Department of Industrial Engineering and Innovation Sciences

Designing an Interoperable and Modular System Architecture to Support the Development of Human Digital Twins for Promoting Healthy Lifestyle Behaviors

MASTER THESIS

Author:
W.J. (Pim) van Vroonhoven
1266691

Program:
Operations Management and Logistics

Supervisors:
First supervisor: Dr. P.M.E. van Gorp
Second supervisor: Dr. L. Genga
Third supervisor: Dr. B. Ozkan
Fourth (daily) supervisor: L.J. James

Eindhoven, December 6, 2024

TU/e, Department of Industrial Engineering &
Innovation Sciences
Master Thesis Operations Management & Logistics
Information Systems Group

Keywords: Human Digital Twins, System Architecture, Interoperability, Modularity, Lifestyle Health Promotion

Designing an Interoperable and Modular System Architecture to Support the Development of Human Digital Twins for Promoting Healthy Lifestyle Behaviors

Wilhelmus J. van Vroonhoven

Abstract—**Digital health applications have become widely available tools for promoting healthy lifestyle behaviors. However, the persistent rise in obesity and chronic health conditions underscores their limitations in achieving long-term user engagement and efficacy. Human Digital Twin (HDT) technology offers a promising solution by enabling adaptive, personalized health interventions through real-time data exchange and dynamic modeling. Despite this potential, existing HDT architectures lack the interoperability and modularity needed to integrate diverse data sources and support the open development of advanced, personalized models for lifestyle health contexts. Building on a comprehensive review of existing HDT frameworks, this paper identifies key architectural gaps and proposes solutions to address them. A proof-of-concept prototype is developed to demonstrate and verify the feasibility of a modular, interoperable HDT system. Extending this prototype, an extensive architecture is proposed that lays a foundation for the future development HDT systems that empower health applications to deliver personalized experiences which dynamically adapt to users' evolving health needs, goals, environmental contexts, and technological advancements, ultimately fostering sustained behavior change and improving long-term health outcomes.**

Index Terms—**Human Digital Twins, system architecture, interoperability, modularity, lifestyle health promotion**

I. Introduction

According to the WHO Regional Office for Europe [1], in 2022, approximately 59% of the adult population in Europe fell into the overweight or obese categories, with 36% classified as overweight (Body Mass Index (BMI) of 25 kg/m^2 or above) and 23% as obese (BMI of 30 kg/m^2 or above). Obesity is linked to an increased risk of cardiovascular diseases, chronic respiratory diseases, cancer, and diabetes, contributing to over 13% of total deaths across European countries [1]. While the BMI metric, which solely relies on an individual's weight and height, has faced criticism for potentially misclassifying individuals [2], [3], alternative metrics like the Fat Free Mass Index and Percentage Body Fat have been shown to classify even more individuals as obese in certain populations [4]. To address these concerning statistics, promoting healthy lifestyle behaviors, such as adopting a balanced diet and engaging in regular physical activity, is essential [1].

Given their availability and ease of use, digital health applications provide an accessible tool to support individuals in maintaining these healthy lifestyle behaviors [5]. However, despite the increasing availability and sophistication of such tools [6], the prevalence of obesity in Europe continues to rise [7]. This paradox highlights a significant challenge: while digital health applications are more accessible than ever, they often fail to engage users effectively in the long term and promote lasting behavioral changes needed for improving one's health [8].

The long-term success of digital health applications is often limited by psychological barriers that hinder sustained user engagement. As noted in a systematic review by Jakob et al. [9], several factors positively influence adherence to digital health applications. The factors include user-friendliness, personalization of content, adaptive reminders, professional support, gamification elements, social features, and automated data collection, processing, and transmission.

One promising approach for addressing the challenges of long-term user engagement and promoting lasting behavioral change in digital health applications is the concept of Human Digital Twins (HDT). Digital Twins (DTs) consist of a physical entity and its virtual twin, connected by bi-directional data flows that enable continuous information exchange. This connection ensures that changes in one are reflected in the other [10], [11]. Unlike static digital models, which offer fixed representations, or digital shadows, which update only in one direction, from the physical entity to its virtual counterpart, DTs support dynamic two-way interaction based on real-time data [12].

As the concept of Digital Twins evolved, it paved the way for Human Digital Twins, where the physical entity being modeled is now a human being. HDT systems hold significant potential for enhancing several of the aforementioned key factors that drive user engagement in digital health applications [9]. For example, HDT systems leverage various methods to optimize and automate data collection, such as wearable devices and sensors [12], [13]. Furthermore, the virtual twin back-ends of HDT systems feature (artificial intelligence) models that generate comprehensive feedback, predictions and actionable insights, which can be used to personalize the content of connected applications [14]–[18].

In addition to addressing key psychological barriers to long-term user engagement, HDTs' personalization capabilities could potentially enable the delivery of more precise and effective lifestyle interventions tailored to an individual's unique physiological characteristics. In practical terms, HDTs could support the continuous optimization of exercise routines, dietary plans, and sleep schedules by adjusting to an individual's real-time data.

Fully harnessing the capabilities of HDT systems has the potential to transform digital tools for promoting healthy lifestyle behaviors, enabling more user-centric, interactive, engaging, and impactful content, driving sustained behavior change and improving health outcomes.

Despite our optimism regarding the potential of HDT technology, several gaps in the system architectures of existing frameworks must be addressed to ensure the effective implementation in the context of lifestyle-oriented health applications.

In the current landscape of digital lifestyle health management, many widely-used applications focus on specific aspects of health, such as nutrition (e.g., MyFitnessPal), exercise (e.g., Strava), or sleep (e.g., Sleep Cycle). While these apps offer robust tracking and feedback, along with some data integrations between them, they lack a centralized approach for synthesizing data across multiple different health dimensions. This lack of comprehensive integration limits their ability to provide a holistic view of the user's overall health and lifestyle [14], [19], [20].

Platforms like Google Fit and Apple Health enable the integration of a wide array of health apps, serving as hubs for health data and providing users with accessible analytics and personalized insights [21], [22]. However, the feedback provided is often constrained by the limitations of each individual app, preventing effective synthesis and cross-analysis of data across different health dimensions. While these platforms offer various analytics, their back-ends are limited in cross-domain analysis and do not support training complex models based on integrated data from multiple health domains.

For advanced, cross-domain models to reach true intelligence, extensive long-term data accumulation and training is required. Therefore, it is crucial that the intelligence built up inside the virtual twin models is preserved, even when users switch between digital health-promoting platforms or opt for different apps in specific health domains.

In this paper, interoperability is defined as an HDT system's capacity to facilitate well-organized data exchange between virtual twin models and various digital health-promoting platforms or applications, allowing data from diverse sources to be accessed in a structured manner. In practical terms, interoperability allows users to seamlessly switch between preferred platforms or applications while retaining and building upon the trained models and generated insights in their virtual twins.

However, most existing health-promoting platforms often "lock" users into their ecosystems. This limitation restricts users from switching between platforms without losing accumulated insights and trained models, ultimately hindering the long-term development and refinement of advanced lifestyle-related virtual twin models. Addressing the interoperability challenge is therefore essential for ensuring that HDT systems can evolve alongside users' changing preferences and technological advancements.

To build a comprehensive virtual twin, it is essential to analyze various dimensions of a person's health and lifestyle, with each aspect requiring a specific model that utilizes diverse data to evaluate the individual's situation and provide feedback accordingly [14], [19], [20]. Addressing these diverse dimensions necessitates a modular system architecture that can integrate specialized models for each domain.

In this paper, modularity therefore refers to an HDT system's capacity to integrate diverse third-party models, which is essential for enabling contributions from domain experts. Such experts could bring specialized knowledge to ensure the accuracy, relevance, and robustness of models in areas such as nutrition, physical activity, and sleep. A modular system architecture would allow experts to develop, train, test, and refine specialized HDT models that capture insights across diverse health domains. Such models could reveal patterns in habits and physiological responses unique to each individual's characteristics, enabling the delivery of highly tailored feedback and recommendations.

However, integrating third-party models into an HDT system presents significant challenges, including ensuring seamless access to standardized data, safeguarding data privacy and security, and balancing the diverse objectives of stakeholders in research and commercial settings.

Addressing the challenges of modularity is pivotal for creating dynamic HDT systems capable of integrating specialized models tailored to diverse health domains. Facilitating modularity enables contributions from third-party domain experts and developers, unlocking significant potential for innovation across both research and commercial settings.

It should be emphasized that the scope of this paper is focused on the design of an interoperable and modular HDT system architecture, rather than the development, implementation and evaluation of specific (AI) models or health interventions.

This paper aims to lay a foundation for the future development of interoperable and modular HDT systems designed to promote healthy lifestyle behaviors. By enabling seamless data integration from diverse sources and supporting the open development and integration of advanced virtual twin models, these HDT systems should empower various health applications and platforms to deliver personalized experiences that dynamically adapt to users' evolving health needs, goals, environmental contexts, and technological advancements, ultimately fostering sustained behavior change and improving long-term health outcomes.

