The DATA BASE for Advances in Information Systems

SynthEthics: Ensuring Digital Ethics and Performance with a Design Theory for Using Synthetic Image Data in Digital Health Deep Learning

Martin Böhmer
Martin Luther University Halle-Wittenberg

Stephan Kuehnel
Martin Luther University Halle-Wittenberg

Johannes Damarowsky
Martin Luther University Halle-Wittenberg

Alfred Benedikt Brendel
Indiana University

Date of Acceptance: 1/26/2025

SynthEthics: Ensuring Digital Ethics and Performance with a Design Theory for Using Synthetic Image Data in Digital Health Deep Learning

Martin Böhmer
Martin Luther University Halle-Wittenberg

Stephan Kuehnel
Martin Luther University Halle-Wittenberg

Johannes Damarowsky
Martin Luther University Halle-Wittenberg

Alfred Benedikt Brendel
Indiana University

Abstract

This paper addresses the need for ethical and effective use of synthetic image data in digital health computer vision. It explores the design requirements and design principles for both responsible use of artificial intelligence in digital health and model robustness, focusing on privacy, ethical compliance, and domain adaptation. Using the design science research paradigm along with value-sensitive design and sociotechnical systems theory, this study presents a design theory that provides actionable guidance for the generation, selection, and integration of synthetic data in digital health. Through heuristic theorizing over two design cycles, the work provides a robust theory artifact and conceptual model to ensure ethical use and improve model performance in digital health through appropriate domain adaptation, generalization, and accuracy. In addition to contributing to theoretical knowledge, this research offers practical implications for health authorities to promote ethical standards and performance in synthetically trained AI applications.

Keywords: Artificial Intelligence; Synthetic Data; Deep Learning; Design Theory; Digital Ethics.

Introduction

Deep learning models for computer vision, which typically contain millions to billions of parameters, rely on extensive datasets to achieve optimal performance and generalizability (Alzubaidi et al., 2021). Nevertheless, the process of collecting authentic training datasets for artificial intelligence (AI) systems is often expensive, prone to errors, and can result in limited or unbalanced datasets (Hinterstoisser et al., 2019; Seib et al., 2020; Zhang et al., 2018). Recently, synthetic image data (artificially generated images) has emerged as a promising alternative, offering scalability, feasibility, and fairly accurate deep learning models (Bird et al., 2020; Hinterstoisser et al., 2019; Krump et al., 2020). In particular, the digital health domain benefits from synthetic data in terms of tailored and innovative applications (Giuffrè & Shung, 2023; Gonzales et al., 2023; Murtaza et al., 2023), but it also raises various ethical and privacy concerns, as sensitive data in AI (e.g., how body parts or environments are modeled) has a high risk of misrepresentation, discrimination, intrusiveness, and bias (Kern et al., 2022; McBride, 2014; Obermeyer et al., 2019; Russel & Norvig, 2021). These concerns are compounded by the fact that while synthetic data is artificial, it is often rooted in real-world datasets through the real-world application settings for which the models are used, which may inadvertently retain identifiable or sensitive characteristics, posing significant privacy risks. While it may seem counterintuitive given the artificial nature of synthetic data, the reliance on real-world references to create such data underscores the importance of proactively addressing these ethical and privacy concerns. However, even if it appears less risky due to its artificial nature, synthetic data does not entirely eliminate privacy risks or ethical challenges; patterns from the source or reference data can be reverse-engineered or misused, and its artificiality can lead to unanticipated harms, such as skewing demographic or clinical representations. In particular, the digital health domain inherently involves sensitive information and human interactions, making it critical to design technologies that are consistent with ethical standards and user values (Vayena et al., 2018; Rajkomar et al., 2018). Therefore, it is imperative that researchers and practitioners in deep learning and digital health carefully consider these ethical implications. Effective strategies must be developed to mitigate risks, including implementing rigorous standards for data anonymization, applying fairness metrics to ensure unbiased AI models, and establishing ethical guidelines for the use and distribution of synthetic image data (Obermeyer et al., 2019; Panch et al., 2019; Raji et al., 2020; Rajkomar et al., 2019; Zhang et al., 2018).

However, while there is considerable meta-level guiding knowledge on ethical research in IS and Design Science Research (DSR) (Myers & Venable, 2014; Herwix et al., 2022), we found that the ethical and effective use of synthetic image data in digital health computer vision in particular is underexplored, leaving both researchers and practitioners navigating a complex ethical AI landscape without a clear compass. We address this gap by responding to the call for value systems (Herwix et al., 2022) by focusing on the specific ethical challenges of synthetic image data in digital health beyond existing meta-level ethical frameworks in IS and DSR. Beyond meta-level ethical guidance, the lack of consensus on standards and guidance for the use of synthetic data in digital health further exacerbates the problem, making it difficult to ensure that the benefits of AI are harnessed without compromising patient privacy or reinforcing existing biases. This gap is significant because it hinders the ability to systematically address the challenges and realize the potential of synthetic data in ways that are both ethically responsible and technologically effective (Panch et al., 2018). Thus, this paper poses the following research question:

> **RQ:** *What are the design requirements and design principles to ethically and effectively utilize synthetic image data in digital health computer vision environments, considering potential tensions between ethical compliance and operational effectiveness?*

We answer this question by adopting a multi-method heuristic design theorizing approach (Gregory & Muntermann, 2014) that combines qualitative methods such as moderated focus groups and think-aloud sessions in the theorizing quadrant of the DSR focus matrix (Brendel et al., 2022) throughout two design iterations. Guided by value-sensitive design and sociotechnical systems theory as our kernel theories, we develop design requirements and principles as a constituted design theory that embodies a general design solution to the problem class (Baskerville & Pries-Heje, 2010) of lacking guidance in developing ethically and performant deep learning models trained on synthetic image data in IS research. Since recent literature explicitly calls for guidance on synthetic data utilization (Murtaza et al., 2023) and design decisions, in general, are not neutral (i.e., having moral and ethical implications) (Findeli, 1994), our design theory addresses key aspects such as ethical compliance, privacy protection, data governance, scene diversity, controlled composition, complexity management, data augmentation, and responsible AI, providing actionable and theory-grounded guidance for researchers and practitioners in the selection, generation, and integration of synthetic image data for digital health computer vision. By adopting the design theory and a derived conceptual model, researchers can ensure the ethical and responsible use of synthetic image data while enhancing model performance, privacy protection, and projectability.

This paper contributes to the theoretical discourse on AI healthcare computer vision systems by presenting a design theory grounded in sociotechnical systems theory and value-sensitive design. Our study addresses the ethical and technical challenges of using synthetic image data for computer vision model training, ensuring that ethical values are embedded in the design process. By integrating core theories, we offer a balanced approach that balances ethical and social imperatives with technical performance, thereby advancing the responsible use of AI in digital health. The following sections cover the theoretical foundations, research methodology, design theory description, evaluation, and discussion.

Theoretical Foundations

Kernel Theories and Theoretical Lens

To ensure scientific rigor and stringency, DSR endeavors can use kernel theories to derive design principles. Broadly speaking, kernel theory functions as a form of justificatory knowledge within the realm of design knowledge development, as indicated by the work of Gregor and Hevner (2013), such as in the form of design principles (Gregor et al., 2020). DSR efforts may employ kernel theories to derive foundational design principles since the genesis of novel artifacts is grounded not merely in prototyping and user participation but fundamentally depends on a kernel theory (Walls et al., 1992; 2004). Thus, this study adopts the *analyze with lens*-mechanism proposed by Möller et al. (2022), drawing upon the theoretical foundations of employing kernel theories as a means of analysis and adhering to the conceptual boundaries of a theory. Using one or more kernel theories as a foundational framework can help clearly define normative prescriptions within design theories, directing the strategies and methods used to find specific solutions through design. Therefore, we argue that the use of a theoretical lens allows us to derive concepts indirectly, guiding the analysis or framing of data within the conceptual borders of a specific theory. This approach aligns with the perspective of Niederman and March (2019) on the theoretical lens, which emphasizes its role in aiding the theorization process, leading to the formulation of design principles or meta-requirements based on a data foundation. In the context of our study, we chose to use two kernel theories that not only informed our research methodology but were also used to inform our design requirements and principles. As shown in Figure 1, we used value-sensitive design theory (Friedman et al., 2002; 2013) specifically for our theory building and to frame our heuristic theorizing process, while sociotechnical systems theory (Trist & Bamforth, 1951; Emery & Trist, 1960) and its concepts were used for our theory grounding.

Value-sensitive design theory (Friedman et al., 2002; 2013) was chosen because it embodies a theoretically grounded approach that considers human values in a principled and comprehensive manner throughout the design process, which aligns with the research goal of developing technology that respects and incorporates user values while ensuring ethically responsible and user-centered design decisions. In general, value-sensitive design encompasses conceptual (Friedman et al., 2013), empirical (Deng et al., 2016), and technical investigations (Denning et al., 2010) of phenomena and their value conflicts and trade-offs (Mueller et al., 2018). Demonstrating its utility in addressing ethical issues in digital health (Dadgar & Joshi, 2018; Detweiler & Hindriks, 2012; Denning et al., 2014), value-sensitive design theory provides a structured approach to incorporating human and ethical values into technology design (Yetim. 2011). We also used value-sensitive design as a guiding framework to structure and evaluate the think-aloud sessions. Ethical considerations such as privacy, fairness, and inclusivity were explicitly embedded in the session design, allowing participants to reflect critically on these values while interacting with prototype artifacts. This approach ensured that the resulting design knowledge was aligned with the theoretical underpinnings of value-sensitive design. In the specific use-case of synthetic image data utilization for deep learning tasks in digital health, this especially connects to the synthetic depiction of humans (including separate or related characteristics, e.g. body parts or demographics), the use-case context in which the synthetically generated image data is used (often human-related, e.g. medicine or surveillance), and the potential ethical implications that arise from the creation and utilization of synthetic images. Given that synthetic imagery often closely mirrors real-image references, we aim to ensure that ethical implications, especially privacy concerns, are systematically addressed throughout the design process. This theoretical lens thus helps not only to derive design principles and requirements but also to establish a rigorous elicitation and evaluation (e.g., with think-aloud sessions) of these principles in the context of digital health. Using this kernel theory ensures that synthetic deep learning technology is not merely functional (i.e., precise, accurate, and reliable) but also ethically value-bound. By grounding the design process in a rigorous theoretical foundation, this approach not only facilitates the development of ethically responsible and user-centered digital health technologies but also contributes to the evolution and refinement of the theory itself, enhancing its applicability and relevance in the rapidly evolving landscape of digital health.

To complement the ethical framework established by value-sensitive design, this study integrates sociotechnical systems theory as an additional kernel theory to ensure a comprehensive approach to the design of synthetic image data in digital health computer vision. Originating from the seminal work of Trist and Bamforth (1951) and further developed by Emery and Trist (1960), sociotechnical systems theory emphasizes the intricate interplay between social subsystems (i.e., organizational structures, cultural practices, and human interactions) and technical subsystems (i.e., tools, processes, and technologies). In the context of using synthetic image data to train deep learning models in digital health, this interplay becomes particularly apparent. The sensitivity of health-related data and the potential for misrepresentation or bias necessitate a design approach that not only incorporates ethical considerations but also acknowledges the sociotechnical dynamics at play. Mumford's (2006) reflections on sociotechnical design underscore the need to align both social and technical elements to achieve sustainable and effective outcomes. This alignment is critical when generating synthetic imagery that must accurately represent diverse populations without perpetuating existing biases. Within the field of IS, Baxter and Sommerville (2011) highlight how sociotechnical systems theory can guide the development of systems that are both technically robust and socially responsive, a principle that resonates with the challenges of ensuring data realism and utility in synthetic datasets. Similarly, Bostrom and Heinen's (1977) examination of MIS failures through a socio-technical lens highlights the importance of considering human factors alongside technical specifications to prevent system failures. By grounding our DSR process in sociotechnical systems theory alongside value-sensitive design, this study not only advocates for ethical and user-centered technology but also ensures that the complex interdependencies between healthcare professionals, patients, and technological systems are thoughtfully integrated. This dual theoretical foundation facilitates the derivation of design requirements and principles that address the multiple challenges of using synthetic image data in deep learning models - balancing technical performance with ethical imperatives and social considerations. Consequently, the integration of sociotechnical systems theory provides a robust framework for navigating the complexities inherent in digital health environments and promotes the development of AI solutions that are not only effective and accurate but also socially attuned and ethically sound.

Synthetic Images in Deep Learning

Given that state-of-the-art deep learning models for computer vision comprise millions, if not billions, of parameters, training such models requires large amounts of data, which is often costly, missing, or unbalanced (Alzubaidi et al., 2021). Synthetic image data (which is synonymous with artificial image data) in deep learning refers to artificially created imagery that is generated by algorithms (e.g., video game engines) to train deep learning models. Several studies have highlighted the effectiveness of synthetic data in various computer vision tasks. Lee et al. (2019) and Krump et al. (2020) utilized synthetic datasets for deep learning-based object detection, specifically in underwater sonar imaging and vehicle detection on unmanned aerial vehicle platforms, respectively, highlighting the effectiveness of synthetic data in various computer vision domains. Additionally, domain adaptation, which is a technique that involves adapting a model trained on one (synthetic) domain of data to perform well on a different but related (real) domain, and transfer learning have been extensively explored in the context of synthetic data. Lahiri et al. (2018), Venkateswara et al. (2017), and Kuhnke and Ostermann (2019) focused on unsupervised domain adaptation for synthetic data, learning transferable feature representations, and domain adaptation for pose estimation, respectively. Seib et al. (2020) conducted a comprehensive review of current approaches that combine real and synthetic data to enhance neural network training, supporting the argument for a combination of training data and data augmentation. Aranjuelo et al. (2021) discussed key strategies for synthetic data generation in people detection from omnidirectional cameras, emphasizing the effective use of both real and synthetic data. Valtchev and Wu (2021) demonstrated the utility of domain randomization for neural network classification, showcasing the effectiveness of synthetic data in training robust models.

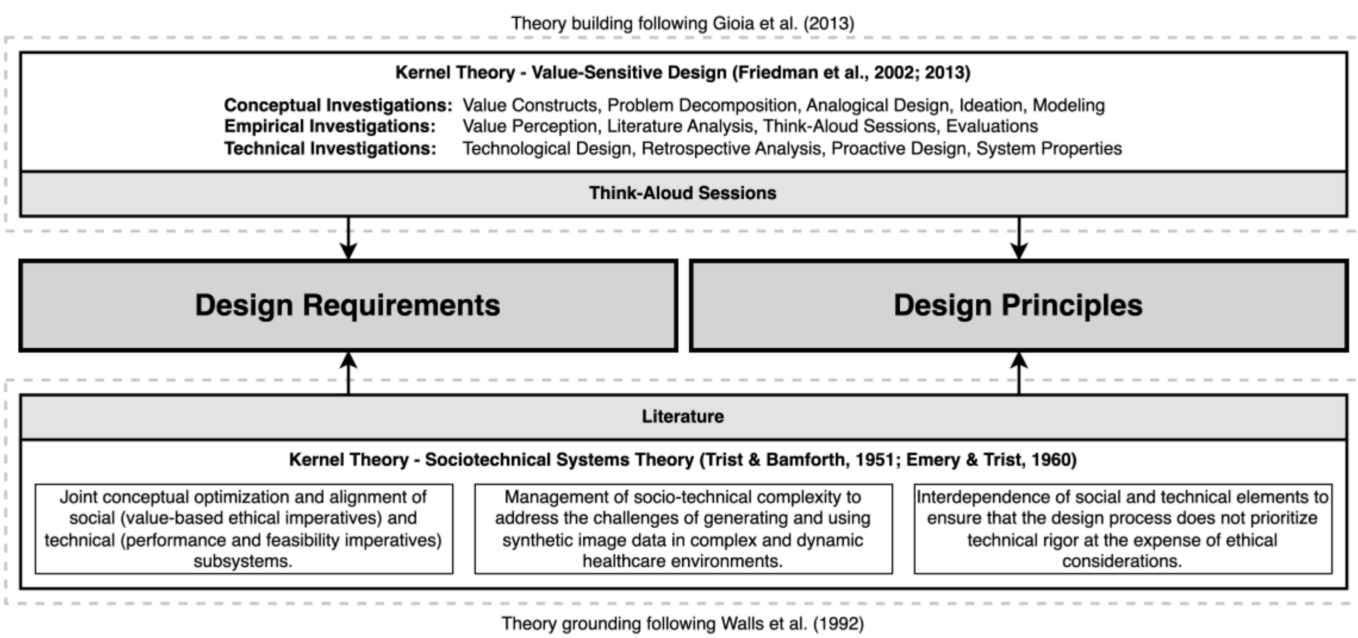


Figure 1. Our Kernel Theories and How They Inform the Design Science Approach of This Study

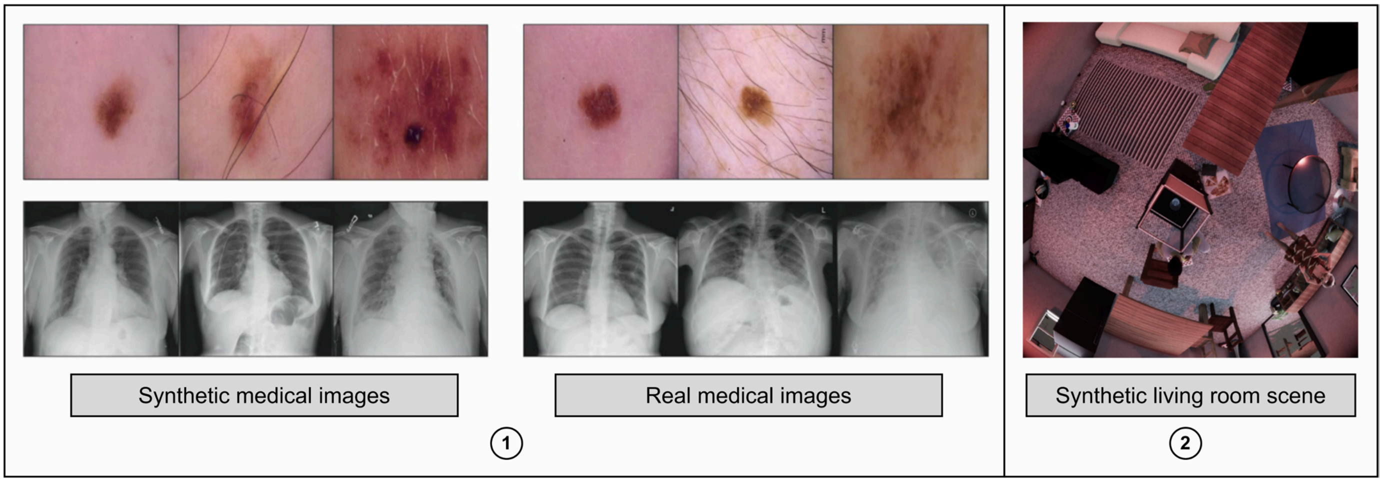


Figure 2. Exemplary Applications of Synthetic Data with 1) A Comparison of Synthetic and Real Image Data in Medicine (Chen et al., 2021) and 2) Synthetic Living Room Scenes in Digital Health (Own Study)

Moreover, the combination of synthetic and real training data has been investigated by several researchers. Wan et al. (2021) and Bird et al. (2020) utilized mixed datasets, comprising both synthetic and real data, for document layout analysis, scene classification, and object detection in augmented reality, respectively. Thereby, these studies highlight the benefits of leveraging both synthetic and real data for training computer vision models. Furthermore, the use of synthetic data generation techniques and simulators has been explored. Müller et al. (2018) introduced a photo-realistic simulator for generating synthetic data for computer vision applications, whereas Zhang et al. (2018) proposed a stacked multichannel autoencoder framework for efficient learning from synthetic data. In addition, Scheck et al. (2020) introduced a synthetic dataset that serves as a valuable resource for training and evaluating deep learning models in overhead object detection.

At the intersection of synthetic data (not limited to image data) and digital health, Giuffrè and Shung (2023) thoroughly examine the technical benefits and challenges of synthetic data in healthcare, focusing on its use in policy, privacy, and predictive analytics. They highlight how synthetic data can improve healthcare by enabling data-driven decision-making while addressing privacy concerns and data scarcity. However, they also identify significant risks, including data bias, the potential for re-identification, and regulatory gaps. As such, they point out that biases in training data can lead to inaccurate or discriminatory results in AI models, which is particularly problematic in healthcare, where decisions directly affect patient outcomes. The paper discusses the need for robust methods to verify the quality of synthetic data, calls for transparency in data generation processes, and proposes the establishment of a digital chain of custody to maintain data integrity and accountability. While these insights are essential, their work primarily addresses technical, privacy, and regulatory challenges, leaving a gap in ethical considerations, particularly as they relate to synthetic image data in digital health. Here, Murtaza et al. (2023) reviewed the generation of synthetic data in healthcare, while Gonzales et al. (2023) discussed the potential applications of synthetic data, focusing on datasets and innovative utility. As shown in Figure 2, synthetic image data can be applied across various medical and clinical healthcare applications such as skin lesion and cancer detection, x-ray diagnosis, or renal cell carcinoma treatment (Chen et al., 2021), providing insights into the generation and use of synthetic data for training deep learning models. At this point, we would like to note that the illustrative examples shown in Figure 2 are only one of many possible forms of synthetic image data applications in digital health, and that, for example, the correct classification rate for cancer detection needs to be significantly higher compared to preventive care use cases, as our instantiation will show later. We argue that a misdiagnosis in cancer detection with synthetic image data is much more fatal than an error in home care interaction with cognitively impaired individuals, for which Figure 2 is only an example of how photorealistic synthetic image data can be and what possible use cases are conceivable.

Despite the considerable research on utilizing synthetic image data for computer vision deep learning models, there remains a notable research gap in terms of a comprehensive framework or guidelines that provide design knowledge to effectively and systematically utilize synthetic data in this context. While individual studies have demonstrated the benefits and effectiveness of synthetic data in specific tasks, there is a lack of unified principles or guidelines that guide researchers and practitioners in the selection, generation, and integration of synthetic image data for training deep learning models. Moreover, relying on synthetic data without a robust ethical framework risks exacerbating existing biases or introducing errors, particularly when synthetic datasets are generated without adequate consideration of demographic diversity or the potential for subtle patterns to be exploited, underscoring the critical need for proactive safeguards to prevent unintended harm in high-stakes scenarios such as cancer detection. Hereby, the absence of such design knowledge hinders the widespread adoption and consistent utilization of synthetic data, leading to potential inefficiencies, suboptimal performance, and challenges in real-world deployment (Giuffrè & Shung, 2023; Müller et al., 2018; Scheck et al., 2020; Zhang et al., 2018).

