Thesis for The Degree of Doctor of Philosophy

Data Engineering for Data-Driven Design

Features, labels, and datasets for accelerated design evaluation

Mohammad Arjomandi Rad

Department of Industrial and Material Science
Chalmers University of Technology
Gothenburg, Sweden, 2025

Data Engineering for Data-Driven Design
Features, labels, and datasets for accelerated design evaluation
Mohammad Arjomandi Rad

ISBN 978-91-8103-221-5
ISSN 0346-718X
Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5679.

Department of Industrial and Material Science
Division of Product Development
System Engineering Research Group
Chalmers University of Technology
SE-412 96 Göteborg,
Sweden
Phone: +46(0)79 033 5053

Printed by Chalmers Digitaltryck,
Gothenburg, Sweden 2025.

دلا دیدی که خورشید از شب سرد

چو آتش سر ز خاکستر بر آورد

نگر تا این شب خونین سحر کرد

چه خنجرها که از دلها گذر کرد

(سایه)

"Did you see, my love, the sun after the cold night,

reborn like fire from ashes beneath the dust?

Look, as this bloody night has turned to dawn,

what knives it has passed through hearts."

Sayeh (H. E.)

Sammanfattning

Produktutvecklingsprocessen är idag starkt beroende av modellering och simulering för att utvärdera hur väl en design uppfyller uppställda krav. Iterativa och simuleringsdrivna processer har minskat kostnaderna för fysiska tester, men inte den ledtid som krävs för att genomföra dessa utvärderingar. Ett exempel är utvecklingen av krockkuddar, där hundratals manuella digitala prototypcykler genomförs innan fysiska tester påbörjas. För att hantera detta krävs effektivare metoder för designutvärdering.

Datadriven design har visat stor potential för att påskynda designcykler, men begränsas av tillgången på data. Storskaliga simuleringar som krävs för att generera labels till dataset är kostsamma, vilket ofta leder till otillräcklig utforskning av designutrymmet. Dessutom skapas dataset individuellt och utan systemperspektiv, vilket begränsar möjligheterna till framtida designändringar. När dataset väl används tvingas designers ofta att följa etablerade konventioner, vilket kan leda till mindre träffsäkra analyser.

Denna avhandling undersöker data engineering som en avgörande drivkraft i datadriven design för att stödja designutvärdering. Målet är att identifiera och mildra centrala problem kopplade till användningen av AI för digital utvärdering, genom att fokusera på nya sätt att konstruera, extrahera och organisera dataset. Arbetet syftar till att effektivisera digitala verifieringscykler och minska ledtiden i ingenjörsdriven produktutveckling.

Att extrahera features från alternativa geometriska representationer, såsom den mediala axeln, föreslås som en metod för att minska beroendet av CAD-parametrisering i prediktiva modeller. Dessa features visade sig vara mer användbara än CAD-parametrar för prediktiva uppgifter. Därför introduceras konceptet ” sovande parametrar” som dataset-features med potential att öka kunskapsinnehåll och överförbarhet i produktstrukturen. Bildregression baserad på designskärmdumpar föreslås som en alternativ metod för att bygga prediktionsmodeller i designutvärdering. Det storskaliga dataset som krävs för detta genereras med hjälp av dynamisk relaxation. För att möjliggöra analys av designförändringar används sovande parametrar i ett ramverk kallat Produktdatasetplattformen”, där komponentdata nyttjas för systemnivåutvärdering. Det visades att detta ramverk möjliggör snabb utforskning av nya designkonfigurationer och påskyndar valideringsprocessen.

Resultaten bidrar till grunden för CAD-CAE-integration genom att erbjuda ett ramverk för designutvärdering. De föreslagna stöden möjliggör prediktiva insikter som kan stödja kopplingen mellan funktion och form samt förbättra utvärderingen efter förändring.

Avslutningsvis visar resultaten en väg mot en datadriven metod för designutvärdering som kan påskynda iterationer och minska ledtiden i ingenjörsdriven produktutveckling.

Data Engineering for Data-Driven Design
Features, labels, and datasets for accelerated design evaluation

Mohammad Arjomandi Rad

Department of Industrial and Material Science
Chalmers University of Technology

Abstract

Today, the design process of products heavily uses modeling and simulation to assess how well they fulfill the requirements. Iterative and simulation-driven design processes have reduced testing costs but not the lead time associated with them. One example is the airbag design process, which involves hundreds of manual and digital prototyping loops before physical tests are conducted. Addressing these issues requires more efficient design evaluation. Data-driven design has shown significant potential in accelerating design cycles; however, it is also hindered by its reliance on the availability of data. The product design process establishes parameterization conventions that must be followed when analyzing. Running large-scale simulations to generate labels in design datasets can be costly, often leading to underexplored design spaces. The datasets generated are created individually without a system perspective, which often restricts future design changes. This thesis investigates data engineering as a critical driver of data-driven design to support design evaluation. The goal is to identify and mitigate the main problems associated with data generation by exploring new methods for constructing, extracting, and organizing features and labels. The thesis aims to streamline digital verification cycles and reduce engineering design lead time. Extracting features from alternative geometric representations, such as the medial axis, is proposed to reduce the reliance of data-driven evaluation methods on model parameterization. Using these parameters is shown to be superior to CAD parameterization for predictive tasks. Therefore, the concept of *sleeping parameters* is suggested as a potentially impactful feature in the dataset, which enhances knowledge encapsulation and transfer within the product structure. Image regression is proposed using design screenshots as an alternative method for building prediction models in design evaluation. The necessary large-scale dataset for this task is created through dynamic relaxation. To address design change analysis in design evaluation, the sleeping parameters concept is used in a framework called *Product Dataset Platform*, where component-level data is leveraged for system-level evaluations. It was shown that these solutions, individually, enable the rapid exploration of novel design configurations and accelerate the validation process. The results contribute to the foundation for CAD-CAE integration by providing a design evaluation framework. The suggested supports enable predictive capacity that can help map function to form and allow more efficient evaluation after the design change. The findings show a way to a data-driven design evaluation method that accelerates design iterations and reduces engineering design lead time.

Keywords Design for Data, Data Engineering, Data-Driven Design, Design Automation, Design Evaluation, CAD/CAE Integration

List of publications

i **Correlation-based feature extraction from computer-aided design, case study on curtain airbags design**
Arjomandi Rad, M., Salomonsson, K., Cenanovic, M., Balague, H., Raudberget, D., Stolt, R.
Computers in Industry, 2022, 22(24), 138, 103634.
(https://doi.org/10.1016/j.compind.2022.103634)

ii **Image regression-based digital qualification for simulation-driven design processes, case study on curtain airbag**
Arjomandi Rad, M., Cenanovic, M., Salomonsson, K.
Journal of Engineering Design , 2023,
https://doi.org/10.1080/09544828.2022.2164440

iii **Datasets in design research: needs and challenges and the role of AI and GPT in filling the gaps**
Arjomandi Rad, M., Hajali, T., Martinsson Bonde, J., Panarotto, M., Wärmefjord, K., Malmqvist, J., Isaksson, O.
Proceedings of the Design Society , 2024,
https://doi.org/10.1017/pds.2024.194

iv **Towards Using Functional Decomposition and Ensembles of Surrogate Models for Technology Selection in System Level Design**
Arjomandi Rad, M., Panarotto, M., Martinsson Bonde, J., Malmqvist, J., Wärmefjord, K., Isaksson, O.
Proceedings of NordDesign , 2024,
https://doi.org/10.35199/NORDDESIGN2024.45

v **Beyond traditional computer-aided design parameterization, feature engineering for improved surrogate modeling in engineering Design**
Arjomandi Rad, M., Panarotto, M., Isaksson, O.
Journal of Computer Aided Design and Applications , 2024,
https://doi.org/10.14733/cadaps.2025.536-554

vi **Product Dataset Platform: System Level Design Performance Evaluation using Feature Engineering and Functional Decomposition, Application case on a Car Front Structure**
Arjomandi Rad, M., Martinsson Bonde, J., Isaksson, O., Panarotto, M., Wärmefjord, K., Malmqvist, J.
Journal of Mechanical Design, ASME, Accepted (Publication id: MD-25-1012)
https://doi.org/xxx/xxx

Acknowledgements

This thesis is the outcome of research conducted between 2019 and 2025. I gratefully acknowledge the financial support from the Swedish Knowledge Foundation (KK-Stiftelsen) for the Butterfly Effect project and the European Union's Horizon 2020 research and innovation program for the CHEOPS project.

I would like to express my sincere gratitude to Chalmers University of Technology, where this work was completed. I am especially thankful to my supervisors, Professor Ola Isaksson, who not only welcomed me into his research group but also supported me through the final and most demanding part of the journey. My co-supervisors, Associate Professor Massimo Panarotto, thank you for your trust and continued support. Kristina Wärmefjord and Professor Johan Malmqvist thank you for your thoughtful guidance and encouragement throughout this process.

This journey began at Jönköping University, and I remain grateful for the support I received there. Special thanks go to Dr. Joel Johansson, Professor Fredrik Elgh, Dr. Roland Stolt, and Dr. Dag Raudberget for their early mentorship and encouragement. I also appreciate the valuable input from Professor Kent Salomonsson from the University of Skövde, Assistant Professor Mirza Cenanovic, and Dr. Tim Heikkinen, whose engagement enriched my work.

Many others have supported me along the way, including colleagues, mentors, and friends in the Departments of Product Development, Production and Design (PPD), and Industrial and Materials Science (IMS), as well as at Autoliv AB. I am especially grateful to Henrik Balagué, Mika Himiläinen, and Christian Svensson for believing in me and my methods and allowing me to continue working with these topics beyond the PhD.

Finally, my heartfelt appreciation goes to my wife, Maryam, for her unconditional support. I also thank my mother, my sister, and her family for their enduring patience and understanding through all these years apart.

List of Figures

List of Tables

Acronyms

AI	Artificial Intelligence	2
ML	Machine Learning	23
CNN	Convolutional Neural Network	29
ANN	Artificial Neural Network	28
CAE	Computer-Aided Engineering	6
CAD	Computer-Aided Design	1
CFD	Computational Fluid Dynamics	11
FEA	Finite Element Analysis	11
FE	Finite Element	62
HEB	High Dimensional, Expensive, and Black-box	28
SDD	Simulation-Driven Design	19
KBE	Knowledge-Based Engineering	20
DA	Design Automation	20
DDD	Data-Driven Design	29
PD	Product Development	5
SE	System Engineering	18
PDM	Product Data Management	19
DE	Data Engineering	25
DfAI	Design for Artificial Intelligence	31
DRM	Design Research Methodology	43
TWB	Thin-Walled Beams	54
FMS	Fluid Management System	43

Contents

Chapter 1

Introduction

Products are evolving in response to rapid technological advancements, shifting societal expectations, and increasingly competitive markets. The rise of high-level complex products characterized by shorter development cycles, more customization, and higher performance demands requires engineering methods to evolve and effectively use more advanced techniques (Isaksson and Eckert, 2020; Jiao et al., 2021). This ongoing evolution underscores the importance of research focused on adapting existing design methods to meet emerging challenges.

Models have historically served as critical tools for conveying conceptual ideas from the "designer" to the "constructor" varying widely from written texts and images to detailed drawings (Jordan and Hoelscher, 1935). The onset of the *information age* (see Figure 1.1) at the beginning of the 1980s brought 3D Computer-Aided Design (CAD) systems and shifted the focus to complete volume models rather than 2D representations, and transformed design tools and environments (Smit and Bronsvoort, 2009). This evolution, enabled by digitalization, has progressively reshaped development environments and design tools and even transformed the interaction between designers and constructors.

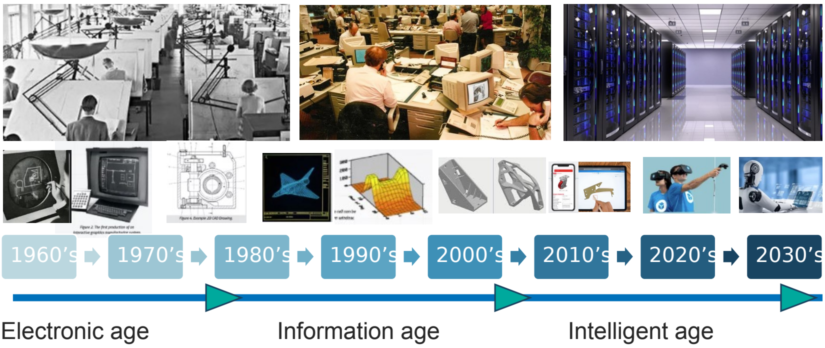


Figure 1.1: Modeling technologies through three main epochs (Wang et al., 2022)

Engineering drawing, a predecessor to modern 3D modeling, dates back to ancient times, with fundamental techniques evolving significantly through the Renaissance (Lagenbach et al., 2015). By the 19th century, engineering drawing had become a critical tool, particularly with the increased documentation and standardization efforts. In the early 20th century, it was established as a universal engineering language (Madsen, 2002). The 1960s marked a turning point with the advent of computer programs. In a way, an emerging technology (specifically computers) contributed to advances in the field of 2D hand drawing and additionally provided an alternative form of design assistance, thereby initiating the era of CAD. (Reffold, 1998). However, the first generation of computer-aided tools was used to automate the paper-based design processes and served as an expensive replacement for the drawing board (Bossak, 1998). Today, hand drawings remain indispensable for brainstorming and quick communication, but have lost the importance that they once held in engineering education (Hilton et al., 2016).

The transformation that occurred with the emergence of CAD is comparable to the situation we face now. The slow and extended transition from manual drawing to computer modeling reflects a trend in technological adoption within engineering. With the advent of the *intelligent age*, Artificial Intelligence (AI)-based technologies are now redefining the "aid" in computer-aided design (Debney, 2020). Most likely, engineering design methods are poised for yet another transformation, potentially redefining the concept design evaluation process and the role of computer-aided design in product realization (Regenwetter et al., 2022; Cooper, 2023). The transition from traditional CAD to intelligent methods seems inevitable, although history reminds us that a complete transition can take several decades. Despite this, design engineering must be prepared with methods to accelerate the integration and aid this transition.

Engineering design has an undeniable role in the current industrial revolution (Tatipala et al., 2021) and will therefore be affected by any future industrial changes as well. After the third industrial revolution, which aimed to initiate the use of computers for digitalization and automation, the latest and fourth industrial revolution, so-called Industry 4.0, is marked by large-scale and integrated digitalized assets such as AI, Big Data, Cloud Computing, Internet of Things, etc (Keleko et al., 2022). Given this context, this thesis becomes highly relevant, as it aims to develop and evaluate methods that effectively integrate advanced digital tools and AI technologies into contemporary engineering design practices, thus directly contributing to the successful realization and acceleration of Industry 4.0 objectives.

As AI models continue to outperform human capabilities on various benchmarks between 1998 and 2024 (Bengio et al., 2024), it underscores the potential for these technologies to alter engineering practices in coding, mathematics, and specific tasks 1.2. The figure shows the performance of AI models on various benchmarks, including Computer vision (MNIST, ImageNet), Speech recognition (Switchboard), Natural language understanding (SQuAD 1.1, MMLU, GLUE), General language model evaluation (MMLU, Big-Bench, and GPQA), and Mathematical reasoning (MATH). It is reasonable to anticipate that it is only a matter of time before a more diverse range of benchmarks (including design science-related ones) are added to this list of benchmarks in which AI models do a far better job than humans.

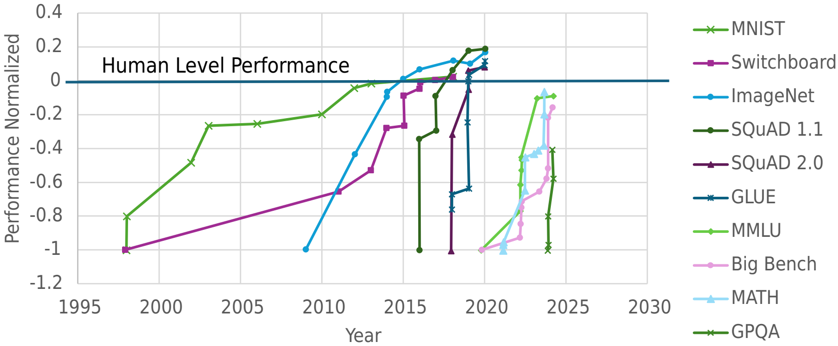


Figure 1.2: Performance of AI models on various benchmarks (Bengio et al., 2024)

Neural scaling laws, which show how AI models improve with bigger sizes, more data, and more computing power, are likely to face economic limits in the future, but have held on the technical side until today (Grace et al., 2018; Henighan et al., 2020). This means AI has gotten smarter by just making models bigger and using more data, which has been a prominent driver of progress. This rapid advancement of AI technology creates new opportunities and challenges for engineering design, especially regarding the role of modeling and simulation (Figoli et al., 2025). This is because models and simulations serve as the primary methods for communicating conceptual ideas from designers to manufacturers.

Some researchers (Will, 1991) suggest that a major decision on the cost (85%) of a product is incurred directly due to decisions made before the product design is released to manufacturing. This figure is criticized by others (Ulrich and Pearson, 1993), and it seems reasonable to assume that the figure will vary based on different factors, one of which is the degree of product complexity. It can be argued that design is a key and influential factor in products, especially those that need to meet today's complex requirements. Therefore, high-level technical products like jet engines and airbags, which require putting more effort into design, have to leverage the potential of AI methods in their development process to stay competitive. Researchers in this field should develop and adopt methods like modeling and simulation to facilitate AI transformation. Given the critical role of modeling and simulation in generating the data necessary for various AI-related activities, studies like this thesis, which serve as integration points for two developed disciplines, seem highly important.

1.1 Background

The goal of product development is to achieve economic success, satisfy a need, or accomplish both (Pahl and Beitz, 2013), which is achieved by iterating the design process to meet various dynamic quality, cost, and sustainability requirements. Most of a product's economic, environmental, and societal impacts are related to the decisions taken during the design process (He et al., 2020). Design iteration is the repetition of decision-making in the product design process, where these decisions share common characteristics (Smith and Eppinger, 1997). This is why design iterations are particularly suitable cases for AI-like adaptable methods that learn from one iteration to the next (Zhou et al., 2023).

Iterations in product development can also be seen as 'bundles of problem-solving cycles' (Fujimoto, 2000). The focus of Decision-makers is on resolving problems by compressing, simplifying, front-loading, or overlapping iteration cycles, all to achieve a shorter lead time. The rule of thumb is that problems are easier to address in the early phases, also known as the *design paradox* in the literature (Ullman, 1992). It takes more cost and time to solve problems later in the development projects, while the lack of knowledge results in lower fidelity in the early simulation models (Clark, 1991). Osborne (Osborne, 1993) reported that iterations accounted for 13 to 70% of the total development effort for nine projects.

As the literature emphasizes the difference between different lead times (Tiedemann et al., 2020), it is essential to note the definitions used in this thesis, see Figure 1.3. The *Total Lead time* for a product realization process, which is also called time to market, includes *Product Development lead time* that includes all activities roughly from the identification of need to start of manufacturing, which is a critical metric in a company's ability to respond to market demands and technological advancements (Clark and Fujimoto, 1989; Petersen, 2010).

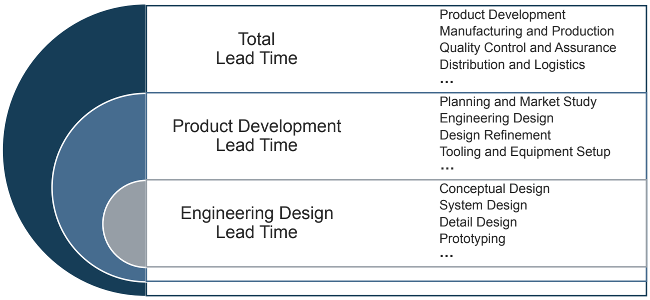


Figure 1.3: The Engineering Design, Product Development, and Total Lead Time

Shorter lead times in design highlight the need for quick feedback, as such an approach improves decision-making during the design process. However, engineering design lead time when it comes to acceleration is often hindered by the extensive computational demands of conventional simulation methods, especially when exploring diverse design alternatives (Edelen et al., 2020). Simulation-driven processes, although essential for design evaluation, are typically constrained by their sequential nature. A failure at one stage requires restarting subsequent analyses, if not the whole design, which increases delays (Karniel and Reich, 2009).

These challenges underscore the potential benefits of integrating AI methods, such as real-time surrogate modeling and prediction tools, which can rapidly qualify designs and significantly reduce iterative times. Thus, AI can be particularly beneficial in streamlining and optimizing design iterations (Zhou et al., 2023; Yüksel et al., 2023).

Despite this importance, adopting AI methods in product design has not been as successful compared to other sectors like manufacturing (Chui et al., 2023; Perrault and Clark, 2024). Comparing the success of AI adoption in the "Manufacturing" sector to "R&D product/service development" in the McKinsey survey, Figure 1.4 shows 55% of respondents reported a reduction in cost with AI adoption in manufacturing but only 31% in Product Development (PD). The revenue increase with AI adoption is also reported to be slightly higher in Manufacturing than in PD.

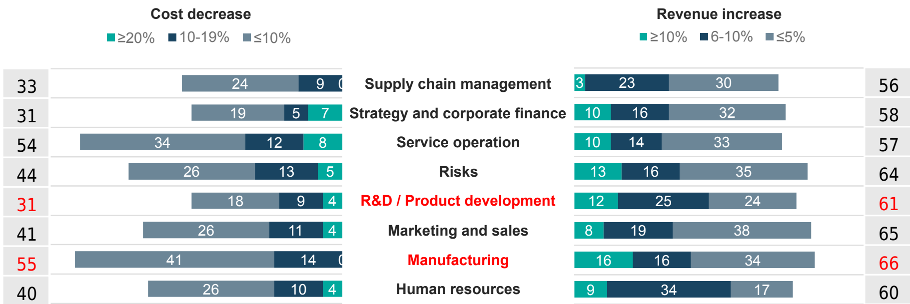


Figure 1.4: The benefits from AI adoption in different sectors (Chui et al., 2023)

This distinction can be elucidated by the characteristics of data within the manufacturing sector, which are inherently more favorable for the application of artificial intelligence. Such data is quantitative, standardized, and frequently updated, allowing for the effective monitoring and optimization of production processes (Arinez et al., 2020). Such data richness enables AI algorithms to identify anomalies, optimize production schedules, and minimize waste. Conversely, product design involves a higher degree of synthesis and qualitative data, which are more challenging for AI to interpret and utilize effectively. Furthermore, cost drivers in manufacturing, such as scrap, inventory management, and overproduction, are tangible and accessible metrics. In product development, however, the primary cost drivers are often related to man-hours for product development lead times. These elements are less quantifiable and involve human factors that AI has yet to effectively replicate or augment.

While product systems must be created, the processes for manufacturing complex products must be discovered and induced (Browning et al., 2006). This means product development uses models to innovate and conceptualize, which involves a greater degree of uncertainty and complexity. Meanwhile, manufacturing uses models to refine and control, with a focus on predictability and minimizing variability. These fundamental differences underscore why AI might be more readily applicable and effective in the manufacturing sector compared to PD, where the creative and exploratory nature of the work presents unique challenges.

Since modeling and simulation are one of the primary data generation sources in product development, addressing the barriers hindering AI adoption in this area is crucial (Diallo et al., 2017). These include the challenges of integrating AI into the iterative cycles of the development process. Additionally, the data derived from models and simulations must be effectively captured, standardized, and made accessible for AI systems to analyze and learn from. Improving the compatibility of AI tools with the dynamic and sometimes unpredictable data generated during modeling and simulation can significantly propel AI's utility in PD, leading to more informed design choices and better optimization of product features and performance (Chiarello et al., 2021a; Hunde and Woldeyohannes, 2022).

There is an ongoing and growing trend in the importance of shortening engineering design lead time. Digital simulation has successfully reduced the costs of physical testing for several decades now (Zorriassatine et al., 2003; Liu et al., 2013). The automotive industry has successfully implemented digital PD tools, such as CAD and Computer-Aided Engineering (CAE), over the last few decades. For instance, the lead time from the freeze of styling to the start of production at Audi Motors has decreased from five years to two years (Roy et al., 2006), while reports from a decade earlier show a reduced time in the automotive industry aimed at three years (Griffin, 1997). A recent survey shows the trend continues to grow now on electric powertrain technology, as they noted that 86% of manufacturers have development cycles at 24-40 months and 4% of them have cycles of 24 months or less (Morley, 2022), which shows a trend toward even shorter cycles, and similar reports from Tesla and Volkswagen confirm the trend. Even on the platform side, the age of platforms (before retirement or substantial overhaul) for various types of vehicles shows a decreasing trend over the last four decades, as it is illustrated in the Figure 1.5.

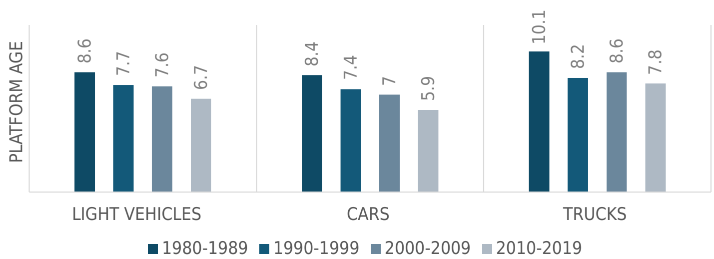


Figure 1.5: The reduced age of platforms in the automotive industry (n.d., 2017)

1.2 Problem Description

A systematic method for structuring problems and systems (VDI-2221, 1987; Wynn and Clarkson, 2018) serves as a guide that allows this thesis to get focused and dive deeper into the core of the problem studied. At the beginning of PhD studies, a series of interviews were performed with case companies, which pointed out that *Engineering Design Lead Time* was one of the problems they experienced often in their design process, and therefore, it was selected as the overall problem area of the studies that followed. However, this problem is broad and needs to be further refined. This is because manufacturing companies that design their products are typically constrained by time factors to explore the plausible designs with sufficient richness. This is particularly true for companies like those studied in this thesis that have a product where design significantly impacts their total lead time.

Engineering design lead time is a wide problem, where the symptom is that products sometimes need to be re-designed very late in the process, or do not even perform as well as they potentially can. One of the contributing mechanisms that enables exploring design alternatives is indeed the ability to evaluate the design performance, regardless of where in the process of concept, system, or detail phase design is being assessed. Therefore, to narrow the scope of the studies, the *Design Evaluation*, as one of the sub-problems that had a considerable effect on the case companies' engineering design lead time, is selected as the main problem under the study in this research. The overall problem and sub-problem (main) studied in this thesis are shown in Figure 1.6.

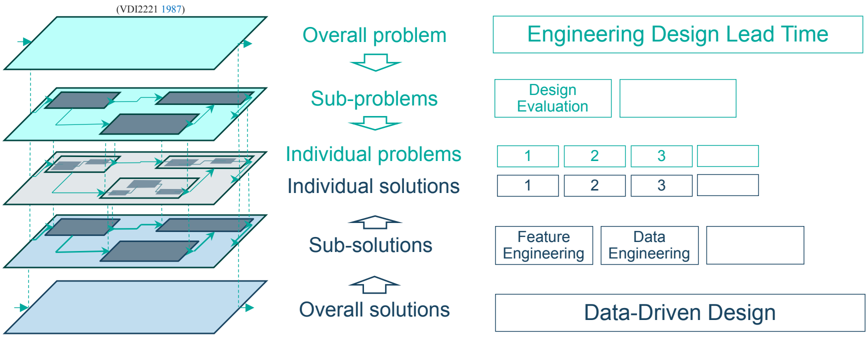


Figure 1.6: Mapping problem areas of this research to solution space

Since iterations are viewed as beneficial for development and optimization, the remedy prescribed for design evaluation in this thesis is to reduce the wait time in each iteration rather than to eliminate them. The reason is that iterations can contribute to a better final solution, even if the product development lead time makes the requirements change over time. On the other hand, recognizing that a shorter lead time does not guarantee a successful product or product development process is essential. However, shorter lead time does reduce engineering man-hours, which in turn enhances competitiveness for companies.

Being able to evaluate concepts digitally has significantly reduced the lead time for physical testing over the last two decades. Nonetheless, this shift has resulted in an increase in the lead time for the digital evaluation of a product as it has reduced the physical testing costs. Consequently, the digital *design evaluation* process now increasingly occupies a larger portion of the overall engineering design lead time and has become one of the primary challenges in product realization today 1.6.

During the first round of studies with our case companies, a development process model for their product with two iterations was developed (Arjomandi Rad et al., 2022). It was concluded that the waste in the form of the waiting times and iterations takes up a big share of their development process, as shown in Figure 1.7. This model shows that different forms of waste can be coupled with each other. For instance, reducing the number of iterations (which results in less design space being explored) can cause a reduction in total waiting time. Given that each iteration in the evaluation process may depend on external input for completion, speeding up the iterations is more critical than reducing their number.

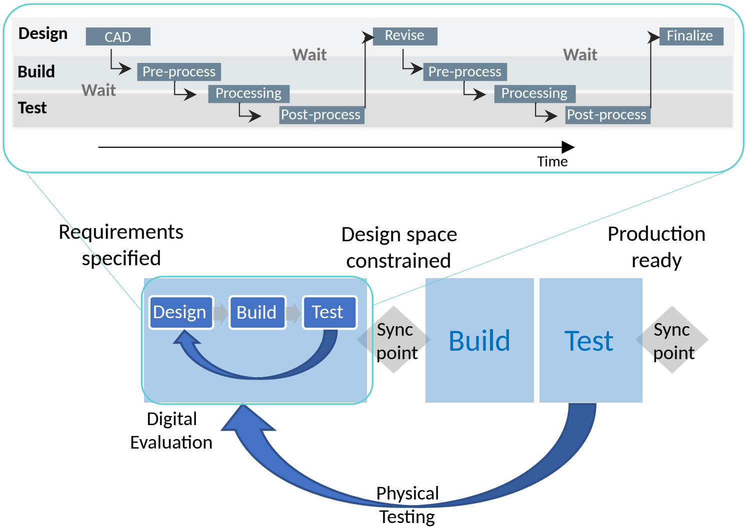


Figure 1.7: Development lead time encompasses digital and physical evaluation loops

As shown in the figure, iterative and simulation-driven products undergo iterations on two loops. The first iteration point occurs just before the gate, where digital models and simulations are validated in collaboration with the customers (first sync point). This is the inner loop shown in Figure 1.7, where the designers iterate the work from the requirements specified and reach a detailed level of the design space. The case company that was studied in this thesis aims to minimize changes after the first sync point and use as few iterations as possible on the second iteration loop with a physical test. In one case, for instance, having 60-80 loops during digital evaluation was reported to be a common practice. After the digital verification loops refine the design concept, the design progresses to the next phase in the second loop.

