robotics

MDPI

Article

The Art of Replication: Lifelike Avatars with Personalized Conversational Style

Michele Nasser [1,*], Giuseppe Fulvio Gaglio [1], Valeria Seidita [1,2] and Antonio Chella [1,2]

[1] Department of Engineering, University of Palermo, 90128 Palermo, Italy; giuseppefulvio.gaglio@unipa.it (G.F.G.); valeria.seidita@unipa.it (V.S.); antonio.chella@unipa.it (A.C.)
[2] High Perfomance Computing and Networking Institute (ICAR), National Research Council (CNR), 90146 Palermo, Italy
* Correspondence: michele.nasser@community.unipa.it

Abstract: This study presents an approach for developing digital avatars replicating individuals' physical characteristics and communicative style, contributing to research on virtual interactions in the metaverse. The proposed method integrates large language models (LLMs) with 3D avatar creation techniques, using what we call the Tree of Style (ToS) methodology to generate stylistically consistent and contextually appropriate responses. Linguistic analysis and personalized voice synthesis enhance conversational and auditory realism. The results suggest that ToS offers a practical alternative to fine-tuning for creating stylistically accurate responses while maintaining efficiency. This study outlines potential applications and acknowledges the need for further work on adaptability and ethical considerations.

Keywords: large language models; avatar; metaverse

Academic Editors: Stefania Costantini, Pierangelo Dell'Acqua, Giovanni De Gasperis and Francesco Gullo

Received: 14 January 2025
Revised: 24 February 2025
Accepted: 11 March 2025
Published: 13 March 2025

Citation: Nasser, M.; Gaglio, G.F.; Seidita, V.; Chella, A. The Art of Replication: Lifelike Avatars with Personalized Conversational Style. *Robotics* **2025**, *14*, 33. https://doi.org/10.3390/robotics14030033

1. Introduction

The metaverse is rapidly emerging as an interconnected virtual space where individuals can engage in immersive experiences, pursue professional endeavors, socialize, learn, and create [1]. This digital environment promises to redefine our relationship with technology, offering new possibilities for interaction and personal representation. The pandemic of the novel coronavirus disease 2019 (COVID-19) has further accelerated the transition to digital solutions for communicating and collaborating, underscoring the potential of virtuality to overcome physical barriers and expand the possibilities for social and economic interaction [2,3]. In this context, the concept of virtual alter egos has gained considerable relevance. Some applications facilitate the creation of autonomous avatars based on images and videos of real people, enabling them to act independently in virtual spaces [4]. These digital alters are regarded as a pivotal component in the forthcoming Internet revolution, potentially influencing social interactions, business models, and lifestyle changes [5]. They facilitate users' exploration of alternative identities and novel experiences in virtual spaces, while interactive installations investigating the concept of digital alter egos offer insights into emotions, facial expressions, and self-perception [6].

The recent advancements in large language models (LLMs) have facilitated the integration of sophisticated artificial intelligence into virtual avatars and extended reality (XR) environments. These AI agents can generate contextual responses accompanied by coherent facial expressions and gestures [7]. Despite the persistence of technical challenges associated with multimodal integration and generation speed, LLM-based avatars demonstrate the growing potential for natural and meaningful conversations [8]. Moreover,

LLM-powered virtual characters have already been effectively utilized in hybrid live events, enhancing the quality of discussions and interactions [9]. These developments indicate a promising future for realistic and intelligent avatars in XR environments while necessitating further efforts to address issues such as privacy protection.

The present work is situated within this innovative scenario, wherein a visually realistic avatar is developed based on the image of a real person and enhanced by an LLM capable of replicating the communicative style of the original subject. This accomplishment is part of the interdisciplinary Research Projects of National Interest (PRIN) ALTEREGO, which investigates the frontiers of artificial intelligence to develop avatars capable of facilitating meaningful interactions in the metaverse. The ALTEREGO project aims to create advanced digital surrogates that represent not only an individual's appearance and behavior but also their intentionality and communicative style. This is achieved by applying concepts such as intentionality, vitality, theory of mind, and embodiment. Among the practical applications of the project are case studies dedicated to the representation of digital influencers and the development of prototypes to test the feasibility of these technologies. ALTEREGO represents a significant advance in digital communication, marking a new era where intelligent and realistic avatars can facilitate enhanced human interactions in virtual environments.

The importance of these developments extends beyond virtual spaces into the domain of robotics, where analogous challenges exist to replicate human-like interaction and communication. The intersection between metaverse research and robotics is an increasingly relevant area of exploration [10]. In the domain of human–robot interaction (HRI), the use of digital avatars guided by artificial intelligence has emerged as a pivotal aspect in the evolution of sophisticated robotic systems [11,12]. The creation of realistic digital avatars with customized language styles, explored in this study, contributes to this field by offering insights into how LLMs can improve the communication capabilities of virtual and robotic agents.

This paper is structured as follows: Section 2 explores the state of the art in LLM-powered conversational avatars, language style generation, and 3D avatars creation; Section 3 shows the process of creating the 3D model and chatbot, with emphasis on replicating the conversational style of the real person; Sections 4 and 5, respectively, concern the presentation of the results obtained in the case study and their discussion; and Section 6 outlines conclusions and future developments.

2. Background

2.1. LLM-Powered Conversational Avatars

LLMs are transforming the landscape of human–machine interactions, facilitating more natural conversations and expanding their use in diverse domains. In self-reported data collection, LLMs are being utilized to develop chatbots that can effectively steer the flow of conversations, thereby enhancing both the user experience and the accuracy of the data collected [13]. In interactions with robots, these models are capable of simulating emotions in real-time, which increases the perception of human likeness and improves the emotional appropriateness of responses [14]. It is important to acknowledge that LLM-based systems exhibit significant differences from human cognition [15]. Indeed, the apparent intelligence of LLMs is often contingent upon the interlocutor's ability to formulate questions, which lends credence to the phenomenon known as the "inverse Turing test" [16]. For this reason, the deployment of LLMs in the development of conversational chatbots and avatars represents a rapidly evolving field of research. These models have transformed our understanding and ability to generate natural language, facilitating more fluid interactions and resembling those between humans [17]. For ex-

ample, Yamazaki et al. [8] developed an open-domain avatar chatbot that addresses the challenges of multimodal integration and rapid response generation. In a related development, Friedman et al. [18] delineated a blueprint for the engineering of LLM-based conversational recommendation systems, introducing sophisticated methodologies for the comprehension of user preferences and the orchestration of more efficacious dialogues. These developments illustrate how LLMs are propelling substantial advancements in the domain of human–machine communication.

In this paper, we aim to address this problem by developing a text generator that can emulate a person's language style in the presence of a limited dataset. Text generation in a specific language style, such as that of a person, represents a relatively unexplored area of research. Existing approaches for LLMs to achieve this include the following:

- **Fine-tuning**, which requires a substantial quantity of data and substantial computational resources, is often impractical in settings where resources are constrained.
- **Style transfer** with standard prompting techniques, as exemplified by studies such as [19–21], prioritizes general stylistic transformations over precise customization, aiming to emulate individual styles.
- Approaches based on **style embedding**, as exemplified by [22], which endeavor to represent the stylistic characteristics of a sentence in numerical vectors, are effective for quantifying style but require further development for broader practical application.

As can be seen from what is listed above, those methods have limitations that necessitate the development of an alternative methodology for leveraging the capabilities of LLMs in the generation of stylistically coherent and personalized texts, eliminating the costs and complexities associated with fine-tuning. Among the most relevant work in this area, the project described in [21] has demonstrated the effectiveness of prompting for stylizing texts in complex linguistic contexts, such as articles in Chinese. Furthermore, recent studies have highlighted how style identification and application can be handled separately from content, which is a key feature in making LLMs versatile and customizable tools.

