiang and Luo (2024), Q. Zhu and Luo (2022), and Q. Zhu and Luo (2023) prompted GPT-2 and -3 to generate design concepts (text-to-text) based on the description of either a concept, problem, or analogy in both a fine-tuning and few-shot learning setting. Similarly, Q. Zhu et al. (2023) fine-tune GPT-3 for bio-inspired design concept generation. Ma et al. (2023) compare design solutions generated with GPT-3 with crowdsourced ones. And S. Jiang and Luo (2024) used GPT-4 for synthetic TRIZ ideation. Other work has focus on design concept evaluation combining Google's pre-trained language model BERT, and image models in a multi-modal one (Song, Miller, & Ahmed, 2023; Yuan et al., 2021). Song, Zhou, and Ahmed (2023) provided an extensive overview of multi-modal machine learning for engineering design. They outline possible applications, but focus on lower level tasks such as text-to-shape or shape-to-text synthesis.

4.2.1.4 The Design process: a goal-oriented argumentative conversation

Digital artifacts include shapes, assembly processes, stress distributions, flow patterns, and more. Until now, however, computer-aided engineering, as practiced in industry, has not included the creative and argumentative process of the PDP itself. It is argued that this process could be digitized and partially automatized next, and it is outlined how this can be achieved (Göpfert et al., 2023).

Many steps in the PDP are performed using computation and are not based on human thought alone. However, humans are needed to integrate these computational processes, such as calculations, simulations, or optimizations, into a meaningful superior PDP. Human thought and world knowledge are required to reduce the solution space in advance and to come up with original ideas that have not been modeled to be computationally accessible before. For example, when a bicycle is designed, the starting point is not a blank slate but an idea of how a bicycle looks and how it has worked well for over a century. If a standardized aerodynamic tube shape across bicycle manufacturers is proposed, it is unlikely that this idea originated from a numerical optimization (Göpfert et al., 2023). Instead, background knowledge and the ability to think and reason are used. Solving engineering problems requires an argumentative dialogue. As such, argumentation is inherent to the PDP. Experiments, calculations and similar activities, inform the dialogue to provide information necessary for PD, thus argumentation is an essential part of the design process.

Göpfert et al. (2023)argues that a goal-driven, argumentative dialogue is at the core of the design process, and propose that it should be represented as a digital artifact. Humans communicate, argue, and reason using natural language. Hence, the argumentative dialogue can be represented in textual form.

Many parts of the design process, however, cannot be represented as text. Thus, the design dialogue is distinguished from external actions (such as performing an experiment or simulation) and other engineering artifacts (such as a drawing, or 3D model). It is formulated that external actions are invoked from within the design dialogue and in turn inform the dialogue, either directly or indirectly, by yielding other engineering artifacts that inform the design dialogue (See Figure 4.4).

Representing the argumentative dialogue as a digital artifact would improve the documentation of the design process. Instead of only archiving the results of process steps (e.g., CAD files or the results of simulation runs), the reasoning process is documented and hence archivable. For a past development process to be efficiently used for the development of a new product generation, past decisions and alternatives must be accessible. Having the reasoning process explicitly documented makes past design decisions traceable. Furthermore, making the reasoning process explicit could improve collective reasoning and therefore collaborative design. Finally, it would allow for machines to participate in the reasoning process.

4.2.1.5 LLMs as the argumentative designer

LLMs and multi-modal models, capable of processing natural language, seamlessly integrate into the largely natural-language-based argumentative dialogue, whether it occurs within or between individuals. World knowledge plays a crucial role in facilitating human interaction and narrowing down solutions to foster creativity. Preliminary findings indicate that LLMs develop comprehensive world representations, even though they are trained on straightforward objectives (K. Li et al., 2023).

Unlike humans, who may be limited by their existing knowledge when employing analogical or biomimetic design approaches, LLMs learn to mimic extensive knowledge during training. Large-scale LLMs demonstrate robust performance across various reasoning tasks and can execute sequential reasoning processes (J. Wei et al., 2023). Multi-modal models handle diverse design representations essential throughout the product development cycle, including text, tables, sketches, and 3D models (Song, Zhou, & Ahmed, 2023). Furthermore, many engineering design tasks surpass the capabilities of pure thought, requiring specialized software and databases, such as CAD and simulation tools, and patent and material databases. Recent studies show LLMs capability to autonomously interact with APIs, enhancing their utility in engineering applications (Schick et al., 2023)

In conclusion, these models are fundamentally suitable for supporting engineering design dialogue. However, a singular model application provides limited benefit. Instead, these models should be integrated into a framework designed to address complex engineering challenges (Göpfert et al., 2023).

```
Prompt: Attached is the design of the current iteration of our aerodynamic road bike
as well as the requirements, simulation results, and wind tunnel results. Please
optimize the seat tube and seat stays.
Thought: What to optimize for?
Thought: Low aerodynamic drag has the highest priority. Other objectives are low
weight, high comfort, high stiffness, low maintenance requirements.
Action: Reinsuring by asking design team: Is this correct?
Prompt: Yes.
Thought: What are the constraints?
Action: Check given requirements file.
Thought: The bike must comply with the rules of the governing body for sports cycling
to be used in professional races.
Action: Look up rule book.
                                        ⋮
Thought: Get inspiration from current aerodynamic road bike designs.
Action: Web search
Thought: Get inspiration from patents.
Action: Patent search
                                        ⋮
Thought: How to reduce the aerodynamic drag?
Action: Look up equation for aerodynamic drag.
Thought: Lowering the air density, frontal area, or drag coefficient reduces
aerodynamic drag. However, the air density is an environmental property which can not
be changed as part of the bicycle design.
Thought: The lower the seat stays the lower the frontal area.
Thought: Define the lowest allowed contact point with the seat tube as a boundary
condition.
Action: Run multi-physics topology optimization to optimize seat stays for
aerodynamics.
                                        ⋮
Thought: How can the drag of the seat tube be reduced?
Thought: Frame and tire should flow as seamlessly as possible into each other.
Thought: What problems does this solution have?
Thought: High pressure zone between tire and seat tube.
Thought: Is there a solution to this problem?
Thought: One might drag the outer shape of the seat tube close to the tire but
envelope it from behind to leave more room between the tire and the seat tube.
Action: Check for patent infringements
Prompt: Can you visualize this idea?
                                        ⋮
```

= Human = Machine

Figure 4.4: Example of design process with an argumentative dialogue (Göpfert et al., 2023).

4.2.1.6 LLMs as an assistant in the design process

There are several studies that show how well LLMs, such as GPT-4, perform as personal assistants. Writing reports, e-mails, taking notes, mind-mapping, idea generation, and summarizing large portions of text are all examples of what an LLM can help with as an assistant (Picard et al., 2023). However, studies depicting LLM assistants that can be useful in the design process specifically in an aerospace application are far less.

A study at the German Aerospace Center (DLR) by Reitenbach et al. (2024) developed an intelligent workflow engine using an AI chatbot. This system merges traditional workflow management with AI, providing a customizable and user-friendly solution for workflow generation. The chatbot, acting as a natural interface, simplifies interactions and enhances usability. This new methodology is integrated into the GTlab software framework, aimed at process automation and collaboration in aircraft propulsion systems.

The AI chatbot were asked to perform various tasks in the integrated workflow. First, the performance and flexibility of the engine were validated through four sub-workflows, each representing unique challenges with specific features and requirements. These workflows demonstrate the engine's ability to manage tasks, like serial and parallel processing of multiple inputs, highlighted in the following scenarios:

- Case A: ”Calculate the value y if x is known”, tests linear sequences.
- Case B: ”Calculate the value z if q and m are known”, examines parallel structure handling.

The evaluation process measured LLM runtime, total workflow creation time, and success rate across different models, including GPT-3.5-Turbo, GPT-4, and GPT-4-Turbo. The tests, performed under varied settings and times, aimed to eliminate time-specific biases. The newer models, particularly GPT-4 and GPT-4-Turbo, consistently produced the desired outcomes, though complex structures slightly lowered the success rate (Lv et al., 2023).

The system's effectiveness was further highlighted through a detailed example involving the compressor system's thermodynamic calculations. It needed specific inputs like inlet enthalpy (Hin) and entropy (PSIin) to compute outlet conditions such as temperature (Tout) and pressure (Pout). This involved calculating the ideal specific compression work (ΔHid) from the actual work and isentropic efficiency, and the outlet enthalpy (Hout) from these computations. The steps were repeated to validate the accuracy of the methodology in interpreting and structuring complex workflows based on user prompts and inquiry such as ”Predict compressor outlet conditions from compressor inlet parameters, work, and efficiency”.
Moreover, the system demonstrated robustness in adapting to various query structures and technical terminologies. The ability to interpret complex prompts and seamlessly integrate workflow elements shows the systems potential in enhancing technical workflow management (Reitenbach et al., 2024).

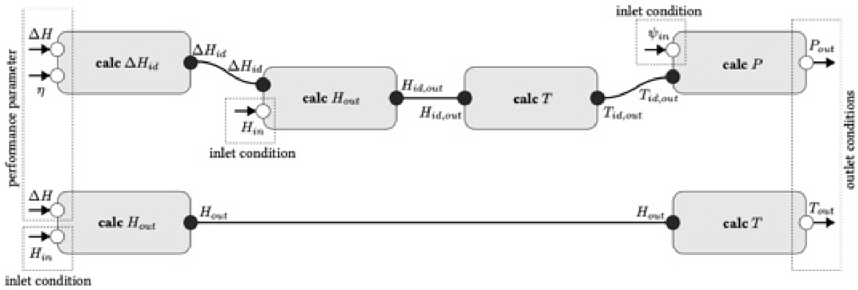

Figure 4.5: The workflow the LLM is asked to perform correct inquiries in (Reitenbach et al., 2024).

However, challenges remain in handling intricate workflows and selecting optimal components from extensive libraries, pointing to future areas for development. These include fine-tuning LLMs for specific tasks and integrating additional functionalities to streamline information retrieval and processing (Lv et al., 2023).

In conclusion, the study by (Reitenbach et al., 2024) showcases the significant potential of AI-enhanced workflow engines in technical domains, with the adaptability and precision required to handle complex engineering tasks effectively.

4.2.2 Limitations with LLMs in Concept Generation and Development

The research identified multiple constraints associated with the application LLMs in concept generation and development. Consensus among the authors reviewed in this study highlights these limitations. The limitations, cited by key articles on "LLMs in Concept Generation and Development," are presented below:
Limitations:

- Interpretability: The generated concepts are not guaranteed to be valid or of good quality, and the development of automatic metrics for evaluating and filtering these concepts is hindered by the lack of a defined quantitative scale for validity (Q. Zhu & Luo, 2023).
- Generalizability: It is challenging to generalize different tasks and associated knowledge necessary for real-world design concept generation due to the need for diverse and extensive datasets (Q. Zhu & Luo, 2023).
- Extendibility: Extending the method to different tasks is limited by the absence of open, high-quality engineering design datasets (Regenwetter et al., 2022).
- Lack of Baselines: The study lacks existing methods for comparison, complicating the evaluation of AI-generated design concepts (Q. Zhu & Luo, 2023).
- Formalizing Engineering Design dialogue: Current approaches to understanding how outputs are formed in deep neural network models are inadequate, as LLMs largely remain black boxes, complicating the formalization of the engineering design dialogue (Beitz et al., 1996; Fricke, 1996).
- Interface Requirements for LLMs: Today's LLMs require engineering software

tools to provide a textual interface, limiting the use of tools that cannot adapt to this requirement (Schick et al., 2023).

- Correlation of Skills and Representations: LLMs struggle to develop representations corresponding to essential engineering skills such as good spatial imagination and a solid understanding of physical processes due to their reliance on textual data (Fricke, 1996).
- Hallucinations: Both ChatGPT and Bard (Google's LLM) Display a significant drawback referred to as "Artificial Hallucinations," where the AI produces seemingly realistic outputs that do not correspond with actual data (Alkaissi & McFarlane, 2023). Specifically, ChatGPT, particularly when processing a large amount of unsupervised data, is prone to such hallucinations (Alkaissi & McFarlane, 2023). This issue emerges from its predictive nature in generating subsequent words, which can lead to content that is inaccurate and hallucinatory (Huh et al., 2023). Concerns about ChatGPT have been raised in critical sectors like education and healthcare, highlighting its inaccuracies and the potential for hallucinated outputs (Hosseini et al., 2023). Research indicates that approximately 45% of responses from ChatGPT are inaccurate (Heck, 2023), and around 30% of research proposals generated by ChatGPT exhibit hallucinated content (Athaluri et al., 2023).

Similarly, the same authors discussed the subsequent steps that is required to be solved to ensure the success of an implementation of this kind:

Future Work:

- Developing an Interpretability Scale: Focus on creating a scale to measure how much sense a generated concept makes and exploring the relationship between interpretability and the concepts' novelty and usefulness (Q. Zhu & Luo, 2023).
- Improving Generalization Capabilities: Addressing the limitation of generalization in AI models by potentially using a dataset with adequate design knowledge and reasoning (Q. Zhu & Luo, 2023).
- Expanding Framework Applicability: Customizing available data to teach additional design reasonings and extend the framework's applicability (Regenwetter et al., 2022).
- Specifying Evaluation Metrics and Datasets: There is a call for design-specific metrics and public datasets to evaluate and compare models' performance, especially in participatory tasks within the argumentative design dialogue (Ma et al., 2023; Regenwetter et al., 2023; Song, Miller, & Ahmed, 2023; Q. Zhu & Luo, 2022).
- Enhancing Machine-Actionable Interfaces: Future research should focus on adapting interfaces of specialized engineering software to meet the evolving needs of LLMs and multi-modal models (Schick et al., 2023).
- Refining Human and Computational Evaluation Metrics: Establish more refined rubrics and explore new methods for computational assessments of design solutions. There's a need to assess the usefulness and feasibility of designs, which currently rely on expert evaluations (Ma et al., 2023).
- Investigating Prompt Engineering Techniques: Future studies should inves-

tigate how designers can use existing prompt engineering techniques to aid early-stage design ideation and explore other LLMs' effectiveness using developed methods (J. Wei et al., 2023).

4.2.3 AI in aerospace applications

The study performed by Hassan et al. (2024), published in April 2024, a review over 300 publications in the area of AI and aerospace. An interesting and highly important note is that this publication does not mention LLMs or NLPs at all. Furthermore, the authors highlight a notable absence of AI/ML methods in the specific area of design, addressing this under ”Research Gaps: Design and Optimization” which is particularly relevant to this thesis. They also outline potential research areas for AI/ML in design and optimization, such as innovational constraints, suboptimal solutions, and inefficient resource utilization.

In the context of AI in the aerospace sector, prior research has explored motivators to AI adoption, overall technological, security, and economic factors. Studies have identified machine learning Yairi et al. (2017) as essential for the successful integration of AI in aerospace (Aksit et al., 2023), particularly noting the sector's data-rich environments is favorable for deep learning. Predictive analytics from machine learning are acknowledged for their role in advancing new aircraft technologies and commercial innovations.

Machine learning and soft computing are pivotal in enhancing aerospace firms capabilities to analyze consumer needs, evaluate new product market potential, and uncover new market opportunities (Ahmad, 2019; Guraksin & Ozcan, 2023). These technologies allow firms to utilize big data to create customized services and products (M. Wang et al., 2019). Despite the increasing need for data-driven insights, the swift advancement of AI technology poses challenges to the timeline and extent of AI adoption. AI adoption in aerospace could revolutionize business models but also introduces substantial challenges such as significant R&D investments, the establishment of industry-wide AI standards, and ethical concerns including job displacement and data privacy (Hassan et al., 2024).

Incorporating AI into aircraft systems offers significant benefits but also presents new risks that must be carefully managed. An exhaustive strategy covering the entire AI lifecycle—from data collection and model development to deployment is vital for mitigating risks and ensuring AI's safe application in aerospace (Becue et al., 2021). Effective data governance and data quality are crucial to maintain the reliability of AI systems. Ensuring data integrity to train and operate AI systems necessitates robust data governance, with data-cleaning techniques crucial to eliminate biases and errors from AI models (Mirchandani & Adhikari, 2020). Ongoing investment in AI research is required to meet the rigorous safety and security demands of the aerospace industry Ali et al. (2020), enabling the sector to protect against AI-related risks while capitalizing on potential advancements in transportation, production, and exploration.

(Hassan et al., 2024) also mentions a few imitations of integration of AI in aerospace companies. Stating that integrating AI is an incremental process requiring a process-oriented approach to fully comprehend factors influencing its acceptance and impacts. Longitudinal studies are advocated to grasp the cognitive dimensions of innovation and decision-making. Employing hybrid methods like sequential exploratory strategies can address identified research gaps. Additionally, interdisciplinary research combining various engineering fields is recommended to enrich understanding and application of AI in aerospace.

4.2.4 Integration and use of AI in aerospace applications

In the article by Reitenbach et al. (2024) they seek to create a enhanced workflow management system (WfMS) in aerospace application by implementing an AI ChatBot. When integrating they used the API RESTful provided by OpenAI. REST (Representational State Transfer) is a widely used architecture for communication between distributed systems using the HTTP protocol. This integration involves the implementation of a special class, the LLM-Handler, within the workflow engine. The LLM-Handler is responsible for managing interactions with the LLM, including the preparation and processing of prompts. To access the OpenAI API, users of the framework must store a personal authentication key, this key is essential for ensuring secure and personalized interaction with the system. In addition, users can select the specific LLM model they want to work with from a range of available options (e.g. GPT-4, GPT-3, etc.). Depending on this selection, the LLM-Handler configures the appropriate API endpoint. The LLM-Handler also provides the ability to adjust various parameters of the LLM, such as temperature and top_p These adjustments allow the response dynamics of the LLM to be more finely controlled and optimized (Holtzman et al., 2019). The temperature parameter in the context of LLMs such as GPT-4 is a crucial factor in controlling the creativity and unpredictability of the produced responses. This parameter is part of the probabilistic nature of LLMs and has a significant influence on how the model responds to a given prompt. The temperature parameter can typically be set between 0 and 2. At lower values, the model tends to produce more certain, predictable and focused responses. This means that the model is more likely to generate text that more closely matches the most common patterns in the training data. Low values of temperature are appropriate for applications where accuracy and clarity are important. Higher values encourage more creative, varied and unpredictable responses. The model is therefore more willing to take risks and use unusual or less common word combinations. High values of temperature are useful when creativity and variety in responses are desired (Reitenbach et al., 2024). Moreover, in the integration performed at DRL, they use the Model Based System Engineering (MBSE) system.

Within MBSE, the necessary vocabulary for the creation of system models can be provided by the Systems Modelling Language (SysML). SysML, a standardized domain specification language for modeling complex systems, was developed by the

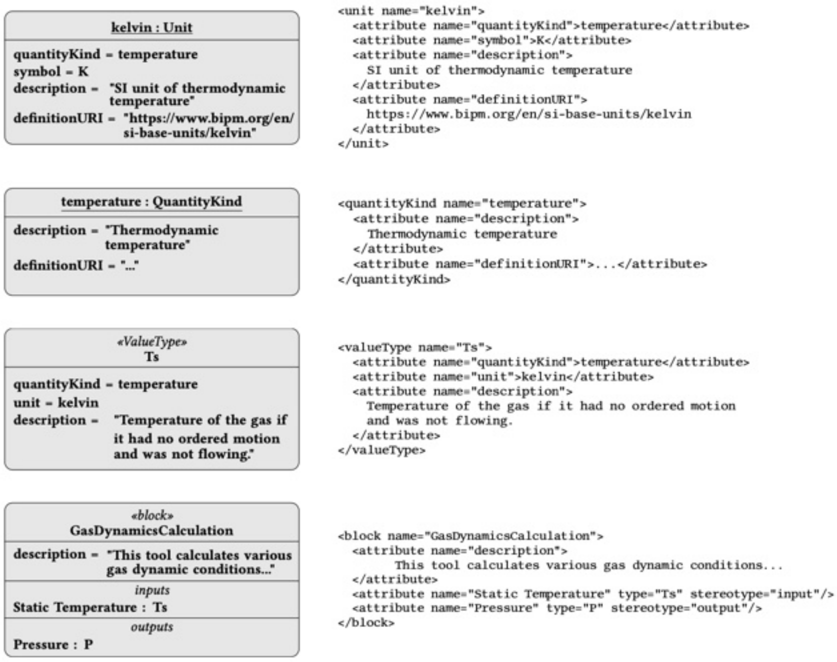

Fig. 2 SysML Serialization to XML-Format

Figure 4.6: SysML to XML conversion (Reitenbach et al., 2024).

Object Management Group. It originated from the adaptation of the Unified Modeling Language (UML), which is a widely recognized language for software modeling. In the GTlab framework, the MBSE paradigm is facilitated through a specialized data processor tasked with organizing and managing the data. All data objects conform to a predefined, standardized object type, ensuring consistent data structuring (Reitenbach et al., 2024). New data objects are integrated using SysML block definitions, which provide a structured and standardised way to extend the system. Then, the SysML format is converted into XML-code in order to provide the LLM with all relevant information. In addition, this includes instructions to the format of the response, specifically structuring the proposed workflow information in JSON (JavaScript Object Notation) format (see Figure 4.6).

These format specifications are essential as they enable the workflow engine to parse the LLMs responses and prepare them for further processing. JSON is a data exchange format that is both readable and writable by humans. It is also parsable and generatable by machines. JSON has become a widely accepted, language-independent standard used in a wide variety of applications and services to exchange data. The advantage of using the JSON format in this case is that there is no need to instruct the LLM in a complicated format definition. This reduces both the context size and the error inclination.

Output verification: in the extended workflow engine, the validation of LLM responses by the validator is essential for ensuring accurate workflow configuration. Initially, the LLM-Handler receives a response string, which is processed to extract JSON data. This data, containing details about workflow elements and their con-

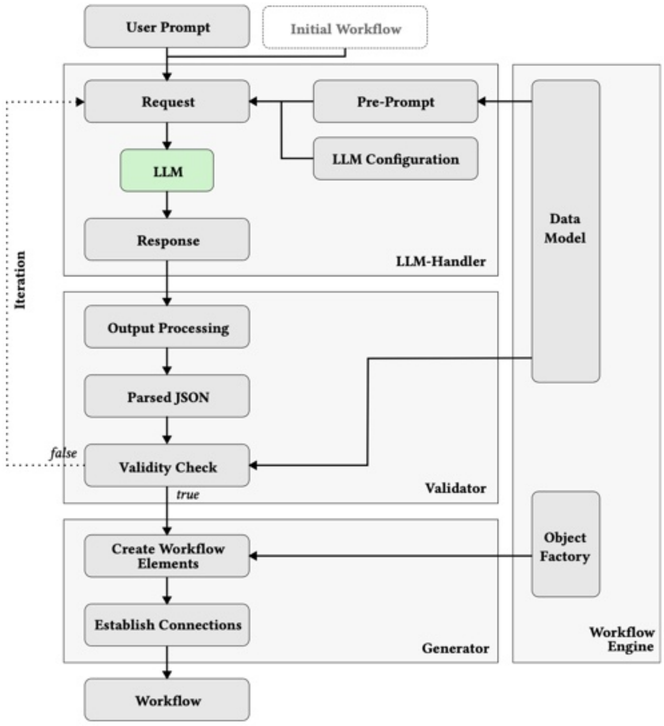

Figure 4.7: The proposed workflow with an integrated LLM, and LLM-handler (Reitenbach et al., 2024).

nections, is crucial for configuring the workflow accurately.

The validator then checks this JSON data to ensure all components and connections suggested by the LLM are feasible and correct, utilizing meta-information from SysML specifications. If discrepancies are found, the validator produces clear, human-readable error messages, pinpointing issues like non-existent tools or incompatible connections. This clarity in communication leverages the LLM's strength in processing natural language, facilitating a better understanding of errors and enabling more precise adjustments in subsequent responses. These error messages are sent back to the LLM-Handler, prompting the LLM to refine its suggestions. This iterative process continues until a valid workflow is achieved or a set number of iterations is reached. If validation fails beyond this point, the workflow creation is deemed unsuccessful.

Upon successful validation, the workflow factory initiates the actual creation of the workflow based on the LLM's validated suggestions. This rigorous validation process ensures that only feasible and correctly configured workflows advance, enhancing the reliability and efficiency of the entire workflow creation process.

