Integrating Personality into Digital Humans: A Review of LLM-Driven Approaches for Virtual Reality

Iago A. Brito*
Advanced Knowledge Center in Immersive Technologies
Federal University of Goiás

Julia S. Dollis†
Advanced Knowledge Center in Immersive Technologies
Federal University of Goiás

Fernanda B. Färber‡
Advanced Knowledge Center in Immersive Technologies
Federal University of Goiás

Pedro S. F. B. Ribeiro§
Advanced Knowledge Center in Immersive Technologies
Federal University of Goiás

Rafael T. Sousa¶
Advanced Knowledge Center in Immersive Technologies
Federal University of Mato Grosso

Arlindo R. Galvão Filho‖
Advanced Knowledge Center in Immersive Technologies
Federal University of Goiás

ABSTRACT

The integration of large language models (LLMs) into virtual reality (VR) environments has opened new pathways for creating more immersive and interactive digital humans. By leveraging the generative capabilities of LLMs alongside multimodal outputs such as facial expressions and gestures, virtual agents can simulate human-like personalities and emotions, fostering richer and more engaging user experiences. This paper provides a comprehensive review of methods for enabling digital humans to adopt nuanced personality traits, exploring approaches such as zero-shot, few-shot, and fine-tuning. Additionally, it highlights the challenges of integrating LLM-driven personality traits into VR, including computational demands, latency issues, and the lack of standardized evaluation frameworks for multimodal interactions. By addressing these gaps, this work lays a foundation for advancing applications in education, therapy, and gaming, while fostering interdisciplinary collaboration to redefine human-computer interaction in VR.

Index Terms: Digital humans, virtual reality, large language models.

1 INTRODUCTION

Virtual reality (VR) technology has emerged as a powerful medium to create and replicate real-world scenarios in digital spaces, offering unparalleled opportunities for simulation and interaction [23, 53]. Recent advancements in artificial intelligence have further enhanced VR environments by enabling more dynamic scenarios, such as generating text-based descriptions for 360-degree VR scenes [56] and user attention-guidance in virtual tours around real locations using multi-modal models [69]. Complementing these advancements, studies on computer-driven characters with human-like characteristics, referred to as digital humans, highlight their significant role in enhancing user engagement within VR environments by providing realistic facial expressions, responsive gestures, and natural conversational abilities, mimicking humans not only in their visual design but also in their personality, decreasing the gap between virtual and real-world interactions [33, 76].

The potential applications of digital humans in VR environment passes through a variety of fields. In educational settings, the use of virtual pedagogic agents with human-like characteristics demonstrate gains in students' learning experience and engagement [23], while in therapeutic context, the use of digital humans combined with therapy techniques demonstrates gains in reducing symptoms of borderline personality disorder (BPD) [18]. These possibilities underscore the importance of further research into designing, implementing, and evaluating such agents in VR.

In parallel, the field of natural language processing (NLP) has witnessed transformative advancements with the rise of large language models (LLMs), such as GPT [4] and LLaMA [77] family of models. Trained on vast datasets containing trillions of words, these models demonstrate emergent abilities (i.e., capabilities that arise without explicit training for specific tasks), like the capacity to simulate human personality and emotions in real-time conversations, providing an unprecedented means for human-computer interaction [68, 40, 80]. These capabilities represent a critical step forward in creating digital agents that can think and respond in ways that feel authentically human.

Despite their potential, most applications of these generative models remain confined to chat-based interfaces. Although such systems offer interactive benefits, they do not simulate real-world communication as effectively as VR technology. The integration of NLP-driven digital humans into VR offers a transformative alternative: enabling face-to-face interactions where digital agents can express themselves through both verbal and non-verbal cues, such as facial expressions, gestures, and personality-driven behavior [75, 50]. This multimodal interaction is closer to real human-to-human communication, offering users a more engaging and satisfying experience than chat interfaces.

Furthermore, studies have shown that digital humans personality directly influences how users perceive and interact with the system, impacting key metrics like immersion, trust, and engagement [66, 40]. By embedding these agents with the ability to convey personality traits and human-like expressions, VR can promote deeper connections and more meaningful interactions, paving the way for future advancements in virtual communication.

This work provides a comprehensive review of methods for enabling digital humans to adopt personas and identities, focusing on approaches developed both within and outside the VR context and exploring their adaptation to immersive environments. It examines how large language models can simulate human-like traits, combining verbal and non-verbal communication to enhance immersion and realism. The paper also highlights the challenges of integrating LLM-driven personality traits into VR, including computational demands, latency issues, and the lack of standardized evaluation frameworks for multimodal interactions. By addressing these gaps, this work lays the foundation for advancing applications in education, therapy, and interactive simulations, fostering more engaging and human-like virtual interactions.

---

*e-mail: iagoalves@discente.ufg.br
†e-mail: juliadollis@discente.ufg.br
‡e-mail: fernandabufon@discente.ufg.br
§e-mail: schindler@discente.ufg.br
¶e-mail: rafaelsousa@ufmt.br
‖e-mail: arlindogalvao@ufg.br

2 DIGITAL HUMANS

Digital humans (DH) have become a focal point of research alongside the rapid expansion of VR technology. These entities are computer-generated representations of human beings, designed to replicate physical appearances, cognitive functions, and emotional responses in digital environments [49, 30]. By simulating human behavior across multiple dimensions, digital humans aim to bridge the gap between artificial systems and natural interactions, unlocking new possibilities for immersive experiences.

Despite their potential, DH often fail to exhibit the nuanced personalities and emotional depth necessary to promote natural and engaging interactions. This absence of relatable human-like traits often results in interactions that feel mechanical and impersonal, making it difficult for users to form meaningful connections with these agents [89, 73]. By designing these agents with more human-like behavior and expressions, these systems can offer greater relatability, which enhances the overall interaction experience and promotes stronger user engagement.

By aligning digital humans behaviors with user expectations and situational needs, these personality-driven agents promote a deeper sense of connection, making interactions more meaningful and enjoyable [66]. Digital humans with empathetic personality demonstrates improvements in learning experiences by creating a supportive environment where users feel understood and motivated, ultimately increasing their willingness to engage and learn [73].

In virtual reality environments, agents with distinct and well-defined personalities, conveyed through both textual and visual features, play a critical role in improving user engagement and immersion. By emulating human-like behaviors and expressions, these agents enable interactions that are both dynamic and natural [73, 58]. Moreover, by embedding distinct personality traits into virtual reality scenarios, they play a pivotal role in enhancing both the realism and the functional effectiveness of digital humans.