2.2. 3D Avatar Creation Techniques

In recent years, a substantial body of research has been conducted to develop novel techniques for generating realistic digital avatars from limited input data. A pipeline proposed in [23] generates expressive avatars from a single image, using deep learning techniques to enhance texture quality and eye rendering. In contrast, in [24] a method to create 3D avatars from frontal RGB images by identifying pose, shape, and semantic details is described.

Among the most rapid solutions is Instant Volumetric Head Avatars (INSTAs), an approach that reconstructs photorealistic avatars in less than 10 min by exploiting neural radiance fields in conjunction with a parametric face model trained on RGB monocular video [25]. These advances are intended to enhance the automation, identification, and usability of avatars. Additionally, in [26] a comprehensive analysis of the evolution of facial capture and tracking techniques, examining processes such as data collection, facial encoding, asset creation, tracking, and rendering is conducted.

In [27] is shown a two-tier neural rendering system that enables the rapid creation of head avatars from a single photograph, thereby significantly increasing the speed of inference. Similarly, in [28] a method to generate 3D avatar heads via short smartphone shots, based on a universal avatar model trained with high-resolution multi-view video data, is proposed.

Another noteworthy contribution is Personalized Implicit Neural Avatar (PINA), which is a technique for the creation of personalized implicit neural avatars capable of representing realistic clothing deformations derived from RGB-D video sequences [29]. In [30] AvatarMe, an approach for reconstructing photorealistic 3-D faces in 4 K–6 K

resolution using "in-the-wild" images is presented. This method successfully addresses the challenges of data sparsity and high-resolution processing.

These methods represent a significant advancement in the creation of increasingly realistic and accessible digital avatars, reducing the input requirements and expanding the possibilities for virtual human representations.

3. Materials and Methods

3.1. Avatar Creation

To create a realistic three-dimensional avatar of the subject, we employed methods and applications that exploit the Structure From Motion algorithm [31], according to the methodologies proposed in the work cited above. This algorithm reconstructs a three-dimensional model of a scene or object from a series of two-dimensional images by identifying homologous points in the various photos and calculating their spatial location through triangulation. The workflow described below can be seen in Figure 1. For image capture, the Polycam application was selected, as it enables the user to capture photographs via the smartphone camera and subsequently process them on the application's servers, thereby generating a three-dimensional model. However, to enhance the resolution and quality of the model, a Reflex D3400 Nikon camera was employed to capture the images, which were then manually uploaded to the app. The generated scene required editing and cropping to isolate the object of interest, focusing exclusively on the upper body (head and shoulders). For optimal results in identifying homologous points, diffuse lighting that reduced marked shadows and contrasted less with the digital lighting of the virtual scene in which the avatar would be placed was essential.

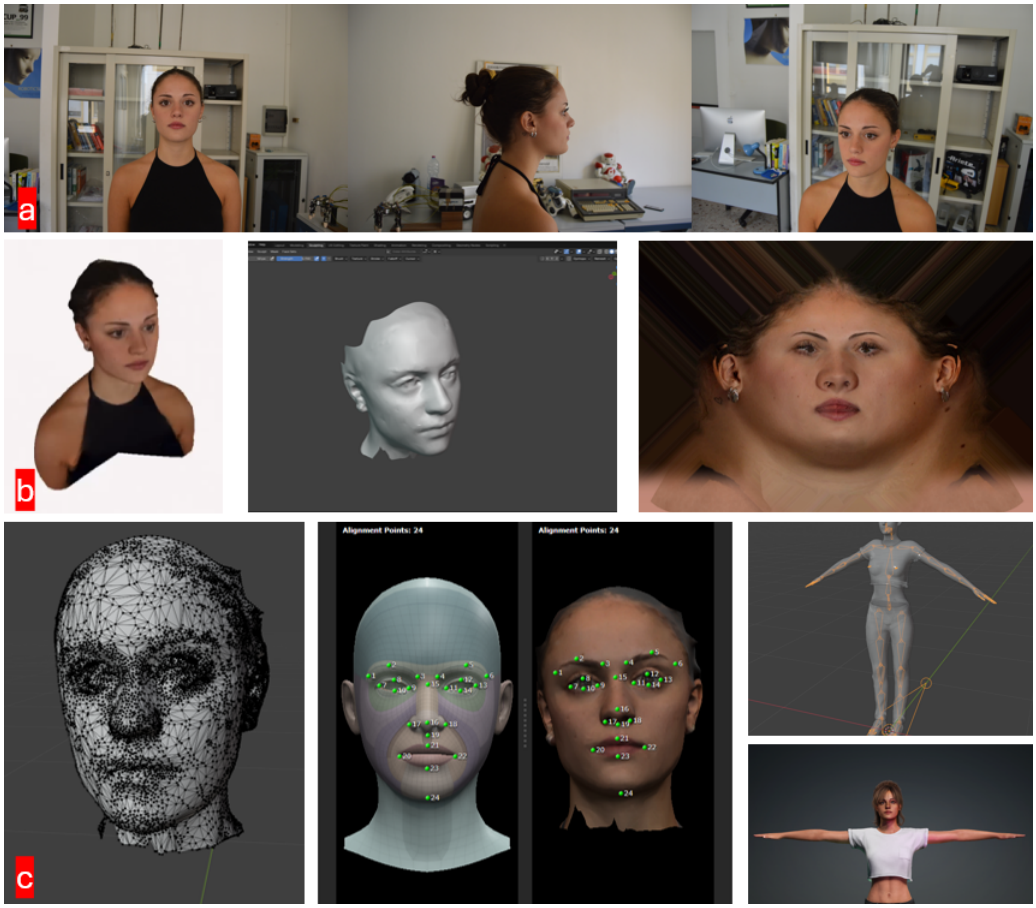

Figure 1. Avatar creation workflow. (**a**) Image acquisition; (**b**) first 3D mesh and texture generation; (**c**) automatic retopology and final avatar generation.

The initial 3D model, complete with textures, was imported into Blender software to correct any imperfections and improve the mesh surface through 3D sculpting operations. Subsequently, the mesh was utilized as the basis for the avatar creation in Character Creator 4 software via the Headshot plug-in. This tool facilitated the generation of a mesh with a neater, cleaner topology, thus enabling the simulation of facial expressions through the recognition of key points on the face and head. The body was replaced by a digital standard, thereby accelerating the process, while the head and shoulders were augmented with accurate textures derived from the original images. The resulting avatar, equipped with a Unity-compatible rig, is thus prepared for utilization in a digital environment (see Figure 2).

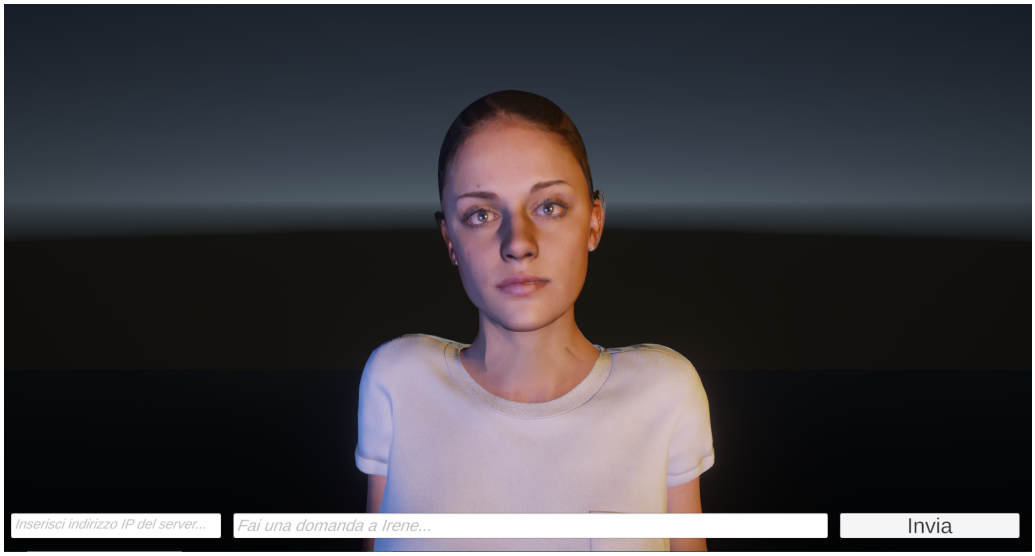

Figure 2. The final avatar in Unity Engine.