Data Collection, Cleaning, and Annotation: the creation of an annotated aerospace corpus is of immense importance because it serves as a labeled dataset to fine-tune the LLM (Tikayat Ray et al., 2024). The first step toward creating a corpus consists of collecting aerospace-domain texts. Since BERT is pretrained on English Wikipedia and BookCorpus (Devlin et al., 2019), it already has some exposure to the aerospace domain from articles about aircraft, airlines, aircraft manufacturers, aviation safety, etc. Hence, this effort augments BERT's aerospace language understanding by collecting scientific aerospace texts and narrows the focus to requirements by gathering examples of requirements from the Federal Aviation Regulations (FARs). (Tikayat Ray et al., 2024)

Tikayat Ray et al. (2024) have performed a study where the LLM BERT was assigned to extract aerospace requirements from text. In the study, they have documented how they have fine-tuned the pre-trained BERT to gain the best possible results in the study. The result is in this study aeroBERT. Different variants of aeroBERT were developed by fine-tuning the BERT variants on the annotated corpus.

A test set of 20 aerospace requirements was created to test aeroBERT and demonstrate the automatic creation of a glossary of terms. The requirements was grouped in the different categories System (e.g exhaust heat exchangers, powerplant, auxiliary power unit), Resource (e.g Section 25–341, Sections 25–173 through 25–177, Part 23 subpart B), Organization (e.g DOD, Ames Research Center, NOAA), Value (e.g 1.2 percent, 400 feet, 10 to 19 passengers), and Date-time (e.g 2013, 2019, May 11,1991).

aeroBERT was able to identify 62.5% of the SYS, 100 % of the RES, and 50% of the DATETIME named entities. Finally, the performance of aeroBERT-NER and BERT aerospace text was compared. aeroBERT was able to identify 71% (32 out of 45) of the relevant named entities. BERT, however, was unable to identify any named entities apart from two subwords (Tikayat Ray et al., 2024). To be of value, however, a language model needs to be fine-tuned to recognize aerospace-specific terms (Tikayat Ray et al., 2024).

4.2.5 Prompt engineering

The result of the LLM depends essentially on how precisely and effectively instructions or questions are formulated. Prompt engineering is about formulating content in such a way that the LLM achieves the best possible results (P. Liu et al., 2023; White et al., 2023).

In the case of Reitenbach et al. (2024) study, the user input, namely the request to the AI to create or modify a specific workflow, is sent to the LLM-Handler via a corresponding input panel within the graphical user interface of the GTlab framework. The user can also select existing workflows or sub-workflows and attach them to the request. This initial workflow setup can be modified or extended by the LLM depending on the user instructions.

The LLM handler utilises the functionalities of the workflow engine's data model to transform the setup into a corresponding text format, which is also used for the output created by the LLM for the proposed workflow. In order for the LLM to understand how to handle the user input, the corresponding context must first be established in a series of preceding prompts before this input is processed. This special prompt engineering technique is referred to as chain of thought (J. Wei et al., 2023).

J. Wei et al. (2023) discovered that this approach improves the LLM ability to answer a complex problem by about 58%. Consequently, the information and instructions of the pre-prompt of the LLM-Handler are transmitted to the LLM in individual steps. The respective response of the LLM is appended to each subsequent transmission in the background in order to utilise the additional thoughts gained by the LLM. It turned out that the chain of thought approach and the way the LLM can respond greatly increases the success rate of the creation of the desired workflow.

4.2.6 Information and knowledge reuse

Effective design processes rely on skilled personnel, efficient information gathering, and knowledge reuse to produce high-quality outputs, emphasizing the critical importance of these elements (Tikayat Ray et al., 2024). In aerospace, where requirements engineering is complex and critical, the ambiguity of natural language can lead to costly errors. Early detection and rectification of these errors during the requirements generation phase are essential (Dalpiaz et al., 2018). A NASA study underscores the high costs of addressing late-stage issues in requirements engineering, emphasizing the need for early intervention (Haskins et al., 2004). Industry data also shows that many defects and rework efforts stem from errors in requirements engineering, with safety-critical systems being particularly vulnerable (Firesmith, 2007).

Given the complexity of modern systems, in this context MBSE systems, an integrated approach to system development is crucial to manage the challenges of large-scale systems effectively (Ramos et al., 2011). Additionally, as systems are used over longer lifespans, there is a significant need for reuse of design and analysis documents to support, maintain, and modernize these systems. Current methods do not fully capitalize on the reuse of engineering knowledge and artifacts, necessitating a shift to new management methods for engineering files, documents, analysis, and data packages across enterprises. If successful, this could enhance agility in design, manufacturing, and system deployment, reducing costs and risks (Kambhampaty et al., 2024).

Organizations are reevaluating how information is stored and retrieved throughout a product's lifecycle due to the increasing complexity of multidisciplinary designs (Fafchamps, 1994). Reuse, defined as any process that applies previously generated work products to new, unrealized goals, plays a crucial role in this context (Sametinger, 1997). Learning from past work to evolve system designs is a common form of reuse in science and engineering, often requiring rework and adaptation (Trujillo et al., 2020).

Reusing components, models, data, and knowledge brings numerous benefits. It increases confidence in product performance, enhances engineering efficiency, and speeds up the acquisition process by focusing efforts on creating new work products rather than rediscovering or recreating existing ones (Kambhampaty et al., 2024). Reuse also aids in knowledge integration across different research and development environments, fostering cross-pollination between directories (Nightingale, 2000). Moreover, it reduces the risk in systems by minimizing the need to design components from scratch (Cusumano & Nobeoka, 1998). Leveraging existing tools and data relevant to a program decreases research and development costs and time, streamlining the path to market (Kotlarsky et al., 2008). With structured and accessible information, less time is spent on version control discussions, thereby accelerating market readiness (Kambhampaty et al., 2024).

LLMs can be a facilitator for knowledge reuse and has the potential to improve automation in knowledge engineering work due to the richness of their training data and their performance at solving NLP tasks (Walker et al., 2024).

Example use case:
In collaboration with the U.S Air Force Research Laboratory, the case study by (Kambhampaty et al., 2024) employs a novel methodology for reusing a vehicle model within the RAAGE datapackage. This approach involves developing a Data Curation system that leverages a graph database built from an uncurated data package, enabling effective extraction and organization of metadata and relationships.

The methodology's ultimate goal is to quantify labor hours saved by applying Digital Engineering and Digital Curation (DC) principles. This is achieved by developing and implementing a practical methodology for curating data packages and facilitating information sharing within an organization. To assess its effectiveness, key metrics such as precision for information retrieval tasks and qualitative usability ratings are introduced.

The RAAGE dataset, stemming from previous AFRL-sponsored research, includes digital artifacts relevant to the design, construction, and operation of attritable air vehicles, supporting air interdiction (AI) and intelligence, surveillance, and reconnaissance (ISR) drone missions. This dataset aids in evaluating proposed design solutions by aligning with sponsor needs and inputting specific architectural requirements.

The study contrasts two reuse approaches: a traditional reuse case and a curated effort using the new methodology. In the traditional approach, engineers rely on existing repository structures and associated reports, navigating through files using standard tools. In the curated approach, a graph database with Neo4j visualization is utilized, improving access to metadata and relationships.

Three reuse tasks assess the impact of extending an aircraft's range by 20%:

1. Designing an aircraft capable of completing a mission with the increased range.

2. Estimating the cost per aircraft for the enhanced range.
3. Calculating the campaign cost per kill for the extended range.

The study by Kambhampaty et al. (2024) finds that several tasks in their pipeline are well-suited to machine-learning frameworks, particularly those involving the classification of metadata keys. This often involves designing representations of informational constructs for models to perform classification tasks effectively. Advances in large-language models, especially zero-shot classifiers, suggest potential for automating much of this process. However, robust methods are needed to assess these models' performance against human metadata writers in classification tasks.

Additionally, the challenge of deduplicating data when used as training material for these models is noted, alongside questions about the most suitable tasks (classification, generation, graph-based property prediction) for the metadata currently being generated. Over time, the aggregation of data regarding the usefulness and the number of times artifacts are reused could inform the development of predictive models. These models could forecast the utility of new or unlabeled artifacts based on metrics like the Summation of Search Queries.

Despite the boredom associated with metadata creation and access, significant benefits are observed when these tasks are integrated and visualized within a graph model, demonstrating the value of this approach in enhancing data management and usability.

4.2.6.1 LLMs in complex system design, use case on aeroengine components

Previously mentioned (Gomez et al., 2024) explores the possibility of generating CAD models by utilizing LLMs. Two use cases are presented, highlighting design processes and results with LLMs. These cases, relevant to aerospace, involve an aircraft OEM and a first-tier aeroengine component supplier. The OEM's use case demonstrates visualizing system architectures with UML diagrams. The supplier's case shows LLMs generating aerospace CAD models through verbal instructions.

ChatGPT can automate configuration rules for architectural design, generating system architectures from required functions, and was used in the study by (Gomez et al., 2024). Using an aircraft hydraulic system as a case study, the process includes defining configuration rules, translating them into UML diagrams, and generating Python code to produce design diagrams. The deterministic and inspectable nature of the code is crucial.

The geometry generation use case demonstrates using LLMs to update aerospace structural components, like modifying a flange geometry on a Turbine Rear Structure (TRS). LLMs enable non-CAD experts to update component geometry quickly using verbal instructions, interacting with CAD software through a Knowledge-Based Engineering (KBE) system (ParaPy). The process involves:

- Model Conditioning: Provide general instructions on the task.

- Few-shot Prompting: Describe geometrical steps and corresponding KBE code.
- Use Case Execution: Verbalize new geometrical definitions and receive corresponding KBE code from the LLM.

This method allows instant CAD geometry updates without requiring team members to be CAD or KBE experts, as demonstrated by automatically generating CAD and Finite Element Models for the TRS component.

The results from Case Study 1 demonstrate that the model can understand user intentions, execute tasks, generalize concepts, and has knowledge in diverse areas like UML and Python. Case Study 2 results shows that the GPT-4 model understands geometrical entities and relationships and can generate KBE code with minimal user input. These capabilities are available without custom machine learning training, supporting the hypothesis that LLMs can assist designers with minimal input.

This study confirms the success of using LLMs in aerospace concept generation, though several iterations and prompt versions were needed to generate valid responses. In aerospace applications, repeatability and solid justification are essential for certification and airworthiness, posing additional challenges. More research is necessary to explore how LLMs can interact with designers without undermining their responsibilities.

Engineering designers must remain accountable for their designs, and LLM-generated results cannot be fully trusted on their own. This risk can be mitigated by combining existing engineering tools with LLMs: tools perform calculations, while LLMs configure them. Although LLMs can generalize and adapt, they lack awareness of specific company or project procedures. Embedding necessary knowledge through fine-tuning the model would be beneficial.

4.2.7 Example: ChatGPT-4V capabilities in design process environments

In a study by (Picard et al., 2023), GPT-4V was tasked with various design process activities. Previous research by (Ma et al., 2023; Q. Zhu et al., 2023) examined earlier versions like GPT-2 and GPT-3. Another study by (Reitenbach et al., 2024) found GPT-3.5's performance unsatisfactory. Thus, evaluating GPT-4V, the latest vision-based model, is crucial for design tasks. GPT-4V boasts over 100 trillion parameters compared to GPT-3's 175 billion, enhancing its capability to handle complex tasks and generate nuanced text. It excels in "few-shot" and "zero-shot" scenarios, understanding nuanced instructions and aligning responses with user intentions, while reducing hallucinations and bias.

The study examines GPT-4V's effectiveness in the areas, *Design Descriptions, Material Selection, Engineering Drawing Analysis, CAD generation, Topology Optimization, Fluid Dynamics Simulation, and Design for Manufacturing.*

Design description and sketches

GPT-4V excelled at matching design sketches to textual descriptions and generating descriptions from sketches. With complete sketches, including handwritten text, it matched descriptions perfectly. Without the text, accuracy averaged 5.33/10, highlighting the importance of combining text and visuals. Removing the "None of the above" option improved accuracy to 7/10.
The model also effectively generated descriptions from low-quality sketches, accurately describing form and function, such as a belt and pulley system. However, engineers should verify outputs to avoid hallucinations. GPT-4V's capabilities can aid in creating searchable design catalogs and generating multimodal datasets.

Material selection:
GPT-4V performed well in general material selection but struggled with specific numerical criteria. It shows promise for broad material family selection in engineering design, particularly in preliminary phases and as an educational tool. However, it needs improvement in handling precise data and complex synthesis, requiring careful oversight in practical applications.

Engineering drawings and CAD generation:
GPT-4V showed mixed results in interpreting and generating CAD from engineering drawings. While it generally recognized components, it often misinterpreted details, such as consistently misidentifying a blind hole as a through hole, achieving accuracy in only one out of nine experiments. However, it successfully extracted all required dimensions in every trial, labeling them correctly 66% of the time, resulting in an average performance score of 96% for dimension extraction and labeling.

In CAD generation, GPT-4V struggled, producing accurate models on the first attempt in only one of nine cases. Subsequent iterations failed to improve accuracy, with persistent errors, such as confusion about hole extrusion direction due to inconsistent dimension labeling.

These findings suggest that while GPT-4V can assist with preliminary design, its precision in detailed CAD tasks is limited. Future research should focus on improving GPT-4V's precision in interpreting detailed drawings and generating accurate CAD models, enhancing training, integration with CAD software, and iterative feedback mechanisms.
Quantitative performance metrics from the experiments showed:

- GPT-4V correctly described a part with a hole in 1/9 experiments when it recognized a "rectangular block with a cylindrical hole."
- In dimension extraction tasks, GPT-4V perfectly scored in 6/9 experiments for extracting dimensions, though it often mislabeled them, especially the hole depth.
- CAD generation was notably poor, with correct CAD produced on the first attempt in only 1/9 experiments using CadQuery, and persistent errors in subsequent iterations across different scripting languages.

Topology Optimization:
GPT-4V shows a basic understanding of topology optimization principles but strug-

gles with detailed interpretations and accurate linkage of diagram sections. It often misinterprets load depictions and specific diagram elements, indicating a superficial grasp of complex visual data.

In precise quantitative tasks like estimating material percentages, GPT-4V's initial attempts were inaccurate. However, using a Python script improved its accuracy, suggesting coding tools can enhance its performance. GPT-4V can identify structural elements in optimization diagrams and respond to engineering queries, but it sometimes fails to recognize the importance of floating or disconnected components.

Fluid Dynamics
GPT-4V demonstrates a strong understanding of fluid dynamics, effectively identifying and analyzing key parameters and flow regimes. However, it has limitations in applying this knowledge precisely to specific simulations. GPT-4V effectively distinguishes between laminar and turbulent flows in CFD simulations, correctly identifies transient regimes, and interprets parameters like Reynolds and Mach numbers. It can compute accurate values using given data, such as estimating the Reynolds number correctly.

In laminar flow analysis, GPT-4V correctly identifies features like Mach number and slower flow patterns but inaccurately suggests shock waves at a maximum Mach number of 0.3, indicating a gap between theoretical knowledge and practical application.

In transient regime analysis, GPT-4V accurately recognizes vortices and transitional flow behavior around an airfoil, noting significant boundary layer effects and the mix of laminar and turbulent flows. It prudently seeks more data when uncertain.

GPT-4V's ability to process and analyze visual information from CFD outputs is promising but inconsistent. While it identifies general dynamics and parameters accurately, it struggles with specific simulation details and sometimes misinterprets shock waves and boundary layer effects. Its vision capabilities enhance its utility in engineering tasks requiring both theoretical knowledge and visual data interpretation.

Design for Manufacturing (Design for Addative Manufacturing)
GPT-4V's performance in Design for Manufacturing (DfM) tasks was limited, particularly in additive manufacturing contexts. The model never fully answered any of the DfM queries accurately, often providing overly cautious or incorrect responses regarding manufacturability.

Additive Manufacturing
GPT-4V consistently predicted that designs would not be manufacturable by additive methods, regardless of their actual feasibility. This uniform negative response occurred across all 60 queries. Even for designs adhering to 3D printing rules, GPT-4V incorrectly suggested violations. It struggled to identify rules by their assigned numbers, correctly identifying violations in only 13 out of 30 problematic designs.

For subtractive manufacturing, GPT-4V identified basic machining features in images correctly in 12 out of 20 cases but performed inconsistently, especially with complex designs. This led to potential misjudgments where precision is crucial.

Overall, GPT-4V's overly cautious approach in additive manufacturing led to overestimations of challenges and inaccurate assessments. In subtractive tasks, while it recognized some features, its inability to handle complex geometries indicates it is not yet reliable for detailed technical evaluations without human oversight.

4.2.8 ZDM with AI assistance

The authors (Leberruyer et al., 2023) highlights that there are several different tools, techniques, technologies, and methods for working with ZDM. The Table 4.1 below illustrates the main techniques that the authors (Caiazzo et al., 2022; Foivos Psarommatis & Kiritsis, 2020; Powell et al., 2022) have identified as having increased interest with the introduction of Industry 4.0.

ZDM Techniques	Description
Artificial intelligence	Data-driven techniques for automated data analysis and decision making
Architecture and Standards	Integration and communication protocols of industrial software
Big data analytics	Elaboration, analysis, and visualization of the massive amount of industrial data
Cyber-Physical Systems and digital solutions	Control strategies combining physical and digital resources
Digital inspection and monitoring	Solutions for the measurement and monitoring of product and process resources
Digital Twin combined with simulation and modelling	Optimization and decision support for processes
Extended Reality and visualization technology	Visualization of information to improve decision making processes
Failure Mode and Effect Analysis	Approach for identifying possible failures in a design, a manufacturing or assembly process, or a product or service.
Process mining	Providing better understanding of process variations that can be decreased and improved

Table 4.1: Table showing the different ZDM techniques the meaning of them.

4.3 Concept evaluation

> Fundamentally, the main purpose of design work is ultimately to "devise courses of action aimed at changing existing situations into preferred ones" - Simon (2019)

4.3.1 AI supported decision-making and evaluation

The initial milestone in AI development was the creation of an intelligent machine capable of emulating human decision-making in chess game play. Since then, the utilization of AI in decision-making processes has stood out as one of the most significant applications in the history of AI (Dwivedi et al., 2021). Researchers in design has thus for a long time also acknowledged the potential of AI tools to revolutionize the practices of engineers in the design field. For instance, design engineers can harness AI to facilitate more automated and intelligent extraction and representation of knowledge, aiding in early-stage design ideation and to uncover solutions to challenges that for a long time remained unsolved (Allison et al., 2022).

Moreover, current AI-algorithms are capable to significantly augment the effectiveness of later-stage system design tasks, particularly those entailing complex, high-dimensional and interrelated detailed design decisions (Allison et al., 2022). AI is argued to play a pivotal role in aiding designers to assess and enhance design concepts according to predefined criteria, including cost, performance, and manufacturability. Through sophisticated algorithms, AI can scrutinize design concepts, proposing alterations or enhancements to optimize them for the desired outcome. By analyzing data and offering insights into the design process, AI empowers designers to make well-informed decisions regarding design choices (Khaleel et al., 2023). Today early phases are generally heavily dependent and influenced from the experience and intuition of designers, and this can create constraints on the design-process due to human cognitive limits. This creates opportunity for AI to support, where human processing capabilities are note sufficient (Allison et al., 2022).

An article by Yüksel et al. (2023) wanted to investigate current progress in AI applications related to design engineering and concept development, during the last 15 years. They found that for a considerable duration, expert systems, fuzzy logic, artificial neural networks, and genetic algorithms have been the predominant methods utilized in the evaluation and optimization processes of design. Nonetheless, the utilization of modern data-driven techniques such as machine learning and its sub-discipline, deep learning, has surged in recent times within the design domain. Various other AI methodologies can also be deployed across different stages of the design process, including idea generation, concept formulation, evaluation, optimization, and decision-making processes (Yüksel et al., 2023).

Additionally, the findings of Picard et al. (2023), indicate that GPT-4V can effectively assist human designers by identifying crucial factors in the design process. While GPT-4V can generate criteria similar to traditional methods, its outputs often require further refinement, such as the categorization of subcriteria. Although GPT-4V grasps the concept of a Pugh chart and formats it correctly, its hesitance to complete the chart without extensive data highlights a limitation. This suggests that while GPT-4V is helpful for organizing and starting the concept selection process, human input is essential for thorough analysis and decision-making. For practitioners, this means VLMs like GPT-4V are useful in early design evaluations but need careful management and more information for complex decision-making tasks.

4.3.1.1 Human bias mitigation

No two designers are the same and abilities and preferred styles of working can differ a lot. Experience level also heavily influences the early development process. Thus, early development is influenced by several human factors, such as intellectual abilities, domain knowledge, detail oriented or not, attitudes towards risk-taking etc. (Lubert, 2005).

An article by Pan and Zhang (2021) investigating AI in construction engineering management found several benefits from AI, one of which being the ability to manage human bias. When manual labor is part of the work process, there is a risk of human human bias impacting that process. According to the authors, AI can automate project management and remove human obstacles such as bias, making it more objective. The technology behind this is machine learning algorithms, utilized to intelligently analyze extensive data sets, uncovering hidden knowledge. These algorithms facilitate automated data-analysis and decision-making, by being integrated into project management software. This ultimately provide tacit knowledge from previous projects and enhancing understanding of the current project (Pan & Zhang, 2021). While this work does not provide a technical take on how human bias is removed, it indicates a general capability of AI.

An article by Yuan et al. (2022) looked at state-of-the-art in concept evaluation and found that performing subjective concept evaluation methods demand substantial manual input, thereby potentially constraining the range of concepts that can be feasibly assessed. Yuan et al. (2022) acknowledges several types of ranking methods that have been tried to evaluate concepts but few have incorporated data-driven approach to incorporate user needs. They claim that current state-of-the-art in concept evaluation still is based on subjectively choosing concepts based on the judgement and expertise of designers. The case studied is footwear.

A comprehensive multimodal design evaluation regression model is proposed (DMDE) by Yuan et al. (2022), aimed at providing designers with a precise and scalable forecast of the overall desirability and attribute-level characteristics of novel concepts. This prediction is derived from extensive user reviews of existing designs. The DMDE model works by utilizing a cutting-edge deep neural network-based model, ResNet-50, pre-trained on Image-Net and later fine-tuned to process orthographic views of existing products. Textual inputs in the shape of product descriptions is also processed by the model using a cutting-edge deep language model, a bidirectional encoder from transformer, which undergoes fine-tuning on an extensive product description dataset. Images and text is thus the training data. The result is a better informed concept evaluation process with promising performance (Yuan et al., 2022).

Traditionally, conceptual design within aerospace is also influenced by subjectivity. Crossley (1999) studied conceptual design within the aerospace discipline and state that even though designers are provided with several computational tools to assist in design evaluation, human subjectivity tend to influence the process. The author acknowledges that out of all design phases, the conceptual design phase is the most difficult as there is typically no starting point. They mean that a set of requirements lacking physical representation must be converted into a device that meets these requirements. Creating this starting point necessitates that the designer employ experience and intuition (Crossley, 1999). This statement is further recognized by Eres et al. (2014), studying the challenging conceptual design process of aircraft and aircraft components. They bring up decision-making for concept selection and highlight some traditional methods such as "Pugh charts" and "Quality function deployment" (QFD), the latter being the most widely used and accepted method. However, QFD relies heavily on subjective judgment when building a house of quality (Eres et al., 2014). Furthermore, conceptual design within aerospace engineering tend to be influenced by dynamic customer requirements (Mavris & Pinon, 2012). Requirements within this discipline are also usually in conflict with each other, making it difficult to come to a decision on what concept to peruse with. Decisions also take into consideration more than just technical aspects. Consequently, the final solution may vary from one decision maker's perspective to another (Mavris & Pinon, 2012).

An article by Khaleel et al. (2023) looked into AI applications in design engineering and acknowledges that conventional concept evaluation can be extremely time consuming were AI methods enable objective and swift evaluations, leading to substantial time and cost savings. The authors also recognize optimization and finding the best product configuration based on several constraints to be an area where designers are finding new ways to incorporate artificial neural networks, swarm intelligence and machine learning. In this early design phase, there exist no starting point which means that

Worth noting is that more articles (Camburn, He, et al. (2020); Yüksel et al. (2023); Allison et al. (2022)) continuously brings up the concept of human bias and subjectivity and how AI can assist in managing this.