2.1 Personality-driven Agents

Personality conditioning refers to the process of equipping virtual agents with human-like traits and behaviors to enhance the quality of their interactions with users [65]. Early approaches to personality conditioning utilized rule-based systems and decision trees to simulate behavior. These systems mapped specific inputs to predefined outputs, functioning within a limited framework of programmed responses [87]. While innovative for their time, these methods were inherently rigid, lacking the adaptability to handle the unpredictability of human interactions. This rigidity often resulted in repetitive, unnatural conversations that failed to engage users meaningfully or address the complexities of real-world scenarios.

The integration of verbal communication with non-verbal cues has proven essential in enhancing interactions between humans and virtual agents, creating experiences that feel more natural and immersive [6]. By combining elements such as facial expressions, gestures, and tone of voice, these systems enable dynamic and responsive interactions. This multimodal approach not only enriches conversational content but also improves conversational flow, fostering mutual understanding and creating a stronger sense of presence in virtual environments.

Building on these foundational principles, recent advancements in embodied virtual agents have pushed the boundaries of natural and effective interactions. For example, [2] developed a system designed to enhance Willingness to Communicate (WTC) in second-language learners. By incorporating context-dependent domain knowledge and user intent detection, the agent managed conversational flow in a way that was both semantic and user-centric. Learners reported increased confidence and a greater desire to communicate in a foreign language after interacting with the system, demonstrating its potential for language acquisition.

However, even with these advancements, capturing the complexity of human personality within these systems remained a challenge. The absence of subtle, context-aware responses can lead to interactions that, while functional, lack the depth and authenticity required for truly lifelike experiences. This highlighted the need for more advanced approaches, which LLMs later addressed by allowing more nuanced and flexible personality modeling capabilities.

3 LARGE LANGUAGE MODELS (LLMs)

The field of natural language processing (NLP) has undergone rapid transformations since the introduction of the Transformer architecture [81]. This breakthrough has fundamentally reshaped NLP, leading to the proliferation of numerous language models, particularly decoder-based architectures, which have revolutionized text generation and paved the way for large language models (LLMs). These advanced neural networks redefine the boundaries of machine understanding and language generation, leveraging extensive datasets comprising billions or even trillions of words from diverse domains to maximize linguistic coverage and generalization [4].

The unprecedented scale of LLMs, characterized by vast numbers of parameters, enables them to capture intricate linguistic patterns, including semantic relationships, syntactic structures, and long-range dependencies [77, 4]. This capacity allows them to perform complex language-related tasks, such as advanced reasoning and contextual understanding. In particular, LLMs demonstrate emergent abilities, performing tasks that are not explicitly programmed but naturally arise from their scale and the complexity of their training process [85, 37].

The development of large language models generally involves two primary stages: pre-training and fine-tuning. The pre-training phase involves exposing the model to extensive volumes of unlabeled general knowledge text data through a language modeling task, most commonly framed as the next token prediction in a sequence [4]. Then, the resulting model passes through the fine-tuning phase, which refines the model's capabilities to perform a more specific set of tasks. This process leverages the pre-trained weights, which encode general knowledge about language, morphology, and syntax, and adjusts them using labeled datasets in a supervised learning framework [16]. This step, commonly referred to as Supervised Fine-Tuning (SFT), aligns the model's performance with the requirements of particular applications, significantly enhancing its ability to tackle domain-specific tasks [52].

While fine-tuning significantly enhances a model's domain-specific performance, it often requires access to a large amount of labeled data, which presents a substantial barrier, especially when dealing with specialized domains. This challenge has motivated research into alternative approaches to task adaptation, like In-Context Learning (ICL) [47], which allows learning different tasks by textual instructions at inference time using techniques such as zero-shot and few-shot learning [4, 77].

This shift of paradigm enables LLMs to execute complex tasks without requiring updates to their parameters or access to extensive labeled datasets in the target domain. Instead, prompts function as contextual frameworks containing either examples of the target task or explicit natural language instructions.

Such adaptability is instrumental in advancing large language models, fostering the development of systems that can dynamically learn, adjust to new contexts, and execute tasks requiring intricate understanding. Among these advancements, a particularly noteworthy avenue is the simulation of human-like behaviors, such as the ability to exhibit nuanced linguistic and contextual adaptations. Although emergent abilities have significantly enhanced natural language understanding and generation, their implications for enabling coherent and consistent personality modeling in virtual agents present a promising yet underexplored research frontier.

3.1 Modelling Personality Through LLMs

Building on the concept of emergent abilities in large language models, the task of represent human personality into virtual agents represents an exciting and emergent domain. Personality modeling seeks to create systems that not only respond contextually but also reflect consistent psychological characteristics, enhancing user engagement and creating more lifelike interactions [65].

Research into personality modeling within LLMs has primarily focused on understanding the intrinsic personality traits of these models and developing methods to control or induce specific traits. Studies such as [57] and [67] have analyzed the underlying characteristics of LLMs using established psychological frameworks, such as the MBTI [54] and Big Five [20] personality traits, highlighting the inherent tendencies of LLMs emerged from the training data and model architecture, and exploring how these tendencies might align or deviate from human-like personality expressions.

Personality modeling in Large Language Models (LLMs) predominantly employs three primary approaches: zero-shot learning, few-shot learning, and fine-tuning.

1. **Zero-Shot Learning:**

   By crafting precise instructions that implicitly convey desired personality traits, these methods guide the model to generate responses aligned with specific personality characteristics. For instance, the approach described in [35] uses descriptive prompts that can elicit specific personality traits by framing the context in a way that enables the model to naturally embody the intended personality characteristics.

2. **Few-Shot Learning:**

   This method involves providing the model with a limited number of carefully selected examples that demonstrates the desired personality characteristics, thus improving the ability of the model to generalize these traits in subsequent interactions. An illustrative example is the study [98] where a limited number of carefully crafted examples from the training data (e.g., responses from the IPIP-NEO-120 questionnaire) are provided within the prompt.

3. **Fine-Tuning:**

   This supervised learning process modifies the pre-trained weights using labeled datasets that reflect the desired personality characteristics, thereby enhancing the model's ability to consistently exhibit these traits across diverse tasks. An example of this approach is [68] which integrates historical context and behavioral traits into the training process to create historically grounded personas. By refining the model's internal representations, fine-tuning ensures that the LLM not only understands general language constructs but also reliably manifests the specified personality attributes in its outputs.

Despite significant advancements in personality modeling using large language models (LLMs), much of the research remains centered on text-based chatbot interactions. For example, studies such as [32] and [63] delve into LLM-driven personalities within text or role-playing contexts. Similarly, [39] investigates personality applications in non-player characters (NPCs) within gaming environments. However, these works often neglect the unique challenges of virtual reality, where embodied agents interact with users in immersive, real-time scenarios. Addressing these challenges requires integrating multimodal communication, ensuring responsiveness, and creating an engaging sense of presence, areas that remain underexplored in the current literature.