3.2. Conversational Module: Text Generation and Speech Synthesis

In light of the recent developments outlined in the background, we propose an innovative methodology that combines a preliminary sentence analysis based on existing natural language processing (NLP) techniques with a new prompting technique, named Tree of Style (ToS) (see Figure 3) by our research team. This approach aims to overcome the limitations of current methods by proposing a conversational module for an avatar capable of emulating an individual linguistic style, responding at appropriate times for a conversation, and utilizing a reduced dataset, i.e., low dimensionality. Unlike other techniques, our method does not involve training a large language model (LLM), but focuses on training a pre-trained model to adopt the target person's language style.

3.2.1. Datasets

To create a personalized voice avatar, two distinct datasets were collected and utilized for different purposes:

- D_1: The dataset employed to calculate the average style embedding (see Figure 3) is divided into two segments: one is utilized to calculate the average style embedding, while the other is employed to conduct evaluations. It is imperative to acknowledge that the second segment does not contribute to the calculation of the average style embedding, with the objective of avoiding any influence on this vector. The average style embedding is then employed as a discriminator of the language style applied by a given model, such as GPT, to the sentences it generates.
- D_2: The dataset under consideration contains audio recordings accompanied by their respective transcripts. These recordings and transcripts are used to replicate the

timbre, pitch, and vocal characteristics of the individual whose voice is the target of the reproduction. This dataset is used for fine-tuning the PiperTTS speech synthesis model, with the objective of obtaining an accurate reproduction of the individual's voice.

The D_1 dataset is comprised of 115 sentences that have been meticulously selected to exemplify an individual's linguistic style. This number has been determined through preliminary laboratory tests conducted on a sample of individuals, to optimize the balance between the size of the dataset and the fidelity of style replication. The objective of this research is to develop a methodology capable of replicating language style without the necessity of fine-tuning. Consequently, the D1 dataset was intentionally constrained in size to assess the efficacy of the proposed methodology, even under conditions of limited resources. The selection process of the D1 dataset was guided by sociolinguistic principles to include lexical and syntactic idiosyncrasies, pragmatic nuances, and characteristic communicative registers. The decision to focus on personalized language style is based on established studies in sociolinguistics, which emphasize that communication is strongly influenced by the following extralinguistic parameters [32–35]:

- **Diamesia**: the selection of a communication channel (in this case, exclusively oral) to ensure greater consistency with speech. Diamesic variation represents a significant aspect of linguistic variation.
- **Diaphasia**: the modulation of communicative style according to context and relationship with the interlocutor. This parameter was pivotal in the construction of the dataset, as it reflects how individuals adapt their linguistic style contingent on the context and the relationship with the interlocutor (e.g., informal register with friends or formal register in professional contexts).
- **Diatopia and Diastratia**: Extreme dialectal or social variations were not included in the analysis, as the objective was to maintain a standard and stable register in line with the profile of the imitated subject.
- **Diachrony**: Temporal variations were not considered, as the pattern reflects the individual's current style.

To construct the dataset, sentences belonging to diverse communicative situations (e.g., family, friendship, professional) were selected, covering a wide range of contexts relevant to the individual in question. Additionally, the subject was requested to record voice messages instead of written text, thereby engaging the sensory faculties inherent to speech. This methodology enables the accurate capture of the subject's natural linguistic style, encompassing its nuances. A particular focus on the diaphasic parameter proved especially efficacious, as the selection of language variety is largely contingent upon contextual and interpersonal factors. To illustrate, a subject might utter the following:

- "Sit down, come on, make yourself comfortable!" in a familiar context;
- "Please, ma'am, have a seat" in a formal context.

The classification of the dataset according to contexts of use enabled the model to learn how to modulate the language register in an appropriate manner, thereby replicating the individual's natural behavior and reinforcing the impression of authenticity.

The second dataset, D_2, comprises approximately 800 recordings of up to 10 s in duration. These data were employed to refine a pre-trained model in English. Despite the base model being trained on British voices, fine-tuning with Italian recordings yielded a remarkably high-quality result. The resulting voice was found to be natural and consistent with the subject's original voice. The absence of a British accent indicated that the adaptation to the new Italian recordings was indeed effective.

3.2.2. Preprocessing and Sentences Analysis

The primary objective of this phase is to develop the Prompt System P_S in such a manner that it provides detailed information about an individual's language style. This will enable the conversational module to have a knowledge base on the style being analyzed, ensuring responses consistent with the register of the relevant communicative context (e.g., family context). In summary, the Prompt System, when applied to a language model (LLM), will enable the LLM to adapt its expressive style to be as similar as possible to that of the individual in a given communicative context. The system prompt is defined as in Figure 4.

Task description:

You are a person named $*name*, $*gender*, $*years*.
Follow the grammatical features described in the "Language Style" section.
Provide information from the "Knowledge Base" section only when asked a question about it.
Use simple language: speak clearly and use an average sentence length of $length words with a standard deviation of $deviation words.
Maintain a natural tone: write as if you were speaking normally.

Linguistic style:

Distribution of 1-grams POS: VERB: $*verb*; PUNCT: $*punct*; PRON: $*pron*; ADV: $*adv*; NOUN: $*noun*; ADP: $*adp*; DET: $*det*; AUX: $*aux*; CCON: $*ccon*; INTJ: $*intj*;
Distribution of 2-grams POS: PRON VERB: $*prn_vrb*; NOUN PUNCT: $*nn_pnct*; VERB ADV: $*vrb_adv*; VERB PUNCT: $*vrb_pnct*; DET NOUN: $*det_nn*; ADV PUNCT: $*adv_pnct*; ADV PRON: $*adv_prn*; PUNCT ADV: $*pnct_adv*; AUX VERB: $*aux_vrb*; VERB DET: $*vrb_det*;
Distribution of 3-grams POS: ...
Distribution of 4-grams POS: ...
Distribution of 5-grams POS: ...
Vocabulary: Types of Words(fundamental: $*fund*; high usage: $*hh*; high availability: $*ha*; non-categorized: $*nc*)
Sentence description: the sentences come mainly from the following context: $*context*
Emotions: joy: $*joy_prc*; sadness: $*sadness_prc*; fear: $*fear_prc*; anger: $*anger_prc*; love: $*love_prc*; surprise: $*suprise_prc*;

Knowledge Base: $*knowledge_base*

Figure 3. System prompt. It consists of three main parts: (1) Task Description provides initial guidance to the LLM; (2) Language Style provides information about the parts of speech obtained from the analysis of style, typical words, emotions conveyed, etc.; and (3) Knowledge Base contains information about the history and personality of the individual. The template is designed to be easily repurposed on another person.

The ultimate goal is to develop a conversational system capable of responding appropriately to the context, modulating the linguistic register accordingly. To simplify the evaluation of the adopted methodologies, the focus was initially on a single communicative context: the family context.

The D_1 dataset was divided according to the communicative situations in which it was collected, including family, friendship, and professional contexts. This classification facilitated the preprocessing and analysis of sentences concerning their communicative context, to extract specific information that reflected the individual's linguistic style in different scenarios. This approach is based on the previously mentioned sociolinguistic studies, which highlight the importance of context and interpersonal relationships in shaping language.

The analysis of sentences in the dataset was conducted considering the following aspects:

- **Distribution of lemmas**: identification of the most frequently used words and their contextual variation.
- **Distribution of POS n-grams**: analysis of grammatical tag sequences (part of speech) with n varying from 1 to 5, to capture typical syntactic patterns.