4.3.1.2 Autonomous decision making

One of the major criticism drawbacks of AI and is that it can be equated with a "black box", resulting in designers feeling scepticism and reluctance towards accepting the results of an AI (Liao et al., 2020). Although the advantages of AI and ML are evident for certain uses, the intrinsic challenge of explaining the actions of these algorithms poses obstacles to potentially implement them, particularly in civil aviation. Without the ability to properly clarify AI decisions or foresee ML outcomes, it is hard to ensure safety or guarantee system performance. While traditional software within civil aviation is deterministic as is is bounded by physics, relying on mathematical stability, AI technologies are considered non-deterministic (Tejasen et al., 2022). Despite this, there are still research that indicates that AI can turn decision-making into an automated process.

An article by Verganti et al. (2020) wanted to look at what happens to innovation and the design process, when decisions that to this day has been taken by humans, start being taken by AI. Their conclusion is that AI is capable of revolutionizing decision-making, by making this an autonomous process, thus driving innovation. By surpassing the constraints of human-centric design, AI can enhance performance in customer centricity, creativity, and speed of innovation. Verganti et al. (2020) explain that the way to achieve autonomous decision-making is through problem solving loops which are AI and ML algorithms/engines based on supervised and unsupervised learning. Problem-solving loops in AI systems autonomously gather real-time data, generating tailored solutions without human input. They improve predictions over time, replacing human effort and scaling easily for diverse solutions with minimal R&D investment. However, to seize the potential of AI-supported decision-making necessitates a fundamental reevaluation of innovation strategies within the organization by manager, according to the (Verganti et al., 2020).

In AI-driven setups, human involvement shifts, from crafting complete solutions to identifying meaningful innovation challenges, framing innovation endeavors, and establishing the necessary software, data infrastructure, and problem-solving mechanisms to address them in real-time. With AI, the purpose of the human shifts from problem solving, to problem finding. As AI is capable of entering the creative space, human design becomes more revolved around sense-making, and understand what problems to focus on, thus bringing design closer to leadership (Verganti et al., 2020). Camburn, He, et al. (2020) acknowledges that concept evaluation is difficult and complex, with perhaps thousands of concepts to take into consideration and as response, they propose an automated solution to design concept assessment of concepts written in natural language. The method was tested empirically and works by:

1. Extracting ontological data from design concepts, utilizing machine learning.
2. Based on ontological data, quantitative metrics and filtering strategy is used to create a creativity rating.

The proposed method by Camburn, He, et al. (2020) offers a potential approach to objectively rate design concepts. Notably, a subset of designs automatically chosen from a vast pool of candidates received higher scores compared to a subset selected by humans, as evaluated by third-party raters. According to the authors, these findings suggest the presence of bias in human design concept selection and advocate for additional research in this area, as there still are risks with this approach (Camburn, He, et al., 2020).

4.3.1.3 Risk and uncertainty management

An article by Mavris and Pinon (2012) studied design challenges within aerospace engineering. Design is an activity focused on problem-solving, where requirements are translated into functions, resulting in decisions that shape a solution tailored to specific needs. This process involves exploring a wide range of possibilities, necessitating the accumulation of knowledge and an understanding of the associated constraints and trade-offs. The design process progresses through the stages of Conceptual, Preliminary, and Detailed design, with each phase increasing in the level of detail and complexity in representations and analyses. As a result, the scope and thoroughness of the analyses and trade-offs, as well as their uncertainty and precision, differ markedly between phases. For instance, Preliminary design employs more precise analyses and tools compared to Conceptual design. Additionally, Conceptual design deals with higher uncertainty (Mavris & Pinon, 2012). Todorov et al. (2022) confirms that deciding upon optimal concepts within the aerospace domain, necessitates evaluating a substantial number of competing engineering solutions and occurs in an environment of uncertainty.

According to Mavris and Pinon (2012), uncertainty seem to heavily influence conceptual design in aerospace engineering. During this stage, there are many different types of uncertainties, at several different levels. For example these uncertainties can originate from:

- Approximations
- Simplifications
- Abstractions
- Estimates
- Lack of knowledge about the problem
- Ambiguous design requirements
- Omitted physics and unaccounted features
- Incomplete information about the operational environment and the technologies available
- Unknown boundary conditions or initial conditions
- Prediction accuracy of the models

Mavris and Pinon (2012) continue to state that uncertainty has a huge impact on the selection of the design concept and state that in order to achieve a robust and reliable design, uncertainties must be managed.

AI in product design is based on several types of algorithms, but case-based reasoning, genetic algorithms, simulated annealing, ant colony optimization, decision tree, association rule mining, Bayesian network, and fuzzy set theory are some of the common types (Lee, 2021). Research by J. Zhu and Deshmukh (2003) looking specifically into Bayesian decision networks found that they are particularly useful in decision-making influenced by high uncertainty, by providing a normative framework for presenting and reasoning about problems related to decisions.

Typically, a Bayesian network structure is constructed using two types of data, either large amounts of historical data or knowledge from experts/engineers and have the potential to support decision-making. It has for instance, being used in the context of construction projects to identify risks, make better decisions, and minimize the probability of failure (Pan & Zhang, 2021).

An article by G.-N. Zhu et al. (2020) looked into fuzzy logic and proposed a fuzzy rough number-based AHP-TOPSIS for design concept evaluation in uncertain environment. Although, the article is not specifically looking into AI, the work is still interesting as this fuzzy logic is one of the more common algorithms in AI in general (Lee, 2021). The authors acknowledge the importance of early stages in product development and how this process tend to be influenced by subjectivity, lack of knowledge and uncertainty. Often times, teams need to rely on subjective opinions from experts, that tend to be imprecise and uncertain, and they propose a framework to handle this:

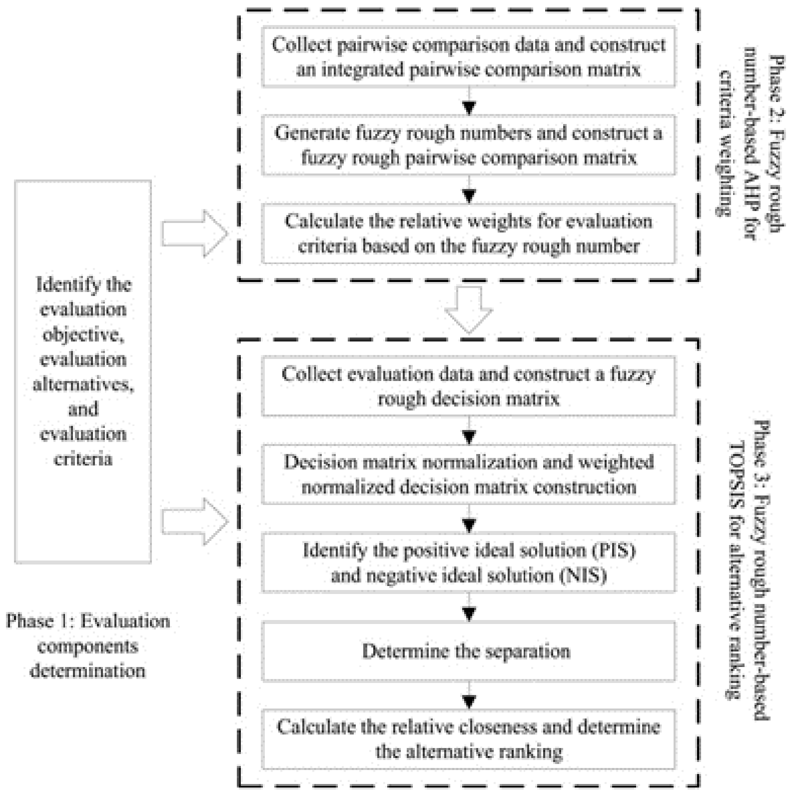

Figure 4.8: Suggested framework (G.-N. Zhu et al., 2020)

The study by G.-N. Zhu et al. (2020) shows large improvements in fuzzy rough number-based methodology in concept evaluation and decision-making, through its ability to deal with uncertainty and objectivity.

Risk mitigation has been found particularly beneficial in construction engineering management according to Pan and Zhang (2021). AI can assist in risk mitigation when uncertainty is high, by identifying risks, assessing risks and prioritizing risks to improve safety, quality, efficiency and cost across teams and work areas. AI-driven risk analysis offers predictive insights on critical issues, empowering project managers to swiftly prioritize risks and proactively mitigate them, rather than merely reacting to them. AI is capable of dealing with complex problems with high uncertainty and return tactical decision-making in this dynamic environment. Behind these capabilities lie various AI techniques such as probabilistic models, fuzzy logic, machine learning, neural networks, employed to analyze data from construction sites (Pan & Zhang, 2021).

4.3.1.4 Decision support system (Human and AI collaboration)

An article by Lee (2021) reviewed studies in AI in product design, and found that one of the main AI is incorporated into the process is through acting as a decision-support system. A decision support system, in the context of the study, denotes an information system facilitating design decision-making processes. They found that, the final decision rests with the designer, while the system provides a set of recommendations to assist in narrowing down the search space for solutions and identifying the most optimal solution (Lee, 2021).

This relates to intelligent optimization which entails the pursuit of the most favorable solution, aiming to either minimize or maximize an objective function within a defined set of constraints (Pan & Zhang, 2021). Optimization problems can either be “single objective optimization” or “multi objective optimization”. The first is focused on finding a single optimal solution, whereas the latter is focused on optimizing several functions simultaneously. In engineering practice, identifying the most optimal solution amidst high complexity, inter-dependency, and non-linearity can be arduous and time-consuming. Over the last few decades, meta-heuristic optimization methods have indicated to a promising alternative to streamline this process. Methods like these can identify near-optimal solutions that still are acceptable and reduce the time taken (Pan & Zhang, 2021).

An article by Saadi and Yang (2023a) visualized and tested how designers and AI can collaborate in early design process. An AI were used to generate concepts, from which humans could select. The selection process consisted of “evaluation” and “selection” and could be described in the following:

- Step 1: Visual examination based on knowledge and experience.
- Step 2: Compare performance by graphing concepts or through finite element analysis.
- Step 3: Create prototypes of results to get a first-hand feel.
- Step 4: Iterative approach. Based on first evaluation, try new parameters and constraints.
- Step 5: Finally, manually choosing of the final concept, based on the final AI-generated ones. In this stage, optimized performance was not the only factor influencing the selection, but experience and knowledge from designers influenced the final decision. For example, a designer might go for a solution with lower performance, in order to improve other factors such as manufacturability.

Saadi and Yang (2023a) found that designers may not achieve comprehensive understanding of how the generative design tool generated the solutions and consequently, a level of trust is placed in both the tool and the designers configuration of the design problem, enabling designers to accept the resulting designs.

An article by Rajagopal et al. (2022) states that there remain scenarios where human judgment remains essential, particularly when decisions hinge on factors beyond structured data analysis. These are decisions related to things like: long term goals, business strategy, organizational ideals, and competitive dynamics. This type of knowledge is not available to AI, as it only exists as thoughts. As an example, while AI could accurately pinpoint the most efficient inventory levels to maximize revenues, in a competitive environment, a company might opt to maintain higher inventory levels to enhance customer satisfaction, even if it means sacrificing short-term profits (Rajagopal et al., 2022). An ideal scenario describing the merge between AI and human decision-making can be seen in figure 4.9 below:

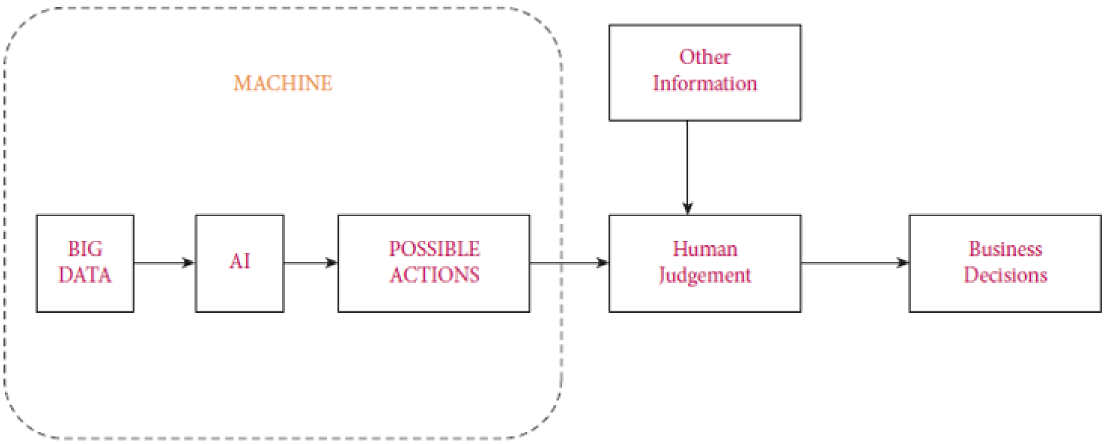

Figure 4.9: Proposed scenario merging AI and human activity (Rajagopal et al., 2022)

Figure 4.9 illustrates a proposed scenario where human- and AI-supported decision-making is combined. The solution is centered around AI supporting in generating several possibilities for humans to choose from. Human decision-makers often opt for choices from a limited set of possibilities based on their existing knowledge base, prioritizing expediency over optimization. The authors argue that better decisions are made when combining the two, rather than relying solely on one (Rajagopal et al., 2022). This goes with what Dwivedi et al. (2021) brings up in an article, namely that human workers play crucial roles in either evaluating and validating AI decision recommendations, implementing the recommended actions suggested by AI, or offering supplementary assistance in the event of errors or failures in AI-enabled automation. Consequently, comprehending the dynamics of human-AI collaboration is essential for realizing the expected benefits of automation (Dwivedi et al., 2021).

Other literature has commented on this phenomenon and mentioned that even though, experiments show that decision-makers make more correct decisions with support from AI, human experience in combination with AI is still vital. The key to humans making accurate decisions with the assistance of AI algorithms lies in the combination of decision-making comprehension and experience in the people (Janssen et al., 2022). While not looking specifically at decision-making, the results by Song, Zurita, et al. (2020) indicate that while AI can be a powerful tool to support the design process, as of right now the best scenario is when humans and AI collaborate together.

Companies stand a lot to gain from focusing on optimizing particularly collaboration between AI and and develop so called fusion skills to optimize human-AI collaboration, according to Wilson and Daugherty (2018). Furthermore, they claim that companies trying to displace employees with AI, will likely experience only short-term gains in productivity (Wilson & Daugherty, 2018). In an interview (Purdy & Williams, 2023) with Harvard Business Review, Matt Johnson, senior scientist at The Institute for Human & Machine Cognition had the following to say about AI in decision-making:

> “If used properly, generative AI could function as a really good teammate, in the same way that I might want to talk through a problem with my colleagues even though I think I already have the solution. It also potentially has a long organizational memory, which is useful for people who may be relatively new to an organization and want to find out how issues were handled previously.” – Johnsson (2023)

While this statement by Johnsson (2023) puts AI in a general decision-making context it could also showcase accuracy and validity that can be expected by an AI bot could should be equalized with an informed colleague. For example, a colleague could potentially provide an interesting input, but by not being sufficiently into the context they could also not be held accountable for what they say. Ultimately, the final decision remains with the designer.

However, what seems to be the closes step towards combining the three concepts of: Human, AI and design engineering is brought up by Demirel et al. (2024). The authors bring up several advantages of AI to perform generative design, things such as efficient evaluation, optimization and enabling expansive design space exploration. However, according to the authors Demirel et al. (2024), current generative design tools do not incorporate human factors, such as product appeal, comfort, and ease of use into the process and thus, limiting early design process. As a response a “Human-centered generative framework”, framed around design thinking enhanced by AI, is proposed:

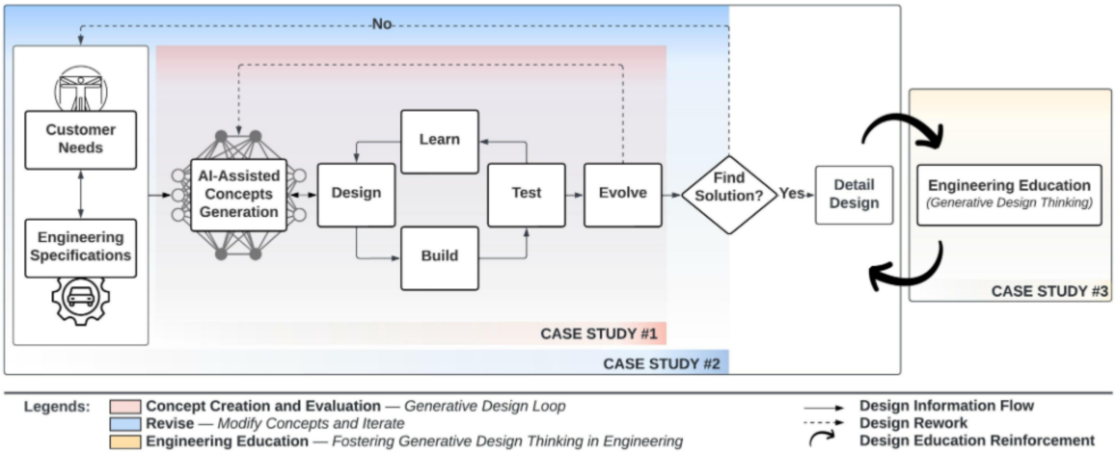

Figure 4.10: Framework for human-centered design framework, utilizing AI (Demirel et al., 2024)

The framework in figure 4.10 is based on a design-build-test-learn loop:

1. Design: Analog and digital process planning and modelling
2. Build: Prototyping in physical, virtual, and mixed environments
3. Test: Optimization through experimentation and simulation
4. Learn: Predicting better design solutions utilizing ML

With this approach, designers are able to capture data for example thanks to IoT, generate concepts with AI, model variants and evaluate assumptions through multi-physics simulation. With this solution, parameters from engineering and human behavior can be incorporated into decision-making of early design, giving engineers the chance to learn from variants of concepts (Demirel et al., 2024).

4.3.1.5 Knowledge support

Developments in numerical simulation and computational methodologies have enabled the generation, collection, and analysis of vast amounts of data, thereby enhancing designers' understanding of the underlying physics of problems. Nonetheless, data on its own is of limited utility unless it is organized and presented in a manner that facilitates actionable insights for designers. Additionally, proper indexing, storage, and management of data are crucial to ensure its availability in the appropriate place, at the right time, and in a suitable format for designers' use (Mavris & Pinon, 2012).

An article by Liao et al. (2020) wanted to investigate AI in early design and development and different roles AI can take in this process. Related to decision-making and evaluation, they found that:

- 1) AI can be used to draw conclusions, based on several conditions. By exposing an AI to design-history-data, AI can establish correlations among past concepts, thereby aiding designers in deciding on potential new ideas for implementation (Liao et al., 2020).

- 2) AI can play a crucial role in connecting designers with existing knowledge and information. By assisting designers in structuring search prompts they can uncover unexpected data, ultimately fostering the development of more innovative design solutions (Liao et al., 2020).
- 3) AI can act as catalysts for design actions, prompting designers to reframe their approaches. Design-problems are rarely well-defined, and solutions tend to co-evolve with the problem itself. Therefore, designers must be able to re-define design-problems continuously. For instance, AI may anticipate when designers encounter obstacles and subsequently offer targeted instructions or pose questions to assist them in overcoming these hurdles. In a broader sense, AI could be integrated into interactive systems that gather data on designers' past and ongoing activities, recommending actions tailored to the current design context (Liao et al., 2020).

5
Discussion

This chapter discusses the results from the interview study and literature review. Each research question is addressed in its own section, covering the answers, implications for designers and the design process, our interpretations and observations, and potential sources of error and their impact on the results.

5.1 RQ1: What are the main challenges when performing concept generation and evaluation?

5.1.1 PD-interviews

Looking at the initial problem, catalyzing this entire study, concept development is a complex process with many different factors to consider. After conducing this interview study, in order to identify and document challenged related to concept development, this initial observation was further strengthened. In addressing the research question, ”What are the main challenges when performing concept generation and evaluation?” the outcomes of the interviews reveal a comprehensive array of obstacles categorized into eight primary areas: data secrecy, time, organizational synergy, information, human factor, balancing demands, conservative culture, and the PD-process. These challenges reflect multifaceted issues that impact both the ideation and assessment phases of concept development.

For specific and detailed answers to RQ1, go to 4.1.1 where all challenges identified challenges are outlined. This chapter will be more focused on interpreting the result in a more general way and comment on what this says about concept development.

One major observation from the interview study, is that there is a lot of information within the organization and this speaks positively for GKN. However, finding necessary information seems to be one of the biggest challenges as almost every interviewee stated that information is hard to find. Lessons learned are documented but as they are difficult to find or even locked away from employees not part of a certain program, documenting this information in the first place can be demotivating. This is not to say that lessons learned are not used, but it seems that the preferred way for design teams to acquire new information, is by involving people with first hand experience from previous projects. Enhancing team diversity and implementing objective evaluation frameworks can mitigate biases, fostering more innovative and effective concept development.

Furthermore, due to restrictions from data secrecy, this can further inhibit the ability to efficiently share and absorb information. These restrictions create significant hurdles in leveraging collective organizational knowledge and inhibit cross-functional collaboration, critical for innovative concept generation and robust evaluation. Reflecting upon the idea of training an LLM on internal company data, can these restrictions potentially stand in the way of successfully training a model in-house? Furthermore, even though there is a lot of information within the organization, certain knowledge exists only in the form of experience and remains undocumented. What does this mean when looking to train an LLM? Not only might this result in even more restrictions when is comes to data availability, this could also potentially introduce human bias into the training data.

As so much information is bound in the form of knowledge and experience of individuals, this can explain why human factors plays such a major role in concept development. Due to this, the arrangement of teams seems to have a large impact on the final outcome, as human subjectivity and bias is introduced to design work. Furthermore, individuals have a significant responsibility to remember certain scenarios and information correctly. This subsequently has an affect of how different demands are balanced, a process that seems to be influenced by high complexity and uncertainty where many different factors need to be taken into account. Finding a solution that satisfies all types of criteria seems extremely complex, as there sometimes is an inverse relationship between certain factors. This meaning that an increase in one variable ultimately results in a decrease in another.

The way demands are balanced is also heavily influenced by the current product development process, where complexity and uncertainty also plays a large role. There are many different opinions on what a good solution is and there is a challenge to find a balance between evaluating many concepts, and also deciding on what the best one is. As the time-factor is so significant, finding the best concept with as little unknown risk as possible, is crucial. Time constraints pressure the development process, often forcing premature decisions or inadequate evaluations, compromising the thorough exploration and validation of concepts. Implementing flexible and adaptive processes, supported by integrated software solutions and proactive risk management, can possibly enhance the efficiency and effectiveness of concept generation and evaluation.

Overall, there is a conservative culture within the organization and probably affects more than just ”PD-process” and ”Human factor” as is currently indicated in figure 4.1. Encouraging a culture that supports calculated risk-taking and open-mindedness towards new ideas can potentially drive more dynamic and forward-thinking concept generation.

The relationship between different challenges seen in figure 4.1 is the result of analysis and processing of the responses given by the interviewees. In order to process the information from the interviews not only simplification had to be made but also interpretation of the responses. The raw data from the interview study for PD-interviews resulted in 86 pages of transcriptions. When selecting key sentences and phrases out of the transcribed material, and storing them in an Excel matrix, the context is removed. While the context is important to capture nuances and the exact meaning of a phrase, this kind simplification had to be made in order to make sense of the information. During this process, a lot is left to our own judgement to determine whether or not a certain sentence is relevant or not.

Trying to group and create sense out of such a vast amount of information is also complex and also requires own interpretation to grasp such a complexity. While some statements are easy to group as they show obvious similarities, relationships between other might not be as clear. This ultimately implies that this kind of grouping is based on our selection of key statements and interpretation of the relationship between these, and might differ if performed by someone else. To put it in another way, this is just one way of organizing, but in fact there are numerous ways the different challenges could be categorized. If the diagram were to be 100% true to reality, there should probably be an arrow between every single challenge.

Furthermore, some respondents provided a lot of relevant information resulting in us selecting many of their statements, and fewer statements from others. This could ultimately slightly angle the outcome towards the views of a few interviewees, even though there were many interviewees a part of the study.

Additionally, frequency of mentioned challenges was considered, but not documented or presented in this study. This removes the ability to showcase whether a challenge is continuously being brought up by several respondents, or if a challenge is brought up once by a certain individual. By providing frequency, we could have been able to more efficiently identify and communicate, not only the challenge itself, but also where the most severe challenges are found. For example, almost every interviewee mentioned that finding relevant information is difficult, but this cannot be deciphered in figure 4.1 and thus not showing the full extent of this challenge.