Ethics in IS and Digital Health

(Digital) Ethics has long been an afterthought in IS research but has gained some momentum in the last decade and is showing various manifestations with a growing number of studies (Kern et al., 2022). Ethics, broadly speaking, is a branch of philosophy that assesses the morality of human actions, offering a thorough examination of interconnected ethical dilemmas, concepts, principles, reasoning, and judgments through various ethical theories (Becker & Becker, 2001; O'Neil, 2004) that orbit around *teleological* (Aristotele et al., 2009), *deontological* (Habermas, 1987; Kant, 1998), and *weak normative/contextual* (Düwell et al., 2011) categories. Similarly, digital ethics focuses on the moral implications associated with the development and application of digital technologies, providing a comprehensive framework to address the ethical and moral challenges posed by digital innovation. Intending to formulate morally good solutions as defined by Floridi and Taddeo (2016), the field encapsulates varying viewpoints; for instance, Spiekermann et al. (2022) emphasize ethical virtues and values, while Schlagwein et al. (2019) advocate for discourse ethics as a foundation for analysis. While existing ethical research frameworks in IS and DSR, such as those proposed by Myers and Venable (2014), provide valuable meta-level guidance, these tend to focus on broad, generalizable principles that may not fully address the unique ethical challenges presented by synthetic image data in digital health computer vision. These unique challenges include the potential for synthetic data to replicate or amplify biases in source datasets, vulnerability to re-identification despite its artificial nature, and ethical ambiguities surrounding consent for its creation and use (i.e., in terms of accountability, fairness, transparency, or compliance). In addition, the perceived infallibility of synthetic data risks fostering complacency, while its increasing accessibility increases the potential for misuse in sensitive digital health contexts. Addressing these issues requires rigorous, domain-specific ethical frameworks that consider the high stakes and sociotechnical complexities of healthcare applications. Similarly, while Herwix et al. (2022) emphasize the importance of value systems design in IS research, there remains a critical need for domain-specific ethical guidelines that not only build upon these existing frameworks but also provide actionable insights tailored to the sociotechnical demands of digital health applications, where the ethical stakes are particularly high. Subsequently, digital health is considered a branch in the interdisciplinary field of digital ethics, which overlaps with, for example, computer science (Mahieu et al., 2018) or the paradigm of computer ethics (Floridi, 2010). More specifically, the IS concept of technology adoption in terms of privacy and e-health records has been reviewed by Hansen and Baroody (2020), while Turja et al. (2020) studied the acceptance of robotic caretakers that pay attention to ethical values and Al-Dhaen et al. (2021) explored the ethical considerations in the internet-of-medical-things use intentions.

On a more conceptual theory level and as shown in Figure 3, digital health ethics can be allocated to *AI ethics* and the design and development of positive/beneficial artificial intelligence (Floridi & Cowls, 2019); *value ethics* and the indication that a person strives for higher intrinsic values (Becker & Becker, 2001); and *IS ethics* that specifically applies to information systems research (McBride, 2014). From this perspective, the developed concepts (Kern et al., 2022) of *artificial intelligence* to act effectively and safely in AI environments (Russell & Norvig, 2021) and *privacy* to determine the extent of personal information exposure (Hung & Cheng, 2009) hereby inform our study of synthetic image data in digital health computer vision in terms of the guiding theoretical orientation. In addition, Rajkomar et al. (2019) delve into the application of AI in healthcare, discussing both its potential benefits and associated challenges. Their comprehensive analysis covers a range of issues, from improving patient care to ethical considerations and data privacy. Focusing on the critical issue of bias in healthcare AI, Obermeyer et al. (2019) provide an in-depth examination of how AI algorithms can inadvertently perpetuate bias and offer methods for identifying and reducing these biases, contributing to the development of fairer and more equitable AI tools in healthcare. In this context, Raji et al. (2020) explored the ethical dimensions of auditing facial recognition technologies, focusing on the methodological challenges and moral implications associated with auditing these AI systems and situating their study within the broader conversation about AI, ethics, and societal impact, with a particular focus on facial recognition technology. Building on this, Vayena et al. (2019) present an ethical framework tailored for AI applications in healthcare. They argue for a multi-stakeholder approach, involving collaboration between technologists, ethicists, and legal experts, highlighting the importance of balancing innovation with ethical responsibility and ensuring that AI advancements enhance healthcare while safeguarding patient privacy and data security. In the larger context of AI research in healthcare, Panch et al. (2019) critically examine the application of AI in healthcare, addressing what they refer to as the 'inconvenient truth' of AI in this domain and highlighting its contributions to the ongoing discussion about the realistic and ethical application of AI technologies in medical settings. Their work serves as a critical reminder of the need to balance enthusiasm for AI with a clear-eyed assessment of its challenges and limitations. Since digital health environments and especially image-based computer vision approaches are exposed to sensitive data, the incorporation of digital ethics is necessary.

However, as digital ethics has gained momentum in recent years, there is a significant lack of research in the area of computer vision healthcare, particularly in the use of synthetically generated image data – both theoretically and practically. The notion of generating synthetic medical images, for instance, creates a potential ethical grey area around consent, data privacy, and the authenticity of information. The stakes are high, as missteps can lead to misinformation, misdiagnosis, or the exploitation of sensitive personal data, such as potential demographic biases or misrepresentations (Giuffrè & Shung, 2023; Gonzales et al., 2023, Murtaza et al., 2023; Obermeyer et al., 2019; Panch et al., 2019). Interestingly, while it could be argued that synthetic data is artificial data and therefore easier to protect privacy, synthetic data often mirrors real-world data, meaning that identifiable patterns or sensitive information from real-world reference datasets may be unintentionally replicated in the synthetic output (Giuffrè & Shung, 2023, Gonzales et al., 2023). This link between synthetic and real data underscores the need for robust ethical frameworks and privacy protections, particularly in the sensitive context of digital health, where the potential for re-identification and misuse remains significant. Moreover, the intersection of AI ethics, value ethics, and IS ethics with the realm of synthetic image data in digital health computer vision needs a thorough exploration to ensure the responsible development and deployment of these technologies. Therefore, design knowledge and guidance are needed to ensure the ethical and performant use of synthetically trained deep learning models in digital health, and to generate compliant image data that is both realistic and value-sensitive (Becker & Becker, 2001; Giuffrè & Shung, 2023; Murtaza et al., 2023).

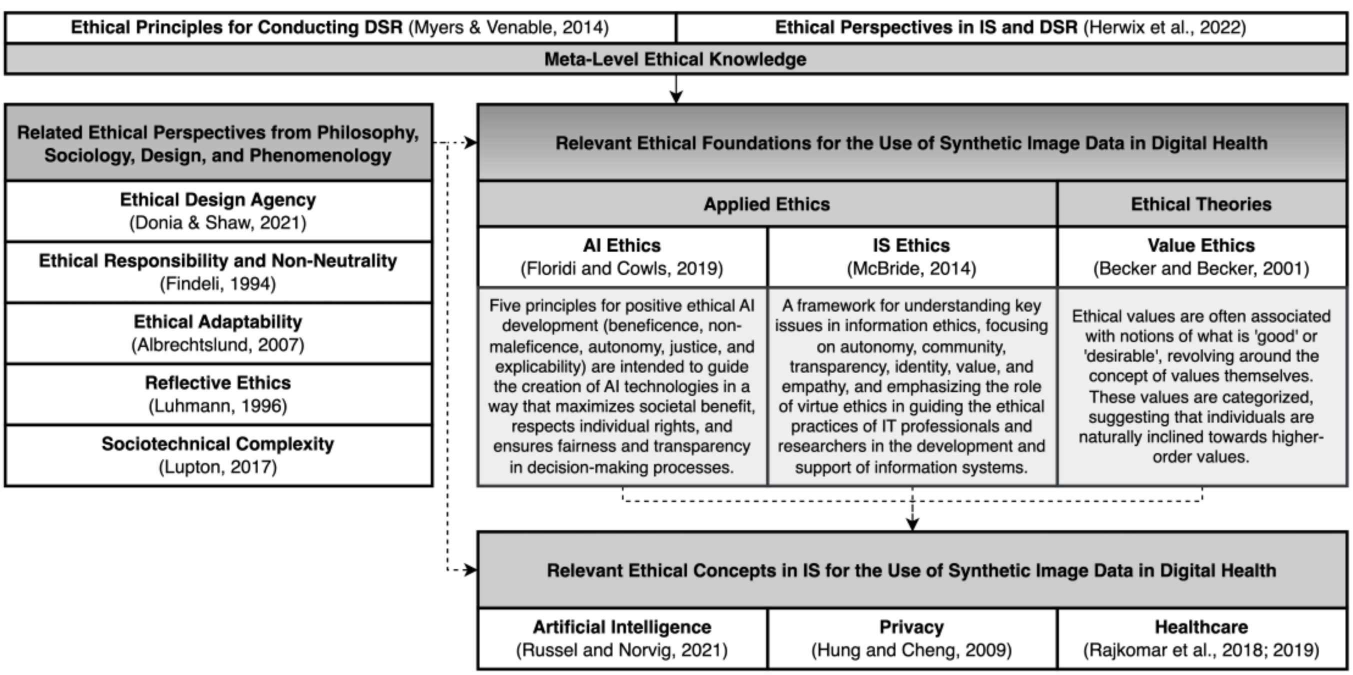


Figure 3. Relevant Ethical Foundations and Concepts for the Use of Synthetic Image Data in Digital Health

From a more philosophical, sociological, and phenomenological perspective, our study is further informed by theoretical foundations and (design) knowledge from related fields to IS and digital ethics. In general, ethical design decisions are not neutral but have inherent moral implications, as highlighted by Donia and Shaw (2021), who argue that the degree of designer agency - constrained by external factors such as corporate/performance pressures - significantly influences the integration of values into design processes. Moreover, the concept of *normative strength* within design emphasizes how strongly ethical principles are embedded in these processes, challenging us to consider how the constraints on designers' agency affect the ethical quality of the resulting artifacts. This aligns with Albrechtslund's (2007) critique of positivist assumptions in value-sensitive design, which argues for a flexible approach that recognizes the multistability of technologies and the unpredictable ways in which they can be used or misused once deployed. This phenomenological approach shifts the focus from predicting ethical use to preparing for ethical adaptability and flexibility in design. In addition, Findeli's (1994) exploration of the intertwined relationship between ethics and aesthetics in design emphasizes that design decisions should take into account both ethical and aesthetic judgments, reinforcing the notion that design in socio-technical systems carries moral responsibility, especially when designing artifacts that mediate between the human and technological realms (biocosm, sociocosm, and technocosm). The term *technoethics* is introduced to discuss the ethical issues arising from technological developments and their impact on professions, including design. Findeli (1994) emphasizes that designers have a professional responsibility for the moral implications of their creations that goes beyond mere functionalism. Designers must consider how their designs affect users, society, and the environment at large, which is consistent with the key assumptions of sociotechnical systems theory. Integrating this with Luhmann's (1996) reflections on ethical systems as adaptive and reflective processes within social contexts underscores the need for continuous ethical deliberation in complex environments such as computer vision and digital health. Finally, Lupton (2017) critiques mainstream design thinking for neglecting sociocultural complexity and advocates for speculative design approaches that address technological systems' broader social and ethical implications. Together, these perspectives deepen our engagement with ethical and sociotechnical frameworks and ground our study in a comprehensive theoretical foundation that not only addresses the challenges of synthetic data but also ensures that ethical considerations are central to the design of health technologies.

Literature Review

As shown in the previous sections, research on deep learning models, especially concerning synthetic training data, and digital ethics in IS has become increasingly sophisticated. While real image models in computer vision depend on the availability and quality of training data, this particular challenge can be addressed by synthetic data, which is scalable, cheap, and less error-prone, but also raises questions of ethical use, especially in digital health. Research has consistently highlighted the complexity and data-intensive nature of deep learning models in computer vision, underscored by the challenges of acquiring the diverse and extensive datasets required for these models (Alzubaidi et al., 2021). This gap in effective and efficient data acquisition, which is particularly relevant in digital health, affects model accuracy and generalizability, which directly impacts human health outcomes. Looking at Research Stream 1, which focuses on synthetic image data in deep learning, it can be conceptually divided into the effectiveness and application of synthetic data (EA-SD) as well as the challenges and strategies in synthetic data utilization (CS-SDU), as shown in Table 1. Studies listed in the former category, such as those by Lee et al. (2019), Krump et al. (2020), and Aranjuelo et al. (2021), focus specifically on the applications and effectiveness of synthetic data in different deep learning and computer vision contexts. On the other hand, CS-SDU includes papers such as those by Lahiri et al. (2018) and Seib et al. (2020), which address domain adaptation, challenges in using synthetic data, and strategies for combining real and synthetic data. Research Stream 2 can then be conceptually divided into the theoretical and philosophical foundations of digital ethics (TPF-DE) and the practical applications and implications in digital health (PAI-DH). Here, studies from the former category, such as those by Becker and Becker (2001) and Floridi and Taddeo (2016), focus on the ethical theories and philosophical foundations relevant to digital ethics, which is different from the practical applications of these theories. On the other hand, the studies listed in PAI-DH, such as those by Hansen and Baroody (2020) and Turja et al. (2020), are specifically focused on the application of ethical considerations in digital health, which is distinct from the theoretical underpinnings of digital ethics. At the intersection of these conceptually divided research streams, general synthetic data has been addressed in terms of application, privacy, and policy (Giuffrè & Shung, 2023; Gonzales et al., 2023) while also covering healthcare settings (Murtaza et al., 2023). By further grounding our theory-driven approach (Schoormann et al., 2024) in philosophical, sociological, and phenomenological perspectives from related fields (Donia & Shaw, 2021; Albrechtslund, 2007; Findeli, 1994; Luhmann, 1996; Lupton, 2017), we ground our study in this prior (design) knowledge.

Table 1. Literature and Research Streams are Categorized Based on Their Conceptual Orientation

Literature	**Research Stream 1:** Synthetic Image Data in Deep Learning		**Research Stream 2:** Digital Ethics in IS	
	EA-SD	*CS-SDU*	*TPF-DE*	*PAI-DH*
Lee et al. (2019)	X			
Krump et al. (2020)	X			
Aranjuelo et al. (2021)	X	X		
Valtchev and Wu (2021)	X			
Wan et al. (2021)	X			
Bird et al. (2020)				
Lahiri et al. (2018)		X		
Venkateswara et al. (2017)		X		
Kuhnke and Ostermann (2019)		X		
Seib et al. (2020)		X		
Müller et al. (2018)		X		
Zhang et al. (2018)		X		
Scheck et al. (2020)		X		
Giuffrè and Shung (2023)		X		X
Murtaza et al. (2023)	X			X
Gonzales et al. (2023)	X			X
Becker and Becker (2001)			X	
O'Neill (2004)			X	
Floridi and Taddeo (2016)			X	
Spiekermann et al. (2022)			X	
Schlagwein et al. (2019)			X	
Vayena et al. (2019)			X	
Donia and Shaw (2021)			X	
Luhmann (1996)			X	
Lupton (2017)			X	
Albrechtslund (2007)			X	X
Findeli (1994)			X	X
Obermeyer et al. (2019)			X	X
Panch et al. (2019)			X	X
Hansen and Baroody (2020)				X
Turja et al. (2020)				X
Al-Dhaen et al. (2021)				X
Floridi and Cowls (2019)				X
Russel and Norvig (2021)				X
Raji et al. (2020)				X
Rajkomar et al. (2019)				X
Hung and Cheng (2009)				X
Our study approach	X	X	X	X
Note	• EA-SD = Effectiveness and Application of Synthetic Data • CS-SDU = Challenges and Strategies in Synthetic Data Utilization • TPF-DE = Theoretical and Philosophical Foundations of (Digital) Ethics • PAI-DH = Practical Applications and Implications in Digital Health.			

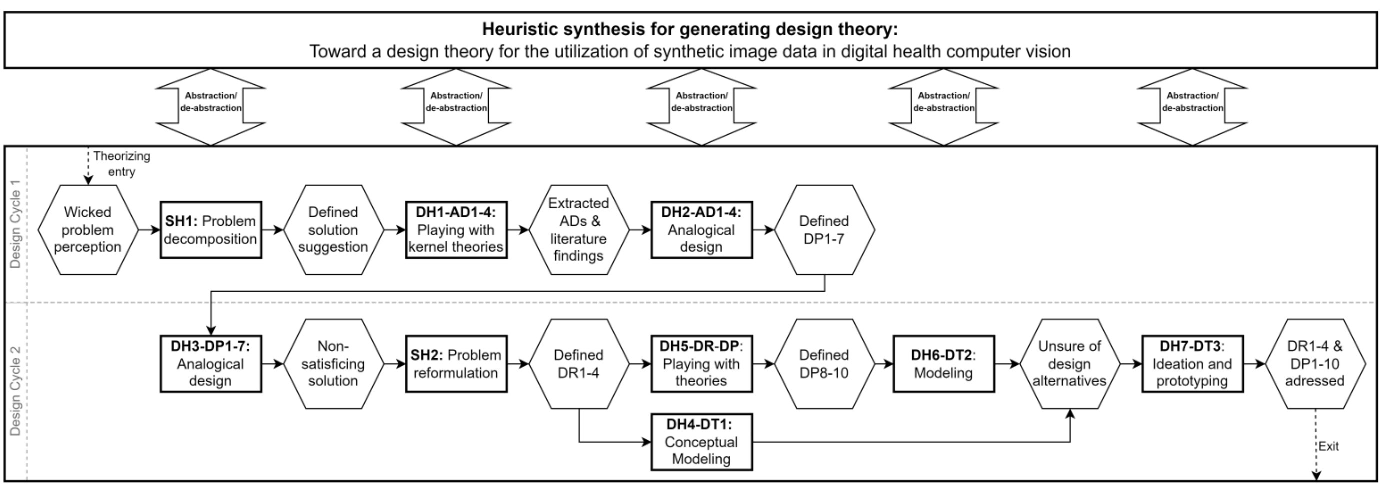
Despite the considerable work that has been done in these categories, there remains a significant gap in synthesizing these diverse streams into cohesive design knowledge that can inform the development of digital health technologies using synthetic image data. This gap is significant because it hinders the ability to systematically address the challenges and realize the potential of synthetic data in ways that are both ethically responsible and technologically effective (Rajkomar et al., 2018; Panch et al., 2018). This is particularly relevant given the rapid evolution of digital health technologies and their increasing reliance on complex deep learning models. The need for unified design knowledge becomes even more apparent when considering the diverse applications and implications of synthetic image data in digital health. Unlike meta-level approaches to ethical guidance in IS and DSR (Myers & Venable, 2014; Herwix et al., 2022), the need for knowledge about the ethical use of synthetic data stems from its growing use in healthcare, where it presents unique ethical challenges and, despite being artificially generated, often reflects real-world data that may inadvertently carry sensitive or identifiable characteristics. For example, using synthetic data to train algorithms for diagnostic purposes requires a socio-technical perspective with technical precision and a deep understanding of ethical considerations such as patient privacy and data security. Similarly, when synthetic data is used to train models for predictive healthcare analytics or in digital health applications aimed at assisting vulnerable populations, it is critical to ensure that the models are not only accurate but also unbiased and fair. Thus, the need for a design theory to address the research gap in the ethical and effective use of synthetic image data in digital health computer vision is multifaceted. First, it has the potential to integrate technical and ethical aspects of healthcare, providing a comprehensive framework for practitioners and researchers (Chandra et al., 2015; Meth et al., 2015). Furthermore, it aims to emphasize ethical compliance and social responsibility, focusing on values such as privacy and fairness in the use of synthetic data (Gregor & Jones, 2007). The design theory could, therefore, improve AI-driven medical diagnosis and treatment, promote innovation in medical imaging, and address risks such as AI bias and legal issues. It could, therefore, also foster interdisciplinary collaboration and reduce indeterminacy in design theory (Burton-Jones et al., 2021; Lukyanenko & Parsons, 2020), ensuring that digital health technologies remain adaptable and responsible.

Research Methodology

We drew on the DSR paradigm (Hevner et al., 2004), as our study seeks to provide prescriptive insights for research on synthetic image data, rather than merely descriptive knowledge, as outlined by Gregor and Hevner (2013). Given that the DSR paradigm can be instantiated through a variety of methodological processes (Venable et al., 2017), and that design theories, in general, embody a form of theorizing in IS research (Brendel et al., 2022; Burton-Jones et al., 2021; Dehling & Sunyaev, 2023; Gregory & Muntermann, 2014; Kane et al., 2021; Lee et al., 2011; Mandviwalla, 2015; Young et al., 2021), our study can be placed in the *theorizing* domain of the DSR focus matrix (Brendel et al., 2022), which implies a higher focus on rigor and derivation of theoretical artifacts. This matrix domain generally involves using higher-level kernel theories (i.e., in our case, value-sensitive design theory and sociotechnical systems theory) to form new design theories for a given problem space (i.e., in our case, the use of synthetic image data in digital health). In this domain, the development of design principles and propositions, similar (but not equal) to hypothesis formulation, plays a central role, emphasizing the importance of grounding new theories in prior knowledge (Brendel et al., 2022; Schoormann et al., 2024). Thus, we followed the grounding and conceptualization step for theory-driven DSR configurations as suggested by Schoormann et al. (2024) to ground our design requirements and principles in prior work and theory based on our theoretical foundations.

As such, we meticulously followed the heuristic design theorizing framework proposed by Gregory and Muntermann (2014) over two design cycles to ensure both transparency and stringency (Figure 4) for design theory replication as discussed by Brendel and Muntermann (2022). In our case, heuristic theorizing is defined as the process of proactively creating design theory from problem-solving experiences and prior theories. This is achieved by continuously iterating between seeking a satisficing (adequate) problem solution (heuristic search) and synthesizing new information generated during the heuristic search through, e.g., literature, theory, or practical findings (Gregory & Muntermann, 2014). Thus, our theorizing approach contained multiple rounds of analogical design (e.g., transferring design knowledge about one design situation to another), reformulating the problem (e.g., for specific domains), playing with kernel theories (i.e., value-sensitive design and sociotechnical systems theory), and modeling (cf. Figures 6 and 7). Here, our design theorizing approach is located in the abstract theoretical domain, seeking an abstract solution that solves an abstract problem, and using abductive reasoning because there is an objective to guide learning and problem-solving (Lee et al., 2011). From a more metatheoretical perspective of our theorizing approach, and in line with the general design theory development directions mentioned above, we followed the *envision*-strategy proposed by Burton-Jones et al. (2021) for next-generation theorizing. In doing so, our proposed design theory helps to study a new phenomenon (i.e., synthetically generated image data in computer vision) emerging in a changing world (i.e., dynamic digital health domain and increased ethical awareness), which effectively leads to a new IS theory.

