Asia Pacific Academy of Science Pte. Ltd.

Metaverse 2026, 7(3), 8553.
https://doi.org/10.54517/m8553

Perspective

Multimodal Ontology: Semantic Substrate for Human Digital Twins in AI-Driven Metaverse Telerehabilitation and Immersive Medical Education

Oleksandr Palagin[1], Serhii Matvieishyn[1], Oleksii Bahatskyi[1], Dmytro Vakulenko[2], Kyrylo Malakhov[1,*]

[1] V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine, Kyiv, Ukraine
[2] I.Ya. Horbachevsky Ternopil National Medical University, Ternopil, Ukraine
*** Corresponding author:** Kyrylo Malakhov, k.malakhov@incyb.kiev.ua

CITATION

Palagin O, Matvieishyn S, Bahatskyi O, Vakulenko D, Malakhov K. Multimodal Ontology: Semantic Substrate for Human Digital Twins in AI-Driven Metaverse Telerehabilitation and Immersive Medical Education. Metaverse. 2026; 7(3): 8553. https://doi.org/10.54517/m8553

ARTICLE INFO

Received: 19 February 2026
Accepted: 10 June 2026
Available online: 15 July 2026

Abstract: The convergence of the metaverse with AI and digital twin technologies is reshaping healthcare delivery and medical education, yet Human Digital Twins (HDTs) – continuously updated virtual patient models – often lack a unifying semantic substrate capable of integrating heterogeneous multimodal evidence (sensor streams, clinical text, imaging, and interaction traces) into actionable meaning for perception, reasoning, and decision support. This viewpoint positions multimodal ontology as that substrate and as a prerequisite for robust AI-driven metaverse workflows in telerehabilitation and immersive medical education. Multimodal ontology is defined at a high level of abstraction and specified via a minimal formalization capturing its essential elements: domain concepts and relations, modality sets, semantic grounding operators that map raw observations to ontological entities, and state-update functions that maintain a temporally evolving HDT knowledge state. Building on recent HDT-enabled telerehabilitation work, the ontology layer is shown to operationalize core metaverse AI functions – environmental perception, semantic understanding, behavioral planning, and human–machine collaboration through interoperable representations, rule- and graph-based inference, and explainable agent behavior across clinician, patient, and intern interfaces. A testable research agenda is outlined for ontology-backed HDTs, emphasizing integration with large-scale multimodal models, validation in simulation and real-world deployments, and evaluation using semantic consistency, reasoning fidelity, user trust, and patient or intern outcomes. The specific contribution is not a new classifier, simulator, or network-optimization mechanism; rather, it is a formal semantic substrate that couples multimodal grounding, confidence-weighted fusion, uncertainty-aware HDT state evolution, and ontology-constrained interaction with large-scale foundation models.

Keywords: Metaverse; Human Digital Twin; Multimodal Ontology; Telerehabilitation; Immersive Medical Education; Knowledge Graph; Semantic AI; Digital Health; Semantic Grounding; Edge-Enabled HDT; Foundation Models; Explainable AI

1. Introduction

The metaverse–a persistent and immersive virtual space–is poised to transform how we work, learn, and receive healthcare by seamlessly blending physical and digital environments [1–3]. Recent advances enable users to engage with rich 3D simulations and avatars in real time, opening new frontiers in telemedicine and education. In healthcare, the metaverse vision converges with the concept of digital twins, virtual replicas of physical entities. A *Human Digital Twin* (HDT) is a continuously updated virtual patient model reflecting an individual's state across molecular, physiological, and behavioral dimensions [4]. HDTs have been heralded as a catalyst for personalized, proactive care–supporting remote monitoring, diagnosis, rehabilitation, and in silico experimentation in a risk-free digital setting [3,4]. For example, Ukraine's recent HDT-based telerehabilitation (telerehab) initiative couples patient digital twins with AI agents to enable remote patient monitoring, personalized therapy adjustments, and simulation-based medical training [3]. This integration allows clinicians and students to interact with virtual patients in realistic scenarios, thereby converging digital health and digital education on a common platform [1,3,5–7].

In this study, Immersive Medical Education (IME) denotes a mode of clinical training in which learners develop diagnostic, procedural, and communication competencies through high-fidelity interactive simulation delivered via VR/AR/MR interfaces (and, when relevant, instrumented physical simulators), while learner actions and multimodal traces are captured for assessment and adaptive feedback (see Definition 1). In the metaverse context, IME naturally aligns with HDT workflows because both depend on coherent state modeling, semantic interoperability across modalities, and closed-loop human–machine collaboration in real time.

AI-driven metaverse technologies play a major role in realizing these scenarios. Intelligent agents endow virtual environments with capabilities for *environmental perception, semantic understanding, behavioral planning, and human–machine collaboration*. In a telerehab use case, computer vision models perceive a patient's movements via camera feed (perception), interpret clinically relevant features like range of motion or gait symmetry (semantic understanding), formulate therapy recommendations (planning), and coordinate with human therapists or other AI assistants (collaboration) [3]. Likewise in an educational simulation, an AI avatar can perceive a trainee's questions or actions, understand the medical context, plan an appropriate response or clinical scenario progression, and interact with the human user in a collaborative dialogue. Achieving this level of intelligence and coherence in the metaverse is challenging because it demands *contextualized, multimodal knowledge*. Raw sensor data alone (images, wearables, electronic medical record/electronic health record (EMR/EHR) entries, etc.) do not yield insight unless grounded in a semantic framework that reflects medical concepts, patient-specific information, and logical relationships.

Computer ontologies – formal representations of knowledge as sets of concepts and relationships – are widely recognized as a solution for integrating heterogeneous data and enabling machine reasoning [8]. Gruber's classic definition [9] describes an ontology as "an explicit formal specification of a conceptualization," capturing domain knowledge through defined entities, attributes, and relations. By providing a shared vocabulary and logical structure, ontologies facilitate interoperability across data sources and support automated reasoning over complex systems. In dynamic domains like the metaverse, an ontology can serve as the "world model" that anchors an AI agent's understanding of the environment [10]. Notably, recent works in *digital health* propose combining digital twins with knowledge graphs or personal ontologies to create semantically enriched, interoperable patient models [11,12]. Such *ontologically augmented digital twins* – sometimes termed *cognitive twins* – incorporate *Semantic Web* technologies to enhance knowledge representation, context-awareness, and decision support. For instance, Carbonaro et al. [12] describe a multi-layer healthcare platform where heterogeneous data streams are harmonized and fed into general-purpose patient digital twins backed by personal knowledge graphs; these ontologized twins then become the foundation for building smart clinical applications. Similarly, in industrial IoT (and IoMT) settings, semantic data spaces (e.g. using the NGSI-LD ontology standard) have enabled traceable, interpretable interactions among modular AI services operating over digital twin models [13]. These trends reflect a broader consensus that injecting symbolic knowledge structures into AI systems greatly improves their ability to understand and explain complex multimodal information [14,15].

Despite the rise of powerful foundation models, such as large language models (LLMs) and multimodal transformers that automatically learn representations from large datasets, a *semantic gap* remains in today's AI. Deep neural networks excel at pattern recognition but their learned "knowledge" is often implicit and opaque [15]. For example, a multimodal model might learn the correlation between a motion sensor time series and a patient's fall risk, or between an MRI image and a diagnosis, but without an external knowledge scaffold it cannot articulate the underlying concepts (like "balance instability" or "tumor size") or reason about causal relationships. Recent research has begun to *extract latent ontologies* from such models – for example, by revealing hidden concept hierarchies in a vision-language foundation model [15], and align them with human-understandable ontologies for validation [15]. These efforts underscore that even advanced AI benefits from an explicit knowledge layer. In the context of HDTs, a *multimodal ontology* is therefore foundational for connecting low-level signals to high-level clinical semantics, ensuring that digital twin states remain both *machine-operationalizable* and *clinically interpretable*.

This perspective advances the argument that a multimodal ontology functions as the semantic substrate for HDTs in AI-driven metaverse applications. Related work is first synthesized across metaverse technologies, digital twins, multimodal learning, and ontologies (Section "Related Works"). A minimal formal model of multimodal ontology is then introduced (Section "Methodological Approach"), specifying core components such as concepts, modalities, semantic grounding, and state-update operators, and clarifying how these elements formally support an HDT system. The "Conceptual Results" section subsequently maps the ontology-augmented HDT to the principal metaverse AI functions, using telerehabilitation and immersive clinical education exemplars to substantiate improvements in perception, semantic understanding, behavioral planning, and human–machine collaboration. A scenario narrative and a tabulated alignment framework are provided to link each metaverse AI function to its ontological mechanisms and resulting system-level benefits. The manuscript concludes with a discussion of implications, open challenges, and a forward-looking research agenda in the "Discussion" section.

The novelty of the proposed semantic substrate is therefore positioned at the operational layer between raw multimodal evidence and higher-level HDT reasoning. Unlike a conventional knowledge graph that primarily stores typed entities and relations, and unlike many cognitive-twin frameworks that emphasize analytics, control, or application orchestration, the substrate specified here defines how modality-specific evidence is grounded into ontology concepts, how confidence, provenance, and time are carried into the evolving HDT state, how ontology axioms constrain state transitions, and how symbolic validation can audit generative modules. Its contribution is thus integrative and governance-oriented: it makes perception, multimodal fusion, temporal state maintenance, explainability, and foundation-model interaction operate over a shared clinically interpretable representation.

Claims are grounded in recent literature, prioritizing peer-reviewed open-access sources and formal studies, and in an anchor use-case reference [3] reporting practical outcomes from a collaborative program conducted jointly by two research teams at

I.Ya. Horbachevsky Ternopil National Medical University (TNMU) and the V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine (Institute of Cybernetics).

This Perspective was prepared within the institutional research context of a contract with I. Ya. Horbachevsky Ternopil National Medical University of the Ministry of Health of Ukraine for the R&D project “Development of an experimental prototype of an expert system with artificial intelligence elements for psycho-emotional screening” (https://prozorro.gov.ua/uk/contract/UA-2026-05-25-003617-a-a1). The study is also related to the broader national R&D initiative “Development of a comprehensive model of psycho-emotional screening with decision-making and patient-routing algorithms at the primary healthcare level,” carried out under a contract between the Ministry of Education and Science of Ukraine and I. Ya. Horbachevsky Ternopil National Medical University (State registration No. 0126U003034). The conceptual and infrastructure background also draws on the R&D projects “To develop theoretical foundations and a functional model of a computer for processing complex information structures” (State registration No. 0124U002317; https://nrat.ukrintei.ua/searchdoc/0124U002317/) and “Develop Means of Supporting Virtualization Technologies and Their Use in Computer Engineering and Other Applications” (State registration No. 0124U001826; https://nrat.ukrintei.ua/en/searchdoc/0124U001826), registered by the National Academy of Sciences of Ukraine.