The answers gotten from the interviewees were certainly heavily influenced on what they were able to come up with on the spot. There is always a risk that they might remember situations incorrectly or not being able to come up with information at all. This is why thorough question preparation is crucial as the answered one gets are only as good as the questions asked. Providing the questions to the interviewees beforehand might partially aid this possible factor.

Initially in the interview process, answers were scattered and each new interview provided new interesting takes on concept development. However, what was interesting was that after having conducted several interviews, answered was starting to repeat and were starting to converge towards a certain direction. This indicated that enough interviewees had been involved in the study.

5.1.2 Decision-making interviews

Something that became apparent in the PD interviews was the conflict regarding who or where people think decisions in the PDP are made. Therefore, it became necessary to conduct two additional interviews with the Senior Vice President of GKN to truly understand who makes these decisions, and more importantly, what considerations are taken into account. It was highlighted that GKN follows a strict decision-making process where all sub-categories of a decision are classified as either ”satisfactory” or ”more information is needed.” When all sub-categories are deemed satisfactory, a project can move forward. In the context of our interviews, it was the Head of Industrial Architects who was one of the key decision-makers in determining whether a new project would proceed to manufacturing. It also became clear that designers themselves do not make decisions on whether a concept is ready to move forward; instead, they report their work, and then a review group evaluates the work and decides if everything is in place or if the concept is mature enough for the next phase in the project.

5.2 RQ2: What are the gaps that are needed to be filled for a successful implementation of AI/ML methods to improve robust conceptual design work at GKN | RQ3: What is GKN's current AI/ML capabilities?

5.2.1 AI/ML interviews

To successfully implement AI/ML methods and improve conceptual design work at GKN, several key gaps must be addressed. The second round of AI/ML interviews highlighted limitations in cultural/organizational aspects, data quality, infrastructure, knowledge/skills, technology and legal compliance/data secrecy.

A significant cultural and organizational gap exists. At the top-down level, there is reluctance and skepticism about adopting AI/ML until its value is proven, preventing rushed strategies. Within the organization, employees' awareness and interest in AI/ML vary, resulting in fragmented exploration rather than a unified approach. Resistance to moving from traditional methods to AI solutions is prevalent, and engineers often lack AI/ML expertise. Quality data and clear problem understanding are essential for AI effectiveness, but these are often missing.

Data quality is another major challenge. Issues with data management have shown that data is not stored effectively for reuse, analysis, or training. GKN needs to focus on data quality, governance, and literacy to ensure high-quality input data for successful AI outcomes. Currently, much of the data is unsorted, hindering efficient use of big data for automation. Robust data management practices are crucial.

Infrastructure limitations also pose a bottleneck. Limited internet bandwidth slows data downloading and analysis, and the high cost of moving large datasets to and from the cloud is a concern. Current hardware is unsuitable for scaling AI/ML operations. Significant upgrades and investments in IT infrastructure are necessary to fully leverage AI capabilities.

Knowledge and skill gaps are another hurdle. While some individuals have advanced AI skills, broader organizational understanding is limited. Many employees view AI as a buzzword and lack practical knowledge. Expanding AI/ML knowledge through training is essential to foster informed perspectives and effective use.

Legal compliance and data secrecy challenges are significant. Data usage restrictions, especially for sensitive information, require in-house model training and extensive infrastructure. Export-controlled data, particularly military data, cannot be processed externally. Opinions vary on outsourcing model training, but anonymizing data could allow external training without compromising security. Compliance with legal guidelines is crucial.

Technology integration presents another gap. There is potential for using advanced AI to generate CAD models from text or spoken inputs, transforming the design process. However, integrating AI into established design guidelines and training AI to align with company-specific thinking is complex. LLMs can handle subjective design elements, but practical integration into workflows is needed.

Addressing these gaps is crucial for GKN to successfully implement AI/ML methods and improve conceptual design work. This requires improving data quality, upgrading infrastructure, expanding AI/ML knowledge, ensuring legal compliance, and integrating advanced AI technologies into the design process. By tackling these issues, GKN can enhance its AI/ML capabilities and achieve more efficient and innovative design processes.

GKN's current capabilities
GKN is actively utilizing AI/ML technologies across various domains, showcasing its commitment to advancing capabilities in this field. Efforts are focused on understanding and integrating AI/ML into both production and enterprise processes.

Current AI/ML applications primarily consist of commercial tools such as CAD programs, Office and GitHub Copilots, and OptiSlang, which is used for exploring design spaces and optimizing designs using ML. AI and ML are used more in the design process than employees might realize, given the significant leap in integrating AI/ML applications in commercial tools used at GKN, such as NX. Furthermore, GKN's research projects collect images, timestamps, and other process data, integrating this with the main data logging system to boost operational efficiency. AI/ML technologies automate production processes and implement image recognition on production lines, enhancing productivity and reducing errors. Additionally, generative AI is being explored to improve data quality in HSE reports, demonstrating GKN's initiative towards new AI methods, and more importantly, LLMs.

GKN has considered using public datasets for on-premise model training to maintain data security while enhancing AI capabilities. Discussions are ongoing about outsourcing model training with anonymized data to ensure confidentiality while advancing AI/ML technologies, which is necessary since the resources for training big models are not available on-site.

While some individuals at GKN possess advanced AI skills, broader organizational adoption is still developing. Efforts are being made to automate data usage more effectively and reduce manual interventions. AI integration in regular processes, such as report generation, is in the early stages, with technologies like OpenAI being explored for broader application.

Overall, GKN demonstrates engagement with AI/ML technologies through projects at different levels of advancement and research initiatives. Similar to the companies that took part in the benchmarking study, GKN also shows interest in AI implementation. However, the tone from the interviews suggests that other companies might have more capabilities for faster integration than GKN.

Their current capabilities include ML applications in design platforms, document and code writing assistance, design optimization, and production process automation. They are proactive in addressing data privacy and security concerns through thoughtful strategies for model training and data usage. Generally, the adoption of AI methods is cautious and in the early stages. One of the main challenges is the lack of broad, enterprise-wide knowledge in this area, as currently, only certain individuals possess advanced expertise, leaving the majority of employees behind. This makes a large-scale adoption of, for example, fine-tuned LLMs unlikely in the near term.

Drawbacks of the AI/ML interviews

After the AI/ML interviews were conducted and additional literature was reviewed, we concluded that having more background knowledge in the AI/ML area prior to the interviews would have enabled us to ask more specific questions. However, since the interviews were semi-structured, we still gathered much useful information. We view this as a minor flaw within the broader context of the study.

Given that the focus of this thesis is to identify AI/ML methods and tools that can aid the early stages of concept development, the number of AI/ML interviews conducted may have been insufficient. In retrospect, it would have been more logical to conduct an equal number of PD and AI/ML interviews. As these two groups combined totaled 16 interviews, having eight PD interviews and eight AI/ML interviews would have been preferable. One reason for this is the greater variability in responses from the AI/ML interviews compared to the PD interviews. While the responses from the 11 PD interviews were quite similar, the AI/ML interviews varied significantly from one another. We conclude that having more AI/ML interviews would have enhanced the quality of this thesis. We do not believe that the knowledge existing at GKN is adequately represented by only these five interviews. However, since one of the directors participated in the study, we were able to gain insight into their plans, intentions, and attitude towards AI/ML implementation, which is a crucial aspect of the study.

Only in one of the interviews did a respondent possess knowledge of both the design process and AI/ML methods, and how these could be beneficial. In contrast, many others primarily focused on AI/ML applications that lie outside of the design process. By increasing the number of AI/ML interviews, there was a potential to gather more useful information in this specific area, which is the primary topic of this thesis.

Generally, the interview study became too extensive, reducing time spent on literature study which may have affected the results. The scope of the study was to wide, resulting in a less thorough investigation of each specific aspect. We would recommend a more directed study next round, focusing on a very specific issue in the concept development at GKN, not the whole process at once.

5.3 Benchmarking interviews

The purpose of the benchmarking study was mainly divided into three parts: 1) Gain access to current state of the art of AI implementation in industry, and compare this to where GKN currently stands, 2) Take part of other organizations motivations for implementing AI, 3) Take part of lessons learned from companies being successful/unsuccessful implementing AI.

The three studied organizations were one GKN Fokker, SKF and C3 being a large life science company in a regulated industry. Even though, none of the organizations is directly comparable to GKN, all organizations worked with product development where product safety, customer safety and information safety were a very high priority.

On one hand for the purpose of this study, an interesting observation was that neither GKN Fokker, SKF use AI support for concept generation and evaluation, hinting that this type of process still is influenced by manual work in large organizations as of right now. On the other hand, C3 gave indications that concept generation and evaluation was influenced by AI applications, but unfortunately this is where that discussion ended. This might suggest that this type of technology is still so new and rare that it is strategically beneficial to keep this information withing the organization. However, this is nothing we could confirm and thus making it difficult to benchmark concept development against C3.

The interview with R20 at SKF was particularly interesting because he did not only have large insight into AI implementation at SKF, but also into several other large engineering organizations and thus was able to make comparisons. He found that SKF is at the forefront of AI implementation, for one main reason: having managerial support and push to make sure that the organization realize that AI is important. However, this benchmarking study was still limited to the firsthand information from just three organizations, and just gives a glimpse into AI-implementation in industry as a whole.

In general, the findings suggest that it's clear that the three organizations studied in this benchmarking study, still are in the early phase of AI-implementation. It is mostly done in small scale, for specific applications and still very much done as trial and error, but early improvements can already be seen in specific areas. Furthermore, companies are still in the experimental phase and continuously discovering new areas where AI can be implemented. Even though each company have face their own specific challenges and do their own take on AI implementation, the overall goal seem to be quite consistent: increase productivity, be more efficient and make optimization easier. What is clear is that all three organizations are well aware of the echoes of AI and are looking into how they are able to introduce AI to suit their particular needs.

5.4 RQ4: How can tools from AI and ML can be used to simplify concept generation, evaluation and to propose a robust design solution?

This section aims to discuss the answers of RQ4.

5.4.1 Implications of the result and how it relates to literature, and interviews

This section aims to adress the implications of the result and how it relates to the literature found.

5.4.1.1 LLMs in Concept Generation

LLMs, such as GPT-4, could revolutionize the engineering field by simplifying concept generation, evaluation, and the development of robust design solutions. Traditionally, engineering design has relied heavily on human expertise, involving iterative dialogues filled with calculations and simulations. However, the integration of LLMs and other digital tools is streamlining these interactions, automating complex processes, and enhancing the reasoning and argumentative processes that are intrinsic to design (Göpfert et al., 2023).

LLMs and multi-modal models are adept at processing a wide range of data types, including natural language, text, tables, sketches, and 3D models. This capability makes them exceptionally good at integrating into the goal-driven dialogues essential to design. Through their extensive training on diverse datasets, these models perform complex reasoning tasks and support decision-making processes, transforming how concepts are generated and evaluated (Song, Zhou, & Ahmed, 2023; J. Wei et al., 2023).

LLMs are at the forefront of advancements in NLP as stated, and have expanded their utility to support creative reasoning tasks that are crucial in engineering, particularly when integrated with computational techniques like topology optimization (Göpfert et al., 2023; A. Wang et al., 2019). Furthermore, AI technologies such as GANs and reinforcement learning systems demonstrate the substantial role AI plays in enhancing design strategies and promoting a deeper understanding of the design process through advanced learning mechanisms (Gyory et al., 2021, 2022; Raina et al., 2021; Regenwetter et al., 2022).

The versatility of LLMs extends to their ability to process and organize vast amounts of design documentation and decompose complex design tasks into more manageable, functional formats. This capability facilitates broader and more effective ideation across various aspects of engineering projects (Qiu & Jin, 2023; B. Wang et al., 2023). Furthermore, LLMs demonstrate remarkable utility in extracting and integrating user needs from product reviews, allowing for a user-centered approach in design that leverages user-generated content effectively (Han et al., 2023).

LLMs also found a role in fields like microfluidics and robotics, where they assist in both ideation and problem-solving. This reflects the adaptable and extensive applications of LLMs across different engineering disciplines (A. Li et al., 2023; Nelson et al., 2023; Stella et al., 2023). Their integration into design heuristics is not only innovative but potentially transformative, offering new paradigms for addressing modern engineering challenges and enhancing the innovation process (Yilmaz et al., 2016). The abilities of LLMs extend to generating design concepts, analyzing sketches, selecting materials, and creating CAD drawings, highlighting their potential to fundamentally transform engineering design (Makatura et al., 2023; Picard et al., 2023; Q. Zhu & Luo, 2023).

The general idea synthesized from the literature is that LLMs act as assistants at every stage, helping designers become more efficient in their work. They spend less time on tasks that do not contribute to the robustness of the concept or the product itself. By providing designers with information quickly, aiding in ideation and creative reasoning, assisting in topology optimization, structuring documents, and even generating simple emails, LLMs allow designers to focus on tasks that require human expertise and evaluation—those intrinsic to the concept generation and design process.

All these factors combined suggest several ways that LLMs can help simplify and streamline design processes, offloading tedious tasks from designers. This increases their cognitive capacity as they can focus on fewer areas, leading to higher concentration and productivity. This human-computer interaction would logically result in more knowledgeable human designers, who are more capable of generating robust concept solutions.

LLMs in archivable design dialogues: The potential to digitize design dialogues into a manageable, archivable format adds another layer of efficiency. By documenting and archiving every step of the design process, including the reasoning and decision-making phases, past designs can be revisited and utilized efficiently for new product developments, ensuring that design decisions are traceable and that the collective reasoning process is enhanced, fostering better collaboration and innovation (Göpfert et al., 2023).

The idea of using traceable past decisions and archiving design dialogues to think like a GKN designer also holds great potential. However, it's important to consider that humans have biases toward different solutions based on their training, academic background, and previous work experience. Training an LLM on designer-specific data and decisions could incorporate these biases. This can be both beneficial and detrimental, as GKN must sometimes prioritize certain biases due to external stakeholder pressures, such as demands for reduced weight that might increase costs.

Another aspect to consider is the selection of design dialogues. Should all dialogues be included, even those involving less experienced designers, or should the focus be on those between experienced designers? This could be a concern, but testing on a variety of results would include the most diverse selection of biases, which could make decisions and workflows more objective.

LLMs in aerospace applications: An example of LLMs in action can be observed in a study at the German Aerospace Center (DLR), where an AI-powered chatbot serves as the core of an intelligent workflow engine integrated within the GTlab software framework (Reitenbach et al., 2024). The study's positive results demonstrate the potential of incorporating LLMs in high-complexity situations, particularly in aerospace applications. For GKN, this serves as a proof of concept, highlighting that the implementation of LLMs in an aerospace workflow can be successful. The system's ability to perform detailed thermodynamic calculations, such as predicting compressor outlet conditions from given parameters, showcases its potential to streamline technical workflows and enhance usability in critical engineering applications.

For designers wanting to quickly adjust a part by increasing or decreasing a parameter, this could be very beneficial. The workflow engine could provide instant feedback to queries without requiring the designer to wait for analysis team feedback. This does not eliminate the need for analysis teams, instead, it offers immediate feedback on ideas that can later be implemented with fully performed and verified calculations.

Generative design tools: which are integrated with AI, are redefining traditional tasks within the engineering sector. These tools enhance the efficiency and creativity of design processes by fundamentally altering how designers interact with both the conceptual and optimization phases of projects (Kazi et al., 2017; Krish, 2011; Saadi & Yang, 2023b; Zhang et al., 2021). While these innovative approaches often require designers to rethink their strategies for setting up design spaces and aligning with design briefs, the benefits include more dynamic and responsive design outcomes (Alcaide-Marzal et al., 2020; Lopez et al., 2018; Vlah et al., 2020). The digitization of engineering design has also advanced significantly, supported by modern tools that enhance every stage of the PDP through improved visualization and interaction capabilities (Göpfert et al., 2023). In particular, generative AI design tools has revolutionized the design of 3D structures, making the mechanical design process more versatile (Jadhav & Farimani, 2024; C. Li et al., 2022; Nichol et al., 2022; Sanghi et al., 2022). Furthermore, generative AI design tools help designers quickly determine if a simple idea could work or not. The instant feedback, combined with human intuition, allows for easy identification of unworkable ideas. Additionally, commercial generative design tools are rapidly integrating new techniques into their software. These tools could further assist designers in the future, streamlining much of the design process. This means GKN might not need to implement their own AI solutions in workflows but could instead purchase commercial products that offer the same capabilities.

The adaptability of these models to different query structures and technical terminologies is crucial, especially in applications like aircraft engine simulations, underlining the profound impact LLMs have in simplifying concept generation, enhancing evaluation processes, and proposing robust design solutions in engineering and aerospace applications.

5.4.1.2 Robust design and Zero Defects with AI

(Leberruyer et al., 2023) highlights several different topics of integration of AI into the ZD subject. AI in general can be used for data-driven techniques for automated data analysis and decision making, such as Failure Mode and Effects Analysis (FMEA) and in detail (Process-FMEA). Approach for identifying possible failures in a design, a manufacturing or assembly process, or a product or service.

Robust design: Throughout increased reasoning, creativity, and idea generation, there is a potential to predicting and generating more attractive attributes that satisfies or surpasses expectations. Predicting delighters is particularly challenging as they often arise from unspoken needs or latent desires. In terms of robust design, delighters could be an optimized solution generating a new unseen concept. We argue that, following the Set-Based way of PD, that a increased knowledge space would also increase the design space, and in extension help propose a robust design solution.

AI/ML can facilitate SBD by quickly generating a wide range of design options based on predefined constraints and objectives. Machine learning algorithms can evaluate these options against multiple criteria, allowing designers to eliminate infeasible or suboptimal solutions early in the process. This enhances the efficiency of the concept development phase by focusing resources on the most promising design paths.

Design for Manufacturing ensures that a product is easy and cost-effective to produce. AI/ML can enhance DfM by analyzing manufacturing constraints and capabilities early in the design process. Predictive models can evaluate the manufacturability of design alternatives, identify potential production issues, and suggest design adjustments to simplify the manufacturing process. This integration reduces the likelihood of defects and rework, contributing to zero-defect manufacturing goals.

Failure Modes and Effects Analysis (FMEA) and its more detailed variant Failure Mode, Effects and Criticality Analysis (FMECA) is a systematic method for identifying potential failure modes and their effects on the system. AI/ML can automate and enhance FMEA by analyzing large datasets to predict possible failure modes and their impacts. NLP can process historical maintenance logs, incident reports, and other textual data to uncover patterns of failure. Machine learning algorithms can then prioritize these failure modes based on their severity, occurrence, and detect ability, providing valuable insights for mitigating risks early in the design process.

5.4.1.3 Concept evaluation

When studying how AI can assist in concept evaluation and decision-making, AI capabilities found were found to be ”Human bias mitigation”, ”Automated decision-making”, ”Risk and uncertainty management”, ”Decision support system” and ”Knowledge support”

Based on the studied literature, it’s clear that human bias and subjectivity plays a large role in product development, and particularly when it comes to early concept evaluation, where several articles state that AI can support in managing this bias. Based on the amount and frequency of this stated phenomenon, human bias and subjectivity mitigation seems to be the main advantage and capability of incorporating AI in decision-making and evaluation.

There are a lot of similarities between a System 1 type of thinking, influenced by habit and intuition Kahneman (2011) and decision-making based on subjective and biased elements found in early concept development. Kahneman (2011) state that for important decisions, basing decisions the System 1 way is not recommended. This might suggest that AI has the capability to turn the current way of evaluating concepts, from a system 1 type of decision-making, into system 2 type of decision-making:

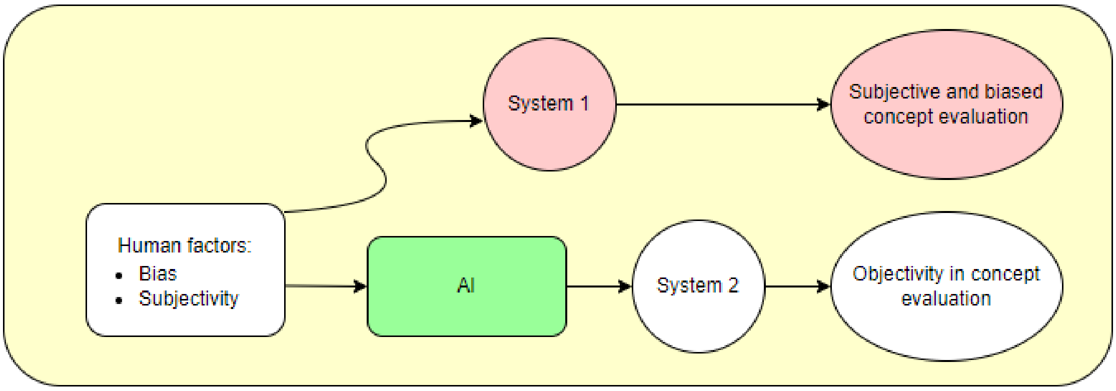

Figure 5.1: Visualization of how AI can trigger System 2 type of decision-making, resulting in objective concept evaluation

Figure 5.1 illustrates a proposed scenario in how AI can come to play a role in design work, paving the way for more proper decision-making.

An interesting take on AI and decision-making is that by Verganti et al. (2020) stating, as decision-making becomes an automated process, the purpose of the engineer shifts from problem solving, to problem finding. This, connected to what was stated by Saadi and Yang (2023a), where they claim that designers might not achieve the same understanding of how a generative design tool generates concepts. Even though AI can be considered a "black box" (Liao et al., 2020), there is also a level of new trust placed on the designer in structuring the problem properly (Saadi & Yang, 2023a). Consequently, design-engineering will become a lot more about giving proper instructions to and AI accepting or not accepting what it produces.

The literature study has indicated that this type of iterative work between human an AI in decision-making is what the solution is, even though a few articles (Verganti et al., 2020)(Camburn, He, et al., 2020) have stated that AI can turn decision-making into an automated process. The majority of studies of AI evaluation in design work has been indicating that human-AI collaboration is going to be the most favourable. Lee (2021) found that the role of AI in decision-making is in acting as a "decision support system" where the final design decision ultimately rests with the engineer, but AI assisting in narrowing down the search space. Saadi and Yang (2023a) illustrated a 5 step approach to incorporating human and AI capabilities, combining the strengths of AI and humans. Demirel et al. (2024) illustrated that human factors doesn't have to be negative in design work, but that factors like these risk being removed when performing generative design, and thus another framework is proposed.

There are still areas where human expertise is essential (Rajagopal et al., 2022), but also areas influenced by high complexity and uncertainty where AI can support as human capabilities are restricted (Lee, 2021). The idea is not necessary to remove human factors, but to aid human restrictions and assist where human capabilities are insufficient. Based on the studied literature, the following framework is proposed:

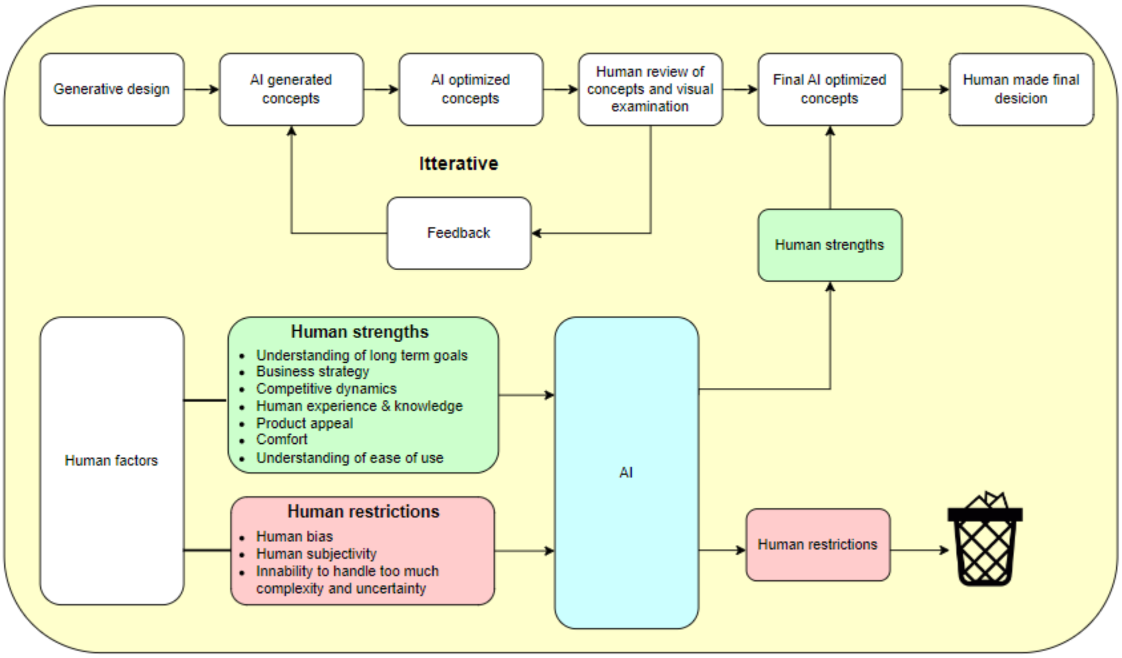

Figure 5.2: Visualization of how AI could support concept evaluation

The figure 5.2 is meant to illustrate, that even though some have found that AI actually make better decisions, it is not perfect, and therefore, humans still need to be in the process. Consequently, AI should be given the role of a decision co-pilot, where human an AI work together.