The second loop of iterations occurs during the physical testing and often loops back to digital iteration, often as a result of the dynamic nature of customer requirements and/or a lack of correlation between digital and physical tests. Overall, the two loops form a highly iterative process for the case company, which is a common characteristic among the groups of products examined in this research. Developers aim for a flawless launch in high-level technical products by consistently correlating digital design evaluation results with physical testing. Thus, our case companies' products with iterative processes take up to two years to launch (their engineering design lead time). This iteration continues until a highly reliable product is achieved, as expected in that group of high-precision and technical products.

Having years of product development lead time, in the automotive industry, where the trend is toward shorter cycles, is problematic. During studies with experts in the company, it was concluded that design evaluation is one of the bottlenecks that vastly affects the engineering design lead time. With the rise of AI, data-driven methods have proven to be one of the promising approaches to resolving design evaluation in iterative tasks in recent years. Therefore, it was natural to further decompose the problem at hand (design evaluation) into identifying the possible hindrances to prevent leveraging AI in design evaluation. Throughout our studies, three such individual problems have been identified (for decomposition see Figure 1.6).

1. Models depend on the initial parameterization convention: The first issue explores the dependency of the design process on CAD model parameterization, which often constrains design flexibility due to its reliance on initial parameterization conventions. The digital evaluation cycles are dependent on designers for data gathering, pre-processing, and post-processing, and therefore can be facilitated by running as many activities as possible in real-time in a transdisciplinary environment. This was an idea that was also pursued in digital twins that create a virtual environment with assets that carry the footprints of a physical system (Boschert and Rosen, 2016). A data-driven prediction model can contribute to the goals of achieving an integrated real-time development process. Yet, these modeling techniques usually use fully defined geometric parameters as low-hanging fruit as their input in their design process. This dependency can stifle innovation and adaptability in the design phase, as during the design iterations, it reduces the design space that can be explored. Reducing this dependency and allowing designers to make modifications beyond the limitations of defined parameters in legacy models, thereby fostering a more creative and efficient design analysis process.

2. Costly testing methods limit large-scale labeling of datasets: Labeling refers to the process of assigning ground truth outputs (e.g., CAE simulation results) to input data (e.g., design geometries or images) that will be used to train a model. Simulation analysis methods are heavily used in the design phase to test and improve design variants and respond to changing requirements. Simulations utilize models to study the behavior and performance of a system, mainly to reduce the chances of failure in meeting requirements. More detailed models are more effective at revealing the occurrence and consequences of iterations, but they can be costly. Models are of a simplified nature and can never capture the full complexity of reality happening iterations in PD (Wynn et al., 2007), and therefore, the tradeoff between modeling cost and the level of desired confidence should be a prioritized study. This is specifically true for the data-driven predictive models with the advent of deep learning. While many CAE tools are commonly used to evaluate input designs in data-driven methods and calculate the performance of designed outputs, the growing demand for larger datasets within data-driven methods and the prohibitive cost of methods like finite element impose limitations on such applications. This gives the second individual problem, the need for evaluation methods that can expedite the design iteration process without extending the time required for each iteration. This challenge is associated with creating outputs of the datasets, which are essential for training machine learning models used in design evaluation. The

poor scalability of the traditional digital evaluations necessitates cheaper evaluation methods for comprehensive design analysis and opens up for exploring alternative methods.

3. Prepared datasets are rigid and prevent design changes: Models in product design and development range from physical to analytical and aim to answer questions like 'How well does it work?' (Eppinger and Ulrich, 2015). As product development progresses, design decisions become more concrete, and the cost and complexity of changing models increase significantly. Early phases often use flexible, conceptual models allowing rapid exploration and iteration. In later stages, models become detailed, constrained by technical specifications, manufacturing considerations, and committed resources, reducing flexibility. Unlike conceptual or embodiment models, which can quickly adapt based on new ideas or design changes, data-driven models rely on historical or simulated data that limit flexibility early on. Expanding their flexibility typically requires significant effort to collect new data, retrain models, or modify prediction frameworks. Creating data-driven models that are easier to change and modify can be a significant accelerator for the digital evaluation of design concepts. Therefore, the last individual problem studied in this thesis examines the rigidity of prepared datasets in the context of design changes. Typically, datasets are prepared and structured in a way that does not easily accommodate modifications once design parameters are established, thereby suppressing the iterative nature of design improvements. Thus, explore methodologies for preparing modular and flexible ways of building datasets that can support dynamic design change analysis. This can enhance the applicability of running data-driven predictors outside of their trained zone to capture the consequences of radical design changes. However, it should be noted that digital evaluation can never fully replace physical tests.

Both digital and physical design loops (see Figure 1.7) play a crucial role in creating a pipeline of raw data that is essential for data-driven design evaluation and analysis. On the data science side, these pipelines are often directed into repositories for easy storage and accessibility for future utilization. Later, the data undergoes two loops for analytical tasks. The first loop focuses on data preparation, which includes steps such as data cleansing, normalization, and transformation, ensuring that the data is accurate and formatted correctly for analysis. The second loop is dedicated to identifying and fine-tuning the most suitable AI algorithms for the specific objectives of the project. This involves experimenting with various algorithms, assessing their performance metrics, and selecting the ones that best align with the intended outcomes. Only after successfully navigating through these stages can the refined data and chosen algorithms be implemented within the overall system. This process illustrates how synergy occurs between design science and data science, as shown in Figure 1.8 through a classical design study of an engineered product on the left and a data-driven problem-solving process on the right.

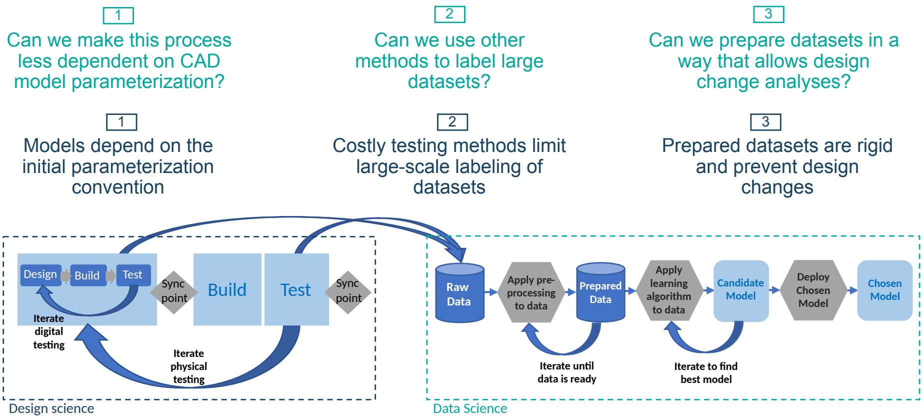


Figure 1.8: The individual problems that exist in AI adoption in product design

Leveraging advanced data-driven approaches could accelerate the design evaluation process in digital testing loops, which in turn could contribute to engineering design lead time. As shown in Figure 1.6, there could very well be other factors that can be added, and also other solutions as well to achieve this goal. However, the three individual problems summarized above are the only hindrances to the application of data-driven methods in design evaluation that have been studied in this thesis. These issues were identified through several studies conducted over the past five years in collaboration with industrial partners involved in the two research projects.

1.3 Aim and Objectives

The overarching aim of this thesis is to enhance the efficiency of the engineering design process by accelerating digital design evaluation loops, thereby reducing engineering design lead time without sacrificing the depth and breadth of design exploration. By selecting the data-driven approach as an avenue to explore, the objective is to identify and mitigate key bottlenecks that hinder the effective implementation of data-driven methods within iterative product design evaluation cycles.

This approach draws inspiration from historical advancements in engineering analysis, such as Finite Element Analysis (FEA) and Computational Fluid Dynamics (CFD). Similar to how these computational techniques did not eliminate the need for physical testing but reduced its frequency and became a tool for designers for faster design space exploration, AI in design evaluation should be viewed as an augmentative tool rather than a substitutive one. Thus, the thesis supports the exploration of what could be termed the "next best model," as illustrated in Figure 1.9. The concept of the next best model revolves around employing AI to develop models and methods that might potentially have a slightly lower confidence level than the most accurate models achievable through expensive and time-consuming methods, but offer substantial cost benefits. If reduced cost and increased speed are achieved with these AI-enhanced evaluations, they may be a valuable asset in the iterative design process, providing rapid insights that are economically feasible and sufficiently accurate for early to mid-stage design decisions.

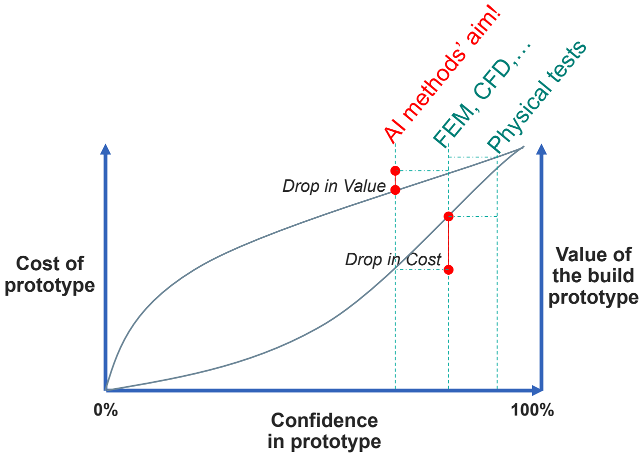


Figure 1.9: The cost and value for different models in PD, Adapted for (Sargent, 1988)

The hypothesis is that utilizing data-driven design evaluation methods can provide practical benefits and actionable insights to designers, contributing to the efficiency and effectiveness of the process. To achieve this overarching aim (Engineering design lead time), the research pursues three specific objectives, i.e., addressing individual problems to fulfill the main design evaluation issue:

1. To investigate and propose methods to reduce the dependency on initial CAD parameterization conventions, thereby increasing design flexibility and innovation potential during iterative digital evaluation cycles.

2. To develop evaluation methodologies that enable large-scale dataset labeling and facilitate efficient generation of training data for AI-based predictive models, mitigating the prohibitive computational costs typically associated with detailed simulation methods such as finite element analysis.

3. To explore and establish frameworks for creating flexible and modular datasets that accommodate dynamic design changes, thereby enhancing the robustness and adaptability of data-driven models in evaluating product concepts, even when subjected to significant design alterations.

By addressing these objectives, the research intends to provide an approach for more streamlined integration of data-driven AI techniques into the engineering design process, ultimately contributing to a more agile, responsive, and competitive product development lifecycle.

1.4 Research Questions

The research is guided by specific research questions. These questions are structured to focus the investigation on critical aspects of integrating data-driven AI techniques within the engineering design evaluation processes:

RQ1: What are the challenges of data-driven design evaluation in the design process of iterative and simulation-driven products?

The question is raised to gain a better understanding of what evaluation aspects need to be addressed, and in particular, in what way a data-driven evaluation approach can be improved.

RQ2: How can data-driven methods in product development processes be customized for the design evaluation of iterative and simulation-driven products?

By answering this research question, factors affecting possible solutions will be identified, and the limitations concerning possible success factors for proposed solutions will be discussed.

RQ3: What data-driven design supports can be developed for more efficient design evaluation?

This question drives the need to develop and test new data-driven approaches while also providing an initial evaluation of the effectiveness of the proposed methods in relation to the project's objectives.

1.5 Scope and Delimitation

This research focuses on the iterative design process, which is a common characteristic among high-level technical products. The research is also limited to concept refinement utilizing computer-aided methods to minimize excessive physical testing. The iterations discussed here are constructive in nature, facilitating improvement, and are not intended to be eliminated. The focus is not on doing fewer simulations but rather on developing a smarter approach to building knowledge from modeling and simulation that enables data-driven AI techniques in design evaluation.

Simulation-driven design encompasses a wide range of knowledge areas involving various engineering disciplines, including mechanical, electrical, chemical, software, and control engineering. All these fields utilize simulations in different forms and scales to aid in developing complex products. In this thesis, the term ”simulation” is used interchangeably to refer to CAE and FEA. As a result, the case studies in this research exclude other types of simulations, such as agent-based, discrete event, and multimethod simulations, which are also commonly used in product development.

The focus of the thesis is on the design process rather than the actual design itself. As such, the examples presented are merely defined to demonstrate and test the suggested methods, and should not be taken as an actual product example. This means that the data presented should not be interpreted as valid for the actual product design case. This is partially due to confidentiality matters surrounding the real business conditions, and partially due to the need to idealise and make assumptions for the purpose of method testing, rather than product evaluation.

Another delimitation for this research is its concentration on classical modeling techniques despite the surge in generative AI research. This decision is driven by the enduring relevance and untapped potential of traditional models, which often require less computational power and offer greater interpretability than their generative counterparts. Additionally, the data engineering perspective in this thesis can promote advancements in both traditional models and emerging AI technologies upon its success.

Chapter 2

Theoretical Framework

The forthcoming sections of this chapter establish a foundation for the theoretical framework that guides this research. The wide scope of methods addressing the design evaluation problem does not allow for discussing all the existing methods here. However, this chapter propels through some of the methods in data-driven design as a mainstream topic on top of which this thesis is built. Figure 2.1 presents a word cloud generated from the text of all included publications in this thesis. The size of each word in the figure reflects how frequently it is used.

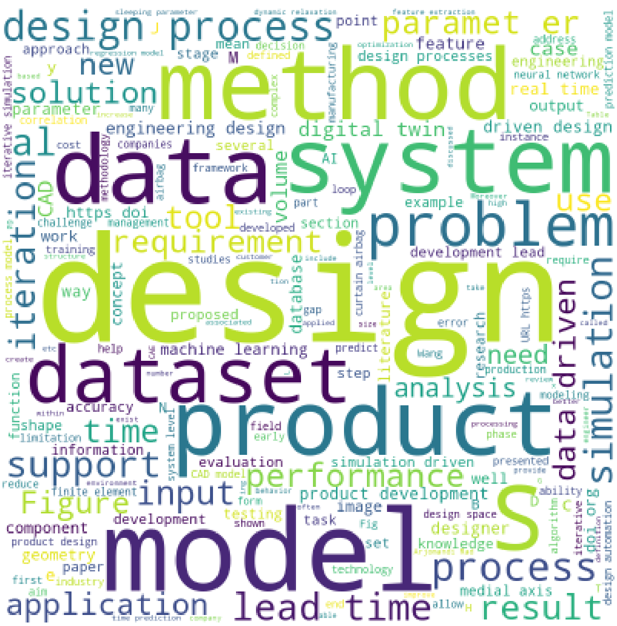


Figure 2.1: Word cloud generated from the text of all included publications

The chapter is structured to progressively build an understanding of the PD process, AI, and their intersections in data-driven design methodologies. It begins with a general overview of PD, providing background information relevant to understanding the context and challenges. The next section, 2.2, delves into how design automation and CAD-CAE integration have been utilized to enhance the efficiency of simulation-driven processes, addressing gaps and limitations in current integration techniques. Following these foundational topics, the chapter moves to Artificial Intelligence and Data Science in section 2.3, introducing core AI concepts and their relevance to data-driven design, including a discussion on data engineering techniques for data preparation, feature extraction, and preprocessing.

Building on these concepts, the chapter continues with Artificial Intelligence in Design Science 2.4, structured to cover Data-Driven Design, Meta/Surrogate Modeling, and Design for AI. This progression highlights how AI methodologies are applied to design processes, particularly in CAD and CAE contexts. The discussion then transitions to data engineering in Data-driven Design in section 2.5, focusing on Data Mining and Feature Engineering methodologies that enhance predictive model performance. The chapter ends in section 2.6 by addressing Datasets in Data-driven Design, emphasizing the importance of structured datasets, identifying existing gaps, and proposing improvements to enhance their applicability in AI-driven design methodologies. This coherent structure ensures that the theoretical framework aligns directly with the research objectives.

2.1 Product Development and Design

Design typically refers to the process of inspiration, ideation, and implementation (Brown, 2008), and engineering design uses this cycle to satisfy requirements and fulfill a need (Chakrabarti et al., 2011). Engineering design is not a single step but a series of interlinked activities that evolve a product and move it from an idea to existence. PD is a broader term that includes the design process but also extends beyond it to cover the entire process of bringing a new product to market (Ulrich and Eppinger, 1995).

Since PD is challenging and costly, as it takes a significant amount of time and resources to gain the necessary knowledge about the forthcoming design. The amount of engineering and the coordination of engineering resources can be a substantial part of the design effort and cost. Therefore, modeling PD stages and interactions between them has been studied for decades. A linear PD process with a handful of methods and guidelines is described in (Ulrich and Eppinger, 1995) with six phases, depicted in Figure 2.2. Emphasizing the generic nature of the exhibited process, the authors argue that each company needs to adapt these steps based on the company's context and the project's challenges. Also, the process might not follow such a sequential fashion, as many factors could cause overlapping between phases, step-backs, or iterative activities. In Ulrich & Eppinger's model, early phases of design refer to the initial stages of the design process where the main focus is on understanding the market, identifying customer needs, and conceptualizing possible solutions. In this thesis, we utilize this model to distinguish between different phases of product development, as it is widely accepted among scholars.

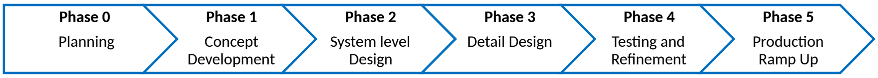


Figure 2.2: Generic PD process model by (Ulrich and Eppinger, 1995)

Another well-known linear PD process, with a narrow focus on the engineering design and problem-solving aspect of product development, is outlined by Pahl and Beitz. They define PD as a general decision-making process and propose a practical and procedural process as illustrated in Figure 2.3. It is mentioned that any process model has to be considered an operational guideline for activities based on the patterns of technical PD and the logic of stepwise problem-solving (Pahl and Beitz, 1984). The reason may be that this model comes from a time when the use of digital support tools was limited; however, the basic logic remains useful and well recognized in engineering design practice.

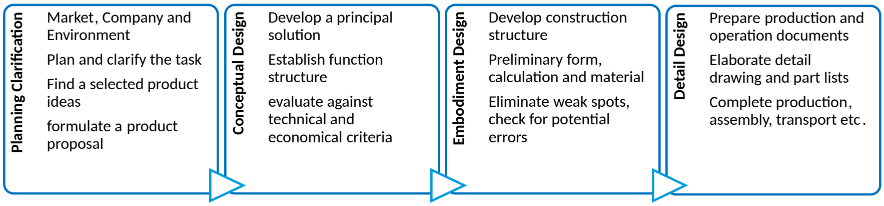


Figure 2.3: Generic PD process model by (Pahl and Beitz, 1984)

In the field of PD, more methodologies distinguish themselves by addressing the complexities of design through structured frameworks. For instance, as closely related to Pahl and Beitz, the VDI-2221 guideline methodically progresses from conceptual to detailed design, emphasizing systematic and orderly design stages (VDI, 1993). Ullman's approach focuses on iterative refinement, using feedback loops to enhance design based on real-world testing and evaluation (Ullman, 1992). Axiomatic Design is another model that applies fundamental design axioms to ensure that solutions meet their objectives while minimizing complexity (Suh, 1998). Hubka & Eder's work outlines a theoretical model that categorizes design activities into classes, enhancing understanding of the design process through a structured scientific approach (Hubka and Eder, 1988). David Stauffer emphasizes the importance of fundamental design principles that guide the development process, ensuring consistency and quality in engineering outcomes (Stauffer and Ullman, 1991). Reviewing a large variety of these methodologies, Wynn & Clarkson categorize PD processes into meso, macro, and micro levels, offering a detailed map that guides the organization of design activities to address diverse project needs effectively (Wynn and Clarkson, 2018).

Although PD models can give intuition to the flow of work or activities, some of them are criticized, for instance, for being too rigid to handle, too planned to be innovative or dynamic, too controlling, and bureaucratic, with too much non-value-added work (Cooper, 2014). Choosing the best process model for a design process is not trivial and needs attention to multiple factors. The degree of formalism is one of the factors to consider when choosing from a wide range of processes (Graner and Mißler-Behr, 2013).

A structured design method increases design transparency and helps achieve complex objectives by preventing unnecessary iterations (Heikkinen, 2021).

In this work, PD models have primarily served as tools to study the design processes within case companies and identify key problem areas in their development. One of the influential models that inspired Figure 1.7 is the design-build-test design model (Wheelwright and Clark, 1992, 1994). All models used have provided a structured lens through which the complexities and challenges faced by the case company in their design activities have been analyzed and understood. Furthermore, the insights gained from this analysis have been instrumental in suggesting targeted supports and interventions aimed at addressing these challenges.

As societal demands and technological capabilities evolve, so too does the field of PD. Modern PD practices are increasingly addressing more complex, systemic issues that reflect broader economic, environmental, and social challenges (Hallstedt et al., 2020). This evolution necessitates a shift towards more integrated and adaptive PD frameworks that can handle today's design problems. Emphasizing, for instance, sustainability (Chatty et al., 2022) and user-centric approaches (Chou and Wong, 2015), contemporary PD models integrate cross-disciplinary knowledge and leverage advanced technologies to foster more resilient and flexible design strategies. Despite the rapid evolution in methodologies, it is crucial to recognize that many traditional theories remain highly relevant to contemporary PD practices. Theories such as the axiomatic design principles (Suh, 1998) and the concept-knowledge theory (Hatchuel and Weil, 2003) continue to serve as fundamental frameworks upon which modern, digitalized, and automated approaches are constructed.

As PD continues to evolve, it increasingly intersects with the principles and practices of System Engineering (SE) (Browning, 2018). This convergence is evident in the holistic approach both disciplines take to address complex, multi-faceted projects. Systems engineering is evolving to integrate commercially available subsystems and leverage this off-the-shelf approach to reduce the time to market (INCOSE, 2023). Modern PD models' shift towards addressing broader product lifecycle considerations, with a focus on more agile, systemic, and human-centric approaches, mastering interventions, and developing products within value networks(de Weck et al., 2011; Isaksson and Eckert, 2020). This highlights the critical role of systems thinking in understanding and managing the interdependencies and complexities inherent in modern PD. These perspectives reinforce the necessity of adopting a systems-oriented approach in PD to ensure comprehensive and coherent product strategies that are capable of adapting to evolving technological and market conditions.

2.2 Simulation-Driven Design

Models provide a simplified version of reality, and simulations are used to validate a concept, idea, or design. Simulations can be considered where observing or testing the real-world experiment is expensive or impossible, and where analytical solutions are too complicated or costly to be validated (Maria, 1997). In this sense, simulations serve not only as substitutes for experimentation but also as tools for early exploration and validation of design alternatives within a virtual environment.

Simulation-Driven Design (SDD), also known as simulation-based design, has been defined as ”a design process where decisions related to the behavior and the performance of the design in all major phases of the process are significantly supported by computer-based product modeling and simulation” (Sellgren, 1999). In essence, SDD transformed traditional design-built-test cycles with an additional step in between for computer-based analyses/syntheses to form a design-simulate-build-test cycle (Bossak, 1998).

In another view, SDD aims to move simulation technology from the middle and the late cycles of a design process to the very front cycles and reduce the time it takes for companies to develop products. For instance, using CFD simulation on jet components, it has been demonstrated that more intensive use of computational simulations in the early design phase leads to fewer tests because it enables analytical evaluation of the design (Isaksson et al., 2000). The activities involved in SDD are inherently generic, meaning they are broadly applicable regardless of the specific design context. However, while the activities themselves are context-independent, the methods and data required to execute them are highly situation-dependent. These elements must be tailored to the particular product, domain, and stage of development, and therefore cannot be prescribed in a generic form. The relation between a generic activity and its associated data and methods, with five steps, is shown in Figure 2.4.

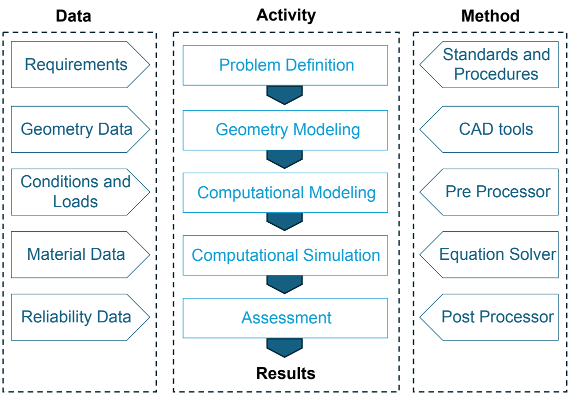


Figure 2.4: Generic SDD process model with five steps (Isaksson, 1998)

Despite its potential, the expansion of SDD and achieving computational efficiency were hindered by integration challenges between the design and simulation activities. The state of Product Data Management (PDM), CAD, and CAE in the 1990s and early 2000s was perceived as inadequate to satisfy the needs of seamless interoperability (Shephard et al., 2004), limiting the effective use of simulations in early development stages. Uncertainties in input parameters or model structures were another concern point in the field of SDD, since they can propagate through simulations and influence outcomes and decision-making processes. Du and Chen (Du and Chen, 2000) proposed integrating robust design principles to handle and mitigate these uncertainties, ensuring reliable design decisions. Wall (Wall, 2007) suggested that effective SDD should actively guide designers toward optimized configurations early in the development stages rather than merely validating predefined solutions.

More recent advancements in simulation-driven approaches emphasize model-centric communication and the integration of Knowledge-Based Engineering (KBE) systems, as demonstrated by Sandberg et al., (Sandberg et al., 2013), to facilitate information exchange and enhance decision-making within PD environments. For instance, the knowledge-based master model approach allows for concurrent management of design and analysis models across disciplines, thereby enabling faster iterations and more efficient exploration of the design space (Sandberg et al., 2017). As these integrated frameworks aim to streamline cross-functional collaboration and reduce modeling overhead, they are increasingly complemented by surrogate-based optimization techniques. Simulation-driven approaches have benefited from surrogate-based optimization techniques, where expensive high-fidelity models are replaced with computationally efficient surrogates, thus reducing computational costs without significantly sacrificing accuracy (Koziel and Leifsson, 2016).

Simulations in PD have grown to be an umbrella term for a diverse range of methods. For instance, in the mechanical design field, CAE was introduced first in the 1980s as a way to provide analytical information in a timely manner in the PD process. Over time, this evolved into the concept of virtual prototyping, where simulations replaced some of the physical prototyping cycles to evaluate performance earlier and more frequently (Kojima, 2000). Modern CAE tools simulate a variety of physical phenomena and are employed not only for validation but also for optimization of both products and processes (Merkel and Schumacher, 2003). However, the effectiveness of CAE is closely linked to how strategically and consistently it is embedded into the product development workflow (Isaksson, 1998; King et al., 2003). Successful implementation requires not only technical capabilities but also alignment with the design process to ensure timely and actionable insights.

In this thesis, SDD is utilized at various levels within the conducted case studies. CAE simulations can be costly and contribute to extended engineering design lead times, which is the overall problem of this thesis. However, the data needed for any potential support also comes from CAE simulations. In this context, the studied products are used with automated CAE simulations to map the design space into the solution space and create the necessary datasets for further analysis.

2.2.1 Design Automation

Design automation was early defined as "use of computers to aid in the design of computers" (Case, 1972), with an initial focus on automating basic calculations and visualizations in early computer-aided systems in the computer-science society. In the PD context, as computers started to be used as "machine elements" in design methodologies and simulations dominated design practice, the Design Automation (DA) stabilized itself as a field (Ragsdell, 1980) in design optimization and numerical methods. Later, the definition evolved to a broader one as "engineering support by the implementation of information and knowledge in solutions, tools, or systems, that are pre-planned for reuse and support the progress of the design process" (Cederfeldt and Elgh, 2005). The definition includes automating tasks linked to the design process, whether directly or indirectly. This encompasses a spectrum of activities, from the creation of individual components to the comprehensive development of complete products. Since the aim of automating design is to improve engineering productivity and reduce engineering cycle time in meeting customer specifications, DA is a suitable approach for the current research problem area.

To enable effective reuse and flexibility in automated design, modern CAD systems require tool-independent, generic modeling strategies (Amadori et al., 2012). A useful distinction in this context is between morphological and topological transformations, as illustrated in Figure 2.5.

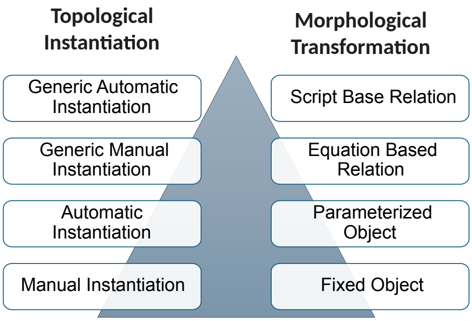


Figure 2.5: Geometry handling in DA with various stages (Amadori et al., 2012)

Morphological transformations deal with how a design instance changes in form, ranging from fixed geometries to parameterized and script-based models. Topological transformations, on the other hand, involve structural changes, such as adding, removing, or replacing model instances. From a practical viewpoint, DA boils down to using programming to connect design tools/assets and create information flow that facilitates the design process. This can go beyond only design and cover up to evaluation steps. For instance, a design asset like a spreadsheet can be connected to a CAD design tool and plugged into a CAE simulation tool, where the information flow is supported by allowing the computer to map design requirement inputs to design objectives. This view, which also reflects how DA has been used in this thesis, reveals some of the limitations of DA that are related to the nature of how it works.