Therefore, a multi-method research approach (Mingers & Brocklesby, 1997) consisting of a focus group session for data collection and two think-aloud sessions for evaluating the theorized artifacts was chosen to address the shortcomings of single methods and ensure validity. While the heuristic design theorizing process does not require formal evaluation (Gregory & Muntermann, 2014), we incorporated these evaluation episodes to align our design theory with the broader principles of DSR, ensuring both theoretical rigor and practical applicability. As such, our iterative, heuristic process of data collection and analysis ensured that the findings from these sessions were continually refined. We acknowledge that sample size is often questioned in qualitative research (Marshall et al., 2013; Hwang & Salvendy, 2010; Noushad et al., 2024), especially when dealing with smaller numbers of participants (see Figure 8 for our study sample), who in our case were selected for their specialized expertise in synthetic (image) data, digital ethics, and AI in digital health. However, theoretical saturation was reached when no new insights emerged to further inform our design theory (Glaser & Strauss, 2017; Charmaz, 2006), while data saturation occurred when additional data collection did not yield new second-order themes or aggregate dimensions (Saunders et al., 2018), signaling that further data would not contribute new insights. Code/thematic saturation (Bowen, 2008; Guest et al., 2020; Hennink et al., 2022) confirmed that the identified themes fully captured the essence of our study focus, even after adding or subtracting experts from the sample. Finally, we reached meaning saturation, where our interpretive understanding of the themes was fully developed, and no new meaning could emerge from additional data or sessions (Hennink et al., 2017; Yang et al., 2022). The depth of engagement with our expert participants allowed us to make meaningful design decisions without the need for a larger participant pool, focusing on the premise that qualitative rigor comes from the richness of the data rather than the number of participants (Bryant & Charmaz, 2007; Sandelowski, 1995). The rather small sample size was used to ensure the participation of a practical and realistically recruitable subset of the target community, allowing for a focused yet meaningful evaluation of the proposed design requirements and principles by relevant practitioners outside the DSR project (Iivari et al., 2021).



Note: SH=structuring heuristics; DH=artifact design heuristics; DR=design requirement; DP=design principle; AD=aggregate dimension; DT=design theory.

Figure 4. Heuristic Design Theorizing Approach, adapted from Gregory and Muntermann (2014)

Table 2. Value-Sensitive Design as a Companion to the Heuristic Design Theorizing Process

Kernel Theory Steps of Value-Sensitive Design	Application in our Health and Ethics Context	Application in the Heuristic Design Theorizing Process
1) *Start with a Value, Technology, or Context of Use*	Synthetic image data in the context of digital health deep learning, focusing on ethical implications, privacy concerns, and the need for unbiased AI models.	Corresponds to the initial entry into the theorizing process where the "wicked problem perception" is identified. Transitions into problem decomposition (SH1).
2) *Identify Direct and Indirect Stakeholders*	Researchers, designers, developers and regulatory bodies as direct stakeholders. Patients and healthcare providers as indirect stakeholders.	Corresponds to the problem decomposition where stakeholders' roles and impacts are considered.
3) *Identify Harms and Benefits for Each Stakeholder Group*	Identified privacy breaches, bias, misrepresentation as potential harms. Identified model performance, accuracy, scalability, and applicability as potential benefits.	Occurred during the problem decomposition and continued through the analogical design and playing with kernel theory stages where potential design solutions were discussed and evaluated.
4) *Map Harms and Benefits onto Corresponding Values*	Aligned identified harms and benefits with the ethical values of privacy and consent, fairness and non-discrimination, accountability and transparency, and beneficence and non-maleficence.	Achieved through analogical design stages where harms and benefits are aligned with the defined ethical values. Subsequently, these were incorporated into design requirements and principles.
5) *Conduct a Conceptual Investigation of Key Values*	Analyzed values and enriched them by theoretical and practical underpinnings to ensure they are embedded in the design requirements and principles.	Conducted in-depth theoretical analysis and aligned these insights with the theory-building dimensions of the think-aloud content analysis.
6) *Identify Potential Value Conflicts*	Addressed conflicts such as performance vs. transparency or anonymity vs. accountability, and balancing these in the design theory.	Addressed through problem reformulation (SH2) and iterative design cycles to balance conflicting values and find adequate design solutions.
7) *Integrate Value Considerations into the Design Process*	Ensured that the design and implementation processes are aligned with the ethical values/standards of the design theory.	Corresponds to the final design and the exit of the heuristic theorizing process.

As shown in Table 2, and according to our analyze-with-lens kernel theory mechanism (Möller et al., 2022), we used value-sensitive design to inform our heuristic theorizing process in terms of the methods used and aspects of the design requirements and principles. By incorporating value-sensitive design theory as a theoretical lens into the heuristic theorizing process, we ensure that ethical values are systematically integrated into both our research methodology and the design artifact. Thus, we based the structured heuristics (SH1, SH2) on problem decomposition and reformulation and their understanding of ethical and technical challenges. Subsequently, our artifact design heuristics (DH1-DH7) ensured that our design artifact embodied the identified values and addressed the key ethical and technical concerns. Thus, the iterative process of value-sensitive design accompanied the equally iterative nature of our DSR methodology.

As shown in Figure 4, our multi-cyclical research approach contains two sequentially completed design cycles, with each of them containing theorized artifacts, whereas the second completed cycle drew on the conclusion of the first one. The design knowledge encompasses the collective design intuitions, design decisions, principles of form and function, and the descriptive insights employed to comprehend the problems and formulate the respective solutions in each cycle. Therefore, the approach and results within this paper are presented for both design cycles.

The first design cycle was dedicated to gathering initial and foundational design knowledge about the use of synthetic image data in digital health computer vision environments, focusing on information security and model performance. In this first cycle, the main artifact proposed was a set of nascent design principles based on the literature and a qualitatively analyzed focus group session with AI experts (n = 11). For the qualitative content analysis, we followed the methodological approach proposed by Gioia et al. (2013) and developed first-order concepts, second-order themes, and aggregate dimensions (AD). This particular content analysis methodology is used to build theory within IS (Magnani & Gioia, 2023) and provides our foundation for design requirement development. This approach to concept development allowed us to balance the often-conflicting goals of developing new concepts inductively and maintaining high standards of rigor in IS research. We used first-order coding (inductive) to capture informant-centered concepts, which were then interpreted into second-order themes through constant comparison and theoretical reflection (abductive), as shown in Böhmer et al. (2023). These themes were then aggregated into four dimensions representing core design constructs: *privacy and ethical compliance* (AD1), *data governance* (AD2), *synthetic scene generation* (AD3), and *robust learning and generalization* (AD4). These ADs, together with the literature, formed the basis of the design principles, which are prescriptive and universal in this context, specifying how an instantiation should be designed to meet the proposed requirements (Fu et al., 2016). In this context, the design principles were derived from a supporting approach (Möller et al., 2020) that complements the conceptual scheme of Gregor et al. (2020) (see Appendix A for the detailed scheme), whose a priori specification suggests a prescriptive formulation (Fu et al., 2016). Moreover, the use of the framework allowed us to formulate accessible, precise, and expressive design knowledge (Gregor et al., 2020). By evaluating the reusability of the design principles in the context of digital health (Iivari et al., 2021), we derived several areas for improvement and revision, which were crucial for initiating the second design cycle. Evaluation feedback from the first cycle included suggestions for strengthening the responsible AI paradigm due to ethical and social concerns, revising ambiguous design principle descriptions, addressing control mechanisms in data generation, and reevaluating the strategy for increasing complexity in synthetic scenes. Concerns were also raised about the specific applicability of these principles in the digital health domain, especially in the context of digital ethics, through potentially missing design knowledge, as their applicability may vary in different support, treatment, or diagnostic contexts, which ultimately triggered the need for a second cycle. Based on the evaluation feedback from the first cycle and its results (Böhmer et al., 2023), we decided to revise the artifact and conduct a second cycle, focusing on a more holistic theoretical approach, incorporating the revision suggestions, and operationalizing our design theory artifact.

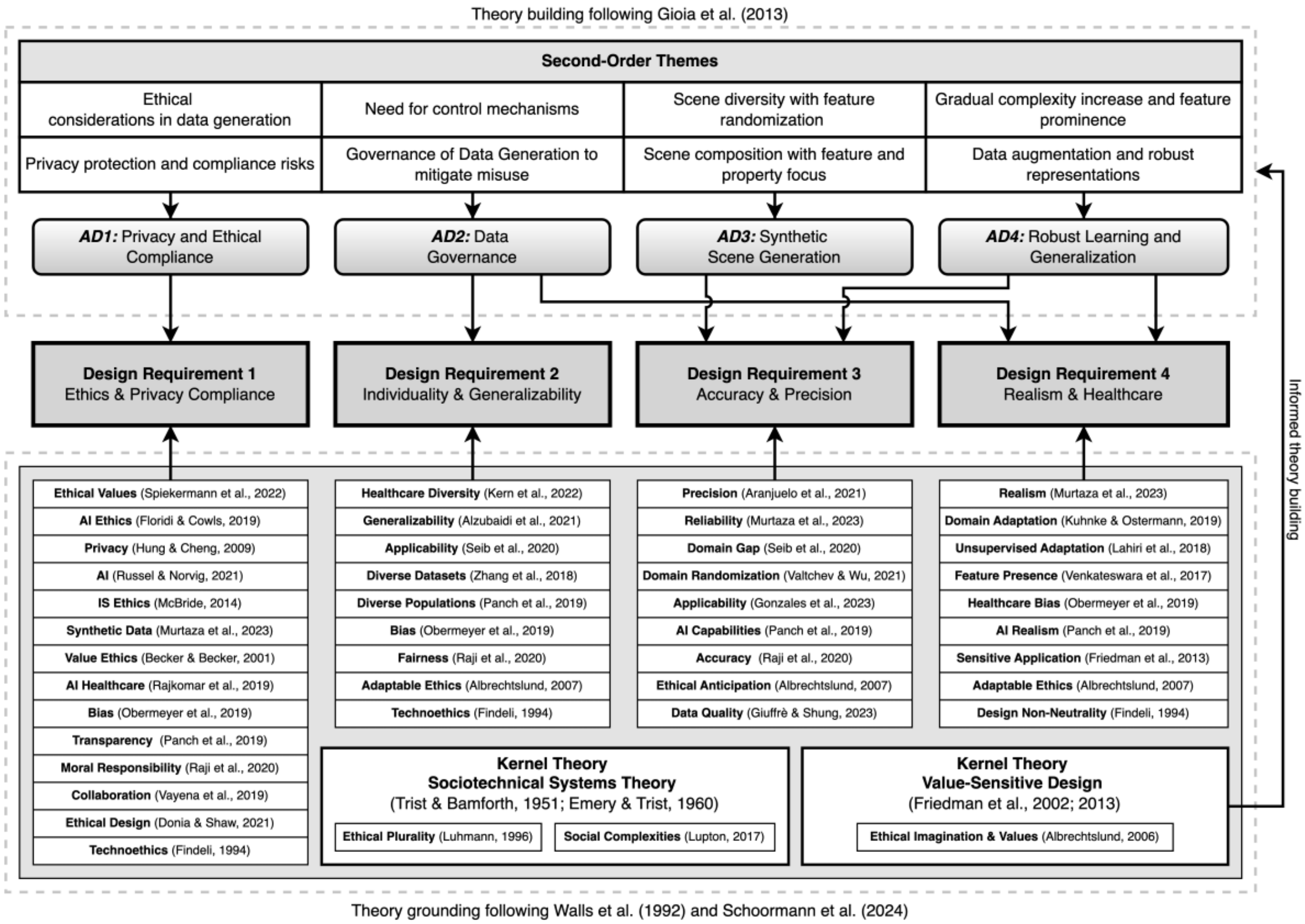
Design Theory

Design Requirements

The design requirements outlined by our design theory serve as the main objectives for the theoretical and conceptual artifact. In a way, these requirements can be seen as meta-requirements, as they guide the development of the subsequent reference architecture (Baskerville & Pries-Heje, 2010; Walls et al., 1992; Schoormann et al., 2024). Following evaluation feedback on missing design requirements from cycle 1 (cf. Table 2), we derived the requirements of *ethics & privacy compliance* (**DR1**), *individuality & generalizability* (**DR2**), *accuracy & precision* (**DR3**), and *realism & healthcare relevance* (**DR4**). Figure 5 shows both the theory building (Gioia et al., 2013) and theory grounding (Walls et al., 1992; Schoormann et al., 2024) of those DRs, while Figure 6 shows a simplified depiction of our design theory and its core components.

When working with synthetic image data in digital health, it is crucial to ensure that all processes related to the generation and use of synthetic imagery are ethical and privacy-compliant (**DR1**), protecting individual privacy rights and ethical norms (Dadgar & Joshi, 2018; Mueller et al., 2018). Drawing from *AD1*, it highlights the need for ethical responsibility in the development and deployment of technologies that handle sensitive data in digital health environments as depicted by value-sensitive design theory (Friedman et al., 2002; 2013), the use of ethical values in IS (Spiekermann et al., 2022), and the general necessity for ethical AI and IS considerations (Floridi & Cowls, 2019; Hung & Cheng, 2009; McBride, 2014; Murtaza et al., 2023; Russell & Norvig, 2021). While it builds on general meta-level guidance on ethical research in IS (Myers & Venable, 2014; Herwix et al., 20222), it embodies the core ethical values, which include adherence to the aspects of privacy and consent, fairness and non-discrimination, accountability and transparency, and beneficence and non-maleficence. Grounded in philosophical, sociological, and phenomenological ethics, Donia and Shaw (2021) emphasize the role of designer agency in embedding ethical values, while Findeli's (1994) concept of "technoethics" emphasizes the moral responsibility of designers to consider the ethical implications of technological decisions and to ensure privacy. Deng et al. (2016) and Yetim (2011) emphasize the need for fairness, inclusiveness, and transparency, reinforcing the importance of ethical guidelines and privacy protection in the generation of synthetic data, especially in understanding stakeholder values (Denning et al., 2014). These tenets ensure compliance with privacy standards such as GDPR, promote equitable treatment, require transparent methodologies, and focus on maximizing benefits while minimizing harms, thereby aligning synthetic data use with societal values and ethical standards (Becker & Becker, 2001; Hung & Cheng, 2009; Floridi, 2010; Russel & Norvig, 2021). In this context, it is important to recognize that while synthetic data is artificially generated, it can still pose risks of re-identification and privacy breaches if not properly managed. This is particularly relevant when the synthetic data is modeled on sensitive real-world datasets, as it may inadvertently retain identifiable features or patterns (Giuffrè & Shung, 2023; Gonzales et al., 2023). For example, synthetic images generated from a dataset of patient scans may still contain patterns or features that, when combined with other data, could potentially be traced back to the original patient. Therefore, DR1 emphasizes the critical need for rigorous data governance and transparency in the synthetic data generation process to avoid a false sense of security (i.e., the misconception that it is just "artificial data"). This ensures that synthetic data does not circumvent ethical and legal obligations, especially when it reflects the sensitive real-world scenarios for which models are trained. As originally proposed by sociotechnical systems theory (Trist & Bamforth, 1951; Emery & Trist, 1960), the need for joint optimization of social and technical subsystems emphasizes preventing the technical subsystem (i.e., data generation) from undermining the social subsystem (i.e., ethical values and practices) by balancing technical capabilities with social imperatives. This approach is based on Mumford's (2006) reflections on the need to align social and technical elements to achieve sustainable outcomes. Hence, Rajkomar et al.'s (2019) insights into the diverse applications of AI in healthcare, including improving patient care and addressing ethical challenges, set the stage for a more informed and balanced adoption of these technologies. The imperative to uphold ethical standards and protect privacy in digital health is further highlighted by Obermeyer et al. (2019) and the importance of addressing bias in healthcare AI to avoid ethical pitfalls, while Panch et al. (2019) inform DR1 in the context of the ethical and practical challenges of implementing AI in healthcare, emphasizing the need for transparency and protection of patient privacy. Therein, the work of Raji et al. (2020) serves as a guiding frame for the moral responsibilities that accompany the deployment of AI technologies, especially those with profound societal impacts. Building up on this, DR1 is further grounded in the framework proposed by Vayena et al. (2019) as their advocacy for a multi-stakeholder approach resonates with the need for collaborative efforts in developing AI solutions that are not only innovative but also ethically sound and respectful of privacy and data security.

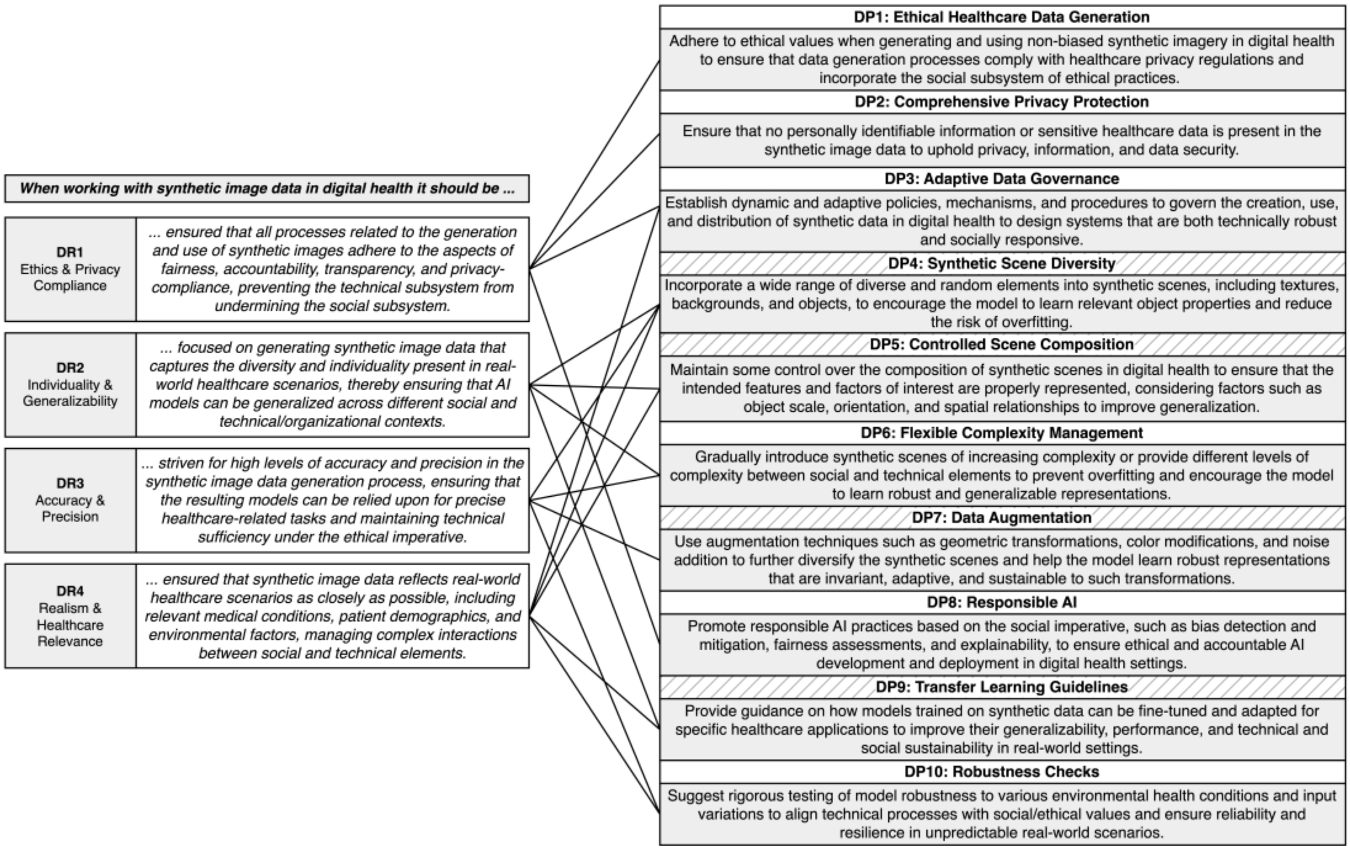
Drawing from *AD2*, **DR2** highlights the importance of capturing the diversity and individuality present in real-world healthcare scenarios (Kern et al., 2022), ensuring that models trained on synthetic data are highly generalizable and applicable to a wide range of situations (Alzubaidi et al., 2021; Seib et al., 2020). These diverse applications can cover areas such as predictive healthcare (Gonzales et al., 2023), evaluation and testing (Murtaza et al., 2023), or classifying diseases (Giuffrè & Shung, 2023), and set the objective of high generalizability in various healthcare settings. The call for comprehensive and diverse datasets (Zhang et al., 2018) further supports DR2, highlighting the need for AI in healthcare to encompass a wide range of patient data to ensure applicability across different patient demographics and conditions, which is consistent with advocacy for AI systems in healthcare that are not only accurate but also versatile in their application to address the nuanced needs of diverse patient populations (Mueller et al., 2018; Panch et al., 2019). Drawing on phenomenological and socio-technical perspectives, Albrechtslund (2007) emphasizes the need for adaptable designs to manage the multistability of technology, while, again, Findeli's (1994) principle of "technoethics" reinforces the ethical obligation to avoid bias in the representation of patient demographics and medical conditions. This further aligns with value-sensitive design's emphasis on autonomy and diversity, which highlights the need for synthetic data models to be generalizable and adaptable to different healthcare scenarios (Deng et al., 2016), as well as the importance of addressing diverse patient values and needs (Dadgar & Joshi, 2018). In addition, the ethical dimension of AI model development, particularly in terms of bias and fairness, as discussed by Raji et al. (2020) and Obermeyer et al. (2019), is integral to DR2. From a sociotechnical systems theory perspective, this means designing data and systems that are responsive to the needs and concerns of patients, healthcare providers, and regulators. This follows Baxter and Sommerville's (2011) emphasis on designing technically robust and socially responsive systems, ensuring that governance frameworks consider the diverse requirements of all stakeholders, thus maintaining the integrity and applicability of synthetic data across different healthcare settings (Rajkomar et al., 2018). In general, the mentioned studies shed light on the need to develop AI tools that are not only high-performing but also fair and unbiased in different healthcare settings. Thus, DR2 embodies the aspect of developing synthetically trained AI models that are ethically sound, highly generalizable, and adaptable to the diverse nature of digital health.



