Next-Gen IoT: Human Digital Twins and Large Language Models for Proactive User Engagement

Luigi Serreli, *Student Member, IEEE* , Massimo Farina, and Michele Nitti, *Senior Member, IEEE*

***Abstract*—In the rapidly evolving landscape of the Internet of Things (IoT), the integration of Human Digital Twins (HDTs) and advanced computational frameworks, such as Large Language Models (LLMs), marks a step towards creating more proactive and user-centric systems. This work explores the development of next-generation IoT solutions that leverage these technologies to enhance user engagement while minimising the need for explicit user intervention. By accurately mirroring user behaviours and preferences, HDTs enable systems to predict needs effectively and automate responses, fostering a more intuitive interaction environment. Our study highlights the potential of these integrated technologies necessary to streamline user experiences and substantially extend IoT systems' capabilities in understanding and reacting to human contexts through verbal language. This strategy lays the foundation for the next generation of IoT systems, characterised by proactive interactions and enhanced digital assistance, finalised to more personalised and efficient user experiences in the connected world.**

***Index Terms*—Proactivity, Human Digital Twins, Large Language Models, IoT, User-Centric**

I. INTRODUCTION

THE rapid advancement of the Internet of Things (IoT) has conducted in a new era of connected devices and intelligent systems. Despite significant technological progress in AI and natural language processing (NLP), a critical challenge persists: although IoT research promotes automatic and responsive systems, they usually focus on analytical intelligence but still lack other types of intelligence, such as emotional and creative intelligence, crucial for everyday interactions [1]. While modern IoT systems increasingly integrate AI, many still rely on reactive approaches, limiting their adaptability to users' evolving needs and personalization capabilities [2]. The evolution of new IoT solutions goes through the concept of Digital Twin (DT), which enables models capable of mimicking object behaviour in operational environments [3]. Human Digital Twins (HDTs) have the potential to extend this concept, creating digital replicas of individual users, capturing interests and preferences for personalised IoT experiences. The IoT ecosystem can act as the perceptive and active framework for creating an HDT, with a multitude of devices and interfaces collecting data from and interacting with users' environments.

Opposed to reactive systems, proactive IoT systems aim to minimise the need for user-initiated actions while learning from users' behaviours and preferences. Understanding both verbal and non-verbal language is fundamental for a genuinely proactive IoT system [4]. While IoT already retrieves non-verbal information (facial expressions, heart rate, movements), verbal language processing is critical for seamless integration into daily life [5]. Verbal language has been the cornerstone of human evolution, serving as the foundation for communication, collaboration, and cultural development. It enables us to convey complex ideas, share knowledge across generations, and build societies. The capacity to utilise spoken language allows for expressing thoughts, emotions, and intentions, promoting understanding and cooperation among individuals. In the context of IoT and HDTs, harnessing oral language goes beyond simple data analytics. It adds a comprehensive understanding of how individuals interact among themselves. The use of spoken language not only enriches the information derived from user interactions but also integrates seamlessly into daily life as a supportive tool, rather than merely serving as an interface for command and response [6]. Nowadays, various devices enable natural conversational interactions with humans. For consistency, we will use the term Personal Digital Assistant (PDA) to refer to assistants utilising NLP and machine learning to interact with users. Serving as a primary interface for user interaction, PDAs acquire and execute tasks on behalf of the user, providing information and recommendations that significantly enhance the IoT experience [7].

This work presents a framework designed to acquire knowledge about the user in order to provide increasingly personalised services. In particular, we focus on the key contribution of verbal language and how it can be integrated to provide seamless proactive applications.

This work provides the following contributions:

1) We propose a novel framework that integrates HDTs, PDAs, and Large Language Models (LLMs) to enhance IoT system proactivity, addressing the current lack of emotional and creative intelligence in traditional IoT solutions.
2) We introduce a novel approach that enables continuous learning from user behaviour, leveraging both verbal and non-verbal communication to create a more comprehensive understanding of user needs and preferences.
3) We address critical privacy concerns inherent in such personalised systems, proposing strategies for context-aware communication and user control over sensitive information.

The following sections explore various aspects of the proposed framework. Section II describes the enabling paradigm.

L. Serreli, M. Farina and M. Nitti are with the Department of Electrical and Electronic Engineering (DIEE), University of Cagliari, Italy.

L. Serreli and M. Nitti are with National Telecommunication Inter University Consortium, Research Unit of Cagliari, Italy.

E-mails: L.S. luigi.serreli@unica.it, M.F. m.farina@unica.it and M.N. michele.nitti@unica.it.

Section III outlines the architecture and components of our proposed framework, and Section IV shows a preliminary evaluation of the system. Section V introduces crucial considerations for ensuring data privacy and confidentiality. Finally, the article concludes with the final remarks in Section VI.

II. Paradigm Shift

This section describes the two enabling paradigms that support the proactive reference platform by collectively enhancing the system's ability to deliver personalised and proactive user experiences.

A. From Digital Twin to Human Digital Twin

DTs have gained significant attention in recent years, with a rapidly growing body of literature exploring their potential applications. A DT is a virtual representation of a physical object or system that can simulate, monitor, and optimise its real-time performance [8]. Similarly, the HDT represents a digital replica of an individual capable of encompassing the entire spectrum of human attributes, from anatomy to behaviour [9]. These dimensions interact dynamically, making accurate modelling of HDTs challenging.

The resulting profile provides insights into their behaviour in real-time, enabling personalised services and recommendations. The transition from Digital Twins to Human Digital Twins is enabled by advanced data integration and representation technologies. These include sophisticated data fusion techniques that merge heterogeneous data sources, such as transformers for multimodal data fusion or attention-based mechanisms for weighting and combining information from diverse inputs. In addition, the progress is towed by advanced embedding representations that transform diverse data types into unified vector expressions. These embeddings can then be used to construct dynamic knowledge graphs, which capture the complex, evolving relationships between different aspects of a user's life and environment.
Key aspects of HDTs include:

Dynamicity: HDTs exhibit remarkable flexibility in creating virtual replicas of individual users across various domains [10]. By leveraging machine learning algorithms, HDTs can dynamically adapt to users' changing statuses and behaviours in real time. These virtual representations continuously analyse historical and current data to create comprehensive, evolving profiles of users by constantly updating their model to reflect the user's current state. This dynamic approach enables highly personalised services and recommendations that evolve with the user over time.

Multimodal Data Fusion: HDTs leverage advanced multimodal perceiving capabilities, combining quantitative and qualitative aspects of a person's life to provide a holistic view of their behaviours, preferences, and emotional states [11]. This multimodal approach significantly upgrades traditional DTs to HDTs. Indeed, HDTs incorporate a multitude of qualitative inputs, allowing algorithms to interpret facial expressions, voice tones, and other physiological indicators to discern emotions, sarcasm, and other important features not measurable quantitatively. This expansion allows for a more nuanced understanding of the user, going beyond simple data points to capture the complexity of human experience and fostering emotional intelligence, allowing the system to generate personalised suggestions and solutions to the user's unique profile and current context.

B. The Rise of LLM-Powered PDAs

Traditional PDAs have long been task-oriented, executing simple commands and providing basic information. However, integrating LLMs marks a significant advancement in human-machine interaction, initiating a new era of conversational AI. LLMs serve as a key enabling technology for next-generation PDAs, providing unprecedented capabilities in understanding and generating human-like text. Based on transformer architectures, these models excel at processing unstructured data like conversations, structured data and application logs. Their understanding of context and nuance allows for more natural and interactive responses. Techniques like few-shot learning and prompt engineering enable LLMs to quickly adapt to specific tasks and user contexts. Moreover, efficient fine-tuning methods such as Low-Rank Adaptation (LoRa) allow for customising these large models to specific domains or individual users without extensive retraining, making them practical for deployment in personalised PDA systems [12]. This shift, grounded in semantic comprehension and the PDAs' pervasiveness, introduces several key advancements to HDT interactions:

Contextual Interpretation: LLM-powered PDAs have the potential to enhance human-machine interpretation. Building upon their ability to process diverse data types, these systems can now integrate information from multiple sources to form a cohesive understanding of the user's context. This includes the content of conversations and the user's historical behaviour, current environment, and even emotional state as inferred from various sensors and data streams.

PDA Pervasiveness: PDAs exhibit pervasiveness across devices and geographical locations, offering context-appropriate interactions. The ubiquity of IoT devices, wearables, and smart home technologies provides a continuous data stream about the user's environment and activities. This pervasiveness, combined with the contextual understanding enabled by LLMs, allows PDAs to offer assistance across various aspects of a user's life.

III. The Envisioned IoT-PDA Framework

In this subsection, we present an IoT-PDA framework to deliver a personalised and proactive user experience, which integrates verbal language. The framework is based on the concept of DTs so that, for every entity in the real world, there is a digital counterpart in the virtual world. The reference user-centric scenario is depicted in Figure 1.

A. Keystones

- **Human Digital Twin:** Whenever a user registers on the platform, the cloud provider offers a virtual user space (dashed black line in the Figure) where the user has

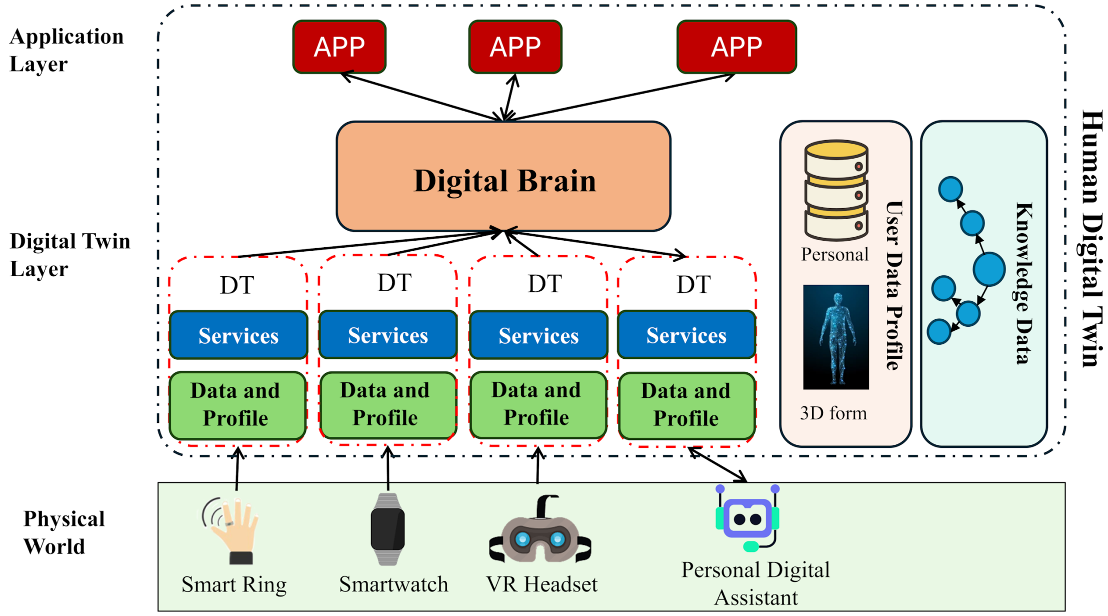


Figure 1. High level reference scenario.

data storage and computing available to contain her/his information and run her/his DTs and applications. This space has the potential to become the users' virtual counterpart, i.e. their HDT. This can be implemented using containerisation technologies and microservices principles, utilising orchestrators like Kubernetes for efficient resource management. Static users' information is contained in the user data profile, which exposes fixed information on an individual in the context and time frame of the scenario of use, such as their personal data or a 3D form of the subject. All the real-world entities of the user are then digitalised in the form of DTs at the Digital Twin Layer representing physical entities and thus exposing their profile and raw data. However, DTs can further enhance the physical entities and thus offer additional services and it is up to the user to decide which services have to be installed and activated.
The HDT is then not only the representation of the information strictly related to the users, as derived by their profile, but a combination and integration of data from diverse and heterogeneous sources, which has to consider all the raw data, services, and applications installed in the virtual user space of the user. Research should investigate how orchestrators could be adapted for HDT-specific requirements, particularly focusing on ensuring scalability, modularity, and secure isolation of user spaces while maintaining efficient resource management.

- **Knowledge Data.** Whenever a user talks to other people or uses its smart devices or applications, the system is able to learn their habits, interests, preferences, and thought processes. This information is stored in the user's HDT in the form of knowledge data, creating what can be considered a Personal Knowledge Graph (PKG) [13]. Unlike traditional knowledge graphs that focus on globally important entities, PKGs are specialised resources that capture structured information about entities personally related to individual users, including those that might not be globally significant but are relevant to the user's daily interactions. We assume that this graph can be expressed using different forms such as Resource Description Framework (RDF), Knowledge Representation Learning (KRL) or Knowledge Graph Embedding (KGE). The goal is to memorise the reflection of the real world's rules of behaviour related to that particular user. In this sense, these graphs are not common to the different users across the platform, but each user will build her/his own knowledge data.
- **Digital Brain.** This is the core of the proposed framework as it enables the HDT to learn from the users' behaviour, share knowledge with the Digital Brain of other users, and auto-evolve, thus proactively creating new services and useful interactions with the users.

B. Digital Brain

Figure 2 shows a focus on the structure of the Digital Brain. The blue modules represent the main flow, which integrates information from verbal, i.e. PDAs, and non-verbal, i.e. IoT devices, communication and how the user utilises the installed apps. In particular, the green modules represent the additional components needed to enable the system to interface with different sources of structured and unstructured data, while the violet blocks focus on sensitive data and privacy concerns. Digital Brain works towards integrating adaptive learning models and neural architectures with LLMs.

- **Acquiring User's Knowledge.** The Digital Brain aggregates user interests through a multi-modal approach. The system could employ triplet extraction, a process that

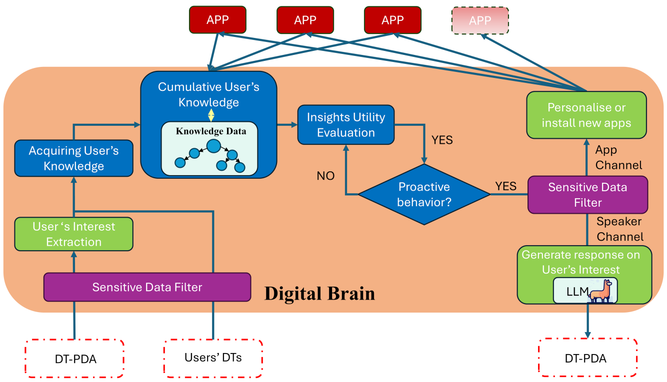


Figure 2. Digital Brain structure.

identifies relationships between entities through subject-predicate-object structures. This fundamental semantic representation transforms raw, unstructured data into structured knowledge that can be efficiently processed and reasoned upon. Building upon the interests extracted, IoT sensor data serves as contextual augmentation, enriching this knowledge with precise environmental parameters. LLMs could serve as a unifying framework to synthesise these diverse data streams. By fine-tuning these models on IoT sensor data, application logs, and processed verbal communications, they could interpret both explicit and implicit user interests across various interaction channels. Topic classification algorithms could then collaborate with these fine-tuned LLMs to refine broad understandings into specific, actionable insights.

- **Cumulative User's Knowledge.** This module creates and updates dependencies between various user interests, communicating with the *knowledge data*. To realize the cumulative knowledge function, we suggest leveraging a dynamic knowledge graph structure. Such a structure could be effectively implemented using graph database technologies like Neo4j, which are well-suited for representing complex, interconnected data. The cumulative aspect could be achieved through a dual approach of immediate updates and long-term learning. For immediate updates, a rule-based system could adjust edge weights in the graph based on the frequency of topic occurrences, while periodic retraining of the models could capture evolving user interests over time. Graph Neural Networks or Graph Attention Networks could be integrated to capture more nuanced relationships over time. These networks have shown promise in learning to assign importance to different node connections, potentially allowing the system to infer complex dependencies between topics that may not be immediately apparent. [14]
- **Insights Utility Evaluation.** This module represents the primary role of the Digital Brain, which is to predict future user needs and anticipate requests. For this purpose, we propose leveraging LLMs to generate contextually relevant responses based on the user's knowledge data. While current models typically rely on simple scoring mechanisms, we suggest a more sophisticated multi-faceted scoring system that could evaluate complex factors, including the user's short-term activities, long-term behavioural patterns, emotional states, and contextual relevance (e.g., location, time of day). These responses could be evaluated through feedback and reflection mechanisms to determine their utility and relevance to the user's current needs and long-term interests. The system could refine its proactive suggestions through an adaptive feedback loop, monitoring explicit user engagement and interpreting subtle implicit cues, thus continuously evolving its understanding of user preferences and intervention effectiveness.

Table I provides an overview of the three modules, highlighting currently available technologies that could serve as building blocks, identifying critical technological gaps that still need to be addressed, and suggesting potential applications. This overview highlights both the immediate possibilities and future research directions.

The Digital Brain is highly time-dependent, which means that topics create connections among themselves with a given utility, but this utility can decrease if the user is poorly related to the topic, e.g. she/he does not mention it too often when speaking, or it can increase if it is recurrent in the user's life.

The system leverages this utility function to detect opportunities for proactive behaviour. When such an opportunity is identified, the system takes action through the PDA, generating contextually relevant responses. These responses range from timely information to more complex interventions. In addition, the system recommends or automatically installs new applications that align with the user's evolving interests and needs. Conversely, it may suggest removing apps that have become less relevant or useful over time. As shown in the Figure, newly installed apps are represented by transparent elements. These apps play a necessary role in the ongoing acquisition and enhancement of the user's knowledge data, contributing to a dynamic and ever-evolving system that continually adapts to the user's changing requirements and interests. The implementation of the system raises important considerations regarding computational efficiency, scalability and privacy. A key feature to manage computational overhead is the dynamic weight mechanism in our knowledge graph, which naturally prunes inactive connections over time. This approach ensures that rarely accessed or outdated information gradually decreases in prominence, reducing both memory usage and processing requirements without manual intervention. Furthermore, the use of LLMs and knowledge models introduces computational overhead that must be carefully managed. Model distillation techniques could significantly reduce LLM parameter counts while maintaining domain-specific performance, while quantisation methods could further optimise memory and processing requirements. Furthermore, the decentralised nature of our proposed architecture, where processing occurs primarily within each user's virtual space, inherently supports scalability for large-scale deployments. This virtual space can be implemented using containerisation technologies, allowing for dynamic resource allocation and horizontal scaling. For IoT-specific scenarios, edge computing

Table I
TECHNICAL OVERVIEW OF DIGITAL BRAIN COMPONENTS

Component	Available Technologies	Still Needed	Potential Applications
Acquiring Knowledge	Speech recognition	IoT-specific embeddings	Speech-to-text conversion
	Fine-tuning techniques	Multi-modal fusion (text + IoT Data)	Interest extraction
	Unsupervised learning	Contextual understanding	Semantic analysis
Cumulative Knowledge	Graph databases	Dynamic weight updates	Knowledge representation
	Graph Neural Networks	Long-term learning	Relationship inference
Insights Evaluation	Multi-faceted scoring	Context-aware triggers	Proactive suggestions
	Feedback mechanisms	Adaptive thresholds	User engagement optimization

principles can process time-critical data closer to the source, while federated learning techniques could enable privacy-preserving knowledge sharing across multiple users. This distributed approach allows linear scaling with user growth rather than facing exponential resource demands.

Moreover, processing conversations raises important privacy concerns. As a result, the Sensitive Data Filter module should use modern privacy-preserving technologies that can be integrated into the framework to either automatically identify and redact sensitive information while preserving the conversation's context and evaluating the sensitivity of responses and selecting appropriate communication channels through tools such as Nemo Guardrails and Llama Guard [15].

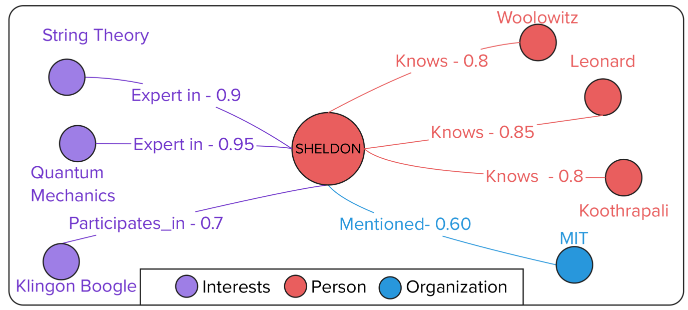


Figure 3. Initial KG constructed from the first analysed scene, showing detected entities (nodes) and their relationships (edges) with confidence scores.

IV. Preliminary Evaluation

In this section, we show the effectiveness of our framework, using as a reference case the analysis of dialogues from "The Big Bang Theory" TV show. We focus solely on Sheldon, the main character of the TV show, using his conversations to construct the HDT. For simplicity of demonstration, we analyse textual data rather than audio processing, specifically examining the third scene of the first episode, where Penny explores Leonard and Sheldon's apartment, prompting awkward attempts at conversation from Howard and Raj, while Sheldon proudly explains his scientific work.

We suppose the system processes the entire scene, considering only Sheldon's sentences, thus making the analysis more difficult. While we acknowledge this represents a controlled dataset rather than diverse real-world scenarios, this methodological choice offers several advantages for our initial proof-of-concept validation. First, it provides consistent character personalities with well-documented behavioural patterns that can be objectively verified against the source material. Second, the episodic nature allows us to track knowledge evolution over time in a controlled environment where external variables are minimised. Third, it serves as an accessible reference point that readers can intuitively understand, facilitating comprehension of our framework's core mechanisms.

Simulating the acquiring User's Knowledge, the textual data is transformed into vector representations using the all-MiniLM-L6-v2 model, configured with mean pooling strategy and standard tokenisation parameters, which generates 384-dimensional embeddings capturing the semantic meaning of each dialogue segment. These embeddings are stored to enable future Retrieval-Augmented Generation (RAG) operations, maintaining contextual awareness across multiple interactions.

To implement the Cumulative User's Knowledge module, we employ a two-stage approach combining LLM analysis with knowledge graph construction. In the first stage, the framework employs Llama 3.1 8B to perform deep semantic analysis of the conversation, extracting entities, relationships, and their attributes. The model was called with specific inference parameters (temperature=0.1, top_p=0.9, max_tokens=1024). The low temperature setting was chosen to ensure deterministic and consistent outputs necessary for reliable entity extraction, while the relatively high *top_p* value allows the model to maintain some flexibility when identifying non-obvious relationships between concepts. These parameters were critical for balancing precision with contextual understanding in our knowledge graph construction. The extracted elements are then fed into the second stage, where they are organised into a dynamic knowledge graph structure that evolves as new information is processed. Each identified entity becomes a node in the graph, while the relationships between them form weighted edges based on their co-occurrence and contextual proximity in the conversation. Weights are initialised with a base value of 0.4 when relationships are first detected and evolve dynamically, strengthening when entities are repeatedly mentioned together and decaying over time if not reinforced.

Figure 3 shows the initial knowledge graph. It successfully identifies different types of entities (shown in different colours). Each relationship represents an edge type (e.g., "Knows", "Expert_in") with an associated confidence score. It's worth noting that with just a few dialogue exchanges, the system captures meaningful insights about Sheldon's friendships, expertise areas, and interests. Nevertheless, there are a few missing elements in the knowledge graph, such as the omission of Penny, who rarely interacts directly with Sheldon in the analysed scene and is therefore excluded by the current system. This highlights significant opportunities for improvement in the construction of knowledge graphs, underscoring the importance of refining these models as a key area for future research efforts. The evolving graph also shows how the framework guides entity classification. For example, "MIT" is correctly identified as an Organization entity. This classification is facilitated by the LLM's world knowledge functioning as a common knowledge source, enabling proper categorisation of entities mentioned in the conversation even without explicit type information in the dialogue itself.

As more scenes are processed, the knowledge graph evolves and becomes richer with new information. Figure 4 shows the graph's evolution after processing the entire first episode, demonstrating a significant increase in complexity from 8 to 19 nodes. Particularly noteworthy is how the system captures subtle relationships, such as the connections between Sheldon's workspace ("Working Spot") and his research areas. The system establishes these connections with varying confidence scores (0.30 with String Theory and 0.35 with Quantum Mechanics), reflecting the nuanced nature of information extraction from dialogue. These weaker connections show the system's ability to make nuanced inferences while appropriately expressing uncertainty when working with incomplete information. Meanwhile, stronger connections emerge where evidence is more direct and explicit in Sheldon's own statements, such as his expertise in these fields (0.90 and 0.95 respectively).

To evaluate the proactive capabilities of the proposed framework, we analysed the system's behaviour in the same evolved graph. The framework detected a significant trigger in the food-related domain. By design, our proposed methodology activates proactive behaviour when at least two branches of the knowledge graph related to the same entity reach high weight values. This occurred following two specific dialogue segments:

f1: "We can't have Thai food, we had Indian for lunch."

f2: "They're both curry based cuisines"

These statements led to the creation of high-weight connections in the food domain (i.e., 0.91 out of 1), as shown in Figure 4. Upon detecting a trigger in the knowledge graph, such as strongly weighted connections between topics or recurring patterns of interest, the Insights Utility Evaluation module initiates an analysis process. First, the LLM is fed with relevant contextual information, including the trigger-related sentences. As a result, the model generates personalized suggestions which are evaluated using the Intersection over Union (IoU) metric to measure how well a new suggestion would integrate with the existing knowledge graph. The IoU score measures the overlap between the entities and relationships that would be introduced by the new suggestion and those already present in the graph. In our implementation, we set an IoU threshold of 80%, meaning that suggestions are considered valid only if they share significant commonality with the existing knowledge while also introducing meaningful new connections. While multiple suggestions were generated during the episode processing, only those exceeding this IoU threshold were considered valid.

In this case, the LLM generated the following suggestion:

f3: "Given the rejection of curry-based cuisines, might I suggest either sushi or Italian food"

Remarkably, this suggestion's validity was confirmed later in the episode when Wolowitz, Sheldon's friend, suggests visiting a sushi bar showing the system's ability to generate contextually appropriate and practical recommendations.

Additionally, subjective evaluation metrics could provide valuable insights into user satisfaction and emotional responses to the system's proactive suggestions. These could include Quality of Experience (QoE) assessments using Absolute Category Rating (ACR) scores and the Self-Assessment Manikin (SAM) scale for evaluating emotional aspects of user interaction.

V. PRIVACY

Ensuring user privacy is a critical aspect of developing an HDT. While these technologies can greatly enhance user experiences through personalized and proactive assistance, they also pose risks related to handling sensitive information, especially in light of regulatory frameworks like the **General Data Protection Regulation (GDPR, EU 2016/679)** and **California Consumer Privacy Act (CCPA)**, which set strict standards for data handling and user consent. For example, an HDT might record a conversation where Bob seeks advice from his digital assistant on managing his financial struggles. This is sensitive information, and the system handles it discreetly. However, if, a few days later, the assistant were to publicly advise against an expensive outing, it would breach Bob's privacy by revealing sensitive details, potentially violating Article 5 of the GDPR, which emphasizes data minimization and purpose limitation. To mitigate such risks, the system must ensure that sensitive information is communicated privately, in compliance with legal standards like **Article 32 of the GDPR**, which mandates appropriate security measures to protect personal data. For instance, the assistant could send a discrete notification to Bob's smartphone rather than making a public statement. This preserves confidentiality while providing assistance, aligning with the principles of **data protection by design and by default** (Article 25, GDPR).