Virtual reality introduces distinct opportunities by enabling interactions that are richer and more immersive than text-based systems. The use of gestures, facial expressions, and gaze enhances the expression of personality, making interactions more dynamic and engaging. As highlighted in [40], such interactions emulate natural human communication, fostering deeper social connections and user engagement. While these advantages highlight VR's potential, they also underscore the need for research focused on optimizing personality-driven digital agents for this medium, ensuring that non-verbal communication aligns seamlessly with verbal outputs to create truly lifelike interactions.

The ability to express personality through both text and non-verbal cues, such as gestures and facial expressions, is crucial for enhancing user experience, underscoring the need for further research into seamlessly integrating these expressive modalities into VR environments. While some studies explore multimodal interactions, they often fall short of fully addressing their application in VR. For instance, [5, 12, 83] investigate simulations where agents communicate through speech and gestures or delve into the role of facial expressions, yet they either briefly mention the potential integration of VR or omit it entirely.

Despite the scarcity of research on incorporating LLM-driven personality traits into digital humans in VR, a few studies have begun exploring this convergence. For example, [28] employs multimodal technologies to enhance human-computer interaction, demonstrating the use of LLMs in real-time, open-domain conversations across multiple languages while incorporating speech synthesis and facial expressions. By aligning verbal communication with non-verbal emotional cues, this work illustrates the feasibility of combining LLMs and multimodal systems to create virtual agents with heightened contextual and emotional responsiveness. Such implementations highlight the potential of these technologies to foster more engaging and lifelike interactions in VR environments.

This absence of literature underscores the necessity for more comprehensive studies that bridge these two fields, paving the way for better and more engaging ways to interact with advanced virtual agents with rich personalities.

4 EVALUATING DIGITAL HUMANS PERSONALITY TRAITS

As digital humans become increasingly integrated into daily life, their behavior and impact on user interactions have become areas of growing interest [91, 55]. In the field of psychology, personality traits are defined as enduring patterns of thoughts, feelings, and behaviors that exhibit consistency in time and situations, underscoring individual differences along a set of basic dimensions, reflecting stability and predictability in behavior [15].

The systematic study of personality traits has a long history in psychology, aiming to understand and categorize individual differences [8]. Trait theories provide structured frameworks to explore how people think, feel, and behave in consistent ways over time and across situations. Broad frameworks, like the Big Five model [14], focus on overarching dimensions of personality, organizing personality into five broad dimensions—openness, conscientiousness, extroversion, agreeableness, and neuroticism, capturing fundamental aspects of how individuals engage with the world. More detailed approaches, such as the Sixteen Personality Factors (16PF) [7], break down personality into 16 specific traits, such as emotional stability, social courage, and sensitivity.

Collectively, these methods provide structured tools to explore the complexity of personality, fostering a deeper understanding of the stable characteristics that shape individual differences. Furthermore, research has consistently shown a strong correlation between personality and language use [29, 61, 60, 41]. This relationship has enabled the evaluation of personality traits through text analysis, allowing researchers to assess individual characteristics by examining the frequency and patterns of words and expressions used.

Although foundational theories commonly applied to human beings have inspired efforts to evaluate personality traits in LLMs

[35, 27, 88, 82], it remains a complex and nuanced challenge, particularly in digital humans, as it involves a highly subjective context-dependent and non-standardized domain. Existing methods for evaluating LLM personality traits include human evaluations [19, 25], LLM-as-judge approaches [45, 44], and adaptations of self-assessment personality tests originally designed for humans [72, 38].

These approaches have primarily focused on text-based interactions, examining whether LLMs exhibit consistent and stable personality traits [71, 72], exploring methodologies for personality assessment [42, 99], investigating the impact of personality on model safety [92], and tailoring personality traits to specific use cases [46, 31, 82]. These efforts have established a foundation for evaluating personality in text-based agents but fall short when considering the multimodal nature of embodied conversation agents. Due to its complex layers of interaction and connection, we are introduced to challenges that existing evaluation methods, which primarily focus on text-based outputs, are not equipped to address. In that way, significant gaps remain in the standardization and application of multimodal evaluation frameworks.

To address this gap, it is crucial to critically analyze existing methods for assessing personality traits in LLMs and consider how these approaches can be adapted for VR contexts and embodied conversational agents. The following sections examine qualitative and quantitative evaluation methods for LLMs-driven personality traits, highlighting their strengths, limitations, and potential applicability to computer-driven characters.

4.1 Qualitative Evaluation

As previously discussed, evaluating the personality traits of digital humans involves complex, nuanced, and non-standardized methods. Among these, qualitative evaluation methods are widely used, particularly in LLM-based conversational models [45]. These methods are primarily subjective, relying on human (or, in some cases, LLM) interpretation to assess an agent's performance, coherence, and adherence to specific personality traits. While qualitative evaluations are valuable for understanding how users perceive an agent's behavior and personality—critical for improving user-agent relationships—they lack standardization [9]. To address the limitations of purely human or purely LLM-based evaluations, many studies adopt a hybrid approach, combining the strengths of both methods [35]. However, the inherent subjectivity of these methods means that experiences and interpretations can vary significantly from person to person (or model to model, in the case of LLM-as-judge evaluations).

Human evaluation has historically been one of the most widely used qualitative methods in research, particularly for assessing alignment in LLM-based models. This approach evaluates whether an LLM's behavior aligns with human intentions, ethical values, and predefined goals [1]. Beyond alignment, it serves as a critical tool for evaluating the personality traits and behavioral consistency of LLMs [82]. This method typically involves interactive tests, where evaluators engage directly with the agent to determine whether it effectively exhibits the desired personality traits and meets user expectations [35]. Human evaluation can be conducted from multiple perspectives, enabling a comprehensive analysis of various dimensions of the agent's behavior. These distinct angles of evaluation provide deeper insights into the agent's overall performance and its adherence to intended personality traits and behavioral standards.

Human evaluation has been widely used to assess how personality traits influence user perceptions of virtual agents during conversational tasks, offering valuable insights into social interactions. Research in this area has explored virtual agents designed to exhibit extroverted or introverted characteristics, expressed through variations in tone, communicative style, and level of engagement [40].

User assessments based on factors such as warmth, empathy, and realism have demonstrated the significant role of personality traits in shaping perceptions and interactions in virtual environments.

However, human evaluation as a methodology presents notable limitations. Its subjective nature introduces variability, as individual evaluators may interpret and score interactions differently, leading to inconsistencies, particularly in studies involving multiple evaluators with diverse perspectives. This variability can undermine the reliability of results and complicate cross-study comparisons. Furthermore, the effectiveness of human evaluation as a primary approach to assess the personality traits of digital humans remains a topic of ongoing debate within the research community [13]. These challenges highlight the need for standardized evaluation protocols and objective metrics to improve consistency and comparability between studies.