Note: AD = aggregate dimension

Figure 5. Theory Building and Theory Grounding of the Design Requirements

Drawing from both *AD3* and *AD4*, **DR3** addresses the need for high levels of accuracy and precision in the synthetic image data generation process. This is crucial in ensuring that the resulting models can be reliably employed for precise healthcare-related tasks (Aranjuelo et al., 2021; Murtaza et al., 2023; Seib et al., 2020; Valtchev & Wu, 2021, Giuffrè & Shung, 2023), ultimately benefitting patient outcomes in a variety of potential applications (Gonzales et al., 2023). Thus, the importance of high accuracy in healthcare AI applications is driven by the importance of realistic assessments of AI capabilities to prevent potential misdiagnoses or treatment errors (Panch et al., 2019). DR3 further addresses the need to maintain accuracy in AI algorithms, especially in sensitive applications such as facial recognition (Raji et al., 2020), drawing a parallel to the criticality of accuracy in healthcare AI. The issue of competence marginalization in repetitive tasks supports the need for synthetic images to be accurate and capable of handling complex healthcare tasks (Deng et al., 2016). Therefore, it raises awareness regarding potential risks or inaccuracies in AI models, especially in terms of perpetuating biases, which can have serious consequences in healthcare decision-making (Obermeyer et al., 2019). This is further rooted in the potential value trade-off between performance and ethics, and the discourse on pragmatic, ethical, and moral issues (Yetim, 2011; Albrechtslund, 2007) that encompass the appropriate use and impact of security designs (Denning et al., 2014). Given this potential value trade-off, a nuanced understanding and management of both the technical and social subsystems is required, maintaining technical sufficiency in terms of accuracy and precision while focusing on the ethical value imperative as per sociotechnical systems theory (Trist & Bamforth, 1951; Emery & Trist, 1960; Mumford, 2006). Therefore, DR3 captures the essence of producing synthetically trained deep learning models that are not only technically proficient but also reliable and safe in their application in healthcare.



Note: AD = aggregate dimension; DR = design requirement; DP = design principle; Hatched DPs = generally applicable across domains.

Figure 6. The Core Constructs and Simplified Representation of our Design Theory

Lastly, **DR4** draws from *AD2* and *AD4* and emphasizes the importance of ensuring that synthetic image data is reflective of real-world healthcare scenarios (Murtaza et al., 2023), accurately representing relevant medical conditions, patient demographics, and environmental factors through adequate domain adaptation (Kuhnke & Ostermann, 2019; Lahiri et al., 2018; Venkateswara et al., 2017). It stresses the importance of addressing biases in healthcare AI, which resonates with DR4's focus on accurately representing diverse patient populations and conditions (Obermeyer et al., 2019). Herein, the critical examination by Panch et al. (2019) of the realistic application of AI in healthcare underscores the necessity for DR4 to ensure that synthetic image data is not only functional but also aligns with the realistic and ethical considerations crucial in the medical and digital health domain. This further ensures that the technology is not only functional but also ethically sound and user-centric, aligning with value-sensitive design theory (Mueller et al., 2018; Friedman et al., 2002; 2013) and maintaining ethical adaptability (Albrechtslund, 2007; Findeli, 1994). Particularly, the value placed on making a meaningful impact drives the need for synthetic data that accurately reflects real-world healthcare contexts (Dadgar & Joshi, 2018; Deng et al., 2016). Further grounded in sociotechnical systems theory, DR4 addresses the need to manage complex interactions between social and technical elements. Bostrom and Heinen's (1977) analysis of MIS failures through a sociotechnical systems lens reinforces the need to consider both human factors and technical specifications to ensure that synthetic image data accurately represents the complexity of real-world healthcare scenarios and can be effectively used to train deep learning models.

Design Principles

Based on evaluation feedback from the first design cycle, we have revised DP1, DP2, DP3, and DP6 to make them more accessible, easier to understand (in regards to general AI domains), and more nuanced in their connection to digital health. As shown in Figure 4, we added DP8, DP9, and DP10 throughout our heuristic theorizing process to ensure a strong ethical and model performance focus. The addressed connections between the design requirements and principles can be seen in Figure 6. **DP1** draws from *AD1* and states that ethical guidelines and principles should be followed when generating and utilizing synthetic image data. Incorporating value-sensitive design theory (Friedman et al., 2002; 2013) with value embedding (Donia & Shaw, 2021) and meta-ethical guidance (Myers & Venable, 2014; Herwix et al., 2022), it is important to align data generation processes with privacy regulations and to show respect for individual privacy rights (Floridi & Cowls, 2019; Hung & Cheng, 2009; McBride, 2014; Murtaza et al., 2023; Russell & Norvig, 2021; Spiekermann et al., 2022). Following Donia and Shaw's (2021) emphasis on designer agency and Findeli's (1994) concept of "technoethics," ethical guidelines should ensure the constraint that every technological act (i.e., using synthetic image data in digital health AI) has an ethical dimension where design decisions are not neutral. DP1 underscores the imperative to recognize that while synthetic data is artificial, it is not inherently private; rigorous oversight is required to prevent the accidental inclusion of sensitive or identifiable information that could compromise privacy based on real-world reference data. Therefore, care should be taken to employ suitable data generation techniques (e.g., via Unity3D) and to refrain from incorporating sensitive information or biases that could potentially compromise the privacy or security of individuals. Informed by the ethical values of DR1, the critical role of transparency in reducing marginalization supports the need for ethical, transparent synthetic image generation processes (Deng et al., 2016; Mueller et al., 2018; Panch et al., 2018). In addition, DP1 addresses Yetim's (2011) advocacy for discourse ethics, supporting transparency and ethical integrity throughout the synthetic data generation process. This means documenting and explaining the decisions made in generating synthetic imagery, such as the selection of source data, the algorithms used, and the rationale behind those decisions (Obermeyer et al., 2019; Panch et al., 2019). Such transparency is critical to building trust among users and stakeholders, and to ensuring accountability in the use of synthetic data (Panch et al., 2018; Raji et al., 2020; Vayena et al., 2018). It further ensures the generation of synthetic data upholds the ethical values as stated in DR1, aligning with Mumford's (2006) emphasis on balancing social and technical elements to prevent the technical subsystem from undermining these values.

DP2 also builds on *AD1* and addresses the need for the synthetic image data to contain no personally identifiable information (PII) or sensitive data, which often depends on how the data is generated and processed. As such, it addresses the ethical principles of public interest, informed consent, and privacy (Myers & Venable, 2014). DP2 is based on Albrechtslund's (2007) call for flexible ethical designs and Luhmann's (1996) reflections on societal norms and requires that synthetic data be thoroughly anonymized to prevent re-identification while also adapting to unforeseen or anticipated ethical challenges. In general, the idea of synthetic data is to create a dataset that mimics real-world data without actually containing sensitive information. However, there are scenarios where PII or sensitive data may still be present when closely derived from real patient data that may risk re-identification or contain sensitive metadata, raising privacy and bias concerns. DP2 emphasizes the critical need to recognize that synthetic data, while artificial, is not automatically free of privacy risks; it is essential to apply robust anonymization techniques to ensure that even subtle, potentially identifiable characteristics of the real image reference are not replicated, thereby protecting against re-identification and privacy breaches. As such, concerns about security and exploitation align with the importance of excluding identifiable information from synthetic data (Deng et al., 2016). Techniques such as imperfect anonymization (e.g., altering key features in scans) or model inversion attacks (i.e., reconstructing patient data) can inadvertently reveal PII. Hence, it is necessary to anonymize or obfuscate any elements that could potentially reveal an individual's identity (Giuffrè & Shung, 2023; Hansen & Baroody, 2020; Hung & Cheng, 2009; Kern et al., 2022). To achieve this, DP2 recommends the implementation of advanced anonymization techniques that go beyond the removal of obvious identifiers such as names or faces (Raji et al., 2020). They extend to subtle features that, in combination, could lead to the identification of an individual (Panch et al., 2019; Vayena et al., 2018). This could include background details, specific patterns, or even color schemes that may be unique to a person's environment or possessions. It is recommended that comprehensive privacy mechanisms be incorporated into the generation and use of synthetic image data, where controlled noise or perturbations are introduced during data generation to prevent individual data points from being distinguished with a high degree of certainty (Seib et al., 2020; Zhang et al., 2018). This approach can protect the privacy of individuals even in the presence of external information.

Based on *AD2*, **DP3** states that mechanisms should be implemented to control and regulate the generation of synthetic image data, such as process frameworks, toolkits, virtual environments, or guidelines (Gonzales et al., 2023; Murtaza et al., 2023). Following DP1 and DP2, DP3 emphasizes that even though synthetic data is artificially generated, it requires strong governance frameworks to ensure that its creation and use are conducted ethically and responsibly, recognizing that artificial data is not inherently free from risks related to privacy, security, and ethical concerns. Advocating the establishment of clear and comprehensive governance structures, these structures should not only enforce regulations but also foster an environment of responsible and ethical use of synthetic imagery (Panch et al., 2019; Vayena et al., 2018). This includes the development of robust process frameworks to guide each stage of data generation, from initial design to final output (Rajkomar et al., 2018), which should be based on common meta-ethical guidance (Myers & Venable, 2014; Herwix et al., 2022). These frameworks should be flexible enough to adapt to different contexts and use cases while maintaining a core set of ethical and privacy standards (Panch et al., 2019; Rajkomar et al., 2019). For example, it makes sense to use automated compliance checks, data anonymization templates, and simulation environments to safely test data generation methods without risking data breaches. Hence, policies and procedures need to be established to govern the creation, usage, and distribution of synthetic data to prevent unauthorized access or misuse, ensuring value-sensitive compliance (Becker & Becker, 2001; Friedman et al., 2002; 2013). In that context, the suggestion of structured deliberation and boundary critique reinforces the need for comprehensive governance frameworks to guide ethical synthetic data generation (Yetim, 2011). As such, the need for ICTs that support patient autonomy and accountability suggests the importance of establishing governance structures that ensure ethical and responsible generation and use of synthetic image data (Dadgar & Joshi, 2018). This design principle underscores the need for socio-technical governance that can flexibly adapt to changes in both the technological landscape and the organizational culture of healthcare institutions (Mumford, 2006; Baxter & Sommerville, 2011).

DP4 stems from *AD3* and specifies that a wide range of diverse and random elements should be incorporated into synthetic scenes, including textures, backgrounds, and objects, to address a variety of security concerns and contexts (Denning et al., 2014). Following Lupton (2017) and Donia and Shaw (2021), such scene diversity must incorporate sociocultural variability to ensure the representation of diverse patient populations and settings without reinforcing biases. By introducing a wide range of diverse elements into synthetic scenes that challenge conventional assumptions and biases (Obermeyer et al., 2019; Raji et al., 2020), models can be trained to recognize and understand objects in a variety of contexts, reducing the likelihood of bias toward specific environments or scenarios and model overfitting. This approach is particularly beneficial for models used in dynamic and unpredictable real-world environments. For example, in digital health, a model trained on a variety of scene elements is better able to recognize medical devices or conditions in a variety of settings, from well-equipped urban hospitals to resource-limited rural clinics (Rajkomar et al., 2019). By varying these factors, the model will be encouraged to learn relevant object characteristics instead of relying on color or other irrelevant cues (Scheck et al., 2020; Seib et al., 2020). To further improve generalization, cross-domain scene randomization should be used, which involves incorporating scene elements from different domains or contexts (e.g., non-healthcare elements in healthcare settings). Introducing unconventional backgrounds, objects, or textures that are not typically associated with the objects of interest can push the model to learn their intrinsic properties, thereby promoting adaptability to real-world scenarios (Seib et al., 2020; Valtchev & Wu, 2021).

DP5 closely connects to DP4 and further relates to *AD3*, stating that while aiming to promote scene diversity (and randomness), it is important to maintain a level of control over the composition of synthetic scenes. Albrechtslund's (2007) and Findeli's (1994) integration of ethical and aesthetic values suggests that synthetic scenes should be carefully composed to balance diversity, adaptability, and realism while respecting ethical dimensions. While it is crucial to introduce variety and randomness, it is equally important to ensure that these elements do not overshadow or distort the primary objects of interest in the scene. This balance is achieved by carefully controlling aspects such as scale, orientation, and spatial relationships of objects within the scene (e.g., the scale and proportions of disease characteristics or living environments). This ensures that the intended features and factors of interest are properly represented, where factors such as object scale, orientation, and spatial relationships should be considered to enhance generalization (Krump et al., 2020; Scheck et al., 2020). Rather than relying solely on changing the appearance of synthetic objects, the focus should be on varying their key features, and changing attributes such as shape, size, material properties, and structural characteristics will challenge the model to learn object representations. By doing so, models are trained to recognize and understand the essence of objects, making them more robust to changes in appearance that might occur in real-world settings. In addition, DP5 recognizes the importance of contextual relevance in synthetic scene composition. Objects should be placed in contexts that are representative of real-world scenarios, even when introducing elements of randomness. This approach ensures that while the model is exposed to a wide range of scenarios, it still learns to associate objects with their typical environments and situations.

DP6 draws from *AD4* and addresses the introduction of synthetic scenes with varying complexity. It emphasizes the strategic introduction and flexible management of complexity in synthetic scenes to optimize the learning process of deep learning models, addressing the value trade-off between security needs and usability (Denning et al., 2014). The approach can start with simpler scenes or a variety of complexity levels, allowing the model to first understand and identify the core characteristics of the objects and factors of interest. As the model becomes more proficient, more complex elements and scenarios can be gradually introduced. This incremental approach helps build a solid foundation before exposing the model to more challenging environments (Alzubaidi et al., 2021; Bird et al., 2020; Seib et al., 2020; Wan et al., 2021). On the other hand, the varied complexity aspect prevents the model from becoming overly specialized in recognizing only simple scenes and instead promotes the development of a more adaptive and versatile learning capability. As such, this design principle emphasizes the importance of managing complexity in a way that is consistent with the socio-technical challenges of digital health and promotes models that are both socially sound and technically feasible (Mumford, 2006; Bostrom & Heinen, 1977). By encountering a wide range of complexity early on, the model is trained to generalize better across different scenarios, increasing its effectiveness in real-world applications. In addition, DP6 emphasizes the importance of monitoring and adjusting model performance as complexity increases. Regular evaluation is necessary to ensure that synthetically trained models are not only able to cope with the increased complexity but also learn effectively from it to be used in the highly precise digital health environment (Panch et al., 2019; Rajkomar et al., 2019). If the model shows signs of struggle or overfitting, the complexity can be adjusted accordingly to find the right balance that promotes learning without overwhelming the model. Finally, DP6 recognizes the need for diversity in the types of complexity introduced. This includes not only quantitative changes (such as more objects) but also qualitative changes (such as different types of interactions or more nuanced object properties). This comprehensive approach to complexity ensures that the model is well-rounded and prepared for a wide range of situations.

DP7 also stems from *AD4* and states that augmentation techniques, such as geometric transformations, color modifications, and noise addition, should be utilized to enhance the diversity of synthetic scenes (Müller et al., 2018; Seib et al., 2020; Zhang et al., 2018). By applying geometric transformations such as scaling, rotating, and mirroring, models can learn to recognize objects regardless of their orientation or position within the scene. This helps build a more versatile model capable of recognizing objects in different spatial arrangements. Color modifications also play a critical role in enhancing synthetic scene diversity. Adjusting brightness, contrast, saturation, and hue can help the model become resilient to changes in lighting conditions and color variations that occur in real-world environments (Seib et al., 2020). This aspect of augmentation ensures that the model's performance is consistent across different visual presentations. By introducing noise through random variations at the pixel level, models are trained to focus on the essential features of an object rather than being misled by minor imperfections or variations in image quality, which correlates with DP4. This type of augmentation is particularly useful in scenarios where the model must perform reliably despite the presence of visual noise, such as in low-resolution images or when dealing with sensor imperfections. DP7 further advocates for the use of more advanced augmentation techniques, such as perspective warping and synthetic occlusion. These methods introduce additional levels of complexity, teaching the model to understand objects even when they are partially obscured or viewed from unusual angles (Seib et al., 2020; Zhang et al., 2018). These techniques simulate real-world variations and assist the model in learning robust representations that remain invariant to such transformations, which further mitigates the risk of model overfitting (Alzubaidi et al., 2021). This design principle reflects the general socio-technical imperative to create synthetic data that can be generalized across different healthcare settings and patient populations, ensuring that AI models are adaptable and sustainable (Emery & Trist, 1960). Nonetheless, DP7 emphasizes the importance of balancing the augmentation process. Over-augmentation can lead to unrealistic synthetic scenes that do not represent real-world conditions, potentially hindering the model's ability to generalize effectively. Therefore, it is critical to find the right mix of augmentation techniques that enhance scene diversity while maintaining realism.

Drawing from *AD1*, **DP8** emphasizes the importance of promoting responsible AI practices, including bias detection and mitigation, fairness assessments, and explainability, to guarantee ethical and accountable AI development and deployment in digital health settings (Alzubaidi et al., 2021; Floridi & Cowls, 2019; Hung & Cheng, 2009; McBride, 2014; Murtaza et al., 2023; Obermeyer et al., 2019; Russell & Norvig, 2021; Vayena et al., 2018). DP8 emphasizes that although synthetic data is used in AI development, it does not inherently eliminate bias or ensure fairness, so it is critical to implement robust bias detection, fairness assessment, and accountability practices. This is critical to maintaining trust and transparency in healthcare AI applications, especially in value-sensitive environments (Friedmann et al., 2002; 2013; Herwix et al., 2022), where the responsibility of AI developers is balanced between technical and social/ethical imperatives (Mumford, 2006, Donia & Shaw, 2021; Luhmann, 1996). As such, the duality of empowerment and marginalization suggests the need for continuous bias detection and mitigation in AI models (Deng et al., 2016; Panch et al., 2018; Rajkomar et al., 2018). Hereby, bias detection and mitigation are key components of DP8, including the use of diverse synthetic datasets for training to ensure that AI models do not inherently favor or disfavor any particular group of patients (Obermeyer et al., 2019; Vayena et al., 2018). In addition, continuous monitoring for bias in AI decisions is recommended, along with the implementation of corrective measures when bias is detected. On the other hand, fairness assessments involve evaluating AI models to ensure that they make equitable decisions across different patient demographics (Obermeyer et al., 2019). The goal is to ensure that all patients receive fair and unbiased medical advice, assistance, or diagnoses, regardless of their background. In addition, explainability in AI models is also emphasized by DP8. In healthcare, medical professionals and patients must understand how AI models arrive at their conclusions. This transparency is critical to building trust in AI systems based on synthetic image data, as it allows users to validate the reasoning behind AI decisions and ensures that AI augments, rather than replaces, human judgment (Vayena et al., 2018). From a more indirect perspective, DP8 also calls for ethical training and awareness among those developing and using AI in healthcare. This includes educating AI professionals about the ethical implications of their work and the importance of considering the diverse needs and values of patients.

DP9 stems from AD4 and highlights the need for clear guidelines on how models trained on synthetic data can be fine-tuned and adapted for specific healthcare applications to enhance their generalizability and performance in real-world environments (Murtaza et al., 2023). Its adaptability is based on sociotechnical imperatives (Emery & Trist, 1960), which emphasize the need for systems to evolve and remain effective in dynamic environments. A major focus of DP9 is to establish protocols for domain adaptation. These are techniques that help the model adapt from the synthetic data environment to the nuances and characteristics of real-world healthcare data, tailoring systems to specific patient needs (Dadgar & Joshi, 2018). Such adaptation is critical for models to maintain high levels of accuracy and reliability when confronted with real patient data, which may differ considerably from the controlled conditions of synthetic datasets and artificial environments. In addition to domain adaptation, DP9 highlights the potential of fine-tuning models on real image data. This process involves adjusting the model parameters based on real digital health data to improve its performance and generalizability in clinical, assistive, or predictive settings. Fine-tuning ensures that the model is not only theoretically sound but also practically effective in diagnosing and treating real patients. This further ensures a working domain adaptation toward real-image model deployments and fine-tuning tasks on real-image data (Kuhnke & Ostermann, 2019; Lahiri et al., 2018; Venkateswara et al., 2017).

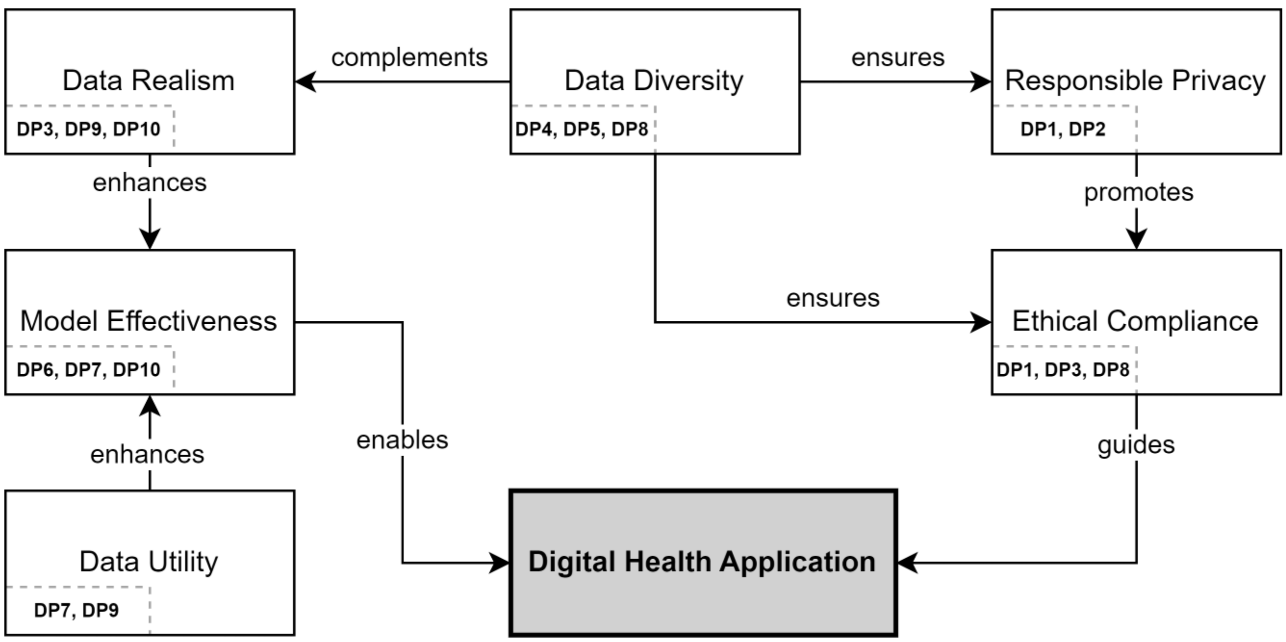
Lastly, **DP10** draws from *AD1* and *AD4*, underscoring the importance of conducting rigorous testing for model robustness against various environmental healthcare conditions and input variations. Emphasized by technological marginalization concerns (Deng et al., 2016), this robustness checking is essential to ensure the reliability and resilience of AI models in unpredictable real-world scenarios, safeguarding against potential errors or malfunctions that could have serious implications for patient care (Floridi & Cowls, 2019; Giuffrè & Shung, 2023; Gonzales et al., 2023; Rajkomar et al., 2019; Russel & Norvig, 2021; Valtchev & Wu, 2021). Hence, a key aspect of DP10 is the simulation of a wide range of environmental conditions during the testing phase. This includes variations in lighting, background noise, and other factors that could affect the performance of AI models in digital health settings. For example, a diagnostic AI tool should be tested for accuracy on different types of medical imaging equipment and under different imaging conditions to ensure consistent performance of the synthetically trained deep learning model. Following on from this, DP10 also emphasizes the importance of testing AI models on a variety of patient data. This includes data from patients of different ages, genders, ethnicities, and health conditions to ensure that the model will perform reliably across the diverse patient population it will serve (Obermeyer et al., 2019; Rajkomar et al., 2019; Vayena et al., 2018). This type of testing is critical to identify and mitigate any biases the model may have and to ensure equitable healthcare outcomes based on the synthetic image data fed to the model. In these scenarios, particularly in the digital health domain, and in light of a common utilitarian view that often overlooks social complexity (Lupton, 2017), it seems reasonable to stress test AI models to assess their resilience to extreme or rare scenarios. This could involve simulating emergencies or rare medical conditions to ensure that the AI model can handle such cases effectively without compromising accuracy or reliability. Finally, models should be continuously monitored and updated after deployment. As real-world conditions and healthcare practices evolve, AI models must be periodically reassessed and updated to maintain their robustness and reliability. As such, not only should their technical performance be assessed, but also their ethical and social implications (Mumford, 2006; Bostrom & Heinen, 1977), to ensure that any emerging issues or changes in health care standards are addressed promptly.