For instance, DA deals with part of the costly development work. It helps reduce manual work, yet the core information processing, such as solving stiffness matrix in FEA, which is time-consuming, still needs to be performed during the process (Arjomandi Rad et al., 2022). This is why many designers view design automation primarily as a means to reduce repetitive tasks and save time in later stages of product development, rather than as a tool to explore solution spaces and support complex design tasks in early phases. Another limitation is related to the nature of automation, which tends to support repetitive and non-creative design tasks and can not handle creative work (Rigger et al., 2018a). DA is usually applied to a specific process that has been well defined to serve a particular purpose. This often involves the use of specialized tools. As a result, there are not many general-purpose automation tools available. Instead, there are process automation tools, such as ISIGHT and MODEFRONTIER, which assist users in automating simulation processes.

While large companies tend to have more resources and may possess some knowledge of design automation, there is an observation that these designers might not be directly involved in implementing automation solutions, leading to a disconnect (Rigger et al., 2018b). On the other side, smaller companies, struggle as development of DA applications is generally undertaken by domain engineers who may not have formal knowledge of engineering or software development training, with subsequent development processes lacking the structure of formalized methodologies, and important principles can be neglected (Van der Velden, 2008). A mapping between approaches in DA finds out that few methods deal with the generation of different design alternatives and laborious design tasks (Rigger et al., 2016). This was used to argue that the application of DA has been limited, most notably in configuration systems such as mass customization.

Today, DA encompasses more sophisticated applications in PD, such as using a knowledge-based approach for creating complex CAD models (Frank et al., 6 30), integration with data-driven and design optimization (Gustafsson, 2022), and as support for additive manufacturing (Wiberg et al., 2023). DA frameworks have been developed to automate design processes for highly customized products (Frank et al., 6 30; Colombo et al., 2008), explore design alternatives (Gustafsson, 2022), and facilitate data-driven design (Chiu and Lin, 2018). These systems can significantly reduce design time and costs while improving product quality (Johansson Joel and Elgh Fredrik, 2013). Implementation strategies include prototyping to evaluate potential benefits (Entner et al., 2019) and developing methodologies tailored to SMEs (Colombo et al., 2008). As DA continues to advance, it promises to enhance design space exploration, integrate manufacturing constraints, and automate repetitive tasks throughout the PD process.

2.2.2 CAD-CAE Integration

The integration of CAD and CAE systems has been a long-standing challenge due to differences in design information and intent loss, inconsistency in data models, differences in mathematical descriptions or geometric representations, and workflows (Khan and Rezwana, 2021; V. et al., 2011). Traditional visions for CAD-CAE integration include either enabling CAD design tools to run simulations, enabling CAE simulation tools to have better design capability, or the master model approach, which provides a consistent model that serves both design and analysis purposes, letting engineers update one file that automatically reflects changes across all stages of development (Smit and Bronsvoort, 2009).

In the conventional sequential up-to-down design process, the idealization of the geometric model (simplification and modification) is implemented before meshing and simulation (Feng et al., 2020). This hinders integration because the initial design model usually contains various detailed features for downstream applications, like log history, tolerance information for manufacturing, and so on, which are redundant information for finite element simulations. Multiple approaches have been proposed to address this issue, including the development of common data models (G. and Yongsheng, 2010), XML-based integration methods (Zhang and Li, 2011), and mixed shape representations (O. et al., 2010). Feature-based multi-resolution and multi-abstraction modeling techniques have been explored to create unified models for both CAD and CAE (Sang, 2005). Despite these efforts, challenges persist in maintaining design intent, data consistency, and mathematical descriptions across platforms. A holistic approach considering product, people, tools, and data dimensions has been suggested to systematically address the integration issues (Deubzer et al., 2005).

The intelligent advisory system called PROPOSE, for design improvements considering FEA results, is an example of another successful integration system (Novak and Dolšak, 2008). The idea was to encode the knowledge and experiences and build an intelligent advisory system to help the designer to perform an analysis-based design improvement process. Regarding any question, the system provides help to the user in the form of explanation or advice to inexperienced designers as to how to change/improve the design in critical areas of a structure after stress-strain or thermal analysis. Recently, Heikkinen et al. proposed a simulation-ready CAD model that essentially integrates pre-processing with CAD work (Heikkinen et al., 2016). It has been argued that this kind of support can address time-pressured technology development in small-sized companies, where building extensive KBE systems are not feasible.

Recent research indicates a growing trend in the field of CAD-CAE towards integrating AI, ML, and deep learning methods that can be regarded as a new vision in the field. This integration aims to enhance design processes, automate tasks, and improve performance evaluation with design data (Hunde and Woldeyohannes, 2022). AI-powered CAD-CAE frameworks can populate design concepts in latent or coded space and thus benefit from lower dimensionality. They also facilitate conceptual design by evaluating them with various inputs, such as images, instead of only depending on parametric models (Yoo et al., 2021). Moreover, hierarchical data repositories as neutral formats are investigated as a bridge between CAD and CAE systems (Khan and Rezwana, 2021) that allows for direct transportation of parametric information to the CAE modeler. In another study, digital twins are proposed to enhance Machine Learning (ML) predictions regarding cutting forces and conditions by leveraging data collected from both CAD and CAE environments (Ozel and Jarosz, 2022). This integration allows for real-time adjustments based on feedback from the machining process, ultimately refining tool path strategies and improving overall efficiency. The digital twin-based integration of CAD and CAE can also be viewed as part of the growing trend of AI adaptation in this field.

2.3 Artificial Intelligence and Data Science

Ever since John McCarthy coined the term "artificial intelligence" in the 1950s, defining it as "the science and engineering of making intelligent machines", the vision has been to emulate aspects of human intelligence. Artificial intelligence is traditionally compared to natural intelligence, in the sense that in our body, senses gather data and human reasoning draws a conclusion from them (de Callataÿ and M, 1986; de Callataÿ, 1992). However, this comparison is up for debate, and there is still no consensus among scholars whether machines really learn and understand like humans (Mitchell and Krakauer, 2023). This is mainly due to a lack of knowledge about how the brain works at that level.

The term ML was also popularized in the 1950s by AI pioneer Arthur Samuel as "the field of study that gives computers the ability to learn without explicitly being programmed." Another well-received definition of the term ML is "a form of applied statistics with increased emphasis on the use of computers to statistically estimate complicated functions and a decreased emphasis on proving confidence intervals around these functions" (Goodfellow et al., 2016). This definition highlights the capacity of computers to learn complex response surfaces, recognize patterns, and make decisions through iterative processes, while placing less emphasis on verifying the underlying functions. This focus aligns well with this thesis, which seeks to utilize machine learning for predictive purposes.

A subset of machine learning is deep learning that focuses on utilizing neural networks with a high number of layers between the input and output layers to perform tasks (Schmidhuber, 2015). To understand the relationships between different AI-related fields (AI, ML, Deep Learning) that enable and employ data-related fields (Data Science, Data Engineering, Data Analytics), Figure 2.6 illustrates a Venn diagram featuring different terms' relations with practical examples relevant to PD and engineering design.

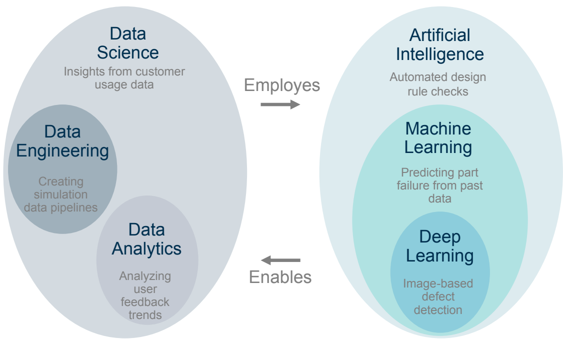


Figure 2.6: A Venn diagram showing AI and Data Science fields with examples

AI and ML can be positioned within Ackoff's and Zeleny's hierarchical framework of data, information, knowledge, and wisdom (DIKW) (Ackoff, 1989; Zeleny, 1987). ML algorithms primarily employ raw data (as discrete, unorganized facts) and transform it into information, as they place the data into meaningful contexts. This process often enables identification of patterns and relationships that help generate knowledge (the ability to interpret and apply insights). Although it has been argued that ML models lack true "understanding" or consciousness, referring to the definition from (Goodfellow et al., 2016), the opposite point of view is that the DIKW framework does not require human-like comprehension, as in this context, knowledge is a functional rather than a philosophical or conscious awareness. In this view, AI extends beyond the acquisition of knowledge by operationalizing it in ways that enable decision-making processes, thereby approaching the realm of wisdom. Through iterative enabling and employing, AI and ML continually refine their understanding, revealing how data can be systematically elevated to higher-level insights that guide more informed decisions.

The concept of data itself is subject to ongoing debate, particularly within interdisciplinary research. Ballsun-Stanton adopts the notion that data is not an objective representation of reality but a socially constructed concept shaped by context and interpretation (Ballsun-Stanton, 2012). Others describe data as things that flow in machines (Al-Fedaghi, 2016), or as inputs that help us understand, regulate, and predict the world (Kitchin and Lauriault, 2014). For the sake of predictive goals in this thesis, the latter is chosen. On the other hand, there are claims that data science is not a science but rather a research paradigm closely related to computer science and statistics (Brodie, 2023). Some others describe it as a significant shift, ushering in the "fourth paradigm" of science (Hey et al., 2009). In this thesis, the definitions provided by Hayashi is adapted, in which *Data Science* is a comprehensive concept that integrates statistics, data analysis, and their related methodologies, as well as the results derived from them (Hayashi, 1998). In this definition of data science, Hayashi discusses three phases for design science activity that include: *design for data*, which concerns planning what data to gather, *collection of data*, that answers the questions like how to gather the data, and finally *analysis of data*, as cornerstones to understand actual phenomena with data.

2.3.1 Data Engineering

Most data scientists reportedly spend up to 80% of their time on data preparation. (Press, 2016), which aligns with reports that consider Data Engineering (DE) as a major time consuming activity in design science (Liu et al., 2024). This has made DE to emerge as an independent discipline in response to the growing complexity and volume of information faced by organizations. As organizations increasingly rely on data for decision-making, they face what has been termed the "information paradox", where companies face both a lack of useful information and an overload of unnecessary data (Königer and Janowitz, 1995). This paradox highlights the challenge organizations face in effectively leveraging available data to support decision-making without being overwhelmed by it. The evolution of DE has been categorized into four distinct, but non-chronological generations (Klettke and Störl, 2022), as shown in Figure 2.7. Each generation builds upon the capabilities of the previous generation, reflecting increasing levels of automation and abstraction in data handling.

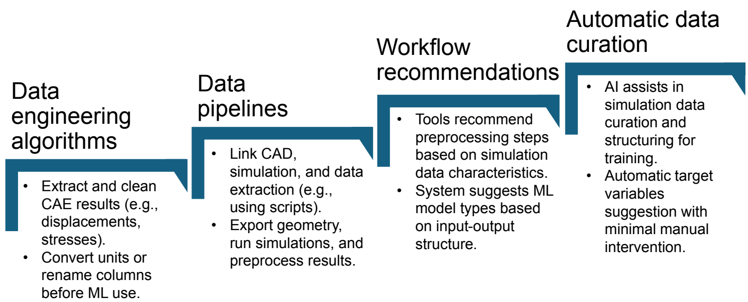


Figure 2.7: Four generations of data engineering applied to product development

- The first generation of DE algorithms involves manual efforts focused primarily on cleaning, transforming, and preparing data. This manual stage laid the groundwork for the subsequent standardization and automation of preprocessing tasks.

- The second generation of DE is the data pipelines. Here, data workflows became standardized and automated, which enhanced the efficiency and reliability of data handling processes.
- The third generation of DE utilizes intelligent adaptation, which introduces algorithmic recommendations and intelligent workflows. Significant aid from domain experts is required to navigate complex data landscapes.
- The fourth generation, Automatic data curation, aims at comprehensive automation of data handling, thus empowering non-technical domain experts to utilize data without extensive engineering expertise (Klettke and Störl, 2022).

Today, DE serves as a foundational component of data science, encompassing the creation, optimization, and maintenance of data infrastructure, architecture, and processing pipelines (Jain, 2003). Within the scope of this thesis, data curation plays a central role by enabling the construction, extraction, and selection of relevant features from simulation and design data.

Effective DE helps improve design and manufacturing processes and supports decision-making by providing clear and timely insights from data (Turney, 2002). The primary focus of contemporary DE is creating reliable, scalable, and adaptable data pipelines. Utilizing advanced data management tools such as SQL and NoSQL databases, Apache Hadoop, and Spark (Jain, 2003). These tools enable organizations to manage growing data volumes and perform advanced analytics. However, they also introduce challenges, including handling heterogeneous data types, meeting real-time processing demands, and ensuring compliance with governance and data quality standards (Achanta and Boina, 2023).

As can also be inferred from Figure 2.7, the latest advancements in data engineering stem from the integration of AI and big data technologies, particularly in support of machine learning applications (Roh et al., 2021). As data volumes continue to grow, the importance of DE in enabling efficient and effective use of data across organizational contexts remains paramount. In this context, synthetic data generation and simulation-based data augmentation are emerging as promising solutions to address the scarcity of labeled data, especially in deep learning contexts (de Melo et al., 2022). These techniques are particularly relevant to this thesis, where curated datasets derived from simulation outputs are critical for training predictive models and enabling design automation.

Data engineering has emerged largely outside the engineering design community. However, these technologies are being applied across various other domains, including medical diagnosis (Chan et al., 2020) and the construction industry (Baduge et al., 2022), paving the way for more intelligent and efficient design and engineering processes. As engineering design continues to evolve toward data-driven practices, data engineering offers essential capabilities for structuring, managing, and extracting value from large volumes of design and simulation data (Liu et al., 2024). In this thesis, these capabilities are leveraged to support the integration of simulation-driven design, machine learning, and design automation through systematic and scalable data handling approaches.

2.4 Artificial Intelligence in Design Science

AI has increasingly shaped the methodologies and outcomes in contemporary design science. This has been done by enabling organizations to surpass the limitations of traditional design processes by improving scalability, broadening scope, and enhancing adaptability (Verganti et al., 2020). The earliest forms of AI in design science emerged in the 1960s and 1970s, primarily in the form of rule-based expert systems designed to emulate human decision-making and capture expert knowledge. However, this focus did not intensify until the 1980s, where they aimed to replicate domain expert reasoning. Over time, AI expanded into other paradigms, such as rule-based, knowledge-based, and data-driven approaches and has found its way into many applications in fields like engineering and design (Reddy and Fields, 2022).

To give several examples, AI applications are being integrated to improve the construction of the design process, considering cognitive design concepts and user behavior (Gameil et al., 2024). In health communication, AI components have shown promise in creating more personalized and interactive solutions (Neuhauser et al., 2013). AI tools are being developed to support the implementation of Axiomatic Design principles, automating early-stage identification of functional requirements and assisting in design decomposition (Akay and Kim, 2021). AI-powered design tools streamline workflows, enable generative design, and facilitate predictive analytics for informed decision-making (Bagnato, 2023). In materials design, AI algorithms are accelerating the discovery of novel materials with optimized properties (Badini et al., 2023).

The growing prevalence of AI systems has raised concerns about their transparency, interpretability, and trustworthiness (Felzmann et al., 2020). Explainable AI (XAI) methods have emerged to address these issues, aiming to make AI models more understandable to various stakeholders (Stoyanovich et al., 2020). In the design realm, the systems engineering paradigm is used to promote the decomposability of engineering designs into interconnected components (Geyer et al., 2024). This approach employs a hierarchical component system to create a deep neural network that incorporates interpretable information and enhances the explainability of predictions. Explainable AI has also been shown to be capable of shedding light on traditional evolutionary algorithms for design space exploration. As an intelligent assistant to designers and engineers, the algorithms help steer the search towards a more desired region, enabling informed decision-making for sustainable outcomes (Dubey, 2024).

2.4.1 Meta/Surrogate Modeling

While surrogate modeling, or meta-modeling, also has its roots in the 1960s and 1970s, it originated primarily as a statistical and optimization-driven approach aimed at efficiently approximating complex engineering simulations (Viana et al., 2021). Over time, however, surrogate modeling techniques have evolved to become closely integrated with modern AI methodologies. Many statistical approximating methods, such as Response surface methodology, Taguchi methods, Artificial Neural Network (ANN), Inductive learning, and Kriging, have been used to predict the output of the computation-intensive design problems (Simpson et al., 2001; Wang and Shan, 2007; Sun and Wang, 2019). These methods function as substitutes for high-fidelity models, enabling the construction of predictive approximations that significantly reduce computational cost and engineering lead time. As illustrated in Figure 2.8, surrogate models serve as an intermediary layer that uses part of the design inputs and replaces the simulation model to approximate the output value, facilitating more efficient iteration and exploration during product development.

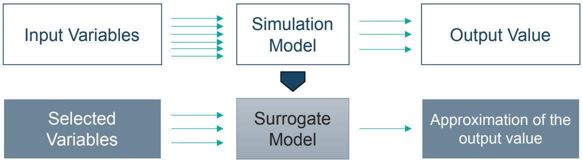


Figure 2.8: Substitive nature of surrogate models (Golzari et al., 2015)

Despite their growing capabilities, surrogate modeling techniques have largely remained confined to the computational analysis stage of the product development process. Their integration into broader phases, such as conceptual design, manufacturing planning, or knowledge capture, remains limited. As a result, their potential to support cross-functional decision-making or to encode tacit knowledge from downstream processes is underutilized (Adler, 2008). This limited scope restricts the broader applicability of surrogate models in facilitating end-to-end data-driven design workflows.

A persistent challenge in surrogate modeling is the curse of dimensionality, which implies that the performance or accuracy of a system is reduced by an increased number of dimensions. In high-fidelity engineering domains, such as vehicle body structures or aircraft design, models often involve hundreds of design variables, leading to highly nonlinear and complex response surfaces. Designers expand design dimensions to address such concerns, resulting in so-called High Dimensional, Expensive, and Black-box (HEB) issues (Shan and Wang, 2010). Another way to attack this problem is to break down the problem into several subproblems (and train several models) but the drawback is that each of the subproblems can have a different correlation with output and optimizing weights is not a trivial practice due to coupled and complex relations and lack of knowledge (Li et al., 2017). While this field has been established for a couple of decades and is supported by a wealth of published research, there continues to be an opportunity for further exploration, particularly in addressing challenges related to dimensionality and parameterization.

One main approach in the literature as a solution for HEB is dimensionality reduction. Dimensionality reduction approaches focus on simplifying geometric or parametric representations while retaining key performance-driving variables. Examples include using minimal yet effective geometric features in 3D printing processes (Wang et al., 2018) or simplified geometry for the road wheel design process (Yoo et al., 2021), or simple design geometry with only three design parameters as well as a less computationally expensive CAE method, modal analysis (Du and Zhu, 2018). Alternatively, increasing the number of training samples has been shown to improve model generalizability, as demonstrated in studies involving 60,000 automotive hood designs (Ramnath et al., 2019, 2020) and 100,000 design samples for predicting aerodynamic coefficients of transport airplanes (Secco and de Mattos, 2017).

The integration of surrogate modeling with advanced AI methodologies offers new opportunities for handling high-dimensional design spaces. A growing body of research explores the use of images as input to predict physical or performance responses (Li et al., 2017; Cunningham et al., 2019). For instance, images of a 2D linear cantilevered beam as input have been used with Convolutional Neural Network (CNN) to predict the stress field as a picture in an end-to-end surrogate model (Nie et al., 2020). In another example, images of automotive wheels are used in a model to predict modal analysis response from finite elements (Yoo et al., 2021). These approaches offer a promising direction for encoding complex geometric variations in a compact form, thereby supporting surrogate modeling in high-dimensional and nonlinear design scenarios.

In this thesis, surrogate modeling serves as a key enabler for integrating traditional and modern AI techniques to investigate the barriers between these methods and engineering design. This approach aims to establish a more efficient data-driven framework for design science. Traditionally, AI in design has often been viewed as a "black box" where inputs are transformed into outputs with little transparency regarding the intervening processes (Kelliher et al., 2018). This perspective has been critically addressed in this thesis by not only elucidating the mechanisms of creation of both the input and output of these black box models but also by enhancing the management and integration of multiple such models within complex design environments.

2.4.2 Data-Driven Design

The concept of Data-Driven Design (DDD), also known as D3, coined initially in the realm of software development (Ward, 1978) and often referred to as 'data-driven software design' during the 1970s and 1990s, which had the idea of treating the data as the core driver of system design, ensuring consistent definitions and structures before coding the logic around them. The aim was to build software that aligns with a specific domain, guided by insights from experts within that domain (Storer, 1988). The term's initial association with PD emerged within the realm of concurrent engineering (Domazet et al., 1995), characterized by dynamic product models in databases and an enhancement to a process-oriented approach aimed at improving design change management. Over the last three decades, DDD has leveraged quantitative and qualitative data throughout the PD process to inform decision-making and refine solutions.

Finding a well-accepted definition for DDD has shown to be difficult (Johnson et al., 2023). Some scholars emphasize the importance of data in their definitions, such as: "the activities that utilize data as the primary enabler for generating value, including design modeling and design reasoning" (Wang et al., 2022). However, other scholars emphasize the use of AI and data analytics as the foundation of the definition, expressing concern that most computer-aided tasks in design are primarily enabled by various PD data (Feng et al., 2020). Frameworks for DDD usually consist of real-time data sensing and acquisition coupled with data processing and storage units, which then input the developing model and perform data mining on results and knowledge (Zhang et al., 2017).

DDD has been used to enable designers to understand user demographics, engagement, and product performance (Kumar, 2019). The process typically involves a combination of requirement identification, modeling, workflow implementation, simulation, data mining, and evaluation. On the industrial front, it has been argued that by integrating data practitioners into design teams, organizations can bridge the gap between data-driven decisions and creative solutions (Noble, 2024). Companies use DDD to create new or improved products and services in different industries. For instance, DDD is used to build long-term customer relationships in a value co-creation manner, adapt to continuous business reconfiguration, or address societal challenges such as sustainability (Lee and Ahmed-Kristensen, 2023).

DDD is increasingly integrated into PD processes, which enables it to offer opportunities for design optimization and innovation (Quiñones-Gómez et al., 2025; Johnson et al., 2023). However, implementing DDD faces challenges as well, particularly in the early design stages and physical PD (Briard et al., 2021, 2023). These challenges include data selection, availability, and capturing (Langner et al., 2024), as well as the integration of data science with traditional design methods (Liu et al., 2024). The shift towards DDD requires designers to acquire new skills and organizations to adapt their processes (Johnson et al., 2023; Cantamessa et al., 2020). While DDD is commonly used for identifying customer needs through text mining of social media and online reviews, there is potential for broader applications, especially in cyber-physical systems and IoT (Bertoni, 2020) that need to be explored. As DDD evolves, it presents both opportunities and ethical considerations, necessitating further research to fully leverage its potential in design practices (Quiñones-Gómez et al., 2025; Liu et al., 2024).

Although some financial and business products benefit from abundant data, this does not hold true for the engineering design of high-level technical products, which is the focus of this thesis. This discrepancy has brought a focus on how to collect big data in the field of solid mechanics, which has been shown to be a bottleneck for the application of DDD in such products (Ramnath et al., 2019). Another reason for this can be the difficulty in effectively managing and integrating heterogeneous data and knowledge across different phases of product design (Feng et al., 2020). As a design knowledge support tool, DDD can be used for data extraction and design realization (Feng et al., 2020). However, the challenge remains in building the amount of data that is required for design and effectively managing it for different tasks. Adaptation of data science techniques to the specific requirements of engineering design tasks has been proposed to be one avenue to fulfill such gaps in the field (Chiarello et al., 2021b). Researchers raise the question "How to teach solid mechanics to artificial intelligence" (Mianroodi et al., 2021) and assert that the application results in a fast solver that can potentially accelerate the calculation of stress distribution in highly non-linear mechanical systems.

DDD, in the context of this thesis, is an engineering methodology in which data serves as the central enabler for creating value through informed decision-making, iterative refinement, and innovation throughout the product development lifecycle. DDD systematically involves the collection, processing, and analysis of quantitative data to accelerate design evaluation. However, real-world applications face bottlenecks and limitations that constrain engineers' ability to fully exploit this methodology. This thesis identifies three such limitations and proposes corresponding solutions to overcome them. Examples include extracting features independent of geometrical representations, employing dynamic relaxation methods as an efficient alternative to traditional finite element analysis, leveraging these methods to create large-scale datasets, and establishing a platform for effective dataset management and surrogate model execution at multiple levels of product architecture.

2.4.3 Design for and/or with AI

Recent research explores the integration of AI in design processes, proposing frameworks to bridge the gap between engineering design and AI. This integration shifts the role of human designers from direct problem-solving (people who design solutions) to defining the parameters and frameworks that guide AI in finding solutions (people who design problem-solving loops) (Verganti et al., 2020). This replacement of objects in the design challenges existing theoretical frameworks around how we design and necessitates a reevaluation of the principles underlying design and creativity in the AI era.

Moving the human role from being at the center of problem-solving to being guides in decision loops raises concerns about control, interpretability, and accountability in the design process. While AI expands what is technically possible, it complicates how designers define problems, evaluate outcomes, and ensure ethical standards. Critics argue that AI systems often operate as opaque "black boxes," making it difficult to trace reasoning or anticipate unintended consequences (Stoimenova and Price, 2020). It has been shown that there is a methodological gap and a lack of integrated and cohesive frameworks that account for the unique demands of designing with *and for* AI.

In this thesis, Design for AI (Design for Artificial Intelligence (DfAI)) is positioned as an extension of the established Design for X (DfX) philosophy, rooted in the German VDI guidelines. DfX has been viewed as a collection of guidelines that incorporates non-functional requirements, given as additional criteria to be included in concept evaluation, configuration, and embodiment phases of the PD process (Pahl and Beitz, 1984). The aim is to include non-intuitive aspects in the design that influence its life, such as durability, adaptability, reliability, and serviceability (Magrab et al., 2009). In a similar way, DfAI is recently defined as "a set of goals, principles, and heuristics that aim to improve the effectiveness, adoption, and innovation of engineering design and manufacturing AI" (Williams et al., 2022). They argue that for industries and academia to reach this goal, they must collaborate to address challenges such as data and expert shortages, enabling the full utilization of AI technology in design practices. This definition of DfAI as an evolving construct aligns its goals with the objectives of the individual problems studied in this thesis.

In recent years, as scholars have begun questioning whether we should design with AI or design for AI (Stoimenova and Price, 2020), another loose use of this terminology has been proposed. Felzmann et al. argue that design for AI should integrate transparency principles from the beginning of development in the system (Felzmann et al., 2020). This is a way to ensure accountability and ethical integrity in automated decision-making systems. For designers, this approach reinforces the need to embed ethical and interpretative considerations into AI-enhanced design processes, ensuring that technological advancements align with societal values. The concept of the "triangle of shared responsibility" further illustrates this shift, which positions the designer, generative AI (e.g., LLMs), and traditional rule-based software as complementary agents in the design process (Pradas Gomez et al., 2024). In this view, the designer sets the creative vision and defines the task, while the AI contributes adaptive reasoning and context-aware responses. Meanwhile, the traditional code provides stability and deterministic execution of specific functions. This triangle, as a collaborative framework, highlights the need to balance creativity, adaptability, and control in AI-supported

Designing with AI emphasizes the collaborative and interactive role of AI in the design process. Rather than treating AI solely as a tool, this perspective positions 'AI as a design material', where it is a dynamic, interactive element that actively shapes both process and outcome (Yildirim et al., 2022). This approach allows designers to leverage AI's capabilities to enhance system efficiency and develop more innovative solutions. Ultimately, treating AI as a core design material highlights that while AI can automate routine interface tasks, its true value lies in augmenting human insight at deeper system and service levels. Shi et al. systematically review designer-AI interactions, revealing that AI can augment creative processes by uncovering latent user needs, generating diverse design alternatives, and supporting iterative refinement. Their findings underscore a critical shift: rather than replacing human creativity, AI is increasingly seen as a complementary partner that empowers designers to push creative boundaries (Shi et al., 2023). Bagnato explores the transformative potential of AI in design through the concept of the "Artificial Intelligence of Objects." His work illustrates that AI-driven generative processes not only automate repetitive tasks but also open up new aesthetic and cultural dimensions in object design. This perspective frames AI as a tool for reimagining design outputs, enabling personalization, rapid prototyping, and a dynamic dialogue between technological innovation and cultural expression (Bagnato, 2023). Together, these perspectives shift AI from a background enabler to an active design agent, one that broadens how designers think, create, and engage with complexity.

While designing with AI emphasizes collaboration, co-creation, and the augmentation of human creativity, designing for AI prioritizes embedding ethical principles, transparency, and accountability into systems from inception. Realizing these approaches in practice requires careful alignment between theoretical ideals and actual development processes. Two key factors that influence the role and impact of AI in design are the level of integration and the level of automation (Zwingmann, 2023). The level of integration refers to how thoroughly AI is embedded within an environment, and the level of automation indicates how much human intervention is needed for AI to function effectively. Together, these dimensions define distinct categories of AI roles, such as AI as assistants, copilots, autopilots, and agents, as illustrated in Figure 2.9. This framework clarifies the varying degrees of AI involvement in designs, as knowing where a system fits is vital for defining its role.