The framework could employ several strategies to ensure privacy:

- **Context-Aware Communication:** Assessing the situation's context and using private channels, like smartphone notifications, for sensitive information instead of public announcements, consistent with the **GDPR's principle of proportionality**.
- **User Consent and Control:** Allowing users to manage their privacy settings, such as opting for encrypted messages or silent notifications for sensitive content, in accordance with **Article 7 of the GDPR**, which requires that consent be freely given, specific, informed, and unambiguous.
- **Anonymisation and Data Minimisation:** Implementing techniques like "incognito mode" to anonymize data and minimize the information collected, reducing the risk

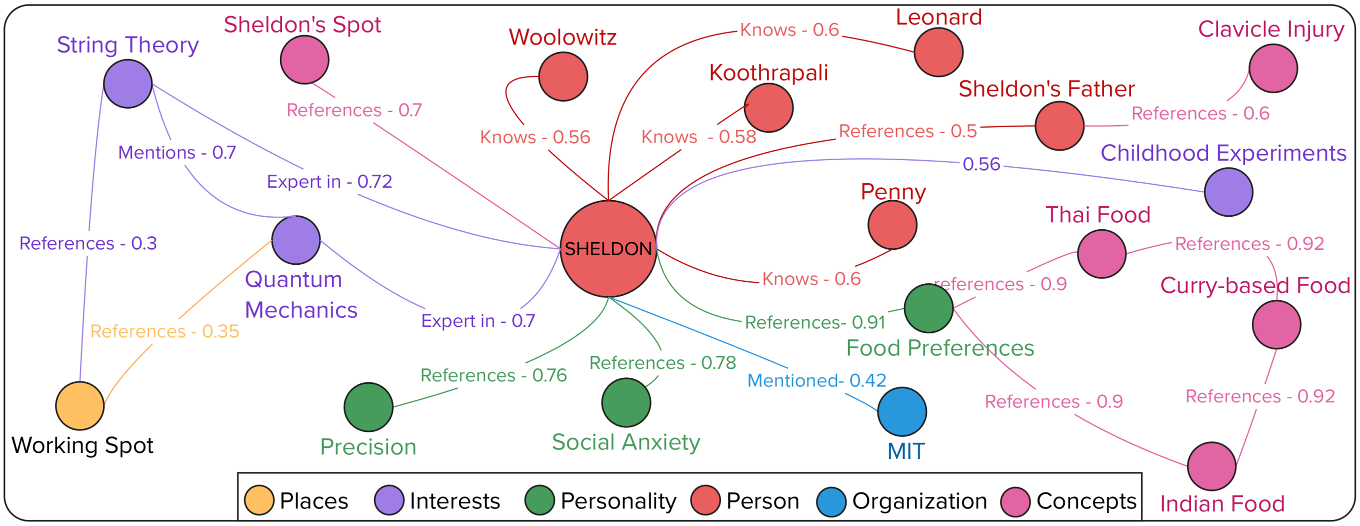


Figure 4. Evolved knowledge graph showing the enriched understanding of Sheldon's character after processing additional scenes.

of exposure and adhering to the **data minimization principle** (Article 5(1)(c) of the GDPR).

VI. Conclusions and Limitations

We presented a preliminary evaluation of user-centric IoT system, proposing a framework that leverages HDTs, PDAs, and LLMs to enhance IoT system proactivity. While we focused on textual data analysis, the framework could be extended to incorporate structured data from IoT sensors, wearables, and smart devices, providing a more comprehensive understanding of user behaviour and context. As this work represents an initial proposal of our framework, future research should focus on developing diverse datasets specifically designed for evaluating knowledge graph construction and proactive recommendation. Such datasets would enable more comprehensive evaluation across multiple domains and user profiles, allowing researchers to assess both the accuracy of entity-relationship extraction and the relevance of proactive suggestions generated by the system. Additionally, investigating scalability considerations for centralised implementations of the framework would be valuable, particularly in scenarios where knowledge sharing across multiple users might enhance the overall system performance. User experience studies would also provide insights into how proactive suggestions are received and perceived in real-world environments.

Acknowledgment

This work has been partially funded by the Ministero dell'Istruzione, dell'Universita e della Ricerca (MIUR) with the PON "Ricerca e Innovazione" 2014-2020 (PON RI) "Azione IV.5 Dottorati su tematiche green", assigned with D.M. 1062 on 10.08.2021 and by the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 1409 published on 14.9.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union – NextGenerationEU– Project Title "METATwin - Metaverse & Human Digital Twin: digital identity, Biometrics and Privacy in the future virtual worlds", – CUP J53D23015030001- Grant Assignment Decree No. 0001382 adopted on 01/09/2023 by the Italian Ministry of University and Research (MUR).

References

[1] J. Navarro-Alamán *et al.*, "Emotiot: An iot system to improve users' wellbeing," *Applied Sciences*, vol. 12, no. 12, 2022, ISSN: 2076-3417. DOI: 10 . 3390 / app12125804.

[2] L. Alkhariji *et al.*, "Semantics-based privacy by design for internet of things applications," *Future Generation Computer Systems*, vol. 138, pp. 280–295, 2023, ISSN: 0167-739X. DOI: 10.1016/j.future.2022.08.013.

[3] R. Minerva, G. M. Lee, and N. Crespi, "Digital twin in the iot context: A survey on technical features, scenarios, and architectural models," *Proceedings of the IEEE*, vol. 108, no. 10, pp. 1785–1824, 2020.

[4] D. Flores-Martin *et al.*, "Sharing knowledge to promote proactive multi-environments in the wot," *Journal of Web Engineering*, vol. 22, no. 2, pp. 327–356, 2023. DOI: 10.13052/jwe1540-9589.2226.

[5] D. Thakur, J. K. Saini, and S. Srinivasan, "Deep-think iot: The strength of deep learning in internet of things," *Artificial Intelligence Review*, vol. 56, no. 12, pp. 14 663–14 730, 2023. DOI: 10.1007/s10462-023-10513-4.

[6] J. Wen *et al.*, "From generative ai to generative internet of things: Fundamentals, framework, and outlooks," *IEEE Internet of Things Magazine*, vol. 7, no. 3, pp. 30–37, 2024. DOI: 10.1109/IOTM.001.2300255.

[7] M. U. Islam and B. M. Chaudhry, "A framework to enhance user experience of older adults with speech-based intelligent personal assistants," *IEEE Access*, vol. 11, pp. 16 683–16 699, 2023. DOI: 10.1109/ACCESS.2022.3230151.

[8] J. Yu *et al.*, "Bi-directional digital twin and edge computing in the metaverse," *IEEE Internet of Things Magazine*, vol. 7, no. 3, pp. 106–112, 2024. DOI: 10.1109/IOTM.001.2300173.

[9] B. R. Barricelli *et al.*, "Human digital twin for fitness management," *IEEE Access*, vol. 8, pp. 26 637–26 664, 2020. DOI: 10.1109/ACCESS.2020.2971576.

[10] B. Wang *et al.*, "Human digital twin in the context of industry 5.0," *Robotics and Computer-Integrated Manufacturing*, vol. 85, p. 102 626, 2024, ISSN: 0736-5845. DOI: 10.1016/j.rcim.2023.102626.

[11] R. Zhong *et al.*, "Construction of human digital twin model based on multimodal data and its application in locomotion mode identification," *Chinese Journal of Mechanical Engineering*, vol. 36, no. 1, p. 126, 2023. DOI: 10.1186/s10033-023-00951-0.

[12] Y. Lin *et al.*, "Towards efficient multi-domain knowledge fusion adaptation via low-rank reparameterization and noisy label learning: Application to source-free cross-domain fault diagnosis in iiot," *IEEE Internet of Things Journal*, pp. 1–1, 2024. DOI: 10.1109/JIOT.2024.3471410.

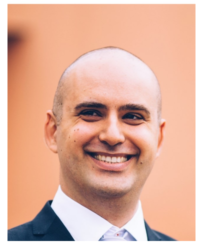
[13] K. Balog and T. Kenter, "Personal knowledge graphs: A research agenda," in *Proceedings of the ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR)*, 2019.

[14] A. G. Vrahatis, K. Lazaros, and S. Kotsiantis, "Graph attention networks: A comprehensive review of methods and applications," *Future Internet*, vol. 16, no. 9, 2024, ISSN: 1999-5903. DOI: 10.3390/fi16090318.

[15] G. Feretzakis *et al.*, "Privacy-preserving techniques in generative ai and large language models: A narrative review," *Information*, vol. 15, no. 11, 2024, ISSN: 2078-2489. DOI: 10.3390/info15110697.

Luigi Serreli holds a B.Sc. and M.Sc. (honours) in Electronic and Computer Engineering from the University of Cagliari. His Ph.D. research focuses on IoT and NLP with an emphasis on Generative LLMs and deep learning. He also works on Digital Twin technologies for IoT and Service Discovery.

Massimo Farina Ph.D. is an Associate Professor of Legal Computer Science and Data Protection Officer at the University of Cagliari. He is, also the Director of the university laboratory "ICT4Law & Forensics" set up at the Department of Electrical and Electronic Engineering, University of Cagliari.



Michele Nitti Michele Nitti is an Assistant Professor at the University of Cagliari, Italy since 2015. Currently, he is a member of the editorial board for the IEEE IoT Journal, Elsevier Computer Networks and MDPI IoT. His main research interests are in architecture and services for the Internet of Things (IoT).