Another qualitative evaluation method involves the use of the LLM-as-judge approach. This emerging framework employs LLMs to evaluate the personality and behavior of other LLMs and conversational agents by simulating human judgment [45]. The LLM-as-judge methodology has gained traction due to its scalability and ability to automate the evaluation process [17, 97], reducing the dependence on human evaluators, which can be time-consuming and costly.

In this approach, an LLM is provided with a prompt containing the text to be assessed along with clear evaluation criteria. For example, the criteria might involve a question designed to assess specific personality traits of the tested LLM. Based on the prompt, the evaluating LLM assigns scores or labels to the text according to the predefined criteria. This method has been widely employed to evaluate conversational agents, models, and chatbots across various domains [84], including the context of this research [80].

Although the LLM-as-judge approach offers notable benefits, such as increased efficiency and reduced costs for evaluating LLM personality traits, it presents several challenges, including inherent biases in the models, sensitivity to different prompts, limitations specific to individual models, and concerns regarding the consistency and reliability of the evaluations [10, 95]. Addressing these issues is essential to enhance the accuracy and fairness of this methodology, highlighting the need for further refinement and the establishment of standardized evaluation frameworks.

4.2 Quantitative Evaluation

Quantitative evaluation methods have become an essential tool for assessing personality traits in LLM-controlled agents, offering a structured and objective approach. These methods focus on analyzing the behavior of language models based on outputs such as responses to questions, storytelling, and other text-based interactions.

Common approaches, such as personality assesments [86] and word frequency analysis [74], aim to provide standardization and reproducibility, addressing the variability and subjectivity often associated with qualitative assessments. However, they still face challenges, including ongoing debates about their reliability in accurately evaluating LLM personality traits [24]. Despite these limitations, quantitative methods provide a valuable foundation for comparing models and identifying consistent patterns in personality expression, making them indispensable for evaluating digital agents in a systematic manner.

Among quantitative evaluation methods, personality assessments originally developed for the psychological evaluation of humans [62, 86] have been extensively employed to analyze a wide range of personality characteristics in LLM-driven agents. These assessments serve as systematic tools designed to measure diverse dimensions of personality, encompassing traits, behaviors, motivations, and interpersonal dynamics. Instruments such as the Big Five Inventory (BFI) [36], the International Personality Item Pool

(IPIP) [22], and the NEO Personality Inventory (NEO-PI-3) [51] are among the most frequently utilized for evaluating both human subjects and virtual agents.

In these structured assessments, individuals or LLMs provide evaluations of behaviors, thoughts, and feelings by responding to a series of predetermined statements or questions. These responses, collected through formats such as Likert scales (e.g., ratings ranging from "strongly disagree" to "strongly agree") [48], true-false options, or forced-choice formats, are scored and analyzed to produce detailed profiles of personality traits. When applied to LLMs, specific prompt instructions are often used to simulate self-reported responses, enabling the modeling of human-like personality traits [34, 67, 64].

Although widely used, these assessments present significant challenges when applied to non-human entities such as LLMs. One notable issue is the sensitivity of LLMs to multiple-choice question (MCQ) formats, as responses are heavily influenced by the order of options presented [93]. Since self-reported questionnaires frequently rely on MCQs, this introduces bias into the responses generated by digital agents, compromising the reliability of the evaluations. Consequently, the administration of self-reported questionnaires to virtual agents often yields unreliable results [24].

To address these challenges, efforts have been made to adapt self-report questionnaires originally designed for humans into formats suitable for assessing personality traits in digital agents [42]. One significant contribution is the Language Model Linguistic Personality Assessment (LMLPA) [94]. This system adapts the Big Five Inventory — a cornerstone in psychometric research [36] — by aligning its format with the operational capabilities of LLMs. Incorporating AI raters and linguistic prompts, the LMLPA quantitatively evaluates personality traits through the linguistic outputs of LLMs, offering a novel, domain-specific approach. Although promising, the methodology requires further empirical validation to ensure its robustness and applicability across diverse scenarios [94].

Another widely employed quantitative evaluation method for assessing personality traits in LLM-based digital agents is the Linguistic Inquiry and Word Count (LIWC) software [59]. LIWC analyzes text by mapping words and phrases to a curated dictionary, categorizing them into dimensions reflecting psychological, emotional, and social constructs [74]. Its latest version, LIWC-22, includes over 12,000 words and expressions across 117 categories, such as personal pronouns, emotion-related vocabulary, and cognitive indicators [3].

Widely applied in psychology [78], and communication research [26], LIWC is a valuable tool for studying personality traits in digital agents [74]. In LLM-based digital agents, LIWC has been applied to assess alignment [21], personality retention, and emotional embedding [79]. These applications highlight LIWC's ability to reveal subtle patterns and provide meaningful insights into how LLMs convey and replicate personality traits through language.

While qualitative methods are effective for analyzing personality traits in text-based agents, they are less suited for evaluating digital humans in VR environments. These agents express personality through a combination of verbal outputs and non-verbal behaviors, such as gestures, facial expressions, and vocal tone. This multimodal nature introduces complexities that demand updated evaluation methods to fully capture their dynamics. Embodied conversational agents, for example, offer a heightened sense of social presence and interaction, resembling face-to-face communication [70]. Solely relying on textual analysis neglects critical non-verbal cues, such as gestures and intonation, which can either reinforce or contradict verbal outputs.

The systematic assessment of personality traits in multimodal contexts remains underexplored. Current methodologies are not yet equipped to address the challenges posed by these environments, where verbal and non-verbal communication intersect. Future research should prioritize developing comprehensive evaluation frameworks that integrate facial expressions, body language, and vocal nuances with traditional textual analysis. Advancing such methodologies will enhance confidence in applying LLMs to digital humans, enabling more natural and effective interactions in VR environments.

5 FUTURE

Despite rapid advancements in the field of digital humans, the use of cutting-edge AI technologies, particularly LLMs, to simulate human identity and personality in realistic, non-chat-based scenarios remains underexplored. The capabilities of generative models to mimic human behavior present a promising avenue for applications across diverse domains. For instance, LLMs can enable adaptive personalities in NPC characters, allowing non-scripted, context-driven conversations. These NPCs could maintain both short-term and long-term memory during interactions, enriching gaming experiences with more dynamic and immersive storytelling [96]. Beyond entertainment, educational applications of such agents have shown potential in improving learning outcomes, such as enhancing user satisfaction and reducing the time required to master new concepts [23, 75].