Following general design theory communication, justification, and cumulation, we adopted the outlined process for design theory "anatomy" proposed by Gregor and Jones (2007). Table 3 provides a summary of this process and shows how we have theorized along it.

Table 3. Components of a Design Theory (Gregor & Jones, 2007)

Component	Description
Purpose and Scope	The current trajectory of synthetic image data utilization in digital health computer vision settings is unguided and inconsistent due to its sudden occurrence in the rapidly evolving deep learning field. As such, this inconsistency may hinder the ethical deployment of such models and their performance in IS digital health applications. We aim to develop a design theory that will not only alter this trajectory but also path the way for ethical, effective, and compliant synthetic image use in such settings. Therefore, the theorized artifact's goals are: ensuring ethical and privacy compliance (**DR1**); focusing on individuality of image data and model generalizability (**DR2**); fostering model accuracy and precision through scene diversity (**DR3**); and ensuring data realism and healthcare feasibility (**DR4**).
Constructs	We conceptualize guiding and framing requirements and principles for the utilization of synthetic image data in digital health computer vision. As such, our proposed design requirements and design principles build the foundation for our conceptual model (Fig. 7), encompassing the constructs of data realism, model effectiveness, data utility, data diversity, responsible privacy, and ethical compliance that are representations of the entities of interest in our design theory.
Design Principles of form and function	Features of current synthetically trained computer vision deep learning models in digital health are fundamentally epitomized by uncontrolled data generation (i.e., potential misrepresentations, privacy data breaches, or ethnical biases), insufficient data and model preparation, overfitting problems, and faulty domain adaptation. The dynamic and rapidly evolving nature of digital health deep learning hence requires novel and contemporary guidance in the form of design knowledge on how to utilize such data in these sensitive environments. Through theorizing, ten relevant design principles were developed to address this goal (Figure 5): ethical healthcare data generation (**DP1**); comprehensive privacy protection (**DP2**); adaptive data governance (**DP3**); synthetic scene diversity (**DP4**); controlled scene composition (**DP5**); flexible complexity management (**DP6**); data augmentation (**DP7**); responsible AI (**DP8**); transfer learning guidelines (**DP9**); robustness checks (**DP10**).
Testable propositions	The design theory offers guidance on how to use synthetic image data in digital health both ethically and effectively. Therefore, the design principles and requirements as the core components of our design knowledge, offer high accessibility, importance, novelty and insightfulness, actability and guidance, and effectiveness. Users can design (synthetically trained) computer vision deep learning solutions in digital health by altering conventional approaches and adhering to our design theory. We view the testable propositions as varying in their degree of generality and as truth statements about our design theory, meaning that these general statements describe expected outcomes (i.e., addressing the DRs) based on the application of the DPs. Thus, we hypothesize that in the context of synthetic image data in digital health computer vision, if System A uses our design theory, it will work or be more ethical and efficient than a System B that does not. On a more granular level, the testable propositions here refer to the relationships between DRs and DPs, as shown in Figure 5, emphasizing that if someone has similar DRs in a different context, the corresponding DPs from our design theory can be applied and their effective relationship can be tested. To give a specific example, if DP4 (Synthetic Scene Diversity) is applied during synthetic image generation, the resulting data will (partially - since multiple DPs address DR3) satisfy DR3 (Accuracy & Precision), leading to improved model performance in diverse healthcare tasks, which can be tested by comparing the accuracy and precision metrics of models trained on diverse versus homogeneous datasets. Thus, the following testable propositions arise: **DR1** → *DP1/DP2/DP3/DP8*; **DR2** → *DP4/DP5/DP6/DP9*; **DR3** → *DP4/DP6/DP7/DP10*; **DR4** → *DP3/DP4/DP5/DP9/DP10*
Artifact mutability	We theorize the framing constructs for the utilization of synthetic image data in digital health as a type of guidance to change the current trajectory in digital health computer vision that works in a complementary relationship to real image-based approaches. Our theorized artifact can be used in various digital health computer vision domains, such as medical imaging, diagnosis, human computer interaction, monitoring, detection, or predictive analytics. In addition, it provides general design principles that can be applied across domains of synthetic image processing, and its applicability is deep learning-model independent.
Justificatory Knowledge	Universal themes of synthetic image data use and value sensitive design theory (Friedman et al., 2002, 2013) together with an initial think aloud session and first cycle evaluation feedback serve as the theoretical and practical foundations of our design theory. The chosen kernel theory and qualitative analysis findings remain consistent across various domains, justifying the derivation of DRs and DPs.

Component	Description
Principles of implementation	The universally design knowledge suggests various implementation criteria to the conventional use of synthetic image data in digital health that will yield ethical and performance outcomes. Thus, the design theory can be used in computer vision project processes as a form of feasible and applicable guiding knowledge. Therefore, we present 20 design features (Figure 8) that illustrate how we operationalized the design theory.
Expository instantiation	We developed a series of DRs and DPs that, alongside a conceptual model for synthetic image use in digital health, show how researchers can apply the proposed constructs to develop ethical and performant computer vision models. Furthermore, we show how the design theory can be used to derive theoretical models or concepts in digital health.



Note: DP = design principles.

Figure 7. Conceptual Model as Abstract Instantiation and Theorizing Step

Conceptual Model

As a result of our heuristic theorizing approach, we conducted various steps of modeling (i.e., Figures 6 and 7) to visually develop different types of representations of the problem solution (Gregory & Muntermann, 2014). Following one of many approaches to graphically depict a conceptual solution schema for design theories (Müller-Wienbergen et al., 2011), we developed a conceptual model based on our design theory for the ethical use of synthetic image data in digital health (Figure 7). Within the theorizing process, such a conceptual model enables a more accurate expression of the underlying assumptions, thoughts, constructs, and implications, and can be seen as an abstract instantiation of the design theory (Bittmann & Thomas, 2013; Schermann et al., 2009). As such, our conceptual model represents the abstract constructs as entities of interest in our design theory (Gregor & Jones, 2007; Kane et al., 2021).

Based on our design theory, and in particular the design principles, the conceptual model graphically depicts how the theorized constructs ultimately affect the application of synthetically trained computer vision models in digital health. Based on DP4, DP5, and DP8, *data diversit*y complements *data realism* (DP3, DP9, and DP10) and ensures both *responsible privacy* (DP1, DP2) and *ethical compliance* (DP1, DP3, DP8). Thus, the conceptual model underscores the importance of a holistic design approach that integrates technical data and model robustness with rigorous ethical standards to promote trust and reliability in digital health settings. In addition, *data realism* and *data utility* (DP7, DP9) both enhance *model effectiveness* (DP6, DP7, DP10) while addressing potential biases to ensure that the use of these models leads to equitable and accurate health outcomes for diverse populations. Finally, *model effectiveness* enables ethical compliance to guide digital health applications, creating a balanced framework where AI-driven solutions not only meet healthcare expectations but also align with ethical values and privacy concerns, paving the way for sustainable innovation in digital health.

Evaluation

To ensure rigor in evaluating our design cycles, the well-established FEDS framework proposed by Venable et al. (2016) was used. The evaluation phase is highly relevant in DSR (Hevner et al., 2004; Venable et al., 2016), as it is necessary to select an appropriate strategic process and determine the constructs to be evaluated. This complementary evaluation approach, while not integral or necessary to the heuristic theorizing process (Gregory & Muntermann, 2014), served to validate the reusability, practicality, and relevance of the design theory, ensuring that it "works" in real-world applications and resonated with stakeholder needs. Given the complexity of human factors in digital health, ethical considerations, and the practical application of technology in healthcare settings, the *Human Risk & Effectiveness* evaluation strategy (Venable et al., 2016) was chosen to address the socio-technical and value-based imperatives of our research (Figure 8).

While our problem-centered research approach was theory-driven (Schoormann et al., 2024; Iivari, 2015), we sought to balance both rigor and relevance in our evaluation strategy. Thus, the goal was to conduct an evaluation episode to complete both design cycles and to move quickly to a summative evaluation result. Despite the relatively small sample sizes of our evaluation episodes, we achieved theoretical, data, code, thematic, and meaning saturation, confirming that further data collection was unlikely to yield new insights. This saturation across multiple dimensions ensured the rigor and robustness of our findings and validated our evaluation approach as both thorough and efficient.

Design Requirement and Design Principle Reusability

To ensure the objectives of feasibility, accessibility, completeness, and applicability, we applied the framework of design principle reusability proposed by Iivari et al. (2021). This framework provides a systematic approach to evaluating the design principles generated during the design cycles, and by assessing the reusability of these principles, researchers can determine their potential for wider application and adoption in similar contexts (Iivari et al., 2021). Since we introduced additional design requirements for our design theory, and the framework was designed for the reusability of design principles, we adapted it to evaluate design requirements as well. Iivari et al. (2021) actively call for adaptations of their framework, for which we have applied it to a different level of design abstraction. Thus, we identified 3 levels of reusability for design requirements (RDR) based on the items in the framework: 1) direct application (items for design principles can be directly mapped to design requirements), 2) item adoption (individual items of a construct can be mapped), and 3) non-applicability (items cannot be mapped to design requirements). Because design requirements embody the objective that design principles address, their evaluation differs to some extent. To ensure content validity based on the framework and questionnaire (Iivari et al., 2021), the constructs of *accessibility* (i.e., understandability, comprehensibility, intelligibility) and *importance* (i.e., real and important problem addressing) were allocated to level 1 RDR and could be asked without tweaking them. The constructs of *actability and guidance*, and *effectiveness* were assigned a level 2 RDR with necessary adjustments. Therefore, we decided to use "sufficient guidance," "sufficient direction," and "sufficient freedom" for *actionability and guidance*, and "usefulness," "design help," and "artifact quality" for *effectiveness*, ensuring content validity by not evaluating items that are explicitly tied to design principles and their empirical validation on these items. Finally, the construct of *novelty and insightfulness* was assigned level 3 RDR and therefore deemed inapplicable due to the goal-oriented formulation of the design requirements. Consistent with our kernel theory of value-sensitive design and beyond the first cycle evaluation, a reconvened qualitative think-aloud session was conducted to address the reusability of the proposed design theory constructs. Therefore, the method of concurrent think-aloud (Van Den Haak et al., 2003) was employed with n=12 AI experts, where the sample size was decided based on the "10±2 rule" for think-aloud sessions (Hwang and Salvendy, 2010). The participants were asked to verbalize their thoughts about the design principles and requirements (to the aforementioned extent) in terms of the reusability categories proposed by Iivari et al. (2021). Table 4 presents the qualitative think-aloud results, including the categories of the reusability framework and the manually aggregated verbalized thoughts of the participants.

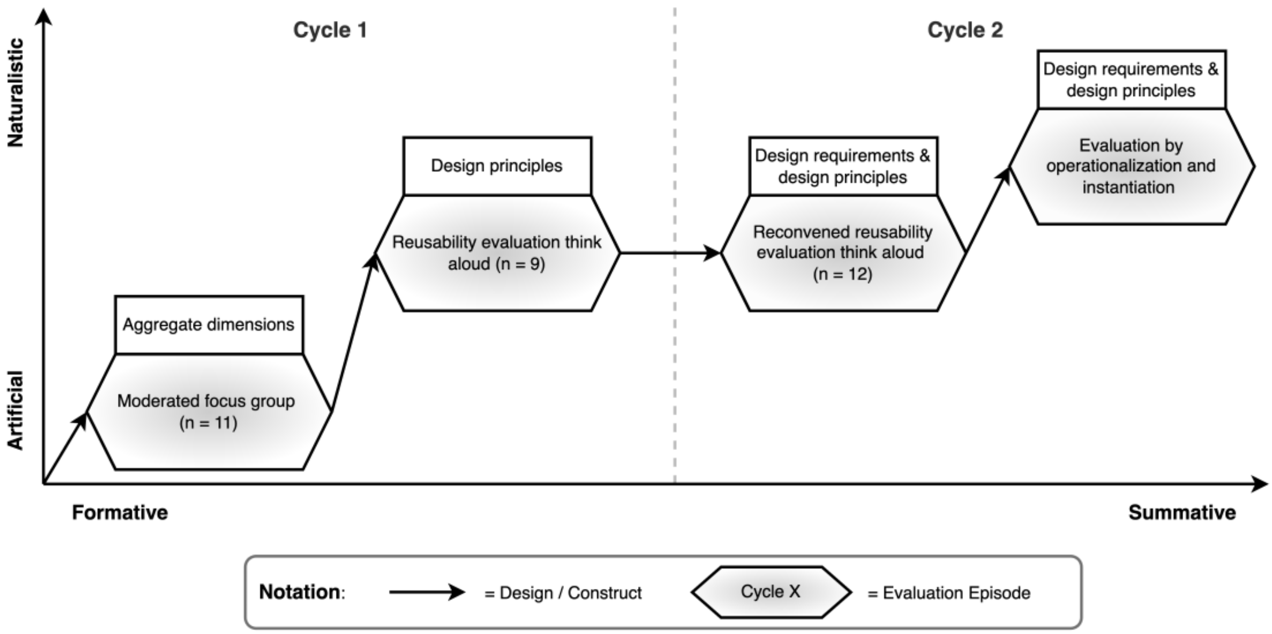


Figure 8. Evaluation Strategy Following the FEDS Framework (Venable et al., 2016)

Table 4. Reusability Framework According to Iivari et al. (2021)

Reusability Category	Design Theory Component	Verbalized Think Aloud Results
Accessibility	**Design Requirements**	The subjects stated the design requirements to be highly accessible in terms of their understandability, comprehensibility, and intelligibility. They particularly highlighted their detailed description and formulation, emphasizing the importance of ethical responsibility in the development and deployment of technologies that handle sensitive data in digital health.
	Design Principles	The subjects appreciated the prescriptive formulation of the design principles, especially the causality of cause and effect, which helped them understand. For some, a shorter description would have been sufficient, but for others, the comprehensive nature of the principles was crucial in ensuring a thorough grasp of the concepts and their application.
Importance	**Design Requirements**	The subjects emphasized that the design requirements address a real and particularly important problem, and highlighted **DR1** and **DR3** as their key requirements. Privacy and accuracy are paramount, and adherence to the DRs would enhance trust and responsible use of synthetic image data, ensuring not only the safeguarding of sensitive information but also the reliability and validity of the data generated for research and development purposes.
	Design Principles	The subjects highlighted the significant importance of DPs related to privacy and responsible AI (**DP1**, **DP2**, **DP3**, **DP8**) and positively assessed their importance. Especially these would address critical foreseeable problems in their professional practice. They acknowledged that incorporating these principles into their workflows would not only mitigate risks but also foster a culture of responsibility and integrity in the handling of such advanced technologies.
Novelty and Insightfulness	**Design Principles**	The subjects appreciated design principles related to synthetic scene diversity, complexity, and composition (**DP4**, **DP5**, **DP6**) as they not only conveyed new ideas to them but also found them insightful. They expressed that these are particularly innovative in regards to synthetic image data generation and its scalability, flexibility, and versatility, emphasizing the potential these principles have in safeguarding the way synthetic image data is utilized across various industries. By incorporating these principles, they foresee a significant enhancement in the quality and applicability of synthetic imagery, paving the way for more advanced, ethical, and diverse applications.

Reusability Category	Design Theory Component	Verbalized Think Aloud Results
Actability and Guidance	**Design Requirements**	The subjects indicated that the DRs would provide sufficient guidance, direction, and freedom for designing and working with synthetic image data in digital health. This would increase standardization across different implementations and lead to reasonably consistent interpretations, thus facilitating a uniform approach to handling data while allowing for innovative uses. The flexibility within the DRs was seen as a key factor in promoting creativity and adaptability in digital health solutions, ensuring that evolving needs and challenges in the field can be met effectively.
	Design Principles	The subjects found design principles related to data generation and domain adaptation (**DP7**, **DP9**, **DP10**) most useful in terms of guidance for designing. They expressed that the majority could easily and realistically be carried out in practice. However, some concerns were raised about restricted design freedom due to the number of DPs, suggesting that a more streamlined set of principles might enhance creativity and experimentation. Some of the subjects indicated a need for a balance between structured guidance and creative flexibility to foster innovative approaches in synthetic image data generation and domain adaptation.
Effectiveness	**Design Requirements**	The subjects expressed a high degree of usefulness and design assistance in terms of artifact (i.e., the design requirements) quality. They mentioned that using the DRs could potentially lead to more effective design and development when working with synthetic image data. They appreciated the comprehensive nature of the DRs, which they believed would not only enhance the quality of design outcomes but also contribute to a deeper understanding and better implementation practices in the field of synthetic image data and digital health.
	Design Principles	The subjects positively emphasized the prevention of ethical/privacy risks (**DP1**, **DP2**) and ensuring exemplary performance (**DP4**, **DP5**, **DP9**, **DP10**), thereby increasing their performance, productivity, effectiveness, and quality at a given task. They noted that the DPs act as a form of guiding knowledge when they need to design such digital health applications efficiently and effectively. However, they expressed concerns about whether DPs would improve the reputation and morale of the organization/company.

Overall, the think-aloud subjects rated the design requirements and principles positively in terms of their theoretical and practical reusability, especially regarding accessibility, importance, novelty & insightfulness, actability & guidance, and effectiveness. Emphasizing digital ethics and precision, these evaluation results indicate a well-founded acceptance and potential for integration of the proposed design theory artifact into existing and future practices within synthetic image data and digital health. The consistent acknowledgment of the detailed and actionable nature of the design requirements and principles by our think-aloud subjects indicates a promising trajectory for the successful application of this theory, particularly in terms of construct understanding and meaning. The articulated concerns regarding the balance between guidance and design freedom, as well as a more nuanced construct description, point to an area for further refinement to ensure that the principles facilitate innovation while maintaining a clear ethical and performance-oriented framework and guiding language.

Evaluation by Operationalization and Instantiation

In addition, the evaluation schema employed in this study takes into account the roles of key stakeholders involved in the formulation of design principles, allowing design science researchers to assess the usability of generated design principles for different user groups. Two critical questions arise from this perspective: first, whether the design principles are theoretically and practically useful, and second, whether they effectively serve the goals of users who implement the resulting instantiations (Gregor et al., 2020). Therefore, *evaluation activity 3* (Sonneberg & vom Brocke, 2012) was used, which describes a validated artifact instance as proof of the applicability of the design theory. To ensure the feasibility and operationality of our design artifact, the evaluation method of *demonstration with a prototype* was chosen (Sonneberg & vom Brocke, 2012).

DP1 Ethical Healthcare Data Generation	**Design Feature 1:** Used data privacy laws and value-sensitive design to guide data generation.
	Design Feature 2: Used non-genuine characters and a variety of ethnicities, cultures, etc. for data generation.
DP2 Comprehensive Privacy Protection	**Design Feature 3:** Used positional reasoning via bounding box coordinates rather than visual representation.
	Design Feature 4: Anonymized subtle features such as background details, specific patterns, and colour schemes.
DP3 Adaptive Data Governance	**Design Feature 5:** Automatically reviewed all generated synthetic images against a set of predefined standards.
	Design Feature 6: Developed a policy document that outlines the ethical creation and usage of that image data.
DP4 Synthetic Scene Diversity	**Design Feature 7:** Used a script for Unity3D to automatically generate random and compliant scenes.
	Design Feature 8: Used a script for Unity3D to randomly add or modify objects, backgrounds, and textures.
DP5 Controlled Scene Composition	**Design Feature 9:** Used scene balancing in Unity3D to maintain the correct scale and orientation of structures.
	Design Feature 10: Used an algorithm that maintains appropriate spatial relationships between objects in a scene.
DP6 Flexible Complexity Management	**Design Feature 11:** Used a script that introduces both quantitative and qualitative complexity changes.
	Design Feature 12: Used an algorithm to ensure that models don't become overly specialized in simpler scences.
DP7 Data Augmentation	**Design Feature 13:** Used a tool that applies geometric transformations such as scaling, rotating, and mirroring.
	Design Feature 14: Used a script that adds controlled noise at pixel level and perspective warping.
DP8 Responsible AI	**Design Feature 15:** Implemented an AI bias monitoring algorithm to ensure that it does not favor or disfavor groups.
	Design Feature 16: Used an AI model evaluation tool to ensure equitable decisions across patient demographics.
DP9 Transfer Learning Guidelines	**Design Feature 17:** Developed a protocol to guide the adaptation of synthetic models to real-world data nuances.
	Design Feature 18: Implemented a process to fine-tune models on real-image data and assessing performance.
DP10 Robustness Checks	**Design Feature 19:** Developed a testing protocol that simulates a wide range of environmental conditions.
	Design Feature 20: Conducted stress tests to evaluate model resilience against unusual or rare conditions.

Figure 9. Operationalized Design Principles as Design Features for Our Prototypical Instantiation

As shown in Figure 9, we operationalized the design principles as design features to validate their feasibility in a real-world application scenario. This iterative process ensured that the design features were not only consistent with the theoretical underpinnings of the design theory but also met the practical needs of users in a dynamic healthcare environment. As a result, the prototype demonstrated the practicality of the design principles and bridged the gap between theory and practice. The design features shown in Figure 9 serve as a transparent description of how we implemented the design theory for a specific digital health scenario, where we used the synthetic image data as a means to locate people with amnestic mild cognitive impairment, monitor their health status, and project appropriate information at the right place and time (see Figure 10). It is meant to be an illustrative example that shows how the design principles can be operationalized by specific actions we took to instantiate the design theory. These design features can be seen as possible characteristics and actions when using the design theory, but they are not meant to be mandatory.