Lastly, AI appears to be able to support designers with knowledge (Liao et al., 2020). By utilizing data from past design-decisions, an AI is capable of creating recommendations in the current work process. It can also help designers find out about knowledge that otherwise might have been overlooked. By letting decision-makers have access to more easily accessible information, one could argue that this consequently will lead to more informed decision. Taking this a step further, perhaps the reason for the fact that human bias and subjectivity plays such a large role in concept evaluation, is the fact that information is hard to find, which makes it natural to use experience instead of more objective information.

Several articles states that AI can assist in decision-making and concept evaluation, yet few propose clear software and techniques that can be implemented (Khaleel et al., 2023)(Yüksel et al., 2023)(Allison et al., 2022). The ones that have proved that AI can improve decision-making, have used it in still relatively simple scenarios. There have been scenarios of evaluating concepts existing in natural text (Camburn, He, et al., 2020) and concept evaluation of footwear (Yuan et al., 2022). Furthermore, the literature studied have been looking at design process in general or at specific examples. However, non of the articles looks into concept evaluation in aerospace engineering. Articles, stating that AI can't be used for decision-making has not been found.
To summarize, there are four main capabilites of AI in design concept evaluation:

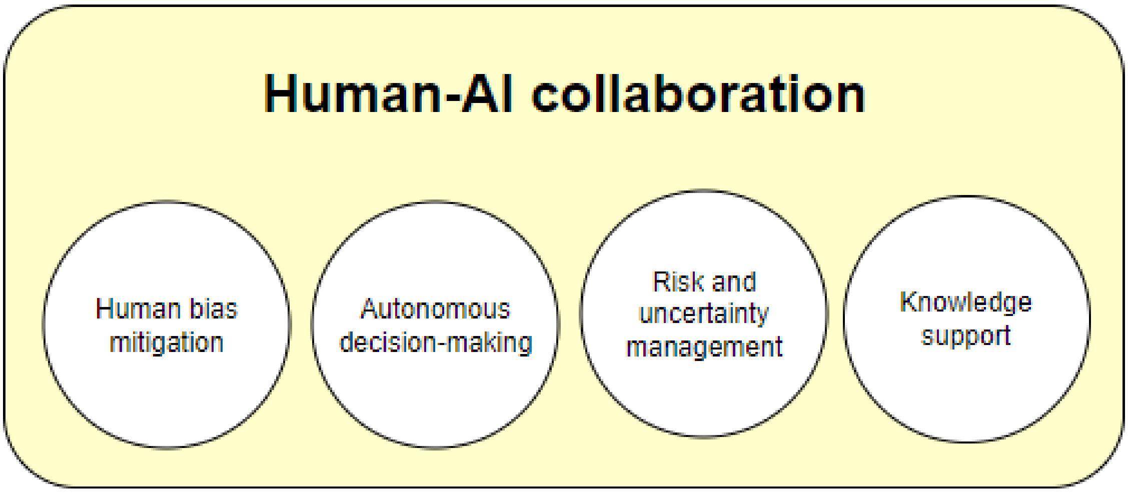

Figure 5.3: Visualization how the four capabilites exist in the contect of Human-AI collaboration

These four capabilities identified in the literature study are not to be seen as individual islands, but all work together, in an overall context of "Human-AI collaboration", as this will be most beneficial. For example, automated decision-making is possible but as it is not perfect, is still needs human support, by iteratively coming up with solutions. However, this is based on the literature found, based on searching with relevant keywords. Worth noting is that few articles published on the subject were newer than 2022 and as this is a fast moving discipline, a lot can change. It would have been beneficial to have found studies from 2022-2024 but unfortunately, no relevant literature published during these years were found.

5.4.2 What affected the results

This section aims to adress the sources of error that could have had an effect on the outcome of the study.

5.4.2.1 LLMs in Concept Generation

There is a lot of literature about AI/ML in the topic of design. CAD, MDO, DSE, workflow enhancements, patent identification with AI/ML support, and so forth. However, the area of AI/ML in aerospace is more limited, especially in the context of LLMs. Additionally, a lot of articles bring up opportunities of the technology, but very few actually describe a process of integration, or whats being used today in practice, and if that is the case, how it is actually used. The conception we get is that there is a lot of will to incorporate AI/ML into design processes and corporations, however the actual integration has not happened yet on an enterprise wide scale according to the literature study. This is a contradicting fact, since the benchmarking study has shown something different. Furthermore it is interesting how articles from just 3 years ago depicts a different AI environment as of now. It has been important to keep in mind the different release dates of commercial and open-to-use AI platforms such as OpenAI and their release of different versions,

because GPT-3 tests are almost obsolete compare to GPT-4 tests after the huge improvement in handling complex tasks after that update. Therefore, it must be said that, while those articles have important insights, that the testing and results cannot be taken into a count from articles published before March 14, 2023.
When currently writing the discussion part May 13 2024, OpenAI have released their newest version GPT-4o. GPT-4o have a lot of the multi-modal capabilities that been discussed in previous works. It has the capability to take in a live video feed from you camera, and have the ability to speak to you in real time, enabling speech-to-speech prompts. Meaning that it can reason across, audio, vision, and text in real-time. Again showing the rapid advancements in the field of AI and LLMs in particular.

5.5 RQ5: What AI based methods have a potential to improve robust conceptual design work at GKN and what are their limitations?

This section discusses how the results from the study answers RQ5 and then sources of error that may have affected the results.

5.5.1 Implications of the result and how it relates to RQ1, RQ2, RQ3

This question will be partially answered by the PD-process, AI/ML and Benchmark interviews, utilizing all respondents from Table 3.1, 3.2, 3.3, 3.4 as well as the literature study, combining information from both to gain a holistic and accurate view. Summaries providing an overarching perspective will be presented from the interviews, along with key citations. The text is structured based on the challenges identified in RQ1 results.

Based on the literature study on concept evaluation and decision-making, no studies were found where a with a clear suggestion of how concept evaluation could be incorporated at an organization like GKN. Several studies were optimistic towards AI in concept evaluation and decision-making and most studies stated that concept evaluation and decision-making were applications of AI. However, few actually provided frameworks and the ones that provided a tangible method for AI supported concept evaluation, still used it with simple concepts like sneaker design and concepts purely existing in text. No studies were found, where concept evaluation was proved to be performed successfully, in an industrial manufacturer like GKN.

However, taking more of a theoretical approach on how AI could support in concept evaluation and decision-making, there are still areas where AI could support in theory, even though there is a lack of a tangible framework ready to implementation. The literature still provided capabilities of AI, even though AI might not be ready to handle the complexity of products developed at GKN. With this in mind, the four capabilities found and illustrated in figure 5.3, are to be used and matched towards challenges outlined in 4.1.1.

5.5.2 Product Development Process

The integration of AI and ML methods in the product development process (PDP) offers significant potential to address the specific challenges identified at GKN. Below, I discuss how the technologies mentioned in the provided text can help tackle each challenge.

CP1: Managing Ambiguity and Diverse Perspectives
LLMs such as GPT-4 can play a pivotal role in managing ambiguity and synthesizing diverse perspectives during the early phases of development. Their capacity to process and generate natural language enables them to integrate and reason about diverse inputs and feedback from different stakeholders. As Göpfert et al. (2023) suggest, LLMs facilitate goal-driven dialogues that are essential in balancing design freedom with timely decision-making. These models can generate multiple design scenarios and solutions, providing a structured way to explore various design options and implications before finalizing decisions (Göpfert et al., 2023).
Furthermore, AutoTRIZ, an LLM-based tool, enhances creative problem-solving by systematically generating innovative solutions while considering diverse engineering principles and patent data. This tool reduces cognitive load and speeds up the decision-making process by automating the generation of detailed solution reports based on user inputs (S. Jiang & Luo, 2024).

CP2: Navigating Theoretical vs. Practical Discrepancies
GANs and reinforcement learning systems can help bridge the gap between theoretical models and practical realities. These AI techniques can simulate and evaluate numerous design iterations rapidly, learning from each iteration to converge on solutions that not only meet theoretical standards but also adhere to practical constraints (Regenwetter et al., 2022).

For instance, GANs could be used for the ideation and brainstorming of 3D models of new products under different manufacturing conditions, identifying potential issues before physical prototyping. This helps in understanding how theoretical designs might perform under real-world manufacturing conditions, thereby mitigating the risk of mistakes and ensuring that the designs are both innovative and manufacturable.

CP4: Complexity and Risk Management
The digitization of design dialogues and archiving of design processes, as enabled by LLMs, contribute significantly to complexity and risk management in project execution. By maintaining detailed records of the decision-making process, AI tools help in tracking the evolution of designs and understanding the rationale behind certain decisions (Göpfert et al., 2023). This documentation is crucial for anticipating and addressing potential deviations during manufacturing.

Moreover, the application of deep learning in the design of 3D structures allows for more versatile and accurate simulations of how designs will behave under various stressors and manufacturing processes (Jadhav & Farimani, 2024). These simulations help in identifying potential risks and complexities early in the design phase, thus allowing for adjustments before costly manufacturing commitments are made.

In terms of concept evaluation the product development process at GKN is characterized by high complexity, risk and ambiguity. Early stages are marked by uncertainty, requiring a balance between creative freedom and planning. Teams must evaluate all concepts while aligning with top management's decisions. Navigating dynamic processes can lead to skipped steps when processes feel cumbersome.

CP1 highlights that there is a challenge with ambiguity and decision-making in concept development. This relates to exploring as many ideas as possible while also coming to a decision and what idea to go forward with. This challenge relates to many of the other challenges [CH2][CB1] as it is brought forward that coming up with a final solutions is difficult. There are also several perspectives to what a good solution is and thus this makes it even more complex. However, it could be argued that there shouldn't be opinion on what a good solution is. In an ideal scenario, it should be possible to state that one solution is objectively the best. With AI, this has the potential to become reality. By deploying AI-human collaboration, AI can assist in providing automated decisions and iteratively develop solutions with human expertise and experience. AI can aid in decision-making and evaluation at GKN, making this process more objective.

CP4 highlights that there is a challenged related to anticipating risk in project management. AI can assist in risk identify, assess, and prioritize risks. As it was stated in [CC2], the problem is not necessarily the risk itself, but not knowing that the risks exist in the first place. AI can potentially assist in this scenario, by providing better identifying risk, providing designers better understanding of the risks and consequently, enabling more tactical decision-making.

5.5.3 Balancing demands

The balancing of demands ultimately comes down to making a decision. While it's possible to continually refine aero-calculations to achieve better results, there comes a point where a trade-off must be made to move the project forward. In this context, AI can serve as a valuable tool in concept evaluation.

Similarly, LLMs can facilitate ideation in concept generation. Often, design processes can become stagnant if certain areas remain unexplored. With the assistance of LLMs, the balancing of demands can be streamlined in two significant ways: firstly, by suggesting solutions to seemingly impossible challenges, and secondly, by aiding in the decision-making process to prioritize performance in one area over another.

CB1: Balancing Ideation and Refinement During Design Exploration: The challenge of balancing ideation and refinement in design exploration is critical, as it requires ensuring a thorough evaluation of potential solutions while maintaining creative exploration. Generative design tools integrated with AI, such as those utilizing GANs and deep learning, are especially well-suited for this task.

Generative Design Tools: These tools employ algorithms to generate multiple design alternatives from initial specifications, which can then be refined and optimized through iterative processes. This not only accelerates the ideation phase by producing diverse solutions but also enhances the refinement phase by allowing rapid prototyping and evaluation of these solutions (Saadi & Yang, 2023b; Zhang et al., 2021). For instance, systems like ClipForge and PointE leverage deep learning to create and modify complex 3D models, enabling engineers to visualize and iterate on design solutions more effectively (Nichol et al., 2022; Sanghi et al., 2022).

LLMs can streamline the refinement process by generating detailed evaluations of designs based on vast amounts of training data, including design specifications, performance criteria, and compliance requirements (Göpfert et al., 2023). These models can simulate expert reasoning, offering insights that typically require significant human expertise, thus ensuring that each design iteration is thoroughly evaluated.

In term of concept evaluation, human factors such as subjectivity and bias ultimately play a role in how demands are balanced and coming up with objectively the best balance is complex as there are numerous factors to consider. Uncertainty also plays a large role here as is it might be difficult to say that balancing factors one way, is better than a similar way.

CB1 highlights that there is a challenge centered navigating trade-offs in design exploration. When there are so many factors to balance and consider, deciding on a solution can be complex. As it is difficult to identify a solution that objectively the best, designers might be forced to introduce human bias and subjectivity in order to create some kind of logic in this complexity. With AI-human collaboration, solutions can iteratively be developed, where an AI generates concepts, optimizes, and decides on the best ones according to certain criteria. Finding the concept with best performance might therefore be simplified, removing the number of factors, and ultimately assisting in balancing demands.

5.5.4 Human factor

CH1: Dependence on Team Composition for Innovative Conceptual Thinking

The dependence on specific team members for innovation poses a significant challenge, as it can lead to inconsistencies in creative output if key individuals are unavailable. Here, LLMs and multi-modal models can play a transformative role.

Decentralizing Creativity with LLMs: LLMs democratize the access to innovative problem-solving capabilities by providing all team members with tools that simulate expert thinking and creativity. For example, AutoTRIZ, an LLM-based tool, applies TRIZ methodologies to automate and enhance creative problem-solving, making innovative thinking accessible to every team member, regardless of their inherent creativity or experience (S. Jiang & Luo, 2024). This approach not only supports individual ideation but also enhances collaborative innovation, ensuring that the design process is not bottlenecked by individual creativity limitations.

Multi-modal Models for Enhanced Collaboration: Multi-modal AI models that process and integrate different data types—text, images, sketches—allow for a more inclusive and comprehensive understanding of design tasks (Driess et al., 2023; Gan et al., 2022). These models enable teams to contribute various forms of input, fostering a collaborative environment where different perspectives are synthesized. This capability is particularly beneficial in addressing complex engineering challenges where diverse viewpoints can lead to more innovative solutions.

Based on the answers provided by (Reitenbach et al., 2024), it is evident that integrating an AI-assisted chatbot could greatly enhance design processes in the initial stages. The early stages of design involve substantial amount of ”napkin-math,” and the implementation of a query-based workflow where specific tasks can be requested shows significant potential to reduce time. Additionally, as LLMs become more multi-modal, incorporating a speech-to-text application could further decrease the time required. For example, you could ask the WfMS questions like, ”Given the diameter of this fan, provide me with the outlet temperature,” to receive immediate feedback during meetings or when brainstorming ideas at the design desk.

In term of concept evaluation, removing human bias and subjectivity is one major capability and benefit of AI in concept evaluation, based on the literature study. For the case of GKN, human factors seem to play a large role when decision are made in the product development process.

CH1 highlights that creating teams can be a challenge as conceptual innovation is dependent on successfully creating teams with members with different qualities and reference frames. This indicates that human factors are important in order to evaluate concept successfully, and that they play a large role at GKN. With AI and human collaboration, teams are able to emphasize the strengths of human factors, while removing human restrictions, ultimately supporting concept development.

CH2 highlights that there is a challenge in achieving objectivity when performing concept scoring as subjective elements tend to influence the process which can result in selecting concepts that are not objectively the best. This is an area can AI can assist by coming up with performance optimized concepts that are objectively the best. However, there might be other factors than performance that is of interest and thus human factors can be added by choosing a concept with lower performance, but that is going to be easier to manufacture, for example.

5.5.5 Time management

CT1: Extensive time required for material strength and aero-performance analysis:

The challenge of reducing the time required for material strength and aero-performance analysis can be addressed effectively through the integration of Deep Learning and Simulation AI technologies.

Deep Learning Models: These models can predict material properties and aerodynamic performance based on vast datasets derived from previous experiments and simulations. For instance, models like those used in ClipForge can be adapted to predict material behavior under different stress conditions and environmental factors, potentially reducing the need for extensive physical testing (Jadhav & Farimani, 2024; Nichol et al., 2022).

GANs can be used to generate synthetic data for training models when actual test data are scarce. This approach can accelerate the development of predictive models for material strength and aerodynamics, ensuring faster and more accurate performance evaluations (Regenwetter et al., 2022).

5.5.6 Organizational synergy

CO1: Inefficient development processes for new technologies due to a lack of coordination across different disciplines:
Addressing inefficiencies in development processes due to poor coordination across different disciplines can be significantly improved by employing LLMs and collaborative AI systems. AI/LLMs address areas where 1) technology development is not common at all and 2) the disciplinary breadth itself is rarely well-suited for a more narrowly focused technology development process.

LLMs can serve as integration tools that facilitate communication and data sharing across disciplines. By processing natural language, technical reports, and design documents, LLMs can provide summaries, align objectives, and highlight interdependencies among various disciplines, enhancing coordination (Göpfert et al., 2023; Schick et al., 2023).

Collaborative AI systems can be configured to manage workflows and integrate inputs from various engineering disciplines into a cohesive development process. Tools like the AI-powered workflow engine in the German Aerospace Center's GTlab can automate the synthesis of inputs from different teams, ensuring that the development process is streamlined and more efficient (Reitenbach et al., 2024).

CO2: Recurring design challenges with the nozzle that persist through multiple iterations, resisting improvement:
For recurring design challenges, such as those with the nozzle, AI can provide deep

insights and innovative solutions to improve design iterations.

AutoTRIZ can be particularly useful in addressing complex recurring problems by applying TRIZ methodologies, which focus on inventive problem-solving strategies. By combining this with computational fluid dynamics (CFD) simulations, engineers can explore a broader range of potential improvements that are data-driven and innovative (S. Jiang & Luo, 2024).

Multi-modal AI Models which integrate data from sketches, CAD models, and simulations, can help in visualizing and analyzing different designs to identify potential flaws early in the design process. By iterating through designs rapidly with AI support, persistent issues can be resolved more efficiently (Driess et al., 2023; Gan et al., 2022).

5.5.7 Information

Information, such as internal documents and knowledge plays a vital role at GKN, as information lays a foundation in how demands are balanced. GKN faces several challenges, including the integration of knowledge and skills, where much information is retained as personal experience, making it hard to find documented lessons and leading to significant time spent on searching. Managing customer expectations and adapting to evolving requirements in product development is another issue, as is maintaining knowledge of past projects and new technologies due to infrequent development cycles.

CI1 highlights that finding necessary information and lessons learned from previous projects is difficult. AI can assist with knowledge support, by providing designers with lessons learned from previous projects, based on design-history-data, making correlations to current work. If GKN can create commonly accessible sets of data that can be shared in between internal projects or programs, there is a good opportunity to use various AI techniques to develop support for an AI to connect designers with existing knowledge and information. If designers at GKN, has the potential to access more documented information, perhaps the might not have to rely on subjective elements in their work, ultimately leading to more objective concept development.

To summarize, The literature review on concept evaluation and decision-making revealed a lack of clear guidance on how to incorporate these processes into an organization like GKN. While many studies are optimistic about the potential of AI in concept evaluation, few provide tangible frameworks applicable to complex industrial contexts. Most AI applications reviewed are limited to simple design tasks, such as sneaker design, and do not address the intricacies of industrial manufacturing.

Despite the absence of ready-to-implement frameworks, AI's theoretical potential to support concept evaluation and decision-making remains significant. The identified capabilities of AI, as illustrated, can be matched to the challenges outlined in the product development process at GKN. AI methods, including LLMs and GANs, can enhance creative reasoning, manage ambiguity, and bridge the gap between theoretical models and practical realities. These tools also aid in complexity and risk management, ensuring thorough documentation and simulations that predict real-world performance.

In conclusion, AI-human collaboration offers promising avenues for improving concept evaluation at GKN. AI can provide objective assessments, assist in risk identification, and enhance decision-making, thereby reducing human bias and subjectivity. However, successful implementation will require careful integration of AI capabilities with human expertise to navigate the complexities of industrial product development.

5.5.8 Limitations of methods in aerospace applications

In the context of aerospace design processes, the limitations of LLMs pose significant challenges that can impact their effectiveness and reliability. The nuanced exploration of these limitations can help in understanding the potential risks and constraints associated with deploying LLMs in highly technical and safety-critical environments like aerospace companies. Here's a detailed discussion based on the limitations identified:

Interpretability Issues
The interpretability of LLM-generated concepts is a major concern in aerospace design, where each component and system must meet stringent standards for safety and functionality. The lack of reliable metrics to validate the quality and viability of generated concepts means that LLM outputs must be closely scrutinized by human experts, potentially negating the efficiency gains expected from using AI in the design process (Q. Zhu & Luo, 2023). The development of robust evaluation metrics is critical to integrating LLMs into aerospace design workflows effectively. Furthermore, the non-deterministic (non repetitive) nature of AI is in contrast to what the aerospace industry normally look for.

Generalizability and Extendibility
LLMs face challenges in generalizing across the diverse tasks involved in aerospace design, which often require deep, domain-specific knowledge (Q. Zhu & Luo, 2023). The absence of comprehensive and high-quality datasets for training further restricts the LLMs' ability to extend their learning to new or less common design tasks (Regenwetter et al., 2022). For aerospace applications, where designs often include novel materials or innovative engineering approaches, the inability of LLMs to effectively generalize or extend their capabilities can limit their usefulness.

Lack of Baselines
The absence of established baselines for comparing the performance of LLMs com-

plicates the assessment of AI-generated designs (Q. Zhu & Luo, 2023). In aerospace, where existing design methodologies are well-tested and validated, introducing AI without clear benchmarks can create uncertainty about the reliability and quality of design outputs.

Formalization of Engineering Design Dialogue
The "black box" nature of LLMs makes it difficult to trace how design outputs are derived, which is a significant issue in environments that demand high levels of documentation and traceability (Beitz et al., 1996; Fricke, 1996). This opacity is a barrier to certifying AI-assisted designs for actual implementation in aerospace projects. Unlike simulations, which eliminate the modeling of uncertainty in the output where a specific set of inputs can be re-run with the same results, an LLM does not. An LLM does not generate the same output when re-runned based on the same input, which is a significant issue within the context of aerospace where results need to be 100% correct every time.

Interface and Integration Challenges
Many aerospace engineering tools do not support textual interfaces required by current LLMs, limiting the integration of these AI models into existing software ecosystems (Schick et al., 2023). The need for specialized interfaces can hinder the adoption of LLMs in aerospace design processes, where seamless integration with CAD tools and simulation software is crucial.

Skill and Representation Correlation
LLMs primarily process textual data and lack the capability to fully comprehend spatial and physical processes that are vital in aerospace engineering (Fricke, 1996). This limitation affects their ability to accurately model complex systems that rely on a deep understanding of spatial dynamics and material properties.

Risk of Hallucinations
The propensity of LLMs to generate "hallucinated" outputs—information that is plausible but incorrect or unverifiable—poses a significant risk in aerospace design (Alkaissi & McFarlane, 2023; Huh et al., 2023). The inaccuracies or fabrications produced by LLMs could lead to flawed design decisions if not adequately vetted, potentially compromising the safety and functionality of aerospace components.

Legal limitations and data secrecy
A major limitation is that AI at GKN cannot train on export-controlled or military data. Consequently, many useful datasets necessary for robust concept development are unavailable for training. This results in trained models lacking essential data, potentially leading to non-useful outputs.

Although the interview study indicated that data could be anonymized before training, making it possible to train on export-controlled data, another respondent contradicted this, making it difficult to draw a final conclusion. Further research on this topic is needed.

While LLMs offer potential benefits in automating and enhancing the aerospace design process, these limitations highlight the need for cautious implementation. Ensuring that AI tools like LLMs are used to complement, rather than replace, human expertise is crucial. Continued research and development into improving the interpretability, reliability, and integration capabilities of LLMs will be essential to mitigate these limitations and safely harness the power of AI in aerospace design.