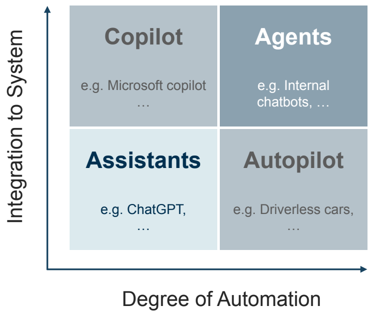


Figure 2.9: Integration and automation in AI framework (Zwingmann, 2023)

Realizing the benefits of designing with and for AI requires bridging theoretical perspectives with practical considerations. Despite conceptual clarity on AI's roles, challenges persist in implementing these ideas effectively in real-world scenarios. Practical obstacles include extended development cycles, insufficient integration of user experience (UX) principles, and the translation of technical outputs into meaningful user experiences (Heier, 2021). Moreover, there is often a gap between theoretical human-centered AI guidelines and their practical applicability, as they frequently lack detailed, context-specific methods. Addressing these implementation barriers—such as enhancing data literacy, fostering stakeholder collaboration, and promoting iterative, user-focused design processes—is critical for effectively integrating AI into design practices and achieving meaningful, user-centric outcomes.

2.5 Data Engineering in Data-driven design

Following discussions on ”Data-Driven Design” and ”Data Engineering in Data Science,” this section aims to define and elucidate ”Data Engineering for Data-Driven Design.” This concept integrates principles from both fields to enhance the capabilities and efficiency of design processes. Here, it is explored how engineering features, labels, and datasets can accelerate design evaluation and improve design outcomes. Previous studies have identified critical synergies between these two fields, highlighting essential tools and algorithms while uncovering significant challenges (Chiarello et al., 2021b). For instance, the innovative use of CAD as input for data-driven methods, the need for meaningful feature representations, the automation of data labeling, and the acceleration of the prototyping process.

Data engineering in data-driven design is a crucial component of modern engineering processes, focusing on the creation, maintenance, and optimization of data architecture, infrastructure, and pipelines (Jain et al., 2023). It involves integrating design theory with data science (Liu et al., 2021) and developing tools for automated data curation (Klettke and Störl, 2021). DE supports decision-making in engineering design by enabling efficient processing of large volumes of data from diverse sources (Petersen et al., 2022). It facilitates advanced analytics, machine learning, and other data-driven operations. The field has evolved through four generations of approaches and is increasingly important in various industries, including automotive systems engineering (Maier et al., 2017; Vlah et al., 2022).

To illustrate how principles from data science can inform and support engineering design, an analogy is made. In 2017, Monica Rogati published a blog post on Hackernoon, drawing an analogy between Maslow's Hierarchy of Needs and data science (Maslow, 1943; Rogati, 2017). This comparison highlights the foundational requirements of data science in a structured manner. Design Hierarchy of Needs is another framework that adapts Maslow's Hierarchy of Needs to design framed by (Bradley, 2010). It prioritizes five levels, from basic to advanced: functionality (does it work?), reliability (is it consistent?), usability (is it easy to use?), proficiency (does it enhance user ability?), and creativity (is it innovative?). A design must meet lower needs before addressing higher ones effectively. Figure 2.10 shows these two hierarchies beside each other.

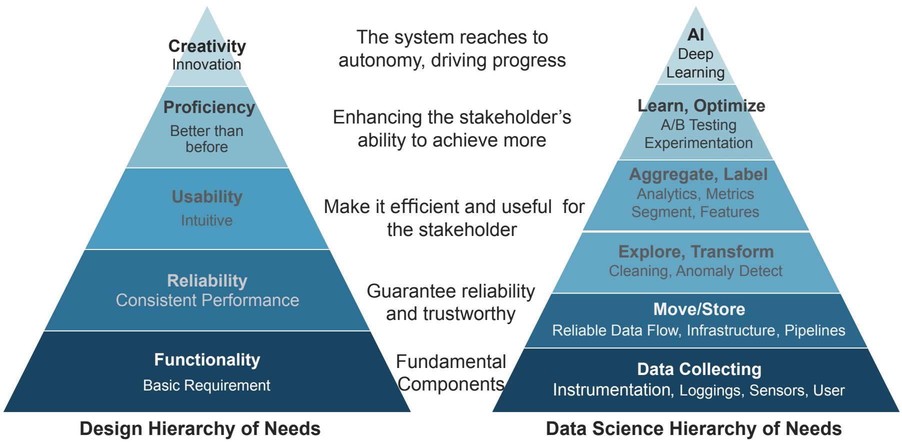


Figure 2.10: Analogy between data science and the design hierarchy of needs

Both hierarchies of needs emphasize a step-by-step progression from foundational requirements to advanced capabilities. At the base, the functionality required in a design is comparable to the data requirement for AI, as both support the smooth operation of their respective systems. Moving up, Reliability in design and infrastructure for moving and storing generated data, both guarantee consistent and trustworthy performance over time. Usability focuses on making a product efficient and satisfying for users; in a similar way, exploring, transforming, and aggregating data provides useful analytics. At the higher levels, Proficiency in design enhances the user's ability to achieve more through the design, and applications offering specific, practical use cases where AI can be leveraged for tasks like automation or decision-making. Finally, Creativity in design mirrors Intelligence in AI, representing the pinnacle of each hierarchy.

2.5.1 Data Mining in Data-driven Design

Data mining in engineering design involves extracting valuable insights from various sources related to design processes and products. Regardless of any possible input data type in ML models - scaler or binary, vector or time series, and matrices or images (Arjomandi Rad et al., 2023) - the data is translated to Real (R) numbers to be used in mathematical computation. Because of their rich data types, simulations are a natural choice for data mining in engineering. In continuous simulations (e.x, Finite Elements), node and shell section information stored in a meshed finite element model can yield input data for analyzing a part's performance after a geometric change (Kuhlmann et al., 2005). Zhao et al. present a framework for data preparation and mining on crash simulation data for studying occupant restraint systems parameters on crashworthiness properties based on attribute importance and decision trees (Zhao et al., 2010). This was reported to reduce the size of the data sets and delete irrelevant features from the data sets, especially in full vehicle model type geometries that have more than hundreds of parameters.

On the other hand, discrete event simulation models are used as cost projectors for estimations in life cycle assessments. Data mining on the history and cost-based features are used in the aerospace industry as tools to characterize cost drivers such as over-performing repair activities (Painter et al., 2006). Such clustering-based simulation mining methods instantiate a vast design space offline. Given new design variants, most similar designs are looked up with a similarity index, and from the simulation results of its 'design neighbors', a behavior valuation for a given simulation is stated without a FEA (Burrows et al., 2011). Generally, simulation data is exploited to learn heuristic connections between the design space and the simulation space, but the effectiveness of this method depends on how well the simulations represent real-world behavior. Bad simulations can lead to bad heuristics. More recently, simulation data mining that uses mesh models has been shown to be effective in assisting designers in tracking design change (Shao et al., 2018).

Graening and Sendhoff suggest several methods for shape mining to enable data mining techniques in engineering design to integrate data across design teams dealing with different simulations, as they argue these techniques are restricted to single design processes and individual design teams (Graening and Sendhoff, 2014). Data mining for such research is more of knowledge discovery by looking at associations (finding dependencies in an analyzed data set), clustering (creating clusters of objects in a way to ensure the highest possible similarity between group members), classification/regression (creating a dependency model between independent variables describing given objects), or description (concise summarizing of analyzed data) (Rogalewicz and Sika, 2016).

2.5.2 Feature Engineering in Data-driven Design

The practical application of AI, machine learning, and data science methods depends fundamentally on the availability and structure of high-quality datasets. In DDD, datasets are systematically constructed to map raw data into structured forms, categorizing information into meaningful *features* and corresponding *labels*. Features in computer-aided design and features in data science are completely referred to as different things. In the data science and ML community, a feature is defined as the numerical encapsulation of the raw data (Zheng and Casari, 2018), serving as characteristics or property of the entities being analyzed (Dong and Liu, 2018). Features serve as structures in a dataset and are meaningful within a scientific or engineering context (Obermaier and Peikert, 2014). Data scientists use algorithms to discover patterns and relationships within mined data to identify patterns, make predictions, or derive insights. Features in data science can be represented as columns in an Excel spreadsheet or as attributes within a dataset in various formats, such as CSV files, SQL databases, or data frames. These features are crucial for training ML models as they provide the necessary information to predict or classify outcomes based on the learned patterns from the data.

On the other hand, there is the notion of *CAD features* or *form features* in the design literature that refer to the fundamental building blocks of a design's form (Shah and Rogers, 1988). These features are essential geometric or functional components of a product, such as holes, slots, bosses, and other standardized shapes that can be combined to create the overall geometry of a product. In addition to the form, features are also defined as entities describing the function (Dong et al., 1991), encapsulating specific engineering significance used to represent attributes and relationships within a part or assembly. Features can also refer to the connection between two parts (Murshed et al., 2010), also known as assembly features like mating and constraints, as well as parametric controllers like dimensions and angles that designers use as relationships that exist between different parts within an assembly. CAD features in parametric modeling allow for easier modifications and optimization of designs, where the relationships among features can be defined to update the entire design when changes are made automatically. Moreover, the terms kinematic features, manufacturing features, and functional features are also introduced in design literature (Cheng and Ma, 2017), which are self-explanatory. User-defined features (UDF) enable users to build their own features in CAD (Tang et al., 2001; Bonde et al., 2022), allowing for greater flexibility and customization in the design process. These features can be tailored to specific engineering requirements or to optimize the manufacturing process, thereby enhancing the functionality and efficiency of the designed products. The definitions of features and the distinctions between Design Science and Data Science are succinctly summarized in Table 2.1. This table highlights how each field approaches the concept of 'features' from its unique perspective.

An overlap between DE in product design and process is the knowledge discovery field that enables understanding a large body of textual datasets. Initially, this included building cyber agents such as web tracers and web organizers to extract needed information for PD (Dagli and Lee, 1997). Such textual features in engineering studies aim to enhance knowledge reuse by introducing computation knowledge extraction in text format from design documents, testing reports, life cycle assessments, customer reviews, sales returns, and so on (Reich, 2005; Romanowski and Nagi, 2001). Semantic literature constructs a tag similarity measure to emulate how humans recall tags from memory. This line of research aims to design information retrieval by utilizing a network of similar semantics (Shi et al., 2017). Features that are being mined are structured information from written historical records (Sexton and Fuge, 2020). Another semantics branch is to analyze sentiments for mining customer requirements in the conceptual design process (Sun et al., 2020; Wu et al., 2022). Features identified here are the emotional tone (positive, negative,

Table 2.1: Definitions for features in data science and design science literature

Data Science	Design Science
Numerical encapsulation of the raw data (Zheng and Casari, 2018)	Form features are fundamental building blocks of a design's form (Shah and Rogers, 1988)
Characteristics or property of the entities being analyzed (Dong and Liu, 2018)	Entity describing both the form and function of a design (Dong et al., 1991)
Structures in a dataset and meaningful in context (Obermaier and Peikert, 2014)	A part feature a shape with specific geometric and topological characteristics, and similarly, assembly feature is a connection between two parts (Murshed et al., 2010)

neutral) expressed within the text. Moreover, there are other kinds of mining-related topics concerning product design, such as reasoning about designs through frequent pattern mining, product design using association rule mining, and text features for mining design rationale, which are all along the same lines. Text mining research in PD continues until today (Park et al., 2023; Yang et al., 2023) with advancements in large language models, but since it goes beyond the scope of this paper, the readers are referred to a review paper (Siddharth et al., 2022) for a comprehensive view on the topic.

2.6 Datasets in Data-driven Design

In the field of computer science, large datasets have been fundamental for improving methods and algorithms. Similarly, datasets, particularly large ones, are becoming increasingly essential for product design and development, especially for testing innovative processes and validating new methodologies. However, designers often struggle to find extensive, well-labeled, public datasets (Arjomandi Rad et al., 2024). This fact has recently been highlighted in several review papers conducted in the field. A compilation of review papers from recent years that acknowledge the challenge of dataset availability is shown in Table 2.2. This compilation echoes a widespread concern regarding the lack of design datasets and sets the context for challenges in dataset acquisition in PD.

Table 2.2: Quotations from recent review papers in the design community

"To shorten long training times, complete and noise-free design datasets created under suitable conditions are required." (Yüksel et al., 2023)
"Most of the research paid more attention to using certain kinds of algorithms to solve certain types of design-related problems, yet seldom has it clarified how to prepare the specific dataset and how to conduct design knowledge-related feature engineering to identify the key design features that are supposed to be learned by the algorithm models." (Yang et al., 2023)
Section on "A Need for Large Multi-modal Design Datasets". "The community should collaboratively construct and maintain expansive design datasets with high-quality labels. This would entail collecting and storing aligned multimodal design data, labeling datasets with design-related attributes, and if available, providing pre-trained embeddings or latent representations, and specifying associated design context." (Song et al., 2023)
"ML-based models can significantly aid in acquiring the massive number of datasets required for typical uncertainty quantification (UQ) procedures, which might not be practical to obtain from simulations and experiments." (Babu et al., 2023)
"We propose that scholars develop standard datasets using design text as a common evaluation platform for future NLP applications." "None of the NLP contributions that we have reviewed in this article leverage a design-specific gold standard dataset for evaluation.", "A gold standard dataset is necessary for NLP applications that aim to measure artifact-level metrics such as novelty, feasibility, and so forth." (Siddharth et al., 2022)
Section on "Datasets", and also writes "Compared to other research fields like Computer Vision, which have massive publicly available datasets, the availability of large, well-annotated, public datasets in engineering is severely lacking." (Regenwetter et al., 2022)
"If the ML-enabled DA is to be attained, larger datasets of real-world designs should be made freely available. Most of the ML algorithms reviewed herein have used training datasets in the order of the hundreds." (Málaga-Chuquitaype, 2022)
"To foster the use of data in the context of engineering design, scholars and practitioners may develop packages especially designed for the ED context, as well as examples and datasets associated to particular methods." (Chiarello et al., 2021a)

Earlier reviews often cited the "need for bigger datasets" as an Achilles heel for DDD methodologies. Design engineers tried to avoid large datasets as they were perceived as less manageable black boxes (Liem, 2007) and calling for reduced-order modeling to address the computational challenges of large-scale statistical problems (Frangos et al., 2010). Yet, as can be inferred from the Table 2.2, the recent discourse reflects a positive shift towards creating larger datasets. More than two decades after the first DDD approaches, data seems to still be a limiting factor in the full-scale applications of ML models (Málaga-Chuquitaype, 2022), but in a different way. This mirrors the paradigm shift in overall AI literature and reflects the recent success of generative models that utilize more complex algorithms and much bigger datasets.

To address this growing challenge, some researchers suggest collaboration between design engineers and data scientists to solve the issue (Chiarello et al., 2021a), and some emphasize the importance of staying lean and small and increasing quality by extracting and managing latent features (Rad et al., 2022). Synthetic datasets are emerging as a viable solution to address the scarcity of real-world data, with guidelines proposed for their creation, annotation, and validation (Picard et al., 2023). There is a cumulative effort for creating guidelines and best practices for data publishing in mechanics and dynamics (Ebel et al., 2025), recognizing its essential role in data-driven engineering design and highlighting challenges, solutions, and examples based on research in AI-supported design tasks (Ahmed et al., 2025).

Chapter 3

Research Methodology

3.1 Research Context

The content of this thesis was developed over the last five years during my studies at two Swedish universities. Figure 3.1 shows a timeline for different activities during this period. The work began with the 'Butterfly Effect' project at Jönköping University in 2019, which lasted for roughly two years. During which papers 1 to 3 were written in the way that is presented in the figure. During the summer of 2022, I got my Licentiate degree from Jönköping University and moved to Chalmers University of Technology to continue my PhD. I was involved in the CHEOPS low power project, which also lasted for another two years, during which papers 4 to 7 were produced in the way that is illustrated in the figure.

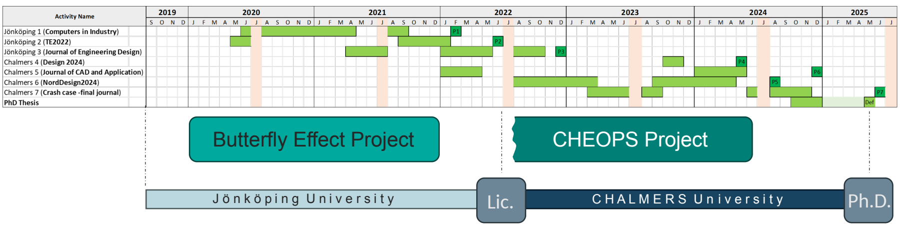


Figure 3.1: Projects involved in the thesis and the timeline for different studies

The Butterfly Effect project was financed through the Swedish Knowledge Foundation (KK-Stiftelsen) with grant number 20180189. The idea for this project was based on an analogy to Lorenz's chaos theory, which aims to forecast unpredictable tornadoes in the future of a system. The Butterfly Effect research project investigated how a small change in the early phases of product development could significantly affect the final product and vice versa.

The Consortium for Hall Effect in Orbit Propulsion System, abbreviated as CHEOPS, was a project funded through the European Horizon project (H2020 project under Grant Agreement 730135). This project aimed to give Europe a competitive edge with respect to technologies related to the design, simulation, and manufacturing of Hall Effect Thrusters, and Chalmers had a small role in it for value and cost assessment for different product architectures designed in the project.

3.2 Industrial Context

It is essential to demonstrate how the developed methodologies and findings of this research have been applied and validated within real-world industrial contexts. Two key industrial sectors involved in this research are the automotive industry, particularly in areas related to passive safety systems such as airbags and crash simulations, and the space industry, specifically concerning fluid management systems. Collaborations within these sectors have provided diverse and challenging application domains to validate the proposed methods and frameworks.

The curtain airbag is an important safety feature that protects people's heads during side-impact and rollover crashes. A CATIA model of a curtain airbag was developed in collaboration with Autoliv AB in the Butterfly Effect project, a top company in automotive safety systems that specializes in airbags, seatbelts, and other safety technologies. This model is used in two case studies, which led to two separate publications that are included in this thesis. Finite element simulations for this model ran with LS-DYNA, the medial axes with Rhino Grasshopper, and dynamic relaxation modeling with Rhino Kangaroo. Figure 3.2 shows all products and cases that have been developed, borrowed, and used for this thesis, and the described Airbag model is numbered a case 1.

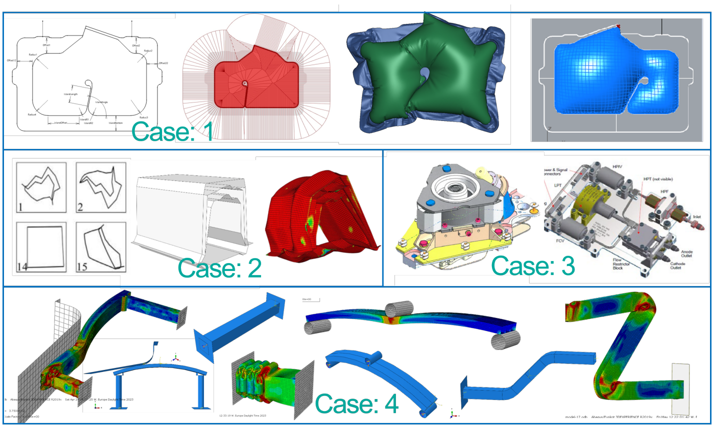


Figure 3.2: Used cases from different industrial context

The crash case studies consist of two distinct analyses. The first case involves investigating the cross-sectional shapes of thin-walled beams for the Toyota RAV4, borrowed from existing literature. These shapes were reconstructed and simulated under identical crash conditions using ABAQUS. Subsequently, the cross-sectional geometries were exported and further utilized in Rhino Grasshopper to facilitate subsequent design exploration and analysis, in a similar way to the airbag case. This case is numbered as Case 2 in the figure. The second crash case (numbered Case 4 in the figure) was the front structure of a car, designed and simulated based on information from the literature. Some of the author's previous publications, completed before starting the PhD, were used.

Case number 3 in the figure relates to the space industry aspect of this thesis, which focuses on Fluid Management System (FMS), specifically related to the CHEOPS project. Due to confidentiality requirements within the space industry, the use of proprietary company data was not possible. Therefore, the FMS datasets are built by considering the actual existing datasets and the underlying physics of the problem. For system-level performance data was collected, comprising 32 thruster results tested under different operating conditions with power ranging from 100 to 2000 W.

Furthermore, I will highlight the contributions of the industrial partners in terms of providing data, domain knowledge, and validation opportunities. Their involvement has been essential to demonstrating the applicability of the research findings to real-world scenarios.

3.3 Research Approach

As briefly mentioned in Section 1.2, A systematic approach for structuring problems and systems is used to break down the problem at hand and explore its core elements (VDI-2221, 1987), which also resembles the V-model in automotive engineering. The overall employed research approach is illustrated in Figure 3.3.

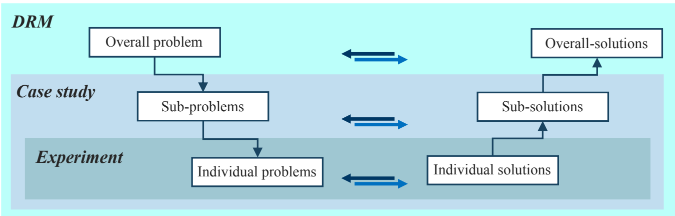


Figure 3.3: Adopted research methods for studying different levels of formulated problems

This figure shows that at a high level, Design Research Methodology (DRM) is employed as a way to organize and guide the direction of the research during the PhD process toward addressing the *Overall Problem*. Beginning from early clarification phases of DRM, the overall problem of engineering design lead time was identified through semi-structured interviews with industrial partners. Additional clarification and descriptive steps reveal that two approaches, design automation and data-driven design, are among the avenues that require exploration.

Case studies are used for further refining the problems and to convey more understanding of the *sub-problems* that are contributing to the overall problem. Together with an industrial partner, it was decided that design evaluation is one of the main contributors to the overall problem and, therefore, was chosen as the main problem for further study. Literature and case studies suggest that a data-driven approach is able to accelerate the design evaluation cycles, but faces challenges, three of which are identified and introduced as *individual problems* in this thesis, see Figure 1.8.

Utilizing the data-driven approach can enhance the analysis and synthesis processes by providing rapid predictions based on previously generated data. While offering substantial speed advantages, it necessitated the availability of extensive, high-quality data. Design automation can be used as an enabler for structuring the datasets required for analysis and enhancing efficiency by minimizing repetitive tasks. However, challenges, such as processing cost, quality, and management of the generated data, limit their utilization and adaptability when applied to complex and evolving design tasks. Therefore, computer experiments are employed to analyze and synthesize these problems that are identified as bottlenecks in addressing the design evaluation sub-problem with a data-driven approach.

The research approach follows a progressive narrowing strategy. Starting from a broad, high-level overall problem, gradually breaking it down into sub-problems, and then further into individual problems considered to be the root causes. To complete the loop on the V-model and validate the findings, the individual solutions that are proposed are validated within their respective case studies. A workshop is designed with industrial partners and experts for validation and verification at the sub-solution level. The validation of the overall solution, as it requires substantial studies and effort, remains for future work. To organize the work during the validation process, Sargent's methodology (Sargent, 1988) is used.

3.4 Research Methodology

3.4.1 Organization of Research with DRM

The work carried out during this thesis is organized using DRM, one of the various methodologies available in design science (Blessing and Chakrabarti, 2009). The framework consists of four stages, as shown in Figure 3.4. The double arrows in the picture show the fact that the researcher is allowed to revisit the stages to make sure the foundation of the research is on solid ground.

The choice of DRM has been partially because it offers a structured and iterative framework for problem identification, development of solutions, and validation of findings. This effectively aligns with this research's aim of reducing engineering design lead time by breaking down this overall problem into sub-problems and individual problems, analyzing them through empirical data collection, and proposing tailored solutions. Additionally, the iterative nature of DRM allows revisiting and refining each stage to ensure robust findings.

The *Research Clarification (RC)* is used in this thesis for clarifying the work to be done and gathering information to support the assumptions (the fact that engineering design

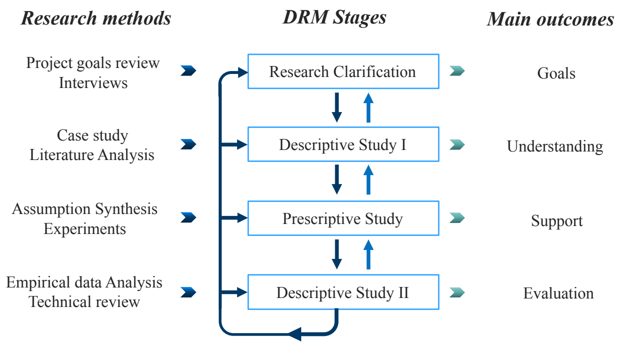


Figure 3.4: DRM framework (Blessing and Chakrabarti, 2009)

lead time is the overall problem). This was achieved by reviewing literature and semi-structured interviews to help explain the research. The outcome of this stage was a description of both the existing and desired situations, which helped us to formulate the overarching aim of the thesis. Additionally, a success criterion for reducing this time is formulated to evaluate the outcome of the research. The measures for success criteria align with the subject under investigation.

In the *Descriptive Study I (DS-I)*, more analyses on empirical data are performed through case studies to investigate influencing factors and to shed light on the current situation. It was noticed that the literature fails to provide answers, and the state of practice falls short. Methods to enhance understanding of the problem, like field observation, are employed at the case company. The gathered data, in combination, provided understanding of the root causes, which are formulated as individual problems, contribute to the sub-problem (see 1.6).

Blessing & Chakrabarti use the term 'support' to represent possible means for improving the process, which are called individual solutions in this thesis. Support can include "strategies, methodologies, procedures, methods, techniques, information sources, software, tools, guidelines, etc., addressing one or more aspects of design" (Blessing and Chakrabarti, 2009). *Prescriptive Study (PS)* was about finding an individual solution for each individual problem and evaluating them. A complete understanding of the existing situation is achieved when moving on with this phase.

Descriptive Study II (DS-II) was about finding out how well the individual solutions work and if they can satisfy the desired success criteria of not only fulfilling the lower level but also the sub-problem level. The generalizability of the provided individual solutions (only two out of three) is investigated by applying the methods to other case studies. In cases where this has occurred, the support's limitations are mentioned, but further investigations are advised to be taken up in future studies. This was mainly because the vision of the desired situation needed to be adjusted accordingly before the tool could be improved.

3.4.2 Understanding the Problem

Semi-structured Interviews

In essence, semi-structured interviews are 'engaging conversations' where the interviewer warmly encourages the respondent to share their feelings and provide unique examples (Merriam and Tisdell, 2015). This approach fosters deeper insights into the challenges they face, moving beyond just surface-level responses. The semi-structured interview features flexible question usage, specific data collection from all participants, a predominant focus guided by a list of questions or topics to investigate, and a lack of predetermined wording or sequence. (Merriam and Tisdell, 2015). It's essential to keep all four aspects in mind, since the primary aim of a semi-structured interview is to create a conversational atmosphere rather than a traditional questioning scenario.

To gain a good understanding of the particular topic of interest, it is important to capture the interviewee's understanding of the topic within their own context. While this may make strict evaluation more challenging, it allows for the inclusion of contextual factors. Challenges and considerations regarding sampling participants (and in general) are discussed in the literature (Cash et al., 2022).

In research involving practitioners in engineering, product development, and production, semi-structured interviews are often utilized to balance rigor and flexibility, especially considering the limited time available to engage with practitioners. Depending on the focus of the study, whether it is preliminary or aimed at detailed validation, different techniques can be employed. A typical process that is also being used in this study can be viewed in Figure 3.5 in the adopted form (Flankegård et al., 2019). In this process, first keywords and phrases were noted when reading the transcripts and then categorized into themes. Next, all quotes exemplifying the codes were copied to a spreadsheet together with the keywords and phrases indicating different themes, and the naming and categorization of the challenges were developed. In the next phase, these categories will be presented at a workshop with the management team and interview respondents at the studied company to receive feedback on the categories.

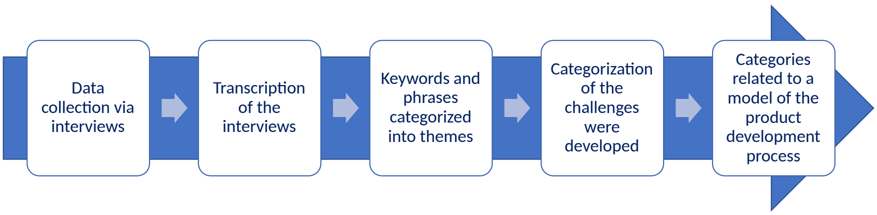


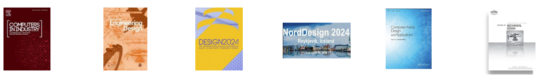
Figure 3.5: Interview process and the analysis afterward (Flankegård et al., 2019)

For this thesis overall, fifteen semi-structured interviews were conducted with experts from a leading airbag manufacturing company (Autoliv AB) to gather qualitative insights. The choice of semi-structured interviews was driven by the need to explore complex, contextual insights into CAD and CAE practices that are not fully understood or documented in the literature. Interviews were conducted over a three-month period, each lasting approximately 60 minutes, and carried out using a digital conferencing tool to accommodate the geographical spread of the participants. One limitation of this approach was the potential for bias in qualitative data, as the interview participants might have had vested interests in promoting specific techniques. Multiple viewpoints were sought to mitigate this by selecting a diverse range of experts.

Case studies

The rationale behind employing a case study approach lies in its profound ability to uncover in-depth insights and complex interactions within real-life contexts. According to Yin (Yin, 2003), a case study methodology is particularly effective in investigating contemporary phenomena within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident. For example, the effectiveness of the data-driven approach could be influenced by various external factors such as the specific simulation tools used, the expertise of the designers, or the type of design problems being addressed. A case study approach allows for the examination of these interdependencies in detail and informs the refinement and validation of the methodology. This depth of analysis is essential for understanding not only whether the methods work but also why and under what conditions they are most effective. Since this research aims to improve design processes by reducing engineering design lead time through data-driven techniques, the case study approach provides a way to observe and analyze how these methods perform when applied within authentic industrial or academic settings.

Four case studies are employed in total to contextualize the overall problem of engineering design lead time. These case studies use empirical data analyses to provide a better understanding and narrow down the sub-problem, namely 'Design Evaluation', to achieve the higher goal. These case studies illustrate how the engineering design lead time, as overall problem defined in the Research Clarification phase (RC), plays out in different cases and how addressing the sub-problem of evaluation of concepts can lead to shorter engineering design lead time. Therefore, to complement the process and refine the problem formulation, three individual problems are identified as hindrances in data-driven design evaluation.



Research Questions			DRM Stages	Paper A	Paper B	Paper C	Paper D	Paper E	Paper F
		RQ1	Research Clarification	●	●	●	●		
	RQ2	RQ1	Descriptive Study I	●	●	●	●		
RQ3	RQ2		Prescriptive Study	●	●		●	●	●
RQ3			Descriptive Study II					●	●

Figure 3.6: Papers connection to research questions and DRM framework

3.4.3 Developed Strategies

The development of strategies for addressing each individual problem is carried out during the prescriptive study phase. By employing the computer experiment as a research method, this phase assesses how various interventions or design changes might affect the performance of the design concept in each case study (Santner et al., 2003). This experimental phase aims to validate the efficacy of each concept under controlled conditions, thus providing empirical support to the theoretical findings derived from the case studies. The interplay between case studies and experimental methods enables the combination of qualitative depth with quantitative rigor. This dual approach ensures that the conclusions drawn are both contextually informed and empirically validated.

Individual problem one was studied on two case studies that have resulted in papers A and E, In the same way, individual problem 3 studied with two case studies resulted in papers D and F. Paper B is the result of a case study with individual problem two and paper C is a review paper that studies opportunities and challenges with datasets in design. Figure 3.6 shows how different outcome papers are connected to different steps in DRM methodology as well as proposed research questions in this thesis.

3.4.4 Validation and Verification Methods

The primary validation framework used is proposed by (Sargent, 1988, 2010, 2020). This approach reinforces that the proposed solutions are effective and efficient concerning the problems and constraints of the thesis, while also adhering to the established standards of academic rigor. The motivation for selecting Sargent's model stems partially from its methodical approach to validation, and also partially because of this method's roots in the modeling and simulation field. Of such reasons, this validation method is highly suitable for ensuring the veracity and applicability of the modeling and simulation work in this thesis. This methodological model (called Sargent's model in this thesis) supports the reliability of the models, simulations, and research findings within four distinct parts illustrated in Figure 3.7.

In this thesis, *Conceptual model validity* is performed to determine that the theories and assumptions underlying the conceptual models are correct and that the model's representation of the problem entity is 'reasonable' for the intended use of the model. This has been done using experts' assessments on models' usability and applicability.

Computerized model verification is used to ensure that the computer programming and implementation of the conceptual models are correct. This has been performed following the Sargent's suggestion for structured walkthroughs/review the code line-by-line with peers, trace the model's execution to detect errors, and dynamic testing of the model under various conditions to verify outputs.

Operational validity has determined that the model's output behavior has sufficient accuracy for its intended purpose or use over the domain of the model's intended application. Sargent suggests accomplishing this by comparing the model's outputs (e.g., performance metrics) to historical data and other validated models by employing similar techniques.

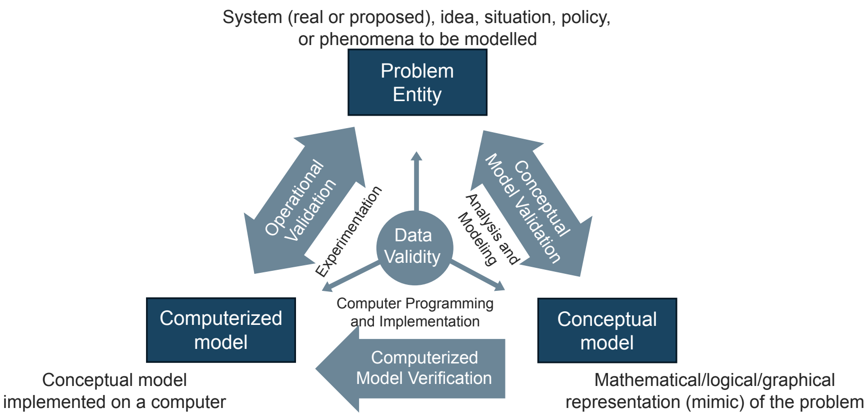


Figure 3.7: Model Validation and Verification by (Sargent, 1988)

An example of graphical comparisons is given to visually assess outputs using histograms or scatter plots.

Data validity step is used to ensure that the data used for model building, model evaluation and testing, and conducting the model experiments are adequate and correct. This is done, for instance, by collecting data relevant to the design task. Test data for accuracy using different error metrics. Screen for outliers, evaluate prediction results with various models.

Although Sargent's model meets classical validation questions regarding effectiveness (is the model designed correctly?) and efficiency (is the model designed right?), it falls short when applied to processes and design process models. For instance, Conceptual Model Validation checks if the simulation's theories are correct for the real system, but it does not focus on the logical consistency of the method itself outside of that context. Operational Validation confirms that outputs match real-world data, but it does not explicitly assess whether the method's structure remains sound across diverse cases. These shortcomings are considered when applying Sargent's methodology to assess findings. Logical consistency of the findings outside the case studies can be investigated by applying the developed supports to a different case. This thesis checks the operational validity of the models by determining that the model's output accuracy remains sufficient for its intended purpose across diverse cases.

Workshop

As mentioned, conceptual model validity in Sargent's model is determined by assessing experts' opinions on the developed models. The assumption is that if the model is deemed applicable and useful by experts, it can be argued that it fulfills its intended purpose. A workshop consisting of three presentations on individual solutions, each lasting about 20 minutes, is designed. Following Sargent's definition, the first goal of the workshop is to assess if experts think the developed supports represent the problem entity. Therefore, before presenting what has been developed, a five-minute survey was conducted to assess how much participants think the problem exist, and therefore how much support is lacking. Figure 3.8 shows the designed workshop that consists of three rows.

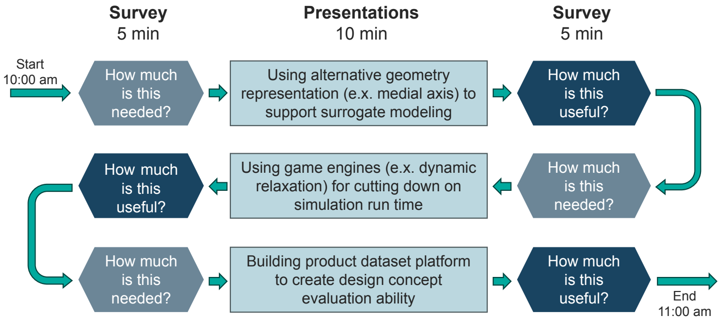


Figure 3.8: Validation workshop with the case company

After the first survey, which inquired about the need, a 10-minute presentation was provided to show how each individual solution addresses that need. This is followed by another five-minute survey that gathered feedback on how useful and applicable they found the support. Almost 10 people participated in the workshop, of whom nearly half had also participated in the interviews conducted earlier. The results of this workshop are presented in the discussion chapter.

In some cases, additional validation checkpoints at the detail level are considered during the work process. However, it is not feasible to provide detailed information about every step. For instance, the willingness of respondents to serve as reliable informants affects the validity of the data collected in semi-structured interviews. In this study, triangulation is considered as a method for such cases. This approach involves not depending on a single source for information and facts but instead validating them through multiple data sources. This is a way of assuring the validity of research through the use of a variety of methods to collect data on the same topic, which involves different types of samples as well as methods of data collection (Yin, 2011). This was accomplished by participating in some internal meetings of the companies and conducting observation sessions with engineers working on the problematic tasks. By comparing conclusions from these observations and interviews, data collection was stimulated with triangulated insights, which provided more trustworthiness for the research.

Chapter 4

Summary of papers

This chapter summarizes six appended papers concerning the problem formulation presented in the Introduction section, as well as their contributions to the research questions. As mentioned, the thesis examines three individual problems in a nonlinear chronological order. To understand how these papers relate to one another, Figure 4.1 illustrates the three problem areas and the sequence in which the papers were produced. This section is written with respect to this illustrated order and will not follow a linear timeline.

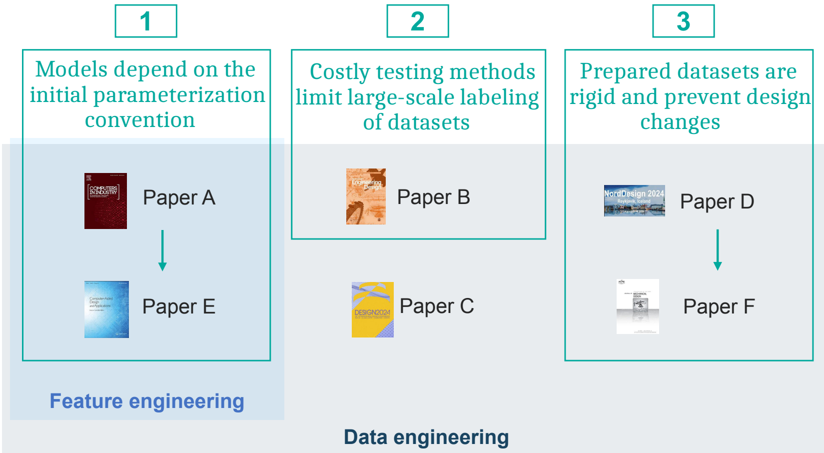


Figure 4.1: Mapping the papers and the individual problems, see Figure 1.8

4.1 Feature engineering on airbag case, Paper A

This paper is the result of the first study performed and is a mix of 'Research Clarification', 'Descriptive Study I', and a 'Prescriptive Study' in DRM methodology. It attempts to answer

RQ1 by presenting a challenge with data-driven design evaluation, RQ2 by highlighting the importance of fast evaluation in the design process, and RQ3 by showing a feature extraction methodology and applying it to the design evaluation of a side airbag. The paper shows that the offered method reduces the evaluation time of a design concept and is, therefore, can be successful in impacting the selected success criteria.

This paper identifies challenges in implementing a prediction tool within the design process, specifically focusing on dimensionality and parameterization, based on a literature review and workshops. A case study on side airbags demonstrated that fully defining a CAD model requires specifying numerous parameters, which can introduce complications. A parametric analysis of these CAD parameters revealed substantial variation in how each one influences the airbag's final volume. This underscores the complex and interdependent relationships among the parameters. Notably, none of the individual CAD parameters showed a direct correlation with the final volume, which may explain the limited accuracy of prediction models trained solely on these inputs.

The search for features that have low dependence on CAD parameterization and high correlation with output commences by calculating the medial axis of the geometrical 2D shape of the airbag. The medial axis, often referred to as the shape's "skeleton," is a method of reducing a shape's complexity by capturing its fundamental structural features without retaining all geometric details. This is achieved by computing the set of all points having at least two closest points on the object's boundary, effectively creating a simplified internal representation of the object. Figure 4.2 shows how the medial axis was generated in Rhino/Grasshopper. The Vonronoi component is used to create circles at equal distances on the edge of the geometry, and then by increasing the radius of these circles and making them create a boundary, the medial axis is generated.

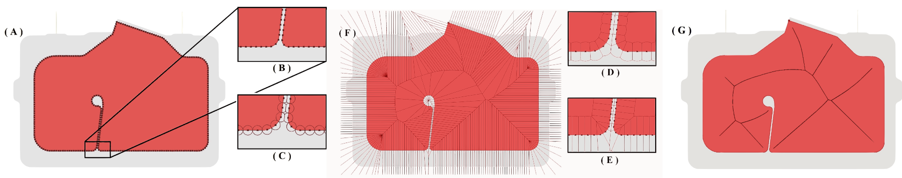


Figure 4.2: Medial axis extraction process from airbag 2D geometry

The medial axis is a geometric entity like other entities such as area, circumference, etc, but is unique for an arbitrary design shape. The paper demonstrates that the medial axis, as an alternative method of representing geometry, can be used to extract high-quality features for prediction analyses. For instance, the length of the medical axis and the sum of the circumferences of all circles inscribed in the design shape are extracted as new parameters. These geometric entities, as new parameters, are referred to as *sleeping parameters*, and are studied as a performance indicator for the inflated curtain airbag. Table 4.1 demonstrates that these new parameters, calculated independently from CAD model parameterization, have a significantly better correlation with the volume output.

Table 4.1: Correlation between CAD and Sleeping parameters with the output

Name of the parameter	R2 Correlation with the output
No. 1 (CAD parameter)	0.037
No. 13 (CAD parameter)	0.018
No. 12 (CAD parameter)	0.0441
Area	0.829
Length of the medial axis	0.752
Sum of circumferences inscribed	0.881

Regression analyses were conducted to assess and validate the performance of extracted parameters within regression modeling. The findings confirm that sleeping parameters enable designers to create straightforward yet precise regression models that utilize fewer features and sample points. A Support Vector Regression (SVR) model was also trained to illustrate that transitioning from basic linear regression to more advanced algorithms does not significantly improve prediction accuracy compared to the benefits provided by sleeping parameters. By enhancing the quality of training features, sleeping parameters enable satisfactory results with Multiple Linear Regression (MLR), eliminating the necessity for complex regression algorithms like SVR or a higher number of training samples.

Table 4.2: Error of trained models using CAD parameters vs sleeping parameters

	Multivariate Linear Regression		**Support Vector Regression**	
Accuracy of the regression model among predicted and expected sets	All 14 CAD model parameters	Selected 3 Sleeping parameters	All 14 CAD model parameters	Selected 3 Sleeping parameters
R2	0.6318	0.9505	0.8027	0.9544
MSE	14.7304	1.8802	14.3419	1.7784

The proposed methodology is transferable to all volume simulations in airbag models that use 2D geometries as inputs, such as knee and side airbags that deploy from the passenger seat. Other inflatable structures that require volume simulation can benefit from the findings of this paper, such as high-pressure vessels, inflatable tunnel plugs, inflatable rubber dams, different kinds of inflatable boats, etc. It can be argued that this methodology is scalable to any performance evaluation that requires good enough accuracy but fast evaluation for decision-making in the early stages of the design process. In the following section, another case study will be introduced to examine if this method applies to other simulations that utilize 2D shapes as inputs.

4.2 Feature engineering on crash case, Paper E

This paper addresses the first individual problem, similar to Paper A, see Figure 1.8, and expands on the proposed solution. It restates the problem and therefore briefly touches on

'Descriptive Study II' but mainly involves 'Prescriptive Study' by introducing new type features that can be extracted, and 'Descriptive Study II' by an attempt to assess the generalizability of the already discussed solution in another industrial case study. The main aim is to offer an organized category of extractable features and demonstrate the method's applicability in another industrial case, thus it attempts to answer RQ2 and RQ3.

In paper A, 'sleeping parameters' are defined in contrast to conventional CAD parameterization as engineered features that are coupled to the geometry of the design but are independent of the geometry creation process. The term *sleeping* emphasizes that these features, while not immediately visible or conventional, have a potential utility that can be awakened through appropriate processes. Unlike *latent*, sleeping parameters can be constructed, extracted, selected, and then processed even if the geometry undergoes drastic changes. This process justifies borrowing and using the feature engineering terminology from the data science field.

Any method's true potential and versatility are revealed when challenged in diverse environments. Therefore, the previously introduced medial axis concept is employed in the analysis of structural components like Thin-Walled Beams (TWB), given their role in vehicle safety and performance. Many automakers use a repetitive design process to evaluate the crash performance of these beams. Thus, applying feature engineering (feature construction, feature extraction, feature selection, and feature processing) to the predictive modeling and analyses in such a case study can lead to an order of magnitude reduction in engineering design lead time. Figure 4.3 illustrates three distinct beam cross-sections utilized in the Raw4 Toyota body frame.

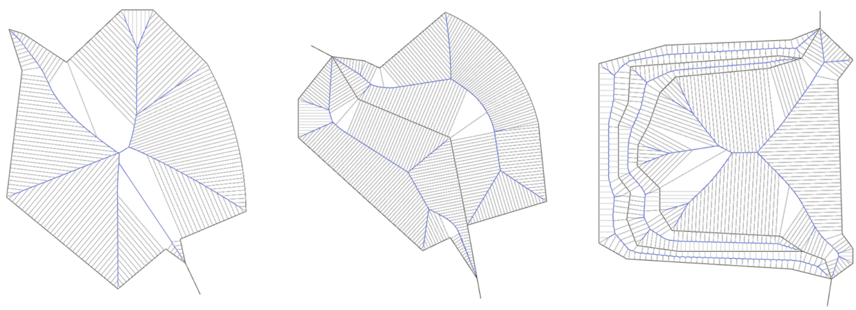


Figure 4.3: Medial axis extraction for three TWB cross sections

In these three figures, the blue line in the middle represents the extracted medial axis, while the gray lines are part of the Voronoi cells used in the process. The features extracted from the medial axis are informed by the beams' crashworthiness and are extractable at any point after the design process. This paper presents various features, such as the number of handles, branching points, and other properties that can be calculated from the medial axis. Table 4.3 presents these extracted features along with their correlation to two crashworthiness outputs, SEA (Specific Energy Absorbed) and PCF (Peak Crushing Force), which serve as performance indicators for the design of these beams. This paper further explores three distinct correlation analyses (linear and non-linear) to ensure that the identified relationship is not reliant on the kind of analysis used.

Table 4.3: Correlation between mined features and two FEA outputs.

	Linear regression score		**Pearson correlation**		**Spearman correlation**	
	SEA	PCF	SEA	PCF	SEA	PCF
Length of Medial Axis	0.83	0.89	-0.91	0.94	-0.93	0.92
Width Information	0.26	0.23	-0.50	0.48	-0.57	0.58
Num. of Handels	0.68	0.77	-0.82	0.88	-0.81	0.81
Branching Points	0.70	0.79	-0.83	0.89	-0.86	0.85
Shape Perimeter	0.87	0.98	-0.93	0.99	-0.99	0.98
Avg. Circle Radius	0.34	0.30	0.58	0.54	0.58	-0.58
Shape Compactness	0.57	0.52	0.75	0.72	0.74	-0.74

The extracted features are derived from the geometric and statistical properties of the shape, and such correlation analyses are a means to rank them for prediction tasks. For instance, the length of the medial axis, the number of branching points (indicative of the shape's complexity), and the perimeter of the shape are shown to be good indicators of crashworthiness. For the length of the medial axis specifically, Pearson correlation values of -0.91 for SEA and 0.94 for PCF were observed.

To explore and understand the diverse impacts of possible geometric properties from the medial axis, this paper categorizes features into region-based, fractal-based, and boundary-based, shown in Table 4.4. This approach facilitates feature selection for machine learning by guiding engineers in selecting the most relevant features for various analyses and ensuring that models are built with the most independent and impactful data, thereby driving more accurate predictions and informed design decisions.

Table 4.4: Categorization of features beyond traditional feature selection in CAD

Feature Type	**Examples**	**Description**
Boundary-based	Perimeter, convex hull, radius of gyration, Euler number, profile, bounding box parameters	Extracted directly from the shape's boundary
Region-based	Area, mean intensity, Eccentricity (elongated or stretched) variance, entropy, texture, Central moments, Hu moments, or Zernike moments can capture shape properties, compactness	Extracted from the interior of the shape
Fractal-based	Fractal dimension, Skeletonization Features (length, branches, loop handles), eigenvalues and eigenvectors of the covariance matrix of shape points, Tortuosity (Measuring the "wiggiliness" of the medial axis), Angles between Branches	Describes the self-similar structure of the shape

One of the most effective features identified and used in both Paper A and E is the *Medial Axis Length*, which can give an idea of the complexity and the extent of the shape. For example, a longer medial axis might correlate with higher energy absorption capabilities as it indicates more extensive internal structuring, which could be beneficial in crash scenarios. Paper E introduces the concept of fractal base features, which are directly related to how Voronoi cells are distributed inside the shape and capture the complexity and detail within the shape. *Number of Branch Points* refers to the number of points where the medial axis splits, indicating complexity and potential stress concentration points, and is another fractal-based feature identified.

Average Circle Radius measures the average radius of circles that can fit within the different parts of the geometry, indicating the uniformity and symmetry of the area. *Width Information* sums or averages of the radii, providing a measure of the overall dimensional spread across the design. Region-based features such as these examples are crucial for models that predict structural integrity and deformation patterns. For instance, a larger average circle radius might indicate larger, more uniform areas that could behave differently under stress compared to areas with smaller, more varied radii. In crashworthiness, such uniform regions might deform more predictably, aiding in better energy absorption predictions.

Perimeter is the total length of the shape's outline, providing a simple measure of size and scale and *Number of Handles* refers to the count of distinct 'loops' or holes within a shape, which can indicate complexity and features like cavities or enclosed spaces. Boundary-based features can significantly impact the predictive modeling of structural behaviors. The perimeter can be used to assess the overall material usage or structural boundary conditions, which are crucial in defining how a shape might behave under external forces. The number of handles affects the flow of forces through a structure; more handles might imply more points for energy distribution, influencing how energy is absorbed in a crash.

4.3 Large-scale labeling of dataset, Paper B

This paper reuses the case study as presented in Paper A, to address another individual problem. Therefore, it touches on the Research Clarification by means of a literature review and then combines 'Descriptive Study I' and a 'Prescriptive Study'. The proposed support satisfies the overall success factor (engineering design lead time) in the case study with limitations. The results provide insights for RQ1 by the results of the literature review, RQ2 by presenting the change in the current practice, and RQ3 by implementing the prescribed changes and the support.

A computational method known as dynamic relaxation is discussed in the literature, which has primarily been used for simulating unstable structures in the past. This method requires several iterations to converge, yet the iterations are computationally inexpensive since there is no need to assemble a stiffness matrix. Nonetheless, the discretization and some steps remain time-consuming. In this paper, an implemented version of dynamic relaxation in a component called Kangaroo within Rhino/Grasshopper is utilized to produce labels for 60,000 CAD models. A methodology for creating a large quantity of labeled data using dynamic relaxation is proposed, which can be applied to a wide range of simulations. Figure 4.4 shows volume visualization in Rhino, which is achieved by using the Kangaroo component in Grasshopper.

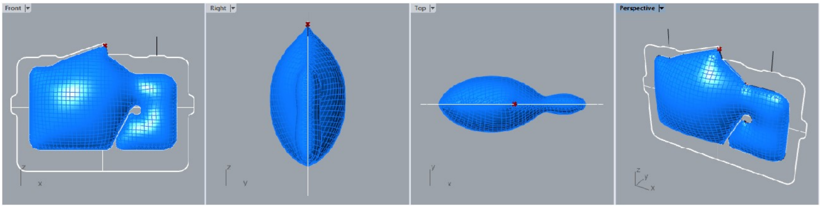


Figure 4.4: Volume simulation visualization in Rhino

Figure 4.5 shows the schematic of the dataset created from two sizes of images as well as their labels which are volumes of the associated geometry. This dataset has been made available to the public in an online repository (Arjomandi Rad, 2022). Most of the machine learning datasets today are based on real-life problems and literature can benefit from engineering datasets like the one presented here for benchmarking purposes.

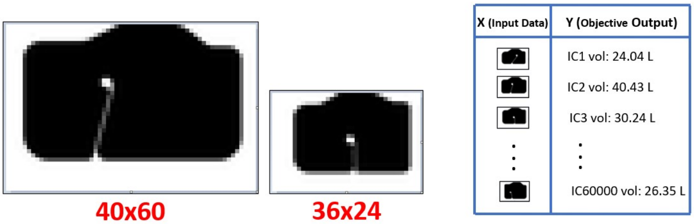


Figure 4.5: The dataset consists of 60,000 labeled images with two sizes

Moreover, an off-the-shelf implemented CNN with three layers is chosen for the training process. The training was performed with 128 a batch size and 50 epochs. As shown in Figure 4.6, the loss value, which represents the summation of the errors in our model (calculated from the cost function) for each case in the testing dataset, is shown on the left. The figure on the right shows accuracy, the percentage of correct predictions applicable only to classification tasks. From the picture, It can be concluded that there is minimal gap in accuracy between training and validation, both converging to roughly 90%. In the loss graph, both training and validation are decreasing effectively and steadily, with an acceptable gap between them (referred to as the generalization gap) across each epoch, ultimately converging to 0.2, indicating a successful model.

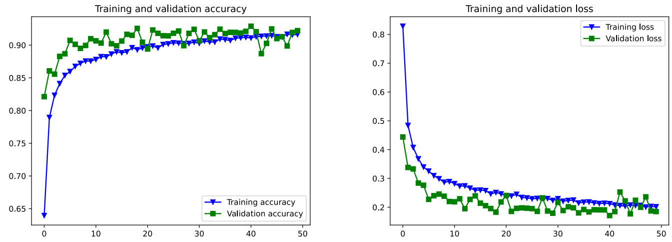


Figure 4.6: Accuracy (Left) and loss (Right) of training and validation dataset

Later, the network accuracy was tested using 10000 new samples that were acquired separately within a similar but separate sampling process. The testing showed 89.42% accuracy and 0.26 loss, which means that from 10000 testing cases, 8943 cases were predicted correctly and 1057 cases were placed in the wrong bin.

4.4 Review of datasets in engineering design, Paper C

After working with features and labels as components of the datasets, this paper comes as a result of the first study after licentiate and therefore includes a revised and broadened 'Research Clarification' and 'Descriptive Study I'. The paper is an attempt to capture the bigger picture by shifting the focus from feature engineering to data engineering as the sub-solution is delivered, and thus revisiting the RQ1 and RQ2 at the beginning of the second part of the PhD process.

The transition from focusing solely on features to addressing the broader data ecosystem is essential for enhancing predictive modeling and analysis in design engineering. Data engineering extends beyond the extraction and optimization of features; it involves the holistic management of the data lifecycle, encompassing acquisition, validation, storage, protection, and processing of data.

The literature in design datasets is growing fast, and by the time of publishing this thesis, the performed literature review is already outdated. A total of 25 design datasets were identified by the publication time of the paper and organized into six categories shown in Figure 4.7. These categories pinpoint their applications at different phases of product development, from conceptualization to refinement. This sets the stage for discussing the overarching need for more comprehensive and scalable datasets that can adapt to the evolving demands of the design industry.

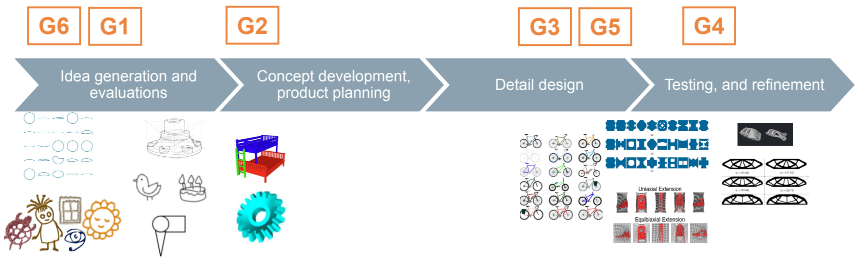


Figure 4.7: Existing design datasets landscape illustrated diversity and the gaps

The paper discusses the challenges in creating design datasets. Issues of data sharing and privacy are paramount, with companies often reluctant to disclose proprietary information that could benefit broader research and development efforts. Moreover, ensuring the quality of datasets is another hurdle, as data often comes from varied sources and may be incomplete or noisy. This necessitates sophisticated data engineering strategies to clean, preprocess, and standardize data, ensuring that it can effectively train AI models.

The economic aspect of data engineering is another issue within the field. Recognizing that acquiring and analyzing large, high-quality datasets can often be cost-prohibitive, particularly for small and medium-sized enterprises. The paper suggests that overcoming these barriers may require innovative approaches, such as the development of synthetic datasets. These datasets, generated through algorithms, can mimic real-world data, offering a practical solution for testing and validation without the associated costs or privacy concerns.

The potential of end-to-end datasets in covering a broader spectrum of the design process is discussed. As the current datasets focus on, for example, identifying customer needs or finalizing design solutions. This approach to building datasets across different phases of product development not only bridges the gap between theoretical and practical aspects of design but also fosters a more integrated development process.

In the concluding sections, the role of AI and specifically Generative Pretrained Transformers (GPT) is discussed. While AI has shown promise in generating innovative design concepts and ideation, its application in quantitative and engineering-specific tasks remains challenging. GPTs' abilities for building design-related synthetic or end-to-end datasets are promising due to their capacity for comprehensiveness, adaptability, making them good at understanding complex design requirements and acquiring large amounts of information in a short time. Additionally, GPTs can assist in labeling, augmenting, and enriching datasets, thereby contributing to more robust and scalable AI-driven design evaluation processes. The paper underscores the need for continued research and development in AI capabilities to fully leverage its potential in creating and utilizing dynamic, comprehensive datasets in design.

4.5 System analysis on component datasets, Paper D

This paper is the first step in addressing the third individual problem, see Figure 1.8, studied in this thesis. It includes a mix of 'Research Clarification' and 'Descriptive Study I' as it explores the last problem primarily through a literature review and previously conducted interviews. The paper aims to capture the bigger picture by shifting the focus from feature engineering to data engineering, thus contributing to RQ1 through the results of the literature review, to RQ2 by proposing a change in current practice, and to RQ3 by providing an initial version of support.

In traditional design processes, disparate teams often choose subsystem technologies independently. This fragmented approach can lead to lengthy design cycles, integration challenges, and suboptimal technology choices due to a lack of comprehensive performance data. Therefore, if designers in the early phases could analyze product performance, many design cycles could have been avoided.

At the core of the proposed framework is functional decomposition through a Function-Mean (FM) tree method, which breaks down the primary function of the system into a hierarchy of subfunctions and associated design means. This structured representation not only clarifies the relationships among subsystems but allows for the mixing and matching of different sub-functions and means to find ideal architecture at an abstract level. This hierarchical structure serves as a blueprint for constructing a modular ensemble of datasets. Each node in the FM tree represents a specific function or design solution, and data collected from simulations or experimental tests are organized to align with these nodes, illustrated in Figure 4.8.

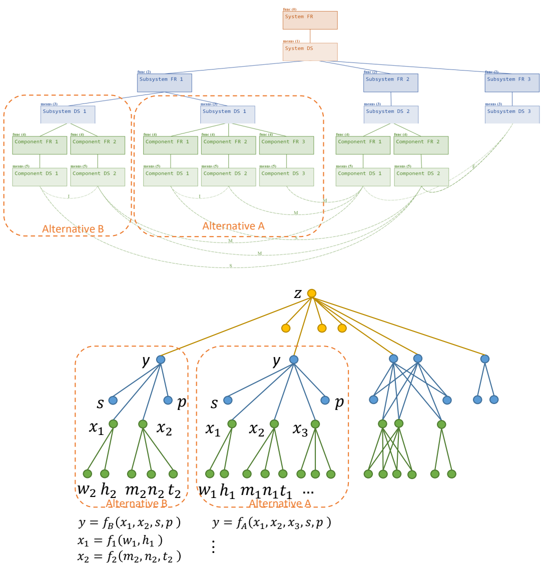


Figure 4.8: Mapping functional decomposition to system-level dataset ensemble

By establishing shared interfaces between these modular datasets, surrogate models can be developed at multiple levels of the design hierarchy. The connected ensemble of datasets thus enables quantitative analysis by linking low-level input parameters (such as component dimensions or operational flow rates) with high-level performance outcomes (such as thrust or energy absorption). This structure allows the overall system behavior to be predicted by aggregating information from the various modules in a coherent and scalable manner.

The proposed support is demonstrated through a case study involving the FMS of Hall Effect Thrusters (HET), illustrated in Figure 4.9. This is a complex, multi-physics system used for satellite propulsion. To generate quantitative predictions of the thrust (system level output) without resorting to high-fidelity simulations at early design stages, surrogate modeling techniques are employed. A dataset comprising performance data for 32 different thruster designs was compiled from the literature. Surrogate models were trained using key component parameters (e.g., anode flow rate, discharge voltage, and current) to predict system-level outputs. A leave-one-out approach trains each surrogate model with one design as the test set, using the others for training. In this study, machine learning models, specifically Gaussian Process Regression models are used. This captures diverse operating conditions and technology variations, which yields approximate system-level performance predictions. Ensemble techniques, including stacking and concatenation in frameworks like Keras, integrate component-level models with system-level predictions, enhancing accuracy and facilitating better design space exploration.

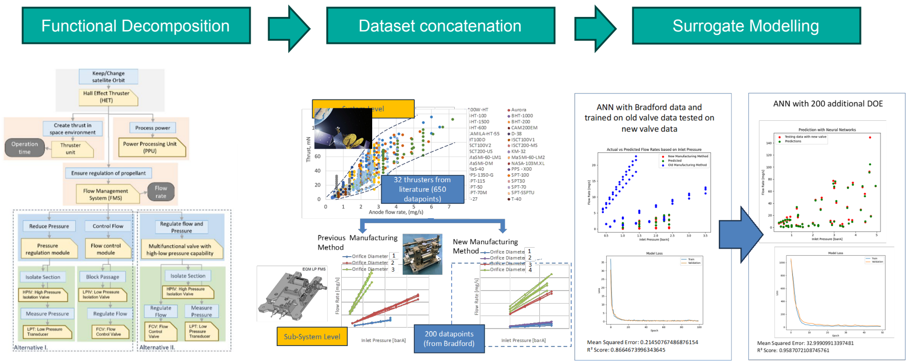


Figure 4.9: HET case datasets in coupled functional and component data

The ensemble approach achieved an average root-mean-square error (RMSE) of 2.85, a mean absolute percentage error (MAPE) of 11.25%, and an R^2 of 0.77. These metrics indicate that the method effectively utilizes low-level operational data to yield reliable predictions of overall system performance. Furthermore, the framework is extended to evaluate subsystem-level performance by analyzing alternative FMS architectures. Neural network models, enhanced with regularization techniques, were trained on datasets from one manufacturing method and tested on data from a different method. The smooth convergence of the loss function and acceptable error margins observed during these experiments demonstrates that the ensemble approach is robust and capable of accommodating design variations.

4.6 Product dataset platform with crash case, Paper F

The most recent paper is a mix of 'Prescriptive Study' and briefly 'Descriptive Study II'. It expands on the proposed solution in Paper D and addresses the shortcomings in demonstrating the support's effectiveness. The paper, therefore, primarily responds to RQ2 by showing how the current practice could be customized and to RQ3 by proving the newer version of the previously introduced support.

The paper presents a novel approach to bridge the gap between early-phase conceptual design and detailed performance evaluation in automotive structures. Recognizing that early design often relies on low-fidelity, qualitative methods due to the prohibitive cost of high-precision simulations, the study proposes the development of a modular dataset platform. This platform is built upon a product model hierarchy that mirrors the functional decomposition of the product architecture. In particular, the integration of function-mean (FM) modeling with datasets acquired from Finite Element (FE) analyses enables a more rigorous, quantitative evaluation of design concepts at the system level. By leveraging shared features extracted from the FM tree, the methodology provides an avenue for front-loading detail-level analysis and supports rapid evaluation of both incremental and radical design changes.

The methodology is illustrated through a detailed case study focused on the front structure of a car, a system comprising key components such as a bumper and a crash box. An FM tree (illustrated in Figure 4.10) is employed as an early design modeling tool to decompose the overall function of a car front structure into hierarchical subfunctions and corresponding design solutions. Each node in the FM tree represents a specific function or design alternative (for example, the energy-absorbing role of a crash box or bumper).

Data obtained from high-fidelity FE simulations, covering different design parameters such as geometry, material properties, and boundary conditions, are mapped onto the FM tree structure. The performance metrics from the component level (such as Energy Absorption and Peak Crushing Force) as additional features for surrogates at high levels serve as *sleeping parameters*. This means that these shared interfaces link lower-level component datasets to the overall system performance dataset. The resulting platform of datasets (illustrated in Figure 4.10) serves as a modular platform that facilitates quantitative analysis by enabling the training of prediction models that capture the influence of component-level innovations on system-level outcomes.

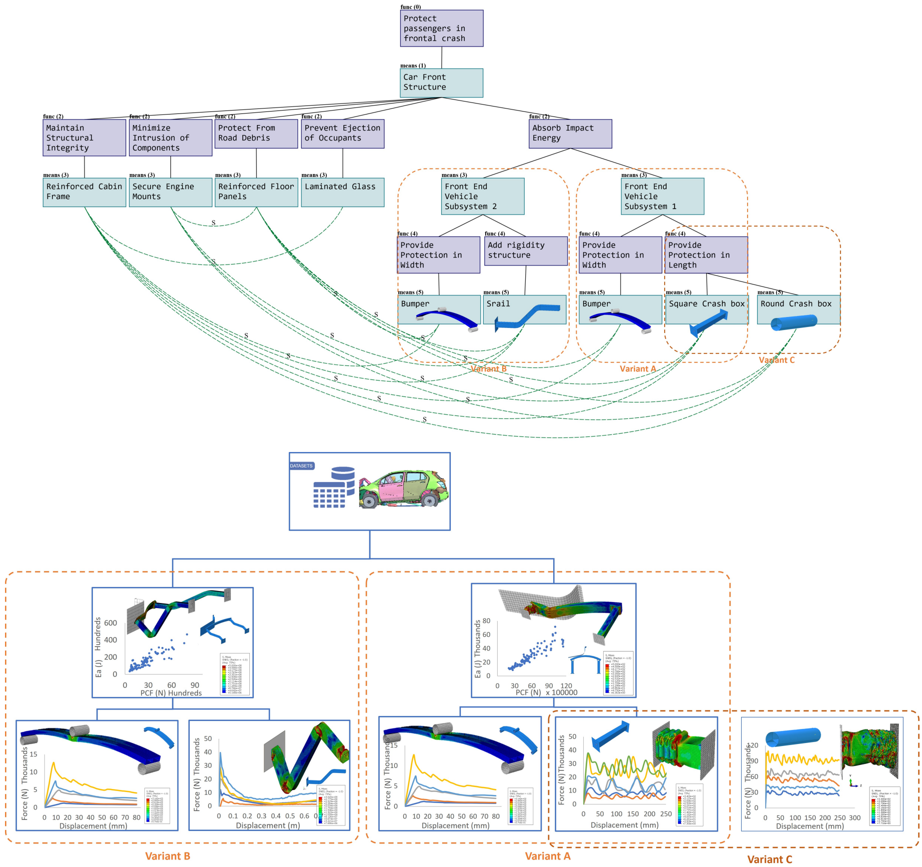


Figure 4.10: The assembled dataset platform for car front structure (Right), Functional decomposition of a car front structure (Left).

The assembled dataset platform is used to train several machine learning models to predict system-level performance indicators. Two distinct types of design modifications are evaluated: modular changes (transitioning from a square crash box to a multi-cell round crash box, denoted by the change from variant A to C) and radical changes (replacing a traditional component with an s-rail configuration, denoted by the change from variant A to B). The results demonstrate that, by leveraging the shared interface provided by the FM tree, the prediction models can roughly estimate key performance outputs. Error metrics such as mean absolute error, mean absolute percentage error, root mean square error, and R^2 values are used to assess the accuracy of the predictions. In the case of modular changes, the models exhibit satisfactory performance, while predictions for radical design changes reveal the challenges associated with extrapolating beyond the original design space.

In addition, similarity analysis is employed to quantitatively assess how closely new design variants align with the legacy configuration. By comparing the extracted feature representations, derived from the shared interfaces of the FM tree, the similarity metric provides an index that helps determine whether a variant falls within the reliable prediction domain of the trained models. Lower similarity distances indicate that the design remains within the incremental modification range, while larger distances suggest that the changes are more radical, potentially challenging the accuracy of the predictions. This analysis is shown in 4.11 and serves as a valuable tool to support informed decision-making by clearly distinguishing between incremental and radical design modifications.

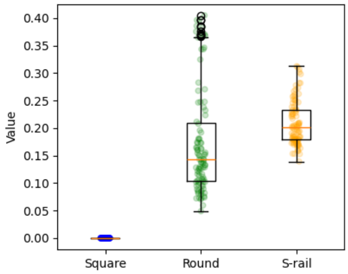


Figure 4.11: Similarity analysis of alternative technologies in the architecture

Chapter 5

Discussion and Conclusion

This section elaborates and synthesizes the results of the papers and then tries to explicitly answer each research question stated at the beginning of this thesis. After a critical reflection on acquired results, the limitations of the proposed three individual solutions were discussed. The final discussion of the chapter is on the validation of the findings with respect to the selected method and attained results.

5.1 Synthesis of Results

The collective results of the findings reported in Chapter 4 form the basis for how data engineering can be introduced into engineering design processes and enable a data-driven design approach. When analyzing the results together and comparing them to the literature, it is relevant to position the data-driven design approach proposed in this thesis in line with more well-known design automation approaches reported over the last decades. The results presented in this thesis rely on using concepts originating in data science and data engineering and combining these in design evaluation disciplines. An important factor to address when proposing how to further introduce data-driven design into practice is to ensure that interpretability, conventions, and definitions used are carefully addressed.

Fully automated systems offer clear advantages, but they do not meet industrial demands for rapid design evaluation, as they do not reduce the computational time. Another challenge with fully automated design approaches is generalizability, i.e., the measure required to fully automate design support requires too much effort for every new design study. On the other hand, data-driven approaches such as surrogate models and sophisticated AI algorithms are challenged by their need for a large amount of training data that, in turn, requires an automated CAE simulation process. Furthermore, the design automation and data-driven approach (separately) require scripting competence, which domain engineers in the industry usually lack. Design automation is usually transparent and allows for debugging, but it is a slow process and is built on the same expensive simulations. Whereas data-driven algorithms lack transparency, once the training has been performed, the speed is almost considered real-time, which makes them more suitable for the early phases of the design process, when high accuracy is traded off for time.

The data-driven approaches presented are an important step to advance design evaluation practice beyond the current state in industry. In the study reported, it has been evidenced that lead time for design evaluation studies can be radically reduced in several different types of evaluation conditions. As such, the ability to evaluate (more) designs in a continuously time-pressed early development phase can be supported. It has been argued that iterations are a natural part of any development process and are even beneficial for optimizing and refining any task. Therefore, in each iteration of such futuristic support, the product requirements are satisfied, and the product is synchronized with the customer before it moves either to the next iteration or to the production phase. Digital models used through this verification process resemble digital twins, and the work can be positioned in the context of Industry 4.0.

5.1.1 Feature Engineering for Data-driven Design Evaluation

Feature engineering enables designers to create straightforward yet precise prediction models using a few features and sample points. Comparing linear regression to support vector regression, paper A concludes that transitioning from simple to more complex algorithms does not resolve the prediction accuracy issue to the extent that sleeping parameters can. Many other researchers share similar opinions about the value of data exceeding that of the model, and this perspective is acknowledged in the literature by the community.

Moreover, straightforward analyses can democratize the use of analyses, allowing designers with diverse backgrounds to have a real-time prediction model. This method, if integrated into a CAD environment, allows designers to swiftly observe how changes in design variables like length, radius, or offset can affect the simulation output, potentially increasing the design evaluation speed. Furthermore, independence from traditional parameterization in CAD will provide the flexibility to innovate with new solutions and explore beyond the design space limitations imposed by those parameterizations.

The results from papers A and E contribute to feature engineering in data-driven design and indicate that employing feature engineering could lead to models that are both less complex and more predictive. This offers a method for the engineering design community to conduct design evaluations independently from the parameterization process, which can also result in a more automated and faster evaluation.

5.1.2 Automated Labeling for Data-driven Design Evaluation

Image regression and surrogate modeling have been proven able to reduce lead time in product development and design evaluation situations, both in literature and in practice. However, most existing design automation techniques show limitations in labeling generated training data due to their reliance on expensive simulation and test data. Dynamic relaxation is proposed as a less computationally intensive simulation to compute the volume of the airbag. When utilized in the proposed framework, it allows the design automation script to level large datasets based on a proportion of the data and thereby facilitates image regression in the design process.

Dynamic relaxation possesses several characteristics that enhance its suitability for automating the simulation output assignment process to training data in data-driven design processes. Despite being less accurate compared to other finite element methods, such as Explicit Time Integration Methods (e.g., Central Difference Method) and Implicit Time Integration Methods (e.g., Newmark-beta Method), it offers computational efficiency and simplicity, making it advantageous for rapid data generation. Firstly, it is less computationally intensive than traditional Finite Element Methods, as there is no stiffness matrix and it avoids solving large systems of equations. Secondly, the simulation setup requires fewer detailed inputs about material properties and boundary conditions, which reduces the preparation time and complexity associated with each simulation run. These benefits collectively make dynamic relaxation an effective tool for supporting faster, more efficient design automation workflows.

The result of Paper B shows a promising way of using novel types of labeling data for state-of-the-art machine learning algorithms in design evaluation. However, more investigations are required to assess this method's generalizability with other case studies.

5.1.3 Data Engineering for Data-driven Design Evaluation

By bridging the gap between component-level details and system-level performance predictions, the framework proposed in papers D and F enables fast performance evaluation in the early phases at the system level, which make it an interesting alternative for early - screening and exploration - design studies where quality of both input and actual design definitions are low.

Using an FM tree to organize and connect diverse datasets creates a coherent data architecture that captures the complex relationships and dependencies within the system. By mapping data across different levels of the product architecture, from component specifics to overall system behavior, the framework enables quantitative analysis that is both modular and scalable. This method supports rapid ”what-if” analyses and facilitates the evaluation of new design alternatives, whether incremental or radical, without the need for extensive re-simulation.

This approach allows for the dynamic integration and updating of elements in product architecture and thereby in the dataset platform as new data becomes available or as design changes or technologies are implemented. The ability to quickly and efficiently draw analysis on the design changes makes this method particularly valuable for the early stages of design, where decisions need to be made rapidly to assess potential impacts on system performance.

In most cases, analyzing a new design change means that ML models stored in the dataset platform should perform extrapolation on the previous dataset. A similarity analysis incorporated in the dataset platform framework enables users to understand how much they can trust the prediction results. This process compares feature vectors derived from new design variants against those stored in a design database. The analysis typically uses distance metrics; based on how similar a new design is to existing designs, users can determine the extent of extrapolation required and allow engineers to make informed choices about when and how to use the platform.

5.2 Challenges and support requirements (RQ1)

RQ1: *What are the challenges of data-driven design evaluation in the design process of iterative and simulation-driven products?*

Data-driven design evaluation has the (demonstrated) potential to accelerate design evaluation, which is a substantial part of the engineering design lead time of those products that heavily rely on simulation for design evaluation. However, despite its promise, several critical challenges impede its effective implementation. Three issues are pinpointed in this thesis, which collectively contribute to extended design evaluation cycles. These challenges not only affect the accuracy and efficiency of predictive models but also limit design flexibility, ultimately influencing the overall product engineering design lead time.

- The dependency of evaluation methods on conventional model parameterization.
- The labeling of large datasets and the lack of a design-embedded evaluation method.
- A scalable and modular method for using design datasets in design change analyses.

The first challenge identified is the dependency of data-driven design evaluation on CAD model parameterization. In the design process of iterative and simulation-driven products, conventional CAD models are built upon predefined parameterization conventions. Often, the designer is required to follow the conventions from legacy models, which leads to a limited design space search. Traditional CAD parameters frequently fail to correlate strongly with simulation outcomes. As a result, ML models built using these parameters may not capture the true complexity or potential performance of a design. The design processes of TWB and curtain airbag are used as case studies in papers E and A to illustrate the challenges posed by reliance on traditional CAD parameterization. It was shown that conventional CAD parameters often result in high-dimensional design spaces that add to the complexity of surrogate modeling. For instance, in the thin-walled beam case, a simple geometric design change (like adding a radius or fillet) can necessitate numerous new parameters, leading to cumbersome model changes, increases in complexity, and a loss of accuracy in predictions. In the airbag study, parameters like offset and island length showed a weak correlation with outputs like volume, yet can cause long finite element FE iterations for design evaluation. This rigidity limits flexibility, as drastic design changes (a new technology with the same function) can render pre-trained models obsolete, escalating computational costs for redesigning and retraining, and consequently prolonging the design cycles.

The second challenge discussed in this thesis is the efficient generation of labeled data that is required for training ML models in data-driven design evaluation tasks. In simulation-driven design, traditional simulations and tests are used as labeling techniques, which tend to be manual, expensive, and not scalable to the large volumes needed. This bottleneck is exacerbated by the sequential nature of the design process, where a single design failure triggers a return to the initial evaluation stage, which prolongs the labeling process even more. This limitation results in delayed updates to the training datasets and compromises the predictive power of the evaluation models. Inaccurate or incomplete labels as a result of testing or simulation limitations can also lead to models that are less effective at design evaluation, thereby hampering rapid decision-making and iterative improvements.

The scarcity of comprehensive datasets hinders data-driven design evaluation of simulation-driven products. Paper C identifies three categories of challenges as the reasons for this challenge: Namely, sharing/privacy, quality assurance, and acquisition economics. Companies are increasingly concerned about privacy issues, and regulations such as the General Data Protection Regulation (GDPR) restrict access to real-world data, which is critical for having wide access to trustworthy data. Meanwhile, competitive reluctance from companies further limits dataset availability as they refrain from publishing their produced data, as noted in security-intensive fields like space design (CHEPOS project). Quality issues, such as noisy or inconsistent data, demand extensive preprocessing, which slows down the iterative design evaluation, while the high cost of collecting and processing large datasets makes it impractical to generate the thousands of samples needed for robust predictive models (touched on in Paper B). These barriers collectively impede the creation of datasets that can support the rapid, data-intensive evaluations required in iterative and simulation-driven design, leaving designers with suboptimal tools for decision-making.

Data-driven design evaluation for Hall Effect Thrusters in the space industry is challenged by interoperability difficulties and limited data availability, as outlined in paper D. The complexity of systems involving multiple subsystems (e.g., thruster units, flow management systems) which are developed by different companies lead to such issues and slows evaluation cycles in the design process. Additionally, the scarcity of comprehensive operational data restricts the ability to train robust predictive models, forcing reliance on physical testing or analytical methods with high error rates. It is worth mentioning that the plasma physics used to simulate the performance of these thrusters is not commercialized and holds a lower Technology Readiness Level (TRL) than other CAE methods. Thus, the space industry is challenged by the inability to share datasets among different teams and digital means of generating accurate data, which leads to an inability to make informed decisions in the early phases.

Therefore, the third problem area concerns the lack of a method that can enable different teams sitting apart from each other and producing their own datasets to work together toward a common design evaluation goal. As shown in Papers D and F, traditional datasets are often constructed around fixed, one-level information at one component or single subsystem level, which restricts the ability to explore alternatives beyond the established design space. This limitation largely arises from designers relying on costly simulations and tests to understand how a design change in the system will impact various outputs. This makes current design change analysis practices resistant to scale. This situation gets worse in larger companies because designers and testers are sitting even further apart, or in complex products where every subsystem is produced by a different company (as is the case in the space industry), which leads to miscommunication and inconsistent assumptions about the design process.

5.3 State of the art and practice (RQ2)

RQ2: *How can data-driven methods in product development processes be customized for the design evaluation of iterative and simulation-driven products?*

The state of the art in engineering design has focused on surrogate modeling techniques to address high-dimensional, expensive black-box problems, such as radial basis functions with high-dimensional model representation and improved kriging surrogates using partial least squares for dimension reduction, as reviewed in Section 2.

Engineering design literature has increasingly leveraged advanced data-driven techniques, including deep learning and image regression, to address challenges in simulation-driven design processes. As shown in Paper B, recent research has explored a variety of input types (scalars, vectors, time series, and images) to build surrogate models that predict simulation outcomes, such as aerodynamic coefficients, stress distributions, and fluid flow fields. Techniques like CNN have gained prominence for handling complex, high-dimensional data, including geometric representations from CAD models, point clouds, and surface meshes. These approaches aim to overcome limitations of traditional parametric surrogate models, such as dimensionality and dependency on pre-existing simulation data, by integrating richer geometric information and advanced algorithms. The literature highlights a shift toward image-based methods to capture intricate design details, driven by the success of computer vision advancements like ImageNet, though engineering applications still lag in handling real-world geometric complexity.

Recent advancements in engineering design have utilized digital twins to enhance decision-making in product design, in addition to monitoring system performance during operation. However, this application faces criticism since there is no physical twin during the concept design phase, leading it to be termed as a digital shadow in the literature. Moreover, in resource-constrained settings, the investment in digital twins might not justify the returns and may be prohibitive for smaller organizations or early-stage projects, as is the case in this thesis.

As for the state of engineering design practice, the use of CAD and CAE simulation is characterized by iterative and resource-intensive processes, compounded by challenges like high dimensionality and parameterization (Discussed in Papers A and E). Designers typically define CAD models with extensive parameters and constraints to cover the design space, which often results in large, cumbersome training datasets for machine learning models. This adherence can limit design flexibility and creativity. The separation between CAD designers and CAE simulation engineers, frequently in different departments, leads to prolonged engineering design lead time due to iterative feedback loops. These inefficiencies are particularly evident in the early design phase, where evaluation tools are lacking, necessitating repeated simulations in later phases.

Simulation-driven processes dominate design evaluation, relying heavily on computationally intensive tools like FEA, CFD, and rigid body dynamics, etc. As mentioned in paper B, the avant-garde companies in these fields use design automation to reduce the burden, however, DA comes with limitations. To alleviate this, industry often supplements commercial tools with in-house solutions to accelerate evaluation, yet the computational burden of processing complex simulations remains a significant bottleneck, particularly for real-time design space exploration. These methods, while accurate, result in long engineering design lead times due to sequential, iterative workflows where designs are repeatedly tested and refined across multiple simulation stages. The empirical data collected for this thesis clearly indicates that the industry requires a next-best model 1.9 that, while less accurate, offers faster evaluations, specifically in early-phase design phases. The tradeoff between evaluation costs and our confidence levels must be considered for future design evaluation methods.

Technology selection within system-level design relies heavily on iterative, resource-intensive processes involving multiple teams working on interdependent subsystems, as discussed in paper E. Design evaluations often depend on physical testing or detailed digital or physical simulations, which extend design cycles due to their computational complexity and the need for extensive validation. Low-fidelity models, such as scoring matrices, are commonly used in early conceptual phases, but they lack the granularity to assess the impact of innovative technologies on system performance. The fragmented nature of subsystem development in early phases creates interoperability challenges, with teams operating in silos, leading to prolonged lead times and difficulties in integrating new technologies. This traditional approach struggles to adapt to rapid design changes or scale effectively across hierarchical system levels, highlighting a gap between practical needs and the ability to explore design spaces efficiently.

5.4 Developed supports (RQ3)

RQ3: *What data-driven design supports can be developed for more efficient design evaluation?*

The design support developed in this thesis centers on the utilization of data and datasets to enhance the data-driven design evaluation processes for more efficient analysis techniques. It prescribes datasets and their elements (features and labels) to enhance the effectiveness of design evaluations in a data-driven context.

The correlation-based feature extraction approach leverages the medial axis as an alternative representation of the geometry to extract and select more impactful features than CAD parameters. This approach is exemplified in two case studies, first on the volume simulation of a curtain airbag and second on the evaluation of TWB for crashworthiness metrics. Through these applications, the approach not only simplifies the computational demands of the design process but also provides a more robust and flexible framework for handling complex design changes and iterations, ensuring more accurate and efficient predictive modeling.

The methodology capitalizes on 'sleeping parameters' (see papers A and E), a concept that involves the extraction of highly relevant features that are typically overlooked in standard CAD-based analyses. These parameters are extracted from the medial axis representation, offering a more profound insight into the design's functional and structural integrity without the constraints of conventional parameterization. The resulting reduction in the dimensionality of design problems significantly reduces the reliance on time-consuming simulations. Enhancing the surrogate modeling techniques allows for rapid prototyping and iterative testing, accelerating the development cycle. It is shown that techniques like the medial axis yield geometric descriptors that can be employed to capture essential structural characteristics, such as length, surface area, or cumulative circumferences, offering a more comprehensive representation of complex shapes. These descriptors excel at capturing the underlying topology and spatial distribution of geometry, which can enhance predictive modeling by providing richer, shape-aware features.

The development of real-time design evaluation support offers a significant enhancement in the efficiency of the design process. To offer such a possibility, this thesis proposes a dynamic relation to label thousands of images, which was not possible by traditional methods. Dynamic relaxation sidesteps traditional FE, which often entail complex setups, including detailed material models, boundary conditions, and intrusion dynamics. Instead, dynamic relaxation treats the simulation as an energy problem within a spring-mass model framework, and such simplification trades off speed for accuracy. Leveraging dynamic relaxation for label creation allows the utilization of state-of-the-art image-based machine-learning models trained on screenshots within a CAD environment. This integration facilitates the transition of design evaluation into a real-time, data-rich phase where a wide range of stakeholders can modify designs and get informed on predictive analytics.

This thesis leverages a framework that combines functional decomposition with surrogate modeling. This innovative approach is particularly beneficial in complex system design processes that involve larger developing teams and more complex product architecture. By applying function-mean modeling and surrogate techniques, designers can predict system-level performance from low-level input parameters. The method is initially applied to a case study from the space industry, where it demonstrated the possibility of synthesizing high-level output (thrust) from lower-level inputs, such as the orifice diameter of the fluid management subsystem, through a shared feature (flow rate). The method is structured as a "Product Dataset Platform" to bridge the gap between different levels of product and perform system-level performance evaluations based on feature engineering and functional decomposition applied to FE datasets. This support links early conceptual models to detailed response evaluations, providing a framework for assessing new technologies and making informed decisions during the system-level design process.

5.5 Critical reflection on acquired results

One major challenge for feature engineering adoption in engineering design is the heterogeneity of data sources. Features used for surrogate modeling could include various sources, types, or qualities. For instance, textual data from requirement documents or descriptive design notes from an older best practice could not be easily used to augment

CAD model parameters in a data-driven evaluation method. Conventional CAD models are built on established parameterization methods that yield a set of numerical features reflective of design intent, such as dimensions or tolerance values, while empirical data might be gathered from physical prototypes or sensors with varying sampling rates and noise levels. Combining these different data types requires careful preprocessing, normalization, and often, the development of bespoke feature extraction techniques tailored to each data source.

Furthermore, the contextual relevance of features may differ based on the design objective. For example, features relevant for predicting structural integrity might differ from those needed for assessing aerodynamic performance. As a result, feature engineering for engineering design involves not only handling heterogeneous data but also identifying the most relevant features for specific tasks within a larger, interconnected design process.

The live prediction model presented in Paper B, despite being able to show the performance of the design correctly, does exhibit some shortcomings. Although dynamic relaxation is computationally less intensive compared to other traditional finite element methods, it still involves iterative discretization steps that can become time-consuming when scaled to very large or highly complex datasets. Moreover, while the CNN-based image regression model achieves close to 90% accuracy on the validation set, its performance diminishes when dealing with more subtle design variations or when applied to simulations involving complex stress distributions that are not as effectively captured by image-like representations. Using larger image resolutions may increase computational costs, and employing deeper networks can introduce additional complexities.

While the dataset platform design (presented in paper F) provides a flexible and modular framework that unifies component-level and system-level data through a hierarchical FM tree, a key shortcoming identified in Papers D and F is its limited ability to accurately extrapolate performance for radical design changes. In practice, the platform works well when modifications remain within the incremental range of the legacy design space; however, when designers push the boundaries with innovative or substantially different alternatives, the shared interfaces and integrated datasets may not capture the full complexity of system interactions. This limitation can lead to degraded predictive accuracy and challenges in confidently assessing the performance of designs that fall far outside the established dataset.