In addition, Figure 10 shows the aforementioned exemplary instantiation of the design principles, operationalized by the design features, in a computer vision deep learning setting for person detection and reasoning, which is therefore concerned with ethical data generation. This example detects people and objects to introduce a seamless, indirect, and intuitive approach to human-computer interaction, where the display of information is triggered when the person moves into a certain area of the living environment. In comparison to the examples of synthetic image data in digital health (Figure 2), i.e., cancer detection and skin conditions, our illustrative example shows a very attenuated form of application, where model errors are certainly not as drastic as in disease diagnosis. Hence, the design principles for such critical applications are even more relevant and decisive.

By following value-sensitive design theory, abstract and non-genuine characters were generated for the scenes, ensuring ethical data generation (DP1) and privacy preservation (DP2). In addition, data gathering and processing for zone alignment only refers to the coordinates of the person's bounding box, not their actual appearance (see Figure 10 and the white rectangle), further ensuring that no sensitive data is displayed or actively used for processing. Using a video game engine and various mechanisms to generate the synthetic image data (DP3), a wide variety of scenes (DP4) and compositions (DP5) were achieved, significantly and intentionally varying the key features. Figure 10 shows the aforementioned domain gap and its adaptation, where our model trained on synthetic image data is applied to real scenes. In addition, several scenes contain varying complexity (DP6) to achieve better generalization and scalability, while data augmentation techniques (DP7) in the form of geometric transformations (i.e., lens distortion) were used to reduce the risk of overfitting. Prior bias detection, e.g., ethnicity or gender, was performed for data generation (DP8), and transfer learning (DP8) and robustness checking (DP10) strategies were employed. By incorporating these design principles that were operationalized by our design features, our instantiation ensures ethical data generation practices in person and activity detection, promoting accurate and reliable results while considering privacy, diversity, generalization, and controlled complexity. Exaggerated by our blatantly privacy-invasive example for positional reasoning in home care settings, the use of ethically generated and applied synthetic image data, properly operationalized by the design features, has the potential to mitigate these privacy-related concerns while still meeting the ethical requirements associated with digital health applications.

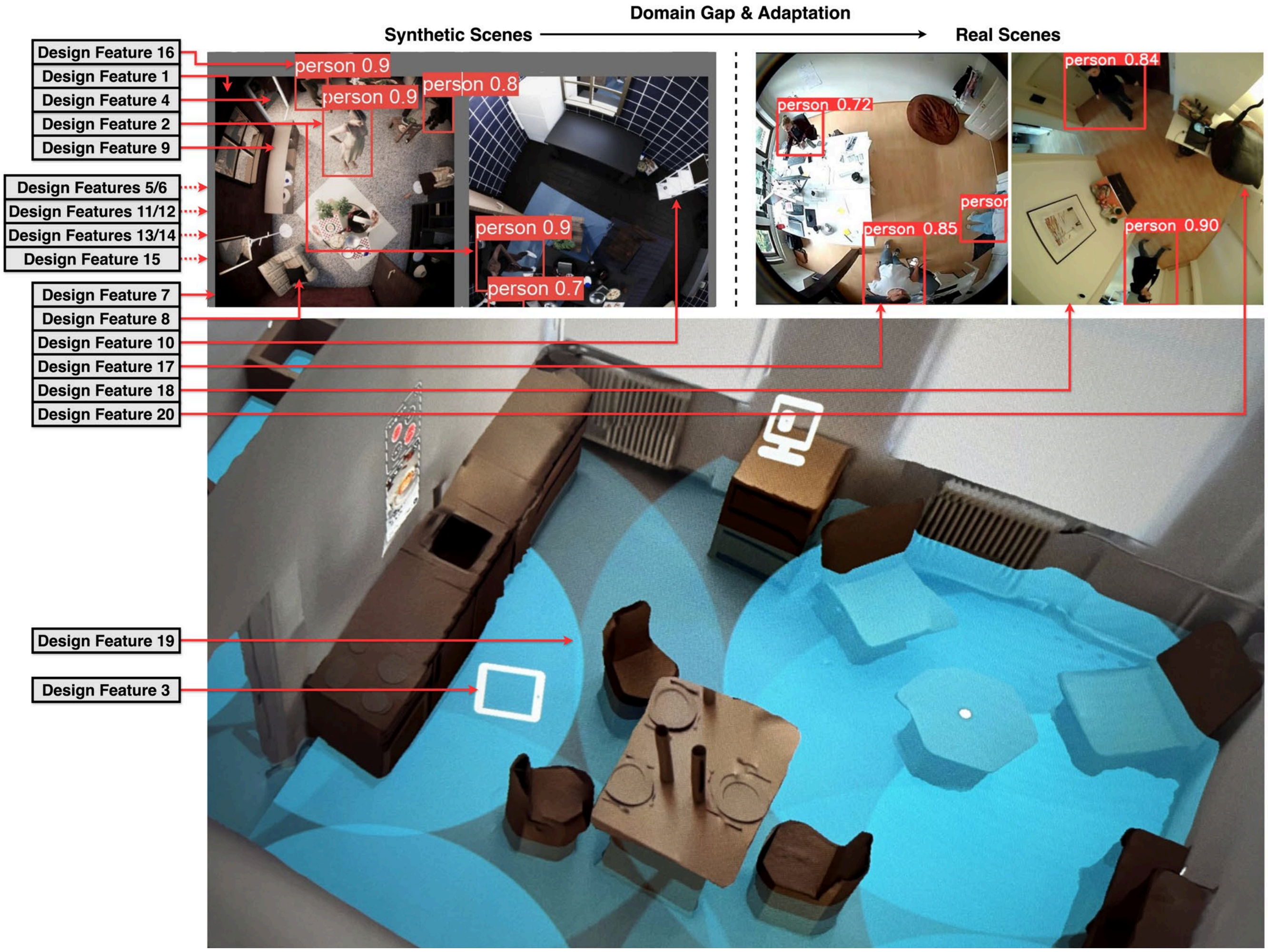


Figure 10. Example of Ethically Generated Synthetic Image Data for Digital Health and Corresponding Design Features From the Instantiation

Note: Dashed lines indicate design features that are not directly visible from this illustration.

Discussion

In this paper, we articulate a design theory that addresses the ethical and technical challenges of using synthetic image data in digital health, particularly for training deep learning models in computer vision. Our core argument is straightforward: if the social and ethical dimensions of health care are not integrated into the technical design of synthetic data systems, the resulting AI models may inadvertently perpetuate biases, compromise patient privacy, and ultimately undermine the very goals they are intended to serve. We are not suggesting that the developers of these systems are acting with malicious intent. Rather, we argue that without a concerted effort to align technical capabilities with ethical imperatives, the deployment of such systems could lead to significant and unintended ethical consequences. This dialog is not only about guiding academic research but also about influencing the policies and frameworks that will determine how these technologies are deployed in real-world settings. The ethical and technical challenges, especially inherent in this sensitive domain, are significant, and our design theory addresses them by proposing a framework that balances ethical integrity with technical performance, particularly in domain adaptation. The overall positive evaluation of our design theory, as evidenced by high levels of agreement across various reusability categories, suggests that our artifact is both feasible and relevant. We argue that the design requirements and principles we developed not only add to the body of IS knowledge but also have the potential to change the way we approach the design and implementation of synthetic image data systems in healthcare. By synthesizing these requirements and principles into an inter-field design theory, as suggested by Darden and Maull (1977), we contribute to the prescriptive knowledge base within the IS community, consistent with the frameworks proposed by Gregor and Hevner (2013) and Woo et al. (2014).

Further locating the theorized artifact within the DSR knowledge framework of Gregor and Hevner (2013), it can be described as a level 3 well-developed design theory about the embedded phenomenon of ethically sound use of synthetic image data in digital health computer vision. While this design theory is presented as a comprehensive theorized artifact, it can also be interpreted as a nascent design theory (Gregor & Hevner, 2013), depending on the lens or context through which it is viewed. We argue that positioning the design theory as a comprehensive framework is more consistent with its multi-cyclical approach to addressing the ethical and technical challenges of synthetic data in digital health, which transcends the characteristics of a nascent design theory. Since our study is located within the *theorizing* area of the DSR focus matrix (Brendel et al., 2022), and according to the classification of DSR artifacts by Gregor and Hevner (2013), the DT can be seen as *Design Theory*, whereas its instantiation as conceptual model is epitomized as *Model*. From a more meta-theoretical perspective, our design theory artifact embodies a *midrange design science theory* (Kuechler & Vaishnavi, 2012) and general midrange theory in IS research (Young et al., 2021).

By adopting the heuristic theorizing approach, we sought to reduce replication errors while maximizing transparency, reproducibility, and traceability. Following the design theory replication framework proposed by Brendel and Muntermann (2022), we effectively addressed various critical aspects of replication in the context of our design theory for digital health computer vision using synthetic image data (Table 5). Encompassing design theory replication aspects of problem class and solution space, robust and generalizable constructs, explicit methods and processes, the rationale for design choices, variability and adaptation, and iterative testing and feedback, our design theory is positioned for effective application and future development in the rapidly changing landscape of digital health and synthetic data use, while providing a theoretical contribution to the field.

Theory Building in DSR

Our study contributes to the accumulation of knowledge in DSR by focusing on midrange theories (Offermann et al., 2011), specifically within the IS design theory genre (Peffers et al., 2018), while aligning with the broader concept of midrange theory in both general theorizing (Young et al., 2021) and DSR (Kuechler & Vaishnavi, 2012). The design theory developed in this study is testable across different populations and settings, although it is not universally applicable. It clarifies the design process and its outcomes by following Gregor and Jones' (2007) design theory framework (Table 3), which includes the essential components of a design theory, addressing user needs (meta-requirements) and the design's response to these (meta-design). This approach bridges theory and practice and emphasizes the importance of design principles in communicating actionable design knowledge (Meth et al., 2015; Chandra et al., 2015). We further address the challenge of design theory indeterminacy by balancing abstraction with actionable principles, following the concerns of Lukyanenko and Parsons (2020). Using the schema of Gregor et al. (2020), along with supportive and prescriptive methods (Möller et al., 2020; Fu et al., 2016), as well as the presentation of application-oriented design features, we mitigate such indeterminacy in the development of design theory.

Table 5. Design Theory Replication Aspects According to Brendel and Muntermann (2022)

Design Theory Replication Aspects	
Replication Aspect	**Our Implementation**
Problem Class & Solution Space	The design theory addresses the problem of ethical and effective utilization of synthetic image data in digital health, focusing on privacy, ethical compliance, bias mitigation, domain adaptation, generalization, and accuracy. The solution space includes design requirements and respective design principles for responsible AI use in digital health computer vision and enhancing model robustness for synthetic image data.
Robust & Generalizable Constructs	To ensure robust constructs, the design theory is based on value-sensitive design and sociotechnical systems theory, incorporating a multi-method (i.e., focus group, think-aloud sessions, qualitative theory building) approach and heuristic theorizing across two design cycles, presenting design requirements and principles as the main artifact and theorized constructs. These primarily apply to the use of synthetic image data in digital health computer vision but can be utilized and adapted across various deep learning domains. Moreover, the design theory contains generally applicable design principles that should be applied regardless of the specific application domain.
Explicit Methodologies & Processes	The development of the design theory followed a structured approach using a heuristic design theorizing methodology (Gregory & Muntermann, 2014) over two completed design science research cycles. In addition, the theory-driven approach of learning from abstract theoretical knowledge (i.e., our kernel theories and the literature) and translating it to solve a problem (Schoormann et al., 2024) and formulating rigorous design principles (Gregor et al., 2020) for conceptualizing fewer abstract artifacts was applied.
Rationale for Design Choices	The multi-faceted study approach ensures that the design requirements and principles are not arbitrary but are grounded in conceptual and theoretical work (Schoormann et al., 2024). The rationale for design choices emphasizes that ethics and privacy are consistent with value-sensitive design and sociotechnical imperatives, preventing the technical subsystem from undermining the social subsystem in digital health AI. We argue that design decisions are never neutral, prioritizing individuality and generalizability to ensure our models' adaptability across diverse healthcare contexts. A focus on accuracy and precision in data generation underpins the reliability and effectiveness of healthcare applications. Finally, our emphasis on realism and healthcare relevance in synthetic data underscores our commitment to creating technically and ethically sound, user-centered solutions.
Variability & Adaptation	Proactively anticipating the need for variability and adaptation in the rapidly evolving nature of both digital health and synthetic image data, we provide guidelines for adapting our theory to evolving technologies and ethical values, ensuring ongoing relevance. Our approach accounts for future advances and changes in the field, both technically and ethically, and maintains the applicability of the theory in diverse and changing scenarios by not only proposing design principles that are generally applicable (e.g., DP4) regardless of the target domain, but also by choosing a wording that is invariant to future developments and maintains clarity (e.g., DP1 - ethical guidelines and laws may change, but adherence to them should not). This foresight underscores our commitment to a flexible, responsive design that can adapt to new ethical imperatives and technological developments, reinforcing the theory's robustness and longevity.
Iterative Testing & Feedback	Given the nature of DSR, we emphasized the importance of iterative testing and feedback, advocating a continuous refinement process beyond the two cycles presented. This iterative approach allows for the consistent incorporation of new insights and technological advancements, ensuring the design theory remains relevant and effective. By incorporating reconvened think-aloud sessions addressing the design theory's constructs, we categorize their reusability in terms of accessibility, importance, novelty & insightfulness, actability & guidance, and effectiveness (Iivari et al., 2021). Hence, we specifically targeted design theory indeterminacy (Lukyanenko & Parsons, 2020), enabling the theory to be applied in various contexts with minimal modifications, extended to cover broader scenarios, and scaled to accommodate different levels of use.
Result Documentation & Sharing	In our design theory, we underscore the significance of meticulously documenting and sharing comprehensive data and results by grounding the design theory's constructs with their theory-grounding and theory-building foundations. Therefore, we chose to unfold our qualitative content analysis following Gioia et al. (2013) to ensure traceability and comprehensibility throughout the design process. This commitment to transparency allows other researchers and practitioners to replicate, validate, and extend our findings, guided by our work. We ensure that every aspect of our research process, from initial theory-grounding, theory-building, and formulation of design requirements and principles to iterative reusability evaluation results, is thoroughly recorded and accessible. By sharing detailed data and results, we foster a deeper understanding of the development and application of our theory and encourage collaborative improvement.

Our approach emphasizes that the projectability of design theories - their applicability in different contexts - determines their relevance and utility in DSR. We argue that while generalizability often emphasizes broad applicability, projectability shifts the focus to ensuring that design principles can be effectively applied in specific but diverse scenarios. As the empirical justification of design principles or theories in DSR is based on their prescriptive projectability rather than descriptive generalizability, this concept is especially critical in the rapidly evolving fields of AI and digital health, where the ability to adapt theories to new technologies and ethical considerations is paramount. By framing projectability as a prescriptive alternative to traditional generalizability (Baskerville & Pries-Heje, 2019), we emphasize its importance in the broader DSR landscape. In this context, high projectability suggests that our design theory is not only applicable within the initial domain of synthetic imagery but is also adaptable to future advances and other domains (vom Brocke et al., 2020). On the other hand, the replication aspect of our design theory (Table 5), in line with the framework proposed by Brendel and Muntermann (2022), further ensures the robustness of the design theory. Nonetheless, the traditional aspects of generalizability and transferability of our design theory are ensured by its emphasis on fundamental design principles that are broadly applicable across domains. Specifically, DP4, DP5, DP7, and DP9 are constructed to be adaptable and effective beyond the initial domain of digital health. These principles are grounded in AI literature (e.g., Alzubaidi et al., 2021; Giuffrè & Shung, 2023; Gonzales et al., 2023; Murtaza et al., 2023; Scheck et al., 2020) and ensure that the synthetic image data generation process captures diverse and varied scenarios, supporting the applicability of the theory across contexts (e.g., fall detection, autonomous driving, or robotics). Interestingly, this could also have implications for the use of generative AI and the feeding of computer vision models with AI-generated (synthetic) images (which would mean that AI applications feed AI applications, the vicious cycle of AI collaboration). As AI models increasingly rely on synthetic data generated by other AI systems/models, we could see the emergence of self-sustaining AI ecosystems that operate independently of human input, which could raise concerns about the oversight and control of such systems – and, as such, require ethical safeguards.

Furthermore, our study contributes to next-generation theorizing in IS research (Burton-Jones et al., 2021; Young et al., 2021) by envisioning a new (design) theory that studies a new phenomenon (i.e., synthetic image data) emerging in a changing world (i.e., the digital health and ethics domain), especially from a DSR focus (Brendel et al., 2022) to adapt to concurrent and emerging research conversations. We extend the application of Iivari et al.'s (2021) design principle reusability framework by proposing three levels of reusability to evaluate design requirements, thereby advancing theoretical evaluation approaches within the DSR domain. We argue that this may lead to more nuanced and multifaceted evaluations of design theories in the future, in line with next-generation theorizing (Burton-Jones et al., 2021). In this context, we advocate for a transparent and replicable approach to design theorizing, emphasizing the importance of heuristic theorizing across *multiple* design cycles alternating between (re)structuring the problem and generating new design constructs (Gregory & Muntermann, 2014; Brendel & Muntermann, 2022). Finally, with our design features and instantiation, we actively worked to mitigate the phenomenon of *theory fetishism* in IS (Ivari, 2020) by promoting a judicious use of theory that bridges the gap between abstract theoretical concepts and practical design applications.

The Value-Sensitive Design Perspective

In the context of our theoretical lens and two kernel theories, our design theory can enrich the theoretical discourse around the kernel theory of value-sensitive design as conceptualized by Friedman et al. (2002, 2013), particularly in the context of the use of synthetic image data in digital health. By addressing the ethical and practical challenges associated with deep learning models, this research not only aligns with the foundational constructs of value-sensitive design but also extends them in meaningful ways. Central to value-sensitive design is the integration of human values such as privacy, fairness, and accountability into the design process. Here, we extend Stilgoe et al.'s (2020) concept of *responsible innovation* by integrating these ethical values into the core design process in the context of IS and DSR, thereby offering a practical application that takes responsible innovation beyond theoretical discourse and aligns with its *anticipatory dimension*, which emphasizes the importance of anticipating the potential impacts of emerging technologies such as AI and synthetic data. Our design theory builds on this by embedding these ethical considerations into the design of synthetic image data systems to ensure that privacy is protected, bias is mitigated, and fairness is promoted, resonating with the principles emphasized by Owen et al. (2013), who advocate for anticipatory, reflective, and inclusive practices in innovation. In addition, our design theory extends the theoretical discourse around value-sensitive design by addressing the need to capture the diversity and individuality inherent in real-world healthcare scenarios, ensuring that AI models apply to diverse populations. This commitment to generalizability and inclusivity is informed by value-sensitive design's emphasis on respecting stakeholder diversity and autonomy, as highlighted in the work of Yetim (2011) and Dadgar and Joshi (2018). Furthermore, our design theory advances value-sensitive design by providing concrete operational steps for implementing ethical principles, particularly in managing complexity and ensuring data realism in synthetic image generation. This practical focus is in line with Denning et al. (2010) and Mueller et al. (2018), who emphasize the importance of balancing technical accuracy with ethical considerations in AI. By introducing structured heuristics for resolving value conflicts - such as those between precision and transparency, or privacy and accountability - our artifact provides a more nuanced approach to integrating ethical values into technology design, thereby expanding the applicability of value-sensitive design in complex ethical landscapes. In addition, the generalizability of the design theory across different AI and computer vision contexts extends the reach of value-sensitive design theory, providing a framework that can adapt to different industries while maintaining its relevance in the evolving landscape of AI and healthcare.

The Sociotechnical Systems Perspective

From a sociotechnical systems theory perspective, our design theory may contribute to the specific demands of digital (health) environments, particularly in how we understand the dynamic interplay between social and technical systems in applying synthetic image data and its ethical imperatives. We drew on the concept of joint optimization (Trist & Bamforth, 1951), emphasizing that ethical considerations (the social subsystem) and the technical requirements of deep learning models (the technical subsystem) must be addressed simultaneously to ensure robust yet ethical AI applications. Our design theory, further informed by Mumford's (2006) perspective on sociotechnical systems, ensures that ethical values (as defined per DR1), privacy protections, and responsible AI practices are not afterthoughts but are embedded at the heart of the design process. For instance, while synthetic image data offers the promise of scalable and diverse datasets necessary for robust AI models, it is the careful calibration of these datasets - through mechanisms such as synthetic scene diversity (DP4) and controlled scene composition (DP5) - that prevents the oversimplification or distortion of real-world healthcare scenarios. Our study thus extends the discourse of sociotechnical systems theory by applying its constructs to the dynamic and evolving field of (synthetically trained) AI in digital health. As such, our design theory not only accommodates but also anticipates the ethical challenges that arise as digital health and AI technologies evolve. This approach is consistent with Bostrom and Heinen's (1977) emphasis on integrating behavioral and organizational considerations into system design, ensuring that the social implications of synthetic data use are addressed as rigorously as the technical ones. Moreover, our design theory goes beyond simply applying the principles of sociotechnical systems theory and actively extends them. For example, Baxter and Sommerville's (2011) concepts of affordances and constraints are used to refine the interaction between users and synthetic data systems, ensuring that these systems are not only technically sound but also socially responsive. This nuanced approach allows us to address the tensions between ethical imperatives and technical requirements, and to propose a framework in which both can be optimized simultaneously, thereby avoiding the pitfalls of ethical negligence. We argue that ethical integration is not simply a matter of adding ethical safeguards; it requires a fundamental rethinking of how we approach the design of digital health systems, especially when the generation of synthetic data is just a click away. This will ensure that such systems are as focused on the social and ethical complexities of healthcare as they are on the technical demands of AI.

Implications for Synthetic Image Data and AI Ethics

Besides methodological and kernel theory perspectives, our design theory adds new perspectives and specific domain applications to synthetic image data in computer vision (Alzubaidi et al., 2021; Aranjuelo et al., 2021; Kuhnke & Ostermann, 2019; Scheck et al., 2020; Seib et al, 2020; Valtchev & Wu, 2021), and introduces and extends such phenomena in digital health (Giuffrè & Shung, 2023; Gonzales et al., 2023; Murtaza et al., 2023) with an explicit focus on digital ethics (Kern et al., 2022; McBride, 2014; Russel & Norvig, 2021). Our study responds to the call for more measures to protect patient well-being and maintain ethical standards while working with synthetic data in healthcare (Giuffrè & Shung, 2023). While Giuffrè and Shung (2023) laid a strong general technical and regulatory foundation for synthetic data in healthcare, our study extends this work by building design knowledge that addresses the socio-technical and critical ethical dimensions, providing a theoretical framework to guide the responsible and equitable use of synthetic data in digital health applications. We argue that while synthetic data is often perceived as inherently private due to its artificial nature, it still has significant ethical implications that must be addressed through rigorous theoretical and practical frameworks. Thus, our design theory challenges the assumption that synthetic data is free from privacy and bias risks and highlights the need for robust mechanisms to prevent re-identification, mitigate bias, and ensure fairness in AI applications.