Conclusion
There is not a way where LLMs can provide a complete design based on an input of requirements and a prompt. However, LLMs have the possibility to enhance the workflow in every iteration of the design process.

5.5.9 Sources of uncertainty and what may have affected the result

This section reflects on the intricacies of the study and how these may have affected the results.

5.5.9.1 LLMs in concept generation

Something that is important to note is that most of the research done when testing LLMs (or GPTs) in concept generation and development applications is that they have mainly focused on GPT-2 and GPT-3. This could be due to the rapid advancements in the releases of new GPT versions, and the time it takes to perform a thorough study of the models. The current version ChatGPT uses is GPT-4 and was released in March 14, 2023 (OpenAI, 2023). Given that GPT-3 is a LLM featuring 175 billion parameters, which was a substantial scale-up from its predecessor, GPT-2, which had 1.5 billion parameters.

Capabilities: GPT-3 made waves for its ability to generate coherent and contextually relevant text across a variety of domains, often producing text that could be indistinguishable from that written by a human. It is capable of understanding and generating natural language or code from a prompt. GPT-4 further advances the capabilities of its predecessors, offering significantly improved performance in generating text, better factual accuracy, and more nuanced understanding of complex instructions.

Limitations: Despite its size and capabilities, GPT-3 sometimes generates plausible-sounding but incorrect or nonsensical answers, a trait referred to as "hallucination." Additionally, it can manifest biases present in the data it was trained on. While GPT-4 scales up to about 2 trillion parameters, a substantial increase from GPT-3, making it more powerful and capable of understanding and generating more complex texts (OpenAI, 2023).

This version has been described as more reliable, with reduced tendencies to generate incorrect information and an ability to handle more abstract and subtle tasks. It also demonstrates a better grasp of user intent and more detailed and specific outputs. Based on the information above, one could argue that GPT-4 would offer better results than GPT-3.

5.5.10 Literature synthesis concept generation

Saadi and Yang (2023b) mentions that computational tools in design can influence the designer's cognitive processes, their design exploration, and overall designs generated. This along with the study that considers AI in concept evaluation, and a generative design process outlined by (Saadi & Yang, 2023b) gave inspiration to create a process of our own. This, along with the results from the interviews, which mapped the PDP and concept generation, has culminated in a comprehensive process flow that incorporates AI/ML methods such as LLMs as a designer, and generative designer tools. By integrating insights from concept generation and evaluation, the authors of this report propose a design process that demonstrates how AI/ML methods can aid in developing robust concept solutions. This proposed process is illustrated in Figure 5.4.

Define Inputs: Objectives, Parameters, and Constraints: The generative design process starts with defining objectives, parameters, and constraints. Objectives are performance metrics, such as minimizing weight or maximizing stiffness. Parameters include material properties, manufacturing methods, safety factors, and loading conditions like forces and moments. Constraints are limiting conditions like maximum weight and geometry restrictions.

Inputs come from user needs, customer requirements, or industry standards. Designers use calculations and intuition to set initial values, such as estimating an upward force using F=MA calculations and conservatively overshooting weight objectives for strength. Tools may limit input types, some only allow force loads, not torques or dynamic loads. While qualitative factors like aesthetics are not official objectives, designers consider them subconsciously and adjust designs later to incorporate these factors. The process involves refining inputs and using the generative design tool to meet specified objectives, parameters, and constraints.

Creative reasoning, argumentative dialogue, ideation: In this stage, the idea is that an AI-based method, such as a LLM trained on internal aerospace-specific data and specific GKN designer data, could improve creative reasoning and thus the output of concepts. Additionally, to ensure all aspects are considered, the process would, in our estimation, become more effective and less time-consuming.

Evaluate and iterate: Designers evaluate generative design results through visual inspection, analytical methods, and prototyping. Visual checks identify features that don't meet specifications. Analytical methods include graphing performance comparisons and using finite element analysis (FEA) to find improvement areas. Prototyping assesses comfort and usability, which screens can't gauge.

Based on evaluations, designers iterate on constraints, parameters, and sometimes

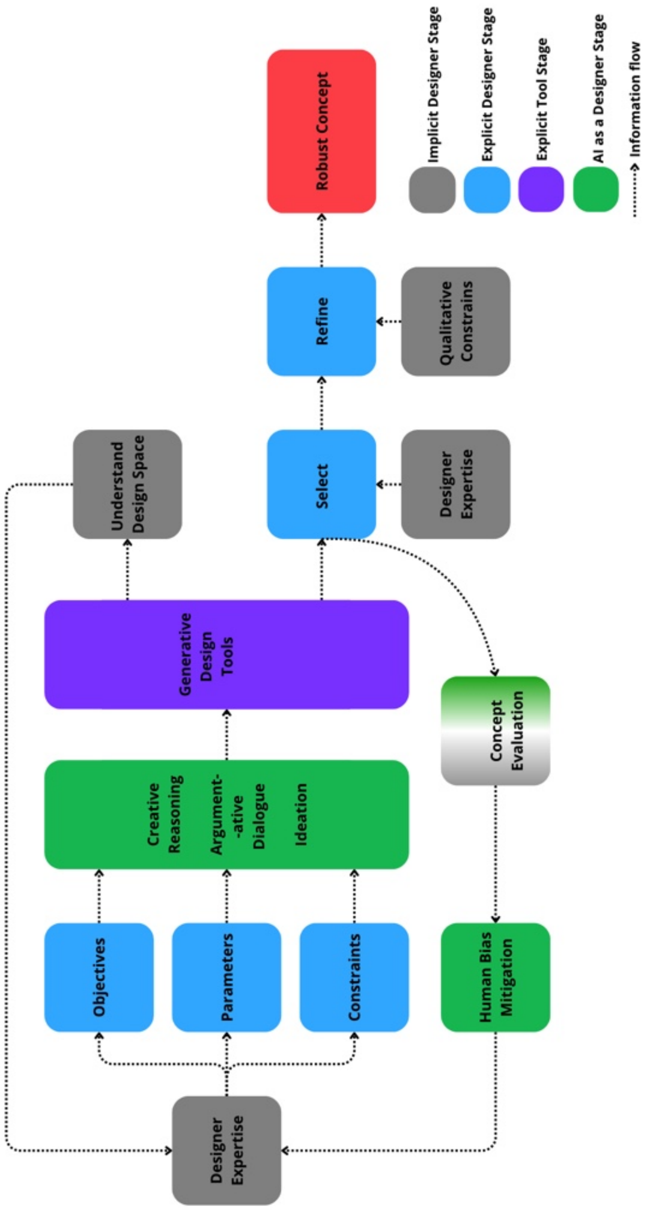

Figure 5.4: The design process illustrating how AI/ML methods can help propose a robust design solution.

objectives, using experience, trial and error, or aesthetic goals. This iterative process can involve many versions to achieve satisfactory results, sometimes up to 37 iterations. Time and computational resources often limit iterations (Saadi & Yang, 2023b).

Despite not fully understanding the tool's final designs, designers trust their setup and the tool's capabilities, tweaking values to refine designs. After iterating, designers manually select a final design from the generated set, balancing performance metrics, manufacturability, aesthetics, and user satisfaction. For example, designers might reject designs that don't align with their intuition if they notice features that seem suspicious.

Designers may choose results based on different performance metrics not represented in the tool, like the moment of inertia, or select lower-performance iterations to improve manufacturability. Aesthetics can also influence the selection process, with lower-performance designs sometimes preferred for their visual appeal.

Selection may also consider context-specific requirements, ensuring the design meets user needs. This range allows the final user to choose the most appropriate geometry.

Designer Expertise: is crucial in the generative design process, influencing all stages with their experience, knowledge, intuition, and understanding of users and context. They define and iterate on objectives, parameters, and constraints, select results, and refine the final design. Expertise in traditional CAD software, engineering, and manufacturing processes is essential to mastering generative design tools.

Designers set relevant objectives, parameters, and constraints at the beginning stages. Their knowledge helps define variables and iterate through them. They choose the best design based on quantitative and qualitative metrics like manufacturability and aesthetics. Designers translate user and context specifications into tool-understandable parameters, combining their knowledge with the tool's computing power to create optimized designs.

Qualitative Considerations: Generative design tools primarily handle quantitative inputs, but qualitative considerations like aesthetics are crucial. Designers influence aesthetics by defining starting geometries, creating a bounding box that impacts the tool's output. While current tools don't accommodate direct aesthetic inputs, designers find workarounds to ensure visual appeal. Some tools exploring aesthetic designs are in development but not widely used (Saadi & Yang, 2023b).

Manufacturing and assembly considerations are also significant. Designers add constraints for assembly tools, like ensuring clearance for a screwdriver. These qualitative factors greatly influence the generative design process outcomes.

Exploring and Understanding Design Space: A significant implicit output of the generative design process is the designer's enhanced understanding of the design space. Through iterative processes, designers gain a deeper comprehension of the design problem and potential solutions.

Designers view this process as a learning experience, building confidence and understanding by exploring different scenarios and questioning the tool's outputs. This trial-and-error approach helps them grasp the intricacies of the design problem better. Initially, designers encounter a learning curve, identifying factors they initially overlooked. This iterative learning helps them recognize and include all relevant constraints, refining their understanding of the design problem and identifying key constraints that drive the solution space.

The generative design tool: outputs multiple designs that meet the specifications, providing designers with a comprehensive understanding of the solution space. This breadth of potential solutions serves as valuable design guidance, unique to the generative design process and not easily achieved through traditional methods.

Concept evaluation: The gradient color showing half grey, half green box symbolises that while AI can help with concept evaluation they will always be overseen by an designer in this stage. The AI will not evaluate concepts on its own, but only provide the designer with information and basis for a decision.

Human bias mitigation: When evaluating concepts, this thesis holds the opinion that there must be one concept that is objectively the best. However, subjectively, another concept could be considered better. The main idea is that designers should be provided with the grading of the objectively best concept, so that any potential bias can be noticed during the evaluation stage.

6 Conclusion and recommendations

GKN encounters various challenges in concept development, such as information accessibility, data secrecy, time constraints, human factors, balancing demands, and a conservative culture. AI/ML methods have the potential to aid in some of these challenges. Addressing these issues through better data management, improved team dynamics, and flexible processes, could help simplify AI/ML implementation.

Current AI/ML capabilities at GKN are in the early stages, with specific applications like design optimization showing promise. However, broader adoption is limited by gaps in strategic direction, data quality, infrastructure, and regulatory compliance concerns. Investments in IT infrastructure and AI/ML expertise would be necessary to fully utilize these technologies.

Generative AI, design tools with integrated AI/ML methods, and LLMs can simplify concept generation and evaluation, enhance creative reasoning, mitigate human bias, and manage complexity and risk. However, to fully leverage these benefits, GKN must facilitate AI-human collaboration to also capitalize on human strengths. AI implementation is not about replacing humans but assisting in areas where human capabilities are lacking.

We argue that a fine-tuned LLM, trained on internal and aerospace-specific data, would enhance the robustness of concept solutions and streamline design processes as well as non-cognitive tasks. This would aid in creative reasoning and ideation, extend the designers knowledge and the design space, and mitigate bias in the evaluation of concepts. However, several prerequisites must be met before this can be realized, including high data quality, robust infrastructure, adequate resources (both human and machine), and strict data secrecy. Given these requirements, implementing such a model in the near term appears unlikely.

To move towards this goal, we recommend establishing effective data governance policies suitable for future model training. Additionally, there should be enterprise-wide education about the capabilities and potential of AI/ML/LLMs. Moreover, commercial generative design tools are making rapid advancements that can independently streamline many processes. The primary challenge lies in the data sharing between GKN and these tools.

Finally, we argue that by enhancing the knowledge space and, by extension, the design space, AI/ML methods can help robust concept development. Generative

AI, design tools with integrated AI/ML methods, and LLMs offer significant opportunities to simplify concept generation. LLMs can assist with ideation, creative reasoning, and offloading of cognitive and repetitive tasks. Fine-tuned LLMs, trained on internal documentation, provide instant feedback on less complex tasks, helping designers explore a broader design space, reduce bias, and enhance knowledge, facilitating the development of robust design solutions.

6.1 Further work

This thesis might be considered a preliminary study on the potential of implementing AI-based methods, such as LLMs, in early conceptual design work. Given that it is a pre-study, all aspects covered in the thesis could be explored in greater depth. For instance, keeping up with and monitoring the rapid advancements in AI is a significant aspect.

Comparing AI implementation at the three benchmarking organizations to where GKN currently stands, it is complex to decipher whether GKN is ahead or behind. However, what it shows is that AI is on all three organizations radar, building an incentive to further investigate how AI actually can be implemented into the organization. While GKN is looking into AI implementation, this should create an urge to further identify ways in which GKN can benefit from this technology.
We recommend that the next step would be to fine-tune an LLM with aerospace-specific and internal data and assess its performance in creative reasoning, ideation, and concept evaluation during concept development.

References

Ahmad, A. (2019). Reliable and fault tolerant systems on chip through design for testability. https://doi.org/10.1109/AICAI.2019.8701390

Ahmed, I., Jeon, G., & Piccialli, F. (2022). From artificial intelligence to explainable artificial intelligence in industry 4.0: A survey on what, how, and where. *IEEE Transactions on Industrial Informatics*, *18*(8), 5031–5042. https://doi.org/10.1109/TII.2022.3146552

Aksit, E., Haapala, K., & Tabei, A. (2023). Digital multiphase material microstructures for image-based ai methods. *Lecture Notes in Mechanical Engineering*, 723–734.

Alcaide-Marzal, J., Diego-Mas, J. A., & Acosta-Zazueta, G. (2020). A 3d shape generative method for aesthetic product design. *Design Studies*, *66*, 144–176. https://doi.org/https://doi.org/10.1016/j.destud.2019.11.003

Ali, M., Hu, Y.-F., Luong, D., Oguntala, G., Li, J.-P., & Abdo, K. (2020). Adversarial attacks on ai based intrusion detection system for heterogeneous wireless communications networks, 1–6. https://doi.org/10.1109/DASC50938.2020.9256597

Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in chatgpt: Implications in scientific writing. *Cureus*, *15*(2).

Allison, J. T., Cardin, M.-A., McComb, C., Ren, M. Y., Selva, D., Tucker, C., Witherell, P., & Zhao, Y. F. (2022). Artificial intelligence and engineering design. *Journal of Mechanical Design*, *144*(2), 020301.

Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2020). A systematic review on supervised and unsupervised machine learning algorithms for data science. *Supervised and unsupervised learning for data science*, 3–21.

Altshuller, G. S. (1999). *The innovation algorithm: Triz, systematic innovation and technical creativity*. Technical Innovation Center Inc.

Andersson, P. (1996). A semi-analytic approach to robust design in the conceptual design phase. *Research in Engineering Design*, *8*, 229–239. https://doi.org/10.1007/BF01597229

Arvidsson, M., & Gremyr, I. (2008). Principles of robust design methodology. *Quality and Reliability Engineering International*, *24*(1), 23–35. https://doi.org/https://doi.org/10.1002/qre.864

Athaluri, S. A., Manthena, S. V., Kesapragada, V. K. M., Yarlagadda, V., Dave, T., & Duddumpudi, R. T. S. (2023). Exploring the boundaries of reality: Investigating the phenomenon of artificial intelligence hallucination in scientific writing through chatgpt references. *Cureus*, *15*(4).

Autodesk. (2024). Generative design ai software [Accessed: 2024-06-11]. https://www.autodesk.com/solutions/generative-design-ai-software

Badillo, S., Banfai, B., Birzele, F., Davydov, I. I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B., & Zhang, J. D. (2020). An introduction to machine learning. *Clinical pharmacology & therapeutics*, *107*(4), 871–885.

Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, *33*(1), 1–17. https://doi.org/10.1007/s12525-023-00680-1

Becue, A., Praça, I., & Gama, J. (2021). Artificial intelligence, cyber-threats and industry 4.0: Challenges and opportunities. *Artificial Intelligence Review*, *54*. https://doi.org/10.1007/s10462-020-09942-2

Beitz, W., Pahl, G., & Grote, K. (1996). Engineering design: A systematic approach. *Mrs Bulletin*, *71*.

Bell, E., Bryman, A., & Harley, B. (2019). *Business research methods: Fifth edition*. Oxford University Press.

Bengesi, S., El-Sayed, H., Sarker, M. K., Houkpati, Y., Irungu, J., & Oladunni, T. (2024). Advancements in generative ai: A comprehensive review of gans, gpt, autoencoders, diffusion model, and transformers. *IEEE Access*.

Boiko, D. A., MacKnight, R., Kline, B., & Gomes, G. (2023). Autonomous chemical research with large language models. *Nature*, *624*(7992), 570–578.

Boothroyd, G. (1987). Design for assembly—the key to design for manufacture. *The International Journal of Advanced Manufacturing Technology*, *2*, 3–11.

Bousdekis, A., Lepenioti, K., Apostolou, D., & Mentzas, G. (2021). A review of data-driven decision-making methods for industry 4.0 maintenance applications. *Electronics*, *10*(7), 828.

Briard, T., Segonds, F., & Zamariola, N. (2020). G-dfam: A methodological proposal of generative design for additive manufacturing in the automotive industry. *Int J Interact Des Manuf*, *14*, 875–886. https://doi.org/10.1007/s12008-020-00669-6

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance? *Available at SSRN 1819486*.

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative ai at work* (tech. rep.). National Bureau of Economic Research.

Brynjolfsson, E., & McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review*, *106*(5), 133–139.

Caiazzo, B., Di Nardo, M., Murino, T., Petrillo, A., Piccirillo, G., & Santini, S. (2022). Towards zero defect manufacturing paradigm: A review of the state-of-the-art methods and open challenges. *Computers in Industry*, *134*, 103548.

Camburn, B., Arlitt, R., Anderson, D., Sanaei, R., Raviselvam, S., Jensen, D., & Wood, K. (2020). Computer-aided mind map generation via crowdsourcing and machine learning. *Research in Engineering Design*, *31*. https://doi.org/10.1007/s00163-020-00341-w

Camburn, B., He, Y., Raviselvam, S., Luo, J., & Wood, K. (2020). Machine learning-based design concept evaluation. *Journal of Mechanical Design*, *142*(3), 031113.

Chandrasegaran, S. K., Ramani, K., Sriram, R. D., Horváth, I., Bernard, A., Harik, R. F., & Gao, W. (2013). The evolution, challenges, and future of knowledge representation in product design systems [Solid and Physical Modeling 2012]. *Computer-Aided Design*, *45*(2), 204–228. https://doi.org/https://doi.org/10.1016/j.cad.2012.08.006

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., et al. (2023). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*.

Chaudhari, A. M., & Selva, D. (2023). Evaluating Designer Learning and Performance in Interactive Deep Generative Design. *Journal of Mechanical Design*, *145*(5), 051403. https://doi.org/10.1115/1.4056374

Chen, H. (2022, May). *Data quality evaluation and improvement for machine learning* [Doctoral dissertation]. https://doi.org/10.13140/RG.2.2.15870.87361

Clark, K. B. (1991). Product development performance: Strategy, organization, and management in the world auto industry. *(No Title)*.

Cooper, R. G. (1990). Stage-gate systems: A new tool for managing new products. *Business Horizons*, *33*(3), 44–54. https://doi.org/https://doi.org/10.1016/0007-6813(90)90040-I

Crossley, W. A. (1999). Optimization for aerospace conceptual design through the use of genetic algorithms. *Proceedings of the First NASA/DoD Workshop on Evolvable Hardware*, 200–207.

Cusumano, M. A., & Nobeoka, K. (1998). *Thinking beyond lean: How multi-project environment is transforming product development at toyota and other companies.* Free Press.

Dalpiaz, F., Ferrari, A., Franch, X., & Palomares, C. (2018). Natural language processing for requirements engineering: The best is yet to come. *IEEE Software*, *35*(5), 115–119. https://doi.org/10.1109/MS.2018.3571242

DARPA. (2017). Darpa perspective on ai [Accessed: 2024-06-02]. https://www.darpa.mil/about-us/darpa-perspective-on-ai

Dasari, S. K. (2021, October). *Predictive modelling to support design and manufacturing in aerospace engineering* (Publication No. 7) [Doctoral Dissertation]. Blekinge Institute of Technology. http://urn.kb.se/resolve?urn=urn:nbn:se:bth-22033

Demirel, H. O., Goldstein, M. H., Li, X., & Sha, Z. (2024). Human-centered generative design framework: An early design framework to support concept creation and evaluation. *International Journal of Human–Computer Interaction*, *40*(4), 933–944.

de Moura, L., & Bjørner, N. (2008). Z3: An efficient smt solver. In C. R. Ramakrishnan & J. Rehof (Eds.), *Tools and algorithms for the construction and analysis of systems* (pp. 337–340). Springer Berlin Heidelberg.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. https://doi.org/10.48550/arXiv.1810.04805

Dong, S., Wang, P., & Abbas, K. (2021). A survey on deep learning and its applications. *Computer Science Review*, *40*, 100379.

Driess, D., Xia, F., Sajjadi, M. S. M., Lynch, C., Chowdhery, A., Ichter, B., Wahid, A., Tompson, J., Vuong, Q., Yu, T., Huang, W., Chebotar, Y., Sermanet, P., Duckworth, D., Levine, S., Vanhoucke, V., Hausman, K., Toussaint, M., Greff, K., … Florence, P. (2023). Palm-e: An embodied multimodal language model.

Dubois, A., & Gadde, L.-E. (2002). Systematic combining: An abductive approach to case research [Markets as Networks]. *Journal of Business Research*, *55*(7), 553–560. https://doi.org/https://doi.org/10.1016/S0148-2963(00)00195-8

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., et al. (2021). Artificial intelligence (ai): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 101994.

Elwakil, S. (2019, March). *Processes and design for manufacturing, third edition*. https://doi.org/10.1201/9780429506635

Eres, M. H., Bertoni, M., Kossmann, M., & Scanlan, J. (2014). Mapping customer needs to engineering characteristics: An aerospace perspective for conceptual design. *Journal of Engineering Design*, *25*(1-3), 64–87.

Fafchamps, D. (1994). Organizational factors and reuse. *IEEE Software*, *11*(5), 31–41. https://doi.org/10.1109/52.311049

Favi, C., Germani, M., & Mandolini, M. (2016). A multi-objective design approach to include material, manufacturing, and assembly costs in the early design phase. *Procedia CIRP*, *52*, 251–256. https://doi.org/10.1016/j.procir.2016.07.043

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative ai. *Business & Information Systems Engineering*, *66*(1), 111–126.

Firesmith, D. (2007). Common requirements problems, their negative consequences, and the industry best practices to help solve them. *Journal of Object Technology*, *6*(1), 17–33. https://doi.org/10.5381/jot.2007.6.1.c2

Foivos Psarommatis, P.-A. D., Gökan May, & Kiritsis, D. (2020). Zero defect manufacturing: State-of-the-art review, shortcomings and future directions in research. *International Journal of Production Research*, *58*(1), 1–17. https://doi.org/10.1080/00207543.2019.1605228

Ford, R. B. (1996). *Process for the conceptual design of robust mechanical systems—going beyond parameter design to achieve world-class quality*. Stanford University.

Formentini, G., Boix Rodríguez, N., & Favi, C. (2022). Design for manufacturing and assembly methods in the product development process of mechanical products: A systematic literature review. *The International Journal of Advanced Manufacturing Technology*, *120*(7), 4307–4334. https://doi.org/10.1007/s00170-022-08837-6

Fragapane, G., Eleftheriadis, R., Powell, D., & Antony, J. (2023). A global survey on the current state of practice in zero defect manufacturing and its impact on production performance. *Computers in Industry*, *148*, 103879. https://doi.org/https://doi.org/10.1016/j.compind.2023.103879

Fricke, G. (1996). Successful individual approaches in engineering design. *Research in Engineering Design*, *8*(3), 151–165. https://doi.org/10.1007/BF01608350

Gan, Z., Li, L., Li, C., Wang, L., Liu, Z., & Gao, J. (2022). Vision-language pretraining: Basics, recent advances, and future trends. *Foundations and Trends in Computer Graphics and Visualization*, *14*(3-4), 163–352. https://doi.org/10.1561/0600000105

GKN Aerospace. (2024a). About us [Accessed: 2024-06-05].

GKN Aerospace. (2024b). Gkn aerospace in sweden [Accessed: 2024-06-05].