- **Sentence length distribution**: calculation of mean length and standard deviation to understand average speech structure.
- **Sentiment analysis**: assessment of the prevailing emotional tone in sentences, useful for stylistic personalization.
- **Distribution according to [36]**: analysis of sentences according to specific lexical categories, i.e., "fundamental words", "words of high usage", and "words of high availability".

To substantiate this description, an exemplar of the P_S System Prompt (Figure 4) has been provided, illustrating the results of the analysis in the familiar context.

Descrizione compito: Impersona Irene, 20 anni. Rispetta le caratteristiche grammaticali descritte nella sezione 'Stile linguistico'. Fornisci le informazioni contenute nella sezione 'Base di Conoscenza' solo quando ti viene fatta una domanda a riguardo. Usa un linguaggio semplice: parla chiaramente e utilizza una lunghezza media di frase di 12 parole con deviazione standard di 5 parole. Mantieni un tono naturale: scrivi come se parlassi normalmente.

Stile linguistico:

Distribuzione dei 1-grammi POS: VERB: 20.28% PUNCT: 17.68% PRON: 14.06% ADV: 13.54% NOUN: 9.45% ADP: 5.13% DET: 4.61% AUX: 3.51% CCONJ: 2.59% INTJ: 2.53%

Distribuzione dei 2-grammi POS: (PRON VERB): 7.04% (NOUN PUNCT): 5.56% (VERB ADV): 5.19% (VERB PUNCT): 5.06% (DET NOUN): 4.01% (ADV PUNCT): 3.83% (ADV PRON): 3.15% (PUNCT ADV): 3.15% (AUX VERB): 2.72% (VERB DET): 2.72%

Distribuzione dei 3-grammi POS: ....

Distribuzione dei 4-grammi POS: ....

Distribuzione dei 5-grammi POS: ....

Vocabolario: Tipi di Parole: {'fondamentale': 76.3%, 'alto uso': 0.9%, 'alta disponibilità': 1.5%, 'non categorizzate': 20.3%}

Descrizione frasi: Le frasi provengono principalmente da conversazioni familiari o informali. Le frasi sono orientate verso azioni quotidiane (es. "andare", "venire", "preparare"), richieste (es. "dimmi", "fammi", "prendere"), e interazioni comuni (es. "mamma", "papà", "nonno"). Potrebbero riflettere un ambiente familiare affettuoso, con un linguaggio informale e diretto. L'uso di parole straniere indica che si tratta di conversazioni informali in cui le persone si esprimono rapidamente.

Emozioni: [joy: 26%, sadness: 12%, fear: 10%, anger: 4%, love: 44%, surprise: 4%]

Base di conoscenza: Spettacolo: A Irene piace il mondo dello spettacolo. Studia recitazione, canto e danza. Fa diversi provini in vari casting per ricoprire la figura di comparsa parlante.

Figure 4. Study case prompt. The figure displays a prompt in Italian, which remains faithful to the original. The prompt is composed of three primary sections: Task Description, which assigns the chatbot the identity of Irene, a 20-year-old individual, through the utilization of simple, natural language; Language Style, which defines grammatical distributions (e.g., VERB: 20.28%) and predominantly informal vocabulary; and Knowledge Base, which describes Irene as an entertainment enthusiast, with details provided only upon request, the latter of which is expandable.

This analysis procedure constitutes a pivotal component of the project, as it will serve as the foundation for evaluating the virtual alter ego in the evaluation phase. In this phase, the linguistic distributions extracted from the individual's real sentences will be contrasted with those derived from the sentences generated by the conversational model. This will facilitate the assessment of stylistic consistency and fidelity to real speech. This comparison will also be instrumental in gauging the model's efficacy in replicating the target individual's linguistic style.

3.2.3. Tree of Style (ToS)

The main contribution of this paper is the creation of what we called Tree of Style (ToS), which is a process for guiding any large language model (LLM) in assuming a specific language style. It is important to note that no model training will be performed at this stage, as the goal of the article is not to apply fine-tuning techniques, but to develop a new prompting methodology. The latter aims to model the linguistic style of an LLM, thereby allowing the model to consistently emulate an individual's style in a given communicative context, without the need to change its structure or basic training. To guarantee precise and uniform emulation of linguistic style, a methodology derived from the Tree of Thoughts (ToT) technique [37] was tailored to the project context. The methodology's fundamental aspect is the utilization of style embeddings, which are vector representations that capture the stylistic characteristics of a text in a manner that is independent of its semantic content.

As detailed in [22], these embeddings are crafted to isolate the stylistic elements of language. The style embedding model employs a RoBERTa-based architecture [38], facilitating multilingual support and substantiating the assertion that stylistic attributes are not contingent on sentence content. To ascertain the stylistic similarity between two sentences, it is first necessary to calculate the respective style embeddings and then to compare them using cosine similarity.

In this project, the model [22] was employed to compute the style embedding of each sentence within the D1 dataset. The D1 dataset consists of sentences produced by an individual in a given communicative context, such as the family context. Following the computation of the style embedding for each sentence, the average of the vectors was calculated to obtain the average style embedding:

$$E_{mean} = \frac{1}{|D_1|} \sum_{x \in D_1} Style\_embedding(x) \quad (1)$$

- $|D_1|$ is the number of sentences in the dataset D_1;
- *Style_embedding(x)* is the embedding of the sentence x.

The average embedding E_{mean} represents the vector representation of the style of the individual to be imitated in a given communicative context. This was employed as a discriminator in the proposed architecture, functioning as a reference for evaluating the stylistic consistency of the generated responses.

The proposed architectural framework is structured in several stages (see Figure 5). The initial stage involves defining a system prompt, designated as P_S. This prompt encompasses a comprehensive description of the task, the stylistic attributes to be employed, and a preexisting knowledge base. This prompt serves to direct the model, such as GPT or LLaMA, to generate responses that align with the stylistic and semantic characteristics of the individual to be emulated.

For each query, the system generates 10 independent responses, thereby capitalizing on the inherently stochastic nature of LLMs. Indeed, these models are inherently indeterministic. When presented with the same input, they produce responses that are semantically equivalent but stylistically different due to variability in syntactic and lexical choices. This behavior is not a limitation but an advantage, as it allows us to obtain a plurality of sentences that maintain content but differ in style. In other words, we can obtain a variety of responses that reflect different stylistic nuances while still maintaining communicative coherence.

In contrast to methodologies necessitating continuous feedback to the model for stylistic correction, this approach enables parallelization of the decision-making process, facilitating rapid generation of responses in accordance with the timing of a natural conversation. Had iterative corrections to the model been required, the process would have been considerably slower, potentially compromising the fluidity and naturalness of the conversation.

Following the generation of ten responses, which are characterized by equivalent semantic content but divergent syntactic and stylistic features, the response that demonstrates the most sophisticated stylistic technique is selected. That is to say, given the absence of variation in semantic content, it is the stylistic features of the response that undergo change. As illustrated in Figure 5, the relative stylistic embedding E_x is calculated for each generated response and then compared with the average embedding E_{mean} via cosine similarity.