Moreover, as synthetic data generation becomes more accessible and automated, the potential for its misuse increases, particularly when assumptions of inherent privacy lead to a lack of critical oversight. We argue that without robust safeguards, synthetic datasets may propagate systemic biases or be exploited in ways that compromise the very ethical standards they are intended to uphold, underscoring the urgency of a proactive and multifaceted governance approach. By operationalizing the ethical values into actionable design principles, our research provides a structured approach to navigating the complex ethical landscape of AI in digital health, ensuring that synthetic image data is used responsibly and ethically. Specifically, it contributes to the discourse on *ethical AI* in healthcare. While existing literature (Mittelstadt et al., 2016; Spiekermann et al., 2020) primarily focuses on the ethical implications of AI, our design theory goes a step further by operationalizing these ethical considerations. As such, our design theory makes a distinctive contribution by addressing the specific ethical risks associated with synthetic image data in digital health through actionable design principles, complementing the broader meta-level frameworks of Myers and Venable (2014) and Herwix et al. (2022), which provide valuable general guidance but do not fully encompass the specific socio-technical challenges in this particular domain, and shifting the conversation from what should be done to how it can be done. Thus, narrowing the scope can enhance value, particularly in sciences that prioritize practical applicability, but if the IS community typically favors a broader scope, we risk lacking the detailed information needed to fully understand IS phenomena (Siponen et al., 2023).

Furthermore, our study draws on Floridi and Cowls' (2019) framework of AI ethics, which emphasizes the importance of AI being beneficial rather than detrimental to humanity. This aligns with our focus on developing AI systems that improve health care while mitigating risks such as bias, privacy violations, and ethical dilemmas. We also integrate insights from Topol (2019) and therefore address the discourse on *deep medicine*, exploring how AI can transform healthcare by personalizing medicine, increasing diagnostic accuracy, improving patient care, and ensuring that our AI designs contribute positively to these aspects. Finally, our design theory advances AI in healthcare by providing a structured, ethical framework for using synthetic image data, addressing key concerns like privacy, bias mitigation, and ethical compliance (Obermeyer et al., 2019; Panch et al., 2019; Rajkomar et al., 2019). It builds on Panch et al.'s (2019) discussion by offering actionable principles, responds to Obermeyer et al.'s (2019) findings on racial bias, and aligns with Raji et al.'s (2020) and Vayena et al.'s (2018) emphasis on ethical challenges and fairness in medical deep learning. Thus, our research addresses the concerns raised in AI and digital health systems about the potential risks of AI exacerbating existing inequities if not properly managed (Panch et al., 2018). In response to calls for transparency in the design, deployment, and evaluation of deep learning models to build trust and accountability (Rajkomar et al., 2018), we advocate for clear documentation and explanation of data generation processes, model design decisions, and ethical considerations to ensure that models do not perpetuate disparities.

Integrating related theoretical perspectives from philosophical, sociotechnical, and phenomenological design ethics deepened our understanding of the moral and ethical responsibilities inherent in using synthetic data for digital health. Donia and Shaw (2021) highlight that design decisions are shaped by the agency of designers, who must navigate external pressures (often commercial and technical) while embedding ethical values in their work. Thus, our study may reinforce the idea that design is never neutral, as moral implications are inherent in every decision, especially in sensitive areas such as healthcare. Findeli's (1994) concept of "technoethics" further emphasizes that this responsibility for ethical implications in system design goes beyond mere functionalism. We argue that adopting a more speculative and adaptive approach to design - one that anticipates and responds to the broader societal impacts of technology - is critical to addressing the evolving challenges posed by synthetic data (Lupton, 2017; Burton-Jones et al., 2021). Through the lens of our sociotechnical systems theory, and as emphasized by Albrechtslund (2006), we contend that the multistability of technology demands ethical adaptability. The unpredictable ways in which innovative technologies, such as synthetic data, will be used once they are deployed underscore the need for flexible design frameworks that can accommodate unforeseen ethical challenges. Finally, in line with Luhmann's (1996) reflections on ethical systems as adaptive and reflective processes, we contend that ongoing ethical deliberation is essential to ensure that synthetic data systems remain ethically sound as they evolve in the dynamic landscape of health care.

Practical Implications

By immersing ourselves in the practical world of healthcare, we have developed a design theory that is not only grounded in the realities of synthetic data use but also anticipates future challenges. It provides practitioners with a robust framework for the ethical use of synthetic data, not just in terms of theoretical considerations but with actionable insights and guidelines that can be directly implemented in real-world scenarios to reduce design theory indeterminacy (Lukyanenko & Parsons, 2020). We argue that the practical implications of our work go beyond immediate applications and serve as a blueprint for how design theories in digital health can be systematically developed and implemented to address complex ethical challenges, thereby setting a precedent for future innovations in health technology. Importantly, recognizing that synthetic image data is not inherently free of privacy risks is the crucial first step - one that, if widely understood and accepted by practitioners, would constitute a significant part of the effort in safeguarding against potential ethical violations. We advocate a sensible and cautious approach to synthetic data that does not consider it inherently risk-free. However, it is certainly less risky than previous approaches using real image data (because privacy is less of an issue), but the problem remains, and we argue that prevention is better than hindsight, especially when working with sensitive healthcare data.

Subsequently, model inversion and adversarial attacks, while likely rare, remain a critical concern, as advanced AI models trained on synthetic data may still be vulnerable to malicious actors who could reverse engineer the synthetic data to extract sensitive information. Although synthetic data is artificial, models may retain traces of the underlying real data, particularly in the case of unexplained AI models, exposing private details. Therefore, we argue that for synthetic data to be trusted, transparency about its origins and associated risks is essential, as stakeholders may overestimate privacy protections if they are unaware of the close link between synthetic and real data. As the field of digital health continues to evolve at a rapid pace, the implications of our design theory extend beyond academic discourse, fostering a culture of ethical mindfulness that is crucial for sustaining the integrity and trustworthiness of health-related AI applications.

At the organizational level, health authorities face several significant challenges when trying to introduce or implement innovative changes or systems (Thakur et al. 2012). Therefore, the implications of our research reach beyond academic theorizing, holding significant promise for real-world healthcare and management applications, as these authorities can develop and implement ethical, effective, and precise computer vision models based on synthetic image data in digital health. Our study highlights the dynamic nature of DSR in line with the principles outlined by vom Brocke et al. (2022). We show that the development of design theory in this context is not a linear process but one that evolves through collaboration, unexpected challenges, and new opportunities. This iterative process, similar to a "dance" between theory and practice, ensures that our approach remains adaptable to the complexities of digital health and ethics.

Limitations & Future Research

In light of the overall positive evaluation of our theorized artifact, some limitations should be considered. First, design artifacts in the form of design requirements and principles, as well as their development, are tied to the subjective creativity of the researcher, even after various data collection episodes and theory grounding. However, not all design decisions can or should be derived from behavioral or mathematical theories, as some degree of creativity is essential to developing an innovative design artifact (Hevner & Chatterjee, 2010; Baskerville et al., 2016), whereas a certain degree of rigor can be implemented through methodological (Fu et al., 2016; Gregor et al., 2020; Möller et al., 2020) or theorizing approaches (Gregory & Muntermann, 2014; Lee et al., 2011). The iterative nature of DSR often means that the final artifact is the result of multiple refinements, which can lead to a divergence from the initial theoretical underpinnings. While this divergence is a natural part of the design process, it could limit the extent to which the final artifact embodies the theoretical constructs it was intended to. Therefore, maintaining a balance between theoretical fidelity and practical utility in the artifact becomes a critical consideration that we encourage future research to address. In this regard, our study acknowledges the potential for refinement in the abstraction and granularity of its design principles, as highlighted by one of the reviewers during the review process. We would like to take up this idea and see future research exploring strategies for optimizing the conciseness and specificity of design principles, depending on their use and actionability in theory and practice. We believe that initiating this ongoing dialogue is essential to advancing the theoretical and practical discourse on the use of synthetic image data in digital health and beyond.

Second, as with any evaluation, the results describe only one sample throughout two completed design cycles, meaning that different results might be expected if a different sample were chosen. It would be presumptuous to assume that design theory contains all the necessary design knowledge to use synthetic image data in digital health computer vision, although we have tried to provide high transparency and rigor with our theorizing approach. Therefore, we encourage other researchers to challenge, adapt, or refine the constructs of our design theory. Here, we aimed to make the design theory highly replicable by following the framework proposed by Brendel and Muntermann (2022), which emphasizes the importance of a clear and structured approach when developing design theories. Future studies could extend our work by applying the design theory to different contexts or replicating the development of design requirements and principles. Furthermore, despite achieving theoretical, data, code thematic, and meaning saturation, we acknowledge that the specialized nature and size of our expert sample may limit the generalizability of our findings. Therefore, further research with diverse samples and in different contexts with different theoretical underpinnings is essential to validate and potentially extend the applicability of our design theory. Future studies should explore larger and more diverse samples to ensure broader applicability while maintaining the rigor of practitioner feedback within relevant subgroups (Iivari et al., 2021). In this context, the incorporation of organizational change management theories could significantly inform the application and refinement of the proposed design theory for synthetic image data in digital health. Theories such as Armenakis and Bedeian's (1999) organizational change framework, Kotter's (1996) eight-step process for leading change, and Lewin's (1947) change management model could provide valuable guidance on how healthcare organizations could effectively adopt and implement the design principles outlined in our study. Given the diverse contexts, cultures, and regulatory environments across countries and healthcare systems, we argue that these organizational change management frameworks are critical for understanding how to tailor change strategies to specific organizational and national contexts. Work such as that of Markus and Robey (1988) on the interaction between technology and organizational change highlights the importance of understanding the complex dynamics at play when implementing new technologies. These insights are particularly valuable in addressing the various challenges of change acceptance, stakeholder involvement, and alignment of new technologies with existing organizational cultures, especially in complex and culturally sensitive healthcare environments. Future studies could explore these aspects, grounding our design theory in practical organizational contexts and enhancing its applicability across diverse healthcare settings to ensure that the transition to synthetic data-driven AI systems is both smooth and sustainable in diverse global contexts.

Third, while the application of the reusability framework (Iivari et al., 2021) covered categories such as effectiveness, these were not empirically validated but only qualitatively through think-aloud sessions. Even after two completed DSR cycles and a certain degree of maturity, our design theory is at an early stage of development. Given the complexity of the topic, it was essential to focus on conceptual development, as theories in scientific research, particularly in DSR, are built on constructs and propositions that provide explanations and prioritize the utility of prescriptive theories over empirical confirmation (Baskerville & Pries-Heje, 2019; Gregor & Jones, 2007). While we propose testable propositions linking design requirements and design principles, validating these propositions relies on artifact instantiation rather than traditional hypothesis testing, which may limit generalizability. Future research should explore more empirical testing of these propositions in diverse real-world healthcare settings further to assess their applicability and effectiveness across different contexts. As such, we recognize the challenges posed by the rapidly evolving nature of digital health and AI, which means that certain aspects of the design theory may change as technology advances. We aim to validate the design theory longitudinally to ensure its continued relevance, adaptability, and feasibility. Specifically, we plan to implement the design theory in a real-world context (e.g., for fall detection, interaction design, or disease detection) to collect data on various metrics such as performance, user satisfaction, and adaptation. As the environmental contexts evolve, the design theory may be revised as we talk to experts and observe technological advances to ensure its validity. In this context, we plan to conduct a longitudinal case study over the lifetime of the aforementioned implementation. Thus, we will periodically reflect on the data collected and compare it to the original design objectives and any changes made to the design theory. Therefore, we encourage researchers, theorists, and practitioners to apply the design theory to new contexts to see if it remains effective. This could mean using the theory to solve different types of problems in different industries or under different conditions. The results of these applications serve as empirical evidence for or against the projectability of the theory. However, we also encourage future research to apply the design theory and compare the results in different case studies. Especially in deep learning, empirical testing of such items or model benchmarking (e.g., comparing synthetically trained models with ground truth models of real images) is an advisable choice.

Conclusion

The ongoing rapid development in both digital health and computer vision requires innovative and ethically inclusive solutions, which can be achieved, for example, through the generation and use of synthetic image data for model training. While the use of artificial imagery provides faster generation, infinite scalability, and photorealism, it often involves ethical concerns (i.e., bias or invasion of privacy) and error-prone application (i.e., poor domain adaptation). Therefore, it is crucial to design ethical, performant, and precise computer vision models based on synthetic image data, especially in the digital health domain with a high emphasis on accuracy, privacy, and feasibility. By theorizing the causality of how to use such artificial data and what effects are to be expected, we demonstrated a design knowledge guidance that contributes to the field of digital health, computer vision, and digital ethics. Therefore, our design theory not only contributes by its feasibility but also serves as a starting point for future research within these domains.

References

Albrechtslund, A. (2007). Ethics and technology design. *Ethics and Information Technology*, *9*, 63–72.

Al-Dhaen, F. et al. (2021). Advancing the Understanding of the Role of Responsible AI in the Continued Use of IoMT in Healthcare. *Information Systems Frontiers*, 1–20.

Alzubaidi, L., Zhang, J. et al. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*, 1–74.

Aranjuelo, N. et al. (2021). Key strategies for synthetic data generation for training intelligent systems based on people detection from omnidirectional cameras. *Computers & Electrical Engineering*.

Aristotele, Brown, L., & Ross, D. (2009). *The Nicomachean Ethics*. Oxford University Press.

Armenakis, A. A., & Bedeian, A. G. (1999). Organizational change: A review of theory and research in the 1990s. *Journal of Management*, *25*(3), 293–315.

Baskerville, R., & Pries-Heje, J. (2010). Explanatory Design Theory, Business Information Systems and Engineering, 2(5), 271–282.

Baskerville, R., & Pries-Heje, J. (2019). Projectability in design science research. *Journal of Information Technology Theory and Application*, *20*(1), 3.

Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. *Interacting with Computers*, *23*(1), 4–17.

Becker, L. C., & Becker, C. B. (2001). Encyclopedia of Ethics, 2. Routledge.

Bird, J. J. et al. (2020), From simulation to reality: CNN transfer learning for scene classification. *IEEE 10th International Conference on Intelligent Systems*, 619–625.

Bittmann, S., & Thomas, O. (2013). An argumentative approach of conceptual modelling and model validation through theory building. *Proceedings DESRIST 2013*, 242–257.

Böhmer, M. (2023). From Pixels to Generalization: Ensuring Information Security and Model Performance with Design Principles for Synthetic Image Data in Deep Learning. In *CIISR@ Wirtschaftsinformatik*, 49–61.

Bostrom, R. P., & Heinen, J. S. (1977). MIS problems and failures: A socio-technical perspective. Part I: The causes. *MIS Quarterly*, 17–32.

Bowen, G. A. (2008). Naturalistic inquiry and the saturation concept: a research note. *Qualitative Research*, *8*(1), 137–152.

Brendel, A. B., & Muntermann, J. (2022). Replication of design theories: Reflections on function, outcome, and impact. *Information Systems Journal*, *32*(6), 1158–1176.

Brendel, A. B., Lembcke, T. B., & Kolbe, L. M. (2022). Towards an Integrative View on Design Science Research Genres, Strategies, and Pivotal Concepts in Information Systems Research. *The DATA BASE for Advances in Information Systems*, *53*(4), 9–23.

Bryant, A., & Charmaz, K. (2007). Grounded theory in historical perspective: An epistemological account. *The SAGE Handbook of Grounded Theory*, 31–57.

Burton-Jones, A., Butler, B. S., Scott, S., & Xu, S. X. (2021). Next-generation information systems theorizing: A call to action. *MIS Quarterly*, *45*(1), 301–314.

Chandra, L., Seidel, S., & Gregor, S. (2015). Prescriptive knowledge in IS research: Conceptualizing design principles in terms of materiality, action, and boundary conditions. *HICSS 2015*, 4039–4048.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage Publications.

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

Dadgar, M., & Joshi, K. D. (2018). The role of information and communication technology in self-management of chronic diseases: An empirical investigation through value sensitive design. *Journal of the Association for Information Systems*, *19*(2), 2.

Darden, L., & Maull, N. (1977). Interfield theories. *Philosophy of Science*, *44*(1), 43–64.

Dehling, T., & Sunyaev, A. (2023). A design theory for transparency of information privacy practices. *Information Systems Research*.

Deng, X., Joshi, K. D., & Galliers, R. D. (2016). The duality of empowerment and marginalization in microtask crowdsourcing. *MIS Quarterly*, *40*(2), 279–302.

Denning, T., Borning, A., Friedman, B., Gill, B. T., Kohno, T., & Maisel, W. H. (2010). Patients, pacemakers, and implantable defibrillators: Human values and security for wireless implantable medical devices. *Proceedings of the*

SIGCHI Conference on Human Factors in Computing Systems, 917–926.
Denning, T., Kramer, D. B., Friedman, B., Reynolds, M. R., Gill, B., & Kohno, T. (2014). CPS: Beyond usability: Applying value sensitive design based methods to investigate domain characteristics for security for implantable cardiac devices. *Proceedings of the 30th Annual Computer Security Applications Conference*, 426–435.
Detweiler, C., & Hindriks, K. (2012). Value-sensitive design patterns for pervasive health care. *IEEE International Conference on Pervasive Computing and Communications Workshops*, 908–913.
Donia, J., & Shaw, J. A. (2021). Ethics and values in design: A structured review and theoretical critique. *Science and Engineering Ethics*, *27*(5), 57.
Düwell, M., Hübenthal, C., & Werner, M. H. (2011). Handbuch Ethik.
Emery, F. E., & Trist, E. L. (1960). Socio-technical systems. *Management Science, Models and Techniques*, *2*, 83-97.
Floridi, L. (2010). *The Cambridge handbook of information and computer ethics*. Cambridge University Press.
Floridi, L., & Taddeo, M. (2016). What Is Data Ethics? Philosophical transactions. Series A, Mathematical, Physical, and Engineering Sciences.
Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society, Harvard Data Science Review.
Findeli, A. (1994). Ethics, Aesthetics, and Design. *Design Issues*, *10*(2), 49–68.
Friedman, B., Kahn, P., & Borning, A. (2002). Value sensitive design: Theory and methods. University of Washington Technical Report, 2(8).
Friedman, B., Kahn, P. H., Borning, A., & Huldtgren, A. (2013). Value sensitive design and information systems. Early Engagement and new Technologies: Opening up the Laboratory, 55–95.
Fu, K. K., Yang, M. C., & Wood, K. L. (2016). Design principles: Literature review, analysis, and future directions. *Journal of Mechanical Design*, *138*(10), 101103.
Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, *16*(1), 15–31.
Giuffrè, M., & Shung, D. L. (2023). Harnessing the power of synthetic data in healthcare: Innovation, application, and privacy. *NPJ Digital Medicine*, *6*(1), 186.
Glaser, B., & Strauss, A. (2017). *Discovery of grounded theory: Strategies for qualitative research*. Routledge.
Gonzales, A., Guruswamy, G., & Smith, S. R. (2023). Synthetic data in health care: A narrative review. *PLoS Digital Health*, *2*(1), e0000082.
Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 337–355.
Gregor, S., & Jones, D. (2007). The anatomy of a design theory. *Journal of the Association for Information Systems*, *8*(5).
Gregor, S., Chandra Kruse, L., & Seidel, S. (2020). Research perspectives: the anatomy of a design principle. *Journal of the Association for Information Systems*, *21*(6).
Gregory, R. W., & Muntermann, J. (2014). Heuristic theorizing: Proactively generating design theories. *Information Systems Research*, *25*(3), 639–653.
Guest, G., Namey, E., & Chen, M. (2020). A simple method to assess and report thematic saturation in qualitative research. *PloS One*, *15*(5), e0232076.
Habermas, J. (1987). *The philosophical discourse of modernity*. Blackwell Publishers.
Hansen, S., & Baroody, A. J. (2020). Electronic Health Records and the Logics of Care: Complementarity and Conflict in the U.S. Healthcare System, *Information Systems Research*, *31*(1).
Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017). Code saturation versus meaning saturation: how many interviews are enough? *Qualitative Health Research*, *27*(4), 591–608.
Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine*, *292*, 114523.
Herwix, A., Haj-Bolouri, A., Rossi, M., Tremblay, M. C., Purao, S., & Gregor, S. (2022). Ethics in information systems and design science research: Five perspectives. *Communications of the Association for Information Systems*, *50*(1), 589–616.
Hevner, A. R. et al. (2004). Design Science in Information Systems. *MIS Quarterly*, *28*(1), 75–105.
Hung, P. C. K., & Cheng, V. S. Y. (2009). Privacy. Encyclopedia of Database Systems. Boston, MA, Springer US, 2136-2137.
Hwang, W. & Salvendy, G. (2010), Number of people required for usability evaluation: The 10±2 rule. *Communications of the ACM*, *53*(5), 130–133.
Iivari, J. (2015). Distinguishing and contrasting two strategies for design science research. *European Journal of Information Systems*, *24*(1), 107–115.
Iivari, J. (2020). A critical look at theories in design science research. *Journal of the Association for Information Systems*, *21*(3), 10.
Iivari, J., Rotvit Perlt Hansen, M., & Haj-Bolouri, A. (2021). A proposal for minimum reusability evaluation of design principles. *European Journal of Information Systems*, *30*(3), 286–303.

Kane, G. C., Young, A. G., Majchrzak, A., & Ransbotham, S. (2021). Avoiding an oppressive future of machine learning: A design theory for emancipatory assistants. *MIS Quarterly*, *45*(1), 371–396.

Kant, I. (1998). *Critique of pure reason*. Cambridge University Press.

Kern, C. J., Noeltner, M., & Kroenung, J. S. (2022). The case of digital ethics in IS research – A literature review. *International Conference on Information Systems (ICIS) 2022*.

Kotter, J. P. (1996). *Leading change*. Harvard Business Review Press.

Krump, M., Ruß, M., & Stütz, P. (2020), Deep learning algorithms for vehicle detection on UAV platforms: First investigations on the effects of synthetic training. *MESAS 2019*, Palermo, 50–70.

Kuechler, W., & Vaishnavi, V. (2012). A framework for theory development in design science research: Multiple perspectives. *Journal of the Association for Information Systems*, *13*(6), 3.

Kuhnke, F., & Ostermann, J. (2019). Deep head pose estimation using synthetic images and partial adversarial domain adaption for continuous label spaces. *IEEE/CVF International Conference on Computer Vision*, 10164–10173.

Lahiri, A., Agarwalla, A., & Biswas, P. K. (2018). Unsupervised domain adaptation for learning eye gaze from a million synthetic images: An adversarial approach. *Indian Conference on Computer Vision, Graphics and Image Processing*, 1–9.