GKN Aerospace. (2024c). Gkn aerospace sweden [Accessed: 2024-06-05].
GKN Aerospace. (2024d). Markets & solutions [Accessed: 2024-06-05].
Goh, T. N. (1993). Taguchi methods: Some technical, cultural and pedagogical perspectives. *Quality and Reliability Engineering International*, *9*, 185–202.
Gomez, A., Krus, P., Panarotto, M., & Ola, I. (2024). Large language models in complex system design. *Proceedings of the Design Society*, *4*, 2197–2206. https://doi.org/10.1017/pds.2024.222
Göpfert, J., Weinand, J. M., Kuckertz, P., & Stolten, D. (2023). Opportunities for large language models and discourse in engineering design.
Gremyr, I., Arvidsson, M., & Johansson, P. (2003). Robust design methodology: Status in the swedish manufacturing industry. *Quality and Reliability Engineering International*, *19*, 285–293.
Guarino, G., Samet, A., & Cavallucci, D. (2022). Patriz: A framework for mining triz contradictions in patents. *Expert Systems with Applications*, *207*, 117942. https://doi.org/10.1016/j.eswa.2022.117942
Guraksin, A. M., & Ozcan, A. (2023). Aco-based approach for integrating product lifecycle management with mro services in aviation industry. *Soft Comput*, *27*, 337–361. https://doi.org/10.1007/s00500-022-07560-4
Gyory, J. T., Kotovsky, K., McComb, C., & Cagan, J. (2022). Comparing the Impacts on Team Behaviors Between Artificial Intelligence and Human Process Management in Interdisciplinary Design Teams. *Journal of Mechanical Design*, *144*(10), 104501. https://doi.org/10.1115/1.4054723
Gyory, J. T., Soria Zurita, N. F., Martin, J., Balon, C., McComb, C., Kotovsky, K., & Cagan, J. (2021). Human Versus Artificial Intelligence: A Data-Driven Approach to Real-Time Process Management During Complex Engineering Design. *Journal of Mechanical Design*, *144*(2), 021405. https://doi.org/10.1115/1.4052488
Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California management review*, *61*(4), 5–14.
Hall, S., Mollan, C., Pandey, V., & Mourelatos, Z. (2022, August). *TRIZ Mapping and Novelty Detection of Engineering Design Patents Using Machine Learning* (Vol. Volume 6: 34th International Conference on Design Theory and Methodology (DTM)). https://doi.org/10.1115/DETC2022-89746
Halpin, J. F. (1966). *Zero defects: A new dimension in quality assurance*. Martin Company.
Han, Y., Nanda, G., & Moghaddam, M. (2023). Attribute-sentiment-guided summarization of user opinions from online reviews. *Journal of Mechanical Design*, *145*, 041402.
Haskins, B., Stecklein, J., Dick, B., Moroney, G., Lovell, R., & Dabney, J. (2004). 8.4. 2 error cost escalation through the project life cycle. *INCOSE International Symposium*, *14*, 1723–1737. https://doi.org/10.1002/j.2334-5837.2004.tb00608.x
Hassan, K., Thakur, A. K., Singh, G., Singh, J., Gupta, L. R., & Singh, R. (2024). Application of artificial intelligence in aerospace engineering and its future directions: A systematic quantitative literature review. *Archives of Computational Methods in Engineering*. https://doi.org/10.1007/s11831-024-10105-7

Heck, T. G. (2023). What artificial intelligence knows about 70 kda heat shock proteins, and how we will face this chatgpt era. *Cell Stress and Chaperones*, *28*(3), 225–229.

Holtzman, A., Buys, J., Du, L., Forbes, M., & Choi, Y. (2019). The curious case of neural text degeneration. *arXiv preprint arXiv:1904.09751*.

Hosseini, M., Gao, C. A., Liebovitz, D. M., Carvalho, A. M., Ahmad, F. S., Luo, Y., MacDonald, N., Holmes, K. L., & Kho, A. (2023). An exploratory survey about using chatgpt in education, healthcare, and research. *Plos one*, *18*(10), e0292216.

Hu, Z., Wang, L., Lan, Y., Xu, W., Lim, E.-P., Bing, L., Xu, X., Poria, S., & Lee, R. K.-W. (2023). Llm-adapters: An adapter family for parameter-efficient fine-tuning of large language models. *arXiv preprint arXiv:2304.01933*.

Huang, J., & Chang, K. C.-C. (2023). Towards reasoning in large language models: A survey.

Huh, Y., Lee, J.-Y., & Finney, S. C. (2023). Not an author but an increasingly proficient secretary. *Episodes Journal of International Geoscience*, *46*(2), 131–131.

IBM. (2024). *Large language models*. https://www.ibm.com/topics/large-language-models

Ilevbare, I. M., Probert, D., & Phaal, R. (2013). A review of triz, and its benefits and challenges in practice. *Technovation*, *33*, 30–37.

Jackson, E., Kang, E., Dahlweid, M., Seifert, D., & Santen, T. (2010). Components, platforms and possibilities: Towards generic automation for mda, 39–48. https://doi.org/10.1145/1879021.1879027

Jackson, E., Seifert, D., Dahlweid, M., Santen, T., Bjørner, N., & Schulte, W. (2009). Specifying and composing non-functional requirements in model-based development, 72–89. https://doi.org/10.1007/978-3-642-02655-3_7

Jackson, E., & Sztipanovits, J. (2009). Formalizing the structural semantics of domain-specific modeling languages. *Software & Systems Modeling*, *8*(4), 451–478. https://doi.org/10.1007/s10270-008-0105-0

Jadhav, Y., & Farimani, A. B. (2024). Large language model agent as a mechanical designer.

Jagadish, Bhowmik, S., & Ray, A. (2019). Prediction of surface roughness quality of green abrasive water jet machining: A soft computing approach. *Journal of Intelligent Manufacturing*, *30*(8), 2965–2979.

Janssen, M., Hartog, M., Matheus, R., Yi Ding, A., & Kuk, G. (2022). Will algorithms blind people? the effect of explainable ai and decision-makers' experience on ai-supported decision-making in government. *Social Science Computer Review*, *40*(2), 478–493.

Jiang, S., & Luo, J. (2024). Autotriz: Artificial ideation with triz and large language models.

Jiang, Y., Li, X., Luo, H., Yin, S., & Kaynak, O. (2022). Quo vadis artificial intelligence? *Discover Artificial Intelligence*, *2*(1), 4.

Johansson, P., Chakhunashvili, A., & Bergman, B. (2006). Variation mode and effect analysis: A practical tool for quality improvement. *Quality and Reliability Engineering*, *22*, 865–876.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus; Giroux.

Kambhampaty, J., Schlichting, G., Coletti, C., Evans, J., Reddy, A., Pinon Fischer, O., Mavris, D., & Graves, R. (2024). Graph-based digital file curation for engineering reuse: Methodology and case study. https://doi.org/10.2514/6.2024-1133

Kang, E., Jackson, E., & Schulte, W. (2011). An approach for effective design space exploration. In R. Calinescu & E. Jackson (Eds.), *Foundations of computer software. modeling, development, and verification of adaptive systems* (pp. 33–54). Springer Berlin Heidelberg.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., et al. (2023). Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and individual differences*, *103*, 102274.

Kazi, R., Grossman, T., Cheong, H., Hashemi, A., & Fitzmaurice, G. (2017). Dreamsketch: Early stage 3d design explorations with sketching and generative design, 401–414. https://doi.org/10.1145/3126594.3126662

Khaleel, M., Ahmed, A. A., & Alsharif, A. (2023). Artificial intelligence in engineering. *Brilliance: Research of Artificial Intelligence*, *3*(1), 32–42.

Kotlarsky, J., Oshri, I., Kumar, K., & van Hillegersberg, J. (2008). Towards agility in design in global component-based development. *Commun. ACM*, *51*(9), 123–127. https://doi.org/10.1145/1378727.1389973

Kotler, P., Wong, V., Saunders, J., & Armstrong, G. (2005). *Principles of marketing* (4th EU Edition). Pearson Education.

Krish, S. (2011). A practical generative design method. *Computer-Aided Design*, *43*(1), 88–100. https://doi.org/https://doi.org/10.1016/j.cad.2010.09.009

Kulin, M., Fortuna, C., De Poorter, E., Deschrijver, D., & Moerman, I. (2016). Data-driven design of intelligent wireless networks: An overview and tutorial. *Sensors*, *16*(6), 790.

Kulin, M., Kazaz, T., De Poorter, E., & Moerman, I. (2021). A survey on machine learning-based performance improvement of wireless networks: Phy, mac and network layer. *Electronics*, *10*(3), 318.

Leberruyer, N., Bruch, J., Ahlskog, M., & Afshar, S. (2023). Toward zero defect manufacturing with the support of artificial intelligence—insights from an industrial application. *Computers in Industry*, *147*, 103877. https://doi.org/https://doi.org/10.1016/j.compind.2023.103877

Lee, K. (2021). A systematic review on social sustainability of artificial intelligence in product design. *Sustainability*, *13*(5), 2668.

Li, A., Wu, J., & Bigham, J. P. (2023). Using llms to customize the ui of webpages. *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. https://doi.org/10.1145/3586182.3616671

Li, C., Pan, H., Bousseau, A., & Mitra, N. (2022). Free2cad: Parsing freehand drawings into cad commands. *ACM Transactions on Graphics*, *41*, 1–16. https://doi.org/10.1145/3528223.3530133

Li, K., Hopkins, A. K., Bau, D., Viégas, F., Pfister, H., & Wattenberg, M. (2023). Emergent world representations: Exploring a sequence model trained on a synthetic task.

Liao, J., Hansen, P., & Chai, C. (2020). A framework of artificial intelligence augmented design support. *Human–Computer Interaction*, *35*(5-6), 511–544.

Liker, J. K., Sobek, D. K., Ward, A. C., & Cristiano, J. J. (1996). Involving suppliers in product development in the united states and japan: Evidence for set-based concurrent engineering. *IEEE Transactions on Engineering Management*, *43*(2), 165–178.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, *55*(9), 1–35.

Liu, X.-Y., Wang, G., & Zha, D. (2023). Fingpt: Democratizing internet-scale data for financial large language models. *arXiv preprint arXiv:2307.10485*.

Lopez, C. E., Miller, S. R., & Tucker, C. S. (2018). Exploring biases between human and machine generated designs. *ASME J. Mech. Des.*, *141*(2), 021104. https://doi.org/10.1115/1.4040702

Lubert, T. (2005). How can computers be partners in the creative process. *International Journal of Human-Computer Studies*, *63*, 365–369.

Lv, K., Yang, Y., Liu, T., Gao, Q., Guo, Q., & Qiu, X. (2023). Full parameter fine-tuning for large language models with limited resources.

Ma, K., Grandi, D., McComb, C., & Goucher-Lambert, K. (2023). Conceptual design generation using large language models.

Madrid, J., Andersson, P., Söderberg, R., Wärmefjord, K., Kveselys, D., Lindkvist, L., & Lööf, J. (2021). Automated and interactive evaluation of welding producibility in an multidisciplinary design optimization environment for aircraft components. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *15*, 1–17. https://doi.org/10.1007/s12008-021-00775-z

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, *9*(1), 381–386.

Mai, G., Huang, W., Sun, J., Song, S., Mishra, D., Liu, N., Gao, S., Liu, T., Cong, G., Hu, Y., et al. (2023). On the opportunities and challenges of foundation models for geospatial artificial intelligence. *arXiv preprint arXiv:2304.06798*.

Makatura, L., Foshey, M., Wang, B., HähnLein, F., Ma, P., Deng, B., Tjandrasuwita, M., Spielberg, A., Owens, C. E., Chen, P. Y., Zhao, A., Zhu, A., Norton, W. J., Gu, E., Jacob, J., Li, Y., Schulz, A., & Matusik, W. (2023). How can large language models help humans in design and manufacturing?

Mankins, J. C., et al. (1995). Technology readiness levels. *White Paper, April*, *6*(1995), 1995.

Mavris, D. N., & Pinon, O. J. (2012). An overview of design challenges and methods in aerospace engineering. *Complex Systems Design & Management: Proceedings of the Second International Conference on Complex Systems Design & Management CSDM 2011*, 1–25.

Meng, D. (2022). Multidisciplinary design optimization of engineering systems under uncertainty: A review. *International Journal of Structural Integrity*, *13*(4), 565–593.

Mirchandani, S., & Adhikari, S. (2020). Aerospace cybersecurity threat vector assessment. https://doi.org/10.2514/6.2020-4116

Myers, D., Mohawesh, R., Chellaboina, V. I., Sathvik, A. L., Venkatesh, P., Ho, Y.-H., Henshaw, H., Alhawawreh, M., Berdik, D., & Jararweh, Y. (2023). Foundation and large language models: Fundamentals, challenges, opportunities, and social impacts. *Cluster Computing*, 1–26.

Nazarenko, A. A., Sarraipa, J., Camarinha-Matos, L. M., Grunewald, C., Dorchain, M., & Jardim-Goncalves, R. (2021). Analysis of relevant standards for industrial systems to support zero defects manufacturing process. *Journal of Industrial Information Integration*, *23*, 100214. https://doi.org/https://doi.org/10.1016/j.jii.2021.100214

Nelson, M. D., Goenner, B. L., & Gale, B. K. (2023). Utilizing chatgpt to assist cad design for microfluidic devices. *Lab on a Chip*, *23*, 3778–3784. https://doi.org/10.1039/d3lc00518f

Nichol, A., Jun, H., Dhariwal, P., Mishkin, P., & Chen, M. (2022). Point-e: A system for generating 3d point clouds from complex prompts. *arXiv preprint arXiv:2212.08751*.

Nightingale, P. (2000). The product-process-organization relationship in complex development projects. *Research Policy*, *29*(7-8), 913–930.

OpenAI. (2023). Chatgpt release notes [Accessed: 2024-05-03]. https://help.openai.com/en/articles/6825453-chatgpt-release-notes

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al. (2022). Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, *35*, 27730–27744.

Pan, Y., & Zhang, L. (2021). Roles of artificial intelligence in construction engineering and management: A critical review and future trends. *Automation in Construction*, *122*, 103517.

Paris, M. (2023). Chatgpt hits 100 million, microsoft unleashes ai bots, and catgpt goes viral. *Forbes*. https://www.forbes.com/sites/martineparis/2023/02/03/chatgpt-hits-100-million-microsoft-unleashes-ai-bots-and-catgpt-goes-viral/

Park, G.-J., Hwang, K.-H., Lee, T., & Lee, K. (2006). Robust design: An overview. *Aiaa Journal - AIAA J*, *44*, 181–191. https://doi.org/10.2514/1.13639

Patel, R., & Davidson, B. (2019). *Forskningsmetodikens grunder: Att planera, genomföra och rapportera en undersökning*. Studentlitteratur. https://books.google.se/books?id=MP7axgEACAAJ

Perez, E., Kiela, D., & Cho, K. (2021). True few-shot learning with language models.

Phadke, M. S. (1995). *Quality engineering using robust design*. Prentice Hall PTR.

Phadnis, V., Arshad, H., Wallace, D., & Olechowski, A. (2021). Are Two Heads Better Than One for Computer-Aided Design? *Journal of Mechanical Design*, *143*(7), 071401. https://doi.org/10.1115/1.4050734

Picard, C., Edwards, K. M., Doris, A. C., Man, B., Giannone, G., Alam, M. F., & Ahmed, F. (2023). From concept to manufacturing: Evaluating vision-language models for engineering design.

Pillai, P. P., Burnell, E., Wang, X., & Yang, M. C. (2020). Effects of Robust Convex Optimization on Early-Stage Design Space Exploratory Behavior. *Journal of Mechanical Design*, *142*(12), 121704. https://doi.org/10.1115/1.4048580

Powell, D., Magnanini, M. C., Colledani, M., & Myklebust, O. (2022). Advancing zero defect manufacturing: A state-of-the-art perspective and future research directions. *Computers in Industry*, *136*, 103596. https://doi.org/https://doi.org/10.1016/j.compind.2021.103596

Provost, F., & Fawcett, T. (2013a). Data science and its relationship to big data and data-driven decision making. *Big data*, *1*(1), 51–59.

Provost, F., & Fawcett, T. (2013b). Data science and its relationship to big data and data-driven decision making. *Big data*, *1*(1), 51–59.

Purdy, M., & Williams, A. M. (2023). How ai can help leaders make better decisions under pressure. *Harvard Business Review*. https://hbr.org/2023/10/how-ai-can-help-leaders-make-better-decisions-under-pressure

Qiu, Y., & Jin, Y. (2023). Document understanding-based design support: Application of language model for design knowledge extraction. *Journal of Mechanical Design*, *145*. https://doi.org/10.1115/1.4063161

Rabie Jaifer, Y. B., & Bhuiyan, N. (2021). New framework for effort and time drivers in aerospace product development projects. *Engineering Management Journal*, *33*(2), 76–95. https://doi.org/10.1080/10429247.2020.1772950

Raina, A., Cagan, J., & McComb, C. (2021). Design Strategy Network: A Deep Hierarchical Framework to Represent Generative Design Strategies in Complex Action Spaces. *Journal of Mechanical Design*, *144*(2), 021404. https://doi.org/10.1115/1.4052566

Rajagopal, N. K., Qureshi, N. I., Durga, S., Ramirez Asis, E. H., Huerta Soto, R. M., Gupta, S. K., & Deepak, S. (2022). Future of business culture: An artificial intelligence-driven digital framework for organization decision-making process. *Complexity*, *2022*, 1–14.

Rajamani, M., & Punna, E. (2020). Enhancement of design for manufacturing and assembly guidelines for effective application in aerospace part and process design. *SAE Technical Paper*. https://doi.org/10.4271/2020-01-6001

Ramos, A. L., Ferreira, J. V., & Barceló, J. (2011). Model-based systems engineering: An emerging approach for modern systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *42*(1), 101–111. https://doi.org/10.1109/TSMCC.2011.2106495

Raudberget, D. S. (2015). *Industrial application of set-based concurrent engineering – managing the design space by using platform system families* (Publication No. 3939) [Doctoral dissertation, Chalmers University of Technology]. Chalmers University of Technology. https://www.chalmers.se/

Regenwetter, L., Nobari, A. H., & Ahmed, F. (2022). Deep generative models in engineering design: A review [Accessed: 2023-06-14]. *Journal of Mechanical Design*, *144*, 071704. https://doi.org/10.1115/1.4053859

Regenwetter, L., Srivastava, A., Gutfreund, D., & Ahmed, F. (2023). Beyond statistical similarity: Rethinking metrics for deep generative models in engineering design.

Reitenbach, S., Siggel, M., & Bolemant, M. (2024). Enhanced workflow management using an artificial intelligence chatbot. https://doi.org/10.2514/6.2024-0917

Romera-Paredes, B., Barekatain, M., Novikov, A., Balog, M., Kumar, M. P., Dupont, E., Ruiz, F. J., Ellenberg, J. S., Wang, P., Fawzi, O., et al. (2024). Mathematical discoveries from program search with large language models. *Nature*, *625*(7995), 468–475.

Saadi, J. I., & Yang, M. C. (2023a). Generative design: Reframing the role of the designer in early-stage design process. *Journal of Mechanical Design*, *145*(4), 041411.

Saadi, J. I., & Yang, M. C. (2023b). Generative Design: Reframing the Role of the Designer in Early-Stage Design Process. *Journal of Mechanical Design*, *145*(4), 041411. https://doi.org/10.1115/1.4056799

Sametinger, J. (1997, January). *Software engineering with reusable components*. https://doi.org/10.1007/978-3-662-03345-6

Sanghi, A., Chu, H., Lambourne, J. G., Wang, Y., Cheng, C.-Y., Fumero, M., & Malekshan, K. R. (2022). Clip-forge: Towards zero-shot text-to-shape generation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18603–18613.

Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., Cancedda, N., & Scialom, T. (2023). Toolformer: Language models can teach themselves to use tools.

Shen, Y., Heacock, L., Elias, J., Hentel, K. D., Reig, B., Shih, G., & Moy, L. (2023). Chatgpt and other large language models are double-edged swords.

Siddharth, L., Blessing, L., & Luo, J. (2022). Natural language processing in-and-for design research. *Design Science*, *8*, e21. https://doi.org/10.1017/dsj.2022.16

Simon, H. A. (2019). *The sciences of the artificial, reissue of the third edition with a new introduction by john laird*. MIT press.

Sobek, D., Ward, A., & Liker, J. (1999). Toyota's principles of set-based concurrent engineering. *Sloan Management Review*, *40*.

Sobek, D. K., & Ward, A. C. (1996). Principles from toyota's set-based concurrent engineering process. *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, *97607*, V004T04A035.

Song, B., Miller, S., & Ahmed, F. (2023). Attention-Enhanced Multimodal Learning for Conceptual Design Evaluations. *Journal of Mechanical Design*, *145*(4), 041410. https://doi.org/10.1115/1.4056669

Song, B., Soria Zurita, N., Zhang, G., Stump, G., Balon, C., Miller, S., Yukish, M., Cagan, J., & McComb, C. (2020). Toward hybrid teams: A platform to understand human-computer collaboration during the design of complex engineered systems. *Proceedings of the Design Society: DESIGN Conference*, *1*, 1551–1560. https://doi.org/10.1017/dsd.2020.68

Song, B., Zhou, R., & Ahmed, F. (2023). Multi-modal machine learning in engineering design: A review and future directions.

Song, B., Zurita, N. S., Zhang, G., Stump, G., Balon, C., Miller, S. W., Yukish, M., Cagan, J., & McComb, C. (2020). Toward hybrid teams: A platform to understand human-computer collaboration during the design of complex engineered systems. *Proceedings of the design society: DESIGN conference*, *1*, 1551–1560.

Stella, F., Della Santina, C., & Hughes, J. (2023). How can llms transform the robotic design process? *Nature Machine Intelligence*, *5*. https://doi.org/10.1038/s42256-023-00669-7

Suh, N. P., & Suh, N. P. (2001). *Axiomatic design: Advances and applications* (Vol. 4). Oxford university press New York.

Svensson, R. B., Feldt, R., & Torkar, R. (2019). The unfulfilled potential of data-driven decision making in agile software development. *Agile Processes in Software*

Engineering and Extreme Programming: 20th International Conference, XP 2019, Montréal, QC, Canada, May 21–25, 2019, Proceedings 20, 69–85.

Taguchi, G., Chowdhury, S., & Taguchi, S. (2000). Robust engineering. *(No Title)*.

Tejasen, K. T., Raju, P., Ryan, W., Jordan, A., & Yang, R. (2022). Defining an initial classification scheme for non-deterministic ai technologies. *2022 Integrated Communication, Navigation and Surveillance Conference (ICNS)*, 1–10.

Thapa, S., & Adhikari, S. (2023). Chatgpt, bard, and large language models for biomedical research: Opportunities and pitfalls. *Annals of Biomedical Engineering*, *51*(12), 2647–2651.

Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. *Nature medicine*, *29*(8), 1930–1940.

Thornton, A. C. (2004). *Variation risk management—focusing quality improvements in product development and production*. Wiley.

Thornton, A., Donnelly, S., & Ertan, B. (2000). More than just robust design: Why product development organizations still contend with variation and its impact on quality. *Research in Engineering Design*, *12*, 127–143. https://doi.org/10.1007/s001630050028

Tikayat Ray, A., Pinon Fischer, O. J., White, R. T., Cole, B. F., & Mavris, D. N. (2024). Development of a language model for named-entity-recognition in aerospace requirements. *Journal of Aerospace Information Systems*, 1–11. https://doi.org/10.2514/1.I011251

Toche, B., Pellerin, R., & Fortin, C. (2020). Set-based design: A review and new directions. *Design Science*, *6*, e18. https://doi.org/10.1017/dsj.2020.16

Todorov, V. T., Rakov, D., & Bardenhagen, A. (2022). Enhancement opportunities for conceptual design in aerospace based on the advanced morphological approach. *Aerospace*, *9*(2), 78.

Trujillo, A., de Weck, O., & Madni, A. (2020). An mbse approach supporting technical inheritance and design reuse decisions.

Ulrich, K. T., & Eppinger, S. D. (2016). *Product design and development*. McGraw-hill.

Van Der Aalst, W., & van der Aalst, W. (2016). *Data science in action*. Springer.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Verganti, R., Vendraminelli, L., & Iansiti, M. (2020). Innovation and design in the age of artificial intelligence. *Journal of product innovation management*, *37*(3), 212–227.

Vlah, D., Žavbi, R., & Vukašinović, N. (2020). Evaluation of topology optimization and generative design tools as support for conceptual design. *Proceedings of the Design Society: DESIGN Conference*, *1*, 451–460. https://doi.org/10.1017/dsd.2020.165

VM, K., Warrier, H., Gupta, Y., et al. (2024). Fine tuning llm for enterprise: Practical guidelines and recommendations. *arXiv preprint arXiv:2404.10779*.

