Department of Industrial Engineering and Innovation Sciences
Human Performance Management

On the Acceptance, Adoption, and Utility of Synthetic Data for Healthcare Innovation

Master Thesis

Robin Daniël van Hoorn

Supervisors:
Dr. A. Ulfert-Blank, Eindhoven University of Technology
Y. de Jong, Philips
Dr. G. Janssen, Philips

January 29, 2024

Keywords: Synthetic Data, Technology Acceptance, Technology Adoption, Technology Utility, Framework Development, Multi-Stakeholder Analysis, Thematic Analysis, Directed Content Analysis.

Abstract

Addressing the difficulty of accessing patient data is key to advancing healthcare innovation. Consequently, healthcare organizations like Philips and the Eindhoven Medtech Innovation Center are exploring the potential of synthetic data. This artificial data, which seeks to mimic the statistical properties of the original data, could be used to share privacy-sensitive datasets or enlarge datasets. Yet, existing research on synthetic data focuses mainly on technical shortcomings, overlooking non-technical factors impacting the realization of its potential.

With the help of a diverse group of experts, this thesis qualitatively explores two perspectives, namely (1) the potential utility of synthetic data for machine learning-based healthcare innovation and (2) the (non-technical) factors influencing the acceptance and adoption of synthetic data for machine learning-based healthcare innovation. Through semi-structured interviews, qualitative analysis methods, and a follow-up survey, two frameworks are created that elucidate insights on both perspectives.

The results show that experts perceive utility in synthetic data for both the research, development, and integration & deployment stages of machine learning-based healthcare innovation. For each stage, the proposed utility framework presents several possible use cases verified by the research participants. In addition, the results show that six factors primarily influence the acceptance and adoption of synthetic data for machine learning-based healthcare innovation.

The resulting frameworks can facilitate more fruitful discussions about the acceptance, adoption, and utilization of synthetic data for machine learning-based healthcare innovation by providing a structured overview of the most important facets and considerations applicable to the process.

Acknowledgements

The completion of this research would not have been possible without the help of my thesis supervisors, Dr. Anna-Sophie Ulfert-Blank (TU/e), Ymke de Jong (Philips), and Dr. Ger Janssen (Philips). I am grateful to Philips for providing me the opportunity to execute the research on their behalf and for supporting me throughout the process. I would also like to thank the rest of the Thesis Assessment Committee, Dr. Philippe van de Calseyde, and Prof. Dr. Evangelia Demerouti. Additionally, I would like to thank Prof. Dr. Arthur Bouwman for his input throughout this project. Second to last, I would like to sincerely thank all twenty-six research participants, without whom this research would not have been possible. Lastly, I would like to thank everyone who has supported me throughout, including colleagues, friends, and family.

Managerial Summary

Over the past decade, machine learning (ML) has emerged as a potent tool for improving healthcare [164, 159, 143, 109]. The increasing demand for healthcare services and the projected shortfall of health workers by 2030 underscore the need for ML-based healthcare innovations [183, 220]. However, the healthcare industry exercises caution in adopting ML-based applications, in part due to the challenges related to accessing and processing sensitive patient data [96, 53, 191, 158]. Large and diverse datasets are crucial for effective ML applications, but obtaining patient data is difficult due to privacy regulations like the GDPR and inherent high complexity and variety of healthcare data [53, 56, 224, 53, 24].

To address these challenges, the thesis focuses on the potential of synthetic data (SD) for ML-based healthcare innovation. SD is artificially generated data that mimics the properties of real data [87]. Various scholars believe that SD can be used for both mitigating arduous data sharing processes or for the augmentation or upsampling of data [13, 25, 39, 43, 49, 87, 99, 166, 215, 216]. Existing literature focuses solely on the technical aspects of the technology, and does not look at other factors that influence the realization of the potential benefits from the technology, such as organizational criteria or cognitive biases during decision making. Consequently, this research looked beyond the technical aspects, and explored (1) the utility of SD for ML-based healthcare innovation, and (2) the practical challenges and opportunities in accepting adopting SD within healthcare innovation. This exploration is performed through the analysis of two perspectives, referred to as the *Utility Perspective* and the *Uptake Perspective*, which have the following research questions:

The Utility Perspective RQ - *How can SD be utilized throughout different stages of ML-based healthcare innovation?*

The Uptake Perspective RQ - *How can SD be accepted and adopted into ML-based healthcare innovation in a human-centered way?*

The analyses of these perspectives employed similar methodologies, and comprised extensive literature research, 24 interviews with experts, directed content analysis for the Utility Perspective, thematic analysis for the Uptake Perspective, framework development, and a follow-up survey for refining and validating the frameworks. The approach enabled the creation of a detailed understanding of (non-technical) aspects influencing the acceptance, adoption and utilization of SD for ML-based healthcare innovation.

The Utility Perspective

In the Utility Perspective, the opinions of experts on the perceived utility of SD for ML-based healthcare innovation were explored. In doing so, the analysis delved into three main aspects: the current way in which ML-based healthcare innovation operates, the involvement of patient data during this process, and the sub-processes that can be expedited using SD.

An initial foundational understanding of the ML-based healthcare innovation process was established by examining existing literature on Technology Readiness Level (TRL) frameworks tailored to ML and healthcare, which led to the identification of the Machine Learning TRL (MLTRL) framework and the construction of a Healthcare TRL (HTRL) framework, outlining the stages of innovation maturity in these fields [132, 119, 46, 68, 78, 94, 101, 182, 189]. Subsequently, the Utility Perspective focused on pinpointing specific data-driven processes leading to the construction of a conceptual model which was then used as a guideline for the 24 expert interviews and the directed content analysis of the expert interviews.

During the interviews, the experts emphasized the potential of SD in areas like algorithm development, data and method exploration, and ML model testing and training. Additionally, a consensus emerged that clinical validation requires real patient data, highlighting the limitations

of SD. The final outcome of the Utility Perspective is a refined and verified framework which takes the basic shape of a simplified TRL framework, visualized in Figure 1. The framework categorizes and ranks processes within the stages of healthcare innovation according to their potential for enhancement through SD. This framework serves as a practical guide for stakeholders considering the adoption of SD, facilitating informed discussions and decisions about its utility in specific contexts.

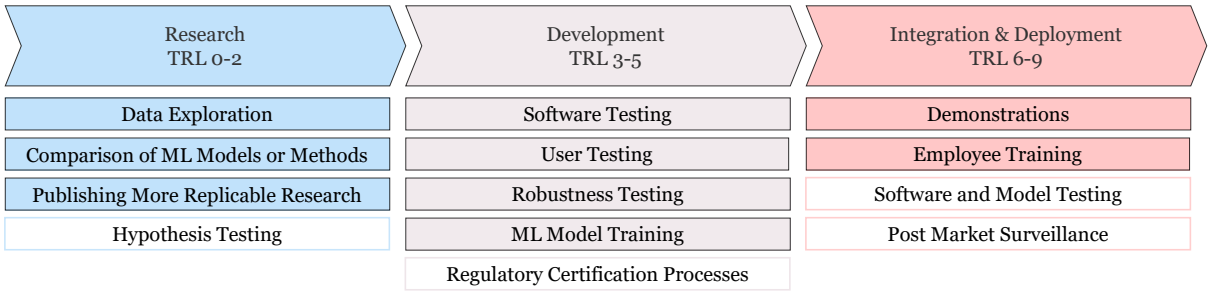


Figure 1: The Utility Perspective Framework. The uncolored use cases were deemed as significantly less likely to benefit from SD by the experts compared to the colored sub-processes.

The Uptake Perspective

In the Uptake Perspective, the challenges and opportunities of adopting and integrating SD for ML-based healthcare innovation were explored. Through a thematic analysis of the semi-structured expert interviews, six distinct themes were created which are central to the acceptance and adoption of SD: the value proposition of SD, the technical barriers to creating SD, the legal adherence when employing SD, the organizational and cultural adoption requirements for SD, the uptake of AI & data-driven innovation, and the importance of stakeholder alignment.

The thematic analysis highlights the diverse motivations and conditions for integrating SD into ML-based healthcare innovation. The findings underscore the necessity of meeting various stakeholder requirements, including understanding and proving the utility, quality, and bias of SD, ensuring legal compliance, educating stakeholders on SD, and aligning SD adoption requirements with the organizations needs and norms. From these insights, a framework was developed to guide healthcare organizations in the nuanced decision-making process regarding SD, visualized in Figure 2. This framework outlines a structured approach to facilitate discussions on the acceptance and adoption of SD.

This research builds on existing literature on technology acceptance and adoption. It confirms the relevance of factors from established models like TAM, IDT, and UTAUT in the context of SD [50, 171, 205, 8, 172, 162, 103, 209, 5, 195, 141, 116, 6, 86]. For example, perceived usefulness from TAM has significant overlap with the value proposition theme [50]. However, the themes also highlights the influential factors in a format that is more actionable for the healthcare sector than the traditional models and their variations. Consequently, the framework can be used to distinguish actionable topics that need to be addressed for the successful acceptance and adoption of SD, and facilitate focused discussions on each individual factor.

The influence of human factors in the uptake of SD was shown to be multifaceted yet less dominant compared to technological and organizational aspects. Human factors, such as experience with SD, cognitive biases such as oversimplification or overconfidence, and trust, play an influential role in shaping stakeholders' perceptions and decisions regarding SD. While these human factors are essential to recognize during the decision-making processes, they are intertwined with and often overshadowed by more prominent technical and organizational requirements and concerns in the overall uptake of SD for ML-based healthcare innovation.

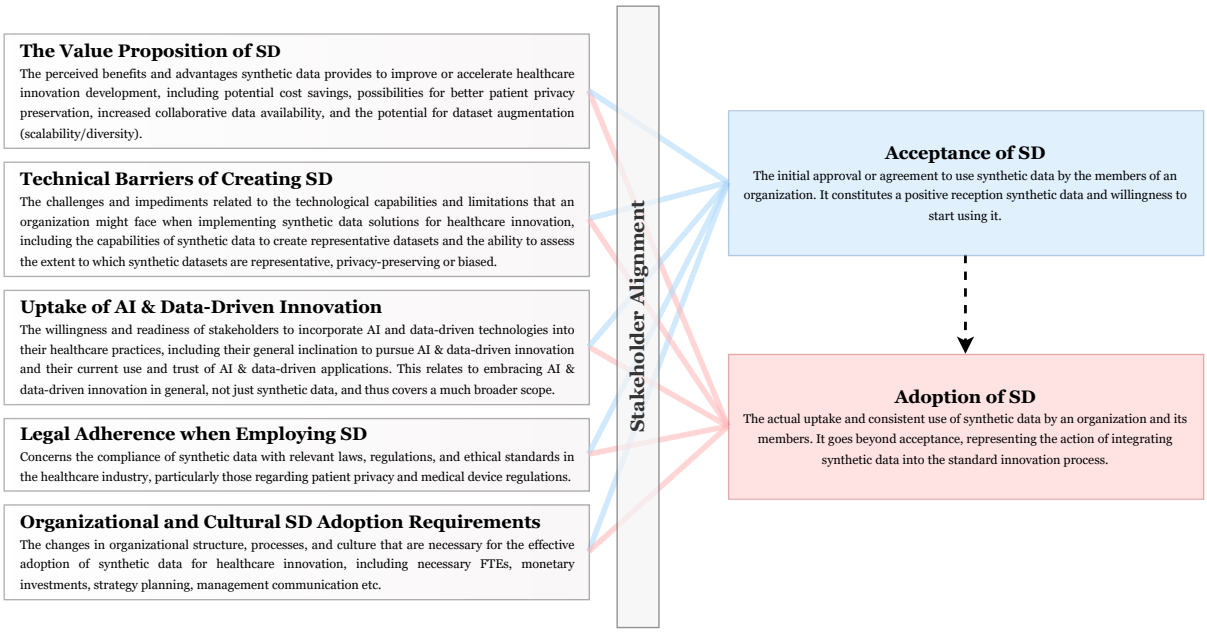


Figure 2: The Uptake Perspective Framework

Implications & Recommendations

First, the Utility Perspective suggests that stakeholders seeking to understand the value of SD in their context should carefully evaluate the identified processes within the proposed framework, considering both the potential benefits and limitations of SD. They should engage in a thorough assessment of their specific requirements, the requirements of their fellow stakeholders, and the extent to which SD can meet those needs or enhance their processes. It is important to balance the enthusiasm for new technological possibilities with a realistic view of where SD can add value, and where and when it cannot. Stakeholders must also remain mindful of the contexts where real patient data is irreplaceable, such as in clinical validation.

In addition, the Uptake Perspective suggests that stakeholders seeking to understand the value of SD in their context should engage in a comprehensive evaluation of the presented six themes, first for the acceptance of SD and subsequently for its adoption. This involves assessing the specific benefits SD can offer their healthcare initiatives, understanding the technical and legal challenges involved, and identifying the organizational changes required for successful integration. They should also prioritize mapping stakeholder requirements to ensure all parties' interests are addressed, creating a collaborative environment supporting the adoption of SD.

Based on this research, organizations seeking to understand and benefit from the potential value of SD are advised to start by understanding the perceived value of SD in their context, utilizing the Utility Perspective framework. Determining the value proposition of SD is the first step towards the uptake of SD for healthcare innovation. Next, this research suggests that discussions should be held between the necessary stakeholders to assess whether they see similar potential value, regardless of the potential barriers. After a consensus has been established on the perceived value, stakeholders are encouraged to map out requirements for each of the factors presented in the Uptake Perspective, which can differ per use case. Subsequently, stakeholders should inquire about the feasibility of addressing the requirements and concerns. Once these requirements and concerns are deemed addressable given the contextual constraints, a state of acceptance to commit can be reached. If requirements and concerns are not deemed addressable, organizations should, at least temporarily, reconsider taking up SD. Once an initial approval or agreement is constructed, stakeholders should seek to adopt the technology.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Over the past decade, machine learning (ML) has emerged as a potent tool for improving healthcare [164]. Having ML-based tools to aid healthcare professionals and automate tasks is critical to sustaining proper healthcare and making healthcare future-proof [220]. To illustrate, the World Health Organization projects a shortfall of 10 million health workers by 2030 [183]. In recent years, ML research has shown potential in many areas, such as the automation of COVID-19 detection [159], the detection of cancer using deep learning [143], or the automated and continual monitoring of in-patient deterioration [109].

Despite their benefits, the healthcare industry has shown understandable caution when integrating these ML-based solutions into their environments, such as the intensive care unit or operating room [96]. A significant hurdle is their reliance on access to and processing of large amounts of sensitive patient data for their training, testing, and validation [53, 191, 158]. Without large and diverse datasets, applications can show poor generalizability [53, 56]. Similarly, limited access to new data will make it difficult to properly update deployed applications for data or concept drifts, which can lead to reduced efficacy [224, 53].

Obtaining patient data is difficult for two primary reasons. First, obtaining access to patient data is often an arduous process which is, in part, the result of rightfully protective privacy regulations such as the General Data Protection Regulation (GDPR) [30, 207, 190]. Second, obtaining useful patient data is also difficult because of the typical characteristics of healthcare data, such as high-class imbalance, limited sample sizes, large data dimensionality, and prevalence of missing values [24]. As many industry professionals and academics alike have identified these hurdles, they have started looking into methods to mitigate these problems and gain easier access to this critically needed data.

One of the methods that has been gaining traction is the utilization of synthetic data (SD). SD is artificially generated data that mimics the properties of real data and can hence be used as a potential substitute. Although some SD generation methods are entirely expert-based, the most recent SD generation methods often generate SD using real data. Empirically, some SD generation methods have shown a potential to preserve the privacy of the original patients and thus to generate SD that cannot be traced back to the original data [58, 60, 97, 98, 147, 219]. SD generated using such methods can be considered non-personal data, which is not subject to privacy regulations and can hence be shared and disseminated without constraints. Similarly, some SD generation methods have shown to be suitable for enlarging or augmenting datasets, leading to improved downstream applications [72, 4, 128]. Consequently, these capabilities of SD are hypothesized to accelerate healthcare innovation by improving the accessibility of useful data.

Existing literature on SD often focuses on either the technical capabilities of generating SD, or the regulatory and privacy aspects of SD. As such, the literature frequently assumes that "synthetic data can significantly accelerate data science projects and reduce the cost of the software development lifecycle" [134]. However, these claims are often merely hypotheses with-

out valid evidence. There is little attention within SD literature for other practical challenges and opportunities that come with the change processes of adopting and integrating SD into innovation development.

1.1 Company Description

The parties associated to this research, Philips and the Eindhoven MedTech Innovation Center (e/MTIC) are interested in accelerating and expediting healthcare innovation.

Philips is a multinational corporation with a rich history of innovation, originally founded in Eindhoven, the Netherlands. Over the past ten years, Philips has transformed from a consumer electronics company into a health technology company [156]. Currently, Philips focuses on three broad topics being diagnosis & treatment, connected care, and personal health [155]. As a health technology company, Philips has first-hand experience with ML innovations and realizes the value of large amounts of data. SD is a potential solution to increase data availability and is, therefore, a potential enabler for faster and better healthcare innovation. Consequently, Philips is interested in understanding the potential of SD.

In addition, e/MTIC is a large-scale strategic healthcare research collaboration with a goal "to create and expand an ecosystem that strongly increases the speed of high-tech health innovation, maximizing value for patients" [36]. e/MTIC was founded in 2018 by the Eindhoven University of Technology, Philips Eindhoven, the Catharina Hospital, Maxima Medical Center, and the Kempenhaeghe Epilepsy and Sleep Center. As SD is a potential enabler for faster and better healthcare innovation, understanding its potential is directly in line with the purpose of e/MTIC.

Both Philips and e/MTIC are interested in the promises of SD. However, they also realize that there is significant uncertainty about the practical opportunities and challenges related to the technology.

1.2 Problem Statement

In the last decade, a growing body of research regarding SD has emerged, focusing mainly on technical enhancement of the technology. The existing literature presents various use cases for SD but often lacks specificity and coherence, with papers taking divergent approaches. For instance, James et al. [93] outline broad and rather vague use cases such as 'machine learning' and 'retention'. Other studies, like those by Gonzales et al. [74], mix topic-based use cases (e.g., 'epidemiological research') with process-based ones (e.g., 'algorithm testing'). For organizations like Philips or e/MTIC, these overviews fail to provide the necessary details on the precise application of SD to speed up or enhance ML-based healthcare innovation. Therefore, a clear problem emerges:

> **Problem Statement 1:** *At present, there is no structured understanding of how, where, and when SD can likely expedite or accelerate ML-based healthcare innovation.*

Recognizing the potential benefits of SD for ML-based healthcare innovation is one thing; effectively leveraging it to achieve these benefits is another. The mere technical ability to generate SD does not guarantee its successful contribution to accelerating healthcare innovation. The suitability of SD varies depending on the specifics of each project. For instance, in projects that seek to use complex electronic health records data, adopting SD may not be advisable. This is because the inherent complexity of such datasets could significantly diminish the efficacy and practical value of the SD. In contrast, projects that seek to use an often researched type of data, such as MRI images, might find using SD more beneficial [72]. The complexity of the

data is just one of many factors that could impact the utility of SD for ML-based healthcare innovation.

There is a noticeable gap in understanding the factors that influence the acceptance and adoption of SD in ML-based healthcare innovation. Most existing research has concentrated on overcoming the technical limitations of SD, with little focus on exploring the broader challenges and opportunities associated with its acceptance and adoption in practical healthcare applications. This lack of comprehensive insight into the non-technical aspects of SD use in healthcare innovation leads to the second problem statement:

> **Problem Statement 2:** *Currently, there is a very limited understanding of the factors that influence the acceptance and adoption of synthetic data for ML-based healthcare innovation, with the (academic) focus being primarily on technical constraints.*

1.3 Research Questions

Despite these problems, a consensus exists among many professionals that SD *holds the potential* to enhance and accelerate the healthcare innovation process [87, 166]. This research aims to explore the two previously outlined problem statements, enhancing the overall comprehension of SD. Both problems highlight distinct yet interrelated challenges in the current understanding of SD in healthcare to benefit ML-based healthcare innovation. The following elaborates upon the research questions that can be formulated to address the aforementioned problems.

The Utility Perspective The first problem statement focuses on the limited understanding of the specific scenarios where SD can expedite or accelerate ML-based healthcare innovation. For organizations like Philips or e/MTIC, it is crucial to comprehend the applications of SD in various stages of ML-based healthcare innovation. This exploration, henceforth referred to as the *Utility Perspective*, involves a three-step process. First, to understand how SD can enhance ML-based healthcare innovation, it is essential to understand how ML-based healthcare innovation operates in general. Knowing how ML-based healthcare innovation operates can subsequently be used to assess what parts of that process involve the use of data. Lastly, the utility of SD can be explored by elucidating which of these data-driven processes can be enhanced or expedited using SD. Accordingly, the following research question and sub-questions are formulated for the Utility Perspective.

RQ1 How can SD be utilized throughout different stages of ML-based healthcare innovation?

- RQ1a How does healthcare innovation, specifically for ML solutions, currently operate?
- RQ1b Which aspects of ML-based healthcare innovation involve the use of patient data?
- RQ1c Which parts of these processes have the potential to be enhanced or expedited using SD?

The Uptake Perspective The second problem statement deals with the limited understanding of how SD is accepted and adopted in ML-based healthcare innovation. This problem, henceforth referred to as the *Uptake Perspective*, considers the broader context of SD usage. Insights from management literature, like the technology-environment-organization framework, suggest that adopting new technologies is influenced by more than only technological characteristics [194]. Therefore, understanding the acceptance and adoption of SD for ML-based healthcare innovation requires a comprehensive analysis of these factors. The analysis of the elements influencing the acceptance and adoption of SD for ML-based healthcare innovation can be segmented into two key areas, namely stakeholder requirements and motivations for using SD and the role of human factors, which have thus far been overlooked in current SD discussions.

First, healthcare innovation can involve healthcare institutions, patients, industry partners, academic institutions, government institutions, regulation agencies, and many more. Each of these stakeholders likely has its own set of perspectives, preferences, motivations, and requirements, but the literature focuses solely on the regulations and technical capabilities of SD [153, 206]. To the best knowledge of the author, literature on SD has never documented the perceptions of healthcare stakeholders with regards to the technology. As healthcare innovation is a multi-discipline business, integrating a new technology or solution is more likely to be successful if all stakeholders believe that the integration is beneficial [154]. If the various stakeholder perspectives are unknown, it is much more difficult to successfully manage the adoption and change process for integrating SD into healthcare innovation processes.

Second, analysis of the acceptance and adoption of novel technologies in the past has shown that human factors can play a variety of roles in the various decision-making and integration processes [205, 9, 178]. All decisions made during the acceptance and adoption processes rely on the interpretations of information available to the various stakeholders, each of whom has their own set of previous experiences, requirements, and goals [69]. Human factors, such as cognitive biases, openness to novel technology, or ability to process unfamiliar information, can play an influential role in the array of decision-making processes [186]. To mitigate unnecessary inefficiencies, it is important to assess whether there are dominant human factors that influence the decision-making processes regarding the acceptance and adoption of SD for ML-based healthcare innovation.

Given the problem statement and the notable absence of stakeholder analysis, including any analysis concerning the influence of human factors, the following research question and sub-questions can be formulated for the Uptake Perspective:

RQ2 How can SD be accepted and adopted into ML-based healthcare innovation in a human-centered way?

- RQ2a What are the requirements or conditions for SD that would make it useful for its stakeholders and increase stakeholder acceptance?
- RQ2b What are the motivations for different stakeholders to, or not to, integrate SD into their ML-based innovation processes?
- RQ2c How do human factors impact the incorporation of SD in the healthcare innovation process?

1.4 Research Scope

This thesis aims to enhance the existing body of literature by creating an understanding of the circumstances under which SD can be beneficial and non-beneficial for ML-based healthcare innovation. The study focuses specifically on the realm of ML. This focus is justified as the applications and implications of SD are predominantly associated with ML-based innovations, which also constitute the main subject of most SD-related research [211, 40, 39, 147, 165].

The analyses of the two perspectives will both be exploratory in nature due to the very limited existing literature on the topic. The existing literature that covers use cases of SD will be integrated, but given the inconclusiveness and incompleteness of the literature, the analysis is still exploratory in nature.

1.5 Objectives

Academically, this thesis aims to enrich the field of SD in healthcare, focusing on its integration into ML-based healthcare innovation. The Utility Perspective aims to enhance the academic literature by providing an overview of use cases of SD, which builds upon the current

literature, while also adding depth through the aggregation of expert opinions. Additionally, the Uptake Perspective aims to enhance academic literature by building an understanding of the (non-technical) factors affecting SD acceptance and adoption, rooting them in established management theory.

The primary practical objective of this thesis is to provide actionable insights for healthcare organizations on the acceptance, adoption, and utilization of SD for ML-based healthcare innovation. This involves creating a framework elucidating when and how SD can enhance the innovation process. In addition, it entails creating a framework that presents factors that are likely to influence the acceptance and adoption of SD.

1.6 Collaborative Partnerships and Researcher Background

The proposed research is to function as a master's thesis for the master's degree Innovation Management within the Human Performance Management research group at the Eindhoven University of Technology. The research is performed in collaboration with Philips and e/MTIC. For the duration of the research, the author was employed under an internship contract within the Data Science & AI department at Innovation & Strategy at Philips. At the start of the research, the author held an MSc degree in Data Science & Artificial Intelligence, which was obtained in June 2023, on the topic of privacy-preserving longitudinal SD generation.

1.7 Structure

The remainder of this thesis follows a standard structure. Chapter 2 provides a broad overview of the relevant literature, exploring key areas of this research such as SD, innovation development processes, and technology acceptance and adoption. Drawing from the insights of this literature review, Chapter 3 presents the methodology for answering the aforementioned research questions and elaborates upon why this approach is taken. Chapters 4 and 5 subsequently elaborate on the analysis of the two perspectives individually. Lastly, Chapter 6 presents an elaborate discussion of the two perspectives and presents the final conclusions.

Chapter 2

Relevant Literature

This section presents a broad review of the main research fields pertaining to the research questions. First, an overview of the literature regarding ML-based innovation in the healthcare sector is provided for general context. Subsequently, SD is shortly elaborated upon, explaining its essence and providing an overview of various SD generation methods. The purpose of the literature on SD is to provide context about SD as an enabler for healthcare innovation and not to go into depth regarding the technical aspects. Following the literature regarding the context of this study, the applicable literature for the Utility Perspective elucidates upon innovation development processes as well as published use cases of SD. Lastly, the applicable literature for the Uptake Perspective is presented, elaborating extensively upon technology acceptance and adoption literature.

2.1 Machine Learning in Healthcare

This thesis investigates how SD can be integrated and used to benefit ML-based innovation processes in healthcare. To answer any research question successfully, an understanding of its context is crucial. ML is a field that concerns itself with the construction and study of systems that can automatically find patterns, structures, and relationships from input data [223]. In the past decade, ML methods have shown significant potential to be effective at detecting infections [159, 163] and diseases such as cancer [143, 181] or heart disease [15, 145]. There are a plethora of different types of healthcare applications for which ML can be applied. Well-known types include predictive analysis systems [150, 180], operational efficiency optimization systems [157, 148], or decision support systems such as image analysis tools [123, 161]. In addition, innovations such as healthcare chatbots [27], medical supply chain management systems [1], surgical navigation systems [208] or healthcare fraud detection systems [20] can all be considered ML-based healthcare innovations.

2.1.1 Stakeholders within ML-based Healthcare Innovation

In healthcare innovation, the integration and utilization of ML is fundamentally driven by clinical experts. They are central to the innovation process, providing essential medical knowledge and context [193]. Research scientists and (ML) engineers collaborate with these clinical experts. Their responsibility often is to develop and refine ML algorithms, drawing upon a deep understanding of ML principles and their application within the healthcare sector [42]. The success of their work is linked to the input and ongoing involvement of the clinical experts. However, besides clinical experts, research scientists, and (ML) engineers, patients also have a role in this ecosystem. They can provide invaluable feedback and are often the primary source of data for these ML algorithms.

Apart from clinicians, researchers, engineering teams, and patients, the landscape of ML-based healthcare innovation includes policymakers, ethical review boards, regulatory bodies, patient advocacy groups, International Standards Organizations, and many others. They have the challenging task of ensuring the ethical and legal integrity of applications, scrutinizing them for issues such as bias, fairness, and privacy [139, 90]. This is especially important in the context of SD, where the ability to generate realistic patient data brings a whole new set of ethical and legal challenges. These stakeholders help to balance the benefits of innovation with the protection of patient rights and the societal implications of these technologies [152].

2.1.2 The Need for Patient Data

ML-based innovations in healthcare are heavily dependent on access to large quantities of high-quality patient data, as the underlying algorithms learn patterns from this data and apply these patterns to make predictions or decisions [23]. For instance, in the development of an ML model to predict disease progression, patient demographic data, clinical measurements, and outcomes can be used to train the model. The accuracy of these models is heavily related to the quantity and quality of the data used for their development [64]. Therefore, the availability of patient data is an important factor for the successful development and implementation of ML-based healthcare innovations.

However, there are significant challenges associated with obtaining useful patient data. The first major challenge relates to privacy and security [202]. The sensitive nature of health data has resulted in rightfully stringent regulations to ensure its safe use and prevent any potential misuse. As a result, patient data-sharing processes are often arduous, which is in part to ensure compliance with these regulations [30, 207]. Balancing the need for data to fuel ML-based innovation with the need for protecting patient privacy is a substantial challenge [37].

Another challenge of acquiring useful patient data is associated with the inherent traits of healthcare data, including significant class imbalances, constrained sample sizes, extensive data dimensionality, and a common occurrence of missing values [24]. These characteristics make for very complex data, resulting in a need for high data availability and diversity to understand these complexities.

When applications are developed with insufficient data, resulting applications might not generalize sufficiently, contain harmful biases, or suffer in outcome quality [61]. Therefore, exploring solutions such as SD, which has the potential to increase data availability, can be very useful in the field of ML-based healthcare innovation.

2.2 Synthetic Data in Healthcare

SD generation can be seen as a sub-field of ML. In essence, SD refers to data that is not collected from real-world events but instead is artificially created through algorithms and computational models [87]. SD generation methods are designed to simulate the behaviors and characteristics of real data, encapsulating its patterns and correlations. SD is often thought to be applicable for a variety of use cases, ranging from preliminary system testing and simulations to training ML algorithms [39]. It is essential to understand that, although SD generation is a form of ML, it cannot directly be applied in healthcare like a prediction model. SD generation methods generate data, which can then potentially enable faster and better innovation. By increasing the availability and accessibility of useful data, SD generation methods are thought to be able to help accelerate healthcare innovation and help build healthcare innovations of better quality.

Although the promises of SD are intriguing, there are also several drawbacks. First, generating SD for healthcare is not easy, as illustrated by the many pieces of literature that have so far been written on improving SD generation algorithms. The main technical difficulties reside in the clinical accuracy of SD, its ability to preserve the privacy of patients, and the uncertainties

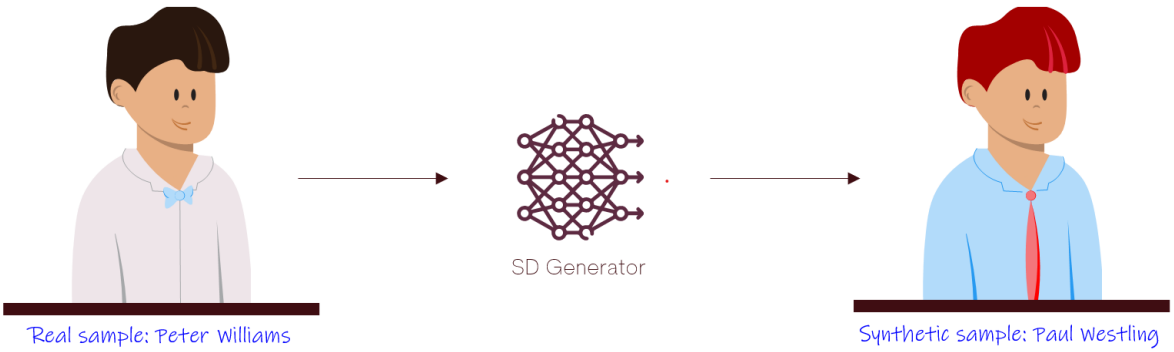


Figure 2.1: A visualization of the fundamental idea of modern SD generation techniques. To generate SD, an SD generator is trained on real data (i.e. Peter Williams), after which it can generate a synthetic sample (i.e. Paul Westling).

surrounding (synthetic) biases in the generated data [199, 17, 117]. Second, there is a high amount of uncertainty with respect to regulations regarding what will or will not be allowed in practice, both with respect to clinical validation and with respect to privacy [39].

2.2.1 Approaches to Synthetic Data Generation

Early SD generation techniques, pre-2010, were largely rooted in statistical methods and simplistic algorithms [175]. Bootstrap methods were widely used, creating resamples of existing data sets to provide additional, though not entirely unique, data [59, 95, 168]. Other statistical procedures, such as Monte Carlo simulation, were also employed, generating SD by running multiple simulations based on probabilistic distributions [140]. In essence, these statistical methods rely heavily on a sound understanding of the underlying statistical distributions and patterns within the original dataset, with the SD reflecting the same characteristics. Statistical SD generation methods are still widely researched and used in the biomedical domain.

In recent years, the advent of more complex deep learning models, such as Generative Adversarial Networks (GANs) [75], Variational Autoencoders (VAEs) [108], diffusion models [89], normalizing flow models [120], and transformer-based architectures [201], has significantly advanced the field of SD generation. The idea of these methods is to use real (patient) data to generate SD, as also visualized in Figure 2.1. These modern approaches to SD generation have opened new possibilities for creating more diverse, more realistic high-dimensional datasets.

Beyond the realm of purely ML-based techniques, there exist numerous methods that blend these approaches with expert knowledge, thereby crafting hybrid models that capitalize on both ML algorithms and domain-specific insights [29, 54, 222]. Such hybrid expert-system ML methodologies effectively harness the computational power of ML, while also incorporating the nuanced understanding that subject matter experts bring. These methods usually attempt to fuse artificial and human intelligence to increase the performance or clinical validity of the models.

2.2.2 An Example of Synthetic Data in Healthcare

One example of how SD can be applied in healthcare is with the work of Gheorghiță et al. [72], who employ SD to augment their training dataset and improve the accuracy of their prediction model. They address the problem of cardiac function quantification by analyzing MRI heart images and attempt to predict clinically relevant volumes from an MRI scan. This task is challenging, partly due to a high class imbalance in their dataset, which lacks diversity in cardiovascular disease phenotypes. To address this, they create additional SD based on the available real data to expand the dataset. They use both the synthetically generated data and

the real data to train their prediction model. Their results show that they can use SD for training an ML model to significantly improve predictions compared to using only the limited available real data. As such, they successfully employ SD to enhance the performance of their healthcare application.

2.3 Relevant Literature for the Utility Perspective

There are two distinct topics for which existing literature is important with regard to the analysis of the Utility Perspective. First, there is a significant body of literature that can provide insights into how ML-based healthcare innovation operates, which is relevant for RQ1a and RQ1b. Second, there is some existing literature that proposes use cases of SD, which is relevant to answering RQ1c. Consequently, this section will first elaborate upon innovation development processes. Subsequently, this section will present a systematic literature review of SD use cases, complemented by a narrative literature search of data-driven processes from ML engineering and systems engineering to improve completeness.

2.3.1 Innovation Development Processes

It is essential to understand how ML-based healthcare innovations are developed, such that a better understanding can be created of how SD can enhance that process. It is first essential to zoom out and understand how innovations, in general, come to be. Developing an innovative healthcare idea into a successful product or service is a complex and challenging process, as highlighted by the funnel of innovation [63]. Innovation, following Fagerberg's definition, is the process that connects invention to practical application [66].

The complex transformation from invention to application necessitates a balance between exploring new ideas and capitalizing on established ones, involving both human and financial resources, as well as embracing associated risks [18]. Diverse research streams have evolved to navigate the innovation process. For instance, implementation science focuses on integrating research into practice [21], yet it generally neglects the research and prototyping phases crucial in healthcare. Similarly, models like the Technology Acceptance Model [50] and the Theory of Innovation Diffusion [170] explore user reception of innovation and thus focus on the user side of what makes an innovation successful.

Given role of SD in expediting healthcare innovation, a focus on individual components or sub-processes of the innovation development process is more pertinent. Identifying these components enables the alignment of SD use with specific stages of the process, providing structured responses to the research question. The following sections will further explore process models and technology readiness assessments that define these components of the healthcare innovation development process.

2.3.1.1 Process Models for Guiding Innovation

The first group of frameworks that provides structures of innovation development processes are called process models. Process models are methods that provide a systematic approach to translating ideas into solutions [92]. In the context of innovation management, process models act as blueprints for managing the life cycle of innovative ideas, from their genesis to their implementation. They break down the journey of an idea, from inception to market delivery, into manageable phases, allowing organizations to effectively allocate resources, manage risks, and measure progress.

The purpose and significance of process models in guiding innovation development cannot be overstated. Process models help mitigate the inherent risks and uncertainties of innovation by providing a step-by-step guide to navigate the innovation journey. They help in identifying potential pitfalls and roadblocks in the process, allowing for proactive risk management.

Moreover, they facilitate effective communication and collaboration within teams, such that team members understand the process and know what to expect at each stage. These models also provide a framework for analyzing and improving the innovation process, leading to more successful innovation outcomes. By using a systematic and structured approach, organizations can increase the predictability and success rate of their innovation initiatives.

There are three fundamental process models: the waterfall model [174], the spiral model [31], and the stage-gate framework [44]. However, there are a plethora of other process models that are inspired by the general concepts of the three fundamental models and usually focus on a specific application field or sub-problem such as healthcare [47, 129, 125].

The Waterfall Model The Waterfall Model, introduced by Winston W. Royce in 1970, is a linear sequential software development methodology, often described as a cascading flow down through the phases of requirements analysis, design, implementation, testing, and maintenance [174]. Each phase is dependent on the delivery and approval of the previous one, with progression moving steadily downward, like a waterfall. While this method has the benefit of simplicity and easy understanding, it has limitations, mainly its lack of flexibility to adapt to changes once the project is underway, which is often a necessity in software development.

The Spiral Model The Spiral Model, proposed by Barry Boehm in 1988, was developed in response to the limitations of the Waterfall Model [31]. This approach combines the features of the Waterfall Model and the iterative nature of prototyping. The Spiral Model uses a risk-driven approach, where each spiral represents a phase of the software process and loops through four stages: objectives determination, risk assessment and reduction, development and testing, and planning the next iteration. Each loop is a fully developed phase of the software, with each subsequent spiral building upon the output from the previous one. This enables early detection and handling of risks, allows for changes during development, and incorporates user feedback early in the project. Despite its advantages, the Spiral Model requires careful management and a clear understanding of risks to be effective.

The Stage-Gate Framework The Stage-Gate model, first introduced by Dr. Robert G. Cooper in the 1980s, is a conceptual and operational framework for managing new product development processes [44]. It organizes the innovation process into a predefined set of stages and gates. In each stage, teams complete a set of related cross-functional tasks such as preliminary market assessment, technical feasibility, and business viability. Between each stage, a gate serves as a decision point where management assesses the completed work, evaluates the business case, and decides whether to continue the project, redirect it, or halt it. This ensures that only high-quality projects that align with the organization's strategy and have a reasonable chance of commercial success proceed further. The Stage-Gate model offers a disciplined, systematic approach to product innovation, reducing uncertainty and improving the chances of launching successful products.

Over time, the Stage-Gate model has evolved to incorporate changes in the business environment and feedback from companies that have implemented it. The original model, often referred to as the 'Classic Stage-Gate', was mostly a linear and rigid process. However, over the years, Cooper introduced several iterations of the model, such as 'Stage-Gate Xpress' and 'Stage-Gate Lite', which offer different possibilities for projects with varying levels of risk and novelty [45]. These newer versions provide more flexibility, such as overlapping stages, spiral development, and concurrent activities to accelerate the process. Furthermore, the model has been adapted to incorporate agile and lean methodologies and now embraces digital transformation, fostering cross-functional collaboration and data-driven decision-making in innovation.

2.3.1.2 Technology Readiness Assessment

In addition to process model research, the research stream of technology readiness assessments also focuses on providing context on what the innovation process looks like. A technology readiness assessment (TRA) is a systematic, evidence-based process used to evaluate the maturity of a particular technology or capability [131]. It is a collective term for tools that provide a common language and standard of measurement for assessing the readiness of a technology to be implemented or commercialized. The concept of TRA was developed by NASA in the 1970s and has since been adopted and adapted by many other industries and organizations [127, 131]. TRA is typically conducted using specific frameworks, the most well-known of which is the Technology Readiness Level (TRL) framework [132].

The importance of TRA stems from its role in managing technological risk and informing decision-making processes. By providing a standardized method for assessing a technology's readiness, TRA can help organizations to make more informed decisions about which technologies to pursue further, when to implement them, and how to manage the associated risks. Moreover, TRA can help in identifying potential gaps or challenges that need to be addressed before a technology can be successfully implemented.

Technology Readiness Level Framework The TRL framework, first developed by NASA in the 1970s, has since become a widely used framework for assessing technology readiness. This system consists of nine levels, each representing a stage in the development and deployment of a technology, from TRL 1 (basic principles observed and reported) to TRL 9 (actual technology proven through successful mission operations) [132]. The objective of the TRL scale is to provide a common language that enables communication and understanding between different stakeholders, such as scientists, engineers, project managers, and decision-makers, about the maturity and risks associated with a specific technology. A summary of the original TRL framework of NASA is provided in Table 2.1.

TRL	Description
1	Basic principles observed and reported
2	Technology concept and/or application formulated
3	Analytical and experimental critical function and/or characteristic proof-of-concept
4	Component and/or breadboard validation in laboratory environment
5	Component and/or breadboard validation in relevant environment
6	System/subsystem model or prototype demonstration in a relevant environment (ground or space)
7	System prototype demonstration in a space environment
8	Actual system completed and "flight qualified" through test and demonstration (ground or space)
9	Actual system "flight proven" through successful mission operations

Table 2.1: A summary of the original TRL framework as published by Mankins et al. [132], representing NASA.

Extensions of the TRL Framework The TRL framework has been adapted and extended in a variety of ways to accommodate the different aspects and stages of technology development and deployment. For instance, the Integration Readiness Level is a scale used to assess the readiness of a system or component to be integrated into a larger system [126]. The Systems Readiness Level goes a step further to evaluate the maturity of the entire system, looking at

factors such as the integration of components and system-level testing [176].

Similarly, the Manufacturing Readiness Level, developed by the U.S. Department of Defense, is designed to measure the maturity of the manufacturing processes, considering aspects like process capability, control, quality management, and more [214, 19]. Another important extension is the Operational Readiness Level, which assesses the readiness of a system to be operated in its intended environment, taking into account factors like operator training, maintenance procedures, and user manuals [46, 78].

In healthcare, variations of the TRL framework are used to evaluate the readiness of medical technologies for clinical use, considering factors like safety, efficacy, and regulatory approval [46, 68, 78, 94, 101, 182, 189]. In the field of ML, TRL variations, such as the recent MLTRL framework by Lavin et al. [119], consider factors like algorithm development, data readiness, and real-world testing under various scenarios. These domain-specific variations of the TRL framework recognize the unique challenges and considerations in each field, providing a more accurate and meaningful assessment of technology readiness.

Evaluating Limitations and Quality of TRL Frameworks While the TRL framework is extensively utilized and recognized in various fields, the basic framework and its various extensions are not without their limitations [185]. Primarily, TRL frameworks emphasize technological readiness, often at the expense of other vital factors such as market readiness, regulatory compliance, and business risks. These elements are integral to the successful deployment and adoption of technology. Moreover, the subjective nature of TRL definitions can lead to inconsistent assessments, complicating the framework's reliability.

In reviewing the literature on TRL framework variations, various notable limitations with respect to quality can also be observed. A common shortfall is the lack of detail in their development [68, 94, 101]. This is in large part visible through the absence of any standardized approach in constructing these frameworks and the inherent difficulty in proving their validity. For example, Fleuren et al. [68] proposes a TRL framework specifically for ML in healthcare, which seems relevant to this thesis. However, the paper's limited scope (just two pages), minimal references, and vague framework development process detract from its authority.

To effectively evaluate the quality of TRL-related publications, one can consider the number of times the work has been cited and the depth of their analysis. Note, however, that recent publications, such as [101], might have fewer citations compared to older works like [189], which is not necessarily an indicator of their quality. A robust evaluation should include the extent of referenced works, the application of grounded theory, empirical validation, and the quality of the writing. Empirical validation could involve interviews with experts [115], or collaborative development among a diverse group of authors, as seen in the MLTRL framework [119]. The MLTRL framework, conceptualized by authors from diverse institutions, provides a more comprehensive and validated approach despite its lack of systematic external validation. The diversity of authors' perspectives and the presentation of detailed use cases add significant depth to the framework and therewith address common limitations found in TRL methodologies.

In summary, while the TRL framework is a valuable tool in assessing technical readiness, its limitations in addressing market, regulatory, and business factors, as well as its potential for subjective interpretation, need careful consideration. The quality of TRL frameworks in literature can be gauged by their citation breadth, theoretical grounding, and empirical validation, which help in mitigating some of these limitations.

2.3.1.3 Comparing Process Models and TRL Frameworks

The original TRL framework, as introduced by NASA, is not a process model. It is a guiding framework for classifying the state of a technology. It is not meant to provide any guidance on the development of technology. Technology is to be classified on the component which is least

advanced [118]. Hence, it is possible that some components are TRL 6, while the technology as a whole is classified as only being TRL 4.

Nevertheless, it is easy to see how one would apply the TRL framework as a guide for the development of innovation by using it as a linear development model, following the ideas of the initial waterfall models [174]. It is no surprise, then, that some publications transform their TRL extension into a type of process model. Again, take the MLTRL framework by Lavin et al. [119], who define specific 'switchback mechanisms' that often occur in the ML innovation development process. The MLTRL framework is not simply a classification framework but features exit-enter criteria and switchback mechanisms to guide the innovation process, effectively making it a process model.

2.3.2 Literature on Use Cases of SD

Apart from needing to understand how ML-based healthcare innovation works, it is essential for the analysis of the Utility Perspective to integrate the already published literature presenting use cases of SD. Understanding which processes are part of the ML-based healthcare innovation process and which of those processes can benefit from SD is crucial to gaining an understanding of exactly how and when SD can be utilized within ML-based healthcare innovation. This section analyzes the literature that identifies use cases of SD through a systematic literature review. This literature is complemented by literature from ML and systems engineering to establish a complete context of the processes that are part of ML-based healthcare innovation development. These two fields encompass a large extent of literature related to developing ML-based healthcare applications.

Systematic Literature Review Specifications In June 2023, a systematic literature review was performed to assess all proposed use cases and applications of SD in the literature. Three key aspects were included in the search strings, namely (1) a synonym of 'use cases', (2) a mention of 'overview' or 'review' and (3) a mention of 'synthetic data'. Additional scopes of 'healthcare' or 'ML' were not added as they resulted in too much of a reduction of the search results. The exact search strings for the systematic literature review are provided in Appendix A.

Three databases were exhausted for literature, namely SCOPUS, PUB-MED, and Web Of Science. In total, SCOPUS returned 72 results, PUBMED returned 18 results, and Web of Science returned 21 results. From the aggregate results, 23 duplicates were deleted. Subsequently, results were skimmed to assess whether use cases of SD were discussed in the paper, leaving 32 publications.

The remaining 32 publications were assessed on their mentions of use cases of SD. Another 12 articles were identified through snowballing on the results of the systematic literature search [212]. A PRISMA specification of this systematic literature review is provided in Appendix A.

Analysis of current taxonomies The systematic literature review revealed that researchers and industry experts foresee numerous uses for SD that can broadly be grouped into cases where real data is (1) scarce or (2) subject to privacy concerns [13, 25, 39, 43, 49, 87, 99, 166, 215, 216]. Both these situations are typical in healthcare. Rare conditions and edge cases are abundant in healthcare data; expanding limited datasets using SD or making data more available can turn a potentially dismissed concept into a plausible solution.

Besides these two uses of SD that are a direct result of their technical characteristics, applied use cases of SD vary widely [74, 93, 135, 217]. Many different taxonomies of use cases for SD have been published, but they often suffer from three problems. First, the individual items in the taxonomies often overlap or have different scopes. Similarly, although one would expect such taxonomies to be then presented in a hierarchical form, no hierarchical taxonomies on the

use cases of SD have been published. For example, Gonzales et al. [74] provides a taxonomy of seven use cases in healthcare, based on a narrative review of the literature, that contains (1) simulation and prediction research, (2) hypothesis, methods, and algorithm testing, (3) epidemiology/public health research, (4) health IT development, (5) education and training, (6) public release of datasets, and (7) linking data. Although the use cases are interesting, the categories 'epidemiology/public health research' and 'simulation and prediction research' clearly overlap. Where 'epidemiology/public health research' is a topic-based use case, 'simulation and prediction research' is a type of research.

Second, because the creation of these taxonomies is poorly documented, it is impossible to assess the quality of the taxonomies properly. Gonzales et al. [74] uses "a narrative review", but does not provide any details on how the taxonomy was created from that review. James et al. [93] does not mention doing a narrative review and simply presents the use cases right after the introduction. Both taxonomies provide valuable insights, but it is difficult to assess their quality.

Lastly, none of the taxonomies provide any reference to previously created taxonomies. Consequently, trying to provide an overview of the taxonomies results in an unstructured clutter of taxonomies that all differ in scope, focus, and quality.

2.3.3 Data Processes in Machine Learning Innovation Development

The literature on SD is not conclusive enough to understand what processes can, and cannot, benefit from employing SD. For example, data-driven processes that cannot be expedited through employing SD are likely not named in the SD literature. Therefore, it is also key to be aware of data-driven processes that are part of ML-based healthcare innovation but not mentioned in SD literature. This section, therefore, supplements the findings of the previous systematic literature review with literature on ML and healthcare innovation processes.

For this, a systematic literature review is not very suitable. A SCOPUS search using terms such as 'machine(-)learning,' 'system(s) engineering,' and 'development' returns 241 results but only includes six potentially interesting papers upon skimming the titles and abstracts. The other databases return fewer results. Consequently, the six remaining papers are used as a basis for a more informative narrative review.

Analysis of Relevant Literature Research on ML innovation can provide additional context and details regarding the use cases identified in the SD literature. For instance, although SD literature specifies 'automated software testing' as a potential use case for SD, the ML testing framework from Google can provide context by specifying different types of ML tests [35]. Subsequently, each of these types of tests can be classified for their potential to be able to benefit from SD.

Interestingly, many papers that introduce process frameworks can provide additional details on ML solution development, as they can feature use case descriptions with elaborate explanations of different processes that are part of the different stages in these frameworks. A good example of this is the MLTRL paper [119], which, after introducing the framework, elaborately describes several use cases and different processes in which data is used.

There is more implementation science or process-oriented framework literature where inspiration can be sought. For example, Nguyen et al. [146] mention the importance of anomaly detection, diagnostics, and prognostics. Similarly, Li et al. [121] underline the importance of environment simulations and modeling. Lastly, Vairo et al. [197] show that data management and data verification are integral parts of the ML innovation process. All such literature adds depth to the previously identified use cases of SD.

2.4 Relevant Literature for the Uptake Perspective

Now that the relevant literature for the Utility Perspective has been elaborated upon, the literature pertaining to the Uptake Perspective will be presented. This research regards the established understanding of how a (novel) technology is accepted and adopted by end-users. Research on technology acceptance and adoption analyzes the antecedents of motivation to integrate novel technologies, as without acceptance and adoption, none of the promised benefits can be reaped. Focusing on technology acceptance is important, particularly in the context of healthcare innovation. The healthcare sector often faces unique challenges in adopting new technologies, such as concerns about data security, ethical considerations, and the need for alignment with existing clinical practices [26, 55]. These challenges make it important to understand not just the technological aspects but also the human factors influencing the acceptance of such innovations. By examining the factors that drive or hinder acceptance, strategies can be developed to facilitate smoother integration into healthcare systems, ensuring that the potential benefits of SD can be fully realized. This focus on acceptance helps bridge the gap between technological capability and practical usability, ensuring that the innovations meet the real-world needs and preferences of healthcare professionals and patients alike.

The following review intends to first dissect the classical theoretical foundations of technology acceptance and adoption models. The classical models will include the Technology Acceptance Model [50], the Unified Theory of Acceptance and Use of Technology [205], Innovation Diffusion Theory [171], the Theory of Planned Behavior [3] and the Technology-Organization-Environment framework [194]. By using the insights from these theories, this review aims to shed light on the psychological, socio-cultural, and contextual factors influencing the acceptance and adoption of technologies in the field of ML and healthcare innovation. Subsequent to the presentation of the theoretical foundation, an elaborate discussion on further derivatives of these models in the context of ML and healthcare will be provided, as well as their applicability to the context of SD acceptance and adoption from a firm-level perspective.

2.4.1 The Technology Acceptance Model

The Technology Acceptance Model (TAM) is one of the most influential models in the technology acceptance literature [50]. Rooted in the Theory of Reasoned Action, TAM focuses on the determinants of acceptance that predict behavior of use. This model posits two primary factors influencing users' decision to use a technology: perceived usefulness and perceived ease of use. Perceived usefulness is the degree to which a user believes that using a particular system would enhance their job performance, while perceived ease of use refers to the degree to which a user believes that using the system would be free from effort [50].

Since its inception, TAM has seen multiple iterations, referred to as TAM2 and TAM3 [204, 203]. These extensions of TAM include additional factors to concretize perceived usefulness and perceived ease of use, including aspects such as job relevance, result demonstrability, voluntariness, and experience. Apart from these iterations on the original model, there are a plethora of studies applying TAM to certain contexts or extending TAM with additional factors such as trust [70].

2.4.2 Innovation Diffusion Theory

Innovation Diffusion Theory (IDT) offers a model for understanding how, why, and at what rate new ideas and technology spread among individuals and communities [171]. According to IDT, four main elements influence the spread of a new idea: the innovation itself, communication channels, time, and the social system. IDT defines innovation as an idea, practice, or project that is perceived as new by an adopter. Communication channels refer to the means by which information about the innovation is transmitted, while time relates to how long it takes for the

innovation to be adopted. Finally, the social system denotes a set of interconnected entities that are engaged in joint problem-solving to accomplish a common goal [171].

IDT's importance stems from its comprehensive perspective, taking into account the technological, social, and individual factors in understanding the diffusion and acceptance of innovations. Human factors underpin the IDT. For instance, the 'innovation-decision process', a concept in IDT, is a process that an individual passes through in deciding to accept or reject an innovation, involving stages such as knowledge, persuasion, decision, implementation, and confirmation. This process is inherently influenced by individuals' perceptions, attitudes, and social interactions [171]. These factors highlight the critical role of human factors in the diffusion and acceptance of new technologies, further underscoring the relevance of IDT in studies of technology acceptance and adoption.

Critical to the adoption of technology, as underscored by IDT, are the perceived characteristics of the innovation, listed under the 'persuasion' stage in the original model [171]. Here, Rogers [171] specified relative advantage, compatibility, complexity, trialability, and observability as key components for persuasion and, therewith, adoption of new technology. Relative advantage is the extent to which the adopter perceives the innovation to be better than the alternatives. Compatibility refers to the extent to which the innovation is compatible with the adopters' values, norms, and perceived needs. Complexity is included to indicate that innovations that key players perceive as 'difficult to use' will be less likely to be adopted. Trialability regards the extent to which innovations can be tested or experimented with before key decisions need to be made. Lastly, observability refers to the extent to which the benefits of an innovation can be observed easily by adopters. Greenhalgh et al. [76], who conducted a systematic review on IDT, added reinvention to this list, referring to the opportunities that adopters have to adapt or modify the innovation to their own needs. In contrast, Agarwal and Prasad [2] found that only relative advantage, complexity, and compatibility are consistently impactful factors of innovation adoption.

2.4.3 The Theory of Planned Behavior

The Theory of Planned Behavior (TPB) is a psychological model that aims to predict human behavior based on a set of beliefs. According to TPB, an individual's behavior is determined by their intention to perform the behavior, which is, in turn, shaped by three factors: attitude towards the behavior, subjective norms, and perceived behavioral control. Attitude towards the behavior refers to the individual's feelings about performing the behavior. Subjective norms reflect the perceived social pressure to perform or not perform the behavior and perceived behavioral control pertains to the individual's perceived ease or difficulty in performing the behavior [3]. It is noteworthy that this theory was originally not considered in the context of technology, innovation, or adoption but much more general as a theory of behavioral prediction. The framework is solely focused on the individual and hence lacks any form of organizational context.

2.4.4 The Technology-Organization-Environment Framework

The Technology-Organization-Environment (TOE) framework is a model used in information systems research to understand the process of technological innovation adoption in organizations [194]. It posits that the decision to adopt innovations is influenced by three contextual factors: technological context (the internal and external technologies relevant to the organization), organizational context (descriptive measures like size, scope, and managerial structure), and environmental context (the structure of the industry, presence of competitors, and access to resources). This framework can be used for analyzing the adoption of complex technological systems where multiple factors interplay to influence organizational behavior and decision-making

and is hence often used to analyze adoption on a firm level, as opposed to, for example, TPB, which takes an individual focus.

2.4.5 The Unified Theory of Acceptance and Use of Technology

The Unified Theory of Acceptance and Use of Technology (UTAUT) was developed as a model for explaining user acceptance and usage behavior [205]. This model combines constructs from eight prominent user acceptance models, including TAM, IDT, and TPB. The UTAUT proposes four core determinants of usage intention and behavior: performance expectancy, effort expectancy, social influence, and facilitating conditions. Performance expectancy refers to the degree to which an individual believes that using the system will help him or her to attain gains in job performance. Effort expectancy is the degree of ease associated with the use of the system. Social influence is the degree to which an individual perceives that important others believe he or she should use the new system. Facilitating conditions are the degree to which an individual believes that an organizational and technical infrastructure exists to support the use of the system [205]. The previously mentioned frameworks are partly subsumed in the UTAUT framework. For example, factors from TAM can clearly be seen in UTAUT as the performance expectancy and effort expectancy concepts from the UTAUT framework are heavily related to the perceived usefulness and perceived ease of use concepts from TAM [105].

UTAUT acknowledges the influence of both human factors and organizational factors on technology acceptance. For instance, the role of social influence in this model highlights how the opinions of peers, supervisors, and other influential individuals within the user's social network can impact their acceptance of a new technology. In contrast, effort expectancy can be hampered by organizational bureaucracy or regulatory impediments. Furthermore, the construct of facilitating conditions underscores the impact of users' perceptions of available resources and support for using the technology. This highlights the importance of considering the users' environment and context in understanding and predicting technology acceptance and use. UTAUT also acknowledges the influence of human factors through similar means as TAM, as the UTAUT performance expectancy and effort expectancy concepts are heavily related to the perceived usefulness and perceived ease of use concepts from TAM [105]. Accordingly, they incorporate the human factor in a similar manner.

2.4.6 Classical Acceptance and Adoption Frameworks in Healthcare

Research in healthcare technology often applies classical acceptance and adoption frameworks like TAM, IDT, or UTAUT. This synthesis is evident from numerous systematic reviews focusing on different aspects, like TAM-UTAUT in healthcare [8] and remote healthcare [172], or TAM in health informatics [162]. These reviews collectively highlight the relevance of these frameworks but also reveal a lack of consensus on an 'optimal version' for healthcare services [162, 172]. Particularly, Rouidi et al. [172] and Khanijahani et al. [103] provide contrasting views on the role of social influence, underscoring the high complexity of technology adoption in healthcare.

Contrasting the widespread use of TAM and UTAUT, IDT (Innovation Diffusion Theory) has seen relatively less application in healthcare, as noted by the absence of comprehensive reviews specifically on IDT. Studies like those by Wang and Lin [209] and Alhasan et al. [5] delve into specific IDT factors like relative advantage and trialability yet leave out the other factors associated with IDT. Unfortunately, there is no significant literature, to the best knowledge of the author, that comprehensively applies IDT to healthcare beyond the application of IDT to a specific use case [209, 5].

In addition to works that focus on a single classical framework, there is plenty of research that combines various classical models together in an attempt to construct a novel framework that better captures the specific healthcare context [172, 12, 62, 100]. For example, Tung et al. [195] combine trust, IDT, and TAM to analyse the adoption of electronic logistics information

systems in the medical industry. These models are related through factors such as relative advantage and perceived usefulness, or complexity and perceived ease of use [141]. They show the importance of compatibility, perceived usefulness, perceived ease of use and trust with regard to behavioral intention to use. They also show that trust and compatibility affect the perceived usefulness, which is commonly regarded as the driving factor for adoption.

Recently, the adoption of AI in healthcare has been a rising focus, with studies exploring how classical models apply to this new context [116, 103, 6, 86]. These studies consistently demonstrate the significance of classical model features, such as TAM's perceived usefulness, in AI adoption. However, they also reveal a tendency to focus on abstract notions like 'organizational' or 'psychological' factors [103], often at the expense of detailed discussions on these aspects. These studies tend to agree that 'psychological factors' play a role in the acceptance and adoption of AI. However, this leaves a lot of room for concretizing how to use that information to increase the chances of the successful adoption of AI. Synthesizing the aforementioned studies, there seem to be common subthemes prevailing throughout the various publications, usually hidden within overarching factors such as 'organizational' or 'psychological' factors. These subthemes include application safety [116, 103, 86], trust [116, 103, 6, 86], the threat of autonomy [116, 103], and diagnostic accuracy [116, 103, 6, 86]. However, limited discussion on these subthemes is provided.

With regards to the TOE framework, there is a large amount of work that focuses on the adoption of AI in the context of healthcare [218, 142, 112, 113]. This can, in part, be attributed to the focus of the TOE framework, which is more on an organizational level. With the adoption of AI, factors such as infrastructure, regulations, leadership support, and the importance of multiple stakeholders seem to incentivize researchers to take a broader perspective on the adoption of such technologies [142, 112, 113]. All works conclude the importance of all the overarching factors of the TOE framework. Kumar et al. [113] further specifies the importance of the factors, stating that technological factors are most important for the adoption of AI for managing the healthcare supply chain. The influence of government regulations is also featured often, but mostly through subsidies and other forms of stimulation [218, 142, 112]. Similarly, all four studies discuss the importance of top management support, with some including it as their primary factor [218, 142, 112, 113]. Interestingly, Yang et al. [218] find that the awareness of the value and benefits of AI healthcare technology is one of the primary factors influencing adoption, which is not mentioned by the other publications. However, it can be argued that 'relative advantage' or 'perceived benefits', included by Mukherjee et al. [142], Kong et al. [112] and Kumar et al. [113], has significant overlap.

2.4.7 Further Findings on Acceptance and Adoption in Healthcare

In addition to the previously mentioned research, there is an abundance of research that looks at barriers and opportunities related to the acceptance and adoption of AI in healthcare without grounding itself in the previously discussed classical theories. There are several barriers that are mentioned relatively often throughout this body of research, including trust, uncertainty about regulation, patient safety, technological maturity, patient privacy, and decision autonomy.

Trust is mentioned by several publications as a barrier, but in various contexts, such as trust in data handling, trust in patient safety, and trust in the clinical validity of the output [188, 73, 41, 81, 11]. Surprisingly, although there is an extensive literary foundation on the creation of trust, works on the acceptance and adoption of AI in healthcare do not seem to connect their findings regarding trust with this literature. Even Gille et al. [73], who specifically focus on the barrier of trust in AI for healthcare, do not link their analysis to established trust research. They highlight that the idea of 'trust in AI' in the context of healthcare often lacks clarity, that research often ambiguously defines the idea of trust and often does not address the complex nature of the matter sufficiently enough.

From the several publications that integrate trust into the acceptance and adoption of

technology in healthcare, the work of Chew and Achananuparp [41] stands out. Chew and Achananuparp [41] present the most extensive overview of other works where trust plays a role in AI adoption in healthcare, stemming from their systematic review, analyzing trust with respect to data privacy, patient safety, and the technology. For example, a primary driving factor for trust regarding patient safety was the validity of the clinical output for rare conditions or unexpected situations. Similarly, trust in data privacy primarily came through unknowingness or uncertainty regarding data handling. Therefore, a common recommendation is to either guarantee anonymity, increase transparency, or improve explainability [41, 11].

Linked to trust of the clinical validity and technology, but worth mentioning separately, is the numerous mentions of application and data quality and the appearance of data and application bias [88, 41, 188, 73, 81, 11, 151]. Although the promise of AI in healthcare is the improvement of health outcomes, it should not do so at the cost of perpetuating or worsening existing biases in the healthcare system. This could relate to cultural or demographic biases, but can can also be much more complex. One of the influential factors of algorithmic bias is the availability and quality of the data that healthcare algorithms are developed with [188, 81, 11, 151].

Lastly, a common occurrence in discussions surrounding the adoption of AI in healthcare regards the amount of accountability that can be allocated to the algorithm [41, 188, 73, 81, 88, 11]. If the clinician keeps complete final responsibility, then one can not rely on the decisions of the algorithm. If there is shared accountability, then the clinician needs to justify every deviating decision made. Lastly, giving full accountability to an algorithm is arguably a bad idea as its power is limited by the information that it is programmed to use for making a decision, whilst external factors might induce a more optimal decision.

2.4.8 On the Generalizability of Aforementioned Healthcare Research

Noteworthy is that, although this study is related to the context of healthcare, there is a gap between the context of this study and the many acceptance and adoption studies in healthcare. Specifically, SD is an enabler for the development of healthcare innovations and not a healthcare innovation itself. As such, it is questionable whether the findings from healthcare-related acceptance and adoption studies, which primarily focus on the acceptance and adoption of novel tools and applications, are generalizable to the context of this study. For example, fear of loss of autonomy is an often considered human factor in research on the acceptance and adoption of AI in healthcare [103, 116, 162]. However, the fear of loss of autonomy is associated with the product or service itself and the use thereof. The fear of loss of autonomy is significantly less directly related to the development process for the product or service. There are, however, other factors mentioned in previous research that have a more direct link to the development process, such as the safety of the application or government regulation on the adoption of SD, factors that have been repeatedly included in the aforementioned research. In all, it is difficult to assess what aspects of the research are directly generalizable to this research.

2.4.9 Acceptance and Adoption in Other Domains

Apart from looking at acceptance and adoption research that is applied to healthcare, it is useful to look at other related literature not directly linked to the context of healthcare. For example, there are a few works on the acceptance and adoption of AI in manufacturing or production companies. As healthcare technology production companies could fall under this context, it is worth looking at the results of Chatterjee et al. [38]. Chatterjee et al. [38] investigated AI adoption in manufacturing and production firms using an integrated TAM-TOE model, further cementing the correctness of TAM but also discussing the importance of employee competency and the need for appropriate knowledge. Finding employees with the right skill sets is also a barrier that is identified by Ulrich and Frank [196], who look at AI adoption (in general) for German SMEs.

Alsheibani et al. [10], Kelly et al. [102] and Gursoy et al. [79] all look at AI acceptance and adoption but do not specify any context. Kelly et al. [102] present a systematic review of AI acceptance, concluding the commonly assumed influence of perceived usefulness, performance expectancy, attitudes, trust, and effort expectancy on the behavioral intention for AI, but also across multiple industries. However, they do not create any overarching framework for AI acceptance. In contrast, Gursoy et al. [79] create a framework for AI Device Use Acceptance (AIDUA), suggesting a three-stage framework consisting of a primary appraisal stage, a secondary appraisal stage, and an outcome stage. Although they include performance expectancy and perceived effort expectancy in their framework, they conclude that these factors lead to a certain emotion, which then determines a willingness or objection to the use of AI devices. Interestingly, they see the willingness or objection to the use of AI devices not as strict opposites and hence include them as separate outcome factors.

Noteworthy is that, again, this research focuses on AI devices and services, whereas SD is an enabler for creating AI devices and services. Consequently, the research is not inherently generalizable towards SD acceptance or adoption. For example, Gursoy et al. [79] concludes in AIDUA that anthropomorphism influences both performance expectancy and perceived effort expectancy. Although people easily anthropomorphize devices they interact with, this relationship is less obvious for digital data.

2.4.10 Acceptance and Adoption of SD

Given that the field of research on SD is rather new, it is unsurprising that there is limited research on the acceptance and adoption of SD. In particular, there is no research that looks at the acceptance and adoption of SD in relation to any of the classical models. The primary purpose of most adoption-related publications in the context of SD is to summarize use cases and highlight potential barriers to its adoption. However, the only potential barriers currently acknowledged by literature are the difficulty of creating SD, validating the quality and bias of SD, and the uncertainty and difficulty surrounding the privacy-preserving characteristics of SD [7, 39, 71, 206, 114]. As such, recommendations from SD literature for the adoption of SD or recommendations for future research solely regard additional technical research or advice to regulators. There is a remarkable absence of any acceptance or adoption literature from an innovation management perspective, further underscoring the need for this study.

Chapter 3

Methodology

To understand how the relevant literature helps in answering the research questions that were presented in Chapter 1, it is essential to concretize the methodology. This chapter elaborates on exactly what approach was taken and why that approach was taken.

The chapter is split up into two Sections. The first section provides an overview of the methodology and subsequently argues why this methodology was chosen. Section 3.2 specifies the details of the methodology, including participant recruitment, data collection, analysis methods, and interview and survey content.

3.1 Understanding the Methodology

Before diving into the details of the various aspects of the methodology, it is important to understand the chosen approach and understand why it was the logical choice for answering the research questions of this study. To that end, a summary of the methodology is presented first, after which this section elaborates upon why this methodology was chosen.

3.1.1 Overview of the Methodology

This thesis attempts to answer two research questions, namely RQ1, which explores the Utility Perspective, and RQ2, which explores the Uptake Perspective. Although the analyses are presented in separate chapters, the approaches that were undertaken to conduct the analyses were heavily intertwined. This study answered both research questions through a combination of interviews and surveys. The following overview summarizes the methodology of this study:

1. Based on the literature presented in the previous Chapter, a conceptual model for the Utility Perspective was developed.
2. Expert opinions were consulted through semi-structured interviews. These interviews had 2 goals:
   (a) Explore the Utility Perspective on how SD can be utilized throughout different stages of the ML-based healthcare innovation process.
   (b) Explore the Uptake Perspective on the barriers, challenges, and opportunities of accepting and adopting SD for ML-based healthcare innovation.
3. The interviews were analyzed using directed content analysis for the Utility Perspective and thematic analysis for the Uptake Perspective.
4. Using the results of the analysis, the conceptual framework for the Utility Perspective was further refined, and a conceptual framework was developed for the Uptake Perspective.

5. A survey was created and conducted to refine and verify the developed frameworks.
6. The results of the survey were analyzed using quantitative methods to improve and finalize the two frameworks.

The remainder of this section explores why this methodology was chosen.

3.1.2 Purpose of the Methodology

The methodology outlined above seeks to answer the research questions outlined in Chapter 1. Because of the dearth of research on the acceptance, adoption, and utility of SD, the exploratory approach to answering the research questions by garnering stakeholder opinions through the use of interviews was logical. The semi-structured interview approach allowed for flexibility and personalization of the conversations, enabling all experts to bring up topics that were of importance to them. This left room for unexpected or unforeseen barriers and opportunities to surface.

Both the Utility Perspective and Uptake Perspective hold significant weight individually, yet their interrelation is undeniable. To realize the potential of SD in ML-based healthcare innovation, it is important to understand the nuances of both these perspectives. Consequently, the perspectives are first analyzed in a standalone manner, and the findings of the Utility Perspective and the Uptake Perspective are hence separately presented in Chapters 4 and 5, respectively. Subsequently, to capture the holistic essence and the interrelated factors, the prevalent and overlapping characteristics of both perspectives are presented. This integrated perspective, presented as a discussion in Chapter 6, elucidates the interconnectedness of these topics and offers guidance for effectively integrating SD into ML-based healthcare innovation.

3.1.3 Theoretical and Inductive Framework Development

For analyzing the Utility Perspective, the aim was to obtain a detailed account of where and how SD can benefit the healthcare innovation process. As there is existing literature that elaborates upon potential use cases of SD, the analysis can best be based on this body of literature. By basing the analysis within a certain scope of literature, the interviews can also be guided to stay within that scope. Hence, it made sense to construct a conceptual model upfront based on the literature analysis and let that further guide the expert interviews. The development of a conceptual framework before refinement through interviews also occurs in other studies [184, 167].

In contrast, the aim of the analysis of the Uptake Perspective was to obtain a rich understanding of the various stakeholder perspectives. This aim was in contrast to the aim of the Utility Perspective. Consequently, it was more appropriate to approach the analysis through an inductive lens and to construct the framework through a bottom-up process from the insights of the experts. Given the lack of existing literature with regard to this specific topic, an inductive approach to creating a framework is logical. It allows complete freedom for participants to introduce concepts and focuses that are unknown to the author.

3.1.4 How the Approach is a Variation on a Delphi Study

The study design takes inspiration from a Delphi study design with regards to framework development and can be seen as a variation of a Delphi study design [48]. In a traditional Delphi study design, a panel of experts is given multiple rounds of questionnaires. After each round, the responses are aggregated and shared with the group. This provides feedback to the experts, allowing them to refine their answers based on the collective input of the group. Over several rounds, the goal is to converge towards a consensus or collective understanding, for example, in the form of a framework.

There is a clear overlap between the traditional Delphi study design and the study design of this thesis, given the two-round iterative approach to developing frameworks. However, there are also two noteworthy differences. First, this thesis only featured two rounds, in contrast to the often more than two rounds of traditional Delphi studies. Second, the first round took the form of a semi-structured interview as opposed to a questionnaire. There are several reasons for these choices.

Semi-structured interviews were chosen as the method of choice for the first round mainly because there were too many unknown variables at the start of the process. There is little published research regarding the Utility Perspective and the Uptake Perspective. Consequently, questionnaires likely would not have provided enough depth to understand the various motivations, thoughts, or arguments of the experts. Semi-structured interviews provided the possibility to really dive into the individual perspectives of experts. Using semi-structured interviews at the start of this research was a logical approach and in line with the aims of this research. Starting a Delphi study with a qualitative first round has precedent in literature [111, 187].

In addition, this thesis only features one round of questionnaires to verify and further refine the frameworks, mostly because there is limited time and resources within the scope of this project, as this thesis was executed as a master thesis. Also, because research participants participated on a voluntary basis, asking for more time upfront would likely have resulted in decreased engagement or availability. Although additional rounds could have potentially yielded further improvements, it will be left for future research. There is precedent for two-round Delphi-based studies [177].

3.2 Specification of the Methodology

3.2.1 Sampling and Participant Recruitment

Answering the research questions of this study was done using extensive input from experts. The study recruited individuals who had relevant knowledge with respect to ML-based healthcare innovation, SD, or both. As experts in the field are scarce, and because this master thesis had no extensive budget for participant recruitment, a purposive sampling methodology was used to select an appropriate sample [65].

To effectively identify potential participants who could valuably contribute, a range of stakeholder groups were pinpointed. These groups were selected based on their likelihood of possessing pertinent expertise and information. This identification process was a collaborative effort between the author and their supervisors, with a focus on ensuring a diverse range of voices and maintaining a balanced number of participants across all stages of innovation development. The following groups were identified and recruited from (with the number of participants in brackets): Healthcare Data Scientists (3), Medical Professionals/Researchers (4), Healthcare IT Professionals (3), Regulatory and Compliance Experts (3), Medical Application/Test Developers (2), Ethicists & Patient Advocacy Groups (2), SD Experts (4), Innovation Program Managers (3) and Data Sharing Experts (2). A large effort was exerted to recruit a diverse sample that included industry experts, academic experts, and clinical experts. However, many of the recruited participants associated with an academic institution were also associated with a hospital or healthcare organization, and vice versa. In total, 24 interviews were conducted.

Two of the interviews were conducted with two experts present instead of the usual one-on-one format. This was done as the experts were expected to have highly overlapping opinions on the topic and thought that having both present could enhance the content of the interview. As such, a total of 26 people were interviewed throughout 24 interviews. All 26 interview participants were sent the questionnaire, of which 20 experts provided a response.

To recruit participants, the extended network of the main author and supervisors was used. Specifically, as the research was conducted in collaboration with Philips and e/MTIC, partici-

pants were recruited from Philips and e/MTIC as well as collaboration partners and contacts of Philips and e/MTIC. Different institutes, organizations, and companies were included to cover academic, clinical, and industry backgrounds.

Potential participants were invited through an email, which included an elaborate explanation of the study. The study explanation included a summary, an information letter, including their rights within the study, and a template consent form. The emails invited the participant to overthink their participation and reply with a response either for or against participation. A total of 34 people were contacted. If the participant wanted to participate, a follow-up email with the consent form and a possible date was sent to initiate the expert participation.

As mentioned, a total of 26 participants were recruited. This sample size is proper for two reasons. First, the sample size is deemed to be appropriate in literature for qualitative studies and is relatively common amongst other qualitative studies [80, 133]. Second, the time and budget constraints limit the sample size from growing any larger. A sample size of 25 usually leads to information saturation for many studies. However, due to the exploratory nature of this study, the expert recruitment from 9 different expert groups, and the employed purposive sampling technique, it cannot be guaranteed that information saturation was reached [33].

3.2.2 Data Collection and Analysis

As elaborated upon before, the study employed a variation of a Delphi study design consisting of two rounds, where the first round took the form of a semi-structured interview. This subsection elaborates upon the data collection method, the analysis method, and the content of the interviews and surveys.

3.2.2.1 Data Collection Method

The collection of data was done in accordance with the biomedical research standards held up by Philips through their Internal Committee Biomedical Experiments (ICBE), involving both a privacy impact assessment (PIA) and an ethical assessment. The ethical review board of the Eindhoven University of Technology also approved the study documents. The semi-structured interviews were, depending on the preference of the participant, conducted physically or through Microsoft Teams. If the interview was conducted physically, audio recordings were made. If the interview was conducted through Microsoft Teams, both audio and video were recorded. Following the interviews, the recordings were transcribed and de-identified from any personal information. Subsequently, the recordings were deleted. The surveys were distributed through email, after which participants had two weeks to fill them in. Reminders were sent after the first week and on the last day.

3.2.2.2 Interview Content and Analysis

The interviews covered both the Utility Perspective and Uptake Perspective. To start the interview, the participants were explained that their answers would be anonymous and that their answers did not need to represent the visions of the organizations they were employed by. Subsequently, the interviewees were first asked a series of participant background questions relating to their expertise, such as their experience with ML-based healthcare innovation and SD. Next, the interviews focused on the Uptake Perspective and covered questions on the benefits, limitations, challenges, and risks of SD in healthcare innovation. This provided an elaborate understanding of the knowledge that the interviewee had on the topic of SD. Following this, the interviews turned to the Utility Perspective, asking questions on both the processes that can benefit from SD, as well as the perspective of the interviewees on which stage of innovation can benefit most from SD. Subsequently, the interviews again turned to the Uptake Perspective to focus specifically on the integration and adoption of SD in healthcare innovation. Lastly,

the interviews were closed of with the recommendations of the interviewees on leading their own organizations through the incorporation of SD for healthcare innovation practices. The interview guide can be found in Appendix C.1.

From the aforementioned structure, it is obvious that the two perspectives are not integrated into the interview protocol separately but that questions regarding the Utility Perspective are in between the questions regarding the Uptake Perspective. The flow of the interview was expected to be better through this approach. In addition, the interviewee was not specifically informed that two perspectives were analyzed. Hence, separating the questions concerning the two perspectives was not necessary.

Analysis Methodology for Utility Perspective For the analysis of the Utility Perspective, the interview transcripts were analyzed using directed concept analysis, a methodology outlined by Hsieh and Shannon [91]. As there is an existing body of literature regarding use cases of SD, that is, to the belief of the author, incomplete, directed content analysis is a suitable choice. The goal of a directed approach to content analysis is "to validate or extend conceptually a theoretical framework or theory" [91]. As such, a conceptual model based on the literature was created before conducting the interviews. This conceptual model introduced several key concepts which were used as initial codes during the content analysis. The directed content analysis consisted of two rounds. In the first round, the key concepts were used to code the data, and uncoded text was sought for additional concepts not found in the original. After the first round, the codes were evaluated and renamed for consistency and quality. Then a second round of coding with the reworked codes was performed to ensure completeness, consistency and quality.

Analysis Methodology for Uptake Perspective For the analysis of the Uptake Perspective, the interview transcripts were analyzed using thematic analysis, a methodology outlined by Braun and Clarke [32]. The authors of the original methodology have stated multiple times that the specific type or approach for a thematic analysis needs to be specified upfront [34]. The goal of thematic analysis for the Uptake Perspective is to obtain a rich description of the perspectives of stakeholders. This can best be done through inductive analysis and latent themes to examine and identify the underlying ideas, assumptions, and conceptualizations of the various experts [34]. Further motivation for this approach is given in Appendix D.

Practical Aspects of Interview Analysis Practically, the interview recordings were transcribed by the main author. Following, the directed concept analysis and thematic analysis were done through the use of the NVivo software tool [22]. The NVivo Tool is provided through the Eindhoven University of Technology. Supervisors were consulted throughout the thematic analysis for feedback on the clarity, content, and consistency of the codes. Having a single researcher for transcription, coding, and thematic analysis is not necessarily a study limitation for thematic analysis, contrary to dominant research values, as the thematic analysis methods set forth by Braun and Clarke [34] emphasize a non-positivist qualitative approach. They firmly believe that researcher subjectivity is a resource for research, rather than a threat to be contained, rejecting the notion of 'researcher bias', and emphasizing researcher reflexivity. Every effort was made to adhere rigorously to the best practices outlined Braun and Clarke [34].

3.2.2.3 Survey Content and Analysis

The content of the survey depended on the results of the analyses of the conducted interviews. The survey consisted of two multiple-choice participant background questions, four questions with multiple-choice sub-questions regarding the Utility Perspective, three questions with multiple-choice sub-questions regarding the Uptake Perspective, and three additional open

questions where participants could add additional thoughts. The following presents the method of analysis for the survey. The types of questions contained in the survey were primarily independent 7-point Likert scale questions and two multiple-response questions with five options. The complete survey is provided in Appendix E.

Participant Background Questions To have more information about the context and knowledge of the participants of the survey, two questions were added to the survey. The first question was 'What stage of healthcare innovation do you associate yourself most with?', providing the options 'research', 'development', 'integration & deployment', and an open option 'other'. The 20 answers to the first question can best be divided into people associating themselves with 'research' (11) and 'not research' (9). To that end, it can be analyzed whether answers to the survey questions differ between these groups in a statistically significant manner.

The second question was 'How familiar are you with synthetic data?', providing the options (A) 'I have heard that synthetic data can be used for healthcare innovation, but I have not looked into it in detail', (B) 'I am quite familiar with synthetic data for healthcare innovation', and (C) 'I have actively researched or developed synthetic data for healthcare innovation myself'. The 20 responses to the second question can best be divided into the three provided categories, as none of the participants used the 'other' option. The responses indicated 8 participants with low familiarity (option A), 6 participants with medium familiarity (option B), and another 6 participants with high familiarity (option C).

Analysis of Likert Scale Questions The analysis of answers to questions employing Likert scales is a much-debated topic. There is a large debate about whether parametric tests (such as independent 2-sample T-tests) can be used to analyze Likert scale data. However, this debate was 'put to bed' by Dr. Geoff Norman, one of the world's leaders in medical education research methodology. Norman [149] concluded that yes, "parametric statistics can be used with Likert data, with small sample sizes, with unequal variances, and with non-normal distributions."

To avoid interpreting the results through means and standard deviations, the first visualization that is provided is that of a vertically stacked bar chart (i.e., Figure 4.5). This provides an intuitive notion of to what extent the experts agreed or disagreed with the statement.

Subsequently, the differences between groups for the first participant background question (2 groups), is analysed using independent 2-sample T-tests and Mann–Whitney U tests [106, 136]. Using such tests is in accordance with the above-mentioned literature and provides a statistical significance and direction of the tested relationship. Both tests are employed to have more robust findings, and have a mix of parametric and non-parametric tests.

Lastly, the differences between groups for the second participant background question (3 groups), is analysed using the Kruskal-Wallis methodology, followed by a subsequent post-hoc Dunn's test for statistically significant differences between groups of the Kruskal-Wallis test [200, 57]. The Kruskal-Wallis test is a parametric test, particularly suitable for this study, as it is usable for comparing groups when sample sizes are small and not necessarily normally distributed.

Analysis of Multiple Response Questions Analysis of multiple-response questions with nominal data is different from the analysis of multiple-choice questions with ordinal data. Tests that assume answer independence (such as T-tests) are not suitable for this analysis, as participants can enter multiple answers, which depend on one another.

First, before doing any statistical testing, it is interesting to assess the frequency of provided answers and the frequency of answer combinations. These can be visualized using histograms and heat maps, and presented separately through the use of tables.

Next, it is relevant to assess between-group differences with regard to the participant background questions. For nominal data, this can be done using χ^2 tests [77, 104], and Fisher's

Exact tests when sample sizes are small. Although the survey had 20 responses, these multiple-response questions had 48 and 44 responses, respectively. As such, depending on the number of groups that the analysis is done over, the χ^2 test is either suitable or not. Consequently, to improve robustness, both tests are performed for both questions. Note that Fisher's Exact tests can only be performed on a 2x2 contingency table. Hence, a question with five answers, divided over two groups, is split up into $\binom{2}{5}$ = 10 different contingency tables. For the second participant background question, this results in a total of $\binom{2}{3} * \binom{2}{5}$ = 30 different contingency tables.

3.2.2.4 Formatting of Results

In the remaining chapters of this thesis, the results, discussion, and conclusion of this thesis are stated. Excerpts and quotes are presented to exemplify the opinions of the experts. There are two noteworthy remarks about how these quotes and excerpts are presented. First, some of the interviews were held in Dutch. When quotes or excerpts from Dutch interviews are presented, the quotes and excerpts are translated into English to the best capacity of the main author. Second, all quotes and excerpts follow the same formatting, being presented in *"italic and between double quotation marks"*. At the end of the quote or excerpt, the interview identification number is stated in a subscript.

Additionally, the remainder of this research incorporates a substantial number of quotes from interviews to support the presented arguments. These quotes are extracted from their complete context for illustrative purposes. In choosing these excerpts, careful consideration was given to ensure that they reinforce the intended messages while faithfully reflecting the experts' original meanings. Nonetheless, it is important to acknowledge that these interpretations, made after the interviews, carry the inherent risk of misrepresenting the experts' intended statements.

Chapter 4

The Utility Perspective: Utilizing Synthetic data for Innovation Processes

This section will focus solely on the Utility Perspective. The primary goal of analyzing this perspective is to ascertain when and how SD can benefit, improve or accelerate ML-based healthcare innovation processes and additionally assess the limitations of the technology. It does so by answering the following research questions, as first introduced in Chapter 1:

RQ1 How can SD be utilized throughout different stages of ML-based healthcare innovation?

- RQ1a How does healthcare innovation, specifically for ML solutions, currently operate?
- RQ1b Which aspects of ML-based healthcare innovation involve the use of patient data?
- RQ1c Which parts of these processes have the potential to be enhanced or expedited using SD?

To answer these research questions, this Chapter continues as follows. First, Section 4.1 builds on the literature presented in Chapter 2 to develop a conceptual model, elucidating processes relying on data in ML-based healthcare innovation and classifying which of these processes can potentially benefit from SD. This framework is subsequently refined through expert input via semi-structured interviews and the follow-up survey in Sections 4.2 and 4.3.

4.1 Towards a Conceptual ML-Healthcare Innovation Development Framework

Healthcare innovation is a complex process, but frameworks exist that can provide a structure of how the process usually operates, as shown in Chapter 2. The aim of this section is to identify a framework that provides general context about what the ML-based healthcare innovation process looks like to act as an answer for RQ1a. Both the previously discussed process models and TRL frameworks can offer a helpful structure (see Sections 2.3.1.1 and 2.3.1.2).

One of the goals of this research is to identify sub-processes within the overarching innovation development process that can benefit from SD. While a significant amount of effort has been spent on developing process models that understand recurrence and feedback loops, these features are not necessarily essential for understanding which sub-processes are part of the larger projects. Instead, it is more important to identify the components of the innovation process because discerning where SD can assist in these processes is integral to answering the research question. The frequency of execution of such a process (i.e. feedback loops) or the subtle considerations for managers deciding when to repeat a process is less urgent.

The details provided in TRL frameworks provide more nuance on the different aspects of the innovation development processes in comparison to other process model frameworks. For example, in the original TRL framework publication, Mankins et al. [132] provides examples with TRL 1 such as "... studies of basic properties of materials (e.g., tensile strength as a function of temperature for a new fiber)." Consequently, the TRL frameworks seem to be better suited for the creation of the conceptual model. Unfortunately, no single TRL framework is suited well for ML-based Healthcare innovation. There are, however, separate frameworks applying TRL to ML innovation [118, 119] and to healthcare-related fields [46, 68, 78, 94, 101, 182, 189].

As previously discussed, the MLTRL framework of Lavin et al. [119] is a high-quality TRL framework for ML innovations and can help tremendously in understanding the ML aspects of innovation for this study. Consequently, the MLTRL framework is considered for the remainder of this research. However, the MLTRL framework, depicted in Table 4.1, only covers the highly technical aspects and not any of the clinical or regulatory aspects of innovating in healthcare.

TRL	Phase	Description
0	Research	First Principles A stage for greenfield research.
1	Research	Goal-oriented Research Moving from basic principles to practical use.
2	Research	Proof of Principle (PoP) Development Active R&D is initiated.
3	Development	Systems Development Sound software engineering.
4	Development	Proof of Concept (PoC) Development Demonstration in a real scenario.
5	Development	Machine Learning "Capability" The R&D to product transition.
6	Integration	Application Development Robustification of ML modules, specifically towards one or more use cases.
7	Integration	Integrations ML infrastructure, product platform, data pipelines, security protocols.
8	Integration	Mission-ready The end of system development.
9	Deployment	Deployment Monitoring the current version, improving the next.

Table 4.1: The Machine-Learning TRL (MLTRL) framework [119].

Consequently, it is important to look at TRL frameworks that focus specifically on healthcare, of which there are quite a few. It is interesting to line these processes up and compare their similarities and differences. Seven relevant TRL adaptations related to healthcare were identified through a literature search using SCOPUS and Google Scholar and snowballed on the obtained results. To start the creation of one overarching healthcare TRL, seven healthcare-related TRL frameworks were aligned [46, 68, 78, 94, 101, 182, 189]. Following their alignment, overlapping characteristics were highlighted and used for the creation of a new framework. Often, reoccurring characteristics and highly overlapping characteristics were included. This led to the TRL framework for healthcare innovation, henceforth referred to as the HTRL framework, depicted in Table 4.2. Details on the creation can be found in Appendix I.

The HTRL framework is not disjoint from the MLTRL framework. Both frameworks go through research, development, integration, and deployment phases. However, the MLTRL framework puts focus on ML-specific development practices, whereas the HTRL framework accentuates the clinical side.

TRL	Phase	Description
1	Research	Concept ideation Identifying problem & possible technology.
2	Research	Concept formulated Publications on analytic studies. Supporting analysis providing scientific information and data to develop research proposals.
3	Research	Proof-of-Concept (PoC) PoC in laboratory established for intended use.
4	Development	Prototyping PoC and safety demonstrated
5	Development	Scaling Scale-up, integration, and limited robustness. Layout regulatory procedures. Ready for clinical trials.
6	Integration	Development for real scenario Clinical trial conducted with small sample.
7	Integration	Good Manufacturing Practice (GMP) Ready Clinical safety and effectiveness trials in operational environment. Determination of adverse events and risks associated with the solution. Final design validated.
8	Integration	Mission Ready System validated, and approved by regulatory agencies.
9	Deployment	Deployment Broad implementation, commercialization, post-market studies.

Table 4.2: Aggregated healthcare TRL (HTRL) framework. This framework is an aggregation of seven different healthcare-related TRL frameworks.

Noteworthy is that although both frameworks feature levels 1-9, the levels do not necessarily line up. Some levels might take longer or shorter to graduate from in one framework compared to the other. For example, obtaining regulatory approval (HTRL 7) might take much longer than integrating an ML infrastructure and solution. It is, hence, also difficult to merge these frameworks. However, these frameworks might be used in a similar vein to how Sauser et al. [176] proposes to assess system readiness with an integration readiness framework and a technology readiness framework. Specifically, Sauser et al. [176] combine two frameworks and assess the readiness of a technology based on its performance in both. For example, a team can be 'done integrating' a solution if their solution is classified as MLTRL 8 and HTRL 7.

4.1.1 Towards a Taxonomy for Utilizing Synthetic Data

In Chapter 2, a systematic literature review of SD use cases was presented, which was complemented by a narrative review of ML and systems engineering research. Here, this literature is aggregated into a taxonomy containing processes that can occur during ML-based healthcare innovation and classifies which of these processes can potentially benefit the integration of SD.

To effectively leverage SD in healthcare innovation, it is important to understand both the potentially beneficial and likely non-advantageous applications. Mapping processes that can and cannot be optimized using SD will give a balanced overview, thus enhancing the value of the research. Understanding for which of the data-related processes SD can be beneficial provides information on how SD can be useful for healthcare innovation. Creating a detailed overview of processes and to what extent each of them can benefit from incorporating SD, provides important context that is essential for understanding the challenges and opportunities that

various stakeholders might identify. Consequently, the overview that will be created includes three different levels of SD utility, namely

1. processes that might benefit from the use of SD,
2. processes that might benefit from the use of SD, but real data is needed to complete the process successfully,
3. and processes that likely cannot benefit from the use of SD.

4.1.2 Synthesizing a Conceptual Model

Creating a *hierarchical* taxonomy provides a logical overview of SD use cases, despite a lack of precedent [74, 93, 135]. Instead of listing use cases such as 'development', at the same level as 'unit test development', they can be related in a hierarchical structure.

To synthesize a taxonomy containing data-driven processes in the innovation development process for healthcare ML-based solutions, the systematic literature review presented in Chapter 2 was used, which elucidated use cases of SD for innovation development. Additionally, the use cases from the complementary narrative review on data-driven processes in ML research and development were also used.

To create the taxonomy, all previously published use cases of SD, as well as data-driven processes from research regarding software development and healthcare, were listed. Through a bottom-up process of categorization, several mid-level and three high-level categories were constructed. The high-level categories are 'research', 'development' and 'integration'. Lower-level processes can be part of more than one high-level category, such as processes that occur in both research and development.

After having constructed the hierarchy, each item was classified as pertaining to one of the three above-defined processes ((1) might benefit from SD, (2) might benefit from SD but needs real data, and (3) likely cannot benefit from SD). The classification was based on the literature discussing the use cases and limitations of SD and the literature on ML-based application development.

It should be noted that the content of the taxonomy, and hence the use cases of SD, are focused on processes in line with the research questions. Not including topic-based use cases is in contrast to the use cases proposed by some papers on SD, such as Gonzales et al. [74] proposing 'epidemiology/public health research or McDuff et al. [135] proposing 'cardiology'. Instead, categories such as 'hypothesis testing' are more appropriate [74].

Furthermore, although publications have identified a plethora of use cases, there was no guarantee that the aggregate is necessarily exhaustive. None of the previous publications claims to be exhaustive, and claiming any set of use cases to be exhaustive would be naive. In addition, the taxonomy for this research includes more than just the use cases for SD, as it also includes processes in which SD cannot be used. The taxonomy includes any process in the innovation development for ML-based healthcare applications that specifically uses data. Hence, effort was exerted to ensure an overview that is as complete as possible. However, as there were possibly unknown processes that might be identified later, the study allowed for additional identification of use cases throughout further analysis.

The conceptual taxonomy is visualized in Figure 4.1. Again, the classification of each item in the taxonomy is based on the literature discussing the use cases and limitations of SD, and the literature on ML-based application development. A list of explanations for each of the items and an explanation for their classification is provided in Appendix H.

4.1.2.1 Identifying Data Sub-Processes within the TRL Frameworks

Given the MLTRL and HTRL frameworks from Tables 4.1 and 4.2, and the SD taxonomy from Figure 4.1, all individual components for the final conceptual model are defined. The

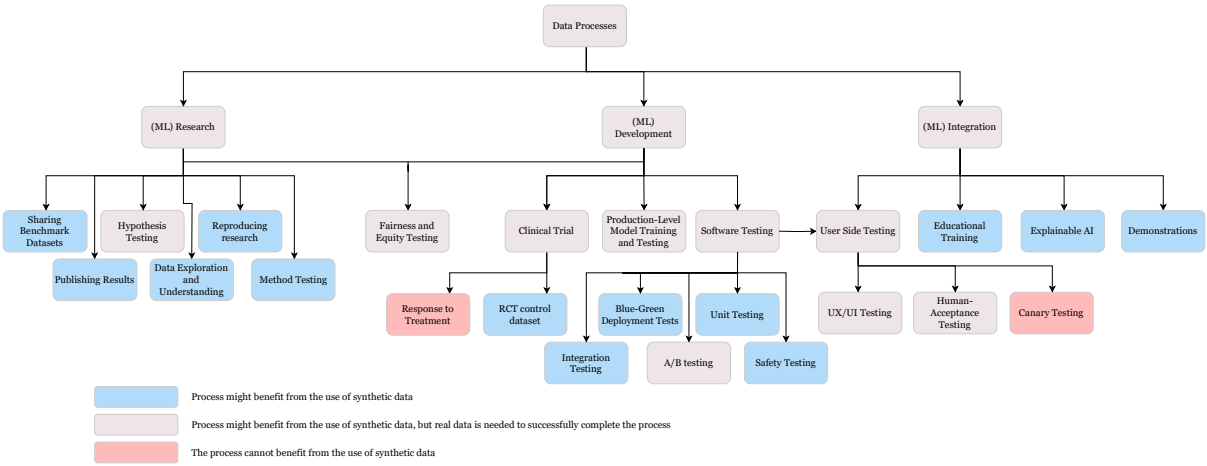


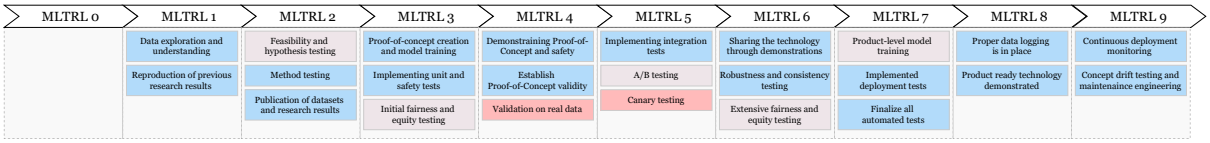
Figure 4.1: The conceptual taxonomy of data processes that are part of ML-based healthcare innovation process. The colors represent their potential to benefit from SD. Low-level items are colored based on the interpretation of the context in which they were mentioned in literature. Higher-level items are colored based on the lower-level items. Appendix Figure J.1 provides a full-page landscape version of the figure.

final conceptual model, visualized in Figure 4.2, assigns different processes in various stages of the MLTRL and HTRL frameworks to show where in the innovation process the sub-process is likely to take place. The color of each item in the conceptual model corresponds to its SD classification, which was also present in the process taxonomy in Figure 4.1.

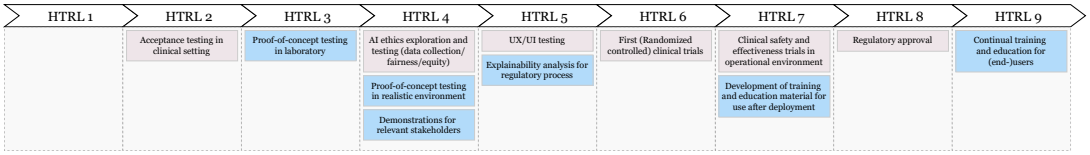
Not all processes within the conceptual model are directly copied from the taxonomy. For one, some processes have been nuanced, such as 'initial fairness and equity testing', as the process is likely to occur throughout the innovation process. Similarly, some processes are slightly adapted to provide more information. For example, MLTRL has the item assigned that reads 'Demonstrating Proof-of-Concept and Safety.' This is inspired by both the MLTRL stage 4, combined with the taxonomy item of safety testing because safety testing is an important feature of MLTRL stage 4 [119]. Lastly, some items are simply added to provide context for the MLTRL and HTRL stages, as they cannot be directly linked to taxonomy items but do include the use of patient data. For example, HTRL stage 3 states 'proof-of-concept testing in the laboratory.' This process will need to use patient data, but cannot easily be subdivided into various taxonomy items.

4.2 Results of Interviews

As the aim of the Utility Perspective is to understand how SD can be utilized throughout different stages of the ML-based healthcare innovation process, questions were asked during the interview in accordance with this aim and the constructed conceptual frameworks. Elaboration on the exact questions can be found in Appendix C. After conducting the interviews, the interviews were coded with the directed content analysis approach outlined in Chapter 3. Based on the discussions during the interviews, the mid-level items from the conceptual taxonomy were used as initial codes and later refined after the first round of analysis. The final codes and the number of interviews in which they were mentioned are provided in Table 4.3. Note that a certain process being discussed during an interview does not provide any indication of whether that has a positive or negative connotation. The rest of this section provides an indication for each of the mentioned processes, going down the table sequentially, and indicating for each the various opinions of the experts.



(a) The MLTRL-SD conceptual model



(b) The HTRL-SD conceptual model

Figure 4.2: The conceptual TRL model presents ways in which SD can benefit the healthcare innovation development process throughout different stages of the (a) MLTRL framework and (b) HTRL framework. The arrows at the top indicate different stages of the TRL frameworks. The boxes, aligned with these different stages, represent different processes in which patient data is used. The color-coding follows that from the conceptual taxonomy model in Figure 4.1. For readers using a printed version, Appendix Figure J.2 provides full-page landscape versions.

4.2.1 Elaboration of Codes and their Content

Clinical Validation Although many of the experts indicated to be positive about the possibilities of SD, there was a consensus amongst twenty-three experts that clinical validation requires real patient data: *"I think no matter what you use to train your algorithms, you have to do a rigorous validation on real world data before you should put it out (#13)."* According to the experts, SD cannot substitute real data for clinical validation. The experts ground their opinion through differing arguments. Fourteen experts argue that without this validation, there is simply no guarantee of quality and safety: *"When you get to the end, then you really want to know, what is the performance of my algorithm [...] Yeah, you may want to do that real validation on real data (#1)."* Three interviewees specify that they think that, without validating on real patient data, that the regulatory bodies would not approve the application: *"For your market release, [...] there you obviously must prove the quality on real clinical data for the foreseeable future. Because an FDA is otherwise not going to approve of it I think (#3)."* One expert referred specifically to the idea that, without validation on real patient data, you cannot convince hospitals that your innovation works: *"Well, when you are going to deploy, you want to deliver proof to the hospital [...] Well, then you are not going to convince anyone by showing them a bit of [...] modeled synthetic data (#11)."* Seven experts did not provide any concrete substantiation for their opinion that SD cannot be used for clinical validation.

Developing Algorithms The next entry in the table, 'Developing Algorithms', has significant overlap with many of the other entries within the table. As could be expected, experts did not answer all questions with equal amounts of focus or specificity. Although there are potential use cases for SD for ML-based healthcare innovation that are not contained within 'Developing Algorithms', the core idea of SD for ML-based healthcare innovation is to substitute or add to real patient data throughout the algorithm development process: *"We see that the value of synthetic data in healthcare primarily resides in for example the development of models (#15)."*

Important to note is that some experts are not in favor of using SD for the development of algorithms. In two of the interviews, the experts objected or were very doubtful of the value of SD for algorithm development. In both cases, the provided argument was the unknown factor of bias: *"Participant: I have my reservations about using synthetic data for algorithm development or the like. Interviewer: Okay, what do you see as the risks in that? Participant: Bias (#20)."* 'Understanding bias' would change their opinion on the usefulness of SD: *"If they tell me that*

Code Name	#Interviews
SD for Clinical Validation	23
SD for Developing Algorithms	16
SD for Data and Method Exploration	13
- Hypothesis testing	3
SD for ML Model Testing	12
- SD for ML Algorithm Robustness Testing	7
SD for ML Model Training	12
SD for Demonstrations	7
SD for Benchmarking and Reproducibility	6
SD for Software Testing (i.e. User Testing, EPD Testing)	5
SD for Regulatory Certification	5
SD for Post-Market Surveillance	2
SD for Explainability	1

Table 4.3: An overview of codes from the interviews and the number of associated interviews. Note that mentions of the code can be in both a positive and negative connotation.

they fully understand all the biases in the data, and they put that in their synthetic data, then [...] I'm all for it. But they have to prove that (#18)."

Data and Method Exploration In thirteen interviews, participants discussed the potential benefit of using SD for 'Data and Method Exploration': *"If [synthetic data] works well, then you can do exploration on that (#19)."* With 'Exploration', most experts refer to some initial process, usually contained within research, where one is attempting to explore the possibilities of what they can do with the data: *"If you can make a dataset in which those relationships are the same, then you can definitely do the first exploratory steps (#19),"* Three experts specifically mention the possibility of testing hypotheses using SD: *"If you have a synthetic environment available, where you can in any case do some first proof of concept [...] Then you do have quick results. [...] You can validate initial hypotheses much quicker. (#15)."*

ML Model Testing & Training The potential benefit of ML model testing was discussed by twelve participants: *"You can use synthetic data for all kinds of testing (#13)."* As also outlined in the conceptual model, model testing comes in a variety of forms, such as quality testing, unit testing, safety testing, and robustness testing. Most experts do not specifically state for which kind of tests SD is particularly beneficial. However, there were seven interviewees who specifically mentioned the usefulness of SD for robustness testing: *"You can say, 'please insert a [specific description of a data point]', and then you can see if the algorithm that you have developed can cope with this kind of [...] data (#4)."*

Twelve interviewees mentioned the potential for SD to benefit ML model training: *"So I think that [synthetic data] is very well usable in the context of the training of models (#6)."* Similarly: *"I think that it is realistic that, ultimately, the actual training of algorithms and in large parts the testing of algorithms can be done on synthetic data (#8)."* One expert specified that they saw value for SD in pre-training models[1] using SD, as they had done so with good results in the past: *"The ambition is not necessarily to make it so realistic that you can't mix it up with real data anymore, but it is more about like pre-training algorithms on such data. And then you have a much better starting point for developing algorithms on such data (#4)."*

[1]Pre-training in this context refers to first using the SD to train the algorithm (neural network) on the SD, to later train the algorithm further using real data. The pre-training helps the algorithm by giving it a 'head start'.

Use Cases Mentioned less than 10 Times Seven interviewees stated the usefulness of SD for demonstrations: *"Upon the purchase of a new algorithm [...] that you can say [...] to a party, show it on this [synthetic] data (#22)."* Similarly, four participants saw the potential for using SD for educational purposes: *"I think it would be very sensible for education. You can show variations that you cannot directly measure but which you can generate that could exist in practice. But not because you have, by chance, a measurement of it (#9)."*

Six experts saw a benefit for SD for benchmarking or reproducibility, for example, for scientific publications: *"When people want to show [their results], and [...] for that need to publicize what they tested on, that they can show that using synthetic datasets (#17)."* This allows other scientists working on the same topic to compare their performance to previously published work.

Software testing was mentioned various times during the interviews, such as load testing or unit testing: *"To test software systems and test a database, when you fill it up with patient data, does it not mix things up, and is it always right (#1)."* Two interviewees discussed the usefulness of SD for user testing: *"For usability testing, how are people handling it, and do they understand the feedback of the system (#1)?"* Interestingly, three interviewees mentioned the usefulness for specifically testing Electronic Patient Dossier (EPD) systems in hospitals. Although the EPD system is not ML-related and thus falls outside the scope of this study, it was mentioned as a use case for hospitals. Often, hospitals test their EPD systems with real patient data. For hospitals, the ability to test their (EPD) systems on synthetic data might present a motivator to invest in SD generation capabilities: *"I already have long expressed the need to have [synthetic data] in certain places. [...] Particularly for medical systems. Look, if we test medical systems [...], if we migrate the EPD [...], then there is a desire to test it on production data [...], but the privacy regulation even states that [...] it's not allowed to use production data (#20)."*

Five interviewees stated that SD might help in obtaining regulatory certification (i.e. FDA clearance): *"There are some developments where, also for FDA clearance, that synthetic data is used (#1)."* Note that this specifically refers to using SD to get regulatory certification, which is different from getting regulatory certification of an application that uses SD. However, the five interviewees who mention it are quite cautious about their statements. When asked whether a clinical trial using SD would be less trustworthy than a normal clinical trial, interviewee #2 stated: *"I would say not inherently [...] it depends on the quality of the data. (#2)."*

Of the remaining codes, only their essence will shortly be explained. Post-market surveillance regards the assessment of model performance and data drift after deployment. Additionally, SD for explainability refers to the potential for SD to help explain how an ML-based healthcare innovation works. Both use cases were scarcely discussed during the interviews.

4.3 Framework Refinement Using the Interview Results

Given these insights from the interviews, it is now key to refine the previously created frameworks from Section 4.1. A first step in this direction is to create an easily interpretable visualization of the interview results. To that end, a word cloud visualization was created that is visualized in Figure 4.3.

Within the word cloud, the color assignments are only a rough indication and strip away some of the nuances from the interview data. Notably, two of the sixteen experts discussing the potential for SD to enhance algorithm development stated that they did not see much potential there as long as bias remained an unknown factor. As such, 'Developing Algorithms' was placed to both overlap the green and the red, but be more skewed towards the green area. However, because the two experts already stated not to believe that SD would provide value for algorithm development, they did not also state their belief regarding processes that are heavily associated with that. They likely also do not believe in the added value of SD for 'ML Model Training', but they did not specifically mention that. Consequently, one should take care interpreting the word cloud visualization, as it might be overly positive.

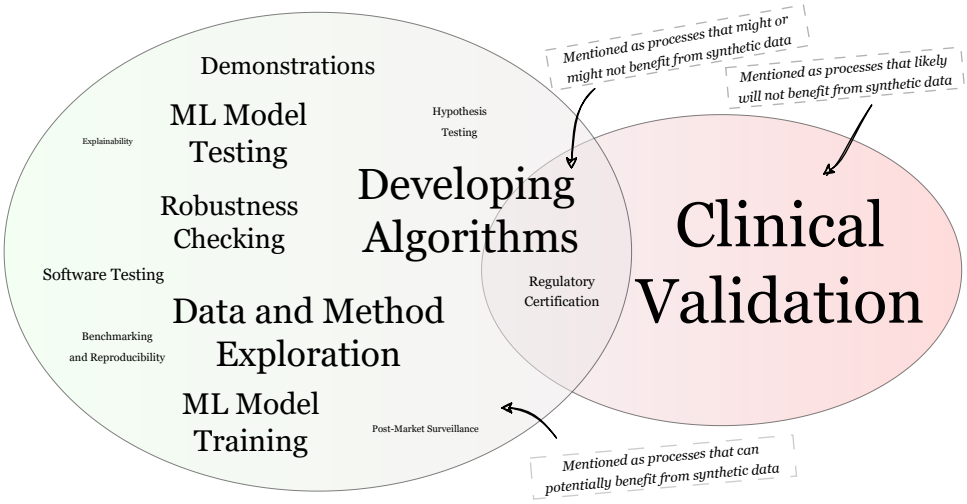


Figure 4.3: A word cloud visualization of the use cases discussed during the interviews. The colors provide an indication of the opinion of the experts. The size of the use cases provides an indication of the number of participants that stated the use case.

4.3.1 Reflecting on the Conceptual Framework

Given the interview results, it can be checked which items from the taxonomy have been mentioned in some way or another. Logically, not all specific low-level use cases were mentioned in the interviews. However, most of the high-level potential use cases were mentioned. An overview of the overlap between the identified use cases from the conceptual taxonomy (see Figure 4.1) and the use cases that were discussed during the interviews are given in Figure 4.4.

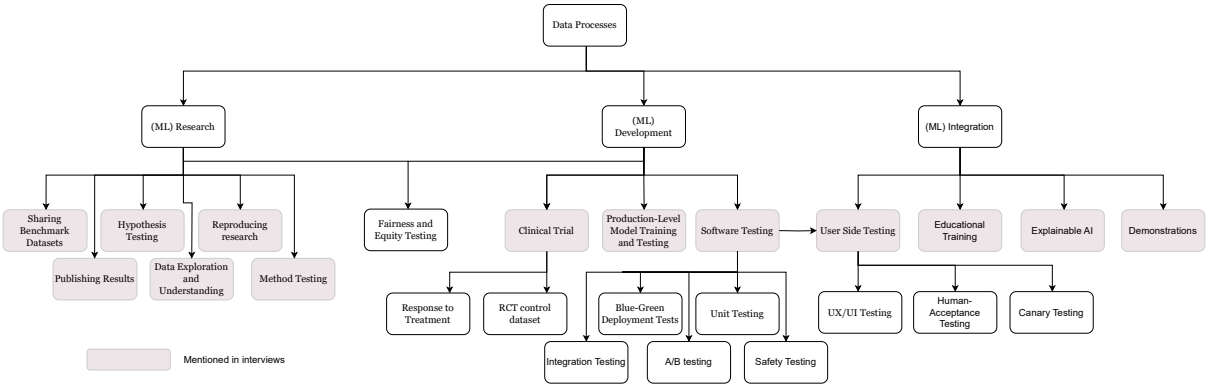


Figure 4.4: An overview of the use cases mentioned in both the taxonomy (literature) and the interviews.

The visualization of the overlap shows that nearly all of the mid-level and high-level use cases identified in the literature were also identified by the experts. Aside from the overlap, the differences are also interesting. Disregarding the specific low-level use cases, Figure 4.4 shows fairness and equity testing not to have been mentioned. As fairness and equity testing is an upcoming topic in data science and ML, especially in healthcare, it is notable that none of the experts mentioned it during the interviews.

Second, the testing of EPD systems was not specifically identified in the literature, although it was mentioned by several interviewees as a good application for SD. Although 'unit testing' or 'integration testing' has significant overlap with this use case, EPD systems were not identified specifically. This is likely due to the fact that this application falls outside the scope of this study. However, it only falls outside of the scope of this study because of an assumption made

in Chapter 1 that the primary application of SD is for ML-based innovation. The author was unaware of this use case before conducting the interviews.

4.3.2 Refining the TRL Framework

Given the interview results, the conceptual TRL framework can also be given further direction. However, there is a gap in specificity regarding the use cases mentioned by the research participants and the use cases in the conceptual TRL framework. The conceptual TRL framework and the taxonomy contain significantly more detailed use cases. To close this gap, it is important to look at the broader interest of this study.

Rethinking the Essence of the TRL Framework The primary aim of examining the Utility Perspective is to explore its role in enhancing various phases of ML-based healthcare innovation. On the surface, this exploration involves a single goal: identifying processes where SD can or cannot be beneficial. However, there is an underlying goal of practicality, namely, to ensure the overview of these processes is sufficiently detailed yet accessible. In the end, this research should facilitate the decision-making process surrounding the acceptance, adoption, and utility of SD, whether that is positive or negative. An overly intricate overview might hinder productive discussions on SD's potential benefits or drawbacks. Thus, a balanced approach is necessary, one that provides a comprehensive yet understandable view of SD applications, enabling stakeholders to evaluate its relevance and potential in their specific contexts effectively.

To that end, the conceptual model might overload a casual reader with information. For a facilitating overview, one likely would not want to see two different frameworks with nine different TRL levels, each of which has an individual definition and description. For those seeking that detailed information, the conceptual model serves as a reference point. However, a streamlined version of the TRL model, categorizing stages into 'research', 'development', and 'integration & deployment', simplifies the framework while retaining its core value. This simplification makes it easier for individuals to grasp the essence of SD's utility in healthcare innovation without getting lost in overly technical details. Within this context, each of these categories can refer to a set of TRL levels according to the MLTRL framework.

Rethinking the Content of the TRL Framework The refined TRL framework should contain use cases that reflect the above-mentioned goal. Consequently, the use cases should encapsulate the essence of the various use cases without delving into excessive detail. Considering that the experts, who are likely the primary stakeholder in the acceptance and adoption decision making process, have suggested use cases based on their experience, their contributions are a good foundation to work with. However, it is important to note that these use cases, as proposed by different experts, vary in their level of specificity.

For example, 'Developing Algorithms' covers the entire scope of research and development in healthcare innovation, likely also encompassing use cases such as 'ML Model Training' and 'ML Model Testing'. Similarly, 'Data and Method Exploration,' 'ML Model Training,' and 'ML Model Testing' have varying definitions across the healthcare innovation process. However, when used within the context of research, they have significant overlap, as one would likely not do 'Method Exploration' without training and testing ML models. Facilitating acceptance and adoption discussions is likely best done when the use cases within the framework are disjoint. Given these considerations, a selection of use cases pertaining to the three stages of healthcare innovation was made. These use cases and accompanying definitions can be found in Table 4.4.

The use cases provided in Table 4.4 reflect a trade-off between the specificity of the original conceptual model and the results from the interviews. Noteworthy is that processes that were only mentioned once or twice are included in the table. The reason for this is that there is a research consensus that interviewees are not very consistent at answering vague questions, and

Research	
Data Exploration	When real data is unavailable, SD can potentially be used to get a feel for the real data and analyze the characteristics of the dataset.
Comparison of ML Models or Methods	When the real data is unavailable, SD can potentially enable the analysis of which ML model or methodology leads to a better performance for a downstream task (i.e. a prediction/classification/other model).
Hypothesis Testing	When real data is unavailable, SD can potentially be used to prove or disprove hypotheses.
Publishing More Replicable Research	Usually, healthcare data cannot be shared when publishing research. When SD is used for research, it can potentially be published alongside the research publication, thereby enabling better reproducibility.
Development	
ML Model Training	SD can potentially benefit the training of ML models by allowing for easier access to (additional) useful data, leading to ML models that are more robust or have equal or better performance than models trained just on the accessible real data.
Software Testing	SD can potentially benefit software testing (i.e. unit testing, integration testing, load testing, etc.) by allowing for easier access to large amounts of (useful) test data.
Robustness Testing	SD can potentially benefit robustness testing by allowing for easier access to diverse data.
Regulatory Certification Processes	SD can potentially benefit regulatory processes, for example, by allowing to more easily show explainability, robustness, and safety and/or having access to more data.
User Testing	SD can potentially create value for usability/user testing by allowing for more representative test scenarios for users to interact with.
Interation & Deployment	
Demonstrations	Demonstrations using SD can potentially be of more value as they can use representative data without potentially compromising the privacy of patients.
Employee Training	SD can potentially aid in training employees by allowing the use of representative data whilst not compromising the privacy of patients.
Software and Model Testing	The testing of new software within hospitals (including those containing ML-based applications) is often done on patient data. SD can potentially help minimize unnecessary use of patient data.
Post Market Surveillance	SD can potentially benefit post-market surveillance, for example, by enabling easier performance monitoring, safety analysis, or model/data drift.

Table 4.4: An overview of relevant use cases per stage of the healthcare innovation process.

'what processes in research do you think SD might benefit' might be perceived as vague [173]. Consequently, use cases discussed by only one or two interviewees might be as valuable as use cases discussed by more interviewees. This needs to be verified.

4.4 Refining the TRL Framework Using the Survey Data

In the follow-up survey, participants were asked the following questions: *To what extent do you agree or disagree that employing synthetic data can enhance the following processes?*. This question was asked for each of the use cases mentioned in Table 4.4, and they were given a 7-point Likert scale ranging from 'Strongly Disagree' to 'Strongly Agree' to indicate their answer. They were given the names of the use cases and the definition as stated in Table 4.4. However, from each of the definitions, the word 'potentially' was deleted. Consequently, the definitions presented a clear statement that 'SD can help this process'. The results of the survey contained no missing data. In addition, the experts were asked to what extent they believed SD could enhance the three stages of healthcare innovation, again given the same 7-point Likert scale. For reference, the survey can be found in Appendix E.

4.4.1 Survey Results

Following the methodology of Section 3.2.2, the results of the survey questions regarding the Utility Perspective are presented in Figure 4.5 through the use of a stacked bar chart.

Following the methodology outlined in Section 3.2.2, differences between groups were analyzed. Regarding the first background question ('Research' vs 'Not Research'), the Mann-Whitney U test showed no statistically significant differences. The independent 2-sample T-test showed only 'Hypothesis Testing' to be statistically significantly different (-2.194, $p = 0.42^{*}$). However, the results should not hold very significant weight for three reasons, namely that the tests were performed on 16 questions, that the tested significance was relatively minimal, and that the Mann-Whitney U test showed no statistically significant difference. The remaining statistical test results are provided in Appendix F.

For the background question on familiarity, a Kruskal-Wallis test was performed. The results showed six questions to have statistically significant differences. These differences were further analyzed using a post-hoc Dunn's test, showing eight between-group differences to be statistically significant. The corresponding themes, their group median scores, and their significance level are presented in Table 4.5. The results of these tests show that experts with high SD familiarity seem to estimate the benefits of SD higher than experts with medium familiarity.

Use Case	**Low Familiarity**	**Medium Familiarity**	**High Familiarity**
Publishing More Replicable Research	2 (M*)	-1 (L*, H*)	3 (M*)
ML Model Training		0*	3*
Robustness Testing	1.5 (H*)	1 (H**)	3 (L*, M**)
Regulatory Certification Processes		-1**	2.5**
Software and Model Testing		-0.5*	3*
Integration & Deployment		-1**	2**

* ≤ 0.05, ** ≤ 0.01
L = Low Familiarity, M = Medium Familiarity, H = High Familiarity

Table 4.5: Overview of the statistically significant different answers between different groups of (perceived) familiarity with SD for questions 4, 5, 6, and 7 of the survey regarding the extent to which experts agreed that SD could benefit these use cases. The data shown are the median values of the group. Non-statistically significant results are not shown.

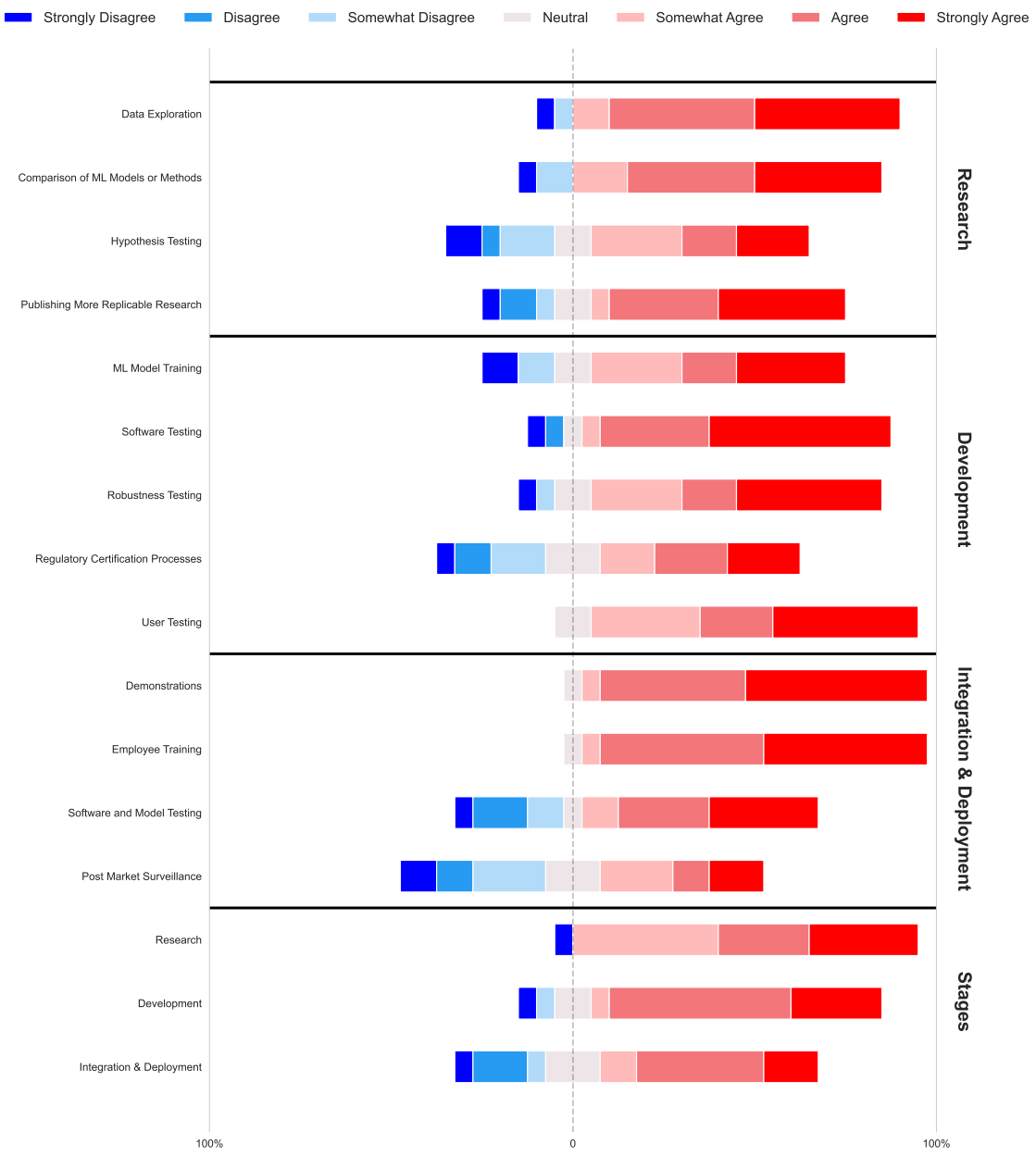


Figure 4.5: A vertically stacked bar chart visualization of the Likert scale answers to the questions regarding the Utility Perspective. When a bar tends more towards one side, it indicates a collective leaning towards that sentiment among respondents. Agreeing answers are visualized right from the middle, with non-agreeing answers being visualized left from the middle. Neutral answers are centered.

4.4.2 Framework Finalization

Now that all expert input has been presented, it can be integrated to finalize the framework. First, the three stages of healthcare innovation were graded differently, where experts seemed to perceive most potential for SD for 'research', followed by 'Development', and having 'integration & deployment' last. However, when testing with an independent 2-sample T-test, no statistically significant differences are observed between the various stages. It is also noteworthy that, although experts see most benefit in 'research' and half of the experts associated themselves most with 'research', the other experts did not grade it in a significantly different manner.

Next, for the framework, each of the use cases can be ranked in importance based on the opinions of the experts. Within each stage of healthcare development, the various use cases can be ranked by a two-stage process: (1) rank on the median value and (2) break ties using the mean value. Mean values were calculated using single integer intervals between the different levels. Because the mean of Likert scale answers has no intrinsic value, tie-brakes based on the mean value should not be attributed much value through visualization.

Next, independent 2-sample T-tests between the 1st ranked process and other processes within the stage are performed to assess whether there is a statistically significant difference. A white fill is applied to use cases that were perceived statistically significantly lower than the top use case. The resulting framework is visualized in Figure 4.6.

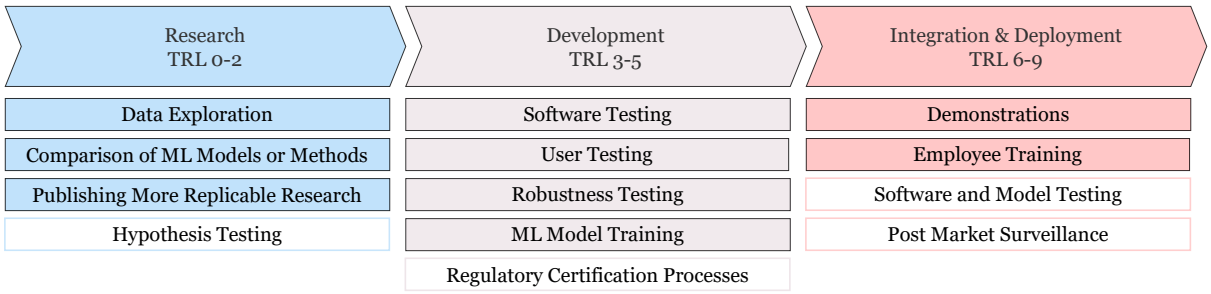


Figure 4.6: Refined and verified TRL framework, indicating for each stage of healthcare innovation what processes might be enhanced by employing SD. The order and colorization of the processes give an indication of the potential to be enhanced, according to the experts.

4.5 Summary of the Utility Perspective

The final framework, presented in Figure 4.6, concludes the analysis of the Utility Perspective. SD is an upcoming technology, and there previously was no systematically created overview of use cases verified by potential end users. For the Utility Perspective, an elaborate analysis was provided of existing literature, after which a framework was conceptualized and further refined with the help of experts in healthcare, ML innovation and SD. The resulting framework presents verified potential use cases of SD, grouped into the different stages of ML-based healthcare innovation. Stakeholders looking into the possibilities of SD can use the framework to assess how they could potentially employ SD and discuss whether they perceive value in these applications of the technology. An elaborate discussion of these results is provided in Section 6.

Chapter 5

The Uptake Perspective: Acceptance and Adoption of Synthetic Data

Now that the analysis of the Utility Perspective has been conducted, this analysis of the Uptake Perspective focuses on the factors that hamper and benefit the realization of the potential benefits of SD and influence its acceptance or adoption within healthcare. The Uptake Perspective analysis attempts to answer the following research questions as previously introduced in Chapter 1:

RQ2 How can SD be accepted and adopted into ML-based healthcare innovation in a human-centered way?

- RQ2a What are the requirements or conditions for SD that would make it useful for its stakeholders and increase stakeholder acceptance?
- RQ2b What are the motivations for different stakeholders to, or not to, integrate SD into their ML-based innovation processes?
- RQ2c How do human factors impact the incorporation of SD in the healthcare innovation process?

To answer these research questions, this chapter continues as follows. First, Section 5.1 presents and elaborates upon the six themes that were constructed through the thematic analysis of the conducted interviews. Subsequently, those results are used in Section 5.2 to construct a framework for the uptake of SD for healthcare innovation. Lastly, Section 5.3 elaborates on the refinement of this framework through the use of the follow-up survey.

5.1 Results of the Interviews

After the interviews were conducted, the conversations were transcribed and analyzed through a thematic analysis approach as previously described in Chapter 3. Through the thematic analysis of the interview data, six themes were constructed that describe the data. These themes, their subthemes, and the associated codes are presented in Table 5.1.

This section elaborates upon the six themes that were constructed from the interview data. Each of the themes will be defined, after which the codes within each of the (sub)themes will be elaborated upon. The themes will be discussed in order of how much of the conversations were spent on the topic, based on the number of references that were assigned to codes during the thematic analysis.

#	Theme (#interviews)	Codes (#interviews)
1	The Value Proposition of SD (24)	
	1.1 The Meaning of SD (24)	Collaborative Data Availability (12) SD for Upsampling (14) Proactive Adoption (7) Frustration of Data Availability (23)
	1.2 The Value Attribution of SD (23)	Awareness and Understanding (21) The Influence of Ethics (4) The Burden of Proof (15)
	1.3 The Perceived Quality of SD (23)	Comparing SD to Real Data (13) Leading (Past) Experience (10) Quality Skepticism (10) A Question of Trust (9)
2	The Technical Barriers to Creating SD (22)	The Need for Algorithms for Creating 'Good' SD (19) The Need for Methodologies for Assessing Whether SD is 'Good' (15) The Need for Understanding and Assessing Real Data and SD Bias (12)
3	Legal Adherence when Employing SD (24)	Privacy-Preservation of SD (8) Regulatory Certification and Clinical Validation (23) Understanding amongst Non-Legal/Privacy Stakeholders (5)
4	Organizational and Cultural Adoption Requirements for SD (22)	Capability Development (11) Dichotomies for SD Development Strategies (14) Requirement Ideas (21)
5	Uptake of AI & Data-Driven Innovation (14)	Interoperability with Existing Systems (8) Trustworthiness of AI (7) Awareness of AI (5)
6	Stakeholder Alignment (23)	Collaboration for SD Development (17) Stakeholder Resistance (12)

Table 5.1: An overview of the themes and most prevalent codes within each theme. Annotated in brackets are the number of interviews from which references were assigned to the theme or code.

5.1.1 Theme 1: The Value Proposition of Synthetic Data (VP)

The first theme, the value proposition of SD, regards the perceived benefits synthetic data provides to improve or accelerate ML-based healthcare innovation development. This includes potential cost savings, possibilities for better patient privacy preservation, increased collaborative data availability, and the potential for dataset augmentation (scalability/diversity). It is noteworthy that the 'value proposition' in this context does not refer to a complete and thorough analysis of the advantages and disadvantages of SD and whether a return on investment can be realized. The value proposition in this context refers to the perception of the proposed value that the concept of SD offers. The theme is split up into three sub-themes, namely (1) motivations for SD, (2) the value attribution of SD, and (3) the perceived quality of SD.

5.1.1.1 Motivations for Synthetic Data

The first sub-theme regards the various motivations that the stakeholders brought up to support their stance on why they believed SD can or cannot benefit the ML-based healthcare innovation process. In all 24 interviews, some form of motivation or meaning was described. Next, these motivations are elaborated upon.

Collaborative Data Availability First, SD is envisioned by twelve interviewees to be useful for collaborative data availability, to enable easier data sharing, and to improve collaboration between stakeholders within healthcare innovation. For example, interviewee #6 said: *"Well, I think it is, partially, a nice solution that makes sharing data easier (#6)."* Similarly, when asked what would be the biggest advantage of SD from their point of view, interviewee #22 stated: *"I think, easier collaboration with other parties (#22)."* Although most simply referred to collaboration and data availability, some specifically mentioned the motivation of patient privacy maximization or the motivation of minimizing the number of people that have access to patient data: *"I see that as a way in which you can exchange data more easily with a company [...] as a hospital, they naturally have to protect their data because that's what you expect as a patient when you come there (#19)."*

Synthetic Data for Upsampling The second motivation for SD is for upsampling, or the process of extending or interpolating data to create more data, as discussed by fourteen participants. Interviewees referred to upsampling through various terminologies such as 'dataset augmentation' or 'creating more data diversity'. For example, participant #4 mentioned: *"So well, then you don't have enough data to train an [application]. But if you synthesized more [...] and you have a more diverse distribution of the data that you need for training, then we would expect that the detection works better.. that the algorithm can be better developed (#4)."* However, there was also skepticism for using SD for upsampling: *"creating synthetic data to expand the sample size is not going to work (#18)."* Apart from having larger datasets, seven participants mention that their motivation for SD stemmed from a promise of data quality or application robustness through additional data. For example, expert #14 stated: *"But perhaps you can use that synthetic data to create much better quality datasets (#14)."*. Similarly, participant #4 discussed: *"Synthetic data could be a very powerful tool for such purposes, boundaries of an algorithm, for instance. [...] Robustness checking, or under which conditions, under which parameters an algorithm works perfect or under which it doesn't (#4)."*

Proactive Adoption A secondary motivation for taking up SD was the proactive adoption of SD. For example, four experts mentioned that SD might be the norm for healthcare innovation in the future: *"At some point, people are going to be surprised if you don't use synthetic data. Maybe not now, but maybe like in 10 years (#4)."* Three other experts mentioned that academic hospitals might want to look into SD purely from the perspective of an innovative endeavor: *"But*

those [academic hospitals] [...], they must be curious about 'okay, [...] what new [technology] can we propose that can help us further to benefit patients' interests (#6)." However, such motivations were never mentioned as a single motivation, but solely in the context of other motivations.

Overarching Frustration of Data Availability During discussions on the different expected purposes of SD technology, interviewees often voiced frustration about low data availability, as mentioned in fifteen interviews. Many stakeholders that are doing research or application development within healthcare innovation struggle with accessing useful data: *"You've been at [research] for almost a year [...] to ultimately ensure that you receive that data from the hospital. That is not an exception, and if you can already start with [...] fake data. [...] Then you will eventually get further (#9)."* This frustration is further compounded by a trend that accessing data is getting more problematic over time: *"I got more skeptical that we'd ever be able to use real data, and so synthetic data became more important. Synthetic data is really our best alternative to all of the challenges with data privacy (#7)."* Similarly: *"I think regulations become stricter in different countries [...] almost daily regulations are being put in place to protect [...] healthcare information. [...] a second factor is that many persons become more skeptic [...] more and more persons just stand up against that whole notion of that their data being monetized while they themselves do not really benefit from it (#2)."*

5.1.1.2 The Value Attribution of SD

This sub-theme regards the considerations of individuals affecting their value attribution to the general concept of SD. For example, someone might attribute more value to a particular type of SD because of its technological characteristics, whereas someone else might not be aware of the potential benefits and consequently not attribute substantial value to the concept of SD. Such considerations are contained within this sub-theme of value attribution of SD.

Awareness and Understanding Amongst the participants, there was a wide range of experience with SD. For example, when asked about their experience with the concept of SD, expert #20 stated: *"No experience.. and I can imagine all kinds of things, but not that I know what it is (#20)."* Similarly: *"The funny thing is that it is indeed very new and many people have never heard of it, but there are also many people who are actually quite advanced in it (#15)."* Thirteen interviewees tended to generalize the capabilities or characteristics of SD based on their past experiences with SD. For example, when talking about the possibilities of SD, interviewee #11 stated: *"Looking at synthetic data, I can say, hey, I want to see a blood vessel here that is 98% closed (#11)."* Although this is true for some forms of SD, it is not true for all. When later asked about different kinds of SD, interviewee #11 stated: *"Well, maybe you're going a step too deep for me [...] to really have knowledge of that (#11)."* Similarly, participant #16 stated *"of course, a lot of synthetic data is already being used (#16)."* When asked about specifics of what kind of SD they referred to, it was specified to be through statistical/physiological models (see Section 2.2.1 on approaches to SD generation). Related to generalization and oversimplification, there is uncertainty amongst eight interviewees about what actually can be regarded as SD: *"is that then synthetic data? [...] that is a question of definition (#5)."*

The Influence of Ethics Although not often mentioned, ethical arguments were occasionally brought up in conversations regarding SD. Only four conversations yielded any form of ethical arguments, three of which stated that ethical arguments should positively affect the acceptance or adoption of SD. They argued that SD provides an opportunity for minimizing the amount of patient data required for research and development of medical applications. Thus, by using SD for innovation, the privacy and safety of patients whose data would otherwise be needed is improved, avoiding potential harm and improving patients' control of their data: *"You will*

need less participants [...] research is inherently risky [...] and you take away some of that risk, at least for participants, and that would be an ethically good thing to do (#2).” However, ethics did not surface in many of the interviews. The fourth and last expert to mention ethics within the interviews stated: *“I think some people will bring up the ethical discussion, but I think that is less relevant (#5).”*. Here, the participant is alluding to the idea that ethical considerations do not heavily influence the uptake of SD.

Burden of proof There was a large demand for proof points that SD can provide useful benefits, as fifteen experts mentioned that they would like to see (more) proof. Although four experts mention the topic in the context of economic viability, fourteen of the participants refer to whether the concept of SD can be beneficial for them, as exemplified by interviewee #1: *“I think that it is important to show good proof points to show where it has yielded acceleration or an improvement of the algorithm. And, that is actually also the point that we currently struggle with (#1).”* Although SD makes interesting promises, interviewees showed an urge for proof points that these promises can come true. In addition, there was also demand for having proof points for specific contexts: *“We shouldn’t be biased by saying ‘oh, it was proven in the heart, so therefore it must work in the brain.’ We don’t know that (#13).”* Lastly, opinions differed on what impact the proof points should show. For example, some seek proof points that show a real business success: *“I think like a business success kind of a win (#07).”*. Others refer to the need for many use cases: *“There too, there will have to be a lot of proof points of the possibilities. Showing many use cases in which it works [...] an algorithm that works in the same way in terms of architecture, but trained on synthetic data in which, for example, it works better (#8).”*

5.1.1.3 The Perceived Quality of SD

This sub-theme specifically regards the perceived value of SD as a result of the perceived attainable quality of SD. Note that the definition of quality of SD depends on the perception of the interviewee, and is thus also heavily dependent on the context of the interviewee. Take the following quote: *“What I am skeptical about is whether you can get the quality of synthetic data good (#17).”* In this quote, there are three ill-defined concepts, namely ‘quality’, ‘synthetic data’, and ‘good’, but the result is that the expert was skeptical of SD as a result of the perceived attainable quality of SD, which is the core discussion within this sub-theme. Although this sub-theme is mentioned separately, it can also be viewed as a sub-sub-theme of the Value Attribution sub-theme. There are several prevalent observations that are worth discussing.

Comparing SD to Real Data First, thirteen participants actively assessed the quality of SD by comparing it to the quality of real data. When comparing SD with real data, nine interviewees tended not to think that handling or working with the quality of SD is inherently a larger problem than handling or working with the quality of real data. When discussing the risks of SD, participant #14 stated: *“the risks seem quite clear to me, but that is not to say that those risks are not also associated to real data (#14).”* Similarly: when asked whether expert #13 saw any risks associated to SD, they said: *“The only risk is with... and it can happen even with real data, is to fool yourself that you have sufficient representation (#13).”*

Past Experience There was a tendency for participants to assess the attainable quality of SD based on their personal experiences with the current state of SD quality, as was done in 10 of the interviews. For example, a negative past experience could lead to scepticism as resembled by the following: *“If you’re trying to do a synthetic dataset of electronic health records which show care paths [...] that’s a really tough place to go. We tried to do some of that [...] we couldn’t get the distributions to line up (#7).”* However, it also worked the other way around, where a positive past experience can lead to a certain belief in SD: *“I already referred to*

'thispersondoesnotexist.com' [...], it has nothing to do with clinical data, but I think it is very interesting that you can generate very realistic things based on existing data (#9)."

Quality Skepticism Ten of the participants stated in some form to be skeptical about the attainable quality of SD. They stated that 'simple' data could be synthesizable. However, when it came to highly-dimensional data, high frequency time-series data, or generally (more) complex data, they were skeptical of the attainable quality: *"the datasets that we use for that [research], are high dimensional.. I don't know them, the algorithms that can generate them (#5)."* Similarly, referring to the preservation of (clinical) value when synthesizing data, participant #18 said: *"The idea that if we can just move a little bit the distribution, then that's enough to de-identify them, and we preserve everything else [...], It's dreaming (#18)."*

Trust Lastly, nine different interviewees brought up the importance or the need for trustworthy SD. This comes in two different forms. Four mentioned the trustworthiness of the quality of SD, whereas six mentioned the trustworthiness of the quality of downstream algorithms that are trained using SD. For example, when asked about the biggest barrier to using SD, interviewee #19 stated: *"So, I think trusting that the quality of synthetic data is already of sufficient quality (#19)."* However, trustworthiness was also viewed as a problem relating to the quality of the downstream algorithms. Interviewee #8, for example, mentioned the following when asked about what would influence hospitals to embrace SD: *"It will be very much a question of trust. The algorithm [...] will have been trained without ever having seen a real patient. How do we know that [...] a real patient is also correctly identified (#08)."* When asked about what will help to create such trust, interviewees responded with varying answers, including the need for proof or demonstrability of quality and experience or knowledge of SD.

5.1.2 Theme 2: The Technical Barriers to Creating Synthetic Data (TB)

This second theme, the technical barriers to creating SD, regards the challenges and impediments related to the technological capabilities and limitations that an organization might face when implementing synthetic data solutions for healthcare innovation, including the capabilities of synthetic data to create representative datasets and the ability to assess the extent to which synthetic datasets are representative, privacy-preserving or biased. There are three major technical barriers that are commonly brought up.

The Need for Algorithms for Creating 'Good' SD Without good SD creation algorithms, the promises of SD will not be realizable. Nineteen participants discussed the importance of being able to create 'good' SD. For example, when asked about the largest barrier to making SD useful, participant #12 stated: *"At the moment.. if you really want to have good data, that it is still very difficult to create it, and that creating good synthetic data may still be [...] too complicated to [...] get that coverage (#12)."*. Various topics were still problematic, according to the experts, such as technological maturity, generating proper diversity, and the existence of techniques for synthesizing different data modalities (types of data, i.e., images, text, etc.). For example, the synthesis of multi-table (database) datasets offers unique problems: *"If you have an electronic patient record database of some 100 tables. [...] If you want to synthesize that [...] There are quite a few issues involved. For example, scalability, referential integrity, and how the tables are linked (#15)."* Different data modalities have different technical barriers. This is also perceived as a barrier by participants: *"I do believe that if you can synthesize the table of one, you can also synthesize the table of the other. But I wouldn't easily believe that for very different types of data (#19)."*

The Need for Methodologies for Assessing whether SD is 'Good' Although creating SD is one aspect, assessing whether SD is actually of good quality is another, as was mentioned by fifteen interviewees. It was deemed essential that there are reliable, explainable, intuitive, and easily interpretable ways with which the quality, representativeness, bias and privacy-preservation of the SD can be assessed: *"it is very difficult to determine when synthetic data, one, [...] can no longer be traced back to the original persons, and two, how can you grade that [...], so how can you attach a number and make sure it's correct (#17)."*. Similarly, participant #19 stated: *"It depends on how you can justify the metric. [...] If you want me to trust it, then I must have a reason to trust it. Otherwise, I can simply say, 'This is a 6, and this is an 8' (#19)."* Not having understandable and interpretable ways to test the quality of SD is experienced as a barrier to the acceptance and adoption of SD: *"That bothers us a bit, because when is it good.. when can you use it.. how can you measure whether something is good or not (#17)?"*

The Need for Understanding and Assessing Real Data and SD Bias The bias in real and synthetic data presents a substantial technological challenge and was brought up in twelve conversations. Stakeholders deemed it essential to recognize such biases in order to prevent harm downstream: *"Data is laced with disparities and inequities, and we need to understand them to prevent them from being encrypted into algorithms. At present, we don't know how to do that (#18)."* One of the fears of experts is that when creating SD based on real data, the bias of the real data can potentially be encrypted into the SD: *"I can imagine that synthetic data, [...] [that] there is a bias in that because you have generated that data, learned from existing data, the model comes from somewhere with that data, there may be a bias in that (#20)."* When asked about the differences between real data bias and SD bias, the problems are perceived as similarly problematic: *"well, it's similar in that no one knows (#18)."* When asked whether SD bias is larger than real data bias, participant #20 answered: *"No, [it's] unknown (#20)"*, later stating (in the same context): *"People have no experience with synthetic data [bias]. They have no feeling with it. And if you don't have a feeling with it, you have less trust with it (#20)."*

5.1.3 Theme 3: Legal Adherence when Employing SD (LA)

The third theme, legal adherence when employing SD, concerns the compliance of SD with relevant laws, regulations, and ethical standards in the healthcare industry, particularly those regarding patient privacy and medical device regulations. The theme includes discussions on the difficulties of accessing real patient data due to privacy concerns and bureaucratic complexities, debates over the definition of privacy preservation of SD, uncertainty about its role in regulatory certification and clinical validation, and a general lack of understanding about legal and privacy issue among non-legal and non-privacy stakeholders.

Privacy-Preservation of SD The first discussion within this theme regards the uncertainty surrounding the privacy-preserving capabilities of SD, as mentioned by eight participants. Defining 'privacy-preserving SD' is both a technical and a legal challenge, influenced by evolving personal data protection regulations, industry standards, and jurisprudence. As interviewee #3 mentioned: *"Then the question is where we set the bar or the limit, that we think it is too close a copy, that can be retrieved again (#3)."* Similarly, interviewee #17 stated: *"It is very difficult to determine when synthetic data is [...] synthetic enough, so it can no longer be traced back to the original persons (#17)."* Although regulations such as GDPR state what is and isn't allowed, there is significant uncertainty on how to interpret those regulations: *"The only thing I can conclude.. ask it to three different legal experts, you will get three different opinions. Ask it to three different data protection officers, you will get three different opinions. [...] That is what they call the shadow of the law, we still have to fill that in (#5)."*

Regulatory Certification and Clinical Validation Strictly separate from the previous discussion is the debate on what is or is not allowed regarding the use of SD in the context of regulatory certification. Ten experts included some elaboration on this topic, showing a range of interpretations and levels of uncertainty: *"At the moment, I do not necessarily see [synthetic data] as something that we can use practically to put new medical solutions in the field, because the European regulators, and also the American regulators, and the Asian regulators, do not yet really know how to deal with this, what percentage of your data can be synthetic and what percentage of your data must be real-world data (#11)."*

In the context of regulatory approval and certification, twenty-three participants indicated that medical application verification necessitates real data: *"At the end of the road.. I think you must have a real dataset to test the algorithm (#1)."* Similarly: *"But ultimately, we will have to validate every kind of new application on our own local data (#21)."* There is an interesting distinction between what various experts interpret the regulators to accept, and what experts themselves might accept. There were seven experts who reasoned that if an application is validated thoroughly on real data, then the risks are acceptable, even if the application was trained on SD. For example, participant #20 stated: *"If in the end the validation [on real data] is done well [...] then you have taken out the risks, or at least limited them [...] to an acceptable level (#20)."* Similarly, when asked about whether they distrust applications developed using SD more compared to applications developed using real data, participant #16 answered: *"If you ultimately validate on the real data and that turns out to work, then no, in my opinion (#16)."*

Perceptions and Misunderstandings among Non-Legal/Privacy Stakeholders Non-legal/privacy experts (experts not recruited specifically for their expertise in legal or privacy, see Section 3) often fail to grasp why SD might not always be privacy-preserving and why there is a significant concern amongst legal and privacy experts about the associated risks: *"that a chest X-ray is personally identifiable health information and cannot be used, right? [...] it's not funny after you fight with the lawyers for a while (#7)."* Similarly: *"To give you an example, a waveform that you create.. that it is so unique that it belongs to a patient in the eyes of the privacy officer [...] personally I totally disagree with that (#3)."* In some cases there is an inclination to better understand the disagreement: *"I still find that very difficult, you know. Someone still has to explain to me that when I just make up data, how it can be.. what is not anonymous about that.. It's just made-up data (#22)."*

5.1.4 Theme 4: Organizational and Cultural Adoption Requirements for Synthetic Data (AR)

The fourth theme, organizational and cultural adoption requirements for SD, regards the changes in organizational structure, processes, and culture that are necessary for the effective adoption of synthetic data for healthcare innovation, such as necessary FTEs, monetary investments, strategy planning, or management communication.

As almost all included experts were still thinking about the possibilities of SD, most did not hold strong opinions on organizational and cultural adoption requirements for integrating SD. When asked at the end of the conversation whether participant #9 had anything to add, they responded: *"Well, you already asked me more than I actually thought [...] I had never really thought about the implementation (#9)."* As a result, there is little tension between stakeholders, and most content within this theme is the ideation on how to approach adoption from an organizational point of view.

Capability Development There is a broad consensus that adopting SD requires investment in capabilities, expertise, and timely execution. When asked about possible steps to embrace adoption, participant #2 mentioned: *"It's also about capacity building by bringing in persons*

(#2)." Similarly, participant #8 stated: *"We will have to have the mindset, and potentially in due time we also need to have the infrastructure and the knowledge on board (#8)."* However, getting access to the right knowledge is also perceived as a large barrier. When asked about the largest barrier to integrating SD, participant #11 simply stated: *"Talent. [...] It is not just someone that is capable of creating synthetic data. But you are also looking for someone with a certain medical background to generate relevant synthetic data (#11)."*

Dichotomies for SD Development Strategies When experts were convinced that SD generation capabilities should be developed in-house and not be outsourced, then they had various opinions on how to approach such development. Logically, each expert bases their idea or recommendation on the context that they are familiar with. However, there seemed to be a dichotomy regarding whether to develop SD centrally within an organization, or use a project-specific development, and only centralize after sufficient experience has been acquired: *"Where the scientists say like 'hey, I want to create this, so I will need training data'. Well, and then they walk back to the department that is going to make the synthetic data (#11)."* This is in stark contrast with, for example, participant #24: *"I do not think that we need specific teams for that. Because the needs are very different (#24)."* Later, the same participant specified: *"I think that the biggest barrier is that we are too ambitious.. So, you just have to look more at practical things. Where do you want to use it, and just start with that (#24)."*

Other Requirement Ideas Apart from investment needs and development strategies, experts brought up a host of ideas. For example, nine interviewees mentioned a need for creating trust amongst stakeholders: *"Yeah, so the market access folks, the payers, and then you have to convince the physician customers, the healthcare providers, that the technology is safe and effective (#7)."* Additionally, ten participants mentioned that there is a need for creating awareness about the potential merit of SD: *"I'm big in favor of proper [...] education of the public (#2),"* or *"Usually it is about knowledge or education anyways. That it is not novel or scary anymore, but normal (#10)."* Nine participants mentioned that deployment of SD presents a noteworthy barrier. During conversation #15, when asked about whether there were barriers related to deployment: *"That indeed depends on the hospital.. It is a pretty large point of attention at the moment to make the deployment process as easy as possible (#15)."* In addition to the aforementioned ideas, there were at least another four ideas mentioned during the interviews. Specifically, six participants mentioned a need for creating industry standards, four a need for change leaders, seven a need for a fitting organizational culture or mindset, and five mentioned considerations for when to outsource SD generation capabilities or when not to.

5.1.5 Theme 5: Uptake of AI & Data-Driven Innovation (AI & DD)

This fifth theme, the Uptake of AI & data-driven innovation, regards the willingness and readiness of stakeholders to incorporate AI and data-driven technologies into their healthcare practices, including their general inclination to pursue AI & data-driven innovation and their current use and trust of AI & data-driven applications. This relates to embracing AI & data-driven innovation in general, not just synthetic data, and thus covers a much broader scope. Whilst there is a general interest in ML & AI amongst the interviewed stakeholders, the uptake of AI is challenged by issues of operational efficiency, technological maturity, trust, and the capability to adopt and integrate these systems.

Factors influencing AI & Data-Driven Innovation Uptake As already presented in the literature review, there are a plethora of factors that influence the uptake of AI & data-driven innovation. A common factor within the interviews, mentioned by eight participants, was the barrier of interoperability: *"The technological debt that a hospital has is just very.. [...] There*

are just systems, you really can't run AI on that (#14)." Not just hardware interoperability is mentioned, software interoperability also presents a significant barrier: *"Purely the integration of new IT systems in the hospital in a way that it also [works] within the healthcare processes.. that is currently the biggest challenge [...] Our patient dossier [system] doesn't do that (#10)."* However, interoperability with the workflow was also noted as critically important, as experts mentioned that the (AI) interoperability with the workflow is where concerns regarding accountability and human-AI cooperation surface: *"What if the computer says no [...] I think that you very much have to look at how you implement it.. what the agreements around it are (#22)."*

Other factors mentioned during the interviews, next to interoperability, included the trustworthiness, interpretability, and explainability of AI, as mentioned by seven interviewees. When asked about barriers for integrating AI applications, interviewee #16 answered: *"One of the barriers is obviously the trust of the doctor. [...] I think explainability of AI is also a very important [factor] (#16)."* Also, AI awareness was stated by five participants as a factor influencing the uptake of AI & data-driven innovation: *"I think anyways that much more awareness needs to come first [...] I think that doctors are generally distrusting towards [...] AI in general (#23)."*

5.1.6 Theme 6: Stakeholder Alignment (SA)

The sixth and final theme, stakeholder alignment, regards the discussions about and congruence between the goals, expectations, and interests of different stakeholders. The definition and inclusion of who are considered stakeholders depends on the context. For a hospital, 'stakeholders' might include hospital staff but can also feature external stakeholders. For a research collaboration, 'stakeholders' might mean multiple different parties and organizations and various potential stakeholders within each of the different parties and organizations.

Do we Need Collaboration for SD Development? When asked whether, apart from themselves, other stakeholders were necessary for the integration of SD, interviewee #7 stated: *"You know that is a good question. [...] I think the honest answer is maybe.. it depends (#7)."* Experts gave no clear answer as to what inter- and intra-organizational cooperation is necessary for the acceptance and adoption of SD. Although seventeen interviewees had something to say about stakeholder cooperation, there was a wide range of opinions. While some perceive that no further collaboration than the already established ones is necessary, others state the need for contractual agreements and cooperation through inter-organizational collaborations.

Stakeholder Resistance Throughout the previous five themes, various barriers and challenges have been discussed that can all lead to forms of stakeholder resistance. For example, quality skepticism can lead to resistance to the acceptance of SD for healthcare innovation: *"Real-world data is not perfect, and we just don't know how to make it better, like replacing it with synthetic data is the wrong approach, because we ourselves do not understand the depth and breadth of data bias (#18)."*. Similarly, skepticism of the added value can lead to resistance to the technology: *"[Integrating SD] only makes sense if it adds something on top of real data (#10)."* This statement was made in a similar vein for both application development and research. Additionally, a noted resistance during the interviews was that of privacy and legal personnel: *"My experience is that in many cases [Data Protection Officers and Chief Information Security Officers] are risk-averse. And if there is any doubt.. [...] So that also blocks innovation (#6)."*

5.2 Framework Development

The Uptake Perspective strives to facilitate the acceptance and adoption of SD in a human-centered way. Through the results of the interviews, it becomes evident that SD is not necessarily useful in all contexts. Thus, the framework should also strive to facilitate the decision-making on

whether SD can be beneficial in a certain setting, which can aptly be defined as the acceptance of SD. The initial approval or agreement to start using SD is likely influenced by the aforementioned themes. Given initial approval or agreement to start using SD, it should then be the question of how to approach the adoption of SD, again influenced by the aforementioned themes. As a result of these observations, a conceptual framework was developed, visualized in Figure 5.1.

The core of the framework shows that it is likely to benefit the organization to first get to a state of acceptance of SD and subsequently move further towards the adoption of SD. The themes likely influence the acceptance and adoption of SD in differing ways. The first five themes are all hypothesized to influence the acceptance. Themes 2, 3, and 4 are hypothesized to influence adoption. Lastly, stakeholder alignment, theme 6, is hypothesized to be required for each of the themes to achieve acceptance and adoption of SD for healthcare innovation.

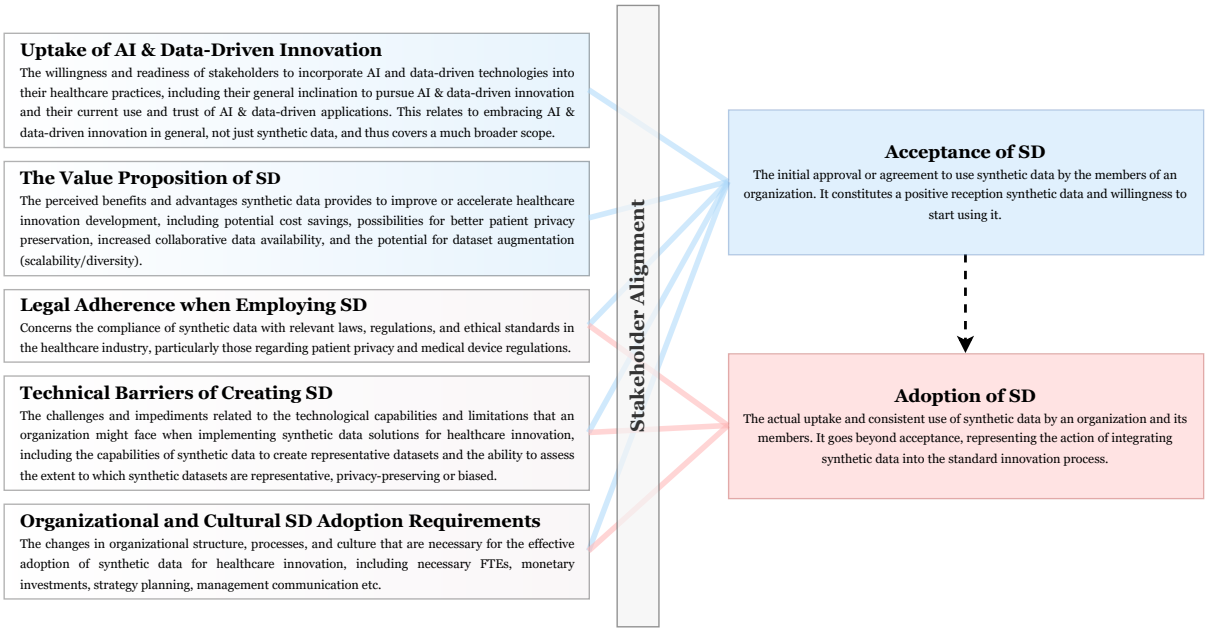


Figure 5.1: The conceptual framework for the Uptake Perspective

5.2.1 Elaboration upon the Hypothesized Connections

In hypothesizing the relationships between the themes and the core components, a balance and trade-off is to be made between usefulness and truthfulness. Truthfully, every theme is likely to have some influence on both themes in some contexts. However, it is more useful only to show connections when they are bound to have a major influence in a majority of the applicable contexts. The following describes the considerations in making this trade-off.

The uptake of AI & data-driven innovation is hypothesized to only influence the acceptance of SD and not the adoption of SD. From the interviews, the theme was primarily mentioned as a facilitating factor. That is, if a healthcare organization is not actively pursuing AI and data innovation, they are prone to resist accepting SD technologies. Consequently, once an organization decides to accept SD to improve its healthcare innovation process, it will presumably mean they are already invested in AI and data-driven innovation. In other words, when deciding whether to accept SD, organizations will have to assess whether their AI & data-driven innovation practices are sufficiently advanced. However, when adopting SD, they already deemed those practices sufficiently advanced and moved forward. Consequently, while a strong inclination towards AI and data-driven innovation is essential for initially accepting SD, it plays a lesser role in the subsequent adoption process, as the commitment to SD likely signifies an existing dedication to AI and data-driven innovation.

The second factor, the value proposition of SD, is also hypothesized to only significantly affect the acceptance of SD. In a similar vein to the previous argument, stakeholders need to align on the value that they seek to extract from adopting SD for healthcare innovation as a prerequisite for adoption and integration. Once stakeholders agree to a certain presupposition of added value, an organization can start adopting with that value in mind. During adoption, the value is a goal to work towards and is likely a relatively constant factor.

Legal adherence when employing SD is hypothesized to be of major influence for both the acceptance and adoption of SD. First, the opinions of legal and privacy personnel presumably hold significant weight in assessing whether the potential risk of SD is acceptable given the context that the organization is in. In addition, once determined that the risk is acceptable, legal and privacy personnel will have to work together with technical staff to ensure that the technical implementation and integration of SD capabilities is in line with their perspective.

Similarly, the technical barriers to creating SD will probably significantly affect both the acceptance and adoption of SD through a similar argument as the previous theme. Specifically, the perceived technical barriers are expected to influence the acceptance as their significance might be perceived as insurmountable or too expensive. If perceived as surmountable or mitigable, they will have to be tackled during the adoption and hence will have a significant influence on the process, timing, and eventual usefulness of the technology.

Lastly, the organizational and cultural SD adoption requirements are presumed to influence both the acceptance and adoption of SD significantly. First, an initial estimation of possible needed investment, talent and knowledge recruitment, and perceived time needed for change management is presumed to heavily influence the perceived feasibility of attaining the added value. In certain contexts, a commitment to a certain investment might be needed before an adoption process is started. However, during the integration process, talent recruitment, change management, and other required organizational aspects will heavily shape the adoption process. For example, if acquiring the right knowledge turns out to be more of a burden than initially expected, it will have a significant influence on adoption.

5.3 Framework Refinement

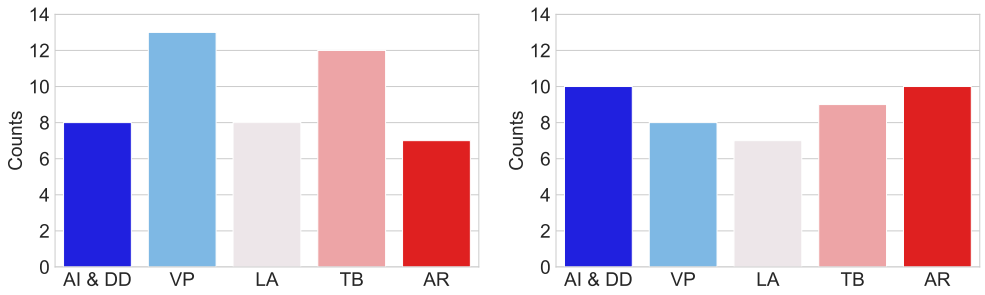
Now that a conceptual framework has been created, it can be refined with further participation of the experts. There are three aspects that are further refined through a quantitative survey.

The first two aspects of the framework that are refined concern the connections of the various themes with the acceptance and adoption of SD. For these aspects, the survey contained two very similar questions. The question on acceptance read: *Which of the following aspects do you think are essential for the acceptance of synthetic data for healthcare innovation?* For adoption, the question read: *Which of the following aspects do you think are essential for the acceptance of synthetic data for healthcare innovation?* Participants were allowed to choose up to three of the six options available. The first five of the six options were the themes from the established framework (except stakeholder alignment). The sixth option was open-ended, allowing experts to specify any other aspect they found more significant than the provided themes.

The third question in the survey regarding the Uptake Perspective regards the last theme, stakeholder alignment, and read: *Please indicate for the following themes to what extent you agree or disagree that obtaining stakeholder alignment is important for the acceptance and adoption of synthetic data for healthcare innovation.* For this third question, experts could indicate their opinion on a 7-point Likert scale. For reference, the survey can be found in Appendix E.

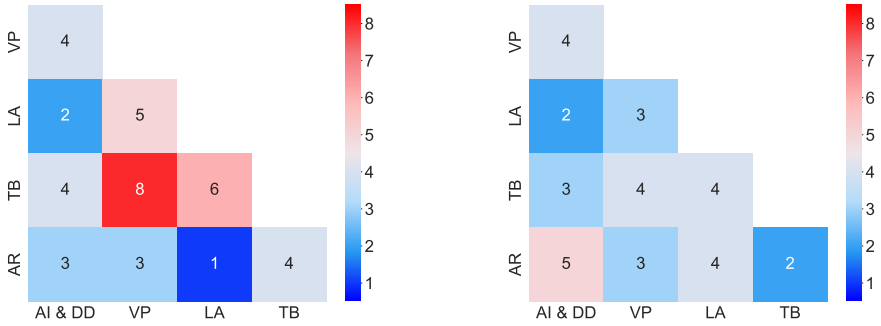
5.3.1 Survey Results

The cumulative answers to the first two questions are visualized through bar charts in Figure 5.2. Noticeably, the 'Other' option is not included in the visualization. There were a total



(a) Acceptance (b) Adoption

Figure 5.2: Bar charts displaying the counts of themes that experts deem most important for the acceptance (plot a) and adoption (plot b). Themes are abbreviated (see Appendix B)



(a) Acceptance (b) Adoption

Figure 5.3: Heatmap visualizations displaying the counts of answer combinations of themes that experts deem most important for the acceptance (plot a) and adoption (plot b). The names of themes are abbreviated (see Appendix B for abbreviations)

of five inputs for the 'Other' option, where experts entered a custom answer. Two indicated that the participant did not think SD should be accepted or adopted, which does not provide any information for factors influencing acceptance and adoption. Hence, these answers were deleted. One custom answer noted a specific part of the technical barrier, and hence, the answer was reclassified to represent this. Lastly, two custom answers indicated a need for proof points. Although not specifically mentioned in the definition of the value proposition theme, it was elaborately discussed as part of the value proposition previously. The two answers were reclassified to reflect this. Apart from this visualization, a χ^2 and Fisher's Exact test was performed to compare the answers regarding the acceptance to the answers concerning adoption. Interestingly, no statistically significant difference was found. Consequently, the survey indicates that there is no inherent statistically significant difference between the influence of the factors on the acceptance and the adoption of SD for ML-based healthcare innovation.

The initial frequency distribution visualization of Figure 5.2 can be further extended by elucidating upon the options often chosen together. To give an intuitive visualization of which of the answers were commonly chosen together, a heatmap visualization of answer combinations is provided in Figure 5.3. A complete tabular overview of single choice and combinatory choice frequency distributions is provided in Appendix G.

After having obtained an indication of the results through visualizations, between-group differences were assessed according to the methodology described in Section 3.2.2. Using χ^2 tests and Fisher's Exact tests, the between-group differences between experts associating themselves with 'Research' and experts associating themselves otherwise were assessed. No statistically

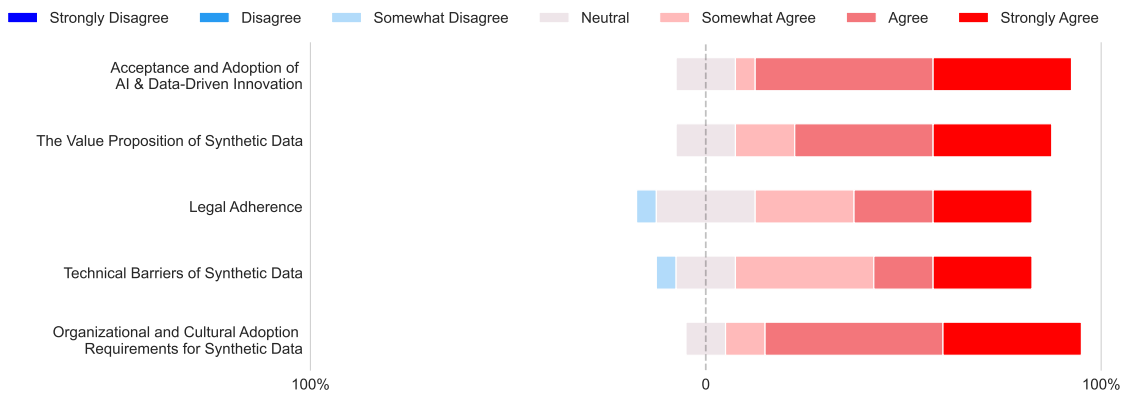


Figure 5.4: A vertically stacked bar chart visualization of the Likert scale answers to survey question 11: *Please indicate for the following themes to what extent you agree or disagree that obtaining stakeholder alignment is important for the acceptance and adoption of synthetic data for healthcare innovation.* A bar tending towards one side indicates a collective leaning towards that sentiment among respondent. Agreeing answers are visualized right from the middle, with non-agreeing answers being visualized left from the middle. Neutral answers are centered.

significant differences were found. Between-group differences regarding familiarity with SD were also assessed. Again no statistically significant differences were found. Consequently, the visualization provided in Figures 5.2 and 5.3 can be interpreted as the perspective of the experts.

Next, the answers to the third question regarding stakeholder alignment are visualized using a vertically stacked bar chart in Figure 5.4. The two missing values, present for the questions on the value proposition and the technical barriers, were ignored in the analysis. Next, between-group differences were analyzed according to the methodology outlined in Section 3.2.2. First, between-group differences for association with research were calculated using independent 2-sample T-tests and Mann-Whitney U tests. Both tests showed no statistically significant differences. Next, between-group differences regarding SD familiarity were calculated using Kruskal-Wallis tests. Three themes showed a statistically significant difference. A post-hoc Dunn's test was performed on these themes to understand which specific differences were significant. The outcome of these tests is presented in Table 5.2.

Use Case	**Low Familiarity**	**Medium Familiarity**	**High Familiarity**
The Value Proposition of SD	1.5*		3*
Legal Adherence for Employing SD	1 (H*)	0.5 (H**)	3 (L*, M**)
Organizational and Cultural Adoption Requirements for SD	2 (H*)	1.5 (H**)	3 (L*, M**)

* ≤ 0.05, ** ≤ 0.01
L = Low Familiarity, M = Medium Familiarity, H = High Familiarity

Table 5.2: Overview of the statistically significant different answers between different groups of perceived familiarity with SD for question 11: *Please indicate for the following themes to what extent you agree or disagree that obtaining stakeholder alignment is important for the acceptance and adoption of synthetic data for healthcare innovation.* The data shown are the median values of the group. Non-statistically significant results are not shown.

5.3.2 Framework Finalization

Now that the results of the survey have been presented, they can be used to finalize the framework. The results from the survey indicate no statistically significant difference between the themes important for the acceptance of SD compared to the adoption of SD. Given the further absence of any statistically significant relation regarding the connection of themes to the central concepts, a new simplified framework can be constructed. This new framework is visualized in Figure 5.5. The refined framework does not make any distinction concerning the themes that influence acceptance or adoption. Compared to the conceptual framework, only two connections are added, namely between the uptake of AI & Data-Driven Innovation and the Adoption of SD, and between the Value Proposition of SD and the Adoption of SD. Because the readability of the framework does not benefit from any specific ordering of the themes anymore, the themes have been reordered and ranked based on the total number of votes from the survey, with ties being broken using the number of interviews that the theme was mentioned in.

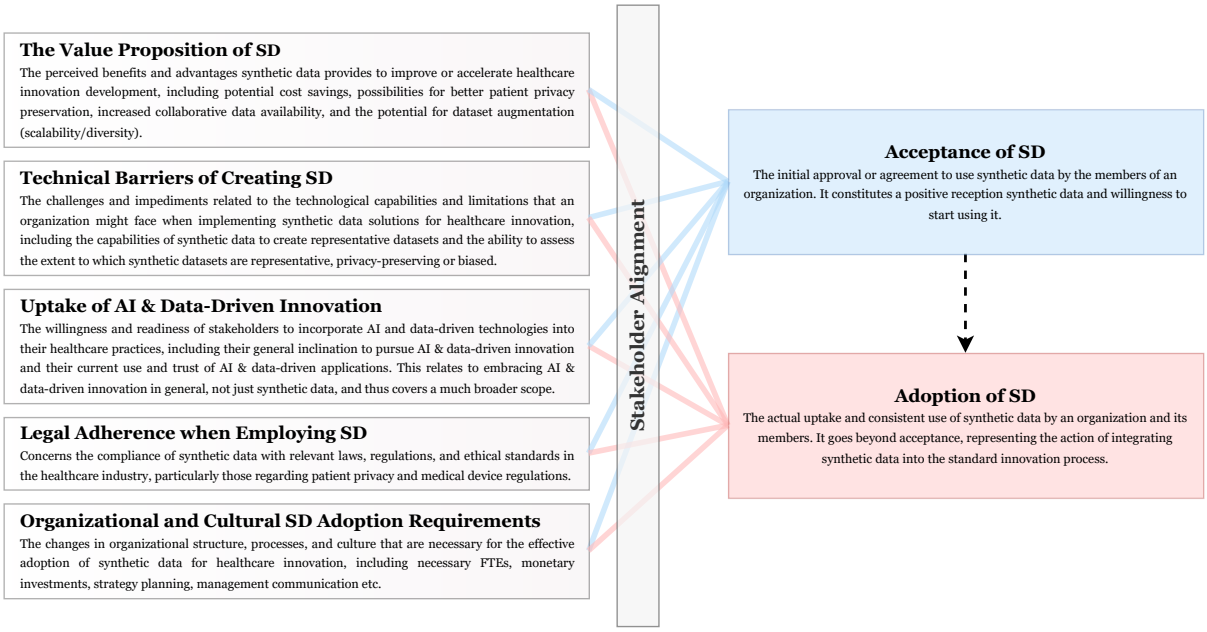


Figure 5.5: The refined framework for the Uptake Perspective, resulting from refining the conceptual framework (Figure 5.1) using the results from the surveys.

5.4 Summary of the Uptake Perspective

The final framework, presented in Figure 5.5, concludes the analysis of the Uptake Perspective. The Uptake Perspective analyzed how SD can be adopted and integrated into ML-based healthcare innovation in a human-centered way. The thematic analysis of the semi-structured interviews revealed six interrelated themes. The contents of the six themes were presented in-depth. Subsequently, a framework was constructed using the six themes. Using a follow-up survey, the framework was further refined and validated, leading to the final framework. Healthcare organizations are advised to recognize these six themes as influential factors and treat them as such for their decision-making processes regarding the acceptance and adoption of SD. The following chapter elaborately discusses the analysis of both the Uptake Perspective as well as the previous Utility Perspective.

Chapter 6

Discussion and Conclusion

The introduction stated that research on SD was overwhelmingly positive about the technology and seemingly overconfident in the capacity of SD to benefit healthcare innovation. Through analyzing the utility and uptake of SD using the input of experts in the field, this study presents a more balanced overview of this capacity. This last chapter will integrate the findings of the two perspectives by answering the research questions and interpreting the results of both perspectives, providing a broad discussion of the two perspectives and finalizing this study with a conclusion.

6.1 Study Summary

This study has researched how SD can be effectively integrated into various stages of the ML-based healthcare innovation process in a human-centered way. It has done so by investigating the matter from two perspectives using the input of 26 experts through semi-structured interviews and a follow-up survey.

The analysis of the Utility Perspective led to the creation of a framework that lists processes for each stage of the ML-based healthcare innovation process. Three stages of ML-based healthcare innovation were identified, namely (1) 'research', (2) 'development', and (3) 'integration & deployment'. Within each of these stages, several processes were identified that can, according to the experts, potentially be enhanced by using SD, such as data exploration in research, software testing during development, or demonstrations during integration & deployment. It presents the first structured, generalizable, verified, and actionable overview of use cases of SD.

In addition, the analysis of the Uptake Perspective led to the creation of a framework that presents six themes that all influence the acceptance and adoption of SD for ML-based healthcare innovation. Specifically, the uptake of SD is influenced by (1) the value proposition of SD, (2) the technical barriers to creating SD, (3) the uptake of AI & data-driven innovation, (4) legal adherence when employing SD, (5) organizational and cultural SD adoption requirements, and (6) stakeholder alignment. Each of the themes contains a set of requirements that, if not met, are likely to hinder the uptake of SD. In addition, the uptake of SD for ML-based healthcare innovation is moderately influenced by human factors, such as the awareness and understanding of SD, trust in SD and other stakeholders, and the inability of stakeholders to handle information overload, leading to cognitive biases during decision-making.

6.2 Answers to the Research Questions

With the results of the analyses, answers to the research questions that were proposed in Chapter 1 can be specified. The research questions of both perspectives were split up into three sub-questions. For both perspectives, the answers to the sub-questions will be provided first, after which these answers are aggregated to provide an answer to the two overarching questions.

6.2.1 The Utility Perspective

The answers to the research questions of the Utility Perspective are relatively straightforward to deduce from the reported results, as individually, they represent the steps undertaken to achieve the results.

RQ1a - How does healthcare innovation, specifically for ML solutions, currently operate? This was a key question in the initial analysis. Literature of process models and TRL models was first assessed in Chapter 2, and TRL frameworks were further worked out during the analysis of the Utility Perspective. This led to the presentation of two frameworks, being the MLTRL framework from Lavin et al. [119], and the HTRL framework, constructed from seven different healthcare-related TRL frameworks. Tables 4.1 and 4.2 present these frameworks, respectively.

The frameworks show that ML-based healthcare innovation likely progresses through three general stages of innovation, being research, development, and integration & deployment. The frameworks provide a general overview of how ML-based healthcare innovation operates. Not all innovation endeavors progress through these steps in a similar manner, but the frameworks provide a useful guideline on the common occurrences within the innovation process.

RQ1b - Which aspects of ML-based healthcare innovation involve the use of patient data? The conceptual taxonomy, visualized in Figure 4.1, provides an array of processes that involve the use of patient data. The conceptual taxonomy does not necessarily provide a complete overview of all processes in ML-based healthcare innovation that use patient data, but the experts identified no additional processes belonging to ML-based healthcare innovation that did not highly overlap with the processes stated in the conceptual model. The taxonomy hierarchically sorts data-driven processes into the three previously mentioned stages of ML-based healthcare development. Examples of processes include 'reproducing research' under the research stage, performing 'clinical trials' under the development stage, or 'educational training' under the integration stage.

RQ1c - Which of these processes have the potential to be enhanced or expedited using SD? Although the conceptual taxonomy already provides an answer to this question, experts were consulted through interviews and a follow-up survey to answer this question more thoroughly. The analysis resulted in a table of potential processes that could be enhanced by employing SD, which is presented in Table 4.4. Subsequently, experts were asked to grade the potential for these processes to be enhanced by employing SD on a 7-point Likert scale. Their answers are visualized using a vertically stacked bar chart in Figure 4.5, giving an insightful overview of what processes experts deem to have the potential to be enhanced or expedited by employing SD. Experts were mostly positive on the potential of SD to enhance or expedite the listed processes. Experts were most skeptical of using SD for hypothesis testing during research, using SD for regulatory certification processes during development, and using SD for software and model testing and post-market surveillance during the integration and deployment stage. However, in general, the majority of the experts believed that SD could enhance or expedite each of the identified stages of ML-based healthcare innovation.

RQ1 - How can SD be utilized throughout different stages of ML-based healthcare innovation? The proposed framework, visualized in Figure 4.6, provides one possible answer to this question. It provides an overview of possible processes for each stage of ML-based healthcare innovation, of which experts have stated that they think there is potential for SD to enhance the process. It does not provide a list of processes that can inherently be enhanced using SD, as the final outcome heavily depends on context and execution. The framework can

be utilized throughout the decision-making processes to pin down the perceived value of SD for each of the stakeholders. Additionally, it can be employed to understand the differences between the perceptions of different stakeholders on their perceived value of SD, which can help to reach a consensus during the decision-making processes.

6.2.2 The Uptake Perspective

The answers to the sub-questions of the Uptake Perspective are less easily deducible from the proposed framework compared to the Utility Perspective. This is in part due to the more exploratory nature of the Uptake Perspective compared to the Utility Perspective. Nevertheless, answers can be formulated for each of the questions based on the results of Chapter 5.

RQ2a - What are the requirements or conditions for SD that would make it useful for its stakeholders and increase stakeholder acceptance? Different stakeholders have different requirements, and hence, this question is not one with a simple answer. However, throughout the results of the thematic analysis, several requirements have been shown to limit experts' acceptance of SD.

To start with the value proposition, the main requirement for experts to perceive value in the technology is the availability and knowledge of applicable and generalizable proof points of the technology. Without such proof points, experts are skeptical of the attainable quality of SD and the concept of the technology as a whole. In an extension of the burden of proof, trust in the technology can be fostered by showing that the technology, if applied correctly, can be used to realize benefits [138, 169, 122]. However, at this moment, applicable and generalizable proof points are limited, and several of the experts were unaware of existing proof points.

Next, the technical barriers consist entirely of requirements for employing SD, namely (1) the need for algorithms for creating 'good' SD, (2) the need for methodologies for assessing whether SD is 'good', and (3) the need for understanding and assessing real data and SD bias. If any of these requirements are not met, stakeholders will be cautious about or resistant to the use of the technology. Existing literature on SD predominantly focuses on these requirements.

In addition, legal adherence when employing SD is a requirement in and of itself. If the technology cannot be employed whilst adhering to laws and regulations, then the technology will unlikely realize any of the promised benefits. Technology experts, privacy experts, legal experts, lawmakers, international standards organizations, patient advocacy groups, and many more stakeholder groups will likely have to collaborate to ensure that the technology is able to comply with relevant laws, regulations, and ethical standards. For example, when SD is to be used for collaborative data availability, it is also essential to establish and validate the positioning of SD as non-personal data within the legal framework. In addition, when SD is to be used for training and testing of healthcare innovations, it is of utmost importance that the development and verification of those applications follow and adhere to the right clinical standards. Such collective agreements are essential for leveraging the full potential of SD in healthcare innovation.

Next, the uptake of AI & data-driven innovation is a requirement in and of itself. Although not as strict a requirement as legal adherence, stakeholders are more likely to perceive a potential value in SD if there is a decent uptake of AI & data-driven innovation. Consequently, the uptake of AI & data-driven innovation can best be seen as a facilitating condition for the uptake of SD.

Furthermore, the theme of organizational and cultural SD adoption requirements contains a diverse set of ideas for requirements. Apart from the almost unanimous understanding that SD requires investment, different stakeholders offered widely different insights regarding what requirements are needed for the successful acceptance and adoption of SD. Different contexts, such as whether it is adopted in a hospital or a health company, also require different organizational and cultural requirements. It is thus likely up to the organization seeking to adopt SD to figure out what their organizational and cultural requirements are.

Lastly, stakeholder alignment is presumably required if an organization wishes to realize any value from SD. In the absence of stakeholder alignment, the uptake of SD will likely be hampered by any number of stakeholders that have not been properly involved in the uptake of SD.

RQ2b - What are the motivations for different stakeholders to, or not to, integrate SD into their ML-based innovation processes? The answer to this question is also multi-faceted but can be summarized in two short points. First, the sub-theme of the value proposition regarding the expected purpose of SD extensively elaborates on various motivations to adopt SD. These motivations primarily fall into (1) SD for collaborative data availability and (2) SD for upsampling. These motivations are grounded in an overall frustration of low data availability. Second, motivations not to integrate SD are almost identical to the answers of the previous research question. Specifically, if any of the requirements is not fulfilled, stakeholders will presumably not agree with or resist to the integration of SD for ML-based healthcare innovation.

RQ2c - How do human factors impact the incorporation of SD in the healthcare innovation process? This research question was initially set up with the expectation that human factors would play a large role in the uptake of SD. The results of the study show that human factors play a role in the uptake of SD, but a relatively small role compared to the influence of technological and organizational factors.

First, experience with SD plays a role in the confidence that experts have regarding their opinions. Experts with more awareness and understanding of SD will likely be able to form a more elaborate opinion and might better know the limitations and possibilities of the technology. There is a consensus in the literature that experience positively influences technology adoption rates [221, 16, 160]. Although awareness and understanding played a role in how experts formed their opinions, experts highly familiar with SD showed no statistically different choices to other groups in questions on what themes influence the acceptance and adoption of SD. Experience does not necessarily need to be positive and can thus also lead to resistance.

Second, cognitive bias refers to any form of deviation of rationality in judgment and includes various types of biases such as confirmation bias, anchoring bias, overconfidence bias, negativity bias, or oversimplification. Cognitive biases play a role in most human-made decisions and can lead to negative long-term outcomes [210]. For example, past experience has been demonstrated to potentially result in overconfidence bias, which has been shown to have a positive and significant influence on investment decisions [107, 179]. To provide a second example, simplification of complex topics can yield a more intuitive understanding, but oversimplification can yield counterproductive theories and take away from obtaining truth and understanding [67]. Consequently, when discussing about the uptake of the complex topic of SD, which has multiple types of applications, multiple yet unresolved technical barriers, and a lot of uncertainty, stakeholders should be cautious of the influence of cognitive biases.

Third, trust is a recurring factor throughout the identified themes. It did not receive its own theme as it plays a relatively small role within each of the themes and takes on different roles throughout the different themes. Although trust is mentioned throughout the various themes, trust is mostly dependent on the knowledge of the participant about SD and the availability and generalizability of proof points. Trust is also mentioned with regard to the quality of the SD, for which many would like to see an intuitive and understandable (trustworthy) explanation of why the SD is of good quality. Participants also state that there needs to be trust in the final application. Note that trust for AI applications in healthcare has been researched already [14, 73]. Experts also noted that trusting other stakeholders to make wise decisions is also important in the uptake of SD for healthcare innovation. There is a consensus in the literature that (inter-organizational) trust helps in the adoption of new technology [84, 124, 52, 28].

RQ2 - How can SD be accepted and adopted into ML-based healthcare innovation in a human-centered way? The Uptake Perspective framework provides one possible answer to this question. It presents six themes that influence the acceptance and adoption of SD for ML-based healthcare innovation. When an organization is interested in better understanding whether SD can provide a benefit for them, the framework can help in breaking up unfocused discussions into six relevant topics, mitigating cognitive biases such as oversimplification or negativity bias. Based on this research, one would be advised to start with discussing the value proposition to ensure that discussions can remain focused on an envisioned goal of realizing the proposed value. In contrast, it is also possible that there is no perceived value for a given context. However, if stakeholders agree on a form of potential value that can be realized, the other themes provide a structure to discuss the necessary requirements for realizing that potential. Although the framework provides a structure for discussions, it leaves out nuance with respect to specific requirements, motivations, and human factors for readability and clarity. However, these nuances are not unimportant in the uptake of SD. Based on this research, stakeholders are advised to take special care in constructing requirements for the adoption of SD, including the creation of trustworthy SD, and be cautious of cognitive biases that might impair successful acceptance and adoption, such as overconfidence or negativity.

6.3 Theoretical Implications

This study builds upon the extensive literature presented in Chapter 2. The following section reflects on how the existing literature is integrated into the proposed frameworks and additionally reflects on how the existing body of literature is enhanced through the contributions of this work.

6.3.1 Theoretical Implications for the Utility Perspective

The aim of the Utility Perspective was to enhance the understanding of SD, and hence the academic literature, by providing an overview of use cases of SD, which builds upon the current literature while also adding depth through the aggregation of expert opinions. The following reflects on how the analysis of the Utility Perspective provides additional depth to the current understanding of SD utility.

As elaborately covered in the research of the Utility Perspective, use cases of SD have, up to now, not been analyzed systematically. Mostly, use cases stated in literature are a direct result of their technical characteristics [13, 25, 39, 43, 49, 87, 99, 166, 215, 216], such as to "overcome the paucity of annotated medical data in real-world settings" [39]. Some literature presents topic-based use cases such as fields where SD potentially adds value, like 'epidemiology research' [74]. If literature does go into detail on potential use cases, these use cases mostly vary in scope or have significant overlap with one another [74, 93, 135, 217].

The analysis of the Utility Perspective has resulted in the first structured overview of processes of ML-based healthcare innovation, which can potentially be enhanced or expedited using SD. Although previous literature presents use cases, the proposed framework is the first to integrate other literature on use cases and subsequently verify these use cases with experts both qualitatively and quantitatively.

Looking at the processes identified throughout the analysis, there is considerable overlap with much of the existing literature, such as algorithm testing [74], education and training [74] or system development [217]. This is especially visible through the overlap between the conceptual taxonomy of processes and the processes mentioned by experts during the interviews, as visualized in Figure 4.4. The visualization indicates that what experts think SD can and cannot be used for has a reasonable overlap with what the published literature suggests.

Use cases, as stated in this research, stand out compared to the existing literature in their

actionability. Previous literature stated use cases such as 'rare events' or 'machine learning' [135, 93]. Although these use cases indeed provide opportunities where SD can provide value, they lack the possibility to use that information to actively adopt the technology. By presenting verified use cases in a format that also provides a generalizable context, stakeholders looking to understand the possibilities of incorporating SD can more easily assess whether they perceive the technology to have potential value in their specific context.

6.3.2 Theoretical Implications for the Uptake Perspective

The academic aim of the Uptake Perspective was to enhance academic literature by building an understanding of the (non-technical) factors affecting SD acceptance and adoption, rooting them in established management theory. The following reflects upon how the established management theory surfaces throughout the proposed framework of the Uptake Perspective.

6.3.2.1 Relation to Classical Acceptance and Adoption Models

As the framework was constructed through a highly exploratory process, the factors often seen in classical frameworks for acceptance and adoption of (novel) technologies are not directly included in the framework. However, the established acceptance and adoption factors are visible throughout the themes of the Uptake Perspective.

Whether a novel technology is perceived to be beneficial is in established literature largely captured through factors such as 'perceived usefulness', 'relative advantage' and 'performance expectancy', which are often the dominant factors for models such as TAM, IDT and UTAUT [50, 170, 205]. There is not a singular theme that encompasses these concepts completely. However, the value proposition theme is a large part of this perceived usefulness, discussing the applicability of SD in a specific context, including the needed quality of SD to be useful for realizing that benefit. Classical models such as TAM, IDT and UTAUT also have the common factor of 'perceived ease-of-use', 'complexity' and 'effort expectancy' [50, 170, 205]. These factors are again captured by multiple themes. For example, the perceived technical barriers to creating SD play a large role in the perceived effort expectancy and, subsequently, in the actual effort during the adoption process. Noteworthy is that 'perceived usefulness' and 'perceived ease-of-use', or their UTAUT variants, are not necessarily disjoint concepts. For example, if legal adherence when employing SD is perceived as a significant barrier, it can limit both the perceived usefulness of SD and the perceived ease of use. This can be viewed in two ways, namely (1) as limiting the perceived ease-of-use, which in turn influences the perceived usefulness, or (2) as influencing both factors separately. These views are also captured in the classical acceptance models as well. In TAM, perceived ease-of-use is proven to influence the perceived usefulness [50]. In contrast, UTAUT does not link 'effort expectancy' to 'performance expectancy' in any of the models (UTAUT 1, 2 or 3). However, some extensions of the theory do make this connection, also in the context of healthcare [82]. For the themes regarding technical barriers, legal adherence, and the uptake of AI & data-driven innovation, it can all reasonably be argued that they likely influence both the perceived usefulness and the perceived ease of use. For example, if the technical barriers are perceived as particularly difficult, it likely diminishes the expected usefulness and presumably also decreases the perceived ease of use. All in all, the perceived usefulness and perceived ease of use are important factors for the acceptance and adoption of SD, as the established literature would also have suggested.

The traditional UTAUT model has two more factors influencing intention to use, namely social influence and facilitating conditions [205]. Social influence is not recognized as a separate theme in the proposed framework for SD acceptance and adoption. In contrast to the literature, social influence was not identified as a dominant and separate factor influencing the acceptance or adoption of SD, in line with findings from Rouidi et al. [172]. There is a separate theme that can be classified as a 'facilitating condition', namely the uptake of AI & data-driven innovation.

Having a solid base of acceptance and adoption of AI is likely to help in the acceptance and adoption of SD technologies. In other words, not having that solid base might lead to resistance of SD.

Looking at IDT specifically, it has three factors in addition to the previously mentioned variations on 'relative advantage' and 'complexity', namely 'trialability', 'observability' and 'compatibility' [170]. These are not easily identifiable from the proposed framework. However, trialability and observability do play a role in the value proposition, as there is a significant need for proof points for the usefulness of SD, as previously discussed in Section 5.1. Compatibility was seen throughout the themes, for example, in the theme regarding organizational and cultural adoption requirements, where participants noted that interoperability and deployment are noteworthy barriers to look at.

Lastly, looking at the TOE framework, there is a visible influence of the technological, organizational as well as environmental factors [194]. Technological factors are clearly featured through the themes regarding the value proposition and technical barriers. Organizational influences are apparent through the adoption requirements and the uptake of AI & data-driven innovation. Lastly, environmental factors are visibly featured through the inclusion of the themes regarding legal adherence and the need for stakeholder alignment, which covers both inter- and intra-organizational stakeholders.

Concluding, many of the factors present in the established technology acceptance and adoption literature can be located in the proposed framework. Noteworthy mentions are the perceived usefulness and perceived ease-of-use (and their variants), facilitating conditions, and organizational factors. The often included factor of social influence is less influential for the acceptance and adoption of SD for ML-based healthcare innovation.

6.3.2.2 Relation to Non-Classical Literature

Apart from further confirming traditionally identified factors from classical models, there are several other factors mentioned in Section 2.4 that previous literature has identified as influential for the acceptance and adoption of healthcare applications or AI. Application safety plays a significant role in the adoption of AI for healthcare [116, 103, 86]. This is also apparent from the results of the Uptake Perspective. For example, there were twenty-three interviews which highlighted the need for clinical validation on real data, particularly for maintaining patient safety. Overlapping with the influence of application safety is the influence of diagnostic accuracy [116, 103, 6, 86]. Evidently, the thematic analysis showed that ten participants were skeptical of the attainable quality of SD, particularly of the conservation of clinical value.

Literature has also identified factors such as infrastructure, regulations, leadership support, and the importance of including all stakeholders to be important in the uptake of novel technologies in healthcare [142, 112, 113]. Infrastructure is apparent in the uptake of SD through the adoption requirement and the uptake of AI & data-driven innovation, where infrastructural needs for both AI and SD are mentioned. The influences of regulation and multiple stakeholders are visible through the inclusion of the themes concerning legal adherence and stakeholder alignment. Although leadership support was often included in previous literature, it was not seen by participants as a highly influential factor or requirement for the uptake of SD during this study [218, 142, 112, 113].

Lastly, looking at the literature regarding the acceptance and adoption of SD, this study significantly extends the number of factors influencing the uptake. Previous literature on SD solely discussed use cases of SD and the technical barriers in realizing those use cases [7, 39, 71, 206, 114]. Although use cases are discussed, they do not discuss the variety of use cases and how technical barriers might change depending on the use case. Similarly, there is little regard for any inclusion of other stakeholders besides technical development stakeholders. For example, there is no indication that authors of privacy-preservation SD literature actively collaborate with privacy or legal experts, to the best knowledge of the author [58, 60, 97, 98, 147, 219].

6.4 Practical Implications and Perspective Integration

The Importance of a Multi-Stakeholder Analysis The introduction of this study noted that previous literature on SD does not actively acknowledge the need for a multi-stakeholder analysis with regard to the acceptance, adoption, and utility of SD for healthcare innovation. Throughout the study, there were several indications that stakeholder alignment is something worth looking at.

First, this research has shown that different stakeholders have different priorities when it comes to SD. When adopting SD, it is likely important to assess the level of accepted risk for every stakeholder. It is important to move beyond the observation that the adoption and integration of SD contain risk and actively assess the multi-dimensional risk factors associated with the technology, such as risks regarding privacy preservation, risks associated with the costs of the technology, and uncertainty regarding the clinical value of the SD. At its current state, adopting SD is inherently risky. Consequently, assessing the level of accepted risk and the methods that could mitigate the risks to an acceptable level is important to mitigate stakeholder resistance.

Second, apart from having different priorities, stakeholders can fundamentally disagree about core aspects of the technology. For example, some experts were aware that they did not agree with privacy and legal experts on the identifiability of SD. Regularly, when differences in perspectives between stakeholders were brought up, this was out of frustration with the opinion of the other group. Most experts who voiced frustrations about stakeholder differences voiced matters of privacy. Apart from frustrations regarding privacy, two experts noted that they disagreed with the SD development approach of others. It is important for the uptake and utilization of SD that stakeholders collaborate to define their requirements and concerns, and seek common ground when disagreements arise.

Lastly, the importance of a multi-stakeholder perspective is clearly visible through the differences of which experts mentioned anything corresponding to the uptake of AI & data-driven innovation in the Uptake Perspective. Factors that were coded to belong within this theme were only mentioned in fourteen of the conducted interviews, as opposed to a minimum of twenty-two different interviews for each of the other themes. Notably, out of those fourteen, only four were from experts recruited from industry, and the majority were mentioned by people from academia and hospitals. Looking at code occurrences, the distinction is even larger. One possibility is that the influence of the uptake of AI & data-driven innovations is most dominantly noted by people from academia and hospitals and that there is a possible dearth of awareness within the industry of this requirement. A possible deviating reason is that the experts associated to industry were more knowledgeable about SD, and conversations with the experts hence steered more towards SD than to AI in general. Nevertheless, the absence of this theme within the interviews with industry experts is notable.

Consequently, this research has clearly shown that the multi-stakeholder perspective is important, as stakeholders hold different priorities and fundamental disagreements and can also be unaware of the concerns of others. When starting to accept, adopt, and utilize SD, it is essential to scope out which stakeholders should be included in that process, map out their priorities and concerns, and attempt to reach a consensus on how to approach the process. By achieving stakeholder alignment, as also highlighted in the Uptake Perspective framework, the acceptance and adoption of SD is expected to have a higher chance of success.

The Dichotomy of SD Utilization One of the core distinctions made in this thesis is the dichotomy between the applications of SD. This is present in both the motivations discussed in the Uptake Perspective and the use cases as discussed in the Utility Perspective. With regards to the sub-theme concerning the motivations for SD, two primary motivations were stated, namely (1) SD for collaborative data availability and (2) SD for upsampling. This dichotomy

is also visible in the framework of the Utility Perspective. Specifically, the use cases associated with research are predominantly related to the idea of collaborative data availability, whilst the use cases associated with development are predominantly related to the idea of upsampling. During research, there is a need for additional data availability to explore and research new ideas. Researchers will likely not be in possession of the real data, hence SD (based on real data) needs to be created by the real data holders. Researchers themselves, hence, likely have relatively limited influence on the availability of SD for this purpose. In contrast, when employing SD for upsampling, in the context of application development, developers likely have a limited real dataset that they wish to augment. Here, developers themselves do have a significant influence on the availability of the SD as they can develop it themselves.

This distinction is a general observation and not true in all contexts. In some cases, for example, it might be beneficial to employ both. Noteworthy is that this distinction in use cases is not necessarily novel, as it has been stated in previous literature [87, 166]. However, the fact that this distinction is so apparent is not reflected in the literature. For example, van Breugel and van der Schaar [198] make a distinction between six primary use cases, including 'privacy' and 'augmentation' on the same level as four other use cases. Similarly, Murtaza et al. [144] present 'data augmentation' next to six others. Although the distinction is very apparent in this research and surfaced through two distinct bottom-up processes, the absence of the specific distinction in many other publications is notable [135, 51, 99, 74, 198, 144].

Naming this specific distinction has merit for improving the acceptance and adoption of SD. First, specifying either motivation for SD focuses discussions and thereby mitigates several cognitive biases during the uptake of SD. According to the experts, there are potential benefits for both applications. Specifying the perceived value to other stakeholders will, based on this research, likely help reach alignment and consensus. Second, specifying the purpose of the research can help to contextualize the research. There is merit in SD methods and metrics for both applications, and they do not need to be similar. Optimizing for data augmentation presumably results in different techniques than optimizing for privacy preservation. However, research on SD often does not specify their purpose according to this distinction, which would be heavily recommended based on this research.

Information Availability The current state of the uptake and utilization of SD is predominantly characterized by low awareness and high uncertainty. These factors are often seen as interrelated for emerging technologies [130]. As SD is an emerging technology, there is limited awareness and a high amount of uncertainty about the opportunities and limitations of SD.

It is important to note that low awareness does not necessarily lead to uncertainty. Nevertheless, based on interviews, experts with a low awareness of the potential applications of SD seemed to have a higher uncertainty about the added potential value of the technology. Similarly, experts with higher unawareness of regulations were generally less concerned about the privacy-preservation capabilities of SD. However, the participating privacy and legal experts were also uncertain about how the various regulations impact the utility of SD whilst usually being highly aware of the contents of these regulations.

The uptake of SD is not an easy endeavor. The technology has low trialability, low observability, high complexity, and a complex to analyze relative advantage. Given the multi-faceted associated risks of SD and the multi-disciplinary nature of healthcare innovation, the uptake and utilization of SD is not straightforward. Consequently, educating stakeholders on SD and engaging in conversations to increase the awareness and understanding of stakeholders is likely to benefit the success of an undertaking to adopt and utilize SD.

Using the Frameworks for the Uptake and Utilization of SD The two perspectives analyzed in this study both explore how SD can be effectively integrated into various stages of the ML-based healthcare innovation process. Subsequently, the frameworks presented as

answers to the respective research questions of the two perspectives can collectively contribute to accepting, adopting, and utilizing SD for ML-based healthcare innovation. The frameworks fulfill distinct roles.

The Utility Perspective framework provides a verified overview of possible use cases of SD. The framework can help stakeholders to understand where SD might provide value or how others might perceive the value of SD differently. SD is often seen as a technology that provides a general solution. However, the immaturity of the technology combined with the complex context of healthcare innovation necessitates the fulfillment of different requirements for different use cases. When a stakeholder with a certain context perceives SD not to be useful because of an unmet requirement, their negative cognitive bias is likely to impede perceiving the value of SD for other use cases. The framework can help to reduce such cognitive bias by providing a structured list of processes, verified by experts, that can potentially be improved by SD.

In contrast, the Uptake Perspective framework provides an overview of factors that influence the acceptance and adoption of SD. This framework can facilitate fruitful discussions by providing a structured overview of the various most influential factors. During discussions, stakeholders should try to contain discussions to one specific theme at a time. Based on this research, one would be advised to first focus discussions on determining what kind of value they would like to extract from SD and contain further discussions on topics such as whether realizing that value is doable given the perceived technical barriers. By helping to set a clear context for discussions, the framework will likely make the decision process more productive. For example, the framework will likely reduce whataboutism during discussions by recognizing and evaluating the relevant aspects of the scenario [85]. Whataboutism is characterized by the way in which counter-accusations may take the form of questions introduced with 'What about [...]?' For example, if one stakeholder would state, 'I think we could benefit from using synthetic data for collaborative data availability', a response grounded in whataboutism could be, 'But what about legal adherence?'. In such cases, whataboutism causes discussions to jump from topic to topic. The framework can help in reducing whataboutism by recognizing the different topics that need to be discussed, such that stakeholders can evaluate the facts of the situation on a per-topic basis [85]. The framework can thus facilitate productive discussions by sketching a guideline for such discussions.

These frameworks can thus both be used during the uptake of SD. When beginning to think about the possibilities of SD and following the advice of this research to start discussing the value proposition to assess whether to accept the technology, then both frameworks can be used. The Utility Perspective framework can help to pinpoint where certain stakeholders do, or do not, perceive value. In contrast, the Uptake Perspective framework can facilitate a focused discussion on the value proposition by highlighting which topics are primarily for subsequent discussions. After agreeing to work out a certain proposed value (i.e., SD for collaborative data availability or SD for upsampling), stakeholders can start assessing the requirements that are part of the other themes represented in the Uptake Perspective framework. The Utility Perspective framework might resurface in importance throughout these discussions, for example, when assessing the requirements of the technical capabilities of SD, which might differ per specific use case. The Utility Perspective framework might also resurface when discussing organizational SD adoption requirements when one use case might require more significant organizational support. Conclusively, the Uptake Perspective framework facilitates discussion topics, whereas the Utility Perspective framework can be featured for lookup during several of these discussions.

Collectively, the frameworks can help to effectively integrate SD into various stages of the ML-based healthcare innovation process in a human-centered way.

6.5 Limitations

This thesis is not without its limitations, which are both academic and practical in nature. The academic limitations primarily feature challenges related to the methodology and interpretation of results. The practical limitations primarily consist of limitations regarding the research population.

6.5.1 Academic Limitations

The academic limitations of the study primarily encompass challenges such as potential cognitive biases in thematic analysis, concerns about coding consistency and comprehensiveness, and the need for thorough justification of the thematic framework. Lastly, there is a note on the verification of the frameworks.

First, the interviews and following analyses were conducted by the main author, constrained by resource limits and the research context. Thematic analysis is known for its flexibility, with multiple analysts typically recommended to validate interpretations [213]. Although the main author regularly consulted supervisors, the primary interpretation of results was independently conducted. This approach potentially amplifies the cognitive bias of the main author in the research outcome beyond merely enriching the analysis through the non-positivist interpretation of qualitative research. Specifically, a single analyst may lead to reduced consistency in coding across data points. While this is less critical for directed content analysis compared to thematic analysis due to its systematic nature, the concern remains relevant.

Second, assessing the alignment between the derived themes in the thematic analysis and the coding framework from the established literature is important. A common issue in thematic analysis is the occurrence of topic summary themes over meaning-based interpretive story themes [34]. Topic summary themes broadly summarize a category, whereas meaning-based interpretive story themes delve into core ideas or meanings that unify observations within the theme. The themes of this study could be classified as both. However, the rationale for utilizing more general themes instead of themes reflecting the dominant expert opinion relates to the aim of the study. The objective was to deeply understand diverse expert perspectives on SD uptake, particularly given the nascent stage of the technology. Choosing a 'dominant' opinion for theme titles might misrepresent the wide range of viewpoints. Similarly, the study does not intend to conclude on the perceived benefits of SD but rather stimulate informed discussions relevant to specific contexts. For instance, initiating a discussion with a theme like 'Distrust of Technical Feasibility' might be according to the dominant thematic analysis literature, but does not optimally facilitate dialogues between healthcare stakeholders.

6.5.2 Practical Limitations

The practical limitations of this study are multifaceted, primarily concerning the sample size and population, including aspects like locality bias, survey dropout, survey population, and potential issues in survey responses.

First, the study involved 26 participants across 24 interviews, with a follow-up survey completed by 20 participants. These participants represented nine distinct stakeholder groups, averaging less than three individuals per group. Such a small group size per stakeholder category may not adequately capture the full range of opinions, leading to a lack of information saturation for each group and the study overall. Additionally, the recruitment strategy focused on Dutch experts in the Eindhoven region, introducing a locality bias. This bias is particularly relevant regarding legal adherence, where European and Dutch laws were the primary legal framework for participants. Therefore, the sample size and locality bias are crucial considerations when generalizing the findings of the study to other contexts.

Second, of the 26 participants, only 20 completed the follow-up survey, indicating a 23% dropout rate. Despite mitigation efforts, including notification emails and reminders within the two-week survey period, this dropout rate is relatively high. Additionally, as the follow-up surveys were considered optionally anonymous for privacy reasons, it cannot be guaranteed that the survey responses represented the originally intended research population well.

Third, the follow-up survey targeted the same participants who were interviewed. A broader and more diverse population could have potentially reduced locality bias and provided more generalizable results. Although this approach aligns with Delphi-study methods, the distribution of the survey might not need to share the practical limitations of the interviews, such as the optional physical presence.

Fourth, the lack of statistically significant differences in factors influencing the acceptance and adoption (see Figures 5.1 and 5.5) can lead to two differing interpretations. The first is that experts genuinely perceive all factors as crucial for both acceptance and adoption. The second interpretation questions the reliability of these findings. With only 20 experts responding and the relative complexity of the survey - introducing up to six new concepts in a single question - there is a risk that responses may not accurately represent the views of the experts. Complex survey questions can hinder accurate translation of perspectives into responses [137]. Consequently, the survey results and their interpretation, reflecting 20 experts who spent an average of 17 minutes each, should be viewed as indicative of diverse viewpoints rather than definitive facts.

6.6 Directions for Future Work

There are considerable opportunities for future research in the domain of SD, mostly linked to the identified requirements in the analysis of the Uptake Perspective. First, future research should continue their current work on the technical barriers. Technical barriers surfaced many times during the interviews and present a barrier to the uptake of SD. For example, at least two of the experts did not want to accept SD technologies until they saw proof that SD bias was sufficiently mitigated. Consequently, future research should aim to improve on generating high-quality SD, assessing the quality of SD, and mitigating SD bias.

Second, future research and industry should conduct extensive research on SD generation techniques to create applicable and generalizable proof points of the technology. A recurrent notion throughout this work was that SD is not necessarily trustworthy. Additionally, it was clear that having good examples in varying applicable domains and with varying data modalities would significantly aid in improving the trustworthiness of the technology.

Third, technical experts should collaborate with privacy and legal experts from industry to assess how the technology fits into current legal frameworks, both on privacy preservation and with regard to clinical certification. This might require the development of methodologies or standards that are deemed sufficient by the collective sector. Without standardization within the legal domain, the benefits of SD might not be easily realizable.

Fourth, if the uptake of SD takes off, future work could analyze whether the organizational and cultural change management requirements for adopting SD can be mapped. In Chapter 5 it was specified that mapping out these requirements is heavily dependent on the applicable context, such as the size of the organization seeking to adopt SD. However, once the sector has developed further, research might want to attempt to assess whether these change management requirements can be generalized to improve broader adoption.

Lastly, due to the scope of this research, the proposed frameworks have not in any way been either quantitatively validated or validated in an empirical setting. The frameworks represent the aggregate of current literature and the opinions of 26 experts. It will be interesting to see whether the frameworks can benefit organizations in practice, for example, through the application of the frameworks in a case study for a specific organization.

6.7 Conclusion

This study has explored two new perspectives on SD for ML-based healthcare innovation, being the Utility Perspective and the Uptake Perspective. The analyses combined the existing literature on innovation processes, SD, and technology acceptance and adoption with the opinions of 26 experts. Both analyses yielded a framework providing both theoretical insights for future research and actionable insights for healthcare organizations.

From this research, it can be concluded that different stakeholders in the healthcare sector perceive various values in the technology of SD. According to the participating experts, all stages of ML-based healthcare innovation can likely benefit from employing SD. The proposed framework for the Utility Perspective provides a structured and verified overview of possible use cases for the technology, grouped per stage of ML-based healthcare innovation. Both academia and industry should use this information to their advantage in understanding and advancing the technology. It is important to specify what value is perceived to mitigate confusion, uncertainty, and stakeholder resistance.

From the results of the Uptake Perspective, it can be concluded that there are six primary factors that influence both the acceptance and adoption of SD. Healthcare organizations should recognize these factors for their SD acceptance and adoption endeavors and not treat the acceptance and adoption of SD as depending solely on its technical characteristics. In addition, academia would do well to also recognize the broader field of acceptance and adoption of SD, starting to incorporate all stakeholders in the advancement of the technology.

Collectively, the frameworks can facilitate more fruitful discussions about the acceptance, adoption, and utilization of SD for healthcare innovation by providing a structured overview of the most important facets and considerations applicable to the process. In general, the majority of experts included in this study stated that they perceive value in SD for ML-based healthcare innovation but remarked the presence of several requirements that currently prohibit the realization of that value.

To conclude, based on this research, organizations seeking to understand and benefit from the potential value of SD are advised to start by understanding the perceived value of SD in their context, utilizing the Utility Perspective framework. Determining the value proposition of SD is the first step towards the uptake of SD for healthcare innovation. Next, this research suggests that discussions should be held between the necessary stakeholders to assess whether they see similar potential value, regardless of the potential barriers. After a consensus has been established on the perceived value, stakeholders are encouraged to map out requirements and concerns for each of the influential factors presented in the Uptake Perspective, which can differ per use case. Subsequently, stakeholders should inquire about the feasibility of addressing the requirements and concerns. Once these requirements and concerns are deemed addressable, given the contextual constraints, a state of acceptance to commit can be reached. If requirements and concerns are not deemed addressable, organizations should, at least temporarily, reconsider taking up SD. Once an initial approval or agreement is constructed, stakeholders should seek to adopt the technology.

Bibliography

[1] K. Abbas, M. Afaq, T. Ahmed Khan, and W.-C. Song. A blockchain and machine learning-based drug supply chain management and recommendation system for smart pharmaceutical industry. *Electronics*, 9(5):852, 2020.

[2] R. Agarwal and J. Prasad. A conceptual and operational definition of personal innovativeness in the domain of information technology. *Information systems research*, 9(2): 204–215, 1998.

[3] I. Ajzen. The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2):179–211, 1991.

[4] Y. Al Khalil, S. Amirrajab, C. Lorenz, J. Weese, J. Pluim, and M. Breeuwer. On the usability of synthetic data for improving the robustness of deep learning-based segmentation of cardiac magnetic resonance images. *Medical Image Analysis*, 84:102688, 2023.

[5] A. Alhasan, L. Audah, I. Ibrahim, A. Al-Sharaa, A. S. Al-Ogaili, and J. M. Mohammed. A case-study to examine doctors' intentions to use iot healthcare devices in iraq during covid-19 pandemic. *International Journal of Pervasive Computing and Communications*, 18(5):527–547, 2022.

[6] S. Alhashmi, S. A. Salloum, and C. Mhamdi. Implementing artificial intelligence in the united arab emirates healthcare sector: an extended technology acceptance model. *International Journal of Information Technology and Language Studies*, 3(3):27–42, 2019.

[7] C. Alloza, B. Knox, H. Raad, M. Aguilà, C. Coakley, Z. Mohrova, É. Boin, M. Bénard, J. Davies, E. Jacquot, et al. A case for synthetic data in regulatory decision-making in europe. *Clinical Pharmacology & Therapeutics*, 114(4):795–801, 2023.

[8] A. A. AlQudah, M. Al-Emran, and K. Shaalan. Technology acceptance in healthcare: A systematic review. *Applied Sciences*, 11(22):10537, 2021.

[9] D. AlSaleh and R. Thakur. Impact of cognition, affect, and social factors on technology adoption. *International Journal of Technology Marketing*, 13(2):178–200, 2019.

[10] S. Alsheibani, Y. Cheung, and C. Messom. Artificial intelligence adoption: Ai-readiness at firm-level. *PACIS*, 4:231–245, 2018.

[11] J. Amann, A. Blasimme, E. Vayena, D. Frey, V. I. Madai, and P. Consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC medical informatics and decision making*, 20:1–9, 2020.

[12] E. Ammenwerth. Technology acceptance models in health informatics: Tam and utaut. *Stud Health Technol Inform*, 263:64–71, 2019.

[13] C. Arnold and M. Neunhoeffer. Really useful synthetic data–a framework to evaluate the quality of differentially private synthetic data. *arXiv preprint arXiv:2004.07740*, 2020.

[14] O. Asan, A. E. Bayrak, and A. Choudhury. Artificial intelligence and human trust in healthcare: focus on clinicians. *Journal of medical Internet research*, 22(6):e15154, 2020.

[15] R. Atallah and A. Al-Mousa. Heart disease detection using machine learning majority voting ensemble method. In *2019 2nd international conference on new trends in computing sciences (ictcs)*, pages 1–6. IEEE, 2019.

[16] R. A. Balgah, D. N. Bwifon, and P. N. Shillie. Coid-19, armed conflict and icts adoption decisions. insights from cameroonian farmers. *Advances in Social Sciences Research Journal*, 2022. doi: 10.14738/assrj.98.12763.

[17] E. Barbierato, M. L. D. Vedova, D. Tessera, D. Toti, and N. Vanoli. A methodology for controlling bias and fairness in synthetic data generation. *Applied Sciences*, 12(9):4619, 2022.

[18] S. Baškarada and J. Watson. Managing the exploitation-exploration tradeoff: how leaders balance incremental and discontinuous innovation. *Development and Learning in Organizations: An International Journal*, 31(4):13–16, 2017.

[19] B. Basu, S. Ghosh, B. Basu, and S. Ghosh. Assessment of technology and manufacturing readiness levels. *Biomaterials for Musculoskeletal Regeneration: Applications*, pages 235–246, 2017.

[20] R. A. Bauder and T. M. Khoshgoftaar. Medicare fraud detection using machine learning methods. In *2017 16th IEEE international conference on machine learning and applications (ICMLA)*, pages 858–865. IEEE, 2017.

[21] M. S. Bauer and J. Kirchner. Implementation science: What is it and why should i care? *Psychiatry research*, 283:112376, 2020.

[22] P. Bazeley and K. Jackson. Qualitative data analysis with nvivo. *Qualitative data analysis with NVivo*, pages 1–376, 2019.

[23] A. L. Beam and I. S. Kohane. Big data and machine learning in health care. *Jama*, 319 (13):1317–1318, 2018.

[24] J. Beinecke and D. Heider. Gaussian noise up-sampling is better suited than smote and adasyn for clinical decision making. *BioData Mining*, 14:1–11, 2021.

[25] S. M. Bellovin, P. K. Dutta, and N. Reitinger. Privacy and synthetic datasets. *Stan. Tech. L. Rev.*, 22:1, 2019.

[26] O. Ben-Assuli. Electronic health records, adoption, quality of care, legal and privacy issues and their implementation in emergency departments. *Health policy*, 119 3:287–97, 2015. doi: 10.1016/j.healthpol.2014.11.014.

[27] N. Bhirud, S. Tataale, S. Randive, and S. Nahar. A literature review on chatbots in healthcare domain. *International journal of scientific & technology research*, 8(7):225–231, 2019.

[28] I. Bile Hassan, M. A. A. Murad, I. El-Shekeil, and J. Liu. Extending the utaut2 model with a privacy calculus model to enhance the adoption of a health information application in malaysia. In *Informatics*, volume 9, page 31. MDPI, 2022.

[29] B. Billot, D. N. Greve, O. Puonti, A. Thielscher, K. Van Leemput, B. Fischl, A. V. Dalca, and J. E. Iglesias. Synthseg: domain randomisation for segmentation of brain scans of any contrast and resolution. *arXiv preprint arXiv:2107.09559*, 2021.

[30] M. Black, J. Wallace, D. Rankin, P. Carlin, R. Bond, M. Mulvenna, B. Cleland, S. Fischaber, G. Epelde, G. Nikolic, et al. Meaningful integration of data, analytics and services of computer-based medical systems: the midas touch. In *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, pages 104–105. IEEE, 2019.

[31] B. W. Boehm. A spiral model of software development and enhancement. *Computer*, 21 (5):61–72, 1988.

[32] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[33] V. Braun and V. Clarke. To saturate or not to saturate? questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health*, 13:201 – 216, 2019. doi: 10.1080/2159676x.2019.1704846.

[34] V. Braun and V. Clarke. Toward good practice in thematic analysis: Avoiding common problems and be (com) ing a knowing researcher. *International Journal of Transgender Health*, 24(1):1–6, 2023.

[35] E. Breck, S. Cai, E. Nielsen, M. Salib, and D. Sculley. The ml test score: A rubric for ml production readiness and technical debt reduction. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 1123–1132. IEEE, 2017.

[36] E. M. I. Center. TU/e - Technische Universiteit Eindhoven, unknown. URL https://www.tue.nl/en/research/research-groups/eaisi/eaisi-business-operations/eindhoven-medtech-innovation-center. [Accessed January 10th, 2024].

[37] D. S. Char, N. H. Shah, and D. Magnus. Implementing machine learning in health care—addressing ethical challenges. *The New England journal of medicine*, 378(11):981, 2018.

[38] S. Chatterjee, N. P. Rana, Y. K. Dwivedi, and A. M. Baabdullah. Understanding ai adoption in manufacturing and production firms using an integrated tam-toe model. *Technological Forecasting and Social Change*, 170:120880, 2021.

[39] R. J. Chen, M. Y. Lu, T. Y. Chen, D. F. Williamson, and F. Mahmood. Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering*, 5(6): 493–497, 2021.

[40] V. Cheng, V. M. Suriyakumar, N. Dullerud, S. Joshi, and M. Ghassemi. Can you fake it until you make it? impacts of differentially private synthetic data on downstream classification fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 149–160, 2021.

[41] H. S. J. Chew and P. Achananuparp. Perceptions and needs of artificial intelligence in health care to increase adoption: scoping review. *Journal of medical Internet research*, 24 (1):e32939, 2022.

[42] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman, et al. Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface*, 15 (141):20170387, 2018.

[43] E. Choi, S. Biswal, B. Malin, J. Duke, W. F. Stewart, and J. Sun. Generating multi-label discrete patient records using generative adversarial networks. In *Machine learning for healthcare conference*, pages 286–305. PMLR, 2017.

[44] R. G. Cooper. Stage-gate systems: a new tool for managing new products. *Business horizons*, 33(3):44–54, 1990.

[45] R. G. Cooper. What's next?: After stage-gate. *Research-technology management*, 57(1): 20–31, 2014.

[46] C. D. Corley, L. L. Pullum, D. M. Hartley, C. Benedum, C. Noonan, P. M. Rabinowitz, and M. J. Lancaster. Disease prediction models and operational readiness. *PloS one*, 9 (3):e91989, 2014.

[47] P. Craig, P. Dieppe, S. Macintyre, S. Michie, I. Nazareth, and M. Petticrew. Developing and evaluating complex interventions: the new medical research council guidance. *Bmj*, 337, 2008.

[48] J. Crisp, D. Pelletier, C. Duffield, A. Adams, and S. Nagy. The delphi method? *Nursing research*, 46(2):116–118, 1997.

[49] F. K. Dankar and M. Ibrahim. Fake it till you make it: Guidelines for effective synthetic data generation. *Applied Sciences*, 11(5):2158, 2021.

[50] F. D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pages 319–340, 1989.

[51] E. De Cristofaro. What is synthetic data? the good, the bad, and the ugly. *arXiv preprint arXiv:2303.01230*, 2023.

[52] G. de Jong and R. K. Woolthuis. The institutional arrangements of innovation: Antecedents and performance effects of trust in high-tech alliances. *Industry and Innovation*, 15:45 – 67, 2008. doi: 10.1080/13662710701858520.

[53] J. W. de Kok, M. Á. A. de la Hoz, Y. de Jong, V. Brokke, P. W. Elbers, P. Thoral, A. Castillejo, T. Trenor, J. M. Castellano, A. E. Bronchalo, et al. A guide to sharing open healthcare data under the general data protection regulation. *Scientific Data*, 10(1):404, 2023.

[54] C. M. de Melo, A. Torralba, L. Guibas, J. DiCarlo, R. Chellappa, and J. Hodgins. Next-generation deep learning based on simulators and synthetic data. *Trends in cognitive sciences*, 2021.

[55] E. de Redon and A. Centi. Realities of conducting digital health research: Challenges to consider. *Digital Health*, 5, 2019. doi: 10.1177/2055207619869466.

[56] G. P. Dexter, S. Grannis, B. Dixon, and S. Kasthurirathne. Generalization of machine learning approaches to identify notifiable conditions from a statewide health information exchange. *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science*, 2020:152–161, 2020.

[57] A. Dinno. Nonparametric pairwise multiple comparisons in independent groups using dunn's test. *The Stata Journal*, 15(1):292–300, 2015.

[58] T. Dockhorn, T. Cao, A. Vahdat, and K. Kreis. Differentially private diffusion models. *arXiv preprint arXiv:2210.09929*, 2022.

[59] G. T. Duncan and R. W. Pearson. Enhancing access to microdata while protecting confidentiality: Prospects for the future. *Statistical Science*, 6(3):219–232, 1991.

[60] C. Dwork, A. Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[61] T. Eche, L. H. Schwartz, F.-Z. Mokrane, and L. Dercle. Toward generalizability in the deployment of artificial intelligence in radiology: role of computation stress testing to overcome underspecification. *Radiology: Artificial Intelligence*, 3(6):e210097, 2021.

[62] O. C. Edo, D. Ang, E.-E. Etu, I. Tenebe, S. Edo, and O. A. Diekola. Why do healthcare workers adopt digital health technologies-a cross-sectional study integrating the tam and utaut model in a developing economy. *International Journal of Information Management Data Insights*, 3(2):100186, 2023.

[63] H. E. Essmann. *Toward innovation capability maturity*. PhD thesis, Stellenbosch: University of Stellenbosch, 2009.

[64] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean. A guide to deep learning in healthcare. *Nature medicine*, 25(1):24–29, 2019.

[65] I. Etikan, S. A. Musa, R. S. Alkassim, et al. Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics*, 5(1):1–4, 2016.

[66] J. Fagerberg, D. C. Mowery, and R. R. Nelson. *The Oxford handbook of innovation*. Oxford university press, 2005.

[67] R. S. Firestone. Oversimplification in philosophy. *Open Journal of Philosophy*, 2019. doi: 10.4236/ojpp.2019.93025.

[68] L. M. Fleuren, P. Thoral, D. Shillan, A. Ercole, P. W. Elbers, and R. D. R. N. C. M. H. B. G. T. L. K. T. G. L. F. R. E. L. S. A. R. Girbes. Machine learning in intensive care medicine: ready for take-off? *Intensive care medicine*, 46:1486–1488, 2020.

[69] R. Frambach and N. Schillewaert. Organizational innovation adoption: a multi-level framework of determinants and opportunities for future research. *Journal of Business Research*, 55:163–176, 2002. doi: 10.1016/S0148-2963(00)00152-1.

[70] D. Gefen, E. Karahanna, and D. W. Straub. Trust and tam in online shopping: An integrated model. *MIS quarterly*, pages 51–90, 2003.

[71] J. Georges-Filteau and E. Cirillo. Synthetic observational health data with gans: from slow adoption to a boom in medical research and ultimately digital twins? *arXiv preprint arXiv:2005.13510*, 2020.

[72] B. A. Gheorghiță, L. M. Itu, P. Sharma, C. Suciu, J. Wetzl, C. Geppert, M. A. A. Ali, A. M. Lee, S. K. Piechnik, S. Neubauer, et al. Improving robustness of automatic cardiac function quantification from cine magnetic resonance imaging using synthetic image data. *Scientific Reports*, 12(1):2391, 2022.

[73] F. Gille, A. Jobin, and M. Ienca. What we talk about when we talk about trust: Theory of trust for ai in healthcare. *Intelligence-Based Medicine*, 1:100001, 2020.

[74] A. Gonzales, G. Guruswamy, and S. R. Smith. Synthetic data in health care: A narrative review. *PLOS Digital Health*, 2(1):e0000082, 2023.

[75] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks, 2014. URL https://arxiv.org/abs/1406.2661.

[76] T. Greenhalgh, G. Robert, F. Macfarlane, P. Bate, and O. Kyriakidou. Diffusion of innovations in service organizations: systematic review and recommendations. *The milbank quarterly*, 82(4):581–629, 2004.

[77] P. E. Greenwood and M. S. Nikulin. *A guide to chi-squared testing*, volume 280. John Wiley & Sons, 1996.

[78] L. Guillier. Predictive microbiology models and operational readiness. *Procedia Food Science*, 7:133–136, 2016.

[79] D. Gursoy, O. H. Chi, L. Lu, and R. Nunkoo. Consumers acceptance of artificially intelligent (ai) device use in service delivery. *International Journal of Information Management*, 49:157–169, 2019.

[80] A. Hagaman and A. Wutich. How many interviews are enough to identify metathemes in multisited and cross-cultural research? another perspective on guest, bunce, and johnson's (2006) landmark study. *Field Methods*, 29:23 – 41, 2017. doi: 10.1177/1525822X16640447.

[81] H. Haider. Barriers to the adoption of artificial intelligence in healthcare in india. Technical report, K4D helpdesk service, 2020.

[82] B. Z. Hameed, N. Naik, S. Ibrahim, N. S. Tatkar, M. J. Shah, D. Prasad, P. Hegde, P. Chlosta, B. P. Rai, and B. K. Somani. Breaking barriers: Unveiling factors influencing the adoption of artificial intelligence by healthcare providers. *Big Data and Cognitive Computing*, 7(2):105, 2023.

[83] E. Hariton and J. J. Locascio. Randomised controlled trials—the gold standard for effectiveness research. *BJOG: an international journal of obstetrics and gynaecology*, 125(13): 1716, 2018.

[84] P. J. Hart and C. Saunders. Power and trust: Critical factors in the adoption and use of electronic data interchange. *Organization Science*, 8:23–42, 1997. doi: 10.1287/ORSC.8.1.23.

[85] R. Haupt and A. J. Shockley. Whataboutery [ethically speaking]. *IEEE Antennas and Propagation Magazine*, 62:118–119, 2020. doi: 10.1109/map.2020.2983980.

[86] M. Hercheui, G. Mech, et al. Factors affecting the adoption of artificial intelligence in healthcare. *Global Journal of Business Research*, 15(1):77–88, 2021.

[87] M. Hernandez, G. Epelde, A. Alberdi, R. Cilla, and D. Rankin. Synthetic data generation for tabular health records: A systematic review. *Neurocomputing*, 2022.

[88] S. E. Hickman, G. C. Baxter, and F. J. Gilbert. Adoption of artificial intelligence in breast imaging: evaluation, ethical constraints and limitations. *British journal of cancer*, 125(1):15–22, 2021.

[89] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[90] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller. Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(4):e1312, 2019.

[91] H.-F. Hsieh and S. E. Shannon. Three approaches to qualitative content analysis. *Qualitative health research*, 15(9):1277–1288, 2005.

[92] I. Huybrechts, A. Declercq, E. Verté, P. Raeymaeckers, and S. Anthierens. The building blocks of implementation frameworks and models in primary care: a narrative review. *Frontiers in Public Health*, 9:675171, 2021.

[93] S. James, C. Harbron, J. Branson, and M. Sundler. Synthetic data use: exploring use cases to optimise data utility. *Discover Artificial Intelligence*, 1(1):15, 2021.

[94] S. Jansen-Kosterink, M. Broekhuis, and L. van Velsen. Time to act mature—gearing ehealth evaluations towards technology readiness levels. *Digital Health*, 8:20552076221113396, 2022.

[95] F. Jelinek. *Statistical methods for speech recognition*. MIT press, 1998.

[96] A. E. Johnson, M. M. Ghassemi, S. Nemati, K. E. Niehaus, D. A. Clifton, and G. D. Clifford. Machine learning and decision support in critical care. *Proceedings of the IEEE*, 104(2):444–466, 2016.

[97] J. Jordon, J. Yoon, and M. Van Der Schaar. Pate-gan: Generating synthetic data with differential privacy guarantees. In *International conference on learning representations*, 2018.

[98] J. Jordon, J. Yoon, and M. van der Schaar. Differentially private bagging: Improved utility and cheaper privacy than subsample-and-aggregate. *Advances in Neural Information Processing Systems*, 32, 2019.

[99] J. Jordon, L. Szpruch, F. Houssiau, M. Bottarelli, G. Cherubin, C. Maple, S. N. Cohen, and A. Weller. Synthetic data–what, why and how? *arXiv preprint arXiv:2205.03257*, 2022.

[100] A. Karahoca, D. Karahoca, and M. Aksöz. Examining intention to adopt to internet of things in healthcare technology products. *Kybernetes*, 47(4):742–770, 2018.

[101] S. B. Kedia, J. C. Baker, R. G. Carbonell, K. H. Lee, C. J. Roberts, J. Erickson, J. E. Schiel, K. Rogers, G. Schaefer, and S. Pluschkell. Biomanufacturing readiness levels [brl]—a shared vocabulary for biopharmaceutical technology development and commercialization. *Biotechnology and Bioengineering*, 119(12):3526–3536, 2022.

[102] S. Kelly, S.-A. Kaye, and O. Oviedo-Trespalacios. What factors contribute to acceptance of artificial intelligence? a systematic review. *Telematics and Informatics*, page 101925, 2022.

[103] A. Khanijahani, S. Iezadi, S. Dudley, M. Goettler, P. Kroetsch, and J. Wise. Organizational, professional, and patient characteristics associated with artificial intelligence adoption in healthcare: A systematic review. *Health Policy and Technology*, 11(1):100602, 2022.

[104] H.-Y. Kim. Statistical notes for clinical researchers: Chi-squared test and fisher's exact test. *Restorative dentistry & endodontics*, 42(2):152–155, 2017.

[105] S. Kim, K.-H. Lee, H. Hwang, and S. Yoo. Analysis of the factors influencing healthcare professionals' adoption of mobile electronic medical record (emr) using the unified theory of acceptance and use of technology (utaut) in a tertiary hospital. *BMC medical informatics and decision making*, 16(1):1–12, 2015.

[106] T. K. Kim. T test as a parametric statistic. *Korean journal of anesthesiology*, 68(6): 540–546, 2015.

[107] A. H. Kind and T. Twardawski. Board overconfidence in mergers & acquisitions: A self-attribution bias, 2016.

[108] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[109] P. Kipnis, B. J. Turk, D. A. Wulf, J. C. LaGuardia, V. Liu, M. M. Churpek, S. Romero-Brufau, and G. J. Escobar. Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the icu. *Journal of biomedical informatics*, 64:10–19, 2016.

[110] M. Klaic, S. Kapp, P. Hudson, W. Chapman, L. Denehy, D. Story, and J. J. Francis. Implementability of healthcare interventions: an overview of reviews and development of a conceptual framework. *Implementation Science*, 17(1):10, 2022.

[111] B. Koekkoek, B. van Meijel, A. Schene, and G. Hutschemaekers. Clinical problems in the long-term care of patients with chronic depression. *Journal of advanced nursing*, 62 6: 689–97, 2008. doi: 10.1111/j.1365-2648.2008.04645.x.

[112] Y. Kong, Y. Hou, and S. Sun. The adoption of artificial intelligence in the e-commerce trade of healthcare industry. In *Digital Health and Medical Analytics: Second International Conference, DHA 2020, Beijing, China, July 25, 2020, Revised Selected Papers 2*, pages 75–88. Springer, 2021.

[113] A. Kumar, V. Mani, V. Jain, H. Gupta, and V. Venkatesh. Managing healthcare supply chain through artificial intelligence (ai): A study of critical success factors. *Computers & Industrial Engineering*, 175:108815, 2023.

[114] S. Kurapati and L. Gilli. Synthetic data: A convergence between innovation and gdpr. *J. Open Access L.*, 11:1, 2023.

[115] S. Kvale. Ten standard objections to qualitative research interviews. *Journal of phenomenological psychology*, 25(2):147–173, 1994.

[116] S. I. Lambert, M. Madi, S. Sopka, A. Lenes, H. Stange, C.-P. Buszello, and A. Stephan. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. *NPJ Digital Medicine*, 6(1):111, 2023.

[117] J. Lamp, M. Derdzinski, C. Hannemann, J. van der Linden, L. Feng, T. Wang, and D. Evans. Glucosynth: Generating differentially-private synthetic glucose traces. *arXiv preprint arXiv:2303.01621*, 2023.

[118] A. Lavin and G. Renard. Technology readiness levels for ai & ml. *arXiv preprint arXiv:2006.12497*, 2020.

[119] A. Lavin, C. M. Gilligan-Lee, A. Visnjic, S. Ganju, D. Newman, S. Ganguly, D. Lange, A. G. Baydin, A. Sharma, A. Gibson, et al. Technology readiness levels for machine learning systems. *Nature Communications*, 13(1):6039, 2022.

[120] J. Lee, M. Kim, Y. Jeong, and Y. Ro. Differentially private normalizing flows for synthetic tabular data generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 7345–7353, 2022.

[121] Q. Li, H. Wei, C. Yu, and S. Wang. Model and data driven complex product development: from v, double vs to triple vs. In *2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS)*, pages 860–864. IEEE, 2019.

[122] X. Li, T. J. Hess, and J. S. Valacich. Why do we trust new technology? a study of initial trust formation with organizational information systems. *The Journal of Strategic Information Systems*, 17(1):39–71, 2008.

[123] Z. Liang, G. Zhang, J. X. Huang, and Q. V. Hu. Deep learning for healthcare decision making with emrs. In *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 556–559. IEEE, 2014.

[124] S. K. Lippert. Investigating postadoption utilization: An examination into the role of interorganizational and technology trust. *IEEE Transactions on Engineering Management*, 54:468–483, 2007. doi: 10.1109/TEM.2007.900792.

[125] J. Logan and I. D. Graham. Toward a comprehensive interdisciplinary model of health care research use. *Science communication*, 20(2):227–246, 1998.

[126] J. M. Long. Integration readiness levels. In *2011 Aerospace Conference*, pages 1–9. IEEE, 2011.

[127] D. Lord. Advanced technology requirements. Technical report, NASA, 1970.

[128] C. Lu, P. J. Ball, and J. Parker-Holder. Synthetic experience replay. *arXiv preprint arXiv:2303.06614*, 2023.

[129] C. V. Lukas, S. K. Holmes, A. B. Cohen, J. Restuccia, I. E. Cramer, M. Shwartz, and M. P. Charns. Transformational change in health care systems: an organizational model. *Health care management review*, 32(4):309–320, 2007.

[130] A. Magruk. Analysis of uncertainties and levels of foreknowledge in relation to major features of emerging technologies—the context of foresight research for the fourth industrial revolution. *Sustainability*, 13(17):9890, 2021.

[131] J. C. Mankins. Technology readiness assessments: A retrospective. *Acta Astronautica*, 65 (9-10):1216–1223, 2009.

[132] J. C. Mankins et al. Technology readiness levels. *White Paper, April*, 6(1995):1995, 1995.

[133] B. Marshall, P. Cardon, A. Poddar, and R. Fontenot. Does sample size matter in qualitative research?: A review of qualitative interviews in is research. *Journal of computer information systems*, 54(1):11–22, 2013.

[134] A. F. Mavrogenis and M. M. Scarlat. Artificial intelligence publications: synthetic data, patients, and papers. *International Orthopaedics*, pages 1–2, 2023.

[135] D. McDuff, T. Curran, and A. Kadambi. Synthetic data in healthcare. *arXiv preprint arXiv:2304.03243*, 2023.

[136] P. E. McKnight and J. Najab. Mann-whitney u test. *The Corsini encyclopedia of psychology*, pages 1–1, 2010.

[137] T. L. Mentzer. Response biases in multiple-choice test item files. *Educational and Psychological Measurement*, 42(2):437–448, 1982.

[138] D. Meyerson, K. E. Weick, R. M. Kramer, et al. Swift trust and temporary groups. *Trust in organizations: Frontiers of theory and research*, 166:195, 1996.

[139] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi. The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2):2053951716679679, 2016.

[140] C. Z. Mooney. *Monte carlo simulation*, volume 116. Sage, 1997.

[141] G. C. Moore and I. Benbasat. Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information systems research*, 2(3): 192–222, 1991.

[142] S. Mukherjee, V. Chittipaka, M. M. Baral, S. K. Pal, and S. Rana. Impact of artificial intelligence in the healthcare sector. *Artificial Intelligence and Industry 4.0*, pages 23–54, 2022.

[143] K. Munir, H. Elahi, A. Ayub, F. Frezza, and A. Rizzi. Cancer diagnosis using deep learning: a bibliographic review. *Cancers*, 11(9):1235, 2019.

[144] H. Murtaza, M. Ahmed, N. F. Khan, G. Murtaza, S. Zafar, and A. Bano. Synthetic data generation: State of the art in health care domain. *Computer Science Review*, 48:100546, 2023.

[145] S. Nashif, M. R. Raihan, M. R. Islam, and M. H. Imam. Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system. *World Journal of Engineering and Technology*, 6(4):854–873, 2018.

[146] K. T. Nguyen, K. Medjaher, and D. T. Tran. A review of artificial intelligence methods for engineering prognostics and health management with implementation guidelines. *Artificial Intelligence Review*, 56(4):3659–3709, 2023.

[147] S. I. Nikolenko. *Synthetic data for deep learning*, volume 174. Springer, 2021.

[148] M. Nikolova-Simons, R. Keldermann, Y. Peters, W. Compagner, L. Montenij, Y. de Jong, and R. A. Bouwman. Predictive analytics for cardio-thoracic surgery duration as a stepstone towards data-driven capacity management. *NPJ Digital Medicine*, 6(1):205, 2023.

[149] G. Norman. Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education*, 15:625–632, 2010.

[150] O. Noy, D. Coster, M. Metzger, I. Atar, S. Shenhar-Tsarfaty, S. Berliner, G. Rahav, O. Rogowski, and R. Shamir. A machine learning model for predicting deterioration of covid-19 inpatients. *Scientific reports*, 12(1):2630, 2022.

[151] A. Owoyemi, J. Owoyemi, A. Osiyemi, and A. Boyd. Artificial intelligence for healthcare in africa. *Frontiers in Digital Health*, 2:6, 2020.

[152] T. Panch, J. Pearson-Stuttard, F. Greaves, and R. Atun. Artificial intelligence: opportunities and risks for public health. *The Lancet Digital Health*, 1(1):e13–e14, 2019.

[153] A. Pereno and D. Eriksson. A multi-stakeholder perspective on sustainable healthcare: From 2030 onwards. *Futures*, 122:102605, 2020.

[154] D. Pesce, P. Neirotti, and E. Paolucci. When culture meets digital platforms: value creation and stakeholders' alignment in big data use. *Current Issues in Tourism*, 22(15): 1883–1903, 2019.

[155] Philips. About us, 11 2023. URL https://www.careers.philips.com/global/en/about-us. [Accessed on January 10th, 2024].

[156] Philips. Our strategic focus — Philips, unknown. URL https://www.philips.com/a-w/about/our-strategy.html. [Accessed on January 10th, 2024].

[157] O. S. Pianykh, S. Guitron, D. Parke, C. Zhang, P. Pandharipande, J. Brink, and D. Rosenthal. Improving healthcare operations management with machine learning. *Nature Machine Intelligence*, 2(5):266–273, 2020.

[158] P. Plsek. Complexity and the adoption of innovation in health care. *Accelerating quality improvement in health care: strategies to accelerate the diffusion of evidence-based innovations. Washington, DC: National Institute for Healthcare Management Foundation and National Committee for Quality in Health Care*, 2003.

[159] M. Poongodi, M. Hamdi, M. Malviya, A. Sharma, G. Dhiman, and S. Vimal. Diagnosis and combating covid-19 using wearable oura smart ring with deep learning methods. *Personal and ubiquitous computing*, pages 1–11, 2022.

[160] M. Radović-Marković, S. Kabir, and E. Jovičić. *Gender and technology adoption among farmers in Banglash*. WorldFish, 2020. doi: 10.5937/INTREV2003012R.

[161] W. Rahane, H. Dalvi, Y. Magar, A. Kalane, and S. Jondhale. Lung cancer detection using image processing and machine learning healthcare. In *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, pages 1–5. IEEE, 2018.

[162] B. Rahimi, H. Nadri, H. L. Afshar, and T. Timpka. A systematic review of the technology acceptance model in health informatics. *Applied clinical informatics*, 9(03):604–634, 2018.

[163] M. Rahimzadeh, A. Attar, and S. M. Sakhaei. A fully automated deep learning-based network for detecting covid-19 from a new and large lung ct scan dataset. *Biomedical Signal Processing and Control*, 68:102588, 2021.

[164] A. M. Rahmani, E. Yousefpoor, M. S. Yousefpoor, Z. Mehmood, A. Haider, M. Hosseinzadeh, and R. Ali Naqvi. Machine learning (ml) in medicine: Review, applications, and challenges. *Mathematics*, 9(22):2970, 2021.

[165] J.-F. Rajotte, R. Bergen, D. L. Buckeridge, K. El Emam, R. Ng, and E. Strome. Synthetic data as an enabler for machine learning applications in medicine. *Iscience*, 25(11), 2022.

[166] D. Rankin, M. Black, R. Bond, J. Wallace, M. Mulvenna, G. Epelde, et al. Reliability of supervised machine learning using synthetic data in health care: Model to preserve privacy for data sharing. *JMIR medical informatics*, 8(7):e18910, 2020.

[167] J. T. Rehling. Conceptualising eco-anxiety using an existential framework. *South African Journal of Psychology*, 52(4):472–485, 2022.

[168] J. P. Reiter. Using cart to generate partially synthetic public use microdata. *Journal of official statistics*, 21(3):441, 2005.

[169] G. Riva, R. M. Baños, C. Botella, B. K. Wiederhold, and A. Gaggioli. Positive technology: using interactive technologies to promote positive functioning. *Cyberpsychology, Behavior, and Social Networking*, 15(2):69–77, 2012.

[170] E. Rogers. Diffusion of innovation 3rd edition, 1983.

[171] E. M. Rogers. *Diffusion of Innovations*. Simon and Schuster, 5th edition, 8 2003.

[172] M. Rouidi, A. Hamdoune, K. Choujtani, A. Chati, et al. Tam-utaut and the acceptance of remote healthcare technologies by healthcare professionals: A systematic review. *Informatics in Medicine Unlocked*, 32:101008, 2022.

[173] J. Rowley. Conducting research interviews. *Management research review*, 35(3/4):260–271, 2012.

[174] W. W. Royce. Managing the development of large software systems: concepts and techniques. In *Proceedings of the 9th international conference on Software Engineering*, pages 328–338, 1987.

[175] D. B. Rubin. Statistical disclosure limitation. *Journal of official Statistics*, 9(2):461–468, 1993.

[176] B. Sauser, D. Verma, J. Ramirez-Marquez, and R. Gove. From trl to srl: The concept of systems readiness levels. In *Conference on Systems Engineering Research*, volume 5. Citeseer, 2006.

[177] U. Schmalz, S. Spinler, and J. Ringbeck. Lessons learned from a two-round delphi-based scenario study. *MethodsX*, 8, 2020. doi: 10.1016/j.mex.2020.101179.

[178] C. O. Seneler, N. Basoglu, and T. U. Daim. A taxonomy for technology adoption: A human computer interaction perspective. In *PICMET'08-2008 Portland International Conference on Management of Engineering & Technology*, pages 2208–2219. IEEE, 2008.

[179] Y. Setiawan, A. Atahau, and R. Robiyanto. Cognitive dissonance bias, overconfidence bias dan herding bias dalam pengambilan keputusan investasi saham. *AFRE (Accounting and Financial Review)*, 2018. doi: 10.26905/AFR.V1I1.1745.

[180] F. Shamout, T. Zhu, and D. A. Clifton. Machine learning for clinical outcome prediction. *IEEE reviews in Biomedical Engineering*, 14:116–126, 2020.

[181] S. Sharma, A. Aggarwal, and T. Choudhury. Breast cancer detection using machine learning algorithms. In *2018 International conference on computational techniques, electronics and mechanical systems (CTEMS)*, pages 114–118. IEEE, 2018.

[182] S. R. Shaw and S. Pecsi. When is the evidence sufficiently supportive of real-world application? evidence-based practices, open science, clinical readiness level. *Psychology in the Schools*, 58(10):1891–1901, 2021.

[183] A. Sherwood. How will generative AI impact healthcare?, 10 2023. URL https://www.weforum.org/agenda/2023/05/how-will-generative-ai-impact-healthcare/.

[184] D. Smith, M. Cartwright, J. Dyson, J. Hartin, and L. M. Aitken. Barriers and enablers of recognition and response to deteriorating patients in the acute hospital setting: A theory-driven interview study using the theoretical domains framework. *Journal of advanced nursing*, 77(6):2831–2844, 2021.

[185] J. D. Smith. An alternative to technology readiness levels for non-developmental item (ndi) software. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, pages 315a–315a. IEEE, 2005.

[186] E. Straub. Understanding technology adoption: Theory and future directions for informal learning. *Review of Educational Research*, 79:625 – 649, 2009. doi: 10.3102/0034654308325896.

[187] U. S. Subri, R. Rus, R. Mustapha, and Z. Hanapi. Delve into the challenges of career retention among women engineer: The application of the modified delphi technique. *International Journal of Academic Research in Business and Social Sciences*, 2019. doi: 10.6007/IJARBSS/V9-I3/5742.

[188] S. Sunarti, F. F. Rahman, M. Naufal, M. Risky, K. Febriyanto, and R. Masnina. Artificial intelligence in healthcare: opportunities and risk for future. *Gaceta Sanitaria*, 35:S67–S70, 2021.

[189] S. C. Tapia-Siles, S. Coleman, and A. Cuschieri. Current state of micro-robots/devices as substitutes for screening colonoscopy: assessment based on technology readiness levels. *Surgical endoscopy*, 30:404–413, 2016.

[190] European Commission. General data protection regulation. https://gdpr-info.eu/, 2018. [Accessed on 05-10-2022].

[191] P. J. Thoral, J. M. Peppink, R. H. Driessen, E. J. Sijbrands, E. J. Kompanje, L. Kaplan, H. Bailey, J. Kesecioglu, M. Cecconi, M. Churpek, et al. Sharing icu patient data responsibly under the society of critical care medicine/european society of intensive care medicine joint data science collaboration: the amsterdam university medical centers database (amsterdamumcdb) example. *Critical care medicine*, 49(6):e563, 2021.

[192] K. Thorlund, L. Dron, J. J. Park, and E. J. Mills. Synthetic and external controls in clinical trials–a primer for researchers. *Clinical epidemiology*, pages 457–467, 2020.

[193] E. J. Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, 25(1):44–56, 2019.

[194] L. G. Tornatzky and M. Fleischer. *The Processes of Technological Innovation*. Lexington Books, 1990.

[195] F.-C. Tung, S.-C. Chang, and C.-M. Chou. An extension of trust and tam model with idt in the adoption of the electronic logistics information system in his in the medical industry. *International journal of medical informatics*, 77(5):324–335, 2008.

[196] P. Ulrich and V. Frank. Relevance and adoption of ai technologies in german smes–results from survey-based research. *Procedia Computer Science*, 192:2152–2159, 2021.

[197] T. Vairo, M. Pettinato, A. P. Reverberi, M. F. Milazzo, and B. Fabiano. An approach towards the implementation of a reliable resilience model based on machine learning. *Process Safety and Environmental Protection*, 172:632–641, 2023.

[198] B. van Breugel and M. van der Schaar. Beyond privacy: Navigating the opportunities and challenges of synthetic data. *arXiv preprint arXiv:2304.03722*, 2023.

[199] R. D. van Hoorn, T. Bakkes, Z. Tokoutsi, Y. de Jong, R. A. Bouwman, and M. Pechenizkiy. Generating privacy-preserving longitudinal synthetic data. In *NeurIPS 2023 Workshop on Synthetic Data Generation with Generative AI*, 2023.

[200] A. Vargha and H. D. Delaney. The kruskal-wallis test and stochastic homogeneity. *Journal of Educational and behavioral Statistics*, 23(2):170–192, 1998.

[201] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[202] E. Vayena, A. Blasimme, and I. G. Cohen. Machine learning in medicine: addressing ethical challenges. *PLoS medicine*, 15(11):e1002689, 2018.

[203] V. Venkatesh and H. Bala. Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*, 39(2):273–315, 2008.

[204] V. Venkatesh and F. D. Davis. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 46(2):186–204, 2000.

[205] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis. User acceptance of information technology: Toward a unified view. *MIS quarterly*, pages 425–478, 2003.

[206] G. Visani, G. Graffi, M. Alfero, E. Bagli, F. Chesani, and D. Capuzzo. Enabling synthetic data adoption in regulated domains. In *2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 1–10. IEEE, 2022.

[207] A. Vlahou, D. Hallinan, R. Apweiler, A. Argiles, J. Beige, A. Benigni, R. Bischoff, P. C. Black, F. Boehm, J. Céraline, et al. Data sharing under the general data protection regulation: time to harmonize law and research ethics? *Hypertension*, 77(4):1029–1035, 2021.

[208] M. von Atzigen, F. Liebmann, A. Hoch, D. E. Bauer, J. G. Snedeker, M. Farshad, and P. Fürnstahl. Holoyolo: A proof-of-concept study for marker-less surgical navigation of spinal rod implants with augmented reality and on-device machine learning. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 17(1):1–10, 2021.

[209] S. L. Wang and H. I. Lin. Integrating ttf and idt to evaluate user intention of big data analytics in mobile cloud healthcare system. *Behaviour & Information Technology*, 38(9): 974–985, 2019.

[210] E. Weber. Breaking cognitive barriers to a sustainable future. *Nature Human Behaviour*, 1, 2017. doi: 10.1038/S41562-016-0013.

[211] E. H. Weissler, T. Naumann, T. Andersson, R. Ranganath, O. Elemento, Y. Luo, D. F. Freitag, J. Benoit, M. C. Hughes, F. Khan, et al. The role of machine learning in clinical research: transforming the future of evidence generation. *Trials*, 22(1):1–15, 2021.

[212] C. Wohlin. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, pages 1–10, 2014.

[213] B. Wolff, F. Mahoney, A. L. Lohiniva, and M. Corkum. Collecting and analyzing qualitative data. *The CDC Field Epidemiology Manual; Oxford University Press: Oxford, UK; New York, NY, USA*, pages 213–228, 2019.

[214] C. Wu, B. Wang, C. Zhang, R. A. Wysk, and Y.-W. Chen. Bioprinting: an assessment based on manufacturing readiness levels. *Critical reviews in biotechnology*, 37(3):333–354, 2017.

[215] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni. Modeling tabular data using conditional gan. *Advances in Neural Information Processing Systems*, 32, 2019.

[216] A. Yale, S. Dash, R. Dutta, I. Guyon, A. Pavao, and K. P. Bennett. Assessing privacy and quality of synthetic health data. In *Proceedings of the Conference on Artificial Intelligence for Data Discovery and Reuse*, pages 1–4, 2019.

[217] C. Yan, Y. Yan, Z. Wan, Z. Zhang, L. Omberg, J. Guinney, S. D. Mooney, and B. A. Malin. A multifaceted benchmarking of synthetic electronic health record generation models. *Nature Communications*, 13(1):7609, 2022.

[218] J. Yang, B. Luo, C. Zhao, and H. Zhang. Artificial intelligence healthcare service resources adoption by medical institutions based on toe framework. *Digital Health*, 8: 20552076221126034, 2022.

[219] J. Yoon, L. N. Drumright, and M. Van Der Schaar. Anonymization through data synthesis using generative adversarial networks (ads-gan). *IEEE journal of biomedical and health informatics*, 24(8):2378–2388, 2020.

[220] J. Zajac. The public hospital of the future. *Medical Journal of Australia*, 179, 2003. doi: 10.5694/j.1326-5377.2003.tb05531.x.

[221] M. A. Zayyad and M. Toycan. Factors affecting sustainable adoption of e-health technology in developing countries: an exploratory survey of nigerian hospitals from the perspective of healthcare professionals. *PeerJ*, 6, 2018. doi: 10.7717/peerj.4436.

[222] A. Zhao, G. Balakrishnan, F. Durand, J. V. Guttag, and A. V. Dalca. Data augmentation using learned transformations for one-shot medical image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8543–8553, 2019.

[223] Z.-H. Zhou. *Machine learning*. Springer Nature, 2021.

[224] I. Žliobaitė, M. Pechenizkiy, and J. Gama. An overview of concept drift applications. *Big data analysis: new algorithms for a new society*, pages 91–114, 2016.

Appendix A

Systematic Literature Review Specifications

A.1 PRISMA Specification

This section documents the specifications of the systematic literature review. The general systematic literature PRISMA scheme is visualized in Figure A.1.

Three databases were used for the systematic literature review, being SCOPUS, PUBMED and Web of Science. The search queries that were used are specified in the next section. Only one inclusion criterion was specified, namely that the article had to be written in English. The systematic literature review search was conducted on 02/Aug/2023, and the searches included articles from any period. The whole systematic literature review, including the retrieval, screening and analysis was done by the (single) author of this thesis, which increases potential bias for the screening and analysis.

Given the complete list of articles, duplicates were removed. Two rounds of analysis were used to determine whether or not to include an article in the review. The first round consisted of an analysis of the title, abstract and keywords (if present) of the article, to analyze whether it potentially featured use cases of SD. If in doubt, the article was included for further investigation in the second round. In the second round, the full text was first searched for mentions of SD, and subsequently skimmed around the results of that search for potential use cases of SD. If one or more use cases of SD were found, the article was included.

For the analysis, all articles were first sought for mentions of use cases for SD, documenting the published use cases together with the article. Subsequently, a list of all mentioned use cases was assembled, including a count of how many times specific use cases were mentioned. Highly overlapping use cases were grouped.

A.2 Search Queries

This section elaborates upon exactly how that search process was done. The databases and search queries can be found in Table A.1. The queries differ slightly per database, depending on the functionalities of the advanced search.

The SCOPUS query is the reference query on which the others are based. However, as PUBMED does not support keywords, and Web of Science does not support searching in Title, Abstract and Keywords, the queries are adapted accordingly.

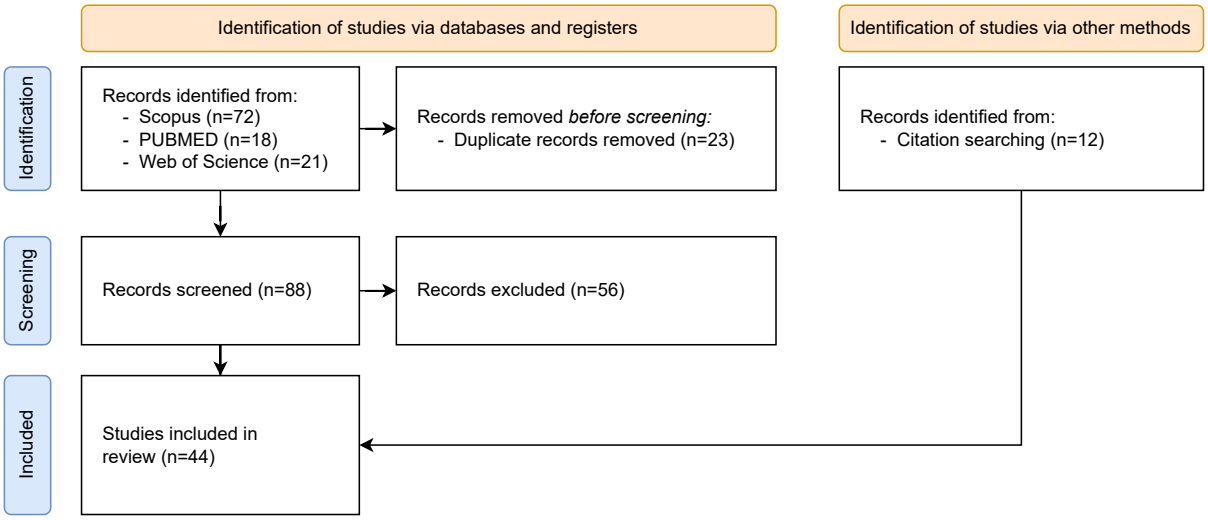


Figure A.1: A visualization of the PRISMA systematic literature review.

Database	**Search Query**
SCOPUS	(TITLE-ABS-KEY (applications OR "use-cases" OR "use cases" OR utility) AND TITLE-ABS-KEY (overview OR review) AND KEY ("Synthetic Data"))
PUBMED	(applications[Title/Abstract] OR "use-cases"[Title/Abstract] OR "use cases"[Title/Abstract] OR utility[Title/Abstract]) AND (overview[Title/Abstract] OR review[Title/Abstract]) AND "synthetic data"[Title/Abstract]
Web of Science	((AB=(applications OR "use-cases" or "use cases" or utility)) AND AB=(overview OR review)) AND AK=("synthetic data")

Table A.1: The databases and according search queries used in the systematic literature review exploring SD use cases.

Appendix B

Abbreviations

Abbreviation	Definition	Usage
SD	Synthetic Data	
AI	Artificial Intelligence	
ML	Machine Learning	
TAM	The Technology Acceptance Model	
IDT	Innovation Diffusion Theory	
TPB	Theory of Planned Behavior	
TOE	Technology-Organization-Environment (Framework)	
UTAUT	The Unified Theory of Acceptance and Use of Technology	
VP	The Value Proposition of SD	Used in Figures 5.2 & 5.3
TB	The Technical Barriers of Creating SD	Used in Figures 5.2 & 5.3
LA	The Legal Adherence when Employing SD	Used in Figures 5.2 & 5.3
AR	The Organizational and Cultural SD Adoption Requirements	Used in Figures 5.2 & 5.3
AI & DD	The Uptake of AI & Data-Driven Innovation	Used in Figures 5.2 & 5.3

Table B.1: An overview of the abbreviations used throughout the thesis

Appendix C

Interview Specification

C.1 Interview Guide

The following is the interview structure for the expert interviews. A separate Dutch translation was created for the interviews that were conducted in Dutch.

1. Introduction to interview.
    (a) Explain what is recorded, and how those recordings will be used and subsequently deleted. Recording will be started after verbal confirmation of the interviewee.
    (b) Explain that the study is seeking for their professional perspectives, not the perspectives of the companies or institutions they work at. Answers will be de-identified, and hence not be traceable back to the company or organization.
2. Background questions.
    (a) To start, can you please share a bit regarding your role and background?
    (b) Can you describe a typical project or process you've been involved with? Please share the general objective, the timeframe, and how your role contributed to its success.
    (c) (If not yet covered) How have you contributed to or been involved with healthcare innovation in your role?
    (d) Which stakeholders, both within and outside your organization, do you collaborate or interact with?
    (e) How familiar are you with the concept of synthetic data?
        i. If not familiar, explain synthetic data to the interviewee: synthetic data refers to data that is artificially generated rather than collected from real-world events. It's designed to mimic the characteristics of actual data, allowing for testing, training, and validation in environments where using real data may be problematic or restricted, such as due to privacy concerns or limited sample sizes.
    (f) (If familiar) Can you please share a bit about your experiences with synthetic data?
    (g) (If not yet covered) In what ways, if at all, have you observed the application of synthetic data within the healthcare sector?
3. Questions regarding the benefits, challenges, and risks of synthetic data.
    (a) If synthetic data were widely accepted and used in healthcare innovation, how would it impact your role or the work that you do?
    (b) (if applicable) How has your view on synthetic data evolved over time, especially as it pertains to its application in healthcare?

(c) How do you perceive the role of synthetic data in healthcare innovation, and what factors influence your acceptance of it?

(d) In your opinion, what potential advantages could synthetic data offer for healthcare innovation? (In addition) How do you hypothesize the utility of synthetic data might compare to real-world data in healthcar einnovation?

(e) What risks or challenges, if any, can you identify for the use of synthetic data in healthcare innovation?

(f) What limitations, if any, can you identify for the use of synthetic data in healthcare innovation?

(g) From your experience in healthcare innovation, do you believe synthetic data could speed up the general healthcare innovation process? Can you explain your reasoning?

(h) What criteria or standards would synthetic data need to meet for you to feel confident in utilizing it for healthcare innovation?

4. Questions to understand and explore the perspectives on processes that can benefit from the usage of SD.

   (a) To get from an initial idea towards a fully integrated and operational healthcare solution, a long process takes place. Within my research, I identify three common stages, namely (1) research - extending from the initial idea to the development of a prototype, (2) development - extending from the prototype to the development of a robust, validated and scaled-up product, ready for integration, and (3) integration & deployment - ranging from the production-ready product to the deployment and active monitoring of applicatoins in the field.

   (b) For each of these stages:

      i. In the context of this stage, which processes do you believe can benefit from leveraging synthetic data?

      ii. In the context of this stage, which processes do you believe cannot benefit from leveraging synthetic data?

5. Questions to understand their perspectives on where they think most potential lies for SD in healthcare innovation.

   (a) Considering the stages of machine learning in healthcare innovation, how would you rank them in terms of potential to benefit from synthetic data?

   (b) Can you elaborate upon your ranking?

6. Questions regarding the integration and adoption of synthetic data in healthcare innovation.

   (a) (if interviewee works in healthcare innovation) Would you consider using synthetic data in your *current* healthcare innovation projects? What key factors would influence this decision?

   (b) What do you see as the main barrier for adopting and integrating synthetic data for healthcare innovation?

   (c) How would you feel if you would, right now, start on a project that uses synthetic data?

   (d) What strategies or steps do you suggest for increasing the acceptance of synthetic data for other stakeholders and coworkers?

(e) How do you envision the optimal integration of synthetic data into the healthcare innovation workflow? Additionally, what strategies might be effective in engaging various stakeholders in this transition?

7. Closing.

(a) How would you recommend leading your organization toward embracing and incorporating synthetic data into healthcare innovation practices? Are there any specific strategies or approaches you'd advocate for?

(b) Finally, do you have any additional insights or remarks concerning healthcare innovation development or the role of synthetic data within it?

C.2 Interview Specification for Uptake Perspective

This section specifies the methodological aspects of the Uptake Perspective analysis, by providing the reasoning behind the outlined interview questions in Appendix C.1.

The questions regarding the Uptake Perspective surround the questions focusing on the Utility Perspective within the interview protocol (see Appendix C.1). There are two sections of questions regarding the Uptake Perspective included within the interview guide. The first section regards questions on the benefits, challenges, limitations and risks of SD. The second section, after the Utility Perspective analysis, focuses specifically on the integration and adoption of SD in healthcare innovation. The questions contained in these sections are informing towards these general goals.

Questions in the first section specifically asked about, for example, the perceived limitations of SD. The goal of these questions was to understand and explore the perspectives of experts on the acceptance (and adoption) of SD. To nudge the interviewees in the direction of acceptance and adoption, over simply focusing on technical aspects, the questions specifically inquire about their 'use of' or 'usage of' synthetic data. In doing so, the interviewee must form an opinion on how they would use SD, in addition to simply analysing the technical capabilities of SD. The last question aimed to summarize the perspective of the interviewee, and asks whether the interviewees believe SD could speed up the general healthcare innovation process, and to explain their stance.

After having discussed the Utility Perspective, the interview will again focus on the Uptake Perspective to ask general questions regarding whether the interviewee would consider using SD for their current healthcare innovation projects. In addition, the interviewees were asked to elaborate on what they think would help to convince other stakeholders and coworkers to accept and use SD, and how to optimally integrate SD into their workflow. These questions directly targeted their intention to use SD for healthcare innovation, and the underlying motivation for or against it.

C.3 Interview Specification for Utility Perspective

A part of the interview will explore the Utility Perspective of the expert. The expert will not be asked to specifically evaluate the conceptual framework but will be asked questions to provoke answers that are likely to attain results which can help refine and validate the conceptual framework. By approaching the Utility Perspective from a blank slate, there is less chance to induce bias in the exploration process. By understanding the perspectives of experts on where they think SD is most beneficial, the model can be adapted accordingly.

Questions pertaining to the Utility Perspective focus on two distinct goals, namely:

1. to understand and explore their perspectives on processes that can benefit from the usage of SD, focusing specifically on different phases of ML-based healthcare innovation (research, development and integration & deployment), and

2. to understand their perspectives on where they think the most potential lies to expedite the ML-based healthcare innovation process using SD.

The first goal is reflected through a series of three questions, repeated for different stages of the healthcare innovation process. The first question specifically inquired about whether they think SD can benefit that stage of innovation. The follow-up questions then dived into what processes come to mind that, according to the interviewee, can and cannot benefit from using SD. Trivially, as the interview is semi-structured, further questions were asked depending on the nature of the answers of the participant.

The second goal regarding the analysis of the Utility Perspective is reflected through a rather simple question. The question asked to rank the stages of ML-based healthcare innovation from 'most potential to benefit from SD' to 'least potential to benefit from SD'. After the ranking was been established, the interviewer explored the motivations of why the interviewee made their ranking.

Through the exploration of these goals, the interviewee established their perspectives on when and how SD can or cannot benefit the ML-based healthcare innovation process. In addition, they were asked to formulate an opinion on what processes they thought could benefit most and should potentially be highlighted in a future framework.

Appendix D

Specification of Thematic Analysis

D.1 Thematic Analysis Specification of Uptake Perspective

Although thematic analysis is often referred to as a general method of analysis, clear distinctions between various types of thematic analysis can be made. According to Braun and Clarke [32], it is essential to specify the choices made for the thematic analysis to have better scientific rigor.

The thematic analysis for the Uptake Perspective took a constructionist view. The goal of the thematic analysis for the Uptake Perspective was to identify the largest barriers, challenges and opportunities for the integration of SD in healthcare innovation. Consequently, the thematic analysis aimed to provide a rich thematic description of the expert interview data, such that the most important themes were elucidated. Themes were sought in a 'bottom-up' fashion through inductive analysis. As such, themes were not grounded or naturally linked to the theoretical topics that were discussed in Chapter 2. After the themes were identified, the most predominant themes were linked to literature through hypotheses or conclusions.

The nature of this thematic analysis is hence also why the interview questions regarding the Uptake Perspective did not directly include questions such as "do you think synthetic data is easy to use". Although this could translate to a more direct application of a framework such as TAM or UTAUT, it should be noted that it was not the goal of this perspective analysis to apply UTAUT to the insights of experts. Rather, it was to explore the differing perspectives of experts and let that guide the creation of a framework.

Appendix E

Survey

This appendix provides the complete survey as presented to the research participants.

Questionnaire on the Acceptance, Adoption and Integration of Synthetic Data

Dear Participant,

Thank you for your continued involvement in our study on the acceptance, adoption, and integration of synthetic data for innovation in healthcare. This questionnaire, an extension of our earlier interviews, aims to refine and validate our findings from the interviews. Your insights will shape the frameworks and conclusions we draw, impacting the guidance we provide to institutions and companies regarding synthetic data in healthcare innovation.

Estimated Time and Format:

- Duration: Approximately 20 minutes.
- Format: A mix of multiple-choice, and open-ended questions. There will be space at the end of every section, and at the end of the questionnaire, to provide additional details and nuance if you wish to do so.

Confidentiality and Use of Data:

- To prevent repeated reminder emails, the first question requests your first name (optional). This information will be detached from your responses and deleted post-closure of the questionnaire.
- As with the interviews, your responses will remain anonymous in any publications or presentations.

Next Steps:

- Post-deadline, we will analyze responses, draw final conclusions, and integrate these into a master thesis, anticipated for defense around January or February. You will receive a copy of the completed thesis, which will include a summary of the results.

Your participation and insights are incredibly valuable to us, and we sincerely appreciate your time and effort in aiding this research. If you have any questions or need assistance, please feel free to contact me. Thank you once again for your invaluable contribution to this study.

Best regards,

Robin van Hoorn

General Questions

1. What is your first name? (Optional)

2. What stage of the healthcare innovation process do you associate yourself most with:*

   O Research

   O Application Development

   O Application Integration & Deployment

   O Other

3. How familiar are you with synthetic data?*

   O I have heard that synthetic data can be used for healthcare innovation, but I have not looked into it in detail.

   O I am quite familiar with synthetic data for healthcare innovation.

   O I have actively researched or developed synthetic data for healthcare innovation myself.

   O Other

Use of Synthetic Data for Healthcare Innovation

4. Please indicate to what extent you agree or disagree that employing synthetic data can enhance the following **research*** processes:

   **The research phase of healthcare innovation for machine learning applications extends from the initial idea to the development of a prototype.*

   The following provides more explanation on the processes:

   (a) **Data Exploration** - When real data is unavailable, synthetic data can be used to get a feel for the real data and analyse the characteristics of the dataset.

   (b) **Comparison of ML Models or Methods** - When the real data is unavailable, synthetic data can enable the analysis of which Machine Learning (ML) model or methodology leads to a better performance for a downstream task (i.e. a prediction/classification/other model).

   (c) **Hypothesis Testing** - When real data is unavailable, synthetic data can be used to prove or disprove hypotheses.

   (d) **Publishing More Replicable Research** - Usually, healthcare data cannot be shared when publishing research. When synthetic data is used for research, it can be published alongside the research publication, thereby enabling better reproducibility.

   Please indicate to what extent you agree or disagree that employing synthetic data can enhance the following **research*** processes:

	Strongly Disagree	Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Agreee	Strongly Agree
Data Exploration	O	O	O	O	O	O	O
Comparison of ML Models or Methods	O	O	O	O	O	O	O
Hypothesis Testing	O	O	O	O	O	O	O
Publishing More Replicable Research	O	O	O	O	O	O	O

5. Please indicate to what extent you agree or disagree that employing synthetic data can enhance the following **development*** processes:

   **The development phase of healthcare innovation for machine learning applications extends from the proof-of-concept to the development of a robust, validated and scaled-up product, ready for integration.*

   The following provides more explanation of the processes:

   (a) **ML Model Training** - Synthetic data can benefit the training of machine learning (ML) models by allowing for easier access to (additional) useful data, leading to ML models that are more robust or have equal or better performance than models trained just on the accessible real data.

(b) **Software Testing** (unit testing, integration testing, load testing, etc) - Synthetic data can benefit software testing by allowing for easier access to large amounts of (useful) test data.

(c) **Robustness Testing** - Synthetic data can benefit robustness testing by allowing for easier access to diverse data.

(d) **Regulatory Certification Processes** - Synthetic data can benefit regulatory processes, for example by allowing to more easily show explainability, robustness and safety, and/or having access to more data.

(e) **User Testing** - Synthetic data can create value for usability/user testing by allowing for more representative test scenarios for user to interact with.

Please indicate to what extent you agree or disagree that employing synthetic data can enhance the following **development*** processes:

	Strongly Disagree	Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Agreee	Strongly Agree
ML Model Training	O	O	O	O	O	O	O
Software Testing	O	O	O	O	O	O	O
Robustness Testing	O	O	O	O	O	O	O
Regulatory Certification Processes	O	O	O	O	O	O	O
User Testing	O	O	O	O	O	O	O

6. Please indicate to what extent you agree or disagree that employing synthetic data can enhance the following **integration & deployment*** processes:

**The integration & deployment phase of healthcare innovation for machine learning applications ranges from the production-ready product to the deployment and active monitoring of applications in the field.*

The following provides more explanation of the processes:

(a) **Demonstrations** - Demonstrations using synthetic data can be of more value as they can use representative data without potentially compromising privacy of patients.

(b) **Employee Training** - Synthetic data can aid in training employees by allowing the use of representative data, whilst not compromising the privacy of patients.

(c) **Software and Model Testing** - The testing of new software within hospitals (including those containing machine learning applications) is often done on patient data. Synthetic data can help minimize unnecessary use of patient data.

(d) **Post Market Surveillance** - Synthetic data can benefit post market surveillance, for example by enabling easier performance monitoring, safety analysis, or model/data drift.

Please indicate to what extent you agree or disagree that employing synthetic data can enhance the following **integration & deployment*** processes:

	Strongly Disagree	Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Agreee	Strongly Agree
Demonstrations	O	O	O	O	O	O	O
Employee Training	O	O	O	O	O	O	O
Software and Model Testing	O	O	O	O	O	O	O
Post Market Surveillance	O	O	O	O	O	O	O

7. Zooming out from the specific processes within the 3 main stages of healthcare innovation, please indicate to what extent you agree that employing synthetic data can enhance these stages of healthcare innovation:

	Strongly Disagree	Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Agreee	Strongly Agree
Research	O	O	O	O	O	O	O
Development	O	O	O	O	O	O	O
Integration & Deployment	O	O	O	O	O	O	O

8. If you have any remarks or nuance to add regarding the above questions, you can use the following space.

Acceptance and Adoption of Synthetic Data

9. Which of the following aspects do you think are essential for the **acceptance*** of synthetic data for healthcare innovation?

   **The initial approval or agreement to use synthetic data by the members of an organization. It constitutes a positive reception of synthetic data and willingness to start using it.*

   Please select at most 3 options

   - ☐ **Acceptance and Adoption of AI & Data-Driven Innovation** - The willingness and readiness of stakeholders to incorporate AI and data-driven technologies into their healthcare practices, including their general inclination to pursue AI & data-driven innovation and their current use and trust of AI & data-driven applications. This relates to embracing AI & data-driven innovation in general, not just synthetic data, and thus covers a much broader scope.
   - ☐ **The Value Proposition of Synthetic Data** - The perceived benefits and advantages synthetic data provides to improve or accelerate healthcare innovation development, including potential cost savings, possibilities for better patient privacy preservation, increased collaborative data availability, and the potential for dataset augmentation (scalability/diversity).
   - ☐ **Legal Adherence** - Concerns the compliance of synthetic data with relevant laws, regulations, and ethical standards in the healthcare industry, particularly those regarding patient privacy and medical device regulations.
   - ☐ **Technical Barriers of Synthetic Data** - The challenges and impediments related to the technological capabilities and limitations that an organization might face when implementing synthetic data solutions for healthcare innovation, including the capabilities of synthetic data to create representative datasets and the ability to assess the extent to which synthetic datasets are representative, privacy-preserving or biased.
   - ☐ **Organizational and Cultural Adoption Requirements for Synthetic Data** - The changes in organizational structure, processes, and culture that are necessary for the effective adoption of synthetic data for healthcare innovation, including necessary FTEs, monetary investments, strategy planning, management communication etc.
   - ☐ **Other** _____

10. Which of the following aspects do you think are essential for the **adoption*** of synthetic data for healthcare innovation?

    **The actual uptake and consistent use of synthetic data by an organization and its members. It goes beyond acceptance, representing the action of integrating synthetic data into the standard innovation process.*

    Please select at most 3 options

    - ☐ **Acceptance and Adoption of AI & Data-Driven Innovation** - The willingness and readiness of stakeholders to incorporate AI and data-driven technologies into their healthcare practices, including their general inclination to pursue AI & data-driven innovation and their current use and trust of AI & data-driven applications.

This relates to embracing AI & data-driven innovation in general, not just synthetic data, and thus covers a much broader scope.

- ☐ **The Value Proposition of Synthetic Data** - The perceived benefits and advantages synthetic data provides to improve or accelerate healthcare innovation development, including potential cost savings, possibilities for better patient privacy preservation, increased collaborative data availability, and the potential for dataset augmentation (scalability/diversity).
- ☐ **Legal Adherence** - Concerns the compliance of synthetic data with relevant laws, regulations, and ethical standards in the healthcare industry, particularly those regarding patient privacy and medical device regulations.
- ☐ **Technical Barriers of Synthetic Data** - The challenges and impediments related to the technological capabilities and limitations that an organization might face when implementing synthetic data solutions for healthcare innovation, including the capabilities of synthetic data to create representative datasets and the ability to assess the extent to which synthetic datasets are representative, privacy-preserving or biased.
- ☐ **Organizational and Cultural Adoption Requirements for Synthetic Data** - The changes in organizational structure, processes, and culture that are necessary for the effective adoption of synthetic data for healthcare innovation, including necessary FTEs, monetary investments, strategy planning, management communication etc.
- ☐ **Other** ☐

11. Please indicate for the following themes to what extent you agree or disagree that obtaining **stakeholder alignment*** is important for the acceptance and adoption of synthetic data for healthcare innovation.

    **Stakeholder alignment is the congruence between the goals, expectations, and interests of different stakeholders that you deem important for the acceptance and adoption of synthetic data for healthcare innovation. For a hospital, 'stakeholders' might only include hospital staff. For a consortium, 'stakeholders' might mean multiple different parties and organizations.*

    The following provides more information on the themes:

    - **Acceptance and Adoption of AI & Data-Driven Innovation** - The willingness and readiness of stakeholders to incorporate AI and data-driven technologies into their healthcare practices, including their general inclination to pursue AI & data-driven innovation and their current use and trust of AI & data-driven applications. This relates to embracing AI & data-driven innovation in general, not just synthetic data, and thus covers a much broader scope.
    - **The Value Proposition of Synthetic Data** - The perceived benefits and advantages synthetic data provides to improve or accelerate healthcare innovation development, including potential cost savings, possibilities for better patient privacy preservation, increased collaborative data availability, and the potential for dataset augmentation (scalability/diversity).
    - **Legal Adherence** - Concerns the compliance of synthetic data with relevant laws, regulations, and ethical standards in the healthcare industry, particularly those regarding patient privacy and medical device regulations.

- **Technical Barriers of Synthetic Data** - The challenges and impediments related to the technological capabilities and limitations that an organization might face when implementing synthetic data solutions for healthcare innovation, including the capabilities of synthetic data to create representative datasets and the ability to assess the extent to which synthetic datasets are representative, privacy-preserving or biased.
- **Organizational and Cultural Adoption Requirements for Synthetic Data** - The changes in organizational structure, processes, and culture that are necessary for the effective adoption of synthetic data for healthcare innovation, including necessary FTEs, monetary investments, strategy planning, management communication etc.

Please indicate for the following themes to what extent you agree or disagree that obtaining **stakeholder alignment*** is important for the acceptance and adoption of synthetic data for healthcare innovation:

	Strongly Disagree	Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Agreee	Strongly Agree
Acceptance and Adoption of AI & Data-Driven Innovation	O	O	O	O	O	O	O
The Value Proposition of Synthetic Data	O	O	O	O	O	O	O
Legal Adherence	O	O	O	O	O	O	O
Technical Barriers of Synthetic Data	O	O	O	O	O	O	O
Organizational and Cultural Adoption Requirements for Synthetic Data	O	O	O	O	O	O	O

12. If you have any remarks or nuance to add regarding the above questions, you can use the following space.

Final Thoughts

13. In this last section, we want to provide you with an opportunity to provide any remaining thoughts, additions, nuances or concerns.

Appendix F

Statistical Test Results

This appendix reports on all test statistics not mentioned throughout the thesis.

F.1 Statistical Test Results of the Utility Perspective

Use Case	Statistic	P-value
Data Exploration	0.32	0.754
Comparison of ML Models or Methods	−0.58	0.568
Hypothesis Testing	−2.19	0.042*
Publishing More Replicable Research	0.16	0.877
ML Model Training	0.57	0.577
Software Testing	0.03	0.980
Robustness Testing	−1.12	0.278
Regulatory Certification Processes	−0.04	0.972
User Testing	0.45	0.657
Demonstrations	−0.46	0.657
Employee Training	−0.16	0.872
Software and Model Testing	−1.44	0.166
Post Market Surveillance	−0.86	0.402
Research	−0.69	0.501
Development	−0.75	0.462
Integration & Deployment	0.17	0.864

Table F.1: T-test statistics between 'research' and 'not research' for all use cases and healthcare innovation stages of the Utility Perspective.

Use Case	Statistic	P-value
Data Exploration	48.0	0.935
Comparison of ML Models or Methods	43.5	0.662
Hypothesis Testing	24.0	0.054
Publishing More Replicable Research	46.0	0.813
ML Model Training	55.5	0.669
Software Testing	46.5	0.873
Robustness Testing	34.5	0.250
Regulatory Certification Processes	50.0	1.000
User Testing	55.5	0.660
Demonstrations	45.0	0.736
Employee Training	42.0	0.557
Software and Model Testing	31.5	0.173
Post Market Surveillance	37.5	0.377
Research	45.5	0.779
Development	35.0	0.251
Integration & Deployment	50.5	0.969

Table F.2: Mann-Whitney U test statistics between 'research' and 'not research' for all use cases and healthcare innovation stages of the Utility Perspective.

Use Case	Statistic	P-value
Data Exploration	2.01	0.366
Comparison of ML Models or Methods	5.08	0.079
Hypothesis Testing	4.58	0.101
Publishing More Replicable Research	6.88	0.032^{*}
ML Model Training	6.72	0.035^{*}
Software Testing	2.20	0.333
Robustness Testing	7.79	0.020^{*}
Regulatory Certification Processes	9.58	0.008^{**}
User Testing	1.34	0.513
Demonstrations	1.15	0.564
Employee Training	4.81	0.090
Software and Model Testing	6.59	0.037^{*}
Post Market Surveillance	1.99	0.369
Research	4.75	0.093
Development	4.68	0.096
Integration & Deployment	6.80	0.033^{*}

Table F.3: Kruskal-Wallis Test statistics between levels of familiarity for all use cases and healthcare innovation stages of the Utility Perspective.

F.2 Statistical Test Results of the Uptake Perspective

The following tables detail the test statistics that resulted from the analysis of the Uptake Perspective.

Use Case	Statistic	P-value
The Uptake of AI & Data-Driven Innovation	1.18	0.254
The Value Proposition of SD	−0.11	0.913
The Legal Adherence when Employing SD	−1.5	0.153
Technical barriers of SD	−0.23	0.817
Organizational and Cultural Adoption Requirements for SD	−0.48	0.636

Table F.4: T-test statistics between 'research' and 'not research' for the answers to question 11 of the survey for all themes of the Uptake Perspective.

Use Case	Statistic	P-value
The Uptake of AI & Data-Driven Innovation	56.0	0.306
The Value Proposition of SD	42.5	0.913
The Legal Adherence when Employing SD	25.5	0.127
Technical barriers of SD	42.0	0.898
Organizational and Cultural Adoption Requirements for SD	35.0	0.450

Table F.5: Mann-Whitney U Test statistics between 'research' and 'not research' for the answers to question 11 of the survey for all themes of the Uptake Perspective.

Use Case	Statistic	P-value
The Uptake of AI & Data-Driven Innovation	2.10	0.35
The Value Proposition of SD	6.56	0.038^*
The Legal Adherence when Employing SD	9.11	0.011^*
Technical barriers of SD	2.68	0.262
Organizational and Cultural Adoption Requirements for SD	6.95	0.031^*

Table F.6: Kruskal-Wallis Test statistics between levels of familiarity for the answers to question 11 of the survey for all themes of the Uptake Perspective.

Appendix G

Frequency Distributions

The following tables present the frequency distributions of questions 9 and 10 from the follow-up survey. These questions are part of the analysis of the Uptake Perspective. The tables make use of abbreviations of the various themes: (1) AI & DD - the uptake of AI & data-driven innovation; (2) VP - the value proposition of SD; (3) LA - legal adherence when employing SD; (4) TB - technical barriers to creation SD; (5) AR - organizational and cultural SD adoption requirements.

Value	**Frequency**	**Cumulative Frequency**	**Percentage**	**Cumulative Percentage**
AI & DD	8	8	16.7	16.7
VP	13	21	27.1	43.8
LA	8	29	16.7	60.4
TB	12	41	25.0	85.4
AR	7	48	14.6	100.0

Table G.1: Frequency Distribution Table for question 9 of the survey regarding the acceptance of SD for healthcare innovation.

Value	**Frequency**	**Cumulative Frequency**	**Percentage**	**Cumulative Percentage**
AI & DD	10	10	22.7	22.7
VP	8	18	18.2	40.9
LA	7	25	15.9	56.8
TB	9	34	20.5	77.4
AR	10	44	22.7	100.0

Table G.2: Frequency Distribution Table for question 10 of the survey regarding the adoption of SD for healthcare innovation.

Value 1	Value 2	Frequency	Cumulative Frequency	Percentage	Cumulative Percentage
AI & DD	VP	4	4	10	10.0
AI & DD	LA	2	6	5	15.0
AI & DD	TB	4	10	10	25.0
AI & DD	AR	3	13	7.5	32.5
VP	LA	5	18	12.5	45.0
VP	TB	8	26	20	65.0
VP	AR	3	29	7.5	72.5
LA	TB	6	35	15	87.5
LA	AR	1	36	2.5	90.0
TB	AR	4	40	10	100.0

Table G.3: Frequency Distribution Table for answer combinations for question 9 of the survey regarding the acceptance of SD for healthcare innovation.

Value 1	Value 2	Frequency	Cumulative Frequency	Percentage	Cumulative Percentage
AI & DD	VP	4	4	11.8	11.8
AI & DD	LA	2	6	5.9	17.6
AI & DD	TB	3	9	8.8	26.5
AI & DD	AR	5	14	14.7	41.2
VP	LA	3	17	8.8	50.0
VP	TB	4	21	11.8	61.8
VP	AR	3	24	8.8	70.6
LA	TB	4	28	11.8	82.4
LA	AR	4	32	11.8	94.1
TB	AR	2	34	5.9	100.0

Table G.4: Frequency Distribution Table for answer combinations for question 10 of the survey regarding the adoption of SD for healthcare innovation.

Appendix H

List of Taxonomy Items

Table H.1 presents a complete overview of all items present in the conceptual taxonomy (see Figure 4.1). It spans several landscape pages for improved readability.

Table H.1: Explanations of the individual taxonomy items, including the references on which these items are based and an explanation for the SD classification per item.

Item Name	**Explanation**	**SD Classification and Reason**
Sharing Benchmark Data [74, 93]	Sharing high-quality datasets as benchmark data can help other researchers validate their research.	(1) Sharing privacy-preserving synthetic datasets should be easier as it is not upheld to the same privacy regulations.
Publishing Results [74, 93]	Publishing the results of research helps others estimate the merit of an idea.	(1) Publishing in-depth results might be easier when the research was performed with SD as privacy regulations are not a problem.
Hypothesis Testing [74, 93]	Assess the plausibility of a hypothesis by using sample data.	(2) SD might accelerate the process of getting access to the data needed for testing the hypothesis. However, only tests on real data can provide definitive proof of a finding.
Data Exploration and Understanding [93]	Determining the patterns and trends in a dataset.	(1) SD might accelerate the process of getting access to the data that needs to be explored.
Reproducing Research [93]	A process in which the results of some published research is reproduced by other people.	(1) If the data in the original study was SD, it might be much easier to gain access to the same data, and hence accelerate the reproduction process.
Method Testing [74, 93]	The process of assessing whether a method has merit as a potential solution.	(1) By having faster access to SD, method testing can be performed earlier.
Fairness and Equity Testing [135]	The process of assessing whether a solution is fair and equitable, i.e. whether a solution treats samples without bias or discrimination.	(2) By having more disseminated and quicker access to SD, fairness and equity testing can be executed better. However, the fairness and equity of the final solution on real data should also be thoroughly assessed.
Treatment Response Measurement [83]	In clinical settings, how humans respond to treatments is key to understanding if that treatment is working as intended.	(3) Human response on a novel treatment (which has not been tested before) cannot be synthesized as there is no representative benchmark.

Continued on next page

Table H.1: Continuation of explanations of the individual taxonomy items

Item Name	Explanation	SD Classification and Reason
Randomized Controlled Trial (RCT) Control Group [192]	For every randomized controlled trial, a treatment group is compared to a control group. This control group is given either a placebo or normal care.	(1) The control group can be completely synthetic in the case where a real control group would have otherwise received normal care. With SD, no candidates need to be recruited and hence the process can be accelerated.
Production-Level Model Training and Testing	The process of training and testing ML-based solutions for the purpose of being used in a production environment.	(2) A model can be pre-trained on SD and fine-tuned on real data. Real data will most likely be necessary for attaining the best quality.
Integration Testing [74, 93, 135]	The second level of software testing, that tests whether individual components of a solution work together as designed.	(1) SD can be used as test (user) data, allowing for easier code sharing and more diverse tests.
Blue-Green deployment Testing [74, 93, 135]	Blue/green deployment is a deployment strategy where two separate environments (new and old) are active at the same time and users are switched over slowly. Blue-green deployment testing is testing to elucidate faults/defects in that strategy.	(1) SD can be used as test (user) data, allowing for easier code sharing and more diverse tests.
A/B Testing [74, 93, 135]	A methodology for testing two versions of a model/application to understand which performs better.	(2) SD can be used as a proper benchmark dataset to easily see which one performs better. For production models, testing on a real dataset might be necessary.
Unit Testing [74, 93, 135]	A process in which the smallest parts of an application, called units, are extensively tested to assess whether they work as intended.	(1) SD can be used as test (user) data, allowing for easier code sharing and more diverse tests.
Safety Testing [74, 93, 135]	The process of evaluating the reliability and potential risks of a solution.	(1) SD can be used as test (user) data, allowing for easier code sharing and more diverse tests.

Continued on next page

Table H.1: Continuation of explanations of the individual taxonomy items

Item Name	**Explanation**	**SD Classification and Reason**
UX/UI Testing [94]	The process of evaluating the user interface and user experience to ensure these solutions are intuitive, user-friendly, accessible, and meet the needs and expectations of healthcare professionals.	(2) SD can be used as example data for testing the UX/UI. However, for getting the opinion on the UX/UI, testing with end-users is necessary.
Human-Acceptance Testing [110]	The process of evaluating how well end-users accept and adapt to the solution.	(2) SD can be used as example data for the human-acceptance testing. However, end-users are necessary for obtaining their opinions.
Canary Testing [118, 119, 35]	The process of testing a new feature/ version with real users in a live environment.	(3) Because the testing occurs in a live environment, no SD can be used.
Educational Training [74, 217]	The process of teaching end-users how to effectively and safely use a new solution.	(1) SD can be used as example data for the educational training process.
Explainable AI [135]	The process of making AI-based solutions transparent and understandable. A necessary process for healthcare applications.	(1) SD can be used as example data for explaining the decision making process of an AI-based solution for regulatory purposes. These samples and insights can be shared without impairing privacy.
Demonstrations [119]	Practical, possibly hands-on, presentations on how a tool/application functions.	(1) SD can be used as example data for the demonstrations.

Appendix I

HTRL creation

For the creation of the HTRL framework (see Table 4.2), seven different healthcare-related TRLs were analysed. This analysis was done in Excel and is hence difficult to translate onto A4 without becoming highly unreadable. However, the publications used for the creation of the HTRL are the following: [46, 68, 78, 94, 101, 182, 189].

For the creation of the HTRL, all frameworks were aligned next to each other. Each of the frameworks was qualitatively assessed on its quality, and given an indicative color. Next, the various indicated phases were colorized. Lastly, an aggregate was created which is presented in 4.2.

The Excel Spreadsheet document is provided as additional material. The first sheet of the document provides the details of the seven different frameworks aligned next to each other, as well as the HTRL framework written out. The second sheet of the document provides both the MLTRL and the HTRL to provide additional details on the activities contained within each of the TRLs.

Appendix J

Larger Figures

This appendix provides figures in landscape mode, for if readers of paper versions are in need of a larger image. As the visualizations are vector images, zooming in on the smaller images throughout the thesis is easily possible. However, for the reading of physical versions, these larger images are supplemented for readability.

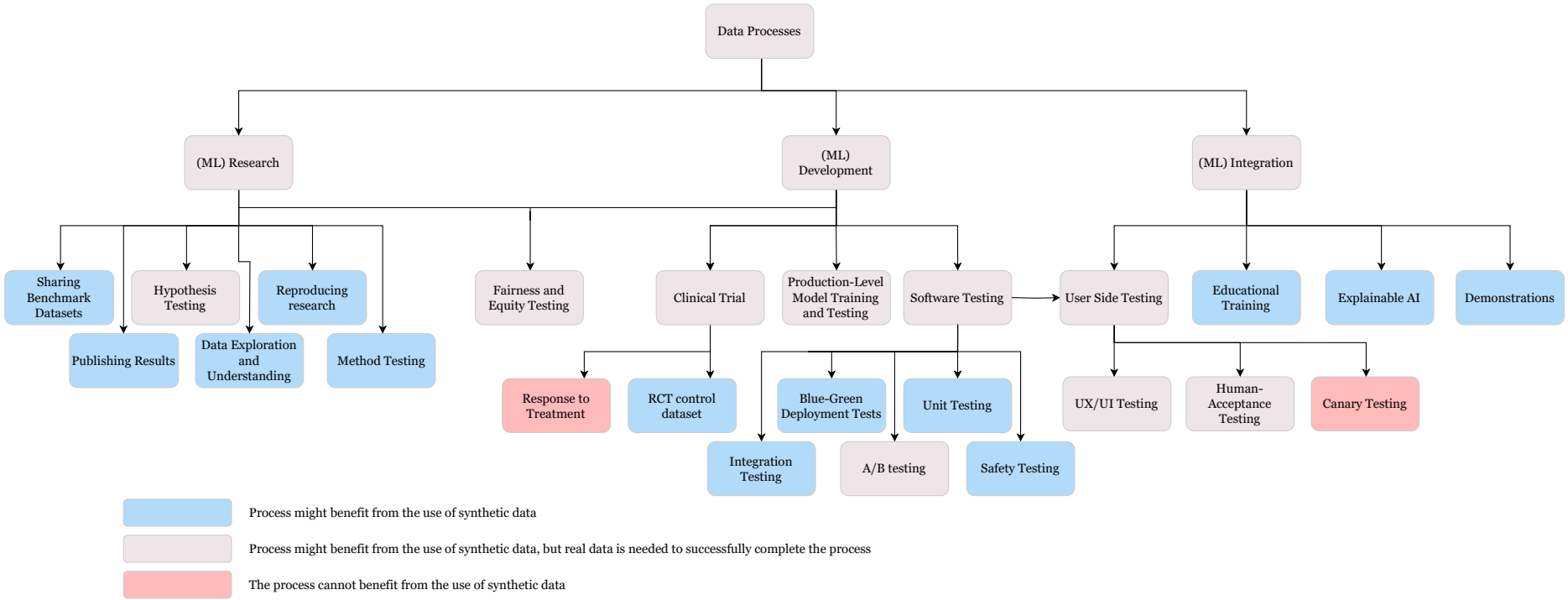
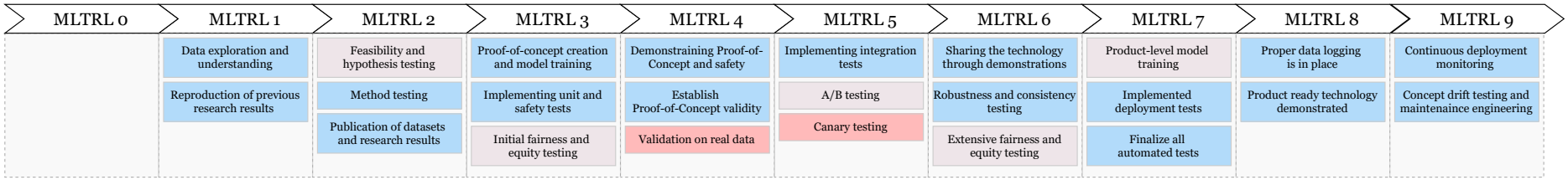
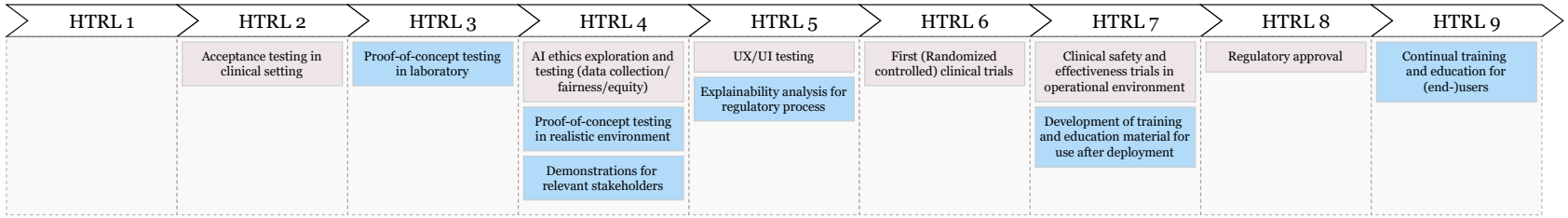


Figure J.1: Conceptual taxonomy model of data processes that are part of the healthcare innovation development process for machine learning-based solutions. The colors of the various items in the taxonomy represent their potential to benefit from SD. Higher-level items are colored based on the lower-level items.



(a) The MLTRL-SD conceptual model



(b) The HTRL-SD conceptual model

Figure J.2: The conceptual TRL model, presenting ways in which SD can benefit the healthcare innovation development process throughout different stages of the (a) MLTRL framework and (b) HTRL framework. The arrows at the top indicate different stages of the TRL frameworks, whereas the boxes, aligned with these different stages, represent different processes in which patient data is used, and can potentially benefit from the use of synthetic data. Color-coding follows that from the conceptual taxonomy model in Figure 4.1.