By uniting insights from knowledge engineering and AI-driven digital health, this article aims to articulate a clear vision: multimodal ontologies will be a cornerstone of next-generation metaverse platforms for healthcare and education, providing the common semantic layer that allows human clinicians, interns, and AI agents to share understanding and cooperate effectively. In an era where the metaverse is increasingly seen as a paradigm for connected, intelligent services, ensuring a strong semantic foundation is not only a technical need but also a step toward safer, more transparent, and human-aligned AI.

2. Related Works

2.1 Metaverse in Healthcare and Education

The metaverse has rapidly evolved from a futuristic concept to concrete pilot applications across various sectors. In healthcare, the metaverse is envisioned as an interconnected network of virtual environments where patients, providers, and students can interact via avatars with a strong sense of presence [2]. *Immersive technologies* such as virtual reality (VR), augmented reality (AR), and mixed reality (MR) provide the interface, while AI provides contextual understanding and content generation.

Definition 1. Immersive Medical Education (IME) is a pedagogical paradigm in which learners acquire clinical knowledge and skills through high-fidelity, interactive simulations supported by VR/AR/MR interfaces (and, where relevant, physical simulators), enabling realistic experiential learning with measurable performance traces, adaptive feedback, and scenario replay within safe, controllable environments.

Recent surveys highlight the potential of metaverse platforms to reshape medical training and clinical care by enabling simulation, remote presence, and enriched data visualization [2]. During the COVID-19 pandemic [16], for example, VR and AR applications were employed to continue medical education when in-person learning was limited [17]. Ghaempanah et al. [17] review how metaverse-related technologies (virtual worlds, lifelogging via wearables, etc.) are being used in health education and practice. They note that virtual worlds like *Second Life* had early success in hosting medical training scenarios and patient support groups, and that the pandemic accelerated acceptance of *telepresence* and *digital simulations* [18] in healthcare. Within such metaverse settings, digital twins are identified as a key enabling technology: "Within the metaverse, digital twins enhance health care by enabling patient–doctor interactions via avatars and AR, transforming care, mental health support, and medical training" [17]. In other words, virtual representations of patients (or clinical scenarios) combined with immersive interfaces allow new forms of interaction that can improve access to care and enrich education/training with realistic, yet safe, experiences. Early examples include VR physiotherapy sessions where a patient's avatar performs prescribed exercises under remote supervision, or AR anatomy lessons where students interact with a 3D organ model that mimics real physiology.

2.2 Human Digital Twins in Personalized Healthcare

The concept of an HDT extends the digital twin paradigm (well established in engineering and Industry 4.0 [19]) to the human body and its health data [4]. An HDT can be viewed as a comprehensive personalized computational model of an individual, continuously fed by data from various sources (sensors, EMR/EHR, genomics, self-reports, etc.) [3,4]. Mokhtari [4] describes an HDT as "a dynamic virtual model that continuously reflects changes in molecular, physiological, emotional, and lifestyle factors" of a person. The HDT not only mirrors the current state for monitoring but also supports *predictive and prescriptive functions* – for example, simulating the outcome of a surgery or rehabilitation plan before it is applied, thus guiding decision-making. In practical terms, an HDT acts as a virtual proxy of the patient: it can be queried, analyzed, and even experimented on (in silico [18,20]) without risk to the actual patient. This capability is profoundly important for personalized medicine. HDTs promise to enable care that is proactive (predicting issues before they manifest), precise (tailored to the individual's unique profile), and participatory (engaging patients via visualizations of their own twin). *Representative use cases* include remote management of chronic diseases via an HDT that aggregates wearable data and alerts clinicians to any anomaly; surgical rehearsal on a patient-specific digital twin derived from imaging data; or, as in the Ukrainian telerehab project [3,21,22], a platform where HDTs of patients and *AI-driven avatars* of clinicians/pedagogical agents interact for therapy and training. Early implementations of HDTs in healthcare remain exploratory, but major health institutions and tech companies recognize their disruptive potential [3,23]. A 2024 scoping review [24] found that while many digital twin projects are still proof-of-concept, they consistently highlight improved monitoring and decision support as benefits, and call out the need for standards in modeling and data integration (to move from bespoke models to interoperable digital twin ecosystems).

2.3 Multimodal Learning and Foundation Models

Contemporary AI is increasingly multimodal, enabling the joint processing and alignment of text, images, and physiological time-series signals. In the HDT context, multimodal learning is foundational because digital twin evidence is inherently heterogeneous, spanning sensor telemetry (motion, vital signs), imaging (radiology, ultrasound), clinical text (diagnoses, notes), and laboratory results. Foundation models such as OpenAI's LLM GPT-5.2 and modern vision–language encoders such as Google's SigLIP 2 have demonstrated strong capabilities in representation learning and cross-modal alignment, supporting tasks from image–text retrieval to downstream multimodal reasoning [25].

With visual inputs, GPT-5.2 can generate structured natural-language interpretations of images, indicating the practical value of coupling visual representations to linguistic and clinical semantics. In safety-critical settings such as autonomous driving, large language and multimodal language models are increasingly used as an explanatory interface layered over perception and planning stacks, enabling natural-language rationales and question answering about decisions, and motivating the use of explicit concept schemata for controllable interpretation [26,27].

These developments support a broader design pattern in which foundation models supply high-capacity perception and prediction, while structured knowledge representations (ontologies, knowledge graphs, and rule systems) supply semantic constraints, interpretability, and partially verifiable reasoning. Purely neural approaches, however, remain vulnerable in specialized domains with stringent safety and accountability requirements, including healthcare, where hallucinations and spurious correlations can translate into clinical risk. Consequently, *knowledge-enhanced multimodal learning* [28] has gained traction, including methods that incorporate knowledge graph structure into training, inference, and retrieval pipelines.

In this direction, López et al. [14] argue that *multimodal knowledge graphs* can provide foundation models with domain context that improves multimodal fusion and yields more meaningful knowledge-enhanced representations in scientific and biomedical discovery workflows.

The implication for HDTs is direct: the analytic core of an HDT should combine data-driven representation learning with explicit domain semantics. A multimodal ontology can be treated as a constrained, clinically curated form of multimodal knowledge graph that mediates interpretation of incoming evidence and enforces consistency with medical concepts and protocols. This ontology layer also strengthens explainability; for example, if an HDT flags impending heart-failure decompensation, anchoring the alert in ontological entities such as edema progression or medication non-adherence can yield an auditable explanatory trace rather than an opaque score.

2.4 Ontologies and Knowledge Graphs in Digital Twins

The use of ontologies in digital twin systems has gained momentum as practitioners recognize the need for semantic interoperability and advanced reasoning. Karabulut et al. [11] conducted a systematic literature review on ontologies in digital twins and introduced the term *"Cognitive Twins"* to denote twins empowered by Semantic Web technology. In their analysis of 82 papers, they found ontologies being used in various twin architectures to integrate data, maintain a formal representation of the twin's state, and enable rule-based or logic-based inference (for example, triggering an alert when certain conditions are met in the twin's state space). Common benefits reported include enhanced *interoperability* (the ontology acts as a lingua franca between IoT devices, databases, and AI modules) and *automatic reasoning* (deducing high-level insights or predictions by applying ontological rules to the twin's data). In manufacturing and smart cities, frameworks like *FIWARE* use standard ontologies (NGSI-LD) for representing digital twin entities, so that different systems can subscribe to and consume twin data with shared semantics [13]. The healthcare domain, with its fragmentation of data formats and critical need for accurate knowledge encoding, stands to gain even more from ontology integration. A salient example is the *CONNECTED* framework proposed by Carbonaro et al. [12]: it defines a multi-level architecture where patient data from wearables, EMR/EHR, and other sources are first integrated and standardized (using health data interoperability standards), then fed into a *Personal Knowledge Graph* (PKG) that underlies a general-purpose patient digital twin. In this design, the knowledge graph (an instantiation of an ontology for the patient's domain) stores semantic relationships – for instance, linking a patient's symptom to possible causes or linking a rehabilitation exercise to the functional ability it targets. The patient's twin state is thus not just raw numbers, but a web of interconnected facts that smart applications can query via APIs. This reflects an emerging consensus – ontologized digital twins are key to moving from static data dashboards to intelligent systems that can reason about "why" and "what-if" scenarios. Indeed, a recent white paper on HDTs for healthcare [29] argues that knowledge graphs can represent complex multimodal healthcare data and serve as a scaffolding for integrating AI analytics, ultimately creating a learning loop where new data refines the knowledge and the knowledge guides interpretation of new data.

In summary, across these intersecting domains, a clear trend emerges toward integrating data-driven AI with symbolic ontologies. The metaverse imposes requirements for real-time, interactive intelligence that purely pre-trained models often cannot satisfy reliably under novelty, distribution shift, or data sparsity. HDTs exemplify a high-stakes class of applications in which patient outcomes and clinical training competence demand capabilities beyond statistical pattern recognition, namely semantically grounded interpretation and accountable reasoning. The reviewed literature provides convergent support that embedding an ontology or knowledge graph within an HDT architecture addresses this requirement by grounding multimodal observations in clinical context, enabling structured inference, and supporting coherent communication among clinicians, patients, interns, and AI agents. On this basis, the subsequent sections formalize multimodal ontology for HDTs and characterize its role as the semantic substrate of AI-driven metaverse systems.

2.5 Edge-Networked and Mobile HDT Infrastructures

Successful HDT deployment also depends on edge-networked infrastructures that maintain low latency, high availability, and secure human-to-virtual connectivity. A first group of studies addresses computation offloading and transmission scheduling. Yi et al. formulated multi-user computation offloading and uplink scheduling for delay-sensitive mobile applications as a mechanism-design problem over a queueing model [30], while Cao and Cai proposed a distributed game-theoretic machine-learning approach for multiuser offloading in cloudlet-based mobile cloud computing [31]. Earlier wireless relay work on semi-distributed user relaying for amplify-and-forward networks [32] is less HDT-specific but remains relevant to practical connectivity design because it reduces global channel-state exchange while preserving relay-selection efficiency.

A second group concerns dynamic edge deployment and distributed construction of the twin. Yang et al. studied dynamic HDT deployment at the edge for task execution through two-timescale accuracy-aware online optimization [33]. Okegbile et al. proposed differentially private federated multi-task learning to enhance human-to-virtual connectivity in HDT systems [34]. Chen et al. later formulated federated digital-twin construction through distributed sensing and overlapping coalitions, emphasizing partial-DT allocation, edge-sensor associations, and long-term optimization under time-varying conditions [35].

A third group focuses on semantic communication and immersive multimodal transmission. Chen et al. surveyed networking architectures and supporting technologies for HDT in personalized healthcare, including pervasive sensing, on-body/beyond-body communication, multi-access edge computing, semantic communication, blockchain, and explainable AI [36]. Zhu et al. framed NextG semantic communication as a pathway from data-mirror DTs to cognitive smart copilots [37]. QoE-aware resource-allocation work for RIS-assisted DT interaction [38] and joint visual-haptic transmission with adaptive compression for immersive HDT interaction [39] further show that metaverse-level experience quality depends on synchronization, bandwidth allocation, compression, and rendering decisions.

A fourth group concerns mobile AIGC and IoT-healthcare HDTs. Mobile AIGC has been proposed as an enabling technology for personalized healthcare HDTs because it can generate rare-disease data, support high-fidelity modeling, and provide customized services at the edge [40]. The IoT-healthcare survey by Chen et al. systematizes generative-AI-driven HDT realization across data acquisition, communication, management, digital modeling, and data analysis [41], while the survey on generative AI over mobile networks emphasizes large-model deployment, personalized model evolution, and immersive interactions in human-centric applications [42].

These studies provide the communication, resource-deployment, privacy, and QoE optimization layers needed for pervasive HDT services. The present perspective addresses a complementary layer: it specifies how evidence carried by such infrastructures becomes a semantically grounded, uncertainty-aware, clinically interpretable HDT state. Across metaverse applications, HDT architectures, multimodal AI, ontology-enhanced twins, and edge-networked HDT infrastructures, a clear convergence is now visible toward integrated neuro-symbolic and systems-aware frameworks.

3. Methodological Approach: A Multimodal Ontology for HDTs

3.1 Ontology Fundamentals and Extension to “Multimodal”

An ontology $\mathcal{O}$ in the context of knowledge representation can be defined as a tuple $\mathcal{O} = (C, R, A)$. Here C is a set of concepts (or classes) representing the key entities and ideas in the domain, R is a set of relations (or predicates) that connect these concepts (including hierarchical relations like is-a and associative relations like affects or part-of), and A is a set of axioms or assertions (which may include instance data, constraints, and rules that encode domain knowledge). This definition aligns with standard ontology definitions in AI: C corresponds to the formally specified concepts, R to relationships and rules, and A to logical sentences that are assumed true (e.g. “every rehabilitation exercise targets at least one functional ability” could be an axiom). Ontologies are typically written in formal languages (OWL, RDF, etc.) but here we remain at the conceptual level.

A *multimodal ontology* augments $\mathcal{O}$ by explicitly incorporating *modalities of data and their connection to the concepts*. We introduce:

- M – set of modalities;
- G – set of grounding functions or mappings.

Each modality $m \in M$ corresponds to a source or format of data (e.g. “textual clinical notes”, “wearable motion sensor data”, “medical imaging”, “user input in VR”, etc.). The grounding G links elements of C to one or more modalities in M by defining how evidence for a concept can be obtained from data. Formally, we can treat G as a set of functions or relations $g_{c,m} : \mathrm{Data}(m) \to \mathrm{Val}(c)$ for selected pairs $(c \in C, m \in M)$, where $\mathrm{Data}(m)$ is the space of raw data in modality *m* and $\mathrm{Val}(c)$ is the value space of concept *c*. In simpler terms, $g_{c,m}$ is a function that interprets raw data of modality *m* to estimate or update the value of concept *c*. For example, if *c* = *KneeFlexionAngle* (concept representing a patient's knee joint flexion) and *m* = *wearable Inertial Measurement Unit (IMU) sensor,* then $g_{c,m}$ could be an algorithm that computes the flexion angle from accelerometer/gyroscope signals on a leg-mounted device. If *c* = *DiseaseDiagnosis* and *m* = *textual EMR*, then $g_{c,m}$ might involve a natural language processing model extracting diagnosis codes from clinical text. A single concept can have multiple grounding functions from different modalities (e.g. a patient's *heart rate* concept could be grounded both in a smartwatch's PPG data and in manual pulse readings noted in text).

Operationally, each grounding function should return more than a verbal mapping. For a modality *m* and concept *c*, the grounding output may be represented as $G_{m,c}(x_m, t) = \langle \hat{v}_{m,c}(t), \alpha_{m,c}(t), \pi_{m,c}(t), \tau_m, s_{m,c}(t) \rangle$, where $\hat{v}_{m,c}(t)$ is the candidate concept value, $\alpha_{m,c}(t)$ is the confidence or calibrated probability, $\pi_{m,c}(t)$ identifies the data source and algorithmic model, τ_m is the observation timestamp, and $s_{m,c}(t)$ records whether the evidence is normal, missing, delayed, contradictory, or outside the validated operating range. Thus, the HDT does not merely assert that raw data map to a concept; it records the evidential basis and reliability of that assertion.

When several modalities ground the same concept, fusion can be performed by normalized weights $w_{m,c,t}$ that combine modality reliability, concept-specific clinical relevance, temporal freshness, and cross-modal agreement. For example, a wearable IMU may receive high weight for *BalanceStability* during a standing-balance task because postural sway, acceleration variance, jerk, and center-of-mass proxies are directly informative; a text medical record may receive high weight for *Diagnosis* when an NLP extractor maps a documented ICD/SNOMED-coded diagnosis to the corresponding ontology class with explicit provenance. Conflicting or low-confidence candidates are not forced into a single deterministic value; they are retained as competing evidence until reconciled by reliability rules, temporal context, or human review.

To illustrate, consider a tiny ontology for stroke rehabilitation:

- *C = Balance, GaitSpeed, ExerciseType, RiskOfFall*. These concepts cover patient balance status, walking speed, type of exercise being done, and risk level of falling;
- *R* might include relations like *affects(ExerciseType, Balance)* or a rule such as "if *Balance* is poor and *GaitSpeed* is low, then *RiskOfFall* is high".
- *M* could have *cameraVideo*, *wearableIMU*, *questionnaire*;
- Groundings *G* would map, for example, video data to *Balance* (via a computer vision model that assesses sway), IMU data to *GaitSpeed* (via step detection algorithm), and perhaps a questionnaire modality to *RiskOfFall* (asking the patient about dizziness or near-falls).

For *BalanceStability* specifically, a grounding may compute sway magnitude, step-width variability, acceleration entropy, and loss-of-balance events from IMU/video windows and then output a graded value such as good, moderate, or poor with confidence and timestamp. For *Diagnosis*, a text-grounding function may combine named-entity recognition, negation detection, section detection, and terminology normalization to map expressions in EMR/EHR notes to ontology classes, while preserving uncertainty when the diagnosis is suspected, historical, ruled out, or contradicted by another record.

The multimodal ontology thus explicitly ties *semantics (concepts)* to *observations (data)*. This is the key to making the HDT "live" in the metaverse: as the patient interacts with the system, their sensors stream data that are immediately interpreted in terms of the ontology's concepts, updating the known state of the twin. Likewise, any action or event in the virtual world (e.g. the patient's avatar attempting an exercise, or a student issuing a treatment command) can be translated into ontology updates.

3.2 State Representation and Update Mechanism

The state of an HDT at time t can be represented as a function $S_t : C \to \mathcal{V}$ that assigns each concept $c \in C$ a value $S_t(c)$ in its associated value space $\mathcal{V}c$. These values can be categorical (e.g. Balance = "poor"), numerical (e.g. GaitSpeed = 0.8 m/s), boolean, or even complex structures, depending on the concept's nature. The collection of all concept values (the image of S_t) constitutes a snapshot of the twin's semantic state at time t. Equivalently, the state may be expressed as a vector or tuple $(vc_1, v_{c_2}, ..., v_{c_n})$ corresponding to $(c_1, ..., c_n) \in C$, though some components may be interdependent as specified by ontology axioms.

The HDT's state evolves over time as new multimodal data arrive and as interventions (by human or AI agents) are applied. A state-update operator F produces the subsequent state from the prior state and newly acquired observations:

$$S_{t+1} = F\left(S_t, O_{t+1}\right), \tag{1}$$

where O_{t+1} denotes the set of observations collected over $(t, t+1]$. Each observation can be modeled as a pair (d, m) indicating data d obtained from modality m. The update operator F can be decomposed into two stages. First, *grounding and interpretation* apply modality-specific grounding functions $g_{c,m} \in G$ to map observed data into candidate updates for one or more concept values, potentially as point estimates or probabilistic assignments. Second, *ontology-mediated fusion* integrates these candidates with the prior state S_t, enforcing consistency with relations and axioms and inferring any logically implied updates. For example, if *ExerciseType* is detected as "tandem stance balance exercise" via a VR-interaction modality, axioms may impose heightened monitoring of *Balance* and a temporary expectation of $GaitSpeed = 0$ because the activity is stationary.

For high-frequency sensor streams, F should not invoke full ontology-wide reasoning for every raw sample. A practical implementation decomposes F into acquisition queues, sliding-window feature extraction at the edge, semantic event abstraction, debouncing of transient artifacts, and incremental reasoning over only the affected ontology subgraph. This design preserves real-time responsiveness because high-rate signals are first reduced to clinically meaningful events or interval summaries, while semantic consistency is enforced at commit time through range constraints, temporal constraints, and domain rules.

The state S_t should therefore be treated as an uncertainty-aware semantic state rather than a flat vector of point values. For each concept c, S_t may store a value or distribution, confidence, provenance, observation time, validity interval, and evidence status. Delayed observations are inserted according to their observation time tau rather than only their processing time, missing data cause confidence decay and stale-state flags, and contradictory evidence is held in a conflict set until resolved by modality reliability, clinical priority, temporal recency, or clinician adjudication.

In many cases, F can combine rule-based reasoning, Bayesian or evidential updating, and temporal-logic constraints. A deterministic view remains useful for exposition, but medical HDTs should represent probabilistic outcomes explicitly, for example as $P(c = v \mid E_t)$, belief intervals, or ranked candidate values attached to the concept c. This prevents premature semantic collapse when evidence is noisy, delayed, or conflicting, and it allows the HDT to expose uncertainty to clinicians and tutors instead of concealing it behind a single categorical state.

3.3 Minimal Mathematical Formalization

Summarizing the above, a multimodal ontology for an HDT can be defined as:

$$\mathcal{O}_{MM} = \left(C, R, A, M, G, F\right), \tag{2}$$

where (C, R, A) is the base ontology and (M, G, F) are the multimodal extensions (modalities, grounding mappings, and state-update function).

Definition 2. A Human Digital Twin is a tuple $(\mathcal{O}_{MM}, S_t)$ consisting of a multimodal ontology $\mathcal{O}_{MM}$ and a current state $S_t : C \to \mathcal{V}$. The HDT operates by evolving S_t over time via F as new multimodal observations are incorporated.

A more implementation-oriented state representation is $S_t(c) = \langle v_{c,t}, P_{c,t}, \alpha_{c,t}, \pi_{c,t}, \tau_{c,t}, \nu_{c,t}, \chi_{c,t} \rangle$, where $v_{c,t}$ is the selected or currently displayed value, $P_{c,t}$ is an optional probability distribution over candidate values, $\alpha_{c,t}$ is confidence, $\pi_{c,t}$ links the value to source observations and models, $\pi_{c,t}$ records observation time, $\nu_{c,t}$ marks the interval over which the assertion is considered current, and $\chi_{c,t}$ stores unresolved incompatible evidence. This formalization makes uncertainty and conflicting evidence first-class elements of the HDT state rather than post hoc annotations.

This formalism ensures that at any point, the HDT's state is well-defined in semantic terms (via *C* and *R*) and is directly connected to raw evidence (via *M* and *G*). It becomes the semantic substrate for all higher-level functionalities: any query or reasoning runs over S_t and *A* (which together act like a knowledge base [43,44]), any learning module can draw features from S_t rather than from unstructured data, and any decision/planning can be formulated as a goal or constraint in terms of the concepts.

To see this in action, consider a step in a *telerehab session*: at time *t*, the patient's HDT state S_t encodes their last known balance as "moderate", gait speed 0.8 m/s, and risk of fall "medium". Now the patient performs a new balance test wearing a motion capture suit (modality m_1) and answers a symptom questionnaire in the VR app (modality m_2). The data d_1 (motion signals) through g_{Balance,m_1} yields an updated Balance assessment = "poor". The data d_2 through $g_{\text{RiskOfFall},m_2}$ (perhaps the patient reports a recent near-fall) yields *RiskOfFall* = "high". The update function *F* takes these and might also, via an axiom/rule, set *GaitSpeed* to 0 (since during the balance test the patient wasn't walking). It thus produces S_{t+1}. Immediately, the HDT's semantic state has changed and this can trigger an alert for the clinician or a re-planning of the therapy (e.g. suggesting easier exercises given the high fall risk).

It is worth noting that the ontology's axioms *A* can encode temporal logic or procedural knowledge as well. For instance, an axiom could enforce that *after* a certain exercise, a specific assessment must be done. This would be part of *F* indirectly (because *F* should ensure *A* isn't violated). For the scope of this paper, we focus on *F* handling observational updates; planning actions will be discussed in the next section conceptually.

3.4 Alignment with the Real System Architecture

Having defined the multimodal ontology framework, we now map it to the architecture of the HDT-based telerehabilitation and education system described in [3]. In the *TD+AI platform* [3], the core components include:

- IoMT data acquisition (sensors and medical devices collecting patient data, e.g., movement and physiological signals);
- Cloud and edge services (which process data and host AI models, e.g., for computer vision and dialogue);
- HDT repository (where the patient's digital twin data is stored and updated continuously);
- AI avatars/agents (including virtual clinician agents, LLM-based dialogue systems, diagnostic models, and virtual patient avatars, which embody the HDT in simulations);
- User interfaces (a web portal for clinicians, a VR interface for patients/students, showing visualizations or avatars from the twin).

In our terms, the multimodal ontology $\mathcal{O}_{MM}$ is effectively the schema that the HDT repository follows. It dictates how raw IoMT data is translated into the HDT state and how that state is structured. The HDT state S_t corresponds to a record (or set of records) in the repository that can be queried by various agents. When Malakhov et al. [3] mention "aggregating heterogeneous patient data from records, sensors, imaging, and wearables into a continuously evolving representation", they are describing exactly what F does under ontology G and axioms A. The AI agents in the system (virtual clinicians, etc.) rely on this semantic layer: for example, a virtual physiotherapist agent might query S_t for the concept *MovementQuality* during an exercise, rather than crunching raw accelerometer readings itself – because S_t already has a value for *MovementQuality* derived from those readings via $g_{MovementQuality,m}$ in G. Likewise, the virtual patient avatar uses the ontology to drive its behavior: if the concept *PainLevel* in S_t becomes high, a rule in A might trigger the avatar to visibly wince or for an AI agent to verbally mention pain, thus informing the student in a training scenario. This aligns with the letter's description: "AI agents assume roles of doctors, nurses, patients… the novelty is uniting HDT-based virtual patients with AI-driven virtual clinicians in a single workflow". The ontology is the common language that unites them – the virtual patient (HDT) and virtual clinicians are interfacing through the shared concepts in S_t. Indeed, during a live virtual concilium, or "virtual multidisciplinary case conference,", multiple agents plus a human clinician review the same HDT state and contribute updates or suggestions via that state. One AI analyzes a video to update a mobility score concept; another AI (an NLP agent) parses patient speech to update a symptoms concept; the human clinician provides input adjusting a therapy goal concept. The ontology harmonizes these inputs and, through reasoning, might reconcile conflicts (e.g. if one agent's suggestion contradicts a known contraindication encoded in A, the system flags it).

By making the ontology explicit, we gain a clearer picture of information flow and decision making in the system architecture:

- Data Layer (sensors/IoMT) → Grounding (G) → HDT State (S_t);
- HDT State + Ontology ($S_t + A$) → Inference/Queries → AI Agent decisions;
- AI Agent actions (including those of human users inputting commands) → State Updates (F) → HDT State.

This cyclic operation is continuously informed by the ontology. Without it, each AI component would have to maintain its own internal understanding, leading to inconsistency and opaqueness. With the ontology, the system achieves what the anchor paper calls *convergence of digital health and education* – essentially a unified data/knowledge model that both clinical care algorithms and training simulations draw from [3].

The subsequent section applies this formal model to articulate how a multimodal ontology enables core AI functions in metaverse settings, connecting theoretical advantages to concrete scenario-level outcomes.

Interaction between the Semantic Backbone and large-scale foundation models can be organized as a constrained protocol rather than as unrestricted prompting. The Semantic Backbone retrieves a task-relevant ontology subgraph and patient-state slice; the foundation model generates a candidate interpretation, question, explanation, or plan; the candidate is normalized into typed ontology assertions; and symbolic validators apply OWL/SHACL-like constraints, clinical rules, provenance checks, and contraindication checks before the result can affect S_t or be shown as advice. Unsupported diagnoses, interventions without evidential grounding, or outputs violating axioms are rejected, downgraded to hypotheses, or routed for human review. In the reverse direction, foundation models may assist ontology completion by proposing candidate classes, synonyms, mappings, or competency questions, but these proposals remain provisional until validated against domain constraints and expert curation.

Formal elements (C, M, G, F) are employed in a conceptually lightweight manner to limit unnecessary mathematical complexity while emphasizing the operational capabilities that the ontology layer confers and that would otherwise be difficult or infeasible to achieve.

Figure 1 and **Figure 2** illustrate the end-to-end information flow and functional decomposition of the proposed metaverse architecture, organized around a multimodal ontology acting as a semantic backbone that interconnects data acquisition, reasoning, decision support, and human interaction.

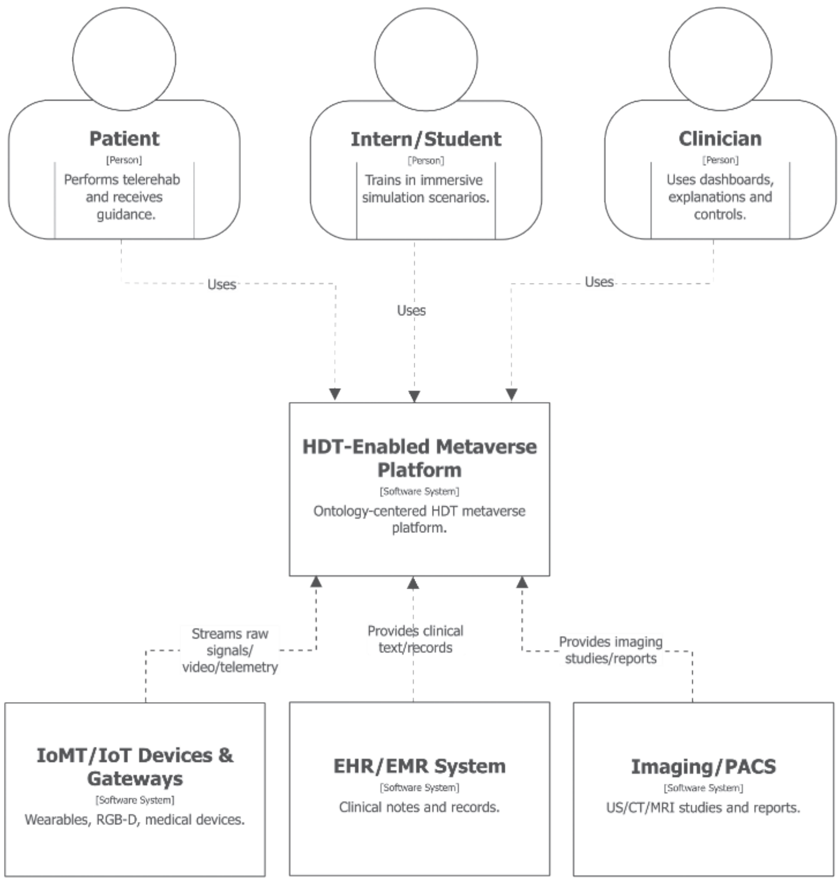

Figure 1. C4 System Context mini-diagram – HDT-enabled metaverse platform.

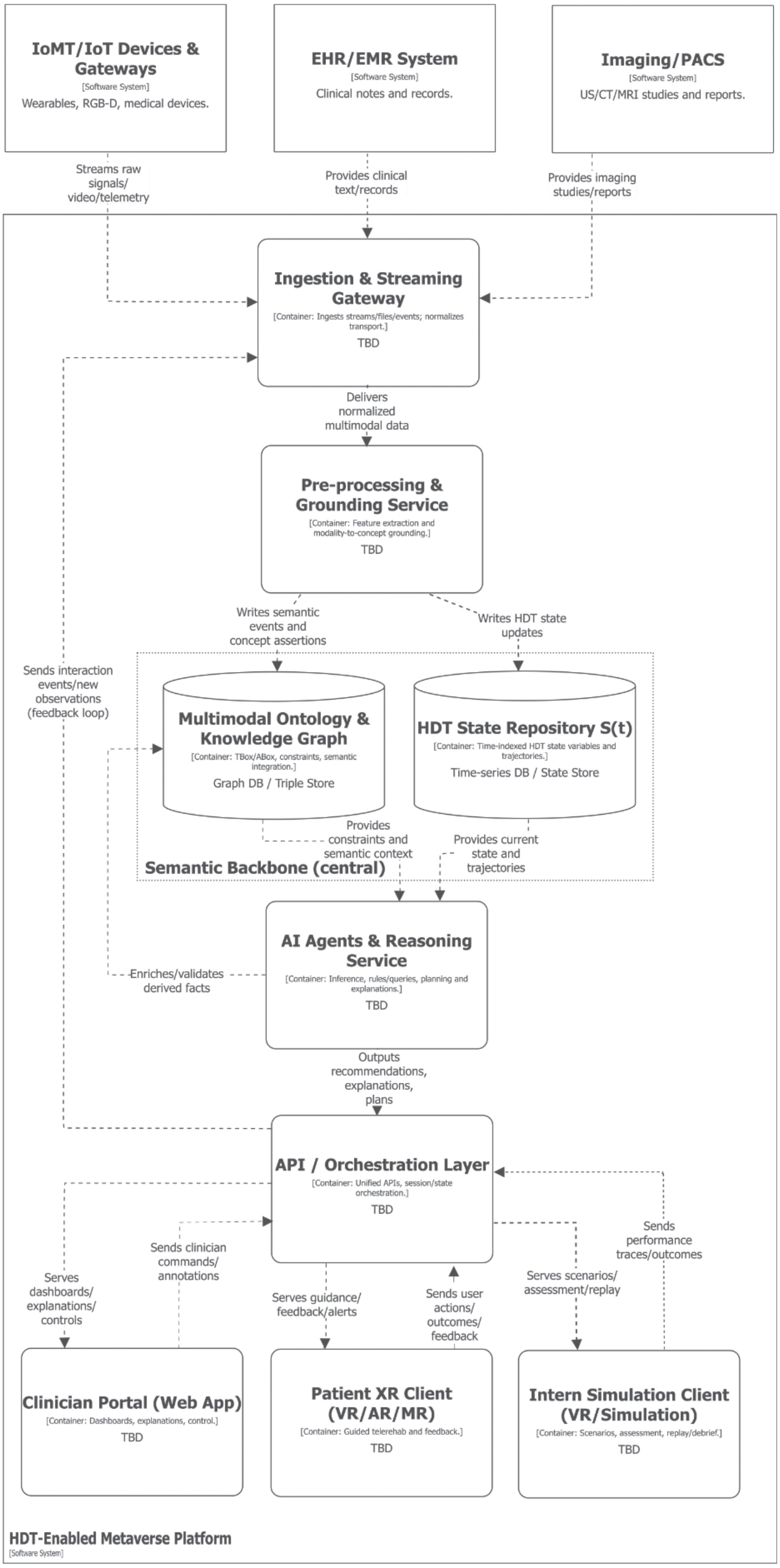

Figure 2. C4 Container View – HDT-enabled metaverse platform: major containers, data stores, and key information flows supporting telerehabilitation and immersive medical education.

4. Conceptual Results: Ontology–HDT–Metaverse Alignment

This section provides a conceptual analysis of how a multimodal ontology underpins HDT operation in metaverse-enabled telerehabilitation and clinical education scenarios. The analysis is organized around four core metaverse AI capabilities: environmental perception, semantic understanding, behavioral planning, and human–machine collaboration, with the ontological contribution to each capability specified explicitly. A unifying use-case narrative spanning a telerehabilitation session and an immersive medical education scenario is used to illustrate how the HDT semantic substrate, namely the ontology coupled with the evolving state, improves system outcomes.

4.1 Environmental Perception in Telerehabilitation

Scenario – Telerehab session. A post-stroke patient is performing balance and gait exercises at home, while connected via a VR and IoT/IoMT setup to a remote rehabilitation center [18]. The metaverse environment includes a virtual therapy room where the patient's avatar mirrors their motions captured by depth cameras and wearable sensors. An AI "exercise coach" agent monitors the scene.

In this context, environmental perception means capturing the patient's physical movements and physiological signals and identifying what is happening – effectively answering: What is the patient doing and how well? Without an ontology, an AI system might rely purely on numeric thresholds or black-box model outputs to judge performance. With the ontology:

a) The relevant ontology concepts C might include *CurrentExercise, RepetitionCount, RangeOfMotion, BalanceStability, HeartRate, FatigueLevel*, etc.;

b) Raw data from the environment (video feed, sensor streams) are mapped to these concepts via grounding functions G. For instance, a computer vision model extracts joint angles and updates the *RangeOfMotion(knee)* concept; an accelerometer on the torso feeds into a *BalanceStability* concept (perhaps defined as sway area or postural control score);

c) By structuring perception this way, the system develops a *contextual understanding* of sensor data. Instead of just "sensor X axis = 0.5", it knows "the patient's knee flexion is 45° which is within normal range for this exercise". In essence, the ontology translates low-level signals into *meaningful events and parameters* [17].

This ontology-grounded perception offers several benefits:

- *Improved accuracy*. The ontology can encode relationships that disambiguate sensor information. For example, if *CurrentExercise* is "sit-to-stand", the system expects knee angle changes; if instead it sees motion when *CurrentExercise* is "standing balance", that motion could be classified as an instability (affecting *BalanceStability* concept) rather than a deliberate movement. The ontology thus guides interpretation;
- *Noise filtering*. Axioms can help ignore impossible or irrelevant readings. If a sensor glitch produces an outlandish joint angle, a constraint in the ontology (e.g., maximum human knee flexion < 180°) prevents S_t from updating to an erroneous value;

During sensor noise or data dropouts, axioms A act as state-transition guards. They can encode anatomical range limits, kinematic continuity constraints, exercise-specific expectations, physiological plausibility bounds, and monotonicity or recovery-rate constraints where clinically justified. Consequently, a transient perception failure cannot create a non-physical HDT transition such as an impossible joint angle, instantaneous gait-speed jump, or sudden normalization of fall risk. The safer default is to mark the affected concept as stale, uncertain, or unavailable, trigger a request for repeated measurement, and shift the interaction to a conservative mode.

- *Comprehensiveness*. Multiple modalities are fused – vision and IMU both provide gait speed estimates; *F* could combine them for a more robust *GaitSpeed* concept value. The ontology ensures they are compared on the same semantic scale.

In the scenario, as the patient performs exercises, the HDT's state fills with high-level perceptions: e.g. *"Exercise = TandemWalk, RepsDone = 2, GaitSpeed = 0.8 m/s, BalanceStability = moderate"*. These become part of the HDT's continuously updated virtual profile [3], accessible to other components.

Figure 3 depicts a use-case where a clinician interacts with an AI avatar and a manikin representing a patient [22,45]. During such an interaction, the AI's ability to perceive the environment (including the manikin's sensor outputs and the clinician's actions) depends on recognizing medically salient events. A semantic ontology helps the AI distinguish, for instance, the clinician's gesture as a "lung auscultation attempt" versus random movement, if concepts for procedures are present.

The ontology empowers environmental perception by providing *contextual labels and relationships for raw data*, turning a chaotic stream of sensor inputs into an organized set of observations about the patient's status in the virtual environment. This is the foundation upon which deeper understanding builds.

Figure 3. Clinician–AI avatar interaction with an instrumented manikin as an immersive clinical education scenario in an HDT-enabled metaverse workflow.

4.2 Semantic Understanding and Situation Assessment

Continuing the telerehab scenario. As the system perceives various parameters, it now must synthesize them into a broader understanding: What is the patient's current status and needs? This is where semantic understanding comes in – drawing inferences, identifying patterns or anomalies, and relating them to clinical knowledge.

Because the HDT's state is anchored in ontology concepts, many such inferences become straightforward logical or rule-based operations. For example:

- The ontology might encode a rule: If *BalanceStability* is poor and the patient's diagnosis is cerebellar ataxia, then *RiskOfFall* is high. When perception updates *BalanceStability = poor*, the rule triggers and sets *RiskOfFall = high*. Without an ontology, an LLM might eventually learn this association, but it would be a black box. Here it's explicit and immediate;
- Semantic queries can be answered: e.g., "Is the patient improving over last week?" can be broken down into concept comparisons (today's *GaitSpeed* vs last week's *GaitSpeed*; presence or absence of certain symptoms, etc.). Since S_t is time-stamped and historical states $S_{t-\Delta}$ can be stored, the ontology provides the structure to calculate improvement (maybe an *ImprovementTrend* concept);
- Uncertainty or ambiguity can be represented. If data is inconclusive, the ontology could have a concept like *BalanceConfidenceLevel*, or it could use probabilistic values in S_t. Semantic understanding includes recognizing when the system doesn't know something (e.g. a sensor offline leading to no recent data for a concept), which can be flagged for human attention.

In our scenario, suppose after some exercises, the patient's HDT state shows: *HeartRate* elevated, *FatigueLevel* high (from a questionnaire), *ExercisePerformance* deteriorating. The ontology might contain a conceptual model of fatigue – e.g. a relation that high fatigue implies performance drop and risk of injury. An axiom could be: *"IF FatigueLevel = high THEN ExerciseIntensity = reduce"* (where *ExerciseIntensity* is a concept representing the prescribed difficulty). The HDT system, through reasoning, thus "understands" that the patient is tired and semantically generates a suggestion to decrease intensity. This is semantic understanding: going from data to an insight with meaning in the domain of discourse (rehab therapy in this case) [3,18]. In practice, the AI coach might say, "Let's take a short break, I see you're getting tired," which is exactly the kind of supportive, context-aware behavior we want.

Switching to an *education scenario*, semantic understanding is equally critical. Imagine a medical student in an immersive virtual clinic, diagnosing a virtual patient (the HDT). The system must interpret the student's actions and the patient's condition. If the student orders a lab test in the simulation, the ontology is updated (e.g. *LabOrder* concept added). The virtual patient's HDT, which contains a disease model, semantically evaluates the findings: e.g., it knows that "chest pain + high troponin = myocardial infarction likely" from a rule or from the structured disease graph it contains. Therefore, it can push the case in that direction, or require the student to recognize it. In essence, the ontology allows the system to maintain a case semantics – linking symptoms, signs, and diagnoses in a way that an expert human would. This leads to an experience far richer than rote Q&A. The student can ask "Why is the patient having chest pain?" and the system (via ontology) can respond, "Because reduced blood flow (concept) in coronary arteries is causing ischemia (concept) – which is part of myocardial infarction pathology." Indeed, large language models integrated with a medical ontology could produce such explanations, making the learning experience robust and safe.

References [3,18] describe the use of AI in a "virtual hospital" configuration where AI agents simulate doctors and patients. Semantic understanding ensures these agents behave plausibly. The virtual patient (HDT) uses its ontology to ensure consistency: if the student administers a certain drug, the ontology might have a drug–effect relation that updates the patient's vital signs accordingly (e.g. *BloodPressure* concept drops if a vasodilator is given, unless contraindicated by another condition in the ontology). This prevents unrealistic or incorrect simulator responses, reinforcing correct learning.

In summary, the multimodal ontology gives the HDT system a form of situational awareness. It can cross-link disparate data points, apply medical knowledge rules, and maintain an interpretable record of the patient's state and trajectory. This is the essence of semantic understanding in an AI-driven metaverse: not just seeing, but making sense of what is happening, in the same conceptual terms a human expert would use.

4.3 Behavioral Planning and Decision Support

Next, we consider *behavioral planning* – the ability of the system (often in collaboration with humans) to plan the next steps, whether it's a therapeutic intervention in telerehab or the progression of a clinical case in an educational simulation. Planning in our context involves both *AI planning* (the system's autonomous choices) and *supporting human decision-making* (recommendations, what-if simulations, etc.). The ontology-augmented HDT greatly facilitates both.

In the *telerehab scenario*, suppose the system has semantically understood that the patient's balance is worsening and fatigue is high. The next behavior could be: adjust the rehab regimen. Planning could occur at multiple levels:

a) *Reactive Planning*. The AI coach agent immediately suggests a simpler exercise or rest (short-term plan). The agent consults the ontology, finds this guideline, and outputs the suggestion. Alternatively, an optimization algorithm [46] could search for the exercise that maximizes expected benefit while keeping fatigue concept within safe range;

b) *Long-term Planning*. At the end of the session, the system might update the patient's rehabilitation plan (stored in the HDT as concepts like *WeeklyExerciseTargets* or *NextVisitSchedule*). The ontology, containing the patient's progress data and possibly comparable data from other patients (if connected to a larger knowledge base/graph of outcomes), can help decide: should the patient advance to more difficult exercises next week or focus on basics? Because the HDT's ontology knows the relationships (e.g. improving *GaitSpeed* might require addressing *BalanceStability* first), it can plan the sequence of goals. This is akin to classical AI planning but with the heavy lifting done by the structured knowledge of "what leads to what" encoded in R and A;

c) *Personalization*. The HDT is individualized; its ontology might be populated with patient-specific thresholds (perhaps learned over time: e.g. this patient's *SafeHeartRateMax* concept). So, plans are tailored. If another patient's HDT with the same ontology but different data wouldn't get tired at that intensity, the AI wouldn't recommend a break for them – it's not one-size-fits-all.

From the system perspective described in [3], this corresponds to the AI assistants "proposing optimal adjustments to the rehabilitation program" in real time. During the "virtual concilium" described, AI agents might literally debate or vote on plan options, using the ontology as the common ground. One agent says "increase balance exercise difficulty" citing improved gait, another says "no, fatigue high – better to rest." The human clinician mediates. All these interactions reference the HDT state (concepts like *FatigueLevel*, *GaitSpeed*, etc.), which ensures transparency and traceability: the clinician can ask why a suggestion was made, and the agent can point to the ontology rule or data that triggered it (e.g. *FatigueLevel = high*, which per guidelines implies rest), providing explainable AI in the decision loop.

In the *education scenario*, behavioral planning plays out as the system controlling the case progression or the patient avatar's behavior:

- The virtual patient (HDT) might have a built-in *illness script* [3] encoded as a state machine or causal graph in the ontology. For example, if untreated, their *ConditionSeverity* concept will progress over (virtual) time. If the student administers treatment, an ontological rule adjusts the trajectory (perhaps improving some vital-sign concept). Planning here is basically simulation: the ontology guides how the patient's state evolves logically from interventions. For instance, the HDT might simulate an asthma exacerbation resolving after medication, following rules akin to "IF bronchodilator given THEN over 5 minutes, oxygen saturation improves". This kind of knowledge can be encoded in an ontology with temporal logic or handled by linking the ontology to a physiological simulation model;
- The AI-driven tutor agent might also plan tutorial interventions. If the student is stuck (detected via the ontology: e.g. they haven't addressed a critical finding concept for too long), the system plans a hint or a nudge. The decision to give a hint can be based on an educational ontology concept like *LearnerPerformance*;
- Metrics from the ontology can be used to generate feedback at the scenario's end: e.g. "You identified 8/10 key findings (concept comparison), and your treatment choice reduced the virtual patient's risk score from high to moderate (ontology inference). To improve, consider the guideline that chest pain with elevated troponin implies immediate anticoagulation (knowledge from ontology)." This feedback is only possible because the ontology tracks which actions corresponded to which concept changes and how they relate to standard knowledge.

The ability to run counterfactual or what-if simulations is a powerful aspect of planning with an ontology-backed HDT. Because the ontology captures causal and correlational knowledge, the system can hypothetically modify a concept and propagate changes to predict outcomes.

The ontology enhances behavioral planning by providing: explicit rules/guidelines for automated decision-making; causal modeling for simulation of interventions; goal structures for AI to navigate; explainability and traceability of decisions.

This ensures that AI-driven decisions in telerehab and training are not just reactive or hard-to-trust outputs of neural networks, but are grounded in clinical reasoning paradigms and best practices represented in the ontology [3,14].

4.4 Human–Machine Collaboration and Explainability

The final, perhaps most important, dimension is *human–machine collaboration*. Both telerehab and education in the metaverse involve close interplay between human users (patients, clinicians, students, instructors) and AI or virtual agents. A core thesis of our viewpoint is that the multimodal ontology significantly improves this collaboration by acting as a shared semantic substrate – essentially a common language and reference point that both humans and AI agents can understand.

In *the telerehab setting*, consider the interaction between the remote clinician and the AI assistants during a live session (the "virtual concilium" described in [3]). The clinician might ask the system, "How is the patient's balance today compared to last session?" Because of the ontology, the AI can respond in human-friendly terms: "The patient's *BalanceStability* is moderate today, down from good last week, and their sway area increased by 20%. This suggests a slight regression, possibly due to fatigue." The use of the concept *BalanceStability* and its qualitative values ("moderate", "good") maps to clinical language, making the AI's analysis transparent. If the system were a black box spitting out "score = 0.7 vs 0.9," the clinician would glean less insight.

Moreover, when the AI agents propose adjustments (like changing an exercise), they can explain the rationale by referencing ontology relations or rules: "We should switch to a seated exercise now because the patient's *RiskOfFall* is high and fatigue is elevated – per rehab protocol, high fall risk indicates using safety measures" [17]. This kind of explanation is only possible because the logic was encoded declaratively (in the ontology or its rule base), so it can be surfaced to the user. Studies in AI safety emphasize the need for *explainable AI (XAI)* in healthcare [15]; our approach inherently supports XAI by using an ontology that is interpretable.

For patients, the ontology can facilitate communication as well. The HDT could power a patient-facing dashboard that shows, in understandable terms, their progress: e.g., "Digital Twin Summary: your walking speed improved to 0.9 m/s (normal for age is ~1.0 m/s), balance is still below average – we'll keep focusing on that. Fatigue was high today; ensure you rest and hydrate." The patient is essentially being presented with the state of their twin in a comprehensible way. This fosters trust and engagement, as the patient can see that the system "knows" them and is tracking meaningful aspects of their health.

In the *educational context*, human–machine collaboration takes the form of the student working through a case with the help of AI facilitators (or an autonomous patient avatar). The ontology allows the virtual patient to be *introspective* and answer the student's questions consistently. For example, a student might ask the virtual patient (via chat), "Where exactly is your pain?" The HDT's ontology has concepts for pain location, quality, intensity. If location is, say, *substernal*, the avatar answers "It's right here in the middle of my chest," possibly even highlighting that area on the 3D avatar. If later the student asks an AI tutor, "Why does the patient have chest pain?", the tutor can reference the ontology of the condition: "Because their heart isn't getting enough oxygen (the HDT has concept *Ischemia = true*), which commonly causes pain in that area." This makes the learning process more like working with a real patient and a supervising physician combined – a rich, explanatory dialogue, rather than a sterile question bank.

Collaboration is also about shared decision-making. In a virtual training scenario, the AI might serve as a co-pilot. Imagine a student formulating a treatment plan – the AI tutor can step in to suggest alternatives or flag issues: "You chose Drug X. According to the patient's HDT, they have an allergy to Drug X (see *Allergy* concept). Are you sure you want to proceed?" Here the ontology provides the factual basis (the allergy is a stored concept), and the AI uses it to prevent an error, similar to a clinical decision support system in hospitals. The difference is the context: this is a safe learning environment, but the principle is the same – an ontology-driven HDT can prevent oversight by the human, and do so by clearly pointing to the relevant knowledge.

For explainable human–machine collaboration, the HDT semantic state can be translated into rationales through explicit mapping rules: observation – grounded concept – confidence/provenance – rule or axiom – inferred state – recommendation. A clinician-facing explanation may therefore state that an exercise was reduced because IMU-derived *BalanceStability* declined, the text questionnaire indicated high fatigue, and the fall-risk rule fired with specified confidence. The same reasoning trace can be rendered in layered form: a plain-language summary for patients and students, an evidence table for clinicians, and a technical provenance graph for auditors. *Contestability* is supported by allowing users to inspect the evidence source, override or annotate an assertion, request repeated measurement, or mark a rationale as clinically inappropriate.

5. Discussion

The proposed framing positions multimodal ontology as a semantic substrate that can materially strengthen HDT architectures deployed in AI-driven metaverse settings. The primary implication is architectural: an explicit knowledge layer that couples an ontology with an evolving HDT state enables modular integration of heterogeneous AI components while preserving stable semantics across perception, inference, and interaction.

5.1 Architectural Implications

Ontology-centered designs support *interoperability* and *component replaceability* by decoupling the knowledge layer from task-specific models. Perception or prediction modules can be upgraded, substituted, or specialized, provided their outputs are grounded into the same ontological concepts (e.g., balance stability, exercise type, fall risk). A shared semantic state also reduces redundant “world models” across agents, enabling complementary mechanisms such as rule-based safety checks, knowledge-graph inference, and foundation-model dialogue to operate over a common representational substrate. In healthcare and clinical education, this single-source-of-truth property is especially salient because it mitigates divergent interpretations and supports auditable provenance of recommendations.

Modularity and reuse can be operationalized through a layered ontology stack: a core HDT module for person, state, observation, intervention, time, provenance, and uncertainty; modality modules for IMU, video, text, imaging, laboratory, and interaction traces; clinical-domain modules for stroke rehabilitation, oncology rehabilitation, cardiopulmonary rehabilitation, or mental health screening; and task modules for assessment, recommendation, simulation, and education. Transitioning from stroke rehabilitation to oncology care would therefore retain the core and modality layers while replacing or extending disease, symptom, protocol, contraindication, and outcome modules. Crosswalks to standards such as FHIR, ICD, SNOMED CT, LOINC, and domain rehabilitation scales can be versioned and tested through competency questions, allowing reuse without assuming that all clinical domains share the same concept hierarchy.

5.2 Technical Constraints and Mitigation Pathways

Three constraints dominate practical deployment. First, *ontology engineering cost* remains substantial, particularly for rehabilitation and training domains that require fine-grained representation of procedures, functional assessments, and context. Practical mitigation includes leveraging existing clinical vocabularies and standards where licensure permits, adopting modular ontology design, and using semi-automated concept elicitation to prioritize high-impact subgraphs for early implementation. Second, HDT state is inherently *dynamic and uncertain*, motivating probabilistic annotations, confidence-aware grounding, and incremental reasoning strategies to maintain real-time responsiveness under high-frequency sensor updates. Third, *validation* must target both semantic correctness (ontology quality and grounding fidelity) and clinical utility (safety, effectiveness, and workflow compatibility). Human-in-the-loop governance remains necessary but insufficient; systematic evaluation against retrospective cohorts, simulation-based stress testing, and prospective studies are required to establish reliability.

5.3 Interoperability and Sociotechnical Considerations

Integration with external health and education infrastructures requires alignment with interoperability standards and robust mapping between operational data models and ontological entities. Beyond engineering, the approach carries nontrivial *ethical and human-factor* implications. Ontology-grounded explanations can improve transparency and contestability, but HDTs consolidate sensitive information and may enable secondary inference of protected attributes. Access control, encryption, and role-based semantic filtering are therefore design requirements rather than optional features. Usability constraints are equally important: ontology-rich systems can amplify information overload unless user interfaces expose layered summaries and permit controlled drill-down aligned with clinical priorities.

5.4 Relationship to Foundation Models

The semantic layer is best interpreted as complementary to large-scale foundation models. Foundation models provide high-capacity perception and natural-language interaction, whereas ontological constraints provide verifiability, consistency checking, and domain-aligned semantics. In safety-critical workflows, ontology-mediated validation of model outputs can function as a guardrail against plausible but incorrect recommendations, supporting a neuro-symbolic architecture in which generative modules propose and symbolic components constrain, justify, and audit.

The decision loop can be closed by treating foundation-model outputs as candidate assertions or candidate actions rather than authoritative decisions. Each generated output is checked for ontology coverage, evidence support, temporal compatibility with S_t, absence of contraindications, and consistency with user role and clinical scenario. If the output passes validation, it is committed with provenance and confidence; if it fails, it is either blocked, reformulated as a hypothesis, or escalated to a clinician. This neuro-symbolic loop also supports hallucination filtering because unsupported entities, invented measurements, and recommendations lacking grounding cannot be promoted to the HDT state.

5.5 Scope and limitations

The argument is primarily conceptual and requires broader empirical substantiation. Practical results reported in the anchor use-case reference [3], arising from collaborative work between I.Ya. Horbachevsky Ternopil National Medical University and the V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine, support feasibility but do not substitute for controlled evaluation across diverse populations, modalities, and deployment environments. Nonetheless, the central claim remains that multimodal ontology offers a principled route to scalable interoperability, accountable reasoning, and human-centered collaboration in HDT-enabled metaverse systems.

In a broader framing, the proposed direction is consistent with Industry 5.0 and Society 5.0 trajectories that prioritize human-centricity, transparency, and resilience in socio-technical systems, particularly where AI is embedded into high-stakes decision processes. As metaverse infrastructures mature through advances in AR/VR, large-scale modeling, and digital twinning, the early incorporation of semantic grounding[47,48] can reduce dependence on opaque end-to-end pipelines and improve accountability, interoperability, and safety. The additional engineering effort required to design and maintain a multimodal ontology is therefore plausibly offset by gains in system reliability, explainability, and stakeholder trust, especially in healthcare and clinical training contexts.

6. Conclusion

This viewpoint has argued that a *multimodal ontology* constitutes the semantic substrate required for HDTs to operate reliably in AI-driven metaverse applications, with emphasis on telerehabilitation and immersive clinical education. A minimal formalization of $\mathcal{O}_{MM}$ was introduced to capture concepts, relations, axioms, modalities, grounding mappings, and state-update dynamics, thereby specifying how heterogeneous evidence can be translated into an evolving, semantically interpretable HDT state. The revised formulation makes explicit that this substrate includes confidence-aware grounding, temporal/probabilistic state evolution, symbolic validation of generative AI, and interoperability with edge-networked HDT infrastructures.

Three conclusions follow. First, multimodal ontology provides a unified representational layer that connects sensor streams, clinical information, and user interaction traces to clinically meaningful constructs, enabling auditable state persistence and semantic interoperability. Second, this layer strengthens the principal metaverse AI functions: environmental perception through contextualized multimodal fusion, semantic understanding through ontology-constrained inference, behavioral planning through goal and constraint formulation over concepts, and human–machine collaboration through explanation and coordination grounded in shared terminology. Third, ontology-backed HDTs define a concrete research agenda centered on ontology development and governance, grounding fidelity, real-time incremental reasoning, and evaluation using semantic consistency, reasoning correctness, user trust, and patient and intern outcomes.

The approach aligns with *Industry 5.0* and *Society 5.0* paradigms that prioritize human-centric, transparent, and resilient AI integration. Embedding semantic grounding early in metaverse-oriented HDT architectures offers a principled route to improve reliability, accountability, and adoption in safety-critical clinical and educational workflows.

Author contributions: Conceptualization, OP and KM; methodology, KM; formal analysis, KM, OB and SM; investigation, KM; writing—original draft preparation, KM, OB, SM and DV; writing—review and editing, OP, OB, DV and SM; supervision, OP and DV; project administration, OP and DV. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Declaration on Generative AI: During the preparation of this manuscript, the authors used gpt-oss-120b (OpenAI's open-weights LLM), deployed locally on institutional hardware, solely for grammar and spelling checks.

Acknowledgments: The authors gratefully acknowledge *Mykola Budnyk*, DSc (Computer Sciences), Full Professor and Chief Researcher, Department of Sensor Devices, Systems and Technologies of Contactless Diagnostics, V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine, and Associate Professor, Department of Computer Science, Sumy State University (Ukraine), Head of the Institute of Cybernetics project team, for strategic leadership, methodological guidance, and sustained support in the development of the TD+AI platform and the associated HDT research program. The authors also express appreciation to colleagues at the V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine, including *Petro Stetsyuk*, Corresponding Member of the National Academy of Sciences of Ukraine, Doctor of Physical and Mathematical Sciences, Head of the Department of Nonsmooth Optimization Methods No. 120; *Tetiana Semykopna*, PhD (Medical Sciences), Senior Researcher, Microprocessor Technology Lab; *Vladislav Kaverinskiy*, PhD, Senior Researcher, Microprocessor Technology Laboratory; and *Illya Chaikovsky*, PhD in Medicine (University of Duisburg–Essen, Germany, magna cum laude) and PhD in Medical Informatics and Cybernetics (International Research and Training Center for Information Technologies and Systems of the National Academy of Sciences of Ukraine, Kyiv), Lead Researcher, Department of Sensor Devices, Systems and Technologies of Contactless Diagnostics. Their contributions in nonsmooth optimization, medical informatics, and sensor-based diagnostics substantially strengthened the scientific and technological foundations of the TD+AI and HDT initiative. The authors further extend sincere gratitude to the team at I.Ya. Horbachevsky Ternopil National Medical University for participation in the implementation and testing of HDT-based scenarios, including standardized virtual patient use cases.

Conflict of interest: The authors declare no conflict of interest.

References

1. Malakhov KS. Innovative Hybrid Cloud Solutions for Physical Medicine and Telerehabilitation Research. International Journal of Telerehabilitation. 2024, 16(1): 1–19. https://doi.org/10.5195/ijt.2024.6635
2. Zhou H, Chen X, Li J, et al. Generative Artificial Intelligence in the Metaverse Era: A Review on Models and Applications. Research. 2025, 8: 0804. https://doi.org/10.34133/research.0804
3. Malakhov KS, Vakulenko DV. Letter to the Editor – Human Digital Twin in Ukraine: Converging Digital Health and Digital Education for Next-Generation Telerehabilitation. International Journal of Telerehabilitation. 2025, 17: 1–10. https://doi.org/10.63144/ijt.2025.6729
4. Mokhtari M. Human Digital Twins in Personalized Healthcare: An Overview and Future Perspectives. 2025.
5. Palagin OV, Kaverinskiy VV, Petrenko MG, et al. Digital Health Systems: Ontology-Based Universal Dialog Service for Hybrid E-Rehabilitation Activities Support. In: Proceedings of the 2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS). 2023, 1: 84–89. https://doi.org/10.1109/IDAACS58523.2023.10348639
6. Litvin A, Palagin OV, Kaverinskiy VV, et al. Ontology-Driven Development of Dialogue Systems. South African Computer Journal. 2023, 35: 37–62. https://doi.org/10.18489/sacj.v35i1.1233
7. Palagin OV, Malakhov KS, Velychko VYu, et al. Hospital Information Smart-System for Hybrid E-Rehabilitation. In: Proceedings of the CEUR Workshop Proceedings. 2022, 3501: 140–157.
8. Abu-Salih B. MetaOntology: Toward Developing an Ontology for the Metaverse. Frontiers in Big Data. 2022, 5: 998648. https://doi.org/10.3389/fdata.2022.998648
9. Gruber TR. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition. 1993, 5: 199–220. https://doi.org/10.1006/knac.1993.1008
10. Fung P, Bachrach Y, Celikyilmaz A, et al. Embodied AI Agents: Modeling the World. 2025.
11. Karabulut E, Pileggi SF, Groth P, et al. Ontologies in Digital Twins: A Systematic Literature Review. Future Generation Computer Systems. 2024, 153: 442–456. https://doi.org/10.1016/j.future.2023.12.013
12. Carbonaro A, Marfoglia A, Nardini F, et al. CONNECTED: Leveraging Digital Twins and Personal Knowledge Graphs in Healthcare Digitalization. Frontiers in Digital Health. 2023, 5: 1322428. https://doi.org/10.3389/fdgth.2023.1322428
13. Juarez MGJ, Giret A, Botti V. Semantic and Modular Orchestration of AI-Driven Digital Twins for Industrial Interoperability