Walker, J., Koutsiana, E., Nwachukwu, M., Meroño Peñuela, A., & Simperl, E. (2024). The promise and challenge of large language models for knowledge engineering: Insights from a hackathon. *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3613905.3650844

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management.

Wan, P., & Leirmo, T. (2023). Human-centric zero-defect manufacturing: State-of-the-art review, perspectives, and challenges. *Computers in Industry*, *144*, 103792. https://doi.org/10.1016/j.compind.2022.103792

Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. (2019). Superglue: A stickier benchmark for general-purpose language understanding systems. *Advances in Neural Information Processing Systems*, *32*, 3261–3275. https://papers.nips.cc/paper%20files/paper/2019/hash/4496bf24afe7fab6f046bf4923da8de6-Abstract.html

Wang, B., Zuo, H., Cai, Z., Yin, Y., Childs, P., Sun, L., & Chen, L. (2023). A task-decomposed ai-aided approach for generative conceptual design. https://doi.org/10.1115/DETC2023-109087

Wang, M., Shi, D., Guan, N., Zhang, T., Wang, L., & Li, R. (2019). Unsupervised pedestrian trajectory prediction with graph neural networks, 832–839. https://doi.org/10.1109/ICTAI.2019.00119

Ward, A., Liker, J. K., Cristiano, J. J., & II, D. K. S. (1995). The second toyota paradox: How delaying decisions can make better cars faster. *MIT Sloan Management Review*. https://sloanreview.mit.edu/article/the-second-toyota-paradox-how-delaying-decisions-can-make-better-cars-faster/

Ward, A. C., & Sobek II, D. K. (2014). *Lean product and process development*. Lean Enterprise Institute.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2023). Chain-of-thought prompting elicits reasoning in large language models.

Wei, X., Juan-Li, W., & Xiao-Tun, W. (2010). Product concept generation and evaluation based on qfd and rough set theory. *2010 3rd International Conference on Information Management, Innovation Management and Industrial Engineering*, *2*, 244–247. https://doi.org/10.1109/ICIII.2010.223

Welfare, K. S., Hallowell, M. R., Shah, J. A., & Riek, L. D. (2019). Consider the human work experience when integrating robotics in the workplace. *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 75–84. https://doi.org/10.1109/HRI.2019.8673139

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382*.

Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: Humans and ai are joining forces. *Harvard Business Review*, *96*(4), 114–123.

Wu, C., Lin, W., Zhang, X., Zhang, Y., Xie, W., & Wang, Y. (2024). Pmc-llama: Toward building open-source language models for medicine. *Journal of the American Medical Informatics Association*, ocae045.

Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L., & Tang, Y. (2023). A brief overview of chatgpt: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, *10*(5), 1122–1136.

Xie, T., Wan, Y., Huang, W., Zhou, Y., Liu, Y., Linghu, Q., Wang, S., Kit, C., Grazian, C., Zhang, W., & Hoex, B. (2023). Large language models as master key: Unlocking the secrets of materials science with gpt.

Yairi, T., Takeishi, N., Oda, T., Nakajima, Y., Nishimura, N., & Takata, N. (2017). A data-driven health monitoring method for satellite housekeeping data based on probabilistic clustering and dimensionality reduction. *IEEE Transactions on Aerospace and Electronic Systems*, *PP*, 1–1. https://doi.org/10.1109/TAES.2017.2671247

Yang, C., Wang, X., Lu, Y., Liu, H., Le, Q. V., Zhou, D., & Chen, X. (2024). Large language models as optimizers.

Yilmaz, S., Daly, S. R., Seifert, C. M., & Gonzalez, R. (2016). Evidence-based design heuristics for idea generation. *Design Studies*, *46*, 95–124. https://doi.org/10.1016/j.destud.2016.05.001

Yu, D., Naik, S., Backurs, A., Gopi, S., Inan, H. A., Kamath, G., Kulkarni, J., Lee, Y. T., Manoel, A., Wutschitz, L., et al. (2021). Differentially private fine-tuning of language models. *arXiv preprint arXiv:2110.06500*.

Yuan, C., Marion, T., & Moghaddam, M. (2021). Leveraging End-User Data for Enhanced Design Concept Evaluation: A Multimodal Deep Regression Model. *Journal of Mechanical Design*, *144*(2), 021403. https://doi.org/10.1115/1.4052366

Yuan, C., Marion, T., & Moghaddam, M. (2022). Leveraging end-user data for enhanced design concept evaluation: A multimodal deep regression model. *Journal of Mechanical Design*, *144*(2), 021403.

Yüksel, N., Börklü, H. R., Sezer, H. K., & Canyurt, O. E. (2023). Review of artificial intelligence applications in engineering design perspective. *Engineering Applications of Artificial Intelligence*, *118*, 105697.

Zhang, G., Raina, A., Cagan, J., & Mccomb, C. (2021). A cautionary tale about the impact of ai on human design teams. *Design Studies*, *72*, 100990. https://doi.org/10.1016/j.destud.2021.100990

Zhou, J., Lu, T., Mishra, S., Brahma, S., Basu, S., Luan, Y., Zhou, D., & Hou, L. (2023). Instruction-following evaluation for large language models.

Zhou, J. (, Phadnis, V., & Olechowski, A. (2020). Analysis of Designer Emotions in Collaborative and Traditional Computer-Aided Design. *Journal of Mechanical Design*, *143*(2), 021401. https://doi.org/10.1115/1.4047685

Zhu, G.-N., Hu, J., & Ren, H. (2020). A fuzzy rough number-based ahp-topsis for design concept evaluation under uncertain environments. *Applied Soft Computing*, *91*, 106228.

Zhu, J., & Deshmukh, A. (2003). Application of bayesian decision networks to life cycle engineering in green design and manufacturing. *Engineering Applications of Artificial Intelligence*, *16*(2), 91–103.

Zhu, Q., & Luo, J. (2022). Generative pre-trained transformer for design concept generation: An exploration. *Proceedings of the Design Society*, *2*, 1825–1834. https://doi.org/10.1017/pds.2022.185

Zhu, Q., & Luo, J. (2023). Generative Transformers for Design Concept Generation. *Journal of Computing and Information Science in Engineering*, *23*(4), 041003. https://doi.org/10.1115/1.4056220

Zhu, Q., Zhang, X., & Luo, J. (2023). Biologically Inspired Design Concept Generation Using Generative Pre-Trained Transformers. *Journal of Mechanical Design*, *145*(4), 041409. https://doi.org/10.1115/1.4056598

+

A
Appendix

A.1 Internal Interview Questions - Product Development Processes

Introduction Questions:

1. Would it be okay to record this interview?
2. Would you like to be anonymous?
3. What is your role at GKN?
4. How long have you been working with your current title? Current position?
5. What tasks are included in your position?
6. How long have you worked at GKN in total?
7. What is your experience within product development, concept development and concept evaluation?
8. How long have you worked within that field?

OMS:

1. OMS divides product development into four major phases: Conceptual Design, Preliminary Design, Detailed Design, and Final Design. Does it feel obvious in which process each type of work belongs?
2. Do you feel comfortable working with concept development?
    a. What challenges do you see in working with concept development?
3. Hur gör vi om vi behöver omtag på ett koncept i "preliminary design" exempelvis, eller senare i processen? Vart vänder vi oss isf? Concept design?

PD methods:

1. What part of the product development process are you most familiar with?
2. What conventional methods are used for concept generation and evaluation at GKN?
3. Is the process always the same or do different programs have different conventional methods?
4.
5. What are the current praxis for decision making in these processes?
6. In your experience, which tasks are the most decision-heavy?
7. What do you think are the greatest obstacles in the concept generation and evaluation process at GKN?
8. In the product development process in general, in your experience, what are the greatest challenges you face?
9. What processes are regularly the most time consuming?
10. Is there a consensus around "best practice"?
11. What (data)programs are utilized for these processes?
12. Have you previously made big changes to your product development process? Did you learn something from that?
13. In the concept-choice-matrix, how do you decide upon the "weighting"? Is the weighting the same? If not, how long is the process of deciding upon weightings? Who is deciding?

Robustness & Zero defects:

1. In which of the phases mentioned in OMS are robustness and zero defects most considered in your experience? Does it clearly focus on one more than another?
2. How do you work with ZD in concept development?
3. How do you work with ZD in concept evaluation?

4. How do you work with robustness in concept development?
5. How do you work with robustness in concept evaluation?

Basis of decision-making:
1. How are decisions in the PD processes made?
2. What information lays the basis of decision making?
3. Could you rank the most influential factors that lay the basis of a major decision?
4. How much previous information from projects and lessons learned is utilized in new projects? If so, is this documented?
5. How much are “internal documents” used when making decisions?
6. Do you find it easy finding the information you need?
7. Is it common that you do not find the information you need?
8. How much time is spent on finding the specific resources you need to make a certain decision (in terms of documentation).

AI:

1. What conventional methods could you see being automated in the future?
2. What processes could you see would benefit from the incorporation of AI/ML?
3. If you were to rank which methods, processes, or decisions could be made by AI what would they be?
4. Are there any obvious activities that you feel could be more effective if you could use a tool like ChatGPT?

End of interview:

In the theme of the questions we have asked you, is there something major we have missed or something you would like to add?

A.2 Internal Interview Questions - AI and ML

Introduction Questions:
1. Would it be okay to record this interview?
2. Would you like to be anonymous?
3. What is your role at GKN?
4. How long have you been working with your current position?
5. What tasks are included in your position?
6. How long have you worked at GKN in total?
7. What is the short term plan for AI/ML at GKN?
8. How long have you worked within that field?

OMS:

1. Where can you find your ways of working in OMS?
2. Which “phases” within OMS do you primarily work within?
3. Does it feel obvious in which process -of you ways of working- each type of work belongs?

AI methodology:

1. In general, how much is AI/ML considered currently at GKN to your knowledge?
2. How far has GKN come in integrated data science in their ways of working?
3. What AI/ML initiatives do we currently have at GKN?
4. Are you aware of other projects in GKN where AI/ML is being used?
5. What is your area of expertise?

Data collection:
1. How is different data collected?
2. Manually or automatically?
3. How is data reported manually?

Opportunities:

1. In your specific area, what is considered to be the next big thing?
2. What does your work wish to achieve within the company in the long run?
3. What opportunities in general do you see with incorporating data science,

AI/ML:

1. What conventional methods could you see being automated in the future?
2. What processes could you see would benefit from the incorporation of AI/ML?
3. If you were to rank which methods, processes, or decisions could be made by AI what would they be?
4. What types of AI do you see would benefit these “insert specific problems”? (LLM, Regression, etc)
5. In the PD process, what types of models, programs, etc could you see generate (specific).

6. How can tools from AI and ML can be used to simplify concept generation, evaluation and to propose a robust design solution?

Integration:

1. What would you say are the greatest challenges that come with integrating AI/ML into these processes?
2. In contrast, what kind of processes could face the least amount of challenges when integrating?
3. Would the training of models need to be conducted in-house?
4. If the training would be outsourced on general aerospace data, not company specific data, could this still be beneficial or would it be much worse than a model trained on company specific data?
5. What would need to be in place at the Trollhättan site in order for an integration of this kind to work. (server centers, programs, PLM systems, OMS, etc).
6. Based on the previous questions, is there something obvious we have missed to consider when it comes to integration?
7. What would you say that AI is in terms of TRL at GKN?

AI implementation wishes from PD interviews, what are your thoughts about them?
How could these be implemented? What are the challenges?
Report writing
Search engine for documents quickly
Document sorting
Powerpoint generation
Patent search and guidance
Defect detection decision making,
Programing
Bought in LLM that is trained by experienced people (supervised learning)
CAD drawings
Ritningsgranskning
A "black box" where you input parameters at an early design phase and it generates a correct output after manufacturing
To generate aero-calculations. A system that computes 10 000 solutions of how Stator 1 could look based on a variation of parameters to gain a sweet spot.

End of interview:
In the theme of the questions we have asked you, is there something major we have missed or something you would like to add?
Is there a person you feel that could answer our questions well?

A.3 External Interview Questions

Interviewing questions - External benchmarking

Introduction
Would it be okay to record this interview?
Would you like to be anonymous?
What is your role at this company?
How long have you been working with your current title?
What tasks are included in your position?
How long have you worked at Company in total?
What is your experience within product development, concept development and concept evaluation?
How long have you worked within that field?

Are you working with AI?
If yes:
How is AI currently being integrated into your PD processes?
Is AI used for concept generation, concept evaluation? How is it used?
What conventional methods have been integrated with AI?

What are key drivers for adopting AI into your operation?

What metrics do you use to measure the success and impact of AI integration?
- Can you provide tangible benefits or improvements resulting from AI implementation?

What challenges have you faced when integrating AI?
How did you overcome these challenges?

What skills or expertise are required for managing and deploying AI technologies within your organization?
- How can you ensure your workforce is equipped with the necessary skills for AI integration?

How can you ensure the quality, reliability of data used for AI applications?

How is AI taken into consideration in regards to, safety, security and handling classified information?
- Do you store data in cloudservices?

Are models outsourced or developed in-house?
How is the infrastructure currently at [your company]?

What are future plans for expanding the use of AI within your company's operations?

If no:
Why not?
What factors have made you reluctant towards AI integration?

AI in general:
In general, how much is AI/ML considered currently at your company to your knowledge?
How far has your company come in integrated data science in their ways of working?
What conventional methods could you see being automated in the future?
What processes could you see would benefit from the incorporation of AI/ML?
If you were to rank which methods, processes, or decisions could be made by AI what would they be?
How has AI improved your ways of working? How much has efficiency increased?
Do you have insight into how other companies are working with AI?

End of interview:
In the theme of the questions we have asked you, is there something major we have missed or something you would like to add?

Thank you

A.4 Technology Readiness Level

The Technology Readiness Level (TRL) is a crucial aspect when discussing the implementation of new initiatives within a corporation, particularly in the aerospace industry.

TRL was first introduced by Mankins et al. (1995) in 1995 and is a systematic measurement system that explains the maturity of a specific technology and facilitates the comparison of maturity between different types of technology. The TRL approach has been employed in NASA's space technology planning for many years and was incorporated into the NASA Management Instruction (NMI 7100) in the 1990s, addressing integrated technology planning at NASA (Mankins et al., 1995). Essentially, TRL helps in understanding how close a particular technology is to being ready for practical use or implementation.
TRL ranges from levels 1-9, and each level will be explained in the table below:

- **TRL 1:** ***Basic principles observed and reported***
  At level one, research begins and gets translated into applicable research and development (R&D). Examples of studies are tensile strength of a material.
  Cost to Achieve: Very Low 'Unique' Cost
- **TRL 2:** ***Technology concept and/or application formulated***
  When information of physical principles have been gathered, practical applications of these characteristics can be identified. An example is the observation of high critical temperature (Htc) superconductivity can be defined. At level two, the application of the concept is still speculative.
  Cost to Achieve: Very Low 'Unique' Cost
- **TRL 3:** ***Analytical and experimental critical function and/or characteristic proof-of concept***
  Active R&D begins at this stage, involving analytical studies to position the technology within suitable contexts. Laboratory studies are conducted to validate the analytical predictions formed. These studies aim to construct 'proof-of-concept' validation for the concepts created at TRL 2.
  To illustrate, consider a concept for High Energy Density Matter (HEDM) propulsion, which may rely on slush or super-cooled hydrogen as a propellant. TRL 3 is achieved when the laboratory attains the concept-enabling phase/temperature/pressure for the fluid.
  Cost to Achieve: Low 'Unique' Cost (technology-specific).
- **TRL 4:** ***Component and/or breadboard validation in laboratory environment***
  After successfully proving the concept, it's essential to integrate basic technological elements to ensure that the components work together to achieve concept-enabling performance levels for a component or breadboard. This validation must be designed to support the earlier formulated concept and align with the requirements of potential system applications. The validation is relatively low-fidelity compared to the eventual system.
  For example, at TRL 4, testing a new 'fuzzy logic' approach for avionics could involve trying out the algorithms in a controlled environment like a lab. This

includes using simulated vehicle inputs and demonstrating it with a combination of computer-based and bench-top components (e.g., fiber optic gyros).

Cost to Achieve: Low-to-moderate 'Unique' Cost (investment will be technology specific, but probably several factors greater than investment required for TRL 3)

- **TRL 5:** ***Component and/or breadboard validation in relevant environment***

  At this stage, the testing of the component requires a significant increase in accuracy. Basic technological elements need to be integrated with reasonably realistic supporting elements. This allows testing the total applications (component-level, sub-system level, or system-level) in a 'simulated' or somewhat realistic environment. The demonstration may involve one or several new technologies.

  For instance, a new type of solar photovoltaic material, promising higher efficiencies, would be used in an actual fabricated solar array. This solar array is integrated with power supplies, supporting structure, etc., and tested in a thermal vacuum chamber with solar simulation capability.

  Cost to Achieve: Moderate 'Unique' Cost (investment cost will be technology dependent, but likely to be several factors greater that cost to achieve TRL 4)

- **TRL 6:** ***System/subsystem model or prototype demonstration in a relevant environment (ground or space)***

  After TRL 5, there's a significant leap in technology demonstration accuracy at TRL 6. Here, a model or prototype system is tested in a relevant environment. If space is the relevant environment, the model or prototype must be demonstrated in space to be considered a true TRL 6. Not all technologies undergo a TRL 6 demonstration; at this stage, the main focus is on ensuring management confidence rather than meeting R&D requirements.

  The demonstration may represent an actual system application or be similar to the planned application, using the same technologies. At TRL 6, several to many new technologies may be integrated into the demonstration. For example, an innovative component on a Space Shuttle could be demonstrated to TRL 6 by flying a working, sub-scale (but scalable) model on a Space Shuttle. In this case, space is the relevant environment because microgravity, vacuum, and thermal effects will determine the success or failure of the system, and the only way to validate the technology is in space.

  Cost to Achieve: Technology and demonstration specific; a fraction of TRL 7 if on ground; nearly the same if space is required.

- **TRL 7:** ***System prototype demonstration in a space environment***

  TRL 7 is a crucial advancement beyond TRL 6, necessitating a prototype demonstration of the actual system in a space environment. While not always implemented in the past, in this case, the prototype should closely match the scale of the planned operational system, and the demonstration must occur in space. Not all technologies and systems will reach this level, and TRL 7 is typically reserved for mission-critical and relatively high-risk technology or subsystem applications. For instance, the Mars Pathfinder Rover serves as a

TRL 7 technology demonstration for future Mars micro-rovers based on its system design.
Cost to Achieve: Technology and demonstration specific, but a significant fraction of the cost of TRL 8

- **TRL 8:** ***Actual system completed and "flight qualified" through test and demonstration (ground or space)***
By definition, all technologies incorporated into operational systems undergo TRL 8. In most instances, this level marks the conclusion of authentic 'system development'. For instance, successfully loading and testing a new control algorithm into the onboard computer on the Hubble Space Telescope while in orbit.
Cost to Achieve: Mission specific; typically highest unique cost for a new technology
- **TRL 9:** ***Actual system "flight proven" through successful mission operations***
All technologies implemented in actual systems undergo TRL 9. In most cases, this represents the final stages of addressing the last 'bug fixing' aspects in true 'system development.' For example, making small fixes or changes to address problems discovered after launch. This could involve integrating new technology into an existing system, such as incorporating a new artificial intelligence tool into operational mission control.
Cost to Achieve: Mission Specific; less than cost of TRL 8.

A.4.1 Design Space Exploration

Design Space Exploration (DSE), as defined in (Kang et al., 2011), is the process of evaluating various design options before their actual implementation. DSE is crucial for many engineering activities such as rapid prototyping, optimization, and system integration due to its capability to manipulate a wide range of possible solutions. DSE is used at GKN Aerospace Sweden currently, and is an available ML tool. A major challenge in DSE is the sheer size of the design space, which could include millions or billions of alternatives, making a complete exploration impractical.
Given the cost constraints, not all possible solutions can be practically tested. It is critical, therefore, for the DSE process to be as efficient as possible. (Kang et al., 2011) introduces an effective DSE framework composed of three fundamental components:

- **Representation:** Essential for automated analysis, this involves a formal representation of the design space that captures complex constraints, including arithmetic, Boolean, and data type constraints.
- **Analysis:** This requires automated tools for identifying and validating potential solutions against these constraints, while also efficiently managing computational costs.
- **Exploration Method:** Post-optimization, this component aids in the exploration of unique design candidates, steering clear of arbitrary enumeration of possibilities.

According to the authors, a solution is considered noteworthy if it differs significantly

from previously explored solutions, based on a user-defined criterion of equivalence. Two solutions are equivalent if they have isomorphic mathematical representations. The FORMULA framework, detailed in (Jackson & Sztipanovits, 2009; Jackson et al., 2009), specifies domain-specific languages (DSLs) for modeling design spaces. DSLs are instrumental in formally representing and enforcing constraints within these spaces, making them ideal for handling complex design configurations. FORMULA leverages DSL composition, facilitated by the Z3 SMT solver (de Moura & Bjørner, 2008), to simplify complex design areas into manageable components.
Representation: In FORMULA, a domain block contains the DSL's data types and constraints (Jackson et al., 2010), including various types like simple sorts, record constructors, and unions.
Analysis: FORMULA enhances automated model validation against domain constraints using constraint logic programming (CLP), which is a simpler alternative to the more complex object constraint language (OCL) (Kang et al., 2011).
Exploration Method: The framework employs a "conforms" query to ensure that models meet all specified constraints, defining the design spaces accordingly.
Modularity and Composition: FORMULA supports modular DSL design that allows for the expansion and combination of various domains, thus facilitating consistent enforcement of constraints.

In essence, FORMULA utilizes formal abstractions and DSLs, enhanced by CLP and modular composition, to manage and navigate complex design spaces effectively.
Solving for Instances: FORMULA converts design specifications into queries for the Z3 SMT solver. Each domain query (D.q) is transformed into a first-order logic formula ([X]), representing models as sets of records that align with the satisfying instances of the formula.

Z3 combines decision procedures across different theories and SAT-based techniques to search for solutions efficiently. The process handles complex queries, such as aggregating the capacities of incoming channels to a device, involving computations with term algebras and linear arithmetic.

FORMULA's translation of models to SMT queries is intricate as it predominantly supports existential logic. By symbolically executing the specifications over symbolic inputs, FORMULA generates a quantifier-free version of the logic, representing various potential configurations within the design space.
Design Space Exploration Method After symbolic execution, FORMULA identifies design space elements via the Z3 solver. To effectively find diverse solutions, a technique that groups similar solutions through isomorphisms over algebraic data types is employed.
Projection-Based Equivalence Partitioning introduces a function known as term homomorphism, which alters constants within records while preserving their structure. This function groups isomorphic models into a single equivalence class, significantly streamlining the exploration process.
Exploration Algorithms: FORMULA utilizes algorithms like ExploreII to optimize the search process, preventing revisits to non-productive areas by learning from past explorations.

This thorough exploration is essential for efficiently navigating through complex design spaces, especially those with stringent constraints, ensuring a broader array of potential solutions is considered.

Evaluation

The performance of two algorithms, Explore and ExploreII, is assessed based on their effectiveness in distributing solutions throughout a design space. We utilize a compact generator set to map the entire space, modifying the model to simplify the number of equivalence classes. The configuration involves defined tasks, devices, conflicts, channels, and bindings, shaping the potential equivalence classes. During testing, each algorithm was restricted to 100 invocations of the SMT solver, with graphical representations illustrating the outcomes. These visualizations depict the explored regions and the adequacy of solutions, using colors to indicate successful and unsuccessful areas.

Randomization

The performance of the algorithms is evaluated in various design spaces, focusing on their ability to sample equivalence classes uniformly without bias. Challenges included balancing random sampling against the computational expense of generating non-isomorphic samples. A refined sampling algorithm significantly reduced bias, enhancing cost-efficiency in design spaces with complex symmetries.

Highly Constrained Design Spaces

In tightly constrained spaces, clustering of non-isomorphic solutions challenges effective random sampling. A probability parameter to balance randomly selected samples and those driven by the solver is introduced. Experimental results demonstrate how different pgen settings impact the distribution and discovery of solutions, highlighting the trade-offs between diversity and efficiency in solution discovery.

DEPARTMENT OF INDUSTRIAL AND MATERIAL SCIENCE
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY