FLeS: A Federated Learning-Enhanced Semantic Communication Framework for Mobile AIGC-Driven Human Digital Twins

Samuel D. Okegbile, Haoran Gao, Oluwasegun Talabi, Jun Cai, Changyan Yi, Dusit Niyato, and Xuemin Shen

Abstract

Innovative mobile artificial intelligence-generated content (AIGC) can support the evolution and updating processes of virtual twins (VTs) in human digital twin (HDT) systems. With a reliable and efficient automatic data generation process, the requirement for a timely physical-to-virtual synchronization in HDT can be satisfied. While such an AIGC-enabled HDT system can facilitate modelling high fidelity VTs, generating content that represents the true states in the physical environment and providing timely customized services, it may suffer from a poor understanding of contexts, a lack of creativity, and various security and privacy concerns. In this paper, we propose a novel framework, which integrates federated learning (FL) and semantic communication (SemCom) to enhance performance in the AIGC-enabled HDT system while improving accuracy and convergence properties. First, we present a holistic architectural framework for the proposed FL-enhanced SemCom (FLeS) solution for mobile AIGC-enabled HDT systems and discuss its design requirements and challenges. We later present key technologies necessary to realize the FLeS solution, followed by elaborating on important technical issues to suggest future directions. Experimental results demonstrate that FLeS not only facilitates reliable and personalized content generation but also shows better performance when compared to existing solutions.

Introduction

Human digital twin (HDT) is a next-generation technology that can revolutionize current human-centric systems [1]. The concept of HDT is drawn from the general idea of digital twins (DTs) used in manufacturing industries, where a digital replica of a physical object, located in the physical environment, is created in the virtual environment for system analysis and optimization. Comparatively, HDT involves creating a comprehensive and dynamic digital model, called the virtual twin (VT), of an individual, called the physical twin (PT), and relies on ultra-reliable connectivity between the physical and virtual environments to maintain a true replica (i.e., VT) of each PT in the virtual space [2]. However, HDT dependence on ultra-reliable, secure and privacy-preserving data sharing for VT evolution makes it challenging to realize a timely PT-VT synchronization. Thus, an innovative approach towards extracting data and information in the virtual environment to support the evolution and updating processes of VTs becomes necessary.

Mobile artificial intelligence-generated content (AIGC) is a very useful paradigm that can address this need owing to its ability to creatively generate, augment, and modify valuable and diverse data, tailored to user needs, using advanced artificial intelligence (AI) algorithms deployed at mobile edge networks. Mobile AIGC leverages machine learning algorithms and neural networks to understand user behavior, preferences, and context, and hence, possesses the capability to automate the information creation. When adopted in HDT to enhance personalized healthcare services (PHS), mobile AIGC can facilitate the generation of content that represents the true status of PTs, modelling high fidelity VTs, building versatile testbeds and providing timely customized services [3]. Moreover, mobile AIGC can streamline communication thereby offering quick and contextually relevant outcomes. Mobile AIGC is, therefore, a promising solution to enable HDT and enhance user experiences.

Despite the benefits of this mobile AIGC-driven HDT (MacHDT) [3], such a solution may suffer from many limitations including insufficient contextual understanding, low precision in content generation, a lack of suitable user engagement, and high ambiguity in user inputs. These can result in low accuracy in the generation of content to represent the actual status of users as well as communication ineffectiveness. In addition, there is a need to achieve cross- and multi-modal content generation while meeting specific requirements including accuracy, inference latency and model size, privacy, security and integrity. Currently, it remains unclear how to achieve optimized transmission and interpretation of generated content, especially in HDT where real-time synchronization, high accuracy and reliability are essential.

Digital Object Identifier: 10.1109/MNET.2025.3526556
Date of Current Version: 16 September 2025
Date of Publication: 6 January 2025

Samuel D. Okegbile is with the School of Computing, University of the Fraser Valley, Abbotsford, BC V2S 7M8, Canada; Haoran Gao, Oluwasegun Talabi, and Jun Cai (corresponding author) are with the Department of Electrical and Computer Engineering, Concordia University, Montreal, QC H3G 1M8, Canada; Changyan Yi is with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 211106, China; Dusit Niyato is with the College of Computing and Data Science, Nanyang Technological University, Singapore 639798; Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

Motivation and Contributions

Semantic communication (SemCom) is a very useful technique to address the challenges preventing the success of MacHDT. Incorporating SemCom into the MacHDT framework can enhance its ability to understand, respond, and adapt to user needs, ultimately resulting in more intelligent, context-aware, and user-friendly content generation. SemCom is able to improve communication efficiency and reliability, and enhance the quality of experience (QoE) for human-oriented services while promoting protocol and syntax-independent communication [4]. However, SemCom itself suffers from various complexities relating to the interpretability and explainability of semantic extractions, knowledge coordination and data protection, multi-user interpretation, inconsistent knowledge bases at semantic sources and destinations, and semantic information detection and processing. In addition, SemCom faces difficulty in achieving reliable semantic knowledge modelling because of the need to maintain and constantly update local knowledge models of the source and destination, with complicated relationships, while capturing the deep meaning of knowledge entities. These present trade-offs between semantic extraction accuracy and communication overhead, and also between SemCom performance and security [5].

In this paper, we propose a novel federated learning (FL)-enhanced SemCom (FLeS) solution, which properly integrates FL into the task-oriented SemCom, to enhance performance measures of the MacHDT framework. Although the integration of FL and SemCom has been considered in some existing studies, how such an integration fits the context of HDT, its specific design requirements and challenges as well as the key technologies necessary to realize the conceptualization and implementation of such framework is currently lacking. The main contributions of this paper are summarized as follows.

- We are the first to propose a holistic system architecture of FLeS for MacHDT with a focus on how this innovative solution can improve the overall performance of HDT. We discuss the general concept of SemCom-enhanced MacHDT before demonstrating how federated multi-task learning (FML) can be integrated to enhance such a framework to provide useful insights to readers.
- We discuss specific design requirements and challenges of FLeS-enabled MacHDT. We also provide insightful details on the key techniques that are necessary to facilitate its realization.
- We present a specific comprehensive use case of this FLeS-enabled MacHDT to demonstrate its performance over existing solutions. Last but not least, we discuss useful open issues to inspire future research directions.

Related Works

Before delving into the architecture of the proposed FLeS-enabled MacHDT, we discuss related studies that have previously considered the integration of SemCom and FL. A deep SemCom system based on FL with dynamic model aggregation was presented in [6] where collaboration among multiple users was shown to improve semantic information extraction. The authors in [7] adopted FL to preserve privacy in the SemCom-based metaverse while a federated semantic learning framework presented in [8] aims to collaboratively train semantic-channel encoders of multiple devices with the coordination of a base station-enabled semantic-channel decoder. Similar work was presented in [9], where an FL training method was employed to train the autoencoder for multiple devices and a server. However, these studies [6], [7], [8], [9] only considered the adoption of traditional FL to improve the accuracy of semantic information extraction among multiple devices and a server. Such a solution is not suitable in MacHDT where joint collaborative and personalized training among multiple sources and destinations is essential. Therefore, this paper incorporates a new multi-source and multi-destination collaborative approach to enable joint training of semantic-channel encoders and decoders while facilitating the training of multiple customized AIGC models, thus addressing the diverse personalization needs across different VTs.

> Mobile AIGC leverages machine learning algorithms and neural networks to understand user behavior, preferences, and context, and hence, possesses the capability to automate the information creation.

A General Overview of SemCom Model for AIGC

SemCom has been demonstrated as a useful method for effective content creation in wireless networks. Unlike traditional source-channel coding where source data is directly transmitted, SemCom focuses on an extraction of semantics using deep learning while transmitting the extracted semantics over physical channels. Hence, SemCom can significantly reduce bandwidth consumption and has been considered a promising way to support emerging application scenarios such as HDT and 6G where delivery of a large amount of data is required. In [10], the authors proposed a comprehensive conceptual model for integrating AIGC and SemCom, where a joint optimization of semantic extraction and evaluation metrics, tailored to AIGC services, was presented. In this paper, we extend and improve the capabilities of the SemCom model in [10] to more general applications, which consist of five layers: training, semantic representation, transmission, semantic interpretation and content generation layers.

- **Training Layer:** This first layer captures the learning processes of multiple semantic nodes and AIGC models to facilitate reliable content generation.
- **Semantic Representation Layer:** After training, the system enters into the operational phase where each user converts its raw data into semantic representations, which play a pivotal role in guiding the generation process. This layer generally encompasses extractions of features and the creation of finely-tuned model parameters tailored to specific contextual requirements.

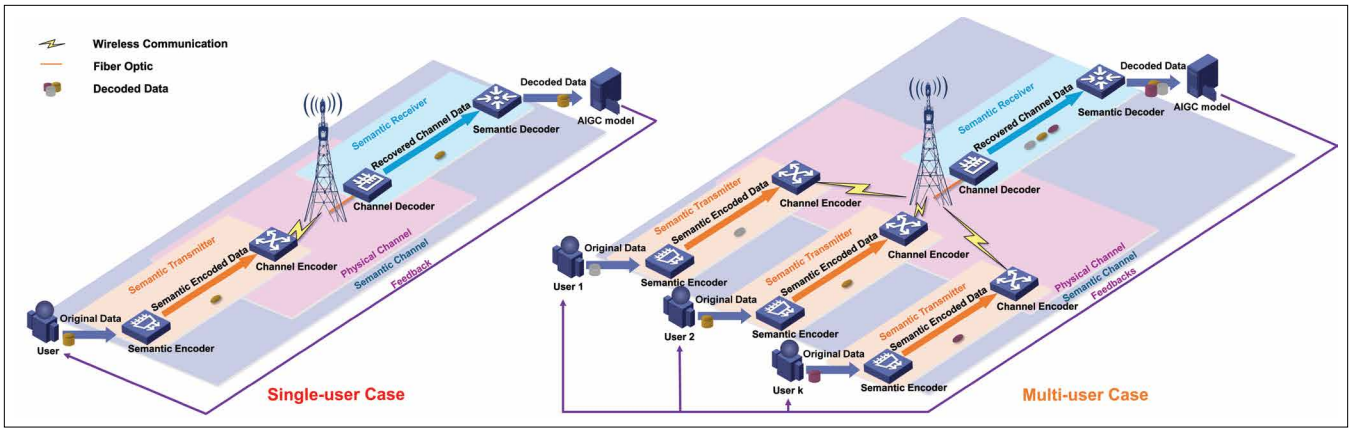

FIGURE 1. Existing SemCom solutions. The single-user case focuses on training a single semantic transmitter-receiver pair. In contrast, the multi-user case involves collaborative training of multiple semantic transmitters, each with diverse modalities of source information, and their associating semantic receiver.

The introduction of feedback mechanisms also ensures that nodes learn from previous mistakes thereby improving the generation of the AIGC model.

- **Transmission Layer:** This layer encapsulates transmission operations within the physical layer, including modules responsible for channel encoding and decoding, and usually involves the transmission of semantic features over a physical channel susceptible to noise.
- **Semantic Interpretation Layer:** This layer consists of modules for the semantic interpretation process at the receiver end and plays a crucial role in the reconstruction and interpretation of the received semantic information.
- **Content Generation Layer:** Content generation layer supports data creation. At this layer, the recovered semantics serve as inputs to the AIGC model and are used to generate content that represents the true status in the physical environment. It leverages semantically encoded information and generates parameters to execute controlled diffusion models, ultimately producing the final content. This final content is then used for the updating and evolution of VTs.

Summarily, these five layers enhance the traditional SemCom model by introducing control and personalization, thereby transforming AIGC into a more adaptable and effective tool for content generation and transmission in HDT.

Generally, existing SemCom frameworks for AIGC can be categorized into: single-user and multi-user scenarios. In the single-user scenario, communication involves only one user as shown in Fig. 1. Hence, each receiver interacts with a single user, and the communication is typically tailored to meet the needs and preferences of a transmitter-receiver pair. While user selection is not a significant concern in such a scenario, each user (i.e., a transmitter) trains its local model using its locally generated data. This results in poor learning and weak semantic information extractions due to limited training data and computing power. As a result, it is natural to introduce collaborations among multiple users, as shown in Fig. 1, where each user maintains a local model while creating a shared understanding [8]. In such a scenario, training of multiple semantic-channel encoders and the associating semantic-channel decoder is performed on a periodic and/or on-demand basis to reduce the effects of changes in the quality of channels over time. The introduction of feedback mechanisms also ensures that nodes learn from previous mistakes thereby improving the generation of the AIGC model. Clear feedback articulates what aspects of the generated content need improvement. This helps users to understand what changes are necessary during semantics extraction and model fine-tuning.

The adoption of single-user and multi-user SemCom models means AIGC service relies on centralized training during the pre-training, fine-tuning, and inference processes [11], making its adoption in HDT impractical. Although the multi-user scenario, presented in Fig. 1, can improve learning performance compared with the single-user scenario, it cannot be adopted in systems with multi-source and multi-destination requirements, such as HDT, where essential collaborations among multiple PTs and VTs necessitate personalized content generation for each VT to accurately capture the true state of its counterpart PT. Thus, a more general SemCom model, which can better match the requirements of MacHDT, becomes necessary.

FLeS Framework for MacHDT

In this section, we propose a new multi-user SemCom model to address the multi-source, multi-destination requirements of MacHDT. This model emphasizes the necessary collaboration between related PTs in the physical space and their corresponding VTs [1], enabling effective data sharing and learning transfer. The training level in FLeS relies on the FML.

Federated Multi-Task Learning

To enable users to generate diverse, personalized and high-quality content required for reliable evolutions of VTs, we incorporate the FML [12] into the mobile AIGC framework to enhance personalization while improving the overall performance by finding a balance between synchronization accuracy and connectivity cost

without compromising security and privacy which are the critical requirements of any HDT network [12]. With FML, each user, participating in the learning process, classifies its features through its domain classifier as either sharable or task-specific features by minimizing the distribution difference - variation between different distributions of data in a given context - between its shared and global parameters. This allows each user to customize its task by training its task-specific model using the global model while contributing to the learning of others by contributing its sharable features during the training of the global model.

The integration of FML is envisioned to enable users to have a real-time interactive environment, context awareness support, and personalized and engaging experiences. Since training is decentralized, such a solution will alleviate the pressure from model and dataset sizes that are often experienced in centralized AIGC services where huge models and datasets sometimes result in critical difficulties during pre-training, fine-tuning and inferring the AIGC model among resource-constrained devices [11]. Furthermore, the issues related to data privacy are significantly eliminated as only gradients are shared while the raw data remains with the users. Therefore, FML is a useful approach that can improve the performance of MacHDT. Next, we present the architecture for FLeS before delving into its design requirements and key techniques.

System Architecture

The conceptual framework of FLeS is shown in Fig. 2, where multi-source and multi-destination collaborative training is performed. In such a scenario, the training process is done in two phases. In the first phase, each transmitting node within a cluster participates in the joint training of multiple semantic-channel encoders and the associated decoder following the FML by exchanging its semantics-enabled sharable models to enhance the knowledge of others while improving the performance of its specific model. In the second phase, each receiving node and the global server, via the FML, similarly participate in the training of multiple decoders and the global AIGC model. The essence of the second phase is to ensure that each receiving node contributes towards enhancing the training of every semantic decoder in each cluster while improving the generation of customized AIGC models using the global one. The multi-source and multi-destination training method presents a decentralized learning approach (through knowledge sharing among multiple encoders and decoders) and ensures the update and maintenance of multiple customized AIGC models necessary for the evolution of the associated VT models.

As in the multi-user scenario, the training in FLeS is performed on a periodic and/or on-demand basis while preserving privacy owing to the adoption of FML. In addition, FLeS can eliminate the operational latency issue often experienced in traditional SemCom due to the presence of a more expansive semantic knowledge base at each semantic node, thus promoting a faster understanding of diverse inputs from users. Generally, the major processes of the FLeS-enabled MacHDT include node selection (also called user association), training, semantic

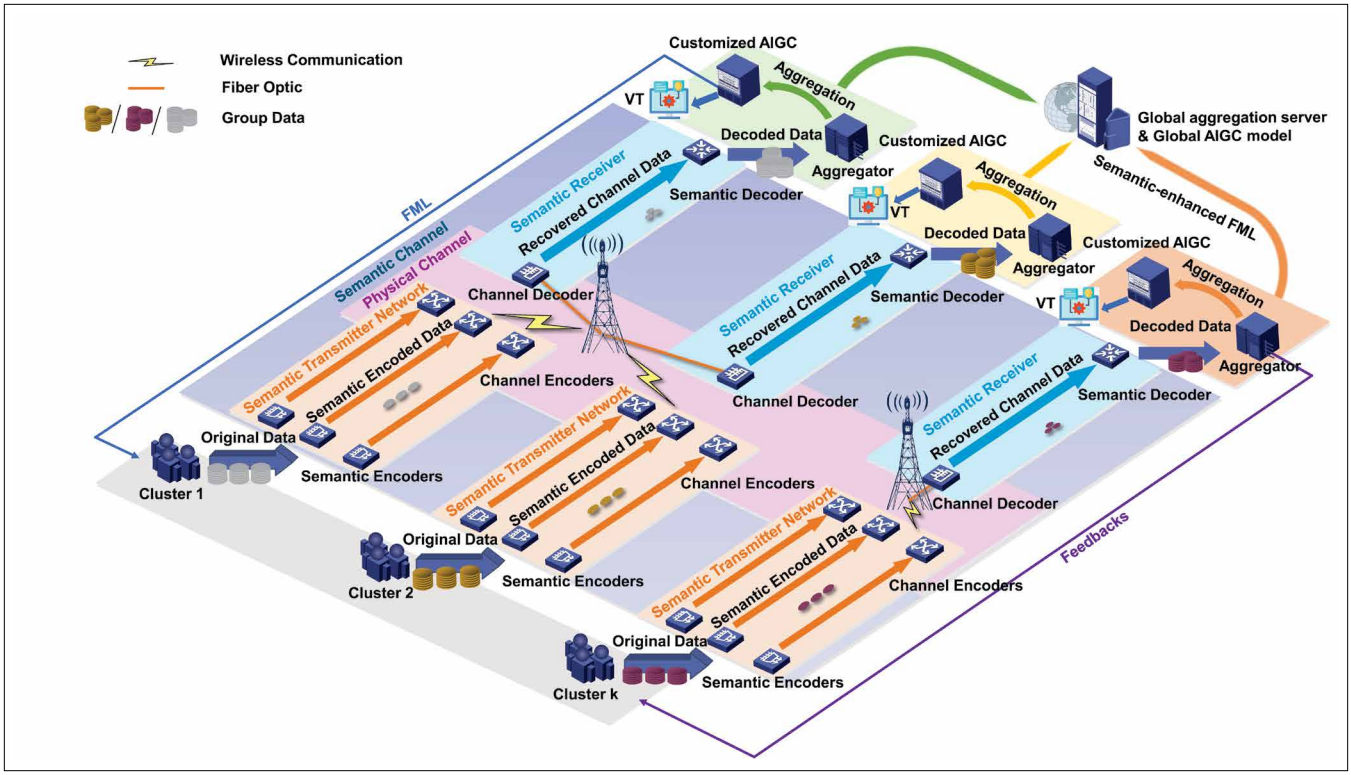

FIGURE 2. Conceptual framework of FLeS. FLeS enables the use of multiple customized AIGC models to facilitate content generation among various VTs with distinct contextual requirements. The second-layer training enhances learning transfer and further expands semantic knowledge bases across multiple clusters.

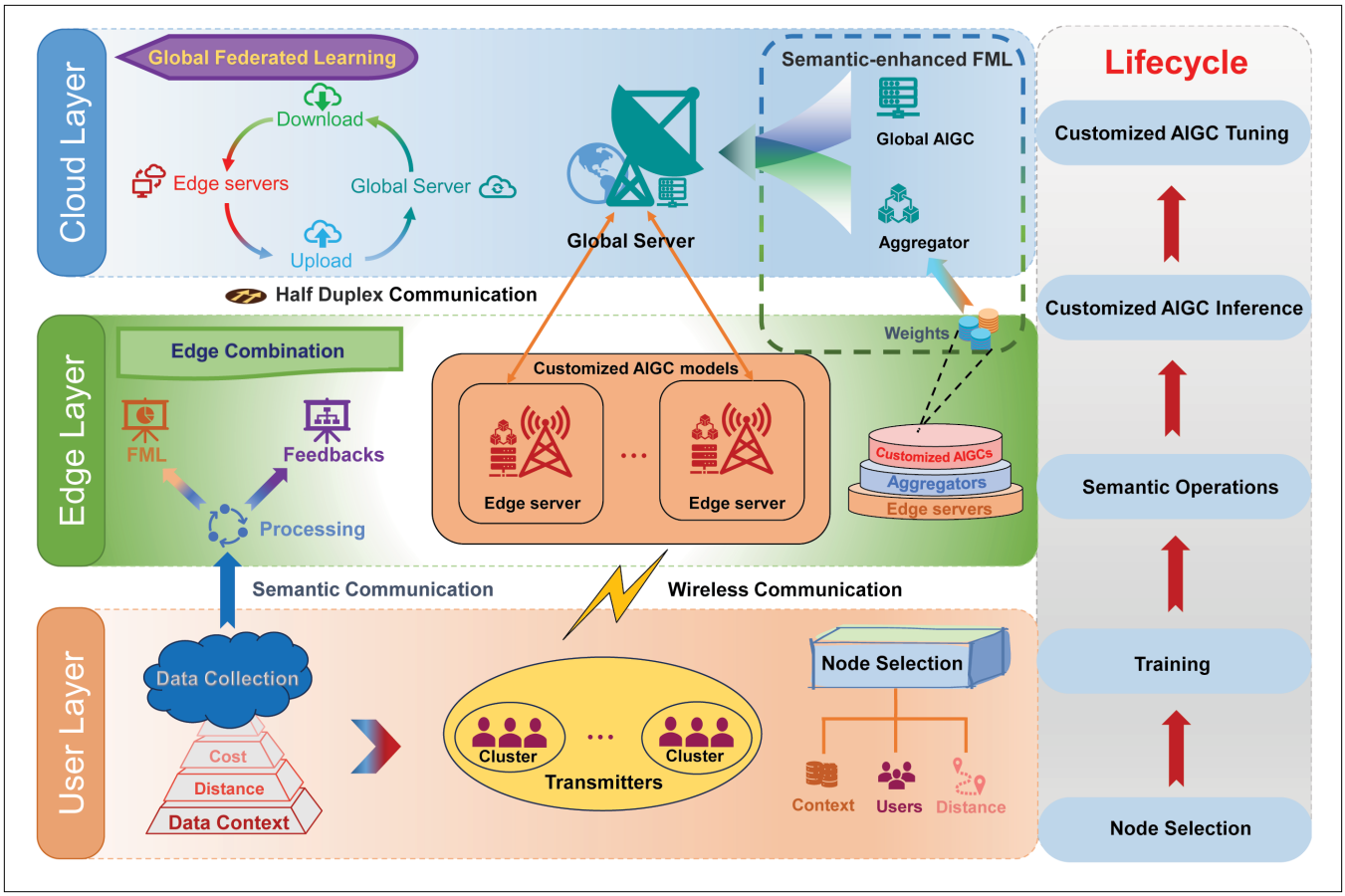

FIGURE 3. Multi-layer architecture of FLeS. The multi-layer training structure adheres to the mobile-edge-cloud collaborative framework, where the initial training layer follows the mobile-edge association, while the subsequent layer is modelled according to the edge-cloud collaborations.

operations, customized AIGC fine-tuning, and customized AIGC inference as shown in Fig. 3.

- **Node Selection:** Node selection is an important process in MacHDT as it involves creating different clusters based on distance, communication and content context. This ensures that users with similar goals are grouped in the same clusters to enable more efficient training while capturing long-range dependencies. It is worth noting that the use of multiple semantic encoders and decoders in a neural network architecture is often associated with more advanced models like transformer-based architectures. Hence, the specific node selection mechanisms depend on the architecture and objectives of the model. For FLeS-enabled MacHDT, the neural architecture search mechanism can effectively guide the node selection process that optimizes the overall performance.
- **Training:** Training involves the joint training of multiple semantic-channel encoders and decoders to improve performance in terms of accuracy and convergence. This stage also includes the pre-training of the global AIGC model as well as the local customized AIGC model in each cluster via the use of semantics-enabled sharable gradients from each transmitting node. Because of the need for high computing power, the global AIGC model is deployed in the cloud while each customized AIGC model is deployed at the edge following the concept of edge-cloud collaboration [13].
- **Semantic Operations:** This process involves the detection and extraction of semantic contents from the source data while compressing or removing irrelevant information. Each semantic encoder identifies the entities in the source data based on the local knowledge at the source and destination before inferring a possible relationship. At the other end, each semantic decoder interprets the information received from the transmitting node (i.e., the semantic transmitter) before recovering the received information that is understandable and useful to the receiving node (i.e., the semantic receiver). Another factor in semantic operations is semantic noise [4] which is usually introduced during the communication process.
- **Customized AIGC Fine-Tuning:** In MacHDT, the customized AIGC fine-tuning stage involves refining the pretrained model for specific tasks and domains. This process tailors the understanding of the model to a particular use case, enhancing its performance and relevance. By utilizing task-specific datasets, adjusting hyperparameters, and selecting or adapting model architectures, this stage ensures the AI model aligns with

the nuances of targeted communication contexts. The iterative fine-tuning, coupled with ongoing monitoring and maintenance, allows for continual adaptation to evolving requirements, resulting in a more effective and context-aware FLeS-enabled MacHDT model.

- **Customized AIGC Inference:** In this stage, the fine-tuned model applies learned knowledge to generate contextually relevant and tailored outputs for specific tasks or domains. Leveraging the training on task-specific datasets, the model interprets input data and produces nuanced, domain-specific content, enhancing its SemCom capabilities. This inference process ensures the adaptability and precision of the model in understanding user queries, providing personalized responses, accurately triggering an evolution of any corresponding VT, and maintaining coherence within the defined context. By emphasizing customization, the AI system optimizes its performance, demonstrating proficiency in generating context-aware content aligned with the intricacies of the intended FLeS-enabled MacHDT domain.

The overall FLeS-enabled MacHDT architecture is presented in Fig. 3 with a focus on its key components.

Design Requirements and Challenges

Clearly, the overall architecture in Fig. 3 demonstrates that implementation of the FLeS-enabled MacHDT is not straightforward. Next, we present the key design requirements and challenges of implementing the FLeS-enabled MacHDT. These challenges can be thus addressed by the key technologies discussed afterward.

1. **User Association:** User association involves the selection of closely related nodes during cluster formation to achieve an improved learning experience and performance. FLeS-enabled MacHDT considers multiple parameters such as semantic context, communication capability, and computation capacity during node selection. However, it is unclear how to establish seamless interoperability and coordination among diverse devices and ensure standardized protocols for efficient data exchange. Additionally, the number of nodes in each cluster is related to both the communication overhead during training and improved learning experiences. Given the dynamic nature of user interactions, an improvement in the learning accuracy among diverse semantic transmitters and receivers is required without incurring a high communication overhead. Balancing personalized content generation for individual nodes while avoiding information overload or misinterpretation adds complexity.
2. **Security:** Despite the adoption of the FL technique to preserve privacy, multi-source and multi-destination collaborative training is susceptible to various security challenges [5], [7]. Importantly, the FML ensures the protection of sensitive data during decentralized training and enables individual contributions to the shared model. However, it is vital to ensure that unauthorized nodes do not have access to the training platform while guaranteeing that all nodes participating in training do not have malicious intentions. Encryption and authentication are imperative to counter potential threats, preventing unauthorized access to user-specific information exchanged during the training phases [2]. Establishing robust access controls and secure communication protocols is critical, mitigating risks associated with malicious actors exploiting vulnerabilities. Advanced techniques against adversarial attacks, model poisoning, and data breaches are essential to guarantee the integrity and confidentiality of FLeS-enabled MacHDT, ensuring user privacy and system reliability.
3. **Multi-Layer Training and Resource Allocation:** The multi-layer training, following the mobile-edge-cloud collaborations as shown in Fig. 3, introduces high training latency. Thus, allocating computing and communication resources efficiently across diverse devices participating in FML is essential for optimal model training. Balancing the computational load, memory usage, and communication bandwidth is crucial to ensure seamless collaboration while preserving user privacy. Moreover, FLeS introduces the need for huge processing capabilities to interpret and respond to various user-generated content. Thus, addressing device heterogeneity, managing communication overhead, and optimizing for energy efficiency are important.
4. **Training-Communication and/or Operations Complexity:** HDT requires extreme ultra-reliable low latency communications (xURLLC) to meet the demand of its latency-critical applications [1]. The training in FLeS-enabled MacHDT is, however, periodic and on-demand depending on changes in the environment which may complicate the communication process between any semantic transmitter and receiver, limiting the possibility of achieving such a stringent requirement. Thus, effective solutions to achieve xURLLC in the presence of periodically scheduled training processes are required.
5. **Knowledge Base Management and Evolution:** Every semantic transmitter-receiver pair is required to maintain and constantly update its local knowledge to encompass the rich meanings of knowledge entities and the intricate relationships among these entities. One possible approach is to employ a graphical framework to represent the semantic connections among diverse entities. Nonetheless, managing a knowledge graph with numerous entities and multi-relational edges can prove exceptionally intricate and challenging [4]. The complexity is further compounded by a fundamental lack of comprehension regarding diverse semantic structures and content meanings, intensifying the difficulties associated with incorporating a semantic knowledge graph into MacHDT frameworks.

Key Techniques

The following techniques can address the specific requirements and challenges associated with the implementation of FLeS-enabled MacHDT.

1. *Intelligent Blockchain (IB):* IB can address the security issues in FLeS-enabled MacHDT. It refers to the integration of AI technologies with blockchain systems, creating a synergistic relationship that enhances efficiency, security, and consensus-based decision-making processes within decentralized networks. By incorporating machine learning algorithms, smart contracts become more adaptive, automating processes based on real-time data analysis [14]. IB systems can dynamically adjust consensus mechanisms and also optimize performance and scalability in the MacHDT. Additionally, IB can enhance security measures by identifying and mitigating potential threats. This fusion of intelligent technologies introduces a new paradigm for blockchain applications, enabling self-learning, predictive capabilities, and improved overall system resilience that is required to prevent unauthorized access and malicious activities during training and communications.
2. *Privacy-Preserving and Distributed Learning Techniques:* In MacHDT, multiple devices or parties collaboratively train various customized models without explicitly sharing their raw data. Techniques such as secure multi-party computation, differential privacy and homomorphic encryption can play pivotal roles in ensuring privacy. By decentralizing the learning process, participants retain control over their sensitive information, mitigating the risks associated with centralized data storage. Furthermore, distributed learning techniques, such as contextual learning [15], self-supervised learning, and reinforcement learning, offer solutions to the challenges of establishing seamless interoperability and coordination among diverse devices during user association. These techniques can also be effective in balancing personalized content generation for individual nodes while mitigating the risk of information overload.
3. *Multi-Source and Multi-Destination Semantic Structure:* This involves exchanging information among multiple sources and destinations while considering the semantic context of the data. Within this complex communication structure, diverse information origins interact with various endpoints, focusing on preserving and conveying meaningful content. Such a structure is crucial in MacHDT, where heterogeneous sources contribute to a collective understanding. Implementing efficient routing and processing mechanisms becomes imperative to ensure seamless communication flow while maintaining semantic coherence. With this structure, batch learning may be adopted to address the training-communication complexities of MacHDT. Additionally, such a structure is significant for handling the management and evolution of semantic knowledge bases across the entire system.
4. *Advanced Generative AI (GAI) Techniques:* Advanced GAI techniques, such as the diffusion model, offer significant potential for optimizing network resource allocation. With its capacity to model complex data distributions and produce high-quality samples, advanced GAI can play a pivotal role in creatively optimizing network resource allocation through generative diffusion models. This optimization leads to enhanced network performance. Generative diffusion models (GDMs) are particularly effective in optimizing the utilization of bandwidth resources and ensuring efficient communication and collaboration among semantic extraction, AIGC inference, and rendering modules. Leveraging GDMs, an optimal bandwidth allocation scheme can be generated based on specific wireless channel conditions and the computing capabilities involved in these modules.

A Case Study of FLeS-Enabled MacHDT

To demonstrate the efficacy of the FLeS solution, we developed a platform for MacHDT and analyzed how multi-source, multi-destination collaboration impacts content generation reliability, focusing on semantic error rate and throughput as performance metrics. The error rate is obtained as the normalized average absolute difference between predicted values and actual target values of the samples [8]. Additionally, throughput refers to the amount of semantic data successfully transmitted and computed per second in each cluster, determined by setting a specific bandwidth within the simulated wireless environment. For comparison, we also implemented both single-user and multi-user frameworks in the context of MacHDT. For demonstration purposes, the single-user case utilized one transmitter-receiver pair, whereas the multi-user case involved three transmitters and one receiver. For FLeS, there are three clusters, each comprising three transmitters and one receiver. The clustering method systematically categorizes users based on data similarity into sedentary, moderate, and light activity clusters.

We employed a single antenna for each semantic transmitter and multi-antennas for each semantic receiver. Each semantic transmitter network comprises an input layer, an LSTM layer with 64 units, a dense layer with 64 units, ReLU activations, and a dedicated section for textual data processing. This section begins with an embedding layer, tailored to accommodate an extensive vocabulary size of *len(tokenizer.word_index) +1*, with an output dimension of 64 to transform texts into dense vectors. Following this, a Conv1D layer with 64 filters and a kernel of size 2, employing ReLU activation, captures textual nuances. A GlobalMaxPooling1D layer then emphasizes salient features, effectively processing the textual data before it is combined with numerical data features.

For the semantic receiver network, we prioritized the integration and further processing of textual and numerical data. The combined feature set from the transmitter network undergoes processing through a dense network architecture: a first layer of 256 units with ReLU activation and L2 regularization, followed by dropout for

regularization, a second dense layer of 128 units, also with ReLU and L2 regularization, and finally, a layer of 64 units with the same configuration, leading to a linear activation output layer designed for varied output requirements. We considered additive white Gaussian noise channels. Using the natural language generation technique, we obtained both customized and global AIGC models.

Using dementia patient healthcare data comprising 1000 samples (80% for training and 20% for operation), sourced from Kaggle, we aimed to generate content including age, weight, and physical activity. This content is intended for updating VT models tailored to cardiovascular and respiratory system domains. Given the presence of numerical and textual data in the input features, we designed a multi-input model — a hybrid neural network incorporating fully connected layers, dropout layers, batch normalization layers, embedding layers, convolutional layers, and global max pooling layers — to ensure model stability while accommodating both textual and numerical inputs. The FML was implemented as in [12].

To evaluate the error rate, we employed the multi-input model under single-user, multi-user, and FLeS scenarios, maintaining uniformity in the training set and learning rate across all experiments. We observed the operation phase by progressively increasing the number of epochs while monitoring the mean absolute error (MAE), mean square error and the resulting error rate. As depicted in Fig. 4, FLeS demonstrates a superior convergence rate compared to the other scenarios, while also exhibiting the lowest error rate, underscoring its superior performance over both single-user and multi-user scenarios. The similarity in the error rates after convergence across the three scenarios can be attributed to utilizing a fixed learning capability, the same dataset, and the selection of identical optimization algorithms in our experiment.

As shown in Fig. 5, the average throughput per cluster confirms that FLeS achieves higher throughput compared to single-user and multi-user scenarios. FLeS also demonstrates superior accuracy and convergence rates over the multi-user framework, while the single-user framework faces performance issues due to a limited learning experience. In multi-destination scenarios, both single-user and multi-user frameworks show significant performance degradation, as only one receiver participates in training. In contrast, FLeS maintains high performance through multi-layer training with multiple receivers. By optimizing node selection and considering user-server data distribution and requirements, FLeS achieves faster processing and outperforms multi-user schemes.

Conclusion and Future Research Directions

In this paper, we introduced a novel FLeS-enabled MacHDT framework to improve PT-VT synchronization and content generation in HDT, supporting personalization and reliable evolution of high-fidelity VTs. We outlined the framework's architecture, design requirements, and challenges, and highlighted key techniques with a comprehensive case study demonstrating its effectiveness.

In contrast, FLeS maintains high performance through multi-layer training with multiple receivers.

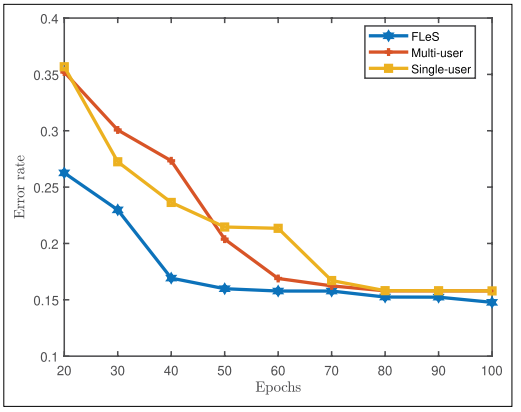

FIGURE 4. Performance of FLeS in terms of error rate.

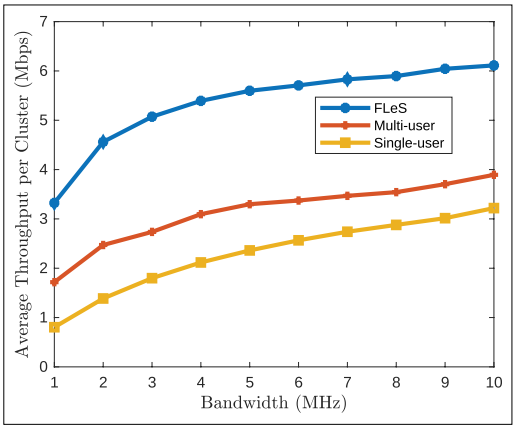

FIGURE 5. Performance of FLeS in terms of throughput per cluster.

Despite promising results, important open issues remain for future exploration.

1. **Effective Node Selection Algorithm:** MacHDT employs multi-layer learning, requiring optimal node selection. This process is not simple, necessitating integration of semantic understanding, modelling, user preferences, and context awareness. Collaborative filtering enhances optimal node selection policies tailored to individual AIGC models. Investigating dynamic node selection with AI and diverse machine learning algorithms is crucial for effective implementation.
2. **Semantics Reconstruction:** Introducing FLeS poses several challenges in recovering meanings from semantics in MacHDT. Investigating methods for accurate recovery becomes crucial. Model aggregation with semantic alignment via knowledge distillation may offer a viable solution. Additional considerations may include federated knowledge transfer, secure multi-party computation, optimized communication protocols, and robust federated optimization to facilitate precise meaning recovery for authorized receivers.
3. **Semantic Knowledge Storage and Management:** Managing knowledge across evolving

systems is challenging. Solutions like content caching, model compression, and knowledge deduplication can be incorporated to prevent duplication. Similarly, an intelligent edge-cloud computing framework may optimize storage across edge and cloud servers while also enhancing efficiency.

Acknowledgment

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant; in part by the Concordia University PERFORM Research Chair Program; in part by the National Research Foundation, Singapore, and Infocomm Media Development Authority under its Future Communications Research and Development Programme, Defence Science Organisation (DSO) National Laboratories through the AI Singapore Programme under Grant FCP-NTU-RG-2022-010 and Grant FCP-ASTAR-TG-2022-003; in part by the Singapore Ministry of Education (MOE) Tier 1 under Grant RG87/22; in part by the NTU Centre for Computational Technologies in Finance (NTU-CCTF); in part by Seitee Pte Ltd.; in part by the RIE2025 Industry Alignment Fund–Industry Collaboration Projects (IAF-ICP) administered by A*STAR under Award I2301E0026; and in part by the Alibaba Group and NTU Singapore through Alibaba-NTU Global e-Sustainability CorpLab (ANGEL).

References

[1] S. D. Okegbile et al., "Human digital twin for personalized healthcare: Vision, architecture and future directions," *IEEE Netw.*, vol. 37, no. 2, pp. 262–269, Mar./Apr. 2022.
[2] J. Chen et al., "Networking architecture and key supporting technologies for human digital twin in personalized healthcare: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 1, pp. 706–746, 1st Quart., 2024.
[3] J. Chen et al., "A revolution of personalized healthcare: Enabling human digital twin with mobile AIGC," *IEEE Netw.*, vol. 38, no. 6, pp. 234–242, Nov. 2024.
[4] G. Shi et al., "From semantic communication to semantic-aware networking: Model, architecture, and open problems," *IEEE Commun. Mag.*, vol. 59, no. 8, pp. 44–50, Aug. 2021.
[5] W. Yang et al., "Semantic communications for future internet: Fundamentals, applications, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 213–250, 1st Quart., 2022.
[6] H. Xing et al., "A multi-user deep semantic communication system based on federated learning with dynamic model aggregation," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2023, pp. 1612–1616.
[7] J. Chen et al., "Trustworthy semantic communications for the metaverse relying on federated learning," *IEEE Wireless Commun.*, vol. 30, no. 4, pp. 18–25, Aug. 2023.
[8] H. Wei et al., "Federated semantic learning driven by information bottleneck for task-oriented communications," *IEEE Commun. Lett.*, vol. 27, no. 10, pp. 2652–2656, Oct. 2023.
[9] H. Tong et al., "Federated learning based audio semantic communication over wireless networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2021, pp. 1–6.
[10] G. Liu et al., "Semantic communications for artificial intelligence generated content (AIGC) toward effective content creation," *IEEE Netw.*, vol. 38, no. 5, pp. 295–303, Sep. 2024.
[11] X. Huang et al., "Federated learning-empowered AI-generated content in wireless networks," *IEEE Netw.*, vol. 38, no. 5, pp. 304–313, Sep. 2024.
[12] S. D. Okegbile et al., "Differentially private federated multitask learning framework for enhancing human-to-virtual connectivity in human digital twin," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 11, pp. 3533–3547, Nov. 2023.
[13] S. D. Okegbile, B. T. Maharaj, and A. S. Alfa, "A multi-class channel access scheme for cognitive edge computing-based Internet of Things networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 9912–9924, Sep. 2022.
[14] S. D. Okegbile et al., "A reputation-enhanced shard-based byzantine fault-tolerant scheme for secure data sharing in zero trust human digital twin systems," *IEEE Internet Things J.*, vol. 11, no. 12, pp. 22726–22741, Jun. 2024.
[15] G. Li et al., "Cross-modal attentional context learning for RGB-D object detection," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1591–1601, Apr. 2019.

Biographies

Samuel D. Okegbile (Member, IEEE) (samuel.okegbile@ufv.ca) received the Ph.D. degree from the University of Pretoria in 2021. He is currently an Assistant Professor with the University of the Fraser Valley, Canada. His current research interests include distributed systems, wireless communications, mobile computing, data sharing, cyber-security, AI, and human digital twins.

Haoran Gao (haoran.gao@mail.concordia.ca) is currently pursuing the M.Sc. degree with the Department of Electrical and Computer Engineering, Concordia University, Canada. His research interests include AI-generated content and federated learning.

Oluwasegun Talabi (oluwasegun.talabi@mail.concordia.ca) is currently pursuing the M.Sc. degree with the Department of Electrical and Computer Engineering, Concordia University, Canada. His research interests include semantic communications and federated learning.

Jun Cai (Senior Member, IEEE) (jun.cai@concordia.ca) received the Ph.D. degree from the University of Waterloo, Waterloo, ON, Canada, in 2004. He is currently with the Department of Electrical and Computer Engineering, Concordia University, Canada, as a Full Professor and the PERFORM Centre Research Chair. His current research interests include edge/fog computing, eHealth, radio resource management in wireless communications networks, and performance analysis.

Changyan Yi (Member, IEEE) (changyan.yi@nuaa.edu.cn) received the Ph.D. degree from the University of Manitoba, Winnipeg, MB, Canada, in 2018. He is currently a Professor with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China. His research interests include game theory, queueing theory, machine learning, and its applications.

Dusit Niyato (Fellow, IEEE) (dniyato@ntu.edu.sg) is currently the President's Chair Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include edge intelligence, machine learning, and incentive mechanism design.

Xuemin (Sherman) Shen (Fellow, IEEE) (sshen@uwaterloo.ca) is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research focuses on network resource management, wireless network security, social networks, and vehicular ad hoc networks. He is a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, and a Chinese Academy of Engineering Foreign Fellow.