$$E_x = Style\_embedding(x) \quad (2)$$

$$x_{selected} = argmax_x(sim\_cos(E_x, E_{mean})) \quad (3)$$

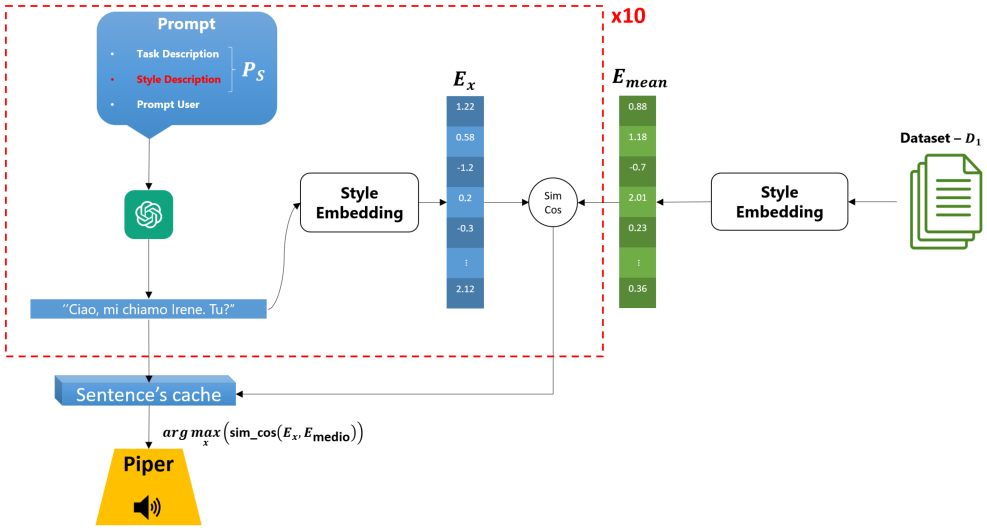

Figure 5. ToS Architecture: The process of stylistic selection of generated responses involves the calculation of the relative style embedding E_x for each sentence produced by the conversational model. This is followed by a comparison of the embedding of each sentence with the average style embedding E_{mean} obtained from the reference dataset. The similarity between the two vectors is measured by cosine similarity, and the sentence with the highest similarity is selected as the final response.This process ensures that the generated response maintains the stylistic consistency of the individual, adapting the syntactic content but preserving the linguistic style.

The sentence that most closely resembles the average style embedding will be selected as the output response in order to facilitate the continuation of the conversation. This process ensures stylistic consistency, as each generated sentence is compared with the average style embedding, and the one that is most similar is selected in order to maintain consistency of language style.

This approach generates a Tree of Style, wherein each node represents a stylistically consistent response emulating the individual in question. The primary benefit of this structure is that, along the selected path, the responses are consistently aligned with the target E_{medium} style, ensuring stylistic consistency throughout the conversation. Furthermore, stylistic continuity enables the model to effectively adapt the tone of the conversation to the responses, enhancing the illusion of interacting with the original individual (see Figures 6 and 7).

3.2.4. Text-to-Speech (TTS)

We decided to use Piper TTS to create an alter ego that faithfully reproduces the voice of a specific person. After selecting the sentence that most closely resembles the style of the person to be imitated, Piper TTS is used to speak the text, generating speech synthesis that reflects the unique characteristics of the original voice. Piper TTS is an open-source speech synthesis system that generates realistic audio from text using the Variational-Inference-Text-to-Speech (VITS) architecture. This end-to-end model combines a generator and discriminator in a deep neural network to produce natural, fluid voices in real-time.

To personalize the avatar's voice, approximately 800 audio samples of the target individual were recorded and used to fine-tune a pre-trained model in English. This fine-tuning process allowed the system to learn the individual's unique vocal characteristics, such as frequency, pitch, and rhythm, ensuring that the avatar not only visually resembled the person represented, but also convincingly reproduced his or her voice.

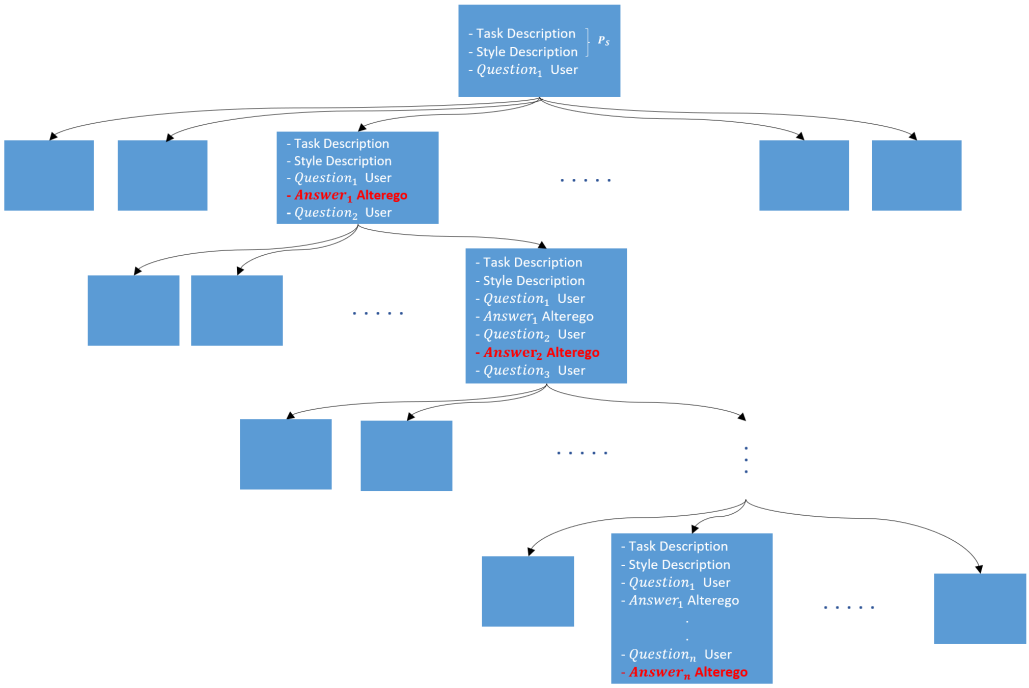

Figure 6. Tree of Style. The implementation of 10 distinct prompts for each inquiry is vital to investigate a comprehensive array of stylistic alternatives. This approach not only enhances stylistic diversity but also guarantees that each selected utterance contributes to the construction of a coherent and genuine dialogue while maintaining consistency with the original style.

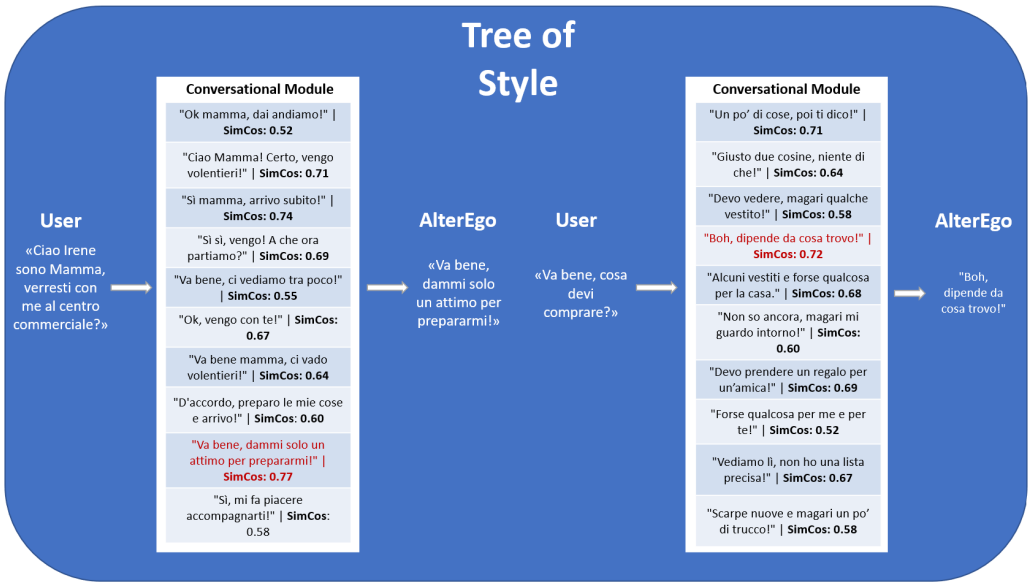

Figure 7. Tree of Style in practice. The image shows an example of a conversation in which the conversational module, shown in Figure 5, selects one among ten generated responses. The chosen response is the one with the highest cosine similarity score relative to the average style embedding. This response is then returned to the interlocutor. The entire question and answer process, repeated over multiple interactions, generates the Tree of Style (Figure 6).