and Optimization. Journal of Industrial Information Integration. 2025, 48: 100959. https://doi.org/10.1016/j.jii.2025.100959
14. Lopez V, Hoang L, Martinez-Galindo M, et al. Enhancing Foundation Models for Scientific Discovery via Multimodal Knowledge Graph Representations. Journal of Web Semantics. 2025, 84: 100845. https://doi.org/10.1016/j.websem.2024.100845
15. Keser M, Schwalbe G, Amini-Naieni N, et al. Unveiling Ontological Commitment in Multi-Modal Foundation Models. 2024.
16. Velychko VYu, Petrenko MG, Semykopna TV, et al. Transdisciplinary Intelligent Information and Analytical System for the Rehabilitation Processes Support in a Pandemic. 2021. https://doi.org/10.31274/isudp.2022.121
17. Ghaempanah F, Moasses Ghafari B, Hesami D, et al. Metaverse and Its Impact on Medical Education and Health Care System: A Narrative Review. Health Science Reports. 2024, 7: e70100. https://doi.org/10.1002/hsr2.70100
18. Petrenko MG, Malakhov KS. Methodology and Practice of Interdisciplinary Knowledge Convergence in Digital Health. Information Technology and Systems. 2025, 3: 76–96. https://doi.org/10.15407/intechsys.2025.03.076
19. Velychko VYu, Voinova S, Granyak V, et al. New Information Technologies, Simulation and Automation. 2022. https://doi.org/10.31274/isudp.2022.121
20. Pammi M, Shah PS, Yang LK, et al. Digital Twins, Synthetic Patient Data, and in-Silico Trials: Can They Empower Paediatric Clinical Trials? The Lancet Digital Health. 2025, 7: 100851. https://doi.org/10.1016/j.landig.2025.01.007
21. Palagin OV, Stetsyuk PI, Sergienko I, et al. Cloud-based Platform for Patient-centered Telerehabilitation of Oncology Patients with Mathematical Related Modeling. 2024. https://doi.org/10.34229/mono.palagin.24.1
22. Malakhov KS. Letter to the Editor – Update from Ukraine: Development of the Cloud-Based Platform for Patient-Centered Telerehabilitation of Oncology Patients with Mathematical-Related Modeling. International Journal of Telerehabilitation. 2023, 15: 1–3. https://doi.org/10.5195/ijt.2023.6562
23. Kurgaev O, Malakhov KS. Next-Generation Biofeedback Rehabilitation Systems: An AI-Enabled Architecture Integrating Brain–Computer Interfaces and Knowledge-Based Control. In: Duncan LT, editor. Advances in Health and Disease. Volume 90. Nova Science Publishers, Inc. 2026, 90: 1–34.
24. Katsoulakis E, Wang Q, Wu H, et al. Digital Twins for Health: A Scoping Review. npj Digital Medicine. 2024, 7: 77. https://doi.org/10.1038/s41746-024-01073-0
25. Tschannen M, Gritsenko A, Wang X, et al. SigLIP 2: Multilingual Vision-Language Encoders with Improved Semantic Understanding, Localization, and Dense Features. 2025.
26. Gao Y, Piccinini M, Zhang Y, et al. Foundation Models in Autonomous Driving: A Survey on Scenario Generation and Scenario Analysis. 2025. https://doi.org/10.1109/OJITS.2026.3660686
27. Hou J, Wang H, Li T, et al. An Explainable End-to-End Autonomous Driving Framework Based on Large Language Model and Vision Modality Fusion: Design and Application of DriveLLM-V. Transportation Research Part C: Emerging Technologies. 2025, 181: 105368. https://doi.org/10.1016/j.trc.2025.105368
28. Lymperaiou M, Stamou G. A Survey on Knowledge-Enhanced Multimodal Learning. Artificial Intelligence Review. 2024, 57: 284. https://doi.org/10.1007/s10462-024-10825-z
29. Micelli L, Montagna S. Human Digital Twin for Healthcare Applications: A White Label Digital Twin Implementation. In: Proceedings of the 2025 International Conference on Information Technology for Social Good. 2025, 90–98. https://doi.org/10.1145/3748699.3749778
30. Yi C, Cai J, Su Z. A Multi-User Mobile Computation Offloading and Transmission Scheduling Mechanism for Delay-Sensitive Applications. IEEE Transactions on Mobile Computing. 2020, 19: 29–43. https://doi.org/10.1109/TMC.2019.2891736
31. Cao H, Cai J. Distributed Multiuser Computation Offloading for Cloudlet-Based Mobile Cloud Computing: A Game-Theoretic Machine Learning Approach. IEEE Transactions on Vehicular Technology. 2018, 67: 752–764. https://doi.org/10.1109/TVT.2017.2740724
32. Cai J, Shen X, Mark JW, et al. Semi-Distributed User Relaying Algorithm for Amplify-and-Forward Wireless Relay Networks. IEEE Transactions on Wireless Communications. 2008, 7: 1348–1357. https://doi.org/10.1109/TWC.2008.060909
33. Yang Y, Shi Y, Yi C, et al. Dynamic Human Digital Twin Deployment at the Edge for Task Execution: A Two-Timescale Accuracy-Aware Online Optimization. IEEE Transactions on Mobile Computing. 2024, 23: 12262–12279. https://doi.org/10.1109/TMC.2024.3406607
34. Okegbile SD, Cai J, Zheng H, et al. Differentially Private Federated Multi-Task Learning Framework for Enhancing Human-

to-Virtual Connectivity in Human Digital Twin. IEEE Journal on Selected Areas in Communications. 2023, 41: 3533–3547. https://doi.org/10.1109/JSAC.2023.3310106

35. Chen R, Yi C, Zhou F, et al. Federated Digital Twin Construction via Distributed Sensing: A Game-Theoretic Online Optimization With Overlapping Coalitions. IEEE Transactions on Mobile Computing. 2025, 24: 12221–12238. https://doi.org/10.1109/TMC.2025.3582755
36. Chen J, Yi C, Okegbile SD, et al. Networking Architecture and Key Supporting Technologies for Human Digital Twin in Personalized Healthcare: A Comprehensive Survey. IEEE Communications Surveys & Tutorials. 2024, 26: 706–746. https://doi.org/10.1109/COMST.2023.3308717
37. Zhu F, Chen J, Wen J, et al. From Data Mirror to Smart Copilot: A Survey on NextG Semantic Communication for Propelling Digital Twin World Into Cognitive Stage. IEEE Communications Surveys & Tutorials. 2026, 28: 4915–4947. https://doi.org/10.1109/COMST.2026.3665395
38. Chen J, Yi C, Gong S, et al. Generative AI-Aided QoE-Aware Resource Allocations for RIS-Assisted Digital Twin Interaction With Uncertain Evolution. IEEE Transactions on Mobile Computing. 2026, 25: 7888–7905. https://doi.org/10.1109/TMC.2025.3645306
39. Wu K, Chen J, Chen L, et al. QoE-Aware Joint Visual and Haptic Signal Transmission With Adaptive Data Compression for Immersive Interactions in Human Digital Twin. IEEE Transactions on Network and Service Management. 2025, 22: 2780–2794. https://doi.org/10.1109/TNSM.2025.3558358
40. Chen J, Yi C, Du H, et al. A Revolution of Personalized Healthcare: Enabling Human Digital Twin With Mobile AIGC. IEEE Network. 2024, 38: 234–242. https://doi.org/10.1109/MNET.2024.3366560
41. Chen J, Shi Y, Yi C, et al. Generative-AI-Driven Human Digital Twin in IoT Healthcare: A Comprehensive Survey. IEEE Internet of Things Journal. 2024, 11: 34749–34773. https://doi.org/10.1109/JIOT.2024.3421918
42. Chen J, Shi Y. Generative AI over Mobile Networks for Human Digital Twin in Human-Centric Applications: A Comprehensive Survey. 2024. https://doi.org/10.36227/techrxiv.172349525.50239637/v1
43. Litvin AA, Velychko VYu, Kaverinskiy VV. A New Approach to Automatic Ontology Generation from the Natural Language Texts with Complex Inflection Structures in the Dialogue Systems Development. In: Proceedings of the CEUR Workshop Proceedings. 2022, 3501: 172–185.
44. Palagin OV, Petrenko MG, Kaverinskiy VV, et al. A Method for Enhancing the Efficiency of RDF/XML-Structure Processing in the Apache Jena Semantic Web Framework. Cybernetics and Systems Analysis. 2025, 61: 469–486. https://doi.org/10.1007/s10559-025-00784-w
45. Integration of AI Agents and HDT into Digital Medicine. Available online: https://youtu.be/Ze8jjnlB5Fs (accessed on 1 June 2023)
46. Vakulenko DV, Palagin OV, Sergienko IV, et al. Algorithmization and Optimization Models of Patient-Centric Rehabilitation Programs. Cybernetics and Systems Analysis. 2024, 60: 736–752. https://doi.org/10.1007/s10559-024-00711-5
47. Palagin OV, Velychko VYu, Malakhov KS, et al. Distributional Semantic Modeling: a Revised Technique to Train Term/Word Vector Space Models Applying the Ontology-related Approach. In: Proceedings of the CEUR Workshop Proceedings. 2020, 2866: 342–353
48. Petrenko MG, Malakhov KS. Systems Science: Digitalization of Transdisciplinary Research. Journal of Robotics, Networking and Artificial Life. 2024, 10: 342–352. https://doi.org/10.57417/jrnal.10.4_342