Lee, J. S., Pries-Heje, J. et al. (2011). Theorizing in design science research. *DESRIST 2011*, 1–16.

Lee, S., Park, B., & Kim, A. (2019). Deep learning based object detection via style-transferred underwater sonar images. *IFAC*, *52*(21), 152–155.

Lewin, K. (1947). Frontiers in group dynamics: Concept, method and reality in social science; social equilibria and social change. *Human Relations*, *1*(1), 5–41.

Luhmann, N. (1996). The sociology of the moral and ethics. *International Sociology*, *11*(1), 27–36.

Lukyanenko, R., & Parsons, J. (2020). Research perspectives: Design theory indeterminacy: What is it, how can it be reduced, and why did the polar bear drown? *Journal of the Association for Information Systems*, *21*(5).

Lupton, D. (2018). Towards design sociology. *Sociology Compass*, *12*(1).

Magnani, G., & Gioia, D. (2023). Using the Gioia Methodology in international business and entrepreneurship research. *International Business Review*, *32*(2), 102097.

Mahieu, R., van Eck, N. J., van Putten, D., & van den Hoven, J. (2018). From dignity to security protocols: A scientometric analysis of digital ethics. *Ethics and Information Technology*, *20*(3).

Mandviwalla, M. (2015). Generating and justifying design theory. *Journal of the Association for Information Systems*, *16*(5).

Markus, M. L., & Robey, D. (1988). Information technology and organizational change: Causal structure in theory and research. *Management Science*, *34*(5), 583–598.

Marshall, B., Cardon, P., Poddar, A., & Fontenot, R. (2013). Does sample size matter in qualitative research? A review of qualitative interviews in IS research. *Journal of Computer Information Systems*, *54*(1), 11–22.

McBride, N. (2014). Active ethics: An information systems ethics for the internet age. *Journal of Information*, *12*.

Meth, H., Mueller, B., & Maedche, A. (2015). Designing a requirement mining system. *Journal of the Association for Information Systems*, *16*(9).

Mingers, J., & Brocklesby, J. (1997). Multimethodology: Towards a framework for mixing methodologies. *Omega*, *25*(5), 489–509.

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

Möller, F., Guggenberger, T. M., & Otto, B. (2020). Towards a method for design principle development in information systems. *Proceedings DESRIST 2020*, Kristiansand, Norway, 208–220.

Möller, F., Schoormann, T., Strobel, G., & Hansen, M. R. P. (2022). Unveiling the cloak: Kernel theory use in design science research. *Proceedings International Conference on Information Systems 2022*.

Müller, M., Casser, V., Lahoud, J., Smith, N., & Ghanem, B. (2018), Sim4cv: A photo-realistic simulator for computer vision applications. *International Journal of Computer Vision*, *126*, 902–919.

Müller-Wienbergen, F., Müller, O., Seidel, S., & Becker, J. (2011). Leaving the beaten tracks in creative work–A design theory for systems that support convergent and divergent thinking. *Journal of the Association for Information Systems*, *12*(11), 2.

Mueller, M., Heger, O., & Niehaves, B. (2018). Exploring ethical design dimensions of a physiotherapeutic mHealth solution through value sensitive design. *Proceedings Americas Conference on Information Systems 2018*.

Mumford, E. (2006). The story of socio-technical design: Reflections on its successes, failures and potential. *Information Systems Journal*, *16*(4), 317–342.

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

Myers, M. D., & Venable, J. R. (2014). A set of ethical principles for design science research in information systems. *Information & Management*, *51*(6), 801–809.

Niederman, F., & March, S. (2019). The “theoretical lens” concept: We all know what it means, but do we all know the same thing? *Communications of the Association for Information Systems*, *44*(1), 1.

Noushad, B., Van Gerven, P. W., & De Bruin, A. B. (2024). Twelve tips for applying the think-aloud method to capture cognitive processes. *Medical Teacher*, *46*(7), 892–897.

O’Neil, P. M. (2004). Ethics, Ethics J.K. Roth (ed.). California, Salem Press.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453.

Offermann, P., Blom, S., Schönherr, M., & Bub, U. (2011). Design range and research strategies in design science publications. *Proceedings DESRIST 2011*, 77–91.

Owen, R., Stilgoe, J., Macnaghten, P., Gorman, M., Fisher, E., & Guston, D. (2013). A framework for responsible innovation. Responsible innovation: managing the responsible emergence of science and innovation in society, 27–50.

Panch, T., Szolovits, P., & Atun, R. (2018). Artificial intelligence, machine learning and health systems. *NPJ Digital Medicine*, *1*(1), 1–3.

Panch, T., Mattie, H., & Celi, L. A. (2019). The “inconvenient truth” about AI in healthcare. *NPJ Digital Medicine*, *2*(1), 77.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77.

Peffers, K., Tuunanen, T., & Niehaves, B. (2018). Design science research genres: Introduction to the special issue on exemplars and criteria for applicable design science research. *European Journal of Information Systems*, *27*(2), 129–139.

Raji, I. D., Gebru, T., Mitchell, M., Buolamwini, J., Lee, J., & Denton, E. (2020, February). Saving face: Investigating the ethical concerns of facial recognition auditing. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 145–151.

Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine*, *169*(12), 866–872.

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347–1358.

Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach*. Pearson.

Sandelowski, M. (1995). Sample size in qualitative research. *Research in Nursing & Health*, *18*(2), 179–183.

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B. et al. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality & Quantity*, *52*, 1893–1907.

Scheck, T., Seidel, R., & Hirtz, G. (2020). Learning from Theodore: A synthetic omnidirectional top-view indoor dataset for deep transfer learning. *IEEE/CVF Winter Conference on Applications of Computer Vision*, 943–952.

Schermann, M., Böhmann, T., & Krcmar, H. (2009). Explicating design theories with conceptual models: Towards a theoretical role of reference models. *Wissenschaftstheorie und gestaltungsorientierte Wirtschaftsinformatik*, 175–194.

Schlagwein, D. et al. (2019). Ethical norms and issues in crowdsourcing practices: A Habermasian Analysis. *Information Systems Journal*, *29*(4), 811–837.

Schoormann, T., Möller, F., Chandra Kruse, L., & Otto, B. (2024). BAUSTEIN - A design tool for configuring and representing design research. *Information Systems Journal*.

Seib, V., Lange, B., & Wirtz, S. (2020). Mixing Real and Synthetic Data to Enhance Neural Network Training - A Review of Current Approaches. arXiv:2007.08781.

Siponen, M., Klaavuniemi, T., & Xiao, Q. (2024). Splitting versus lumping: Narrowing a theory’s scope may increase its value. *European Journal of Information Systems*, *33*(4), 561–570.

Sonnenberg, C., & vom Brocke, J. (2012), Evaluation patterns for design science research artefacts. *Practical Aspects of Design Science: EDSS 2011*, Leixlip, Ireland, 71–83.

Spiekermann, S. et al. (2022). Values and ethics in information systems. *Business & Information Systems Engineering*, *64*(2), 247–264.

Stilgoe, J., Owen, R., & Macnaghten, P. (2020). Developing a framework for responsible innovation. In The Ethics of Nanotechnology, Geoengineering, and Clean Energy, 347–359.

Thakur, R., Hsu, S. H., & Fontenot, G. (2012). Innovation in healthcare: Issues and future trends. *Journal of Business Research*, *65*(4), 562–569.

Topol, E. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Hachette UK.

Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the longwall method of coal-getting: An examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system. *Human Relations*, *4*(1), 3–38.

Turja, T. et al. (2020). Robot Acceptance Model for Care (Ram-Care): A principled approach to the intention to use care robots. *Information & Management*, *57*(5).

Valtchev, S. Z., & Wu, J. (2021). Domain randomization for neural network classification. *Journal of Big Data*, *8*(1), 94.

Van Den Haak, M. et al. (2003), Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour & Information Technology*, *22*(5), 339–351.

Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. *PLoS Medicine*, *15*(11), e1002689.

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A framework for evaluation in design science research. *European Journal of Information Systems*, *25*, 77–89.

Venable, J. R., Pries-Heje, J., & Baskerville, R. L. (2017). Choosing a design science research methodology. *Proceedings ACIS 2017*.

Venkateswara, H. et al. (2017). Deep-learning systems for domain adaptation in computer vision: Learning transferable feature representations. *IEEE Signal Processing Magazine*, *34*(6), 117–129.

Vom Brocke, J., Winter, R., Hevner, A., & Maedche, A. (2020). Special issue editorial–Accumulation and evolution of design knowledge in design science research: A journey through time and space. *Journal of the Association for Information Systems*, *21*(3).

Vom Brocke, J., Weber, M., & Grisold, T. (2022). Design science research of high practical relevance: Dancing through space and time. In *Engineering the Transformation of the Enterprise: A Design Science Research Perspective*, 115–135.

Wan, L., Zhou, J., & Zhang, B. (2020). Data synthesis for document layout analysis. *International Symposium on Emerging Technologies for Education*, 244–252.

Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (1992). Building an information system design theory for vigilant EIS. *Information Systems Research*, *3*(1), 36–59.

Walls, J. G., Widermeyer, G. R., & El Sawy, O. A. (2004). Assessing information system design theory in perspective: How useful was our 1992 initial rendition? *Journal of Information Technology Theory and Application*, *6*(2), 6.

Woo, C., Saghafi, A., & Rosales, A. (2014). What is a contribution to IS design science knowledge? *Proceedings International Conference on Information Systems (ICIS) 2014*, Auckland.

Yang, L., QI, L., & Zhang, B. (2022). Concepts and evaluation of saturation in qualitative research. *Advances in Psychological Science*, *30*(3), 511.

Yetim, F. (2011). Bringing discourse ethics to value sensitive design: pathways toward a deliberative future. *AIS Transactions on Human-Computer Interaction*, *3*(2), 133–155.

Young, A. G., Majchrzak, A., & Kane, G. C. (2021). Reflection on writing a theory paper: How to theorize for the future. *Journal of the Association for Information Systems*, *22*(5), 1212–1223.

Zhang, X. et al. (2018), Stacked multichannel autoencoder - An efficient way of learning from synthetic data. *Multimedia Tools and Applications*, *77*, 26563–26580.

About the Authors

Martin Böhmer is a doctoral student and research assistant at the Chair for Business Information Management at Martin Luther University Halle-Wittenberg. His research focuses on digital health, artificial intelligence, and ambient assisted living, with a methodological emphasis on design science research and theorizing. A former dean's list scholar at North Carolina State University Raleigh, he now investigates AI- and AR-driven solutions to support vulnerable populations, particularly individuals with neurodegenerative diseases and the elderly. His work has been published in leading peer-reviewed conference proceedings, such as the *European Conference on Information Systems*, and has received multiple awards, including Best Paper Awards at *Pacific-Asia Conference on Information Systems* and *Internationale Tagung Wirtschaftsinformatik*.

Stephan Kuehnel is a postdoctoral researcher and head of a research group focusing on design science research (DSRG) at the Institute for Information Systems and Operations Research at the Martin Luther University Halle-Wittenberg. He received his PhD in Business Informatics with a focus on the economic aspects of compliance in business processes. His research interests include current challenges in information security and compliance, data science project models, as well as design science research in a wide range of application areas (e.g., augmented reality, cloud computing, e-health). His research has been published in journals, such as *ACM Computing Surveys*, *Business & Information Systems Engineering*, *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *Behaviour & Information Technology*, *Computers & Security*, *Journal of Decision Systems*, and *Pacific Asia Journal of the Association for Information Systems*, as well as leading peer-reviewed conference proceedings, such as the *European Conference on Information Systems*, *International Conference on Conceptual Modeling*, and *International Conference on Business Process Management*.

Johannes Damarowsky is a postdoctoral researcher at the Chair for Business Information Management at Martin Luther University Halle-Wittenberg. His research focuses on workflow management, augmented reality, intelligent assistance systems, and knowledge management, particularly in the context of digital transformation in healthcare and nursing facilities. He has led multiple research projects, including the development of augmented reality-driven process and workflow management systems. Johannes earned his doctorate summa cum laude in 2023 and has received several academic awards, including Best Paper Awards at *ECIS* and *PACIS*. He is actively involved in academic self-administration and serves on program committees for international conferences.

Alfred Benedikt Brendel is an associate professor at the department for Operations & Decision Technologies at the renowned Kelley School of Business, Indiana University. Alfred holds a doctor's degree in management science, specializing in Business Information Systems, from the University of Goettingen. His research focuses on the application of Design Science Research to develop novel and innovative design principles and theories. His main areas of research are digital health, smart mobility, and digital work. His research is forthcoming or has been published in the proceedings of leading conferences, such as the *International Conference on Information Systems*, and journals, such as *Journal of Information Technology*, *European Journal of Information Systems*, and *Journal of the Association for Information Systems*.

Appendix

Design Principle Scheme Following Gregor et al. (2020)

Table A.1. Design Principle 1

Design Principle Title	**Ethical Healthcare Data Generation**
Aim, interpreter, and user	To ensure that synthetic imagery complies with ethical and healthcare privacy regulations, maintaining alignment with both technical standards and social ethical frameworks.
Context	In digital health systems where synthetic image data is used for analysis, AI training, or decision support, while ensuring ethical values and the sociotechnical balance between innovation and ethical healthcare practices.
Mechanism	Incorporate ethical values and sociotechnical mechanisms to ensure that users account for both social and technical subsystems when making design decisions.
Rationale	Drawing on the theory of value-sensitive design (Friedman et al., 2002; 2013), ethical AI theories (Floridi & Cowls, 2019), and meta-level ethics (Myers & Venable; Herwix et al., 2022), this principle recognizes that healthcare data generation processes are not inherently private (Giuffrè & Shung, 2023) and must respect the ethical imperative to protect individuals' rights. Since design decisions are not neutral (Findeli, 1994), the principle stresses the balance between social and technical elements (Mumford, 2006), ensuring that ethical considerations, such as bias mitigation and privacy, are not overlooked in technical implementations.

Table A.2. Design Principle 2

Design Principle Title	**Comprehensive Privacy Protection**
Aim, interpreter, and user	To ensure the protection of sensitive healthcare data, upholding privacy, data security, and alignment with technical and social/ethical imperatives.
Context	In digital health systems where synthetic image data must adhere to strict privacy protection frameworks, ensuring compliance with healthcare privacy regulations and ethical values.
Mechanism	Apply robust monitoring techniques to ensure that the synthetically generated image data does not contain sensible real-world reference data to prevent re-identification or leakage of sensitive information, taking into account both technical and social aspects.
Rationale	Rooted in sociotechnical systems theory (Mumford, 2006) and based on flexible ethical designs (Albrechtslund, 2007) as well as privacy protection theories (Hansen & Baroody, 2020; Raji et al., 2020), this principle emphasizes that synthetic data, though artificial, is not inherently free from privacy risks. Rigorous privacy measures must be implemented to safeguard sensitive information, as mandated by meta-ethical frameworks (Myers & Venable, 2014).

Table A.3. Design Principle 3

Design Principle Title	Adaptive Data Governance
Aim, interpreter, and user	To establish adaptive governance frameworks that ensure the ethical and secure creation, use, and distribution of synthetic data in computer vision healthcare settings.
Context	In dynamic computer vision digital health environments where governance frameworks must respond to both technological changes and evolving social/ethical values.
Mechanism	Implement adaptive policies and governance mechanisms that can flexibly respond to emerging ethical challenges while ensuring compliance with both social imperatives and technical standards, recognizing that synthetic image data is not inherently free from risks related to privacy, security, and ethical concerns.
Rationale	Informed by socio-technical systems theory (Mumford, 2006) and adaptive ethical healthcare concepts (Panch et al., 2019; Rajkomar et al., 2018; 2019) for ethical values (Friedman et al., 2002; 2013), this principle emphasizes the need for ethical governance structures to address both privacy concerns and ethical dilemmas that arise during synthetic data generation and usage (Giuffrè & Shung, 2023; Vayena et al., 2018).

Table A.4. Design Principle 4

Design Principle Title	Synthetic Scene Diversity
Aim, interpreter, and user	To enhance model generalization and performance by incorporating diverse synthetic scenes that align with both social and technical learning imperatives.
Context	In computer vision models for healthcare, where synthetic scenes must be both technically proficient and represent a variety of social, cultural, and environmental contexts without reinforcing biases.
Mechanism	Introduce diverse textures, objects, and backgrounds into synthetic scenes to ensure models learn across a variety of scenarios, avoiding overfitting and social biases.
Rationale	Based on bias mitigation strategies (Obermeyer et al., 2019; Raji et al., 2020) and the need for cross-domain generalization (Scheck et al., 2020), this principle ensures that computer vision models trained on diverse data (i.e., scenes) perform well in varied real-world scenarios (Seib et al., 2020; Valtchev & Wu, 2021), reflecting socio-cultural variability (Donia & Shaw, 2021).

Table A.5. Design Principle 5

Design Principle Title	Controlled Scene Composition
Aim, interpreter, and user	To ensure that key features of synthetic scenes in healthcare are represented accurately, with a balance between scene randomness, diversity, and control.
Context	In the generation of synthetic image data for healthcare systems, where scene diversity is required, but important features must remain under control to prevent distortion and falsification.
Mechanism	Control object scale, orientation, and spatial relationships in synthetic scenes to ensure accurate representation of key healthcare factors, balancing randomness with precision and accuracy.
Rationale	Drawing on ethical-aesthetic aspects in design (Albrechtslund, 2007; Findeli, 1994) and scene composition techniques (Krump et al., 2020), this principle stresses that while randomness is beneficial, it must not overshadow core healthcare features, which are crucial for model generalization (Scheck et al., 2020).

Table A.6. Design Principle 6

Design Principle Title	Flexible Complexity Management
Aim, interpreter, and user	To optimize social and technical model learning by managing the complexity of synthetic scenes through gradual introduction and balancing.
Context	In computer vision models for healthcare, which require adaptive learning environments to handle the dynamic complexity of healthcare scenarios from both a social and technical perspective.
Mechanism	Gradually introduce synthetic scenes of varying social and technical complexity and monitor model performance to prevent overfitting and encourage robust learning.
Rationale	Drawing from the value trade-off between security needs and usability (Denning et al., 2014), this principle emphasizes the dynamic introduction of complex elements in synthetic scenes to help models build foundational learning before confronting more challenging, social or technical, scenarios (Alzubaidi et al., 2021; Bird et al., 2020). Based on adaptive complexity management in socio-technical systems theory (Mumford, 2006; Bostrom & Heinen, 1977), this principle ensures that models encounter diverse complexity levels, building robust learning mechanisms (Alzubaidi et al., 2021).

Table A.7. Design Principle 7

Design Principle Title	Data Augmentation
Aim, interpreter, and user	To enhance model robustness by applying augmentation techniques that simulate and promote variations in synthetic image scenes.
Context	In computer vision healthcare models where synthetic image data must simulate a range of real-world conditions to improve generalizability and reliability.
Mechanism	Use geometric transformations, colour modifications, noise introduction, and random elements to diversify synthetic scenes, ensuring that models learn to adapt to various conditions and focus on the key characteristics.
Rationale	Drawing from augmentation techniques in AI (Müller et al., 2018; Seib et al., 2020; Zhang et al., 2018), this principle highlights the importance of diverse data augmentation strategies to prevent overfitting and ensure robustness across different healthcare environments. It reflects the general socio-technical imperative to create synthetic data that can be generalized across different healthcare settings and patient populations, ensuring that computer vision models are adaptable and sustainable (Emery & Trist, 1960).

Table A.8. Design Principle 8

Design Principle Title	Responsible AI
Aim, interpreter, and user	To promote ethical and accountable AI practices in healthcare computer vision, focusing on bias detection, fairness, and explainability.
Context	In digital health systems where computer vision models must be both technically reliable and socially responsible, aligning with ethical healthcare practices.
Mechanism	Implement fairness assessments, bias detection, and transparency mechanisms to ensure that computer vision systems in healthcare operate ethically and are accountable to their users.
Rationale	Based on ethical AI practices (Floridi & Cowls, 2019; McBride, 2014; Russel & Norvig, 2021) and bias mitigation theories (Obermeyer et al., 2019), this principle emphasizes the importance of ensuring that healthcare AI models are fair, transparent, and free from bias, supporting trust in digital health technologies (Vayena et al., 2018). It emphasizes that while synthetic data is used in AI development, it does not inherently eliminate bias or ensure fairness, which is critical in value-sensitive environments (Friedmann et al., 2002; 2013; Herwix et al., 2022), where the responsibility of AI developers is balanced between technical and social/ethical imperatives (Mumford, 2006, Donia & Shaw, 2021; Luhmann, 1996).

Table A.9. Design Principle 9

Design Principle Title	Transfer Learning Guidelines
Aim, interpreter, and user	To guide the domain adaptation of models trained on synthetic data to real-world healthcare applications, ensuring they generalize and perform well.
Context	In healthcare settings where computer vision models must be fine-tuned for specific applications to ensure they perform effectively in real-world scenarios.
Mechanism	Provide protocols for domain adaptation and fine-tuning, ensuring that models trained on synthetic data and their parameters are optimized for real-world healthcare data and environments.
Rationale	Based on domain adaptation and transfer learning foundations (Murtaza et al., 2023; Kuhnke and Ostermann, 2019; Lahiri et al., 2018); Venkateswara et al., 2017), this principle ensures that models maintain high performance when transitioning from synthetic data environments to real-world healthcare applications. Its adaptability is based on socio-technical imperatives (Emery & Trist, 1960), which emphasize the need for systems to evolve and remain effective in dynamic environments.

Table A.10. Design Principle 10

Design Principle Title	Robustness Checks
Aim, interpreter, and user	To rigorously test the robustness of computer vision models against varying healthcare conditions and environmental inputs to align technical processes with social/ethical values.
Context	In healthcare computer vision models where reliability, adaptability, and resilience are critical to ensuring performance in unpredictable and ethically-sensitive real-world scenarios.
Mechanism	Conduct thorough model testing across a wide range of environmental conditions and healthcare scenarios to ensure computer vision model robustness and reliability.
Rationale	Based on technological marginalization concerns (Deng et al., 2016) and the need for robustness testing in models trained on synthetic data (Giuffrè & Shung, 2023), this principle emphasizes the need for extensive testing to ensure that computer vision systems can handle real-world healthcare variations without sacrificing accuracy/performance and ethical aspects (Floridi and Cowls, 2019; Giuffrè and Shung, 2023; Gonzales et al., 2023; Rajkomar et al., 2019; Russel and Norvig, 2021; Valtchev and Wu, 2021). In these scenarios, particularly in the digital health domain, and in light of a common utilitarian view that often overlooks social complexity (Lupton, 2017), it seems reasonable to stress test AI models to assess their resilience to extreme or rare scenarios.