The ability to faithfully clone an individual's voice is critical to avatar authenticity. As previously written, diaphasic variation, which refers to the adaptation of speech to the context and relationship between speakers, plays a key role in the perception of speech authenticity. In addition, paralinguistic elements such as tone, intonation, and rhythm contribute to the perception of credibility and effectiveness of communication. Accurately reproducing these features ensures that the speech style is perceived as authentic and consistent with the context, facilitating user identification and empathy.

We also chose Piper TTS because of its easy integration with Unity, which made implementation in our project efficient and quick. In addition, its versatility and ability to generate high-quality voices made it ideal for our needs in creating immersive interactive experiences in the metaverse.

4. Results

The results obtained seem promising and suggest that the virtual alter ego integrated into the metaverse has credible and functional characteristics. Indeed, the avatar visually mirrors the selected individual, providing a similarity that enhances the user's immersion. In addition, the response time of the conversation module is adequate, with no long pauses that might interrupt the natural flow of the conversation.

Concerning the conversation module, the stylized text generation convincingly approximates the communicative style of the represented individual, and the text-to-speech system reproduces the original user's voice with good fidelity.

Table 1 compares several techniques for generating the text: Zero-Shot, ToS, Prompt+ToS, and direct fine-tuning. The table includes different similarity measures to evaluate the stylistic and content-related correspondence between the generated sentences and the target individual's sentences. These measures include lexical, syntactic, and sentiment-based metrics, offering a comprehensive assessment of performance. Analysis of the data showed that the proposed method, Prompt+ToS, achieved comparable or better performance than fine-tuning, a technique often considered more complex and burdensome to implement. In particular, the values highlighted in red indicate the highest similarity scores for each metric. It is noteworthy that Prompt+ToS achieves the highest values in some cases and remains close to fine-tuning in others, whereas Zero-Shot and ToS methods generally yield lower similarity scores. For instance, in the Sim-js sentiment metric, Prompt+ToS with gpt-4o-mini reaches 0.857, which is very close to the highest value (0.939) obtained with fine-tuning. Similarly, in the Sim-cos Content metric, Prompt+ToS scores the highest value (0.642), surpassing all other approaches.

An important point to mention is that the values shown in the table represent the average of 10 tests performed with the same input. This methodological choice was made because large language models (LLMs) are inherently indeterministic, meaning that responses may vary slightly even with the same initial conditions. Averaging the results over multiple iterations provides a more stable and reliable estimate of system performance. Overall, these findings suggest that Prompt+ToS represents a viable and cost-effective alternative to fine-tuning for implementing the conversational module of virtual alter egos. It maintains a high level of quality in text generation while achieving stylistic and semantic similarity close to, or even exceeding, that of fine-tuning in certain aspects.

Table 1. Style and Content Testing. The table shows the similarity results between the sentences generated by the two LLMs in the three different prompting techniques and fine-tuning with the target individual's sentences collected to form the training dataset. The values highlighted in red represent the highest values obtained. It should be noted that, in the Prompt+ToS method, the values obtained in some cases exceed those obtained by the fine-tuning method. However, in other cases, while not exceeding them, they are the closest, compared to the other Zero-Shot and ToS methods, to the red-highlighted values of the fine-tuning method.

Technique	Model	Sim-js NVdb	Sim-js Lemmas	Sim-js 1-POS	Sim-js 2-POS	Sim-js 3-POS	Sim-js 4-POS	Sim-js 5-POS	Sim-js sent_len	Sim-js Sentiment	Sim-cos Content
Fine-tuning	gpt-4o-mini	0.955	0.372	0.873	0.730	0.521	0.347	0.241	0.825	0.939	0.604
Prompt + ToS	llama-3.3-70b-versatile	0.942	0.375	0.853	0.702	0.519	0.345	0.238	0.798	0.894	0.626
	gpt-4o-mini	0.937	0.387	0.864	0.720	0.534	0.352	0.231	0.835	0.857	0.642
ToS	llama-3.3-70b-versatile	0.935	0.357	0.810	0.667	0.493	0.322	0.223	0.852	0.832	0.580
	gpt-4o-mini	0.930	0.357	0.843	0.708	0.533	0.341	0.230	0.836	0.806	0.635
Zero-Shot	llama-3.3-70b-versatile	0.923	0.340	0.798	0.660	0.497	0.325	0.224	0.805	0.814	0.594
	gpt-4o-mini	0.906	0.331	0.804	0.672	0.505	0.330	0.226	0.811	0.788	0.603

5. Discussion

This study constitutes a preparatory step for the subsequent development of conversational models that reproduce a person's unique communicative style. The primary objective of the study was to test and refine the methodology, which will then be applied to another subject in more advanced stages of the project. Because of this preparatory nature, a limited dataset and number of LLMs were used to establish the basic aspects of style adaptation and response generation. Subsequent phases will involve expanding the dataset and refining the model for broader applications.

In general, when creating an LLM that can replicate a particular communicative style, most evaluation metrics focus on Text Style Transfer (TST), measuring concepts such as Style Transfer Accuracy, Content Preservation, and Fluency [39]. These tools analyze the ability of a model to transform the style of a text while preserving its content and fluency. However, this approach has a significant limitation in our context: it evaluates the stylistic transformation of a text, but not the ability of the model to adopt a particular style.

Conventional metrics such as BLEU and ROUGE [40,41] were not utilized due to their inadequacy in measuring style. These metrics, which are widely used to evaluate the performance of LLMs, focus primarily on overlapping words and n-grams, but do not capture the more complex nuances of style and emotion that our alter ego seeks to model. Even a metric like METEOR [42], while balancing precision and recall through the use of synonyms and rephrasing, is not specifically designed to measure style.

In light of the aforementioned limitations, a decision was made to adopt an alternative approach, centering on the assessment of the similarity of both style and content to that of the individual. The primary emphasis was placed on the evaluation of style, with the similarity of content being considered in terms of its completeness. This approach entails the utilization of metrics designed to analyze model-generated style and content preservation, without a focus on style transfer. Additionally, the assessment of fluency was omitted, given that LLMs have already demonstrated a high level of fluency.

To conduct the analysis, we asked the target individual to answer a specific set of questions. The same set of questions was then used to generate responses from his alter ego. The results obtained were compared to assess the extent to which the alter ego could replicate the behavior and style of the real individual. The analysis was conducted according to two main aspects:

- **Similarity of Style**: The evaluation of style was conducted through a comprehensive analysis of the distributions obtained in the preprocessing phase of the responses. This analysis encompassed various aspects, including lemma distribution, POS n-gram distribution, sentence length, sentiment analysis, and NVdb-based distribution. The examination of user and alter responses was conducted separately, followed by a comparative analysis using the Jensen–Shannon Divergence (*JSd*) [43], adapted to calculate similarity rather than divergence:

$$Sim_{JS} = 1 - JSd \tag{4}$$

  The selection of Jensen–Shannon Divergence was driven by its capacity to provide a well-defined, symmetrical measure, enabling comparison between probabilistic distributions even in the presence of null values in some categories. Furthermore, it offers greater numerical stability due to the averaging of the analyzed distributions, making it particularly effective for studying complex stylistic distributions characterized by high variability and sparsity. The modification of the formula for calculating similarity Sim_{JS} has been shown to facilitate a more intuitive interpretation of results, thereby enhancing the clarity of the degree of similarity between the analyzed distributions. This approach has been demonstrated to enable more precise assessment of stylistic similarity between responses and to facilitate comparison of the behavior of different models with different techniques.
- **Similarity of Content**: The assessment of content preservation was conducted through the utilization of cosine similarity, which was calculated on the semantic embeddings generated by Sentence-BERT [44]. This methodological approach enabled the quantification of the degree of similarity between the content of the responses generated by the alter and that produced by the real individual. Ultimately, the average cosine similarity value over all responses was calculated to obtain an overall score.

Although the techniques used have proven to be robust for evaluating content and style preservation, human evaluation remains irreplaceable for more comprehensive and qualitative judgments of style quality. Automated metrics provide valuable support for quantitative analysis, but they cannot match the accuracy and sensitivity of human judgment.