The discussion in Paper B addresses image regression techniques utilizing image-based datasets for engineering applications, outlining various advantages and disadvantages associated with these methods. It has been argued that engineering-based datasets possess unique characteristics, necessitating tailored treatments regarding their data generation and handling process. One characteristic of that source is the geometry. Like an airbag shape screenshot example shown in Paper B, using DOE with CAD to generate a large dataset results in skewed training data compared to images that are usually used in computer science datasets.

Conversely, design datasets compared to computer science benchmark datasets offer advantages as well. For instance, since all the images are generated digitally, much of the pre-processing typically required for datasets can be covered faster. By uniformly cropping all the images using a script for their generation, it's possible to align all the constant pixels in one position, which can greatly reduce training time in the learning process. This is a process that can be very time-consuming for other datasets.

5.6 Validation of the results

The three techniques for data-driven design evaluation support proposed in this thesis are linked to established modeling techniques and result from a combination of several approaches. The Medial Axis and Sleeping Parameters merge CAD modeling for feature extraction with AI models for predictions. Dynamic Relaxation, as a modeling tool, aligns with CAE modeling, while the CNN models are the AI model aspect of this support. The Dataset Platform leverages various modeling, CAE models to generate datasets, AI as the analysis component, and functional modeling, which is used to recreate the product architecture. To understand the environment of the models developed in this thesis and position them in that environment, Figure 5.1 visualizes all the models developed for the supports.

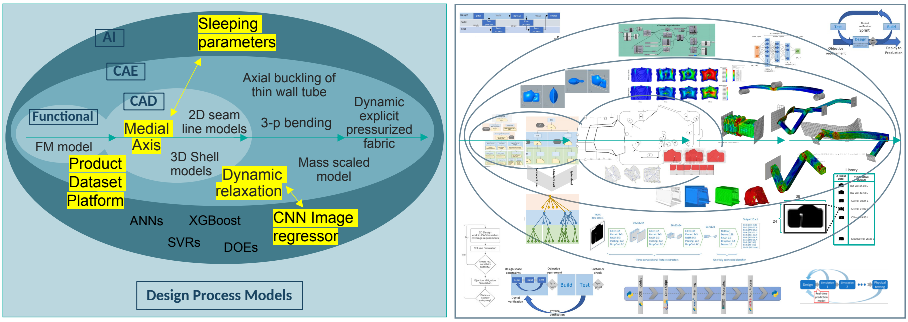


Figure 5.1: Categorization of developed and/or used models

Depending on the type of models used for each design evaluation support, the validation questions that need to be asked may vary in scope. Therefore, the four steps of Sargent's validation model (see Figure 3.7 are applied differently over the three developed design evaluation supports. For the first step in all supports, whether it was of the type a CAD, CAE, or AI model type, it was ensured that the model represented the system to the intended degree. For example, the design evaluation supports 'Medial Axis and Sleeping Parameters' and 'Image Regression and Dynamic Relaxation', are performed on the airbag case study, where the CAD models are developed according to the case companies' internal guidelines and are therefore validated separately through testing for other internal applications. Modifications of the models for developing these supports are performed in collaboration with company specialists. Therefore, the *Computerized model verification* in Sargent's model, which is about making sure the models are correctly implemented in the computer, is deemed completed. Similarly, for the Product Dataset Platform, which is applied to the crash case study, all the CAE models utilized are selected from experi-

Table 5.1: Action has been taken for various parts of validation/verification

	Medial axis and sleeping parameters	**Image regression and dynamic relaxation**	**Product dataset platform**
Conceptual model validation	Investigated through a workshop in the case company		
Computerized model verification	Review/tracing execution, Error debugging		
	CAD models developed with experts	CAD models developed with experts	CAE models built based on literature
Operational validation	Method applied to a different case	AI predictions compared with simulation, Used physically tested CAE	Method applied to a different case, Used physically tested CAE
Data validity	ML models and error metrics are compared, Correlation checked	Ensure dataset is representative	Error metrics compared, Declined accuracy discussed, Data leakage is prevented

mentally validated examples that exist in the literature. Table 5.1 shows the other three steps in Sargent's model and how they have been achieved in this thesis.

The subsequent step in the validation method is *Operational validity*, which involves ensuring that the model's output behavior has sufficient accuracy for its intended purpose. While alternative methodologies may exist to verify the correct implementation of our produced CAD, CAE, and AI models, a potentially more reliable approach is to apply the support to a different case. Consequently, two of the three supports are tested in this manner, as illustrated in the table. When utilizing CAE models in conjunction with the Product dataset platform and Image regression methods, the approach to ensure accurate results was to implement a simulation setup that has been validated through physical testing. This is accomplished by researching existing literature and consulting with the case company to obtain an appropriate model.

Data validity is achieved through involving various ML models and the evaluation of their performance using different error metrics. By comparing these models and analyzing correlation metrics, the consistency and reliability of the data are assessed. Additionally, in the case of the Product dataset platform, steps are taken to prevent data leakage, ensuring that the models are trained and tested on appropriately separated datasets to avoid artificially inflated performance metrics.

Conceptual model validity refers to the process of ensuring that the underlying theories, assumptions, and abstractions used in developing conceptual models are accurate, appropriate, and applicable to the intended problem domain. It involves verifying that the models adequately represent the relevant phenomena and that the simplifications or assumptions made during model development do not compromise their validity for the intended purpose. In this thesis, conceptual model validity is established by demonstrating the usefulness and applicability of the developed design evaluation supports. This involves showing that the proposed methods and models are both theoretically sound and practically effective in addressing the specific needs of the engineering design process.

This process requires industrial experience and practical knowledge regarding how these supports can be beneficial in a real industrial setting. Therefore, as detailed in the methodology section, a workshop is created and conducted with an industrial partner to assess this component. The results of the workshop are presented in Figure 5.2, which shows that most of the participants expressed themselves positively about the need for the support and also the usefulness and applicability of the them in their work.

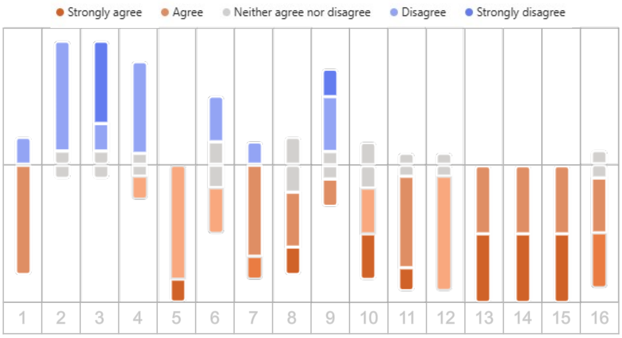


Figure 5.2: Results of validation verification workshop from the case company

The 16 questions that are asked of participants of the workshop are listed below (Legend of the Figure 5.2), which shows 5 questions per design evaluation support.

1. Designers in the company need to know more about AI technology and its applications.
2. You are using AI in your daily work in the company one way or another?
3. Alternative geometry representations are frequently used in your design and/or analysis workflows.
4. Adhering to a fixed parameterization convention adds unnecessary delays in design and/or limits your design freedom.
5. Medial axis-based feature extraction can be extended and useful for more product design processes in the company.
6. Precision that lies in long FEM simulations goes to waste because design iterations happen in early fuzzy stages.
7. Do you think the company is doing way too many simulations and way too much manual development?
8. Some of the manual development tasks in the company should be done with AI

9. Do you use or take benefit in any form from game engine simulation methods in the company?
10. How useful do you think using screenshots and AI can be for evaluating some of the simulations in the company?
11. Some of the simulations in the company can also benefit from the dynamic relaxation method (or similar game engines).
12. There is a relation between the scale of the change in the design process and the time it will take to address it.
13. Ability to know the effect of your decisions at the system level is less than that at the component level.
14. Connecting functional modeling to product architecture will be useful for the company products when it comes to addressing design changes.
15. Dataset platform could be applicable to resolving the design change problem at the system level in the company.
16. Dataset management increase the application of AI in product development?

The techniques for data-driven design evaluation support proposed in this thesis have proven effective in close case situations, reducing the time required for design evaluation (see papers A and B). However, to implement the proposed method in the PD process model level within the company' s real environment, these supports must be incorporated into the design tools currently being used in the company. This aspect will remain to be tackled by companies in the near future.

The validation study conducted has assessed the usefulness and applicability of the proposed methods within a relevant industrial context. To achieve meaningful industrial impact, the methodologies and tools developed and demonstrated in this thesis must be adopted more broadly in practice, where their influence on real product development performance can be thoroughly evidenced. While such extensive validation is beyond the scope of this PhD study, the potential for practical applicability and impact has been evaluated to a possible extent through the conducted studies.

5.7 Conclusion

Design automation has been successfully applied in various aspects of the design process, including design evaluation to free engineers from mundane tasks and reduce the engineering design lead time. However, the continuous growth of product complexity, the ever-changing customer requirements, and the general industrial shift for shorter design and product cycles necessitate new smart ways to evaluate design concepts and reduce engineering design lead time. More recently, AI algorithms and data-driven approaches have shown great potential to take up design automation to the next level, by using machines to learn, generate, and analyze design variants. However, the application of data-driven methods to the design evaluation process poses a few challenges that are addressed in this thesis. In this thesis, three challenges are identified, each addressed by a distinct solution, and applied across four case studies using a design research methodology. First, feature engineering is proposed to make the evaluation less dependent on parameterization, then dynamic relaxation is used to label a large design dataset, which paves the way for large-scale dataset generation. Finally, combining features and labels, a data engineering framework is outlined called dataset platform design that offers a faster and cheaper design evaluation method, trading off accuracy. From a scientific perspective, this thesis contributes to the advancement of data-driven design evaluation by providing methods for leveraging data engineering techniques within design evaluation processes. The methods are demonstrated to be effective in early phases of the design process, such as technology selection. The studies highlight limitations associated with each proposed design evaluation support, which provides a basis for further investigations. Despite these limitations, validation results indicate that the proposed methods are helpful for design evaluation in industrially relevant scenarios. While direct impacts on engineering design lead time have not been measured, the positive feedback on usability and effectiveness suggests that the approach addresses critical challenges in iterative, simulation-driven design processes. This indicates potential applicability across other high-level, technical design environments.

5.8 Future work

Based on the findings and limitations identified throughout this thesis, there are several areas where further work is recommended. The suggested areas for further research are structured into four key themes:

Mitigating Identified Weaknesses

While the proposed methodologies have shown promise, certain limitations were identified during the studies. Addressing these weaknesses requires further refinement and enhancement of the current approaches.

For future image regression methods, Other sampling methods can be explored to generate design variants that efficiently cover the design space. Our analysis indicates that relying solely on parameterized methods may hinder thorough exploration of the design space, limiting the potential for innovative solutions.

Further simulations can be conducted utilizing dynamic relaxation techniques, as the scope of this thesis primarily focuses on the inflation process of airbags.

Implemented CNN mode designed for classification problems, which needs modification for engineering regression tasks. The network architecture will be designed.

Exploring Alternative Approaches

This thesis focused on specific methods for data-driven design evaluation; however, alternative approaches could further enhance the capabilities of design automation.

Alternative extractable information from geometrical and simulation models can be explored as input for building an AI model is an area that requires more exploration. Cloud-based and voxel-based representations of geometrical objects have shown promising results in the literature. These geometrical representations could provide a new data type for training AI models. Testing several data types on different models will determine what kind of data performs best with which AI method.

Physics-informed neural networks (PINNs) have recently shown generalizability for broader applications. PINNs include the physical laws (for example, a differential equation) governing the behavior under study in the cost function of the ML model. Since there are analytical solutions for crash simulation, these equations can be tested on PINNs to improve the learning and reduce the training size.

It is recommended to explore the integration of feature engineering techniques into 3D modeling and to expand this approach to other engineering fields. In 3D geometry, the medial axis is represented in a plane rather than as a line, as it is in 2D geometry. This understanding can lead to new geometric representations that benefit a wider range of products.

Integrating unsupervised or semi-supervised learning techniques to enhance model training where labeled data is limited. Unsupervised learning methods enable the model to identify patterns and structures within the data without the need for labeled examples, allowing for the extraction of valuable insights from unannotated datasets, which makes them relatable to geometrical shape designs.

Further investigating the use of transfer learning to apply learned knowledge from one domain to another. This includes analyzing how pre-trained models can be adapted to new design tasks, the role of fine-tuning in enhancing model performance, and the impact of domain similarities on learning efficacy.

Advancing Research into Practice

The industrial relevance of the proposed methods suggests the potential for broader practical application.

The developed surrogate models have been successful and validated. The next step is to create software tools with a graphical user interface that can be implemented in companies. This will simplify the utilization of sophisticated AI algorithms for non-engineers. This was needed because less technical users struggle to maintain datasets and models.

Conducting more extensive validation studies in collaboration with industry partners to ensure practicality and usability is also recommended to advance research into practice.

Generalizing Findings and Contributing to Knowledge

While the methods proposed in this thesis address specific challenges, generalizing these findings to broader theoretical frameworks and methodologies would strengthen their scientific contribution.

In future studies, the expectation is to find out what kind of data types are more suitable for which types of simulations and geometries, be it solid models, sheet metal, etc. Finally, moving from real-time predictions to real-time analysis, a generalized framework for performing performance analysis in real-time can be developed and validated with relevant validation methods.

Developing generalized theories and guidelines for generating and evaluating datasets in data-driven design is essential. This includes focusing on best practices for dataset curation, such as establishing criteria for data relevance and quality. To effectively evaluate these datasets, we need clear metrics and standards that allow designers to assess their impact on fostering innovative solutions.

Chapter 6

References

Achanta, A. and Boina, R. (2023). Evolving Paradigms of Data Engineering in the Modern Era: Challenges, Innovations, and Strategies. *International Journal of Science and Research (IJSR)*, 12(10):606–610.

Ackoff, R. L. (1989). From data to wisdom. *Journal of applied systems analysis*, 16(1):3–9.

Adler, R. M. (2008). Knowledge engines for critical decision support. In *Knowledge management strategies: a handbook of applied technologies*, pages 143–169. IGI Global Scientific Publishing.

Ahmed, F., Picard, C., Chen, W., McComb, C., Wang, P., Lee, I., Stankovic, T., Allaire, D., and Menzel, S. (2025). Special Issue: Design by Data: Cultivating Datasets for Engineering Design. *Journal of Mechanical Design*, 147(040301).

Akay, H. and Kim, S.-G. (2021). Artificial Intelligence Tools for Better Use of Axiomatic Design. *IOP Conference Series: Materials Science and Engineering*, 1174(1):012005.

Al-Fedaghi, S. (2016). Toward a philosophy of data for database systems design. *International Journal of Database Theory and Application*, 9(10):47–62.

Amadori, K., Tarkian, M., Ölvander, J., and Krus, P. (2012). Flexible and robust CAD models for design automation. *Advanced Engineering Informatics*, 26(2):180–195.

Arinez, J. F., Chang, Q., Gao, R. X., Xu, C., and Zhang, J. (2020). Artificial Intelligence in Advanced Manufacturing: Current Status and Future Outlook. *Journal of Manufacturing Science and Engineering*, 142(110804).

Arjomandi Rad, M. (2022). Airbag cad to cae volume. https://www.kaggle.com/mohammadrad/airbag-cad-to-cae-volume. Data files.

Arjomandi Rad, M., Cenanovic, M., and Salomonsson, K. (2023). Image regression-based digital qualification for simulation-driven design processes, case study on curtain airbag. *Journal of Engineering Design*, 34(1):1–22. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/09544828.2022.2164440.

Arjomandi Rad, M., Hajali, T., and Martinsson Bonde, J. (2024). Datasets in design research: needs and challenges and the role of AI and GPT in filling the gaps. In *Proceedings of the Design Society, 4: 1919-1928*.

Arjomandi Rad, M., Raudberget, D., and Stolt, R. (2022). Data-driven and real-time prediction models for highly iterative product development processes. In *Transdisciplinarity and the Future of Engineering*, pages 463–472. IOS Press.

Babu, S., Mourad, A.-H., Harib, K., and Vijayavenkataraman, S. (2023). Recent developments in the application of machine-learning towards accelerated predictive multiscale design and additive manufacturing. *Virtual and Physical Prototyping*, 18(1).

Badini, S., Regondi, S., and Pugliese, R. (2023). Unleashing the Power of Artificial Intelligence in Materials Design. *Materials*, 16(17):5927.

Baduge, S. K., Thilakarathna, S., Perera, J. S., Arashpour, M., Sharafi, P., Teodosio, B., Shringi, A., and Mendis, P. (2022). Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Automation in Construction*, 141:104440.

Bagnato, Prof. Ass., V. P. (2023). Artificial Intelligence for Design The Artificial Intelligence of Objects. *FORUM A+P Interdisciplinary Journal of Architecture and Built Environment*, pages 30–35.

Ballsun-Stanton, B. (2012). Asking about data: Exploring different realities of data via the social data flow network methodology.

Bengio, Y., Privitera, D., Besiroglu, T., Bommasani, R., Casper, S., Choi, Y., and Mindermann, S. (2024). International Scientific Report on the Safety of Advanced AI (Interim report). *Gov. uk Department for Science, Innovation and Technology and AI Safety Institute*.

Bertoni, A. (2020). Data-DRIVEN DESIGN IN CONCEPT DEVELOPMENT: Systematic REVIEW AND MISSED OPPORTUNITIES. *Proceedings of the Design Society: DESIGN Conference*, 1:101–110.

Blessing, L. T. and Chakrabarti, A. (2009). *DRM: A design reseach methodology*. Springer.

Bonde, J. M., Brahma, A., Panarotto, M., Isaksson, O., Wärmefjord, K., Söderberg, R., Kipouros, T., Clarkson, P. J., Kressin, J., and Andersson, P. (2022). Assessment of weld manufacturability of alternative jet engine structural components through digital experiments. In *2022 Proceedings of ISABE*. International Society for Air Breathing Engines (ISABE).

Boschert, S. and Rosen, R. (2016). Digital twin—the simulation aspect. In *Mechatronic futures*, pages 59–74. Springer.

Bossak, M. A. (1998). Simulation based design. *Journal of Materials Processing Technology*, 76(1-3):8–11.

Bradley, S. (2010). Designing For A Hierarchy Of Needs. Section: General.

Briard, T., Jean, C., Aoussat, A., and Véron, P. (2023). Challenges for data-driven design in early physical product design: A scientific and industrial perspective. *Computers in industry (Print)*.

Briard, T., Jean, C., Aoussat, A., Véron, P., Le Cardinal, J., and Wartzack, S. (2021). Data-DRIVEN DESIGN CHALLENGES IN THE EARLY STAGES OF THE PRODUCT DEVELOPMENT PROCESS. *Proceedings of the Design Society*.

Brodie, M. L. (2023). Defining data science: a new field of inquiry. arXiv:2306.16177 [cs].

Brown, T. (2008). Design thinking. *Harvard business review*, 86(6):84.

Browning, T. R. (2018). Building models of product development processes: An integrative approach to managing organizational knowledge. *Systems Engineering*, 21(1):70–87. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sys.21421.

Browning, T. R., Fricke, E., and Negele, H. (2006). Key concepts in modeling product development processes. *Systems Engineering*, 9(2):104–128. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sys.20047.

Burrows, S., Stein, B., Frochte, J., Wiesner, D., and Muller, K. (2011). Simulation data mining for supporting bridge design. In *Australasian Data Mining Conference*.

Cantamessa, M., Montagna, F., Altavilla, S., and Casagrande-Seretti, A. (2020). Data-driven design: the new challenges of digitalization on product design and development. *Design Science*, 6.

Case, P. (1972). Evolution of Design Automation. *Computer*, 5(3):21–22.

Cash, P., Isaksson, O., Maier, A., and Summers, J. (2022). Sampling in design research: Eight key considerations. *Design Studies*, 78:101077.

Cederfeldt, M. and Elgh, F. (2005). Design automation in smes–current state, potential, need and requirements. In *DS 35: Proceedings ICED 05, the 15th International Conference on Engineering Design, Melbourne, Australia, 15.-18.08. 2005*.

Chakrabarti, A., Shea, K., Stone, R., Cagan, J., Campbell, M., Hernandez, N. V., and Wood, K. L. (2011). Computer-based design synthesis research: An overview. *Journal of Computing and Information Science in Engineering*, 11(21003).

Chan, H., Hadjiiski, L. M., and Samala, R. K. (2020). Computer⌧aided diagnosis in the era of deep learning. *Medical Physics*, 47(5).

Chatty, T., Harrison, W., Ba-Sabaa, H. H., Faludi, J., and Murnane, E. L. (2022). Co-Creating a Framework to Integrate Sustainable Design into Product Development Practice: Case Study at an Engineering Consultancy Firm. *Sustainability*, 14(15):9740. Number: 15 Publisher: Multidisciplinary Digital Publishing Institute.

Cheng, Z. and Ma, Y. (2017). Explicit function-based design modelling methodology with features. *Journal of Engineering Design*, 28(3):205–231. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/09544828.2017.1291920.

Chiarello, F., Belingheri, P., and Fantoni, G. (2021a). Data science for engineering design: State of the art and future directions. *Computers in Industry*, 129:103447.

Chiarello, F., Belingheri, P., and Fantoni, G. (2021b). Data science for engineering design: State of the art and future directions. *Computers in Industry*, 129:103447.

Chiu, M.-C. and Lin, K.-Z. (2018). Utilizing text mining and Kansei Engineering to support data-driven design automation at conceptual design stage. *Advanced Engineering Informatics*, 38:826–839.

Chou, W. H. and Wong, J.-J. (2015). From a Disciplinary to an Interdisciplinary Design Research: Developing an Integrative Approach for Design. *International Journal of Art & Design Education*, 34(2):206–223. Publisher: John Wiley & Sons, Ltd.

Chui, M., Yee, L., Hall, B., and Singla, A. (2023). The state of AI in 2023: Generative AI' s breakout year. Technical report, McKinsey & Company. Publisher: McKinsey & Company.

Clark, K. B. (1991). Product development performance: Strategy. *Organization, and Management in the World Auto Industry*.

Clark, K. B. and Fujimoto, T. (1989). Lead time in automobile product development explaining the Japanese advantage. *Journal of Engineering and Technology Management*, 6(1):25–58.

Colombo, G., Pugliese, D., and Rizzi, C. (2008). *Developing DA Applications in SMEs Industrial Context*, pages 69–82. Springer US.

Cooper, R. G. (2014). What's next?: After stage-gate. *Research-Technology Management*, 57(1):20–31.

Cooper, R. G. (2023). The artificial intelligence revolution in new-product development. *IEEE Engineering Management Review*. Publisher: IEEE.

Cunningham, J. D., Simpson, T. W., and Tucker, C. S. (2019). An investigation of surrogate models for efficient performance-based decoding of 3d point clouds. *Journal of Mechanical Design*, 141(12).

Dagli, C. H. and Lee, H.-C. (1997). Impacts of data mining technology on product design and process planning. In Plonka, F. and Olling, G., editors, *International Conference on Computer Applications in Production and Engineering (CAPE97) 5-7 November 1997, Detroit, Michigan, USA*, IFIP - The ation for Information Processing, pages 58–70. Springer US.

de Callataÿ, A. M. (1992). *Natural and artificial intelligence : misconceptions about brains and neural networks / Armand M. de Callataÿ.* ELSEVIER SCIENCE PUBLISHERS B.V. Edition: New, expanded ed.

de Callataÿ, A. M. and M, A. (1986). *Natural and artificial intelligence. Processor systems compared to the human brain*. Elsevier Science Pub. Co., Inc., New York, NY.

de Melo, C. M., Torralba, A., Guibas, L., DiCarlo, J., Chellappa, R., and Hodgins, J. (2022). Next-generation deep learning based on simulators and synthetic data. *Trends in Cognitive Sciences*, 26(2):174–187.

de Weck, O. L., Roos, D., Magee, C. L., and Cooper, S. (2011). *Engineering Systems: Meeting Human Needs in a Complex Technological World*. The MIT Press.

Debney, P. (2020). The engineer of the future is a centaur. *The Structural Engineer: journal of the Institution of Structural Engineer*, 98(1):24–30. Publisher: Institution of Structural Engineers.

Deubzer, F., Herfeld, U., Kreimeyer, M., and Lindemann, U. (2005). A structured holistic approach for the integration of cad and cae environments. In *ProSTEP iViP Science Days 2005-Cross Domain Engineering*.

Diallo, S., Durak, U., Mustafee, N., and Mittal, S. (2017). Special issue on modeling and simulation in the era of big data and cloud computing: theory, framework and tools. *SIMULATION*, 93(4):271–272. Publisher: SAGE Publications Ltd STM.

Domazet, D., Choong, F., Sng, D., Ho, N., and Lu, S.-Y. (1995). Active data-driven design using dynamic product models. *CIRP Annals - Manufacturing Technology*, 44(1):109 ‑112. Cited by: 15.

Dong, G. and Liu, H. (2018). *Feature engineering for machine learning and data analytics*. CRC press.

Dong, X., DeVries, W. R., and Wozny, M. J. (1991). Feature-based reasoning in fixture design. *CIRP Annals*, 40(1):111–114.

Du, X. and Chen, W. (2000). Methodology for managing the effect of uncertainty in simulation-based design. *AIAA Journal*, 38(8):1471–1478. Publisher: American Institute of Aeronautics and Astronautics _eprint: https://doi.org/10.2514/2.1125.

Du, X. and Zhu, F. (2018). A new data-driven design methodology for mechanical systems with high dimensional design variables. *Advances in Engineering Software*, 117:18–28.

Dubey, R. (2024). Explainable AI Assisted Evolutionary Search of Engineering Designs. In *2024 IEEE 36th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 447–453. ISSN: 2375-0197.

Ebel, H., Delden, J. v., Lüddecke, T., Borse, A., Gulakala, R., Stoffel, M., Yadav, M., Stender, M., Schindler, L., Payrebrune, K. M. d., Raff, M., Remy, C. D., Röder, B., Raj, R., Rentschler, T., Tismer, A., Riedelbauch, S., and Eberhard, P. (2025). Data Publishing in Mechanics and Dynamics: Challenges, Guidelines, and Examples from Engineering Design. *Data-Centric Engineering*, 6:e23. arXiv:2410.18358 [cs].

Edelen, A., Neveu, N., Frey, M., Huber, Y., Mayes, C., and Adelmann, A. (2020). Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems. *Physical Review Accelerators and Beams*, 23(4):044601. Publisher: American Physical Society.

Entner, D., Prante, T., Vosgien, T., Zăvoianu, A.-C., Saminger-Platz, S., Schwarz, M., and Fink, K. (2019). Potential identification and industrial evaluation of an integrated design automation workflow. *Journal of Engineering, Design and Technology*, 17(6):1085–1109.

Eppinger, S. and Ulrich, K. (2015). *Product design and development*. McGraw-Hill Higher Education.

Felzmann, H., Fosch-Villaronga, E., Lutz, C., and Tam\' o Larrieux, A. (2020). Towards Transparency by Design for Artificial Intelligence. *Science and Engineering Ethics*, 26(6):3333–3361. Publisher: Springer Science and Business Media LLC.

Feng, Y., Zhao, Y., Zheng, H., Li, Z., and Tan, J. (2020). Data-driven product design toward intelligent manufacturing: A review. *International Journal of Advanced Robotic Systems*, 17(2):1729881420911257.

Figoli, F. A., Bruggeman, R., Rampino, L., and Ciuccarelli, P. (2025). *AI-against-design* map: A systematic review of 20 years of AI-focused studies in design research. *Design Studies*, 96:101279.

Flankegård, F., Johansson, G., and Granlund, A. (2019). Challenges with supplier involvement in product development: A supplier' s perspective. *Proceedings of the Design Society: International Conference on Engineering Design*, (1):2179–2188.

Frangos, M., Marzouk, Y., Willcox, K., and van Bloemen Waanders, B. (2010). Surrogate and Reduced-Order Modeling: A Comparison of Approaches for Large-Scale Statistical Inverse Problems. In *Large-Scale Inverse Problems and Quantification of Uncertainty*, pages 123–149. John Wiley & Sons, Ltd. Section: 7 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470685853.ch7.

Frank, G., Entner, D., Prante, T., Khachatouri, V., and Schwarz, M. (2014-06-30). Towards a generic framework of engineering design automation for creating complex CAD models. *International Journal of Advances in Systems and Measurements*, 7:179–192.

Fujimoto, T. (2000). Shortening lead time through early problem-solving—a new round of capability-building competition in the auto industry. In *New product development and production networks*, pages 23–53. Springer.

G., G. and Yongsheng, M. (2010). Generative CAD and CAE integration using common data model. *IEEE International Conference on Automation Science and Engineering*.

Gameil, A., Younes, S., and Elhebary, R. (2024). The Utilization of Artificial Intelligence Applications as An Approach in Constructing the Design Process, Considering the Overlap of Perceptual Design Concepts. *Journal of Design Sciences and Applied Arts*, 5(1):210–220.

Geyer, P., Singh, M. M., and Chen, X. (2024). Explainable AI for engineering design: A unified approach of systems engineering and component-based deep learning demonstrated by energy-efficient building design. *Advanced Engineering Informatics*, 62:102843.

Golzari, A., Sefat, M. H., and Jamshidi, S. (2015). Development of an adaptive surrogate model for production optimization. *Journal of Petroleum Science and Engineering*, 133:677–688.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

Grace, K., Salvatier, J., Dafoe, A., Zhang, B., and Evans, O. (2018). When Will AI Exceed Human Performance? Evidence from AI Experts. arXiv:1705.08807 [cs].

Graening, L. and Sendhoff, B. (2014). Shape mining: A holistic data mining approach for engineering design. *Advanced Engineering Informatics*, 28(2):166–185.

Graner, M. and Mißler-Behr, M. (2013). Key determinants of the successful adoption of new product development methods. *European Journal of Innovation Management*, 16(3):301–316. Publisher: Emerald Group Publishing Limited.

Griffin, A. (1997). The Effect of Project and Process Characteristics on Product Development Cycle Time. *Journal of Marketing Research*, 34(1):24–35. Publisher: SAGE Publications Inc.