The remainder of this paper is organized as follows: Section II describes the methods used to address the interoperability and modularity challenges, Section III presents an overview of the relevant literature on HDT systems, resulting in a general literature-based HDT system architecture, Section V describes the development of a prototype lifestyle-oriented HDT system, Section IV describes how the prototype can be extended towards a newly proposed HDT system architecture, Section VI provides a discussion, highlighting limitations and directions for future work and finally Section VII concludes the paper.

II. Methods

The methods used to address the challenges identified in the introduction are as follows:

1) To deepen our understanding of HDT systems, a literature review is performed to explore them from an information systems perspective, focusing on their architectural frameworks. The review aggregates information regarding the key layers and components of HDT systems as applied in various domains and constructs a general HDT system architecture based on existing literature.
2) Next, several suggestions and requirements are formulated for adapting the literature-based HDT system architecture such that it is better able to address challenges related to interoperability and modularity. Considering these requirements, an extensive system architecture is proposed to support future development of HDTs for lifestyle health promotion. The proposed architecture enables seamless integration of user data from various health-promoting platforms and applications while allowing for the open development, testing, and integration of advanced, comprehensive HDT models. Data privacy, security, and user-controlled access management remain integral to the system's design.
3) A proof-of-concept prototype is developed to demonstrate and verify the practical feasibility of the proposed architecture and its associated requirements. This prototype emphasizes seamless data exchange and processing as foundational components.

III. Literature Review

The collection and evaluation strategy for this review prioritized identifying and selecting papers that propose representations of HDT system architectures or discuss key design considerations and technological requirements for developing effective HDT systems. The selection process focused on open access works published in English, sourced from peer-reviewed journal articles or conference proceedings. This approach resulted in a final selection of 25 papers, primarily from 2022 or later, highlighting the recent and evolving nature of research on HDT systems.

Most of these studies address HDT applications within clinical or medical settings, reflecting the technology's advancements in healthcare. However, only two of the selected papers specifically address behavior management [17], [23], highlighting a significant underrepresentation in the literature concerning the application of HDTs to lifestyle behavior management.

Therefore, this review first aggregates insights on the key layers and components of HDT systems as applied in various domains (Section III-A). These insights are then used to construct a general literature-based HDT system architecture (Section III-B). This architecture serves as a foundational framework for the proposed architecture and prototype described in subsequent chapters, aiming to bridge the underexplored domain of lifestyle-oriented HDT systems.

A. Key Layers and Components of HDT Systems

This review adopts a universal HDT framework that incorporates six key layers: *(1) data collection, (2) communication, (3) computation, (4) data management, (5) data analysis and modeling, and (6) an application layer.* Although frameworks in the literature vary slightly in structure and naming, most agree on these foundational layers [13], [20], [24], [25]. The following sections will discuss each layer's key technologies, methods, design principles, and role in the HDT framework.

1) Data Collection Layer: When aiming to construct a comprehensive HDT, various aspects of the physical entity are to be captured, including its actions, performance, and state within its environment [26]. The range of human attributes that can be modeled can be divided into eight categories: physical, physiological, perceptual performance, cognitive performance, personality characteristics, emotional state, ethical stance, and behavior [25]–[27].

Combining data from various sources is crucial for gaining a comprehensive understanding of all aspects of the physical entity [20], [24]. Key data collection methods include the use of: wearable devices, (biomedical) sensors, electronic health records, social media, tests, manual logging, location tracking, and measuring environmental factors [13]–[18], [20], [23]–[38].

Despite advancements in sensor technology, several challenges persist in the data collection process. A significant issue is the accuracy-privacy trade-off: while more accurate sensors enhance data quality, they often come at the cost of increased intrusiveness and exposure of sensitive user information [33]. Furthermore, when fully automated data collection is not feasible, encouraging consistent self-reporting poses a challenge due to reporting fatigue [33], with strategies such as gamification offering promising solutions [14], [33], [37]. Finally, there remains a critical need to balance the technical capabilities of collection devices with their user-friendliness, ensuring they remain accessible and convenient for users [18], [25].

2) Communication Layer: The communication layer facilitates the secure and efficient exchange of data between the digital and physical counterparts of HDT systems, functioning as the bridges between all other layers [13], [18], [20], [25], [27], [35]. Two distinct tiers of communication technologies can be identified: on-body and beyond-body communication [13].

On-body communication involves short-range interactions between sensors and local edge devices, such as smartphones, using technologies like Bluetooth Low Energy, Near Field Communication (NFC), and ZigBee [12], [13], [20], [29], [36]. These protocols are designed for energy efficiency and minimal power consumption, making them ideal for wearable and implantable devices that require continuous operation without frequent recharging.

Beyond-body communication extends data transmission from local gateways to remote servers, supported by technologies such as WiFi, 4G/5G, and wired networks [13], [16], [20], [25], [39]. These technologies ensure faster data transmission and connectivity over longer distances.

For efficient data handling, the bi-directional flow of sensitive data in HDT systems should conform to communication standards, such as ISO/IEEE 11073 (on-body) and HL7 FHIR (beyond-body), ensuring compatibility and secure information exchange across devices and platforms [12], [23], [29]. Secure communication protocols like HTTPS should be used to encrypt data in transit, protecting it from interception or tampering [31], [35]. Lightweight protocols such as MQTT and Zenoh can also be employed to manage data transmission efficiently in resource-constrained environments [35]. Ideally, data should be exchanged in a standardized format, such as JSON or XML, to facilitate efficient transmission between edge nodes and cloud servers, thereby enabling seamless data sharing across the system [34], [36].

3) Computation Layer: This layer provides the computational power needed to perform various tasks in the data management and data analysis layers of HDT systems. Two main types of computing resources can be distinguished: edge computing and cloud computing. Each serving a different purpose, many sources suggests a suitable form of edge-cloud collaboration, using smart allocation of tasks between the two to optimize performance and resource usage [13], [15], [18], [25], [30], [35].

Edge computing, using devices located close to the user, are best suited for handling less computationally intensive and more time-sensitive tasks, such as generating outputs from pre-trained models [18], [30]. Mobile edge computing and local storage of trained models can potentially solve some mobility issues of HDTs, allowing devices to quickly and reliably generate insights at any location [13]. While edge computing offers low latency and real-time processing benefits, it also presents challenges such as high energy consumption, increased burden on edge servers' resources, and potential data privacy issues [13].

Cloud computing offloads more resource-intensive processes, such as machine learning training, to the cloud, where powerful central servers can handle them efficiently [15], [18], [30]. However, cloud computing introduces its own limitations, such as high communication costs, network congestion, and privacy risks associated with centralized storage [13]. Furthermore, the choice between public and private cloud servers involves trade-offs between control, privacy, scalability, and costs [35].

Edge-fog-cloud collaboration expands the edge-cloud framework by introducing an intermediary 'fog' layer, where semi-local servers handle pre-processing tasks before sending data to the cloud. This setup further reduces latency and bandwidth usage, enhancing performance for time-sensitive applications [18], [35].

Together, the intelligent integration of edge, cloud, and fog computing forms a robust computational infrastructure, ensuring that HDT systems remain responsive, cost effective, secure, and capable of handling a wide range of tasks efficiently.

4) Data Management Layer: The data management layer in HDT systems has three main functions: pre-processing incoming data, structured and efficient storage of data, and safeguarding the privacy and security of the data.

Pre-processing is essential for transforming the raw, heterogeneous, multi-source, and high-noise data into a standardized format (such as HL7 FHIR [29]), making it accessible to all developers. Potential pre-processing steps include data cleaning, imputation of missing values, data reduction, data transformation, data segmentation, data augmentation, and data fusion [12], [13], [15], [17], [24]–[26], [29]–[31], [33], [36], [38]. These steps ensure consistency across the system, which is fundamental for effective data sharing, analysis, and modeling.

Structured and efficient data storage plays a critical role in managing different types of data, including raw data, standardized pre-processed data, trained models, and their outputs. Edge nodes can serve as temporary cache storage, enabling basic pre-processing before offloading data to the cloud [35]. Vats et al. [31] describe a system where incoming data is initially stored in a cloud-based raw database. Following pre-processing, the standardized data is used to train HDT models, with both the trained models and their outputs securely stored within the virtual twin for future applications.

Across the literature, various storage solutions have been proposed depending on the nature of the data: relational databases are well-suited for structured data that requires complex querying, while non-relational databases, object-based storage systems, and distributed file systems can handle unstructured or semi-structured data [12], [20], [35].

Privacy and security are paramount in HDT systems due to the sensitive nature of the data involved. HDT systems are vulnerable to various threats across the physical, virtual, and communication spaces, such as Denial of Service (DoS) attacks, malicious data injection, and data transmission risks such as spoofing or eavesdropping [27].

To safeguard user data and ensure the integrity of the virtual twin, proper data protection techniques, such as authentication [13], [27], access control [25], [27], anonymization [12], [27], and encryption protocols [31] are critical. Additionally, requiring users to explicitly opt-in for each data collection device allows them to select privacy-safe options, building trust in the system as they gradually adopt more sensors [33].

Personal data storage solutions facilitate user-centric data management by enabling individuals to store and control their data, giving them the option to share it with third parties as needed [27]. Ensuring compliance with regulations such as GDPR and the EU Unique Device Identification system is vital for secure data exchange and transparency [12], [18], [23], [34], [36].