The results of this study are in line with ongoing research in human–robot interaction (HRI) and artificial intelligence-driven conversational systems. Recent studies have explored how LLMs enhance the ability of robotic assistants to engage in naturalistic conversations, adapt to user preferences, and personalize interactions [45]. In addition, the application of LLMs in digital avatars has been explored for its role in enhancing the emotional intelligence and social cues of virtual and robotic agents [46,47]. Our findings support previous work indicating that personalized conversational patterns contribute to more engaging and effective interactions, particularly in assistive robotics and telepresence. Robot embodiment studies suggest that integrating dialogue systems that adapt style could improve the acceptance and usability of AI-driven robots in healthcare, customer service, and education [48].

6. Conclusions

This work demonstrates the successful integration of advanced LLM technologies with 3D avatar modeling to produce realistic digital alter egos. The Tree of Style (ToS) technique provides a scalable, efficient, and effective means of replicating an individual's conversational style, thereby eliminating the need for extensive fine-tuning. The scalability of the technique is attributed to its capacity to adjust the number of nodes in the tree, which is contingent on the complexity of the style to be emulated. For instance, a conversational style, which is less intricate, necessitates a smaller number of nodes compared to a specialized style, thereby allowing for the optimization of computational resources. In this study, the style to be emulated is that of a female subject employing informal, colloquial language in a familiar context. This style did not prove excessively onerous to replicate, and 10 nodes represented a reasonable compromise to achieve optimal results without unduly burdening resources. The results underscore the system's capacity to generate responses that are stylistically and contextually appropriate, facilitated by a high-fidelity speech synthesis module. The visual realism and linguistic accuracy of the avatar contribute to its potential as a valuable tool for enhancing interactions in virtual environments.

Whilst this study focuses on the metaverse, the methodologies developed have significant potential applications in robotics. The Tree of Style (ToS) methodology, designed for personalized conversational responses, could be implemented in robotic systems to enhance their linguistic adaptability. Robots in social and assistive contexts, such as healthcare companions or customer service agents, could benefit from style-adaptive conversational AI, rendering their interactions more natural and user-specific.

Nevertheless, the ethical implications of replicating human likeness and style warrant further exploration. Future research will concentrate on enhancing context awareness, improving adaptability to different applications, and resolving privacy and consent challenges in the use of digital surrogates. Concerning the conversational module, a future implementation could involve the development of a dynamic prompt. This prompt would adapt to the communicative context in which the avatar finds itself, including the specific situation (e.g., a family, professional, or social setting) and the interlocutor with whom the avatar is interacting. The prompt would ensure precise and contextually appropriate responses. To illustrate this, consider a scenario where an avatar is conversing with a family member; in this instance, the language would remain informal and colloquial. Conversely, if the interlocutor were an employer or friend, the avatar would adopt a different tone and content to align with the relationship. This dynamic updating of the knowledge base would ensure focused and relevant communication, avoiding digressions. These efforts will pave the way for more responsible and widespread adoption of personalized avatars in the metaverse and beyond.

Author Contributions: Conceptualization, methodology, and software, M.N. and G.F.G.; validation, M.N.; data curation, M.N.; 3D avatar, G.F.G.; writing—original draft preparation, G.F.G. and M.N.; writing—review and editing, G.F.G. and M.N.; supervision, V.S. and A.C.; project administration, V.S. and A.C.; funding acquisition, V.S. and A.C. All authors have read and agreed to the published version of the manuscript.

Funding: Funded by European Union—"Next Generation EU"-PNRR M4-C2-investimento 1.1: Fondo per il Programma Nazionale di Ricerca e Progetti di Rilevante Interesse Nazionale (PRIN)-PRIN 2022 cod. 2022MM8LKM Title: "ALTEREGO: how to emulate intentionality and awareness in remote communications by means of software surrogates" CUP B53D23013140006.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patient(s) to publish this paper.

Data Availability Statement: The TOS source code can be found at the link https://github.com/MicheleNas/Tree-Of-Style.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Wang, H.; Ning, H.; Lin, Y.; Wang, W.; Dhelim, S.; Farha, F.; Ding, J.; Daneshmand, M. A Survey on the Metaverse: The State-of-the-Art, Technologies, Applications, and Challenges. *IEEE Internet Things J.* **2023**, *10*, 14671–14688. [CrossRef]
2. Gatto, L.; Gaglio, G.F.; Augello, A.; Caggianese, G.; Gallo, L.; La Cascia, M. MET-iquette: Enabling virtual agents to have a social compliant behavior in the Metaverse. In Proceedings of the 2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Dijon, France, 19–21 October 2022; pp. 394–401. [CrossRef]
3. Gaglio, G.F.; Augello, A.; Pipitone, A.; Gallo, L.; Sorbello, R.; Chella, A. Moral Mediators in the Metaverse: Exploring Artificial Morality through a Talking Cricket Paradigm. In Proceedings of the CEUR Workshop, Rome, Italy, 4 September 2023; pp. 30–43.
4. Rong, R.; Kitamura, Y.; Kitaoka, S.; Kato, H.; Kishi, K.; Nakashima, K.; Yamazaki, K.; Horiuchi, S.; Shioda, T.; Kamakura, K.; et al. Osaka developing story: An application of video agents. In Proceedings of the 7th International Conference on Advances in Computer Entertainment Technology, Taipei, Taiwan, 17–19 November 2010; pp. 71–74. [CrossRef]
5. Rojas Galeano, S.A. *The e-Alter Ego: Ghosts Inside the Internet*; Universidad Distrital Francisco José de Caldas: Bogotá, Colombia, 2000.
6. Wright, A.; Shinkle, E.; Linney, A. Alter ego: Computer reflections of human emotions. In Proceedings of the 6th Digital Art Conference, Copenhagen, Denmark, 1–3 December 2005.
7. Wan, H.; Zhang, J.; Suria, A.A.; Yao, B.; Wang, D.; Coady, Y.; Prpa, M. Building LLM-based AI Agents in Social Virtual Reality. In Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 11–16 May 2024; pp. 1–7. [CrossRef]
8. Yamazaki, T.; Mizumoto, T.; Yoshikawa, K.; Ohagi, M.; Kawamoto, T.; Sato, T. An Open-Domain Avatar Chatbot by Exploiting a Large Language Model. In Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Prague, Czechia, 11–15 September 2023; pp. 428–432.
9. Shoa, A.; Oliva, R.; Slater, M.; Friedman, D. Sushi with Einstein: Enhancing Hybrid Live Events with LLM-Based Virtual Humans. In Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents, Würzburg, Germany, 19–22 September 2023; pp. 1–6. [CrossRef]
10. Kerdvibulvech, C.; Chang, C.C. A New Study of Integration Between Social Robotic Systems and the Metaverse for Dealing with Healthcare in the Post-COVID-19 Situations. In Proceedings of the International Conference on Software Reuse, Montpellier, France, 15–17 June 2022.
11. Shivani, S.; Sharada, A.; Ratkal, B. Simulation of Human Robot Interaction Through Avatar. *Int. Res. J. Mod. Eng. Technol. Sci.* **2023**, *5*, 977–982.
12. Dragone, M.; Holz, T.; O'Hare, G.M.P. Mixing robotic realities. In Proceedings of the International Conference on Intelligent User Interfaces, Sydney, Australia, 29 January–1 February 2006.
13. Wiederhold, B.K. Treading Carefully in the Metaverse: The Evolution of AI Avatars. *Cyberpsychol. Behav. Soc. Netw.* **2023**, *26*, 321–322. [CrossRef]
14. Mishra, C.; Verdonschot, R.; Hagoort, P.; Skantze, G. Real-time emotion generation in human-robot dialogue using large language models. *Front. Robot. AI* **2023**, *10*, 1271610. [CrossRef] [PubMed]
15. Shanahan, M.; McDonell, K.; Reynolds, L. Role play with large language models. *Nature* **2023**, *623*, 493–498. [CrossRef] [PubMed]
16. Sejnowski, T.J. Large language models and the reverse turing test. *Neural Comput.* **2023**, *35*, 309–342. [CrossRef]
17. Pappula, S.R.; Allam, S.R. LLMs for Conversational AI: Enhancing Chatbots and Virtual Assistants. *Int. J. Res. Publ. Rev.* **2023**, *4*, 1601–1611. [CrossRef]
18. Friedman, L.; Ahuja, S.; Allen, D.; Tan, Z.; Sidahmed, H.; Long, C.; Xie, J.; Schubiner, G.; Patel, A.; Lara, H.; et al. Leveraging Large Language Models in Conversational Recommender Systems. *arXiv* **2023**, arXiv:2305.07961. [CrossRef]
19. Tsubota, Y.; Kano, Y. Text Generation Indistinguishable from Target Person by Prompting Few Examples Using LLM. In Proceedings of the 2nd International AIWolfDial Workshop, Tokyo, Japan, 17–19 September 2024.
20. Chen, Z.; Moscholios, S. Using Prompts to Guide Large Language Models in Imitating a Real Person's Language Style. *arXiv* **2024**, arXiv:2410.03848. [CrossRef]
21. Tao, Z.; Xi, D.; Li, Z.; Tang, L.; Xu, W. CAT-LLM: Prompting Large Language Models with Text Style Definition for Chinese Article-style Transfer. *arXiv* **2024**, arXiv:2401.05707. [CrossRef]