Gustafsson, E. (2022). Exploring Data-Driven Methods to Enhance Usability of Design Optimization. *Linköping Studies in Science and Technology. Licentiate Thesis*.

Hallstedt, S. I., Isaksson, O., and Öhrwall Rönnbäck, A. (2020). The Need for New Product Development Capabilities from Digitalization, Sustainability, and Servitization Trends. *Sustainability*, 12(23):10222. Number: 23 Publisher: Multidisciplinary Digital Publishing Institute.

Hatchuel, A. and Weil, B. (2003). A NEW APPROACH OF INNOVATIVE DESIGN : AN INTRODUCTION TO C-K THEORY. *DS 31: Proceedings of ICED 03, the 14th International Conference on Engineering Design, Stockholm*, pages 109–110 (exec.summ.), full paper no. DS31_1794FPC.

Hayashi, C. (1998). What is Data Science ? Fundamental Concepts and a Heuristic Example. In Hayashi, C., Yajima, K., Bock, H.-H., Ohsumi, N., Tanaka, Y., and Baba, Y., editors, *Data Science, Classification, and Related Methods*, pages 40–51, Tokyo. Springer Japan.

He, B., Li, F., Cao, X., and Li, T. (2020). Product sustainable design: a review from the environmental, economic, and social aspects. *Journal of Computing and Information Science in Engineering*, 20(4).

Heier, J. (2021). Design Intelligence - Taking Further Steps Towards New Methods and Tools for Designing in the Age of AI. In Degen, H. and Ntoa, S., editors, *Artificial Intelligence in HCI*, pages 202–215, Cham. Springer International Publishing.

Heikkinen, T. (2021). *Extended product models supporting multidisciplinary design automation*. PhD Thesis, Jönköping University, School of Engineering.

Heikkinen, T., Johansson, J., Elgh, F., et al. (2016). Assessment of simulation ready cad models in a set-based concurrent engineering context. In *DS 84: Proceedings of the DESIGN 2016 14th International Design Conference*, pages 649–658.

Henighan, T., Kaplan, J., Katz, M., Chen, M., Hesse, C., Jackson, J., Jun, H., Brown, T. B., Dhariwal, P., Gray, S., Hallacy, C., Mann, B., Radford, A., Ramesh, A., Ryder, N., Ziegler, D. M., Schulman, J., Amodei, D., and McCandlish, S. (2020). Scaling Laws for Autoregressive Generative Modeling. arXiv:2010.14701 [cs].

Hey, T., Tansley, S., and Tolle, K. M. (2009). *The fourth paradigm: data-intensive scientific discovery*, volume 1. Microsoft research Redmond, WA.

Hilton, E., Li, W., Newton, S. H., Alemdar, M., Pucha, R., and Linsey, J. (2016). The Development and Effects of Teaching Perspective Free-Hand Sketching in Engineering Design. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers Digital Collection.

Hubka, V. and Eder, W. E. (1988). Theory of technical systems: A total concept theory for engineering design. *Berlin and New York*.

Hunde, B. R. and Woldeyohannes, A. D. (2022). Future prospects of computer-aided design (CAD) -A review from the perspective of artificial intelligence (AI), extended reality, and 3D printing. *Results in Engineering*, 14:100478.

INCOSE (2023). *INCOSE systems engineering handbook*. John Wiley & Sons, 5th edition.

Isaksson, O., , Sven, K.-S., , and Eppinger, S. D. (2000). Evaluation of design process alternatives using signal flow graphs. *Journal of Engineering Design*, 11(3):211–224. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/09544820075002099б.

Isaksson, O. (1998). *Computational support in product development: applications from high temperature design and development*. PhD Thesis, Luleå tekniska universitet.

Isaksson, O. and Eckert, C. (2020). Product development 2040. *Design Society Report DS107*.

Jain, A., Sumit, Monga, C., and Mittal, S. (2023). Data Engineering: An Overview from a Future Perspective. In *2023 6th International Conference on Contemporary Computing and Informatics (IC3I)*, pages 653–659. IEEE.

Jain, R. (2003). Out-of-the-box data engineering events in heterogeneous data environments. In *Proceedings 19th International Conference on Data Engineering (Cat. No.03CH37405)*, pages 8–21.

Jiao, R., Commuri, S., Panchal, J., Milisavljevic-Syed, J., Allen, J. K., Mistree, F., and Schaefer, D. (2021). Design Engineering in the Age of Industry 4.0. *Journal of Mechanical Design*, 143(070801).

Johansson Joel and Elgh Fredrik (2013). How to Successfully Implement Automated Engineering Design Systems: Reviewing Four Case Studies.

Johnson, J., Hurst, A., and Safayeni, F. (2023). MANAGING DATA-DRIVEN DESIGN: A SURVEY OF THE LITERATURE AND FUTURE DIRECTIONS. *Proceedings of the Design Society*, 3.

Jordan, H. H. and Hoelscher, R. P. (1935). *Engineering Drawing*. Wiley; London, Chapman & Hall.

Karniel, A. and Reich, Y. (2009). From DSM-Based Planning to Design Process Simulation: A Review of Process Scheme Logic Verification Issues. *IEEE Transactions on Engineering Management*, 56(4):636–649. Conference Name: IEEE Transactions on Engineering Management.

Keleko, A. T., Kamsu-Foguem, B., Ngouna, R. H., and Tongne, A. (2022). Artificial intelligence and real-time predictive maintenance in industry 4.0: a bibliometric analysis. *AI and Ethics*, pages 1–25.

Kelliher, A., Barry, B., Berzowska, J., O'Murchu, N., and Smeaton, A. (2018). Conversation: Beyond black boxes: tackling artificial intelligence as a design material. *DRS Biennial Conference Series*.

Khan, M. T. H. and Rezwana, S. (2021). A review of CAD to CAE integration with a hierarchical data format (HDF)-based solution. *Journal of King Saud University - Engineering Sciences*, 33(4):248–258.

King, G., Jones, R., and Simner, D. (2003). A good practice model for implementation of computer-aided engineering analysis in product development. *Journal of Engineering Design*, 14(3):315–331.

Kitchin, R. and Lauriault, T. P. (2014). Towards critical data studies: Charting and unpacking data assemblages and their work. In *The Programmable City Working Paper 2; pre-print version of chapter to be published in Eckert, J., Shears, A. and Thatcher, J. (eds) Geoweb and Big Data.* University of Nebraska Press.

Klettke, M. and Störl, U. (2021). Four Generations in Data Engineering for Data Science. *Datenbank-Spektrum*, 22(1):59–66.

Klettke, M. and Störl, U. (2022). Four Generations in Data Engineering for Data Science. *Datenbank-Spektrum*, 22(1):59–66.

Kojima, Y. (2000). Mechanical cae in automotive design. *R&D Review of Toyota CRDL*, 35(4):1.

Koziel, S. and Leifsson, L. (2016). *Simulation-Driven Design by Knowledge-Based Response Correction Techniques*. Springer International Publishing, 1 edition.

Kuhlmann, A., Vetter, R.-M., Lübbing, C., and Thole, C.-A. (2005). Data mining on crash simulation data. In Perner, P. and Imiya, A., editors, *Machine Learning and Data Mining in Pattern Recognition*, pages 558–569. Springer.

Kumar, R. (2019). Data-Driven Design. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, pages 1–2. ACM.

Königer, P. and Janowitz, K. (1995). Drowning in information, but thirsty for knowledge. *International Journal of Information Management*, 15(1):5–16.

Lagenbach, J., Wächter, M., Lohrengel, A., Müller, N., et al. (2015). Teaching engineering drawing in a technology changing environment. In *DS 82: Proceedings of the 17th International Conference on Engineering and Product Design Education (E&PDE15), Great Expectations: Design Teaching, Research & Enterprise, Loughborough, UK, 03-04.09. 2015*, pages 162–167.

Langner, C., Paliyenko, Y., Müller, B., Roth, D., Guertler, M. R., and Kreimeyer, M. (2024). Challenges for capturing data within data-driven design processes. *Proceedings of the Design Society*.

Lee, B. and Ahmed-Kristensen, S. (2023). Four PATTERNS OF DATA-DRIVEN DESIGN ACTIVITIES IN NEW PRODUCT DEVELOPMENT. *Proceedings of the Design Society*, 3:1925–1934.

Li, Y., Wang, H., and Liu, J. (2017). Can cnn construct highly accurate models efficiently for high-dimensional problems in complex product designs? *arXiv preprint arXiv:1712.01639*.

Liem, R. P. (2007). *Surrogate modeling for large-scale black-box systems*. Thesis, Massachusetts Institute of Technology. Accepted: 2008-05-19T15:01:11Z.

Liu, A., Lu, S., Tao, F., and Anwer, N. (2024). Integration of data science with product design towards data-driven design. *CIRP annals*, 73(2):509–532.

Liu, A., Wang, Y., and Wang, X. (2021). *Data-Driven Engineering Design*, pages 1–22. Springer International Publishing.

Liu, B., Campbell, R., and Pei, E. (2013). Real⊠time integration of prototypes in the product development process. *Assembly Automation*, 33(1):22–28. Publisher: Emerald Group Publishing Limited.

Madsen, D. A. (2002). *Engineering drawing and design*. Cengage Learning.

Magrab, E. B., Gupta, S. K., McCluskey, F. P., and Sandborn, P. (2009). *Integrated Product and Process Design and Development*. CRC Press.

Maier, J. F., Eckert, C. M., and Clarkson, P. J. (2017). Model granularity in engineering design - concepts and framework. *Design Science*, 3:e1.

Maria, A. (1997). Introduction to modeling and simulation. In *Proceedings of the 29th conference on Winter simulation*, pages 7–13.

Maslow, A. H. (1943). A theory of human motivation. *Psychological review*, 50(4):370.

Merkel, M. and Schumacher, A. (2003). An automated optimization process for a cae driven product development. *J. Mech. Des.*, 125(4):694–700.

Merriam, S. B. and Tisdell, E. J. (2015). *Qualitative Research: A Guide to Design and Implementation*. John Wiley & Sons.

Mianroodi, J. R., H Siboni, N., and Raabe, D. (2021). Teaching solid mechanics to artificial intelligence—a fast solver for heterogeneous materials. *npj Computational Materials*, 7(1):1–10.

Mitchell, M. and Krakauer, D. C. (2023). The debate over understanding in AI’s large language models. *Proceedings of the National Academy of Sciences*, 120(13):e2215907120.

Morley, C. (2022). Innovation and Development of Electric Vehicle Powertrain Technology: An Analysis | Jabil.

Murshed, S. M. M., Dixon, A., and Shah, J. J. (2010). Neutral definition and recognition of assembly features for legacy systems reverse engineering. In *ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pages 615–628. American Society of Mechanical Engineers Digital Collection.

Málaga-Chuquitaype, C. (2022). Machine Learning in Structural Design: An Opinionated Review. *Frontiers in Built Environment*, 8.

n.d. (2017). Automotive Product Development Cycles and the Need for Balance with the Regulatory Environment. Center for Automotive Research.

Neuhauser, L., Kreps, G. L., Morrison, K., Athanasoulis, M., Kirienko, N., and Van Brunt, D. (2013). Using design science and artificial intelligence to improve health communication: Chronologymd case example. *Patient Education and Counseling*, 92(2):211–217.

Nie, Z., Jiang, H., and Kara, L. B. (2020). Stress field prediction in cantilevered structures using convolutional neural networks. *Journal of Computing and Information Science in Engineering*, 20(1):011002.

Noble, J. (2024). Measuring Services and Inferring Causes. *Touchpoint*, 15(2):10–12.

Novak, M. and Dolšak, B. (2008). Intelligent fea-based design improvement. *Engineering Applications of Artificial Intelligence*, 21(8):1239–1254.

O., H., J., L., F., G., and B., F. (2010). Method, models and tools for cad-cae integration. *Recent Patents on Mechanical Engineering*, pages 106–130.

Obermaier, H. and Peikert, R. (2014). Feature-based visualization of multifields. In Hansen, C. D., Chen, M., Johnson, C. R., Kaufman, A. E., and Hagen, H., editors, *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization*, pages 189–196. Springer.

Osborne, S. M. (1993). *Product development cycle time characterization through modeling of process iteration*. PhD thesis, Massachusetts Institute of Technology.

Ozel, T. and Jarosz, K. (2022). Machine Learning Approaches towards Digital Twin Development for Machining Systems. *International Journal of Mechatronics and Manufacturing Systems*, 15(2/3):1.

Pahl, G. and Beitz, W. (1984). Engineering design, the design council. *London, Springer*, 12:221–226.

Pahl, G. and Beitz, W. (2013). Engineering design: A systematic method.

Painter, M. K., Erraguntla, M., Hogg, G. L., and Beachkofski, B. (2006). Using simulation, data mining, and knowledge discovery techniques for optimized aircraft engine fleet management. In *Proceedings of the 2006 Winter Simulation Conference*, pages 1253–1260. ISSN: 1558-4305.

Park, S., Joung, J., and Kim, H. (2023). Spec guidance for engineering design based on data mining and neural networks. *Computers in Industry*, 144:103790.

Perrault, R. and Clark, J. (2024). Artificial Intelligence Index Report 2024. Technical report, Stanford, Human Centered Artificial Intelligence (HAI). Publisher: < bound method Organization. get_name_with_acronym of< Organization ⋯.

Petersen, K. (2010). *An Empirical Study of Lead-Times in Incremental and Agile Software Development*, pages 345–356. Springer Berlin Heidelberg.

Petersen, P., Stage, H., Langner, J., Ries, L., Rigoll, P., Philipp Hohl, C., and Sax, E. (2022). Towards a Data Engineering Process in Data-Driven Systems Engineering. In *2022 IEEE International Symposium on Systems Engineering (ISSE)*, pages 1–8. IEEE.

Picard, C., Schiffmann, J., and Ahmed, F. (2023). DATED: Guidelines for Creating Synthetic Datasets for Engineering Design Applications. arXiv:2305.09018 [cs].

Pradas Gomez, A., Panarotto, M., and Isaksson, O. (2024). Evaluation of Different Large Language Model Agent Frameworks for Design Engineering Tasks. In *DS 130: Proceedings of NordDesign 2024, Reykjavik, Iceland, 12th - 14th August 2024*, pages 693–702.

Press, G. (2016). Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says. Section: Tech.

Quiñones-Gómez, J. C., Mor, E., and Chacón, J. (2025). Data-Driven Design in the Design Process: A Systematic Literature Review on Challenges and Opportunities. *International Journal of Human-Computer Interaction*. Publisher: Taylor & Francis.

Rad, M. A., Salomonsson, K., Cenanovic, M., Balague, H., Raudberget, D., and Stolt, R. (2022). Correlation-based feature extraction from computer-aided design, case study on curtain airbags design. *Computers in Industry*, 138:103634.

Ragsdell, K. M. (1980). Design and automation. *Journal of Mechanical Design*, 102(3):424–429.

Ramnath, S., Haghighi, P., Kim, J. H., Detwiler, D., Berry, M., Shah, J. J., Aulig, N., Wollstadt, P., and Menzel, S. (2019). Automatically Generating 60,000 CAD Variants for Big Data Applications. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume Volume 1: 39th Computers and Information in Engineering Conference.

Ramnath, S., Haghighi, P., Ma, J., Shah, J. J., and Detwiler, D. (2020). Design Science Meets Data Science: Curating Large Design Datasets for Engineered Artifacts. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume Volume 9: 40th Computers and Information in Engineering Conference (CIE).

Reddy, B. and Fields, R. (2022). From past to present: a comprehensive technical review of rule-based expert systems from 1980 – 2021. In *Proceedings of the 2022 ACM Southeast Conference*, ACMSE '22, pages 167–172, New York, NY, USA. Association for Computing Machinery.

Reffold, C. N. (1998). Teaching and Learning Computer-Aided Engineering Drawing. *International Journal of Engineering Education*, 14(4):276–281.

Regenwetter, L., Nobari, A. H., and Ahmed, F. (2022). Deep Generative Models in Engineering Design: A Review. arXiv:2110.10863 [cs, stat].

Reich, Y. (2005). Data mining of design products and processes. In Maimon, O. and Rokach, L., editors, *Data Mining and Knowledge Discovery Handbook*, pages 1167–1187. Springer US.

Rigger, E., Münzer, C., Shea, K., et al. (2016). Estimating the potential of state of the art design automation-tasks, methods, and benefits. In *DS 84: Proceedings of the DESIGN 2016 14th International Design Conference*, pages 421–432.

Rigger, E., Shea, K., and Stankovic, T. (2018a). Task categorisation for identification of design automation opportunities. *Journal of Engineering Design*, 29(3):131–159.

Rigger, E., Vosgien, T., et al. (2018b). Design automation state of practice-potential and opportunities. In *DS 92: Proceedings of the DESIGN 2018 15th International Design Conference*, pages 441–452.

Rogalewicz, M. and Sika, R. (2016). Methodologies of knowledge discovery from data and data mining methods in mechanical engineering. *Management and Production Engineering Review*, 7(4):97–108. Publisher: Polska Akademia Nauk. Czytelnia Czasopism PAN.

Rogati, R. (2017). The AI Hierarchy of Needs | HackerNoon.

Roh, Y., Heo, G., and Whang, S. E. (2021). A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective. *IEEE Transactions on Knowledge and Data Engineering*, 33(4):1328–1347.

Romanowski, C. J. and Nagi, R. (2001). A data mining-based engineering design support system: A research agenda. In Braha, D., editor, *Data Mining for Design and Manufacturing: Methods and Applications*, pages 161–178. Springer US.

Roy, R., Kerr, C. I., and Sackett, P. J. (2006). Requirements management for the extended automotive enterprise. *Advances in Design*, pages 269–279.

Sandberg, M., Tyapin, I., Kokkolaras, M., Lundbladh, A., and Isaksson, O. (2017). A knowledge-based master model approach exemplified with jet engine structural design. *Computers in Industry*, 85:31–38.

Sandberg, S., Lundin, M., Näsström, M., Lindgren, L.-E., and Berglund, D. (2013). Supporting engineering decisions through contextual, model-oriented communication and knowledge-based engineering in simulation-driven product development: an automotive case study. *Journal of Engineering Design*. Publisher: Taylor & Francis.

Sang, H. L. (2005). A CAD-CAE integration approach using feature-based multi-resolution and multi-abstraction modelling techniques. *Comput. Aided Des.*

Santner, T. J., Williams, B. J., and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer Series in Statistics. Springer New York, New York, NY, 1 edition.

Sargent, R. G. (1988). A tutorial on validation and verification of simulation models. In *Proceedings of the 20th conference on Winter simulation - WSC '88*, pages 33–39, San Diego, California, United States. ACM Press.

Sargent, R. G. (2010). Verification and validation of simulation models. In *Proceedings of the 2010 Winter Simulation Conference*, pages 166–183. ISSN: 1558-4305.

Sargent, R. G. (2020). Verification And Validation Of Simulation Models: An Advanced Tutorial. In *2020 Winter Simulation Conference (WSC)*, pages 16–29. ISSN: 1558-4305.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61:85–117. Publisher: Elsevier.

Secco, N. R. and de Mattos, B. S. (2017). Artificial neural networks to predict aerodynamic coefficients of transport airplanes. *Aircraft Engineering and Aerospace Technology*.

Sellgren, U. (1999). *Simulation-driven design: motives, means, and opportunities*. PhD thesis, KTH.

Sexton, T. and Fuge, M. (2020). Organizing tagged knowledge: Similarity measures and semantic fluency in structure mining. *Journal of Mechanical Design*, 142(31111).

Shah, J. J. and Rogers, M. T. (1988). Functional requirements and conceptual design of the feature-based modelling system. *Computer-Aided Engineering Journal*, 5(1):9–15. Publisher: IET Digital Library.

Shan, S. and Wang, G. G. (2010). Metamodeling for high dimensional simulation-based design problems. *J. Mech. Des.*

Shao, Y., Liu, Y., Ye, X., and Zhang, S. (2018). A machine learning based global simulation data mining approach for efficient design changes. *Advances in Engineering Software*, 124:22–41.

Shephard, M. S., Beall, M. W., O'bara, R. M., and Webster, B. E. (2004). Toward simulation-based design. *Finite Elements in Analysis and Design*, 40(12):1575–1598.

Shi, F., Chen, L., Han, J., and Childs, P. (2017). A data-driven text mining and semantic network analysis for design information retrieval. *Journal of Mechanical Design*, 139(111402).

Shi, Y., Gao, T., Jiao, X., and Cao, N. (2023). Understanding Design Collaboration Between Designers and Artificial Intelligence: A Systematic Literature Review. *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW2):368:1–368:35.

Siddharth, L., Blessing, L., and Luo, J. (2022). Natural language processing in-and-for design research. *Design Science*, 8:e21. Publisher: Cambridge University Press.

Simpson, T. W., Poplinski, J., Koch, P. N., and Allen, J. K. (2001). Metamodels for computer-based engineering design: survey and recommendations. *Engineering with computers*, 17(2):129–150.

Smit, M. S. and Bronsvoort, W. F. (2009). Integration of design and analysis models. *Computer-Aided Design and Applications*, 6(6):795–808.

Smith, R. P. and Eppinger, S. D. (1997). A predictive model of sequential iteration in engineering design. *Management Science*, 43(8):1104–1120.

Song, B., Zhou, R., and Ahmed, F. (2023). Multi-modal Machine Learning in Engineering Design: A Review and Future Directions. arXiv:2302.10909 [cs].

Stauffer, L. A. and Ullman, D. G. (1991). Fundamental processes of mechanical designers based on empirical data. *Journal of Engineering Design*, 2(2):113–125.

Stoimenova, N. and Price, R. (2020). Exploring the Nuances of Designing (with/for) Artificial Intelligence. *Design Issues*, 36(4):45–55.

Storer, R. (1988). Data-driven software design using inversion. *Information and Software Technology*, 30(2):99 -107. Cited by: 0.

Stoyanovich, J., Van Bavel, J. J., and West, T. V. (2020). The imperative of interpretable machines. *Nature Machine Intelligence*, 2(4):197–199. Publisher: Springer Science and Business Media LLC.

Suh, N. P. (1998). Axiomatic design as a basis for universal design theory. *Universal Design Theory*, pages 3–24.

Sun, G. and Wang, S. (2019). A review of the artificial neural network surrogate modeling in aerodynamic design. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 233(16):5863–5872.

Sun, H., Guo, W., Shao, H., and Rong, B. (2020). Dynamical mining of ever-changing user requirements: A product design and improvement perspective. *Advanced Engineering Informatics*, 46:101174.

Tang, M., Wen, Y., Mi, X., and Dong, J. (2001). Parametric modeling with user-defined features. In *Proceedings of the Sixth International Conference on Computer Supported Cooperative Work in Design (IEEE Cat. No.01EX472)*, pages 207–211.

Tatipala, S., Larsson, T., Johansson, C., and Wall, J. (2021). The influence of industry 4.0 on product design and development: Conceptual foundations and literature review. *Design for Tomorrow—Volume 2*, pages 757–768.

Tiedemann, F., Johansson, E., and Gosling, J. (2020). Structuring a new product development process portfolio using decoupling thinking. *Production Planning & Control*, 31(1):38–59.

Turney, P. D. (2002). Data Engineering for the Analysis of Semiconductor Manufacturing Data. *ArXiv*.

Ullman, D. G. (1992). *The mechanical design process*, volume 2. McGraw-Hill New York.

Ulrich, K. and Eppinger, S. (1995). *Product design and development*. Tata McGraw-Hill Education.

Ulrich, K. T. and Pearson, S. A. (1993). Does product design really determine 80% of manufacturing cost?

V., S., I., T., and A., G. (2011). Geometric Issues in Computer Aided Design/Computer Aided Engineering Integration. *Journal of Computing and Information Science in Engineering*.

Van der Velden, C. (2008). *Application of knowledge based engineering principles to intelligent automation systems*. PhD thesis, RMIT University.

VDI, V. (1993). 2221. design handbook 2221, systematic approach to the development and design of technical systems and products.

VDI-2221 (1987). *Systematic Approach to the Design of Technical Systems and Products*. Verein Deutscher Ingenieure. Google-Books-ID: Rm3IzQEACAAJ.

Verganti, R., Vendraminelli, L., and Iansiti, M. (2020). Innovation and design in the age of artificial intelligence. *Journal of Product Innovation Management*, 37(3):212–227.

Viana, F. A. C., Gogu, C., and Goel, T. (2021). Surrogate modeling: tricks that endured the test of time and some recent developments. *Structural and Multidisciplinary Optimization*, 64(5):2881–2908.

Vlah, D., Kastrin, A., Povh, J., and Vukašinović, N. (2022). Data-driven engineering design: A systematic review using scientometric approach. *Advanced Engineering Informatics*, 54:101774.

Wall, J. (2007). Simulation-driven design of complex mechanical and mechatronic systems. Publisher: Blekinge Institute of Technology.

Wang, G. G. and Shan, S. (2007). Review of metamodeling techniques in support of engineering design optimization. *ASME, Journal of Mechanical Design*.

Wang, J., Das, S., Rai, R., and Zhou, C. (2018). Data-driven simulation for fast prediction of pull-up process in bottom-up stereo-lithography. *Computer-Aided Design*, 99:29–42.

Wang, Z., Zheng, P., Li, X., and Chen, C.-H. (2022). Implications of data-driven product design: From information age towards intelligence age. *Advanced Engineering Informatics*, 54:101793.

Ward, D. L. (1978). Data driven design on a dynamic, nonsequential problem. *IAHS-AISH Publication (International Association of Hydrological Sciences-Association Internationale des Sciences Hydrologiques)*, page 94 - 98. Cited by: 0.

Wheelwright, S. C. and Clark, K. B. (1992). *Revolutionizing product development: quantum leaps in speed, efficiency, and quality*. Simon and Schuster.

Wheelwright, S. C. and Clark, K. B. (1994). Accelerating the design-build-test cycle for effective product development. *International Marketing Review*.

Wiberg, A., Persson, J. A., and Ölvander, J. (2023). A Design Automation Framework Supporting Design for Additive Manufacturing. In *Volume 2: 43rd Computers and Information in Engineering Conference (CIE)*. American Society of Mechanical Engineers.

Will, P. M. (1991). Simulation and modeling in early concept design: An industrial perspective. *Research in Engineering Design*, 3(1):1–13.

Williams, G., Meisel, N. A., Simpson, T. W., and McComb, C. (2022). Design for Artificial Intelligence: Proposing a Conceptual Framework Grounded in Data Wrangling. *Journal of Computing and Information Science in Engineering*, 22(6):060903.

Wu, X.-Y., Hong, Z.-X., Feng, Y.-X., Li, M.-D., Lou, S.-H., and Tan, J.-R. (2022). A semantic analysis-driven customer requirements mining method for product conceptual design. *Scientific Reports*, 12(1):10139. Publisher: Nature Publishing Group.

Wynn, D. C. and Clarkson, P. J. (2018). Process models in design and development. *Research in engineering design*, 29(2):161–202.

Wynn, D. C., Eckert, C. M., Clarkson, P. J., et al. (2007). Modelling iteration in engineering design. In *DS 42: Proceedings of ICED 2007, the 16th International Conference on Engineering Design, Paris, France, 28.-31.07. 2007*, pages 693–694.

Yang, M., Jiang, P., Zang, T., and Liu, Y. (2023). Data-driven intelligent computational design for products: method, techniques, and applications. *Journal of Computational Design and Engineering*, 10(4):1561–1578.

Yildirim, N., Kass, A., Tung, T., Upton, C., Costello, D., Giusti, R., Lacin, S., Lovic, S., O'Neill, J. M., Meehan, R. O., Ó Loideáin, E., Pini, A., Corcoran, M., Hayes, J., Cahalane, D. J., Shivhare, G., Castoro, L., Caruso, G., Oh, C., McCann, J., Forlizzi, J., and Zimmerman, J. (2022). How Experienced Designers of Enterprise Applications Engage AI as a Design Material. In *CHI Conference on Human Factors in Computing Systems*, pages 1–13, New Orleans LA USA. ACM.

Yin, R. K. (2003). *Case Study Research: Design and Methods*, volume 5. SAGE Publication.

Yin, R. K. (2011). *Applications of case study research*. sage.

Yoo, S., Lee, S., Kim, S., Hwang, K. H., Park, J. H., and Kang, N. (2021). Integrating deep learning into cad/cae system: generative design and evaluation of 3d conceptual wheel. *Structural and Multidisciplinary Optimization*, 64(4):2725–2747.

Yüksel, N., Börklü, H. R., Sezer, H. K., and Canyurt, O. E. (2023). Review of artificial intelligence applications in engineering design perspective. *Engineering Applications of Artificial Intelligence*, 118:105697.

Zeleny, M. (1987). Management support systems: Towards integrated knowledge management. *Human Systems Management*, 7(1):59–70. Publisher: SAGE Publications.

Zhang, Y. and Li, H. (2011). A method of CAD/CAE data integration based on XML. *International Conference on Crowd Science and Engineering*.

Zhang, Y., Ren, S., Liu, Y., Sakao, T., and Huisingh, D. (2017). A framework for big data driven product lifecycle management. *Journal of Cleaner Production*, 159:229–240.

Zhao, Z., Jin, X., Cao, Y., and Wang, J. (2010). Data mining application on crash simulation data of occupant restraint system. *Expert Systems with Applications*, 37(8):5788–5794.

Zheng, A. and Casari, A. (2018). *Feature engineering for machine learning: principles and techniques for data scientists*. " O'Reilly Media, Inc.".

Zhou, C., , Xiyuan, Z., , and Yu, C. (2023). How does AI promote design iteration? The optimal time to integrate AI into the design process. *Journal of Engineering Design*, 0(0):1–28. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/09544828.2023.2290915.

Zorriassatine, F., Wykes, C., Parkin, R., and Gindy, N. (2003). A survey of virtual prototyping techniques for mechanical product development. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 217(4):513–530. Publisher: IMECHE.

Zwingmann, T. (2023). The Integration-Automation AI Framework.