For network security, firewalls, intrusion detection systems, and IP address-based access control protect against unauthorized access and potential breaches [12], [20], [27]. Moreover, advanced technologies like blockchain and federated learning offer further protection for securing data sharing and decentralized model training, eliminating the need to transmit sensitive data to centralized locations [13], [30].

5) Data Analysis and Modeling Layer: To build a comprehensive HDT, it is essential to analyze the various dimensions of a person, with each aspect requiring a specific model that utilizes diverse data to evaluate the individual's situation and provide feedback accordingly [14], [19], [20]. The tasks performed by these models can vary in complexity, from straightforward calculations using established formulas and theories to more advanced tasks such as predictions leveraging artificial intelligence (AI) models, and inference tasks that integrate both computational and predictive outcomes into actionable insights [25].

The literature highlights various of AI models—such as regression, classification, deep learning, reinforcement learning, and generative models—that can be deployed to handle tasks of varying nature and complexity [12], [14], [20], [24]–[27], [29]–[31], [34]–[36]. Which model is chosen depends on the specific task requirements, such as regression models for predicting continuous outcomes or reinforcement learning models for tasks involving decision-making in dynamic environments [25].

From a privacy perspective, the ideal scenario involves developing user-specific HDT models that rely solely on an individual's data. However, obtaining sufficient training data for such personalized models poses a significant challenge. To address this limitation, AI techniques such as transfer learning, cascaded classification, and semi-supervised learning offer promising solutions [33].

6) Application Layer: The Application Layer leverages insights from the Data Analysis and Modeling Layer to generate tailored content for individuals, including visualizations, feedback, recommendations, challenges, and rewards. This versatility is reflected in the wide range of use cases found in the literature, including

The Application Layer utilizes insights from the Data Analysis and Modeling Layer to deliver personalized content to individuals. This content includes visualizations, feedback, recommendations, challenges, and rewards. The variety of applications is evident in use cases across the literature,, including healthcare (e.g., patient monitoring and personalized treatment), manufacturing (e.g., human-robot collaboration), human performance (e.g., optimizing training regimens), education (e.g., personalized learning strategies), and mobility (e.g., driver type classification) [12], [14], [20], [24]–[27], [29]–[31], [34]–[36].

B. Literature-Based HDT System Architecture

Based on insights from the literature regarding the key layers and components HDT systems, a general literature-based HDT system architecture was developed. This architecture, shown in Fig. 1, demonstrates the relationships and data flows between the key layers in HDT systems.

Raw user data is collected through various methods to capture information about the physical twin and its environment (1). This data is initially stored in a dedicated raw database (2). Upon user authorization, this data goes through preprocessing, resulting in standardized data that can be used for data analysis and modeling (3). Subsequently, the generated insights and trained models are securely stored in the cloud-based virtual twin (4). To close the loop, the application layer leverages data- and model-driven insights (5) to deliver a range of personalized content back to the user (6). The entire system is supported by the underlying computation and communication layers, which facilitate essential processing power and connectivity between all other layers.

IV. Interoperable and Modular System Architecture for Lifestyle-Oriented HDTs

This chapter first outlines several suggestions and requirements for adapting the literature-based HDT system architecture (Fig. 1) such that it is better able to address challenges related to interoperability and modularity (Section IV-A).

Considering these requirements, an extensive system architecture is proposed to support future development of HDTs for lifestyle health promotion (Section IV-B).

A. Design Requirements

1) Integration of Data Collection into Application Layer: The first major architectural adaptation involves merging the data collection and application layers. In the context of HDT systems designed for promoting healthy lifestyle behaviors, both data collection and the delivery of personalized content frequently occur through interaction with a single application front end. Consequently, data collection can be regarded as an integrated functionality of the application layer, marking a shift from the literature-based framework, where these layers are often distinct and positioned at opposite ends. Integrating these layers into a single, interactive interface enhances coherence and better aligns with practical, real-life use cases, such as tracking physical activity or food intake through mobile applications.

2) Data Standardization: To ensure the effective aggregation of data from diverse sources, the HDT system must establish a robust data exchange standards. However, the data generated by most health applications does not (yet) conform to any widely accepted standard, such as HL7's FHIR [40]. Therefore, the HDT's core infrastructure must include extensive pre-possessing capabilities for transforming heterogeneous, multi-source, and noisy data into a standardized format that adheres to predefined agreements on data structures, required fields, data types, and naming conventions.

Standardization enables interoperability by allowing diverse platforms and applications to contribute data, regardless of the specific subset of properties each source provides. For instance, users may log their food intake using different applications, each with slightly varying data structures or naming conventions but significant overlap in included fields. Aggregating and consolidating this data into a unified format ensures consistency across sources and enables capturing the full spectrum of potentially relevant metrics across lifestyle health domains.

In addition to supporting interoperability, data exchange standards promote modularity by providing third-party developers and domain experts with a consistent data structure. With uniformly formatted data, developers can efficiently access all available information of a specific type, streamlining the creation and refinement of advanced models. This consistency facilitates innovation by lowering barriers for integrating diverse, domain-specific models into the HDT system.

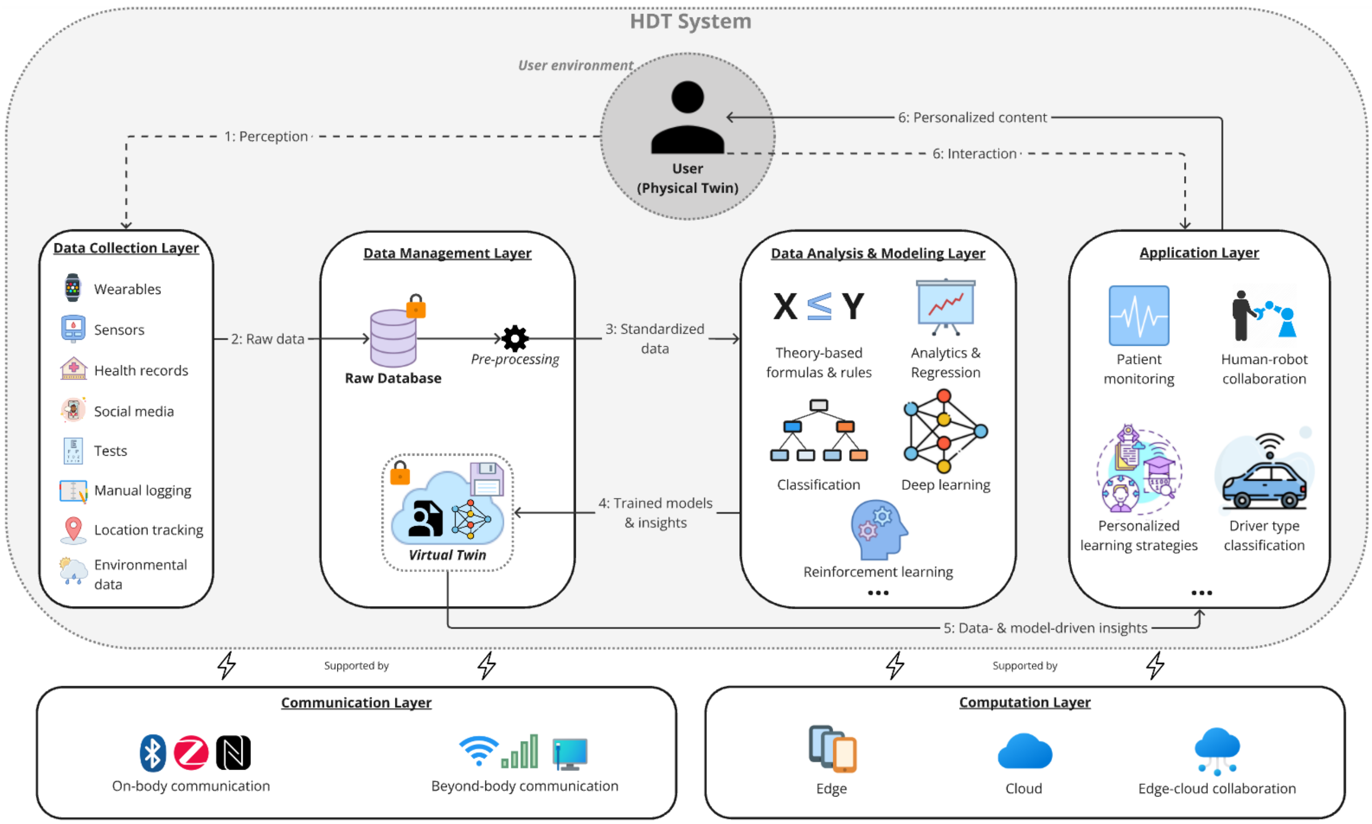

Fig. 1. Generalized, literature-based HDT system architecture. This figure presents the six core layers of HDT systems, emphasizing the data flows between them and showcasing some of the most common technological components and functionalities within each layer.

3) Storage of Trained Models and Resulting Outputs: For authorized health platforms and applications to personalize their content effectively, they need easy access to the data- and model-driven insights in the virtual twin. However, creating a strict standardized interface for storing trained models and their resulting outputs remains difficult due to the significant variability in model structures, formats, inputs and outputs. This complexity necessitates a flexible approach that allows them to be stored and accessed effectively within the virtual twin.