An emerging research area involves leveraging the generative capabilities of LLMs to produce text that not only conveys conversational content but also encodes directives for corresponding visual actions. By integrating emotion recognition and generation, these systems could embed instructions for digital humans to perform specific facial expressions or body movements aligned with the intended emotional tone of the interaction. This approach enables the generated text to act as both a communication medium and a control mechanism for non-verbal expressions, enhancing the realism and interactivity of virtual environments [28]. Such multimodal integration bridges the gap between verbal and non-verbal communication, creating a more seamless and human-like interaction experience.

However, while LLMs exhibit remarkable capacities for simulating and adapting human personalities, they are hindered by significant computational demands. The high inference time associated with auto-regressive generation methods translates into latency challenges, requiring substantial GPU infrastructure to ensure real-time responses [90, 43]. On the other hand, recent research on small LLMs (SLLMs) suggests that these models can perform tasks previously reserved for larger models [11]. Although still in its initial research, this field holds potential for addressing the computational challenges of LLMs, enabling the development of lightweight systems capable of delivering personality simulation and emotion-driven expressions in real time. Further research is needed to evaluate the effectiveness of SLLMs in these complex tasks, particularly in scenarios requiring both verbal and non-verbal adaptability.

6 CONCLUSION

The integration of large language models with virtual reality technologies represents a transformative step in enhancing the realism and interactivity of digital humans. By combining the verbal capabilities of LLMs with non-verbal cues such as facial expressions and gestures, these systems offer the potential to create more engaging, lifelike interactions. Despite significant advancements, this domain remains underexplored, particularly in adapting LLM-driven personality traits to human-like computer-driven characters in immersive VR environments. Addressing challenges such as computational demands, latency, and the absence of standardized evaluation frameworks is critical for realizing the full potential of these technologies.

This paper has reviewed methods for enabling digital humans to adopt nuanced personalities, highlighting the contributions of both rule-based approaches and advanced LLM techniques, including zero-shot, few-shot, and fine-tuning strategies. It also underscored the importance of multimodal integration for fostering richer user experiences and explored the opportunities and limitations of leveraging emerging technologies such as small LLMs to overcome computational constraints.

Our findings provide a foundation for further exploration, encouraging interdisciplinary collaboration between natural language processing, virtual reality, and human-computer interaction to redefine how we interact with digital humans in immersive settings.

REFERENCES

[1] B. Abeysinghe and R. Circi. The challenges of evaluating llm applications: An analysis of automated, human, and llm-based approaches. June 13 2024. doi: 10.48550/arXiv.2406.03339 4

[2] E. Ayedoun, Y. Hayashi, and K. Seta. A conversational agent to encourage willingness to communicate in the context of english as a foreign language. *Procedia Computer Science*, 60:1433–1442, 12 2015. doi: 10.1016/j.procs.2015.08.219 2

[3] R. L. Boyd, A. Ashokkumar, S. Seraj, and J. W. Pennebaker. The development and psychometric properties of liwc-22. *Austin, TX: University of Texas at Austin*, 10, 2022. 5

[4] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, eds., *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901. Curran Associates, Inc., 2020. 1, 2

[5] Z. Cai, J. Jiang, Z. Qing, X. Guo, M. Zhang, Z. Lin, H. Mei, C. Wei, R. Wang, W. Yin, L. Pan, X. Fan, H. Du, P. Gao, Z. Yang, Y. Gao, J. Li, T. Ren, Y. Wei, X. Wang, C. C. Loy, L. Yang, and Z. Liu. Digital life project: Autonomous 3d characters with social intelligence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 582–592, June 2024. 3

[6] J. Cassell. Embodied conversational agents: Representation and intelligence in user interfaces. *AI Magazine*, 22(4):67, Dec. 2001. doi: 10.1609/aimag.v22i4.1593 2

[7] H. E. Cattell and A. D. Mead. The sixteen personality factor questionnaire (16pf). *The SAGE handbook of personality theory and assessment*, 2:135–159, 2008. 3

[8] R. B. Cattell and P. E. Kline. *The scientific analysis of personality and motivation.* Academic Press, 1977. 3

[9] G. H. Chen, S. Chen, Z. Liu, F. Jiang, and B. Wang. Humans or LLMs as the judge? a study on judgement bias. In Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, eds., *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 8301–8327. Association for Computational Linguistics, Miami, Florida, USA, Nov. 2024. doi: 10.18653/v1/2024.emnlp-main.474 4

[10] G. H. Chen, S. Chen, Z. Liu, F. Jiang, and B. Wang. Humans or llms as the judge? a study on judgement biases. *arXiv preprint arXiv:2402.10669*, 2024. 4

[11] L. Chen and G. Varoquaux. What is the role of small models in the llm era: A survey. *arXiv preprint arXiv:2409.06857*, 2024. 5

[12] N. Cherakara, F. Varghese, S. Shabana, N. Nelson, A. Karukayil, R. Kulothungan, M. A. Farhan, B. Nesset, M. Moujahid, T. Dinkar, et al. Furchat: an embodied conversational agent using llms, combining open and closed-domain dialogue with facial expressions. *arXiv preprint arXiv:2308.15214*, 2023. 3

[13] E. Clark, T. August, S. Serrano, N. Haduong, S. Gururangan, and N. A. Smith. All that's 'human' is not gold: Evaluating human evaluation of generated text. In C. Zong, F. Xia, W. Li, and R. Navigli, eds., *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 7282–7296. Association for Computational Linguistics, Online, Aug. 2021. doi: 10.18653/v1/2021.acl-long.565 4

[14] B. De Raad. *The big five personality factors: the psycholexical approach to personality.* Hogrefe & Huber Publishers, 2000. 3

[15] E. Diener and R. E. Lucas. Personality traits. *General psychology: Required reading*, 278, 2019. 3

[16] G. Dong, H. Yuan, K. Lu, C. Li, M. Xue, D. Liu, W. Wang, Z. Yuan, C. Zhou, and J. Zhou. How abilities in large language models are affected by supervised fine-tuning data composition. In L.-W. Ku, A. Martins, and V. Srikumar, eds., *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 177–198. Association for Computational Linguistics, Bangkok, Thailand, Aug. 2024. doi: 10.18653/v1/2024.acl-long.12 2

[17] Y. R. Dong, T. Hu, and N. Collier. Can LLM be a personalized judge? In Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, eds., *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 10126–10141. Association for Computational Linguistics, Miami, Florida, USA, Nov. 2024. doi: 10.18653/v1/2024.findings-emnlp.592 4

[18] C. J. Falconer, P. Cutting, E. Bethan Davies, C. Hollis, P. Stallard, and P. Moran. Adjunctive avatar therapy for mentalization-based treatment of borderline personality disorder: a mixed-methods feasibility study. *BMJ Ment Health*, 20(4):123–127, 2017. doi: 10.1136/eb-2017-102761 1