22. Wegmann, A.; Schraagen, M.; Nguyen, D. Same Author or Just Same Topic? Towards Content-Independent Style Representations. In Proceedings of the 7th Workshop on Representation Learning for NLP, Dublin, Ireland, 26 May 2022; pp. 249–268. [CrossRef]
23. Van Der Boon, M.; Fermoselle, L.; Ter Haar, F.; Dijkstra-Soudarissanane, S.; Niamut, O. Deep Learning Augmented Realistic Avatars for Social VR Human Representation. In Proceedings of the ACM International Conference on Interactive Media Experiences, Aveiro, Portugal, 22–24 June 2022; pp. 311–318. [CrossRef]
24. Beacco, A.; Gallego, J.; Slater, M. Automatic 3D avatar generation from a single RBG frontal image. In Proceedings of the 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Christchurch, New Zealand, 12–16 March 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 764–765.
25. Zielonka, W.; Bolkart, T.; Thies, J. Instant volumetric head avatars. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 4574–4584.
26. Vilchis, C.; Perez-Guerrero, C.; Mendez-Ruiz, M.; Gonzalez-Mendoza, M. A survey on the pipeline evolution of facial capture and tracking for digital humans. *Multimed. Syst.* **2023**, *29*, 1917–1940. [CrossRef]
27. Zakharov, E.; Ivakhnenko, A.; Shysheya, A.; Lempitsky, V. Fast Bi-Layer Neural Synthesis of One-Shot Realistic Head Avatars. In *Computer Vision—ECCV 2020*; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M., Eds.; Series Title: Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2020; Volume 12357, pp. 524–540. [CrossRef]
28. Cao, C.; Simon, T.; Kim, J.K.; Schwartz, G.; Zollhoefer, M.; Saito, S.S.; Lombardi, S.; Wei, S.E.; Belko, D.; Yu, S.I.; et al. Authentic volumetric avatars from a phone scan. *ACM Trans. Graph.* **2022**, *41*, 1–19. [CrossRef]
29. Dong, Z.; Guo, C.; Song, J.; Chen, X.; Geiger, A.; Hilliges, O. Pina: Learning a personalized implicit neural avatar from a single rgb-d video sequence. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 20470–20480.
30. Lattas, A.; Moschoglou, S.; Gecer, B.; Ploumpis, S.; Triantafyllou, V.; Ghosh, A.; Zafeiriou, S. AvatarMe: Realistically Renderable 3D Facial Reconstruction "in-the-wild". In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 760–769.
31. Özyeşil, O.; Voroninski, V.; Basri, R.; Singer, A. A survey of structure from motion. *Acta Numer.* **2017**, *26*, 305–364. [CrossRef]
32. Schøsler, L. The Diasystematic Status of the Diatopic Axis. *Scand. Stud. Lang.* **2023**, *14*, 55–72. [CrossRef]
33. Wüest, J.T. La notion de diamésie est-elle nécessaire? *Trav. De Linguist.* **2009**, *59*, 147–162. [CrossRef]
34. Tekavčić, P. Roberto Gusmani, Itinerari linguistici, Scritti raccolti in occasione del 60. compleanno, a cura di Raffaella Bombi, Guido Cifolrtti, Sara Frdalto, Fabiana Fusco, Lucia Innocente, Vincenzo Orioles; Edizioni dell'Orso, Alessandria, 1995; XXVII+382 pagine. *Linguistica* **1997**. Available online: https://hdl.handle.net/11390/684709 (accessed on 10 March 2025).
35. Virdis, M. Problemi di diatopia e di diacronia della lingua sarda. Un'ipotesi di sociolinguistica storica. In *Il Sardo in Movimento. Vandenhoeck & Ruprecht Unipress, Göttingen;* Vienna University Press: Wien, Republik Österreic, 2020.
36. De Mauro, T. Il Nuovo Vocabolario di Base Della Lingua Italiana. Internazionale. 2016. Available online: https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana (accessed on 28 November 2020).
37. Yao, S.; Yu, D.; Zhao, J.; Shafran, I.; Griffiths, T.; Cao, Y.; Narasimhan, K. Tree of thoughts: Deliberate problem solving with large language models. *Adv. Neural Inf. Process. Syst.* **2024**, *36*, 11809–11822.
38. Yinhan, L.; Myle, O.; Naman, G.; Jingfei, D.; Mandar, J.; Danqi, C.; Omer, L.; Mike, L. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv* **2019**, arXiv:1907.11692.
39. Ostheimer, P.S.; Nagda, M.K.; Kloft, M.; Fellenz, S. Text Style Transfer Evaluation Using Large Language Models. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), Torino, Italy, 20–25 May 2024.
40. Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.J. Bleu: A method for automatic evaluation of machine translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 6–12 July 2002; pp. 311–318.
41. Lin, C.Y. Rouge: A package for automatic evaluation of summaries. In Proceedings of the Text Summarization Branches Out, Barcelona, Spain, 25–26 July 2004; pp. 74–81.
42. Banerjee, S.; Lavie, A. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, Ann Arbor, MI, USA, 29 June 2005; pp. 65–72.
43. Menéndez, M.L.; Pardo, J.; Pardo, L.; Pardo, M. The jensen-shannon divergence. *J. Frankl. Inst.* **1997**, *334*, 307–318. [CrossRef]
44. Reimers, N. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *arXiv* **2019**, arXiv:1908.10084.
45. Han, D.; McInroe, T.A.; Jelley, A.; Albrecht, S.V.; Bell, P.; Storkey, A.J. LLM-Personalize: Aligning LLM Planners with Human Preferences via Reinforced Self-Training for Housekeeping Robots. In Proceedings of the International Conference on Computational Linguistics, Torino, Italy, 20–25 May 2024.
46. Kang, H.; Moussa, M.B.; Magnenat-Thalmann, N. Nadine: An LLM-driven Intelligent Social Robot with Affective Capabilities and Human-like Memory. *arXiv* **2024**, arXiv:2405.20189.

47. Lee, Y.K.; Jung, Y.Y.; Kang, G.; Hahn, S. Developing Social Robots with Empathetic Non-Verbal Cues Using Large Language Models. *arXiv* **2023**, arXiv:2308.16529.
48. Ritschel, H.; Janowski, K.; Seiderer, A.; Wagner, S.; André, E. Insights on usability and user feedback for an assistive robotic health companion with adaptive linguistic style. In Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments, Island of Rhodes, Greece, 5–7 June 2019; pp. 319–320.