A primary requirement for storing trained models is to establish a standardized metadata schema that records essential model details such as version, input data types, expected outputs, usage specifications, and performance metrics. This metadata serves as integrated documentation, enabling each model to be accurately understood by external applications, fostering interoperability across platforms. For storing historical model outputs, a similar schema should capture key result properties, including data types, timestamped entries, and interpretation guidelines, ensuring that applications can easily access and utilize these insights in an appropriate manner.

In addition to metadata, each model's input-output structure can be organized around predefined categories or labels, which represent the general purpose of the model (e.g., activity recommendation, dietary analysis, sleep pattern recognition). This systematized categorization support both interoperability and modularity by providing model developers and health platforms with consistent access points, regardless of specific model architecture.

4) User Control and System Continuity: Ensuring that the virtual twin is persistent, independent, and equipped with robust authorization and authentication measures, such as OAuth 2.0 [41], is essential for enabling users to retain full control over their data, models, and insights [13], [25], [27]. By implementing these controls, users can explicitly decide which third-party services or applications can access specific portions of their virtual twin. This setup allows users to freely choose and switch between platforms throughout their personalized health journey while maintaining the continuity of their virtual twin's accumulated intelligence. As a result, newly adopted applications can immediately benefit from the insights and models already present within the virtual twin.

This approach is similar to the SMART on FHIR framework used in healthcare, where interoperability is achieved by enabling applications to securely access electronic health records with user consent [42]. However, in the context of HDTs, the focus shifts from enabling apps to interact with clinical electronic health records to allowing apps to interact with virtual twins.

5) Supplementary Documentation: To further complement standardization efforts, extensive documentation regarding the broader data organization of the virtual twin is required for third-parties to contribute effectively to the system. This documentation should detail the available data and models, their metadata, and accessibility options, allowing both developers of health applications and models to navigate the virtual twin infrastructure effectively.

B. Proposed System Architecture

Building on the outlined design requirements and addressing additional challenges such as regulatory compliance and the potential commercial interests of third parties in open, real-world environments, this subsection proposes an interoperable and modular architecture to support future development of HDT systems for lifestyle health promotion (Fig. 2).

Colors are use to depict the different (types of) parties responsible for managing the various modules in the system. Additionally, layers are used for indicating that there can be multiple instances of the same type of party or module.

To aid comprehension of the proposed architecture, all data flows and interactions are numbered, where the first number represents an event or a group of events that happen roughly simultaneously, the decimal represents the sequencing of events that are grouped together, and the letters represent different variants of similar events, involving different modules. The remainder of this chapter will guide the reader trough the proposed architecture (Fig. 2).

Users can interact with various health applications (1a) and/or health platforms (1b) that support integrations with various health applications (1c, 1d).

Collected user data can be stored in app-specific databases (2.2a, 2.3a), or aggregated into a health platform's database (2.1, 2.2b, 2.2c, 2.3b).

Achieving interoperability in HDT systems also opens up opportunities for including external data sources beyond health applications or platforms to enhance the relevance of recommendations by adapting to the user's environment (2.2d, 2.3c). For instance, incorporating real-time weather data could prevent recommendations for outdoor activities during adverse conditions, instead suggesting suitable indoor alternatives. Or information regarding local events could be used to suggest suitable activities according to one's profile. While not illustrated in Fig. 2, such contextual data could be integrated in health applications of platforms, such that it can be directly coupled with specific user data before being accessed by the HDT API for modeling purposes.

At any time, model developers can request specific standardized data via the HDT API (3.1). Upon receiving a request, the HDT API first verifies the developer's authorization to access specific user data. For each authorized user, the API identifies the relevant app or platform he/she uses for the requested health dimension and forwards the data request to the appropriate app/platform APIs (3.2a, 3.2b). Responses from these APIs can often be large and noisy (3.3a, 3.3b). To address this, the HDT API processes the incoming data, extracting only the relevant information and formatting it according to a unified predefined data standard, before returning the refined data to the model developers (3.4).

These processing capabilities also allow different health applications to provide overlapping subsets of data for a specific health dimension. For instance, nutrition tracking applications like MyFitnessPal and Cronometer may both supply nutrient intake data. By categorizing this data under a unified label such as "nutrition," the HDT API enables third-party model developers to easily access all available nutrition data in a standardized and aggregated format, regardless of its source application.

However, additional requirements must be addressed when providing data to third-party developers. Even with explicit user consent, the HDT API must ensure that only the minimum amount of data necessary for the development of the authorized model is provided [43]. Additionally, anonymization and pseudonymization techniques can further enhance security by protecting sensitive user information while still allowing for meaningful data analysis and model development [12], [27], [44]. These safeguards are essential for maintaining user trust and ensuring compliance with regulations such as GDPR [45].

Data retrieved through the HDT API can be used for model training purposes. Fig. 3 zooms in on the model training segment (4.1) in Fig. 2, presenting an overview of several modeling techniques for HDTs, each with unique advantages and limitations. The figure uses color-coding to distinguish models likely owned or managed by third-party developers from those managed within the core HDT system.

The first category involves theory- or formula-based models (Fig. 3a), which are based scientific principles and deterministic rules, requiring no training.

Population-level AI models (Fig. 3b) are ideal for an HDT systems early deployment stages when user-specific data is scarce. These models can be trained on large datasets outside the HDT system and provide generalized insights suitable for classification tasks or analyzing trends. Techniques like using synthetic data can enable third-party developers to train models without access to real user data, mitigating privacy concerns [46]. However, such models may overfit to population biases and struggle with personalization.

In contrast, user-specific models (Fig. 3c) are tailored to individual behaviors and needs, trained using data exclusively from a single user. These models excel in personalization but often face challenges in achieving high accuracy due to limited data availability. Unlike population-level models, synthesized data is often unsuitable for user-specific models, as it is more likely to compromise their accuracy and user-specificity. Therefore, allowing third party developers to contribute directly to the development of such user-specific models while maintaining privacy and security remains challenging.

A promising intermediate approach is transfer learning (Fig. 3d). This technique uses pre-trained models developed on large, population-level datasets and fine-tunes them with individual user data. Transfer learning effectively addresses the "cold-start problem," allowing models to adapt to user-specific data while leveraging global insights [47]. While transfer learning facilitates collaboration and personalization, it offers limited value back to the population-level third-party model, as the fine-tuned user-specific models remain with the user.

Personalized federated learning (Fig. 3e) builds upon transfer learning by combining individual user-specific training with global population-level insights. In this approach, the fine-tuned parameters from user-specific models are aggregated to enhance the global model without exposing raw user

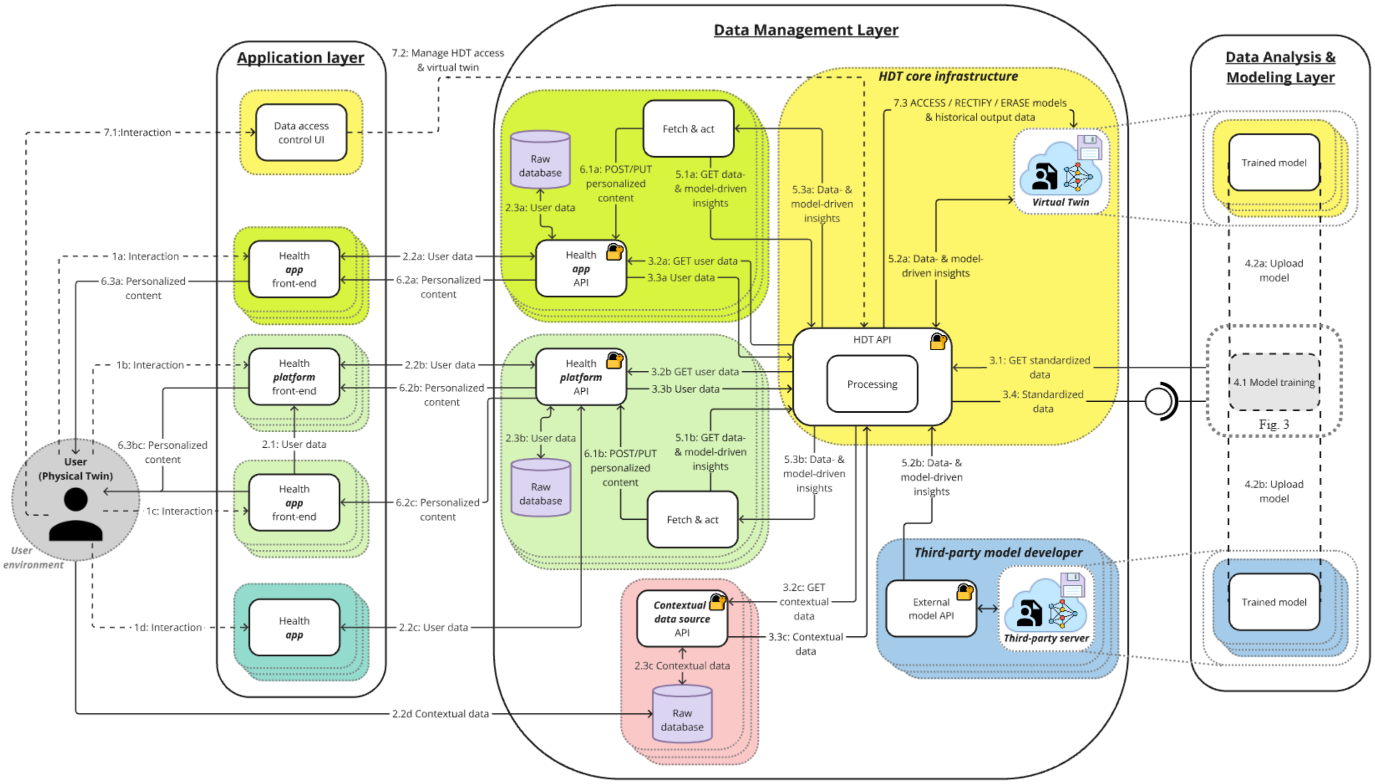

Fig. 2. Proposed interoperable and modular system architecture for lifestyle-oriented HDTs