[19] K. Feng, K. Ding, K. Ma, Z. Wang, Q. Zhang, and H. Chen. Sample-efficient human evaluation of large language models via maximum discrepancy competition. *arXiv preprint arXiv:2404.08008*, 2024. 4

[20] D. W. Fiske. Consistency of the factorial structures of personality ratings from different sources. *The Journal of Abnormal and Social Psychology*, 44:329–344, 1949. doi: 10.1037/h0057198 3

[21] I. Frisch and M. Giulianelli. LLM agents in interaction: Measuring personality consistency and linguistic alignment in interacting populations of large language models. In A. Deshpande, E. Hwang, V. Murahari, J. S. Park, D. Yang, A. Sabharwal, K. Narasimhan, and A. Kalyan, eds., *Proceedings of the 1st Workshop on Personalization of Generative AI Systems (PERSONALIZE 2024)*, pp. 102–111. Association for Computational Linguistics, St. Julians, Malta, Mar. 2024. 5

[22] L. R. Goldberg, J. A. Johnson, H. W. Eber, R. Hogan, M. C. Ashton, C. R. Cloninger, and H. G. Gough. The international personality item pool and the future of public-domain personality measures. *Journal of Research in personality*, 40(1):84–96, 2006. 5

[23] F. Grivokostopoulou, K. Kovas, and I. Perikos. The effectiveness of embodied pedagogical agents and their impact on students learning in virtual worlds. *Applied Sciences*, 10(5), 2020. doi: 10.3390/app10051739 1, 5

[24] A. Gupta, X. Song, and G. Anumanchipalli. Self-assessment tests are unreliable measures of LLM personality. In Y. Belinkov, N. Kim, J. Jumelet, H. Mohebbi, A. Mueller, and H. Chen, eds., *Proceedings of the 7th BlackboxNLP Workshop: Analyzing and Interpreting Neural Networks for NLP*, pp. 301–314. Association for Computational Linguistics, Miami, Florida, US, Nov. 2024. doi: 10.18653/v1/2024.blackboxnlp-1.20 4, 5

[25] F. Guzmán, A. Abdelali, I. Temnikova, H. Sajjad, and S. Vogel. How do humans evaluate machine translation. In O. Bojar, R. Chatterjee, C. Federmann, B. Haddow, C. Hokamp, M. Huck, V. Logacheva, and P. Pecina, eds., *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pp. 457–466. Association for Computational Linguistics, Lisbon, Portugal, Sept. 2015. doi: 10.18653/v1/W15-3059 4

[26] E. M. Haber. On the stability of online language features: How much

text do you need to know a person? *arXiv preprint arXiv:1504.06391*, 2015. 5

[27] T. Hagendorff. Machine psychology: Investigating emergent capabilities and behavior in large language models using psychological methods. *arXiv preprint arXiv:2303.13988*, 1, 2023. 4

[28] M. Hasan, C. Ozel, S. Potter, and E. Hoque. Sapien: Affective virtual agents powered by large language models*. In *2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pp. 1–3, 2023. doi: 10.1109/ACIIW59127.2023.10388188 3, 5

[29] J. B. Hirsh and J. B. Peterson. Personality and language use in self-narratives. *Journal of research in personality*, 43(3):524–527, 2009. 3

[30] B. Hu, L. Ma, W. Zhang, G. Salvendy, D. Chablat, and F. Bennis. Predicting real-world ergonomic measurements by simulation in a virtual environment. *International Journal of Industrial Ergonomics*, 41(1):64–71, 2011. doi: 10.1016/j.ergon.2010.10.001 2

[31] M. Huang, X. Zhang, C. Soto, and J. Evans. Designing llm-agents with personalities: A psychometric approach. *arXiv preprint arXiv:2410.19238*, 2024. 4

[32] Y. Huang. Orca: Enhancing role-playing abilities of large language models by integrating personality traits. *arXiv preprint arXiv:2411.10006*, 2024. 3

[33] N. Hube, M. Reinelt, K. Vidackovic, et al. A study on the influence of situations on personal avatar characteristics. *Vis. Comput. Ind. Biomed. Art*, 7(1):23, 2024. doi: 10.1186/s42492-024-00174-7 1

[34] G. Jiang, M. Xu, S.-C. Zhu, W. Han, C. Zhang, and Y. Zhu. Evaluating and inducing personality in pre-trained language models. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23. Curran Associates Inc., Red Hook, NY, USA, 2024. 5

[35] H. Jiang, X. Zhang, X. Cao, C. Breazeal, D. Roy, and J. Kabbara. PersonaLLM: Investigating the ability of large language models to express personality traits. In K. Duh, H. Gomez, and S. Bethard, eds., *Findings of the Association for Computational Linguistics: NAACL 2024*, pp. 3605–3627. Association for Computational Linguistics, Mexico City, Mexico, June 2024. doi: 10.18653/v1/2024.findings-naacl.229 3, 4

[36] O. P. John, E. M. Donahue, and R. L. Kentle. Big five inventory. *Journal of personality and social psychology*, 1991. 4, 5

[37] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020. 2

[38] S. R. Karra, S. T. Nguyen, and T. Tulabandhula. Estimating the personality of white-box language models. *arXiv preprint arXiv:2204.12000*, 2022. 4

[39] L. J. Klinkert, S. Buongiorno, and C. Clark. Driving generative agents with their personality. *arXiv preprint arXiv:2402.14879*, 2024. 3

[40] L. O. Kroczek, A. May, S. Hettenkofer, A. Ruider, B. Ludwig, and A. Mühlberger. The influence of persona and conversational task on social interactions with a llm-controlled embodied conversational agent. *arXiv preprint arXiv:2411.05653*, 2024. 1, 3, 4

[41] C. H. Lee, K. Kim, Y. S. Seo, and C. K. Chung. The relations between personality and language use. *The Journal of general psychology*, 134(4):405–413, 2007. 3

[42] S. Lee, S. Lim, S. Han, G. Oh, H. Chae, J. Chung, M. Kim, B.-w. Kwak, Y. Lee, D. Lee, et al. Do llms have distinct and consistent personality? trait: Personality testset designed for llms with psychometrics. *arXiv preprint arXiv:2406.14703*, 2024. 4, 5

[43] B. Li, Y. Jiang, V. Gadepally, and D. Tiwari. Llm inference serving: Survey of recent advances and opportunities. *arXiv preprint arXiv:2407.12391*, 2024. 5

[44] D. Li, B. Jiang, L. Huang, A. Beigi, C. Zhao, Z. Tan, A. Bhattacharjee, Y. Jiang, C. Chen, T. Wu, et al. From generation to judgment: Opportunities and challenges of llm-as-a-judge. *arXiv preprint arXiv:2411.16594*, 2024. 4

