PersonaVLM: Long-Term Personalized Multimodal LLMs

Chang Nie[1] Chaoyou Fu[1,♠] Yifan Zhang[2] Haihua Yang[2,†] Caifeng Shan[1]
[1]Nanjing University [2]ByteDance
changnie@smail.nju.edu.cn, bradyfu24@gmail.com

Abstract

*Multimodal Large Language Models (MLLMs) serve as daily assistants for millions. However, their ability to generate responses aligned with individual preferences remains limited. Prior approaches enable only static, single-turn personalization through input augmentation or output alignment, and thus fail to capture users' evolving preferences and personality over time (see Fig. 1). In this paper, we introduce **PersonaVLM**, an innovative personalized multimodal agent framework designed for long-term personalization. It transforms a general-purpose MLLM into a personalized assistant by integrating three key capabilities: (a) **Remembering**: It proactively extracts and summarizes chronological multimodal memories from interactions, consolidating them into a personalized database. (b) **Reasoning**: It conducts multi-turn reasoning by retrieving and integrating relevant memories from the database. (c) **Response Alignment**: It infers the user's evolving personality throughout long-term interactions to ensure outputs remain aligned with their unique characteristics. For evaluation, we establish **Persona-MME**, a comprehensive benchmark comprising over 2,000 curated interaction cases, designed to assess long-term MLLM personalization across seven key aspects and 14 fine-grained tasks. Extensive experiments validate our method's effectiveness, improving the baseline by 22.4% (Persona-MME) and 9.8% (PERSONAMEM) under a 128k context, while outperforming GPT-4o by 5.2% and 2.0%, respectively. Project page: https://PersonaVLM.github.io.*

1. Introduction

Multimodal Large Language Models (MLLMs) are increasingly integrated into the daily lives of millions of users [1, 46], serving as assistants, creative partners, and companions [19, 44, 47]. As their adoption grows, user expectations are shifting from general-purpose problem-solving towards personalized and empathetic long-term experiences [20, 42]. This shift poses a critical question: ***How can we evolve a general MLLM into a truly personalized assistant that accurately infers user intent, dynamically aligns its behavior with individual preferences and personality, and persistently remembers user-specific multimodal information over time?*** Addressing this question not only enhances user satisfaction and trust but also unlocks the significant value of MLLMs in domains like recommendation [38], healthcare [3], and education [48], to name a few.

Even advanced proprietary models exhibit limited capabilities in generating responses that cater to a user's unique preferences and characteristics [6, 14, 50]. This challenge stems from two primary factors: on the model side, they are predominantly optimized within fixed windows and a one-size-fits-all paradigm [21]; on the user side, an individual's preferences and personality are inherently diverse and dynamic, continuously evolving throughout ongoing interactions [14]. As illustrated in Fig. 1, a user initially expresses a preference for *Sprite* but subsequently shifts to *Coca-Cola* to mitigate anxiety in a multimodal interaction. When the user later expresses stress, a retrieval-augmented response fails to capture this shift, resulting in a misaligned recommendation. Furthermore, a generic aligned response may feel overly extraverted, failing to accommodate the introverted and neurotic user whose personality traits are often revealed subtly across many unrelated dialogues.

The root of these failures is that current personalization strategies are designed for static interactions. Specifically, input augmentation-based MLLMs like Yo'LLaVA [28] and RAP [11] specialize in recognizing user-specific concepts, but lack mechanisms