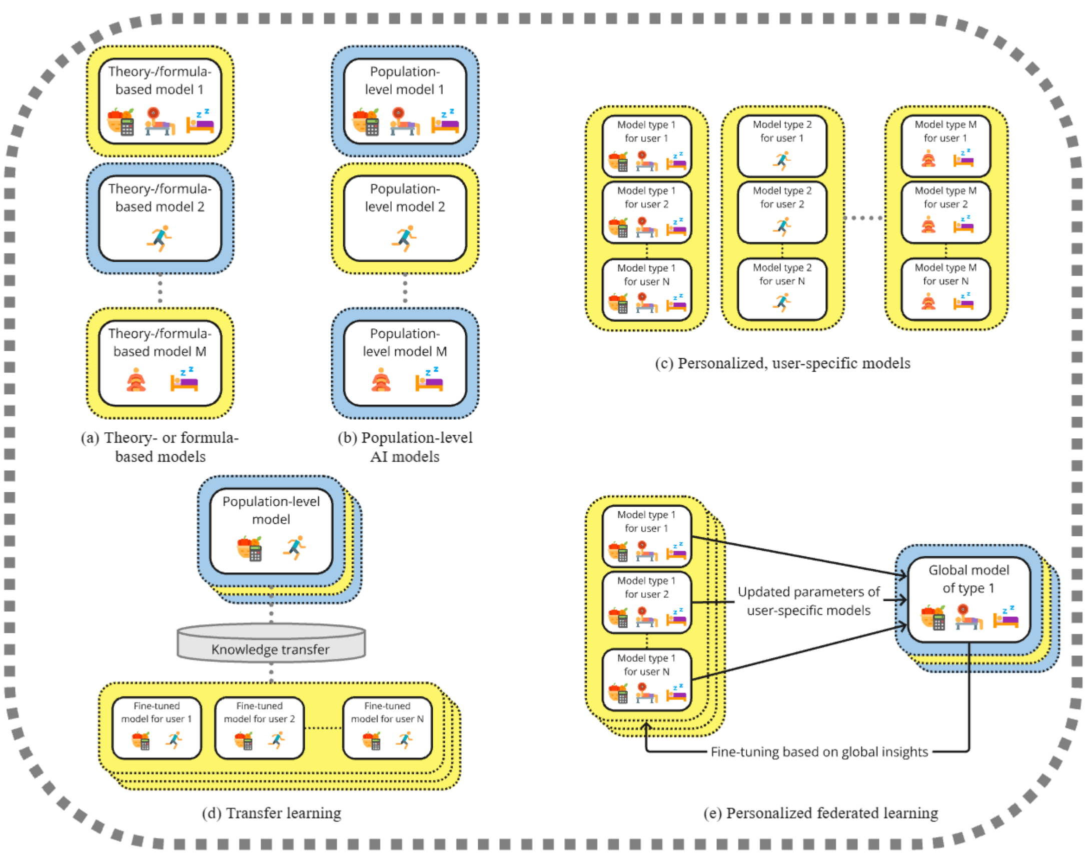

Fig. 3. Various types of modeling techniques for modular HDT systems

data [48]. This dual-layered approach balances personalization with collective knowledge generation, creating value for both users and developers. However, it introduces challenges, including managing synchronization, communication costs, and versioning complexities across diverse user bases.

This overview highlights the variety of modeling techniques available for HDTs, showcasing their potential applications and limitations. However, it is not intended to provide exhaustive coverage, as numerous other modeling techniques and hybrid combinations also exist.

After training or refinement, models are uploaded to the virtual twin. User-specific models are stored within the the virtual twin storage of the core HDT infrastructure (Fig. 2, 4.2a), ensuring that users retain full control over their personal models. Additionally, third-party models can also be hosted on these centrally managed servers. However, third-party developers with commercial interests may view their trained models as valuable intellectual property and choose to keep them on their own servers (4.2b). In this case, the central HDT API functions as a marketplace, facilitating connections between health application or platform developers and third-party virtual twin services.

To support such a marketplace, each model upload should be accompanied by comprehensive metadata. This metadata should not only include details on versioning, usage specifications, and accessibility but also provide extensive information on quality-of-service metrics and pricing. By offering this structured metadata to application and platform developers, they can effectively evaluate and compare virtual twin models with similar functionality, and select the one that best suits their specific needs.

While the internal structure of health apps and platforms may differ, the ‘Fetch & act’ modules are used to represent their ability to requests data- and model-driven insights via the HDT API (5.1a, 5.1b).

Depending on the model queried, these requests should be accompanied by the respective input data required. However, for comprehensive models that use data from multiple lifestyle dimensions (nutrition, physical activity, sleep, etc.) individual health apps may not be able to provide the input required for these model. Therefore, if these types of virtual twin models are also to be used by separate health apps, without any integration with a central health platform, some form of (temporary) data storage and aggregation capabilities are necessary inside the core HDT infrastructure.

If the correct input is supplied, the resulting data- and model-driven insights are returned (5.2a, 5.2b, 5.3a, 5.3b) and simultaneously stored in the virtual twin (5.2a, 5.2b).

By leveraging these insights, personalized content can be delivered to the user, either through separate health applications (6.2a) or health platforms (6.2b) and their integrated applications (6.2c).

Given that the HDT system manages sensitive health and lifestyle-related user data, numerous data flows within the system necessitate explicit user consent. Consequently, a centralized user interface is essential, enabling users to view and manage which entities have access to specific components of their virtual twin (7.1, 7.2). Users must maintain full control over their data and models at all times, with the ability to access, rectify, and erase them, in accordance with GDPR guidelines [45].

V. Prototype of an Interoperable and Modular Lifestyle-Oriented HDT System

A proof-of-concept prototype was developed to demonstrate and verify the feasibility of the proposed interoperable and modular system architecture for lifestyle-oriented HDTs by addressing the system requirements outlined in Section IV-A, thereby establishing a foundation for future extensions and improvements.

To achieve this, the prototype builds upon an existing HDT system, adapting and extending it to align with the proposed architecture. The selected HDT system and its components are first described. Next, the introduction of the separate HDT API is discussed. Lastly, the key modules and data flows of the resulting prototype are outlined, discussing how they fulfill the system requirements.

1) Description of the selected HDT system: The HDT system by De Oliveira et al. [49] is designed to facilitate the delivery of personalized content to patients with type 2 diabetes. It lets players interact with a health-promoting tool, an educational tool, and a knowledge test tool.

The health-promoting tool in this HDT is GameBus. This platform allows users to log a wide variety of health-related activities, either manually or automatically through integrations with third-party applications, such as the educational and knowledge testing tools [50]. The platform includes challenges that award points for participation in specific activities, enabling users to compete on leaderboards, gain levels, and/or open loot boxes [50].

Integrated with GameBus is the educational tool SugarVita, a virtual board game designed to educate newly diagnosed diabetes patients. The game simulates a day in the life of a diabetic patient, challenging players to maintain healthy blood sugar levels [51].

Complementing SugarVita, Trivia Quizzes feature questionnaires for testing one’s knowledge regarding diabetes. Here, hints can be used, that are obtained through previously earned points in GameBus.

Based on playthrough data from SugarVita and the Trivia Quizzes, two machine learning models are trained: one to classify a user’s player type and another to assess their diabetes health literacy. These models’ outputs enable personalization of content, such as adjusting quiz difficulty or tailoring rewards in GameBus.

2) Addition of a separate HDT API: While GameBus provides input interoperability by integrating data from SugarVita and Trivia Quizzes into its central database, the original system by De Oliveira et al. [49] retrieves and processes raw data to a usable form for model training within the virtual twin modules.

However, in the context of this paper, processing of app- or platform-specific data should not be performed inside the virtual twin modules, as this hinders the models’ ability to effectively access similar data provided by diverse sources (for example if health platforms also support the integration of SugarVita and/or the Trivia Quizzes).