[45] H. Li, Q. Dong, J. Chen, H. Su, Y. Zhou, Q. Ai, Z. Ye, and Y. Liu. Llms-as-judges: A comprehensive survey on llm-based evaluation methods. *arXiv preprint arXiv:2412.05579*, 2024. 4

[46] W. Li, J. Liu, A. Liu, X. Zhou, M. Diab, and M. Sap. Big5-chat: Shaping llm personalities through training on human-grounded data. *arXiv preprint arXiv:2410.16491*, 2024. 4

[47] Y. Li. A practical survey on zero-shot prompt design for in-context learning. In R. Mitkov and G. Angelova, eds., *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, pp. 641–647. INCOMA Ltd., Shoumen, Bulgaria, Varna, Bulgaria, Sept. 2023. 2

[48] R. Likert. A technique for the measurement of attitudes. *Archives of Psychology*, 1932. 5

[49] Y. Lin, L. Chen, A. Ali, C. Nugent, I. Cleland, R. Li, J. Ding, and H. Ning. Human digital twin: A survey. *Journal of Cloud Computing*, 13(1):131, 2024. 2

[50] M. Maslych, C. Pumarada, A. Ghasemaghaei, and J. J. LaViola Jr. Takeaways from applying llm capabilities to multiple conversational avatars in a vr pilot study. *arXiv preprint arXiv:2501.00168*, 2024. 1

[51] R. R. McCrae, P. T. Costa, Jr, and T. A. Martin. The neo–pi–3: A more readable revised neo personality inventory. *Journal of personality assessment*, 84(3):261–270, 2005. 5

[52] N. Mecklenburg, Y. Lin, X. Li, D. Holstein, L. Nunes, S. Malvar, B. Silva, R. Chandra, V. Aski, P. K. R. Yannam, et al. Injecting new knowledge into large language models via supervised fine-tuning. *arXiv preprint arXiv:2404.00213*, 2024. 2

[53] S. Mihai, M. Yaqoob, D. V. Hung, W. Davis, P. Towakel, M. Raza, M. Karamanoglu, B. Barn, D. Shetve, R. V. Prasad, H. Venkataraman, R. Trestian, and H. X. Nguyen. Digital twins: A survey on enabling technologies, challenges, trends and future prospects. *IEEE Communications Surveys & Tutorials*, 24(4):2255–2291, 2022. doi: 10.1109/COMST.2022.3208773 1

[54] I. B. Myers and M. H. McCaulley. *Myers-Briggs type indicator: MBTI*. Consulting Psychologists Press Palo Alto, 1988. 3

[55] D. Noever and S. Hyams. Ai text-to-behavior: A study in steerability. *arXiv preprint arXiv:2308.07326*, 2023. 3

[56] E. A. Oliveira, D. Silva, and A. G. Filho. Improving vr accessibility through automatic 360 scene description using multimodal large language models. In *Anais do XXVI Simpósio de Realidade Virtual e Aumentada*, pp. 289–293. SBC, Porto Alegre, RS, Brasil, 2024. 1

[57] K. Pan and Y. Zeng. Do llms possess a personality? making the mbti test an amazing evaluation for large language models. *arXiv preprint arXiv:2307.16180*, 2023. 3

[58] M. Pan, A. Kitson, H. Wan, and M. Prpa. Ellma-t: an embodied llm-agent for supporting english language learning in social vr. *arXiv preprint arXiv:2410.02406*, 2024. 2

[59] J. W. Pennebaker, M. E. Francis, and R. J. Booth. *Linguistic Inquiry and Word Count: LIWC*. Lawrence Erlbaum Associates, Mahwah, NJ, 2001. 5

[60] J. W. Pennebaker and A. Graybeal. Patterns of natural language use: Disclosure, personality, and social integration. *Current Directions in Psychological Science*, 10(3):90–93, 2001. 3

[61] J. W. Pennebaker and L. A. King. Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, 77(6):1296, 1999. 3

[62] S. Pilbeam, L. Allison, and S. White. Psychometric assessment and personality questionnaires. *Contemporary Themes in Strategic People Management: A Case-Based Approach*, p. 76, 2012. 4

[63] Y. Ran, X. Wang, R. Xu, X. Yuan, J. Liang, Y. Xiao, and D. Yang. Capturing minds, not just words: Enhancing role-playing language models with personality-indicative data. In Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, eds., *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 14566–14576. Association for Computational Linguistics, Miami, Florida, USA, Nov. 2024. doi: 10.18653/v1/2024.findings-emnlp.853 3

[64] H. Rao, C. Leung, and C. Miao. Can ChatGPT assess human personalities? a general evaluation framework. In H. Bouamor, J. Pino, and K. Bali, eds., *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 1184–1194. Association for Computational Linguistics, Singapore, Dec. 2023. doi: 10.18653/v1/2023.findings-emnlp.84 5

[65] L. S. Endres. Personality engineering: Applying human personality theory to the design of artificial personalities. In Y. Anzai, K. Ogawa,

and H. Mori, eds., *Symbiosis of Human and Artifact*, vol. 20 of *Advances in Human Factors/Ergonomics*, pp. 477–482. Elsevier, 1995. doi: 10.1016/S0921-2647(06)80262-5 2, 3

[66] D. F. Sacco and M. M. Ismail. Social belongingness satisfaction as a function of interaction medium: Face-to-face interactions facilitate greater social belonging and interaction enjoyment compared to instant messaging. *Computers in Human Behavior*, 36:359–364, 2014. doi: 10.1016/j.chb.2014.04.004 1, 2

[67] M. Safdari, G. Serapio-García, C. Crepy, S. Fitz, P. Romero, L. Sun, M. Abdulhai, A. Faust, and M. Matarić. Personality traits in large language models. *arXiv preprint arXiv:2307.00184*, 2023. 3, 5

[68] Y. Shao, L. Li, J. Dai, and X. Qiu. Character-LLM: A trainable agent for role-playing. In H. Bouamor, J. Pino, and K. Bali, eds., *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 13153–13187. Association for Computational Linguistics, Singapore, Dec. 2023. 1, 3

[69] P. V. S. Silva, A. R. S. Vitória, D. F. C. Silva, and A. R. G. a. Filho. Attention guidance through video script: A case study of object focusing on 360° vr video tours. In *Proceedings of the 26th Symposium on Virtual and Augmented Reality*, SVR '24, p. 247–251. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3691573.3691618 1

[70] H. J. Smith and M. Neff. Communication behavior in embodied virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173863 5

[71] X. Song, Y. Adachi, J. Feng, M. Lin, L. Yu, F. Li, A. Gupta, G. Anumanchipalli, and S. Kaur. Identifying multiple personalities in large language models with external evaluation. *arXiv preprint arXiv:2402.14805*, 2024. 4

[72] X. Song, A. Gupta, K. Mohebbizadeh, S. Hu, and A. Singh. Have large language models developed a personality?: Applicability of self-assessment tests in measuring personality in llms. *arXiv preprint arXiv:2305.14693*, 2023. 4

[73] S. Sonlu, B. Bendiksen, F. Durupinar, and U. Güdükbay. The effects of embodiment and personality expression on learning in llm-based educational agents. *arXiv preprint arXiv:2407.10993*, 2024. 2

[74] Y. R. Tausczik and J. W. Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54, 2010. 4, 5

[75] M. Taylor, Z. Muwaffak, M. Penny, B. Szulc, S. Brown, A. Merritt, and S. Hilton. The rise of the ai scientist: Unleashing the potential of chat-gpt powered avatars in virtual reality digital-twin laboratories. *ChemRxiv*, 2023. 1, 5

[76] S. Thomas, Y. Ferstl, R. McDonnell, and C. Ennis. Investigating how speech and animation realism influence the perceived personality of virtual characters and agents. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 11–20, 2022. doi: 10.1109/VR51125.2022.00018 1

[77] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023. 1, 2

[78] W. Tov, K. L. Ng, H. Lin, and L. Qiu. Detecting well-being via computerized content analysis of brief diary entries. *Psychological assessment*, 25(4):1069, 2013. 5

[79] N. I. Tripto, S. Venkatraman, D. Macko, R. Moro, I. Srba, A. Uchendu, T. Le, and D. Lee. A ship of theseus: Curious cases of paraphrasing in LLM-generated texts. In L.-W. Ku, A. Martins, and V. Srikumar, eds., *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 6608–6625. Association for Computational Linguistics, Bangkok, Thailand, Aug. 2024. doi: 10.18653/v1/2024.acl-long.357 5

[80] Y.-M. Tseng, Y.-C. Huang, T.-Y. Hsiao, W.-L. Chen, C.-W. Huang, Y. Meng, and Y.-N. Chen. Two tales of persona in LLMs: A survey of role-playing and personalization. In Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, eds., *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 16612–16631. Association for Computational Linguistics, Miami, Florida, USA, Nov. 2024. doi: 10.18653/v1/2024

.findings-emnlp.969 1, 4

[81] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds., *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc., 2017. 2

[82] H. Vu, H. A. Nguyen, A. V. Ganesan, S. Juhng, O. N. Kjell, J. Sedoc, M. L. Kern, R. L. Boyd, L. Ungar, H. A. Schwartz, et al. Psychadapter: Adapting llm transformers to reflect traits, personality and mental health. *arXiv preprint arXiv:2412.16882*, 2024. 4

[83] H. Wang, V. Gaddy, J. Beveridge, and F. Ortega. Building an emotionally responsive avatar with dynamic facial expressions in human—computer interactions. *Multimodal Technologies and Interaction*, 5:13, 03 2021. doi: 10.3390/mti5030013 3

[84] H. Wei, S. He, T. Xia, A. Wong, J. Lin, and M. Han. Systematic evaluation of llm-as-a-judge in llm alignment tasks: Explainable metrics and diverse prompt templates. *arXiv preprint arXiv:2408.13006*, 2024. 4

[85] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, E. H. Chi, T. Hashimoto, O. Vinyals, P. Liang, J. Dean, and W. Fedus. Emergent abilities of large language models. *Transactions on Machine Learning Research*, 2022. Survey Certification. 2

[86] I. B. Weiner and R. L. Greene. *Handbook of personality assessment*. John Wiley & Sons, 2017. 4

[87] J. Weizenbaum. Eliza—a computer program for the study of natural language communication between man and machine. *Commun. ACM*, 9(1):36–45, Jan. 1966. doi: 10.1145/365153.365168 2

[88] Z. Wen, Y. Yang, J. Cao, H. Sun, R. Yang, and S. Liu. Self-assessment, exhibition, and recognition: a review of personality in large language models. *arXiv preprint arXiv:2406.17624*, 2024. 4

[89] J. Wester, S. de Jong, H. Pohl, and N. van Berkel. Exploring people's perceptions of llm-generated advice. *Computers in Human Behavior: Artificial Humans*, 2(2):100072, 2024. doi: 10.1016/j.chbah.2024.100072 2

[90] H. Xia, Z. Zheng, Y. Li, D. Zhuang, Z. Zhou, X. Qiu, Y. Li, W. Lin, and S. L. Song. Flash-llm: Enabling cost-effective and highly-efficient large generative model inference with unstructured sparsity. *Proc. VLDB Endow.*, 17(2):211–224, Oct. 2023. doi: 10.14778/3626292.3626303 5

[91] S. Yang, S. Zhu, R. Bao, L. Liu, Y. Cheng, L. Hu, M. Li, and D. Wang. What makes your model a low-empathy or warmth person: Exploring the origins of personality in llms. *arXiv preprint arXiv:2410.10863*, 2024. 3

[92] J. Zhang, D. Liu, C. Qian, Z. Gan, Y. Liu, Y. Qiao, and J. Shao. The better angels of machine personality: How personality relates to llm safety. *arXiv preprint arXiv:2407.12344*, 2024. 4

[93] C. Zheng, H. Zhou, F. Meng, J. Zhou, and M. Huang. Large language models are not robust multiple choice selectors. In *The Twelfth International Conference on Learning Representations*, 2023. 5

[94] J. Zheng, X. Wang, S. Hosio, X. Xu, and L.-H. Lee. Lmlpa: Language model linguistic personality assessment. *Computational Linguistics*, pp. 1–41, 2025. 5

[95] L. Zheng, W.-L. Chiang, Y. Sheng, S. Zhuang, Z. Wu, Y. Zhuang, Z. Lin, Z. Li, D. Li, E. Xing, et al. Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in Neural Information Processing Systems*, 36:46595–46623, 2023. 4

[96] S. Zheng, K. He, L. Yang, and J. Xiong. Memoryrepository for ai npc. *IEEE Access*, 12:62581–62596, 2024. doi: 10.1109/ACCESS.2024.3393485 5

[97] L. Zhu, X. Wang, and X. Wang. JudgeLM: Fine-tuned large language models are scalable judges. In *The Thirteenth International Conference on Learning Representations*, 2025. 4

[98] M. Zhu, Y. Weng, L. Yang, and Y. Zhang. Personality alignment of large language models. In *The Thirteenth International Conference on Learning Representations*, 2025. 3

[99] H. Zou, P. Wang, Z. Yan, T. Sun, and Z. Xiao. Can llm" self-report"?: Evaluating the validity of self-report scales in measuring personality design in llm-based chatbots. *arXiv preprint arXiv:2412.00207*, 2024. 4