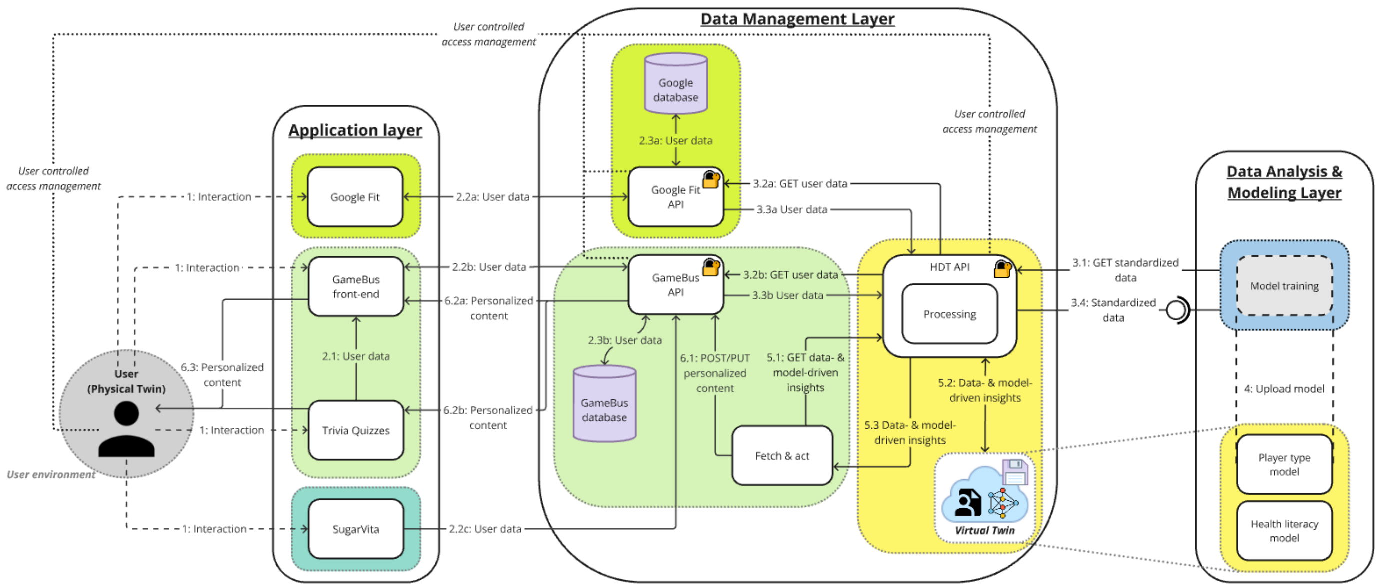

Fig. 4. Prototype of an interoperable and modular HDT system architecture using GameBus as the digital health-promoting platform, incorporating models for player type classification and health literacy assessment in type 2 diabetes management.

Therefore, the prototype introduces a dedicated HDT API, decoupled from GameBus, to act as a central data management module, enabling seamless data exchange between between health platforms and applications on one side and various virtual twin models on the other. The introduction of this HDT API allows for integration of other health platforms, like Google Fit.

3) Key Modules and Data Flows: The architecture of the resulting prototype is shown in Fig. 4, illustrating its key components and the data flows between them.

Again, colors are used for representing the different parties involved and numbering is used to guide the reader through the figure.

Users interact with Google Fit, GameBus, Trivia Quizzes, and SugarVita (1). Data from these interactions is stored in the respective raw databases, managed by Google (2.2a, 2.3a) and GameBus (2.1, 2.2b. 2.2c, 2.3).

At any time, model developers can request specific standardized data via the HDT API (3.1). Upon receiving a request, the HDT API verifies the developer's authorization for the requested data and identifies the appropriate connected health apps for the user. Based on the user's configured data sources, the HDT API dynamically redirects the request to the relevant APIs, such as the Google Fit API (3.2a) or the GameBus API (3.2b). This flexible routing mechanism ensures that data can be aggregated seamlessly from diverse platforms.

Responses from these external APIs are often extensive and include detailed information about their respective internal data structures (3.3a, 3.3b). To simplify usage for model developers, the HDT API parses these responses, extracting only the relevant properties and formatting them according to a predefined data standard. This standardized data is then returned to model developers (3.4), enabling consistent and efficient access to the required information, regardless of the source. An example of the standardized JSON response for Trivia Quiz data retrieval is provided in Fig. 5.


```
[
  {
    "data": {
      "latest_activity_info": {
        "id": 793916,
        "timestamp": "2023-09-15 13:53:22"
      },
      "trivia_results": {
        "NO_HINT_TYPE_OF_ANSWER": {
          "CORRECT": 41,
          "INCORRECT": 16
        },
        "WITH_HINT": {
          "FALSE": 57,
          "TRUE": 4
        }
      }
    },
    "user_id": 1
  },...
]
```


Fig. 5. Part of an example JSON response of the HDT API call requesting Trivia quiz playthrough data

Standardized data retrieved from the HDT API is used to train or refine virtual twin models, which are subsequently uploaded to the virtual twin (4). Each model is accompanied by metadata—detailing its version, input-output specifications, performance metrics, and accessibility options—enabling external platforms to make informed decisions about model usage.

Although not explicitly displayed in 4, the HDT API includes two metadata endpoints that provide system transparency and accessibility for developers. The first metadata API provides detailed information about the available data retrieval endpoints, including their functionality, input requirements, and output structure. This allows model developers to understand the data accessible through the HDT API and tailor their requests accordingly. The second metadata API offers insights into the virtual twin models, describing the stored models' metadata, such as version, input requirements, expected outputs, and performance metrics. These endpoints collectively serve as an initial layer, enabling developers to explore and navigate the system's capabilities efficiently.

Health platforms (such as GameBus and Google Fit) can query virtual twin models via the HDT API (5.1). This can be done either periodically, or be triggered by internal events. Depending on the model queried, this query may require specific input data. The resulting data- and model-driven insights are returned (5.2, 5.3) and simultaneously stored in the virtual twin, such that historical logs of model outputs remain available (5.2), enabling longitudinal analyses. An example of the JSON response for diabetes-related health literacy scores is presented in Fig. 6.


```
{
  "health_literacy_score": {
    "name": "diabetes",
    "score": 0.7721689397272327,
    "sources": {
      "sugarvita": 0.85,
      "trivia": 0.7202815662120545
    }
  },
  "latest_update": "2024-12-05T23:27:47Z",
  "user_id": 1
}
```


Fig. 6. An example JSON response of the HDT API call requesting the diabetes related health literacy scores for user 1

Insights from the virtual twin enable personalized content delivery to users through the GameBus platform (6.2a), integrated third-party applications (6.2b), and potentially other platforms. Since these applications function as both data providers and content delivery mechanisms, the prototype clearly demonstrates how data collection is inherently embedded in the application layer.

The lock and key icons in Fig. 4 represent the authentication and authorization mechanisms in place to ensure privacy and security. User consent governs data flows, including third-party integrations and data/model access requests via the HDT API. Consent mechanisms allow users to control access at a granular level, deciding exactly which external parties have access to which API endpoints, upholding transparency and trust.

Throughout the prototype, JSON is utilized as the primary data exchange format due to its flexibility, readability, and compatibility with diverse systems [34], [36].

To fulfill the final requirement, supplementary documentation for the prototype is made publicly accessible via GitHub [52]. This repository includes the complete codebase for all components of the prototype, alongside Swagger documentation for the API endpoints [53], encouraging local deployment, testing, and further development.

By addressing the system requirements for interoperability and modularity, this prototype demonstrates the feasibility of the proposed system architecture.

VI. Discussion

This chapter discusses the key challenges and opportunities related to the concept of interoperable and modular HDT systems for promoting healthy lifestyle behaviors. It highlights limitations and outlines directions for future work:

- The importance of standardized data exchange protocols cannot be overstated when discussing the development of interoperable and modular HDT systems. While standards like HL7's FHIR provide comprehensive frameworks, widely recognized in clinical healthcare settings, their extensive scope and complexity could create implementation barriers for developers of consumer-oriented lifestyle health applications, operating in domains that do often not require the detailed clinical data structures that FHIR was designed to support [40].
  Recognizing these challenges, initiatives like Open mHealth have shown great promise for facilitating a more lightweight data exchange standard, specifically designed to integrate diverse, non-clinical data types [54]. When appropriate, Open mHealth refers to established healthcare terminology standards such as SNOMED CT or LOINC, but it also uses specific naming conventions better suited for lifestyle health applications [54]. While Open mHealth represents a step in the right direction, it currently lacks the widespread recognition and support of standards like FHIR.
  In the absence of universally adopted data exchange standards, early-stage lifestyle-oriented HDT systems must rely on intermediary processing and mapping capabilities. These can be implemented either within health platforms (e.g., Google Fit, Apple HealthKit, GameBus) or within the HDTs core infrastructure, as demonstrated in this thesis. While the central HDT API in this work includes mechanisms for processing and structuring data from diverse, non-standardized sources, it does not fully map this data to any specific, widely accepted standard. This limitation leaves room for future work to explore the adoption or integration of established standards. However, the computational overhead associated with these transformations may limit the efficiency and scalability as the volume and diversity of data sources grow.
  Therefore, the widespread adoption of data exchange standards among health applications would not only simplify the integration of data from multiple sources but also enhance the efficiency and scalability of interoperable, modular HDT systems.
- The concept of SMART on FHIR [42], mentioned in this paper as an example framework for enabling secure and standardized access to virtual twins, could itself be leveraged to integrate clinical data into lifestyle-oriented HDT systems. This integration would enable the training of advanced virtual twin models capable of predicting health risks, such as assessing the likelihood of developing Type 2 diabetes based on combined activity levels, dietary habits, and lab results. In the future, such an approach holds the potential to bridge the gap between clinical and consumer healthcare domains.
- Ensuring the discoverability of virtual twin services requires some form of structured registry, such as frameworks like UDDI [55], to enable developers to identify

available models and make informed choices based on detailed metadata. This metadata should include service descriptions, data requirements, performance metrics, pricing, and integration guidelines. While the prototype presented in this paper addresses this need by incorporating an initial metadata API layer, future work should aim to develop more extensive registries and promote their widespread adoption.

- Incorporating gamification elements (e.g., challenges, rewards, progress tracking) [56] and social features (e.g., leaderboards, community support) [57] into health applications has shown significant promise for enhancing user engagement [9]
  While some digital health applications and platforms already include gamification and social features, they often lack the adaptability needed to fully engage users with diverse health profiles and evolving goals. This limitation can lead to challenges that are too difficult or too easy relative to the rewards offered, or misaligned with users' personal objectives, often resulting in frustration or disengagement for those who feel disadvantaged [58]. Additionally, this also means that it remains difficult to enable fair and motivating competition among users of varying fitness levels.
  HDT technology, while not inherently responsible for creating gamification or social features, has the potential to enhance their effectiveness. By leveraging real-time data- and model-driven insights, HDTs could enable dynamically adaptive challenge difficulties, rewards sizes, and social interactions that are better aligned with individual users' abilities, preferences, and progress [59]. Given this potential, future research should implement and evaluate the effectiveness of combining HDT technology with gamification and social features in health applications to enable more tailored experiences that align with users' evolving progress in their unique health journeys.
- Finally, to achieve true real-time digital twins for lifestyle health behaviors, data exchange and processing must be automated and streamlined even further to ensure that new user data triggers the generation of insights and the updating or fine-tuning of virtual twin models seamlessly. Future research should explore methods to optimize and automate the initiation of data flows to support this continuous process effectively.

VII. Conclusion

The findings of this paper lay a foundation for the future development of interoperable and modular HDT systems designed to promote healthy lifestyle behaviors. By enabling seamless data integration from diverse sources and supporting the open development and integration of advanced virtual twin models, these HDT systems should empower various health applications and platforms to deliver personalized experiences that dynamically adapt to users' evolving health needs, goals, environmental contexts, and technological advancements, ultimately fostering sustained behavior change and improving long-term health outcomes.

References

[1] WHO Regional Office for Europe, "Who european regional obesity report 2022," pp. 5–7, 2022.

[2] D. O. Okorodudu, M. F. Jumean, V. M. Montori, A. Romero-Corral, V. K. Somers, P. J. Erwin, and F. Lopez-Jimenez, "Diagnostic performance of body mass index to identify obesity as defined by body adiposity: a systematic review and meta-analysis," *International Journal of Obesity*, vol. 34, no. 5, pp. 791–799, 2010.

[3] A. Javed, M. Jumean, M. H. Murad, D. Okorodudu, S. Kumar, V. K. Somers, O. Sochor, and F. Lopez-Jimenez, "Diagnostic performance of body mass index to identify obesity as defined by body adiposity in children and adolescents: a systematic review and meta-analysis," *Pediatric Obesity*, vol. 10, no. 3, pp. 234–244, June 2015.

[4] J. Wong, S. O'Neill, B. Beck, M. Forwood, and S. Khoo, "Comparison of obesity and metabolic syndrome prevalence using fat mass index, body mass index and percentage body fat," *PLoS ONE*, vol. 16, no. 1, p. e0245436, 2021. [Online]. Available: https://doi.org/10.1371/journal.pone.0245436

[5] G. Flores Mateo, E. Granado-Font, C. Ferré-Grau, and X. Montaña-Carreras, "Mobile phone apps to promote weight loss and increase physical activity: A systematic review and meta-analysis," *Journal of Medical Internet Research*, vol. 17, no. 11, p. e253, Nov. 2015. [Online]. Available: http://www.jmir.org/2015/11/e253/

[6] S. J. Iribarren, T. O. Akande, K. J. Kamp, D. Barry, Y. G. Kader, and E. Suelzer, "Effectiveness of mobile apps to promote health and manage disease: Systematic review and meta-analysis of randomized controlled trials," *JMIR mHealth and uHealth*, vol. 9, no. 1, p. e21563, 2021.

[7] M. Evans, J. de Courcy, E. de Laguiche, M. Faurby, C. L. Haase, K. S. Matthiessen, A. Moore, and J. Pearson-Stuttard, "Obesity-related complications, healthcare resource use and weight loss strategies in six european countries: The resource survey," *International Journal of Obesity*, vol. 47, no. 8, pp. 750–757, 2023.

[8] L. Irvin, L. A. Madden, P. Marshall, and R. V. Vince, "Digital health solutions for weight loss and obesity: A narrative review," *Nutrients*, vol. 15, no. 8, p. Article 8, 2023.

[9] R. Jakob, S. Harperink, A. M. Rudolf, E. Fleisch, S. Haug, J. L. Mair, A. Salamanca-Sanabria, and T. Kowatsch, "Factors influencing adherence to mhealth apps for prevention or management of noncommunicable diseases: Systematic review," *Journal of Medical Internet Research*, vol. 24, no. 5, 2022.

[10] M. Grieves, "Digital twin: Manufacturing excellence through virtual factory replication," 2014.

[11] M. Grieves and J. Vickers, *Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems*. Springer International Publishing Switzerland, 2017, pp. 85–113.

[12] M. Lauer-Schmaltz, P. Cash, J. Hansen, and A. Maier, "Towards the human digital twin: Definition and design - a survey," *arXiv*, vol. 2402.07922, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2402.07922

[13] J. Chen, C. Yi, S. D. Okegbile, J. Cai, and X. Shen, "Networking architecture and key supporting technologies for human digital twin in personalized healthcare: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 1, pp. 706–746, 2024. [Online]. Available: https://doi.org/10.1109/COMST.2023.3308717

[14] E. I. Aboulsafa, G. A. E. Khayat, and S. A. Elmorsy, "An educational human digital twin proposed model for personalized e-learning," in *2023 IEEE Afro-Mediterranean Conference on Artificial Intelligence (AMCAI)*, 2023, pp. 1–8. [Online]. Available: https://doi.org/10.1109/AMCAI59331.2023.10431503

[15] J. Chen, C. Yi, H. Du, D. Niyato, J. Kang, J. Cai, and X. Shen, "A revolution of personalized healthcare: Enabling human digital twin with mobile aigc," *IEEE Network*, pp. 1–1, 2024. [Online]. Available: https://doi.org/10.1109/MNET.2024.3366560

[16] S. D. Okegbile and J. Cai, "Edge-assisted human-to-virtual twin connectivity scheme for human digital twin frameworks," in *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, 2022, pp. 1–6. [Online]. Available: https://doi.org/10.1109/VTC2022-Spring54318.2022.9860619

[17] B. R. Barricelli, E. Casiraghi, J. Gliozzo, A. Petrini, and S. Valtolina, "Human digital twin for fitness management," *IEEE Access*, vol. 8, pp. 26 637–26 664, 2020. [Online]. Available: https://doi.org/10.1109/ACCESS.2020.2971576

[18] Y. W. Song, "Human digital twin, the development and impact on design," *Journal of Computing and Information Science in Engineering*, vol. 23, no. 060819, 2023. [Online]. Available: https://doi.org/10.1115/1.4063132

[19] E. Montini, A. Bettoni, M. Ciavotta, E. Carpanzano, and P. Pedrazzoli, "A meta-model for modular composition of tailored human digital twins in production," *Procedia CIRP*, vol. 104, pp. 689–695, 2021.

[20] W. Shengli, "Is human digital twin possible?" *Computer Methods and Programs in Biomedicine Update*, vol. 1, p. 100014, 2021. [Online]. Available: https://doi.org/10.1016/j.cmpbup.2021.100014

[21] G. Inc., "Google fit — google developers," https://developers.google.com/fit, accessed: November 15, 2024.

[22] A. Inc., "Healthkit — apple developer documentation," https://developer.apple.com/documentation/healthkit, accessed: November 15, 2024.

[23] M. W. Lauer-Schmaltz, P. Cash, J. P. Hansen, and A. Maier, "Designing human digital twins for behaviour-changing therapy and rehabilitation: A systematic review," vol. 2. Scopus, 2022, pp. 1303–1312. [Online]. Available: https://doi.org/10.1017/pds.2022.132

[24] R. Zhong, B. Hu, Y. Feng, H. Zheng, Z. Hong, S. Lou, and J. Tan, "Construction of human digital twin model based on multimodal data and its application in locomotion mode identification," *Chinese Journal of Mechanical Engineering*, vol. 36, no. 1, p. 126, 2023. [Online]. Available: https://doi.org/10.1186/s10033-023-00951-0

[25] B. Wang, H. Zhou, X. Li, G. Yang, P. Zheng, C. Song, Y. Yuan, T. Wuest, H. Yang, and L. Wang, "Human digital twin in the context of industry 5.0," *Robotics and Computer-Integrated Manufacturing*, vol. 85, p. 102626, 2024. [Online]. Available: https://doi.org/10.1016/j.rcim.2023.102626

[26] M. E. Miller and E. Spatz, "A unified view of a human digital twin," *Human-Intelligent Systems Integration*, vol. 4, no. 1–2, pp. 23–33, 2022. [Online]. Available: https://doi.org/10.1007/s42454-022-00041-x

[27] G. Sirigu, B. Carminati, and E. Ferrari, "Privacy and security issues for human digital twins," in *2022 IEEE 4th International Conference on Trust, Privacy and Security in Intelligent Systems, and Applications (TPS-ISA)*, 2022, pp. 1–9. [Online]. Available: https://doi.org/10.1109/TPS-ISA56441.2022.00011

[28] W. Hafez, S. Elshamy, A. Farid, and R. Camara, "Transforming ai solutions in healthcare—the medical information tokens," in *2023 IEEE Conference on Artificial Intelligence (CAI)*, 2023, pp. 300–301. [Online]. Available: https://doi.org/10.1109/CAI54212.2023.00134

[29] Y. Liu, L. Zhang, Y. Yang, L. Zhou, L. Ren, F. Wang, R. Liu, Z. Pang, and M. J. Deen, "A novel cloud-based framework for the elderly healthcare services using digital twin," *IEEE Access*, vol. 7, pp. 49 088–49 101, 2019. [Online]. Available: https://doi.org/10.1109/ACCESS.2019.2909828

[30] S. D. Okegbile, J. Cai, D. Niyato, and C. Yi, "Human digital twin for personalized healthcare: Vision, architecture and future directions," *IEEE Network*, vol. 37, no. 2, pp. 262–269, 2023. [Online]. Available: https://doi.org/10.1109/MNET.118.2200071

[31] T. Vats, S. K. Singh, S. Kumar, B. B. Gupta, S. S. Gill, V. Arya, and W. Alhalabi, "Explainable context-aware iot framework using human digital twin for healthcare," *Multimedia Tools and Applications*, 2023. [Online]. Available: https://doi.org/10.1007/s11042-023-16922-5

[32] M. Viceconti, M. De Vos, S. Mellone, and L. Geris, "Position paper from the digital twins in healthcare to the virtual human twin: A moon-shot project for digital health research," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 1, pp. 491–501, 2024. [Online]. Available: https://doi.org/10.1109/JBHI.2023.3323688

[33] E. Vildjiounaite, J. Kallio, J. Kantorovitch, A. Kinnula, S. Ferreira, M. A. Rodrigues, and N. Rocha, "Challenges of learning human digital twin: Case study of mental wellbeing: Using sensor data and machine learning to create hdt," in *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments*, 2023, pp. 574–583. [Online]. Available: https://doi.org/10.1145/3594806.3596538

[34] S. Davila-Gonzalez and S. Martin, "Human digital twin in industry 5.0: A holistic approach to worker safety and well-being through advanced ai and emotional analytics," *Sensors*, vol. 24, no. 2, p. 655, 2024. [Online]. Available: https://doi.org/10.3390/s24020655

[35] Z. Wang, R. Gupta, K. Han, H. Wang, A. Ganlath, N. Ammar, and P. Tiwari, "Mobility digital twin: Concept, architecture, case study, and future challenges," *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17 452–17 467, 2022. [Online]. Available: https://doi.org/10.1109/JIOT.2022.3156028

[36] P. Thamotharan, S. Srinivasan, J. Kesavadev, G. Krishnan, V. Mohan, S. Seshadhri, K. Bekiroglu, and C. Toffanin, "Human digital twin for personalized elderly type 2 diabetes management," *Journal of Clinical Medicine*, vol. 12, no. 6, p. 2094, 2023. [Online]. Available: https://doi.org/10.3390/jcm12062094

[37] M. W. Lauer-Schmaltz, I. Kerim, J. P. Hansen, G. M. Gulyás, and H. B. Andersen, "Human digital twin-based interactive dashboards for informal caregivers of stroke patients," in *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments*, 2023, pp. 215–221. [Online]. Available: https://doi.org/10.1145/3594806.3594824

[38] A. Sharotry, J. A. Jimenez, F. A. M. Mediavilla, D. Wierschem, R. M. Koldenhoven, and D. Valles, "Manufacturing operator ergonomics: A conceptual digital twin approach to detect biomechanical fatigue," *IEEE Access*, vol. 10, pp. 12 774–12 791, 2022. [Online]. Available: https://doi.org/10.1109/ACCESS.2022.3145984

[39] A. Gokhale and Others, "Tactile internet: Enhancing connectivity for real-time communication," *Journal of Network Innovations*, vol. 34, no. 2, pp. 123–145, 2022.

[40] K. Gopinathan, N. A. Kaloumenos, K. Ajmera, A. Matei, I. Williams, and A. Davis, "FHIR FLI: An Open Source Platform for Storing, Sharing and Analysing Lifestyle Data," in *Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health*. Funchal, Madeira, Portugal: SCITEPRESS - Science and Technology Publications, 2018, pp. 227–233.

[41] A. Parecki, "OAuth 2.0 Simplified," 2018, accessed: 2024-11-25. [Online]. Available: https://oauth2simplified.com/

[42] K. D. Mandl, J. C. Mandel, and I. S. Kohane, "Driving innovation in health systems through an apps-based information economy," *Cell Systems*, vol. 1, no. 1, pp. 8–13, Jul. 2015.

[43] A. Goldsteen, G. Ezov, R. Shmelkin, M. Moffie, and A. Farkash, "Data minimization for gdpr compliance in machine learning models," *AI Ethics*, vol. 2, no. 3, pp. 477–491, Aug. 2022.

[44] M. Hintze and K. E. Emam, "Comparing the benefits of pseudonymisation and anonymisation under the gdpr," *Journal of Data Protection and Privacy (JDPP)*, vol. 2, no. 2, p. 145, Dec. 2018.

[45] "General data protection regulation (gdpr) – official legal text," https://gdpr-info.eu/, 2018, accessed: 2024-11-25.

[46] T. Marwala, E. Fournier-Tombs, and S. Stinckwich, "The use of synthetic data to train ai models: Opportunities and risks for sustainable development," *United Nations University*, 2023, accessed: 2024-11-25. [Online]. Available: https://unu.edu/publication/recommendations-use-synthetic-data-train-ai-models

[47] L. Chen, F. Yuan, J. Yang, X. He, C. Li, and M. Yang, "User-specific adaptive fine-tuning for cross-domain recommendations," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 3, pp. 3239–3252, 2023.

[48] H. Li, J. Xu, M. Jin *et al.*, "Personalized federated learning with global information fusion and local knowledge inheritance collaboration," *Journal of Supercomputing*, vol. 81, p. 158, 2025.

[49] C. D. D. Oliveira, L. James, A. Khanshan, and P. V. Gorp, "Toward enhancing diabetes self-management with personalization through human digital twins for behavior change," in *Proceedings of Ninth International Congress on Information and Communication Technology*, ser. Lecture Notes in Networks and Systems, X.-S. Yang, R. S. Sherratt, N. Dey, and A. Joshi, Eds., vol. 1054. Singapore: Springer, 2025, iCICT 2024.

[50] P. Van Gorp and R. Nuijten, "8-year evaluation of gamebus: Status quo in aiming for an open access platform to prototype and test digital health apps," *Proceedings of the ACM on Human-Computer Interaction*, vol. 7, no. EICS, pp. 171:1–171:24, 2023.

[51] "Sugarvita: Het spel," https://www.sugarvita.nl/het-spel-2/, accessed: 2024-11-15.

[52] W. J. van Vroonhoven, "Interoperable and modular hdt system prototype," https://github.com/pimvanvroonhoven/Interoperable-and-modular-HDT-system-prototype, 2024, accessed: 2024-12-05.

[53] W. J. Vroonhoven, "Hdt api documentation," https://pimvanvroonhoven.github.io/Interoperable-and-modular-HDT-system-prototype/, 2024, accessed: 2024-12-05. [Online]. Available: https://pimvanvroonhoven.github.io/Interoperable-and-modular-HDT-system-prototype/

[54] Open mHealth, "Schema library," 2019, accessed: 2024-11-25. [Online]. Available: https://www.openmhealth.org/documentation/#/schema-docs/schema-library

[55] IBM, "Universal description, discovery, and integration (uddi)," 2024, accessed: 2024-11-25. [Online]. Available: https://www.ibm.com/docs/en/rsm/7.5.0?topic=standards-universal-description-discovery-integration-uddi

[56] S. Deterding, R. Khaled, L. Nacke, and D. Dixon, "Gamification: Toward a definition," in *Proceedings of the CHI 2011 Gamification Workshop*, 2011, pp. 12–15.

[57] C. A. Maher, L. K. Lewis, K. Ferrar, S. Marshall, I. De Bourdeaudhuij, and C. Vandelanotte, "Are health behavior change interventions that use online social networks effective? a systematic review," *Journal of Medical Internet Research*, vol. 16, no. 2, p. e40, 2014.

[58] J. Nakamura and M. Csikszentmihalyi, "The concept of flow," in *Flow and the Foundations of Positive Psychology*. Dordrecht: Springer Netherlands, 2014, pp. 239–263.

[59] R. Orji, G. F. Tondello, and L. E. Nacke, "Personalizing persuasive strategies in gameful systems to gamification user types," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Apr. 2018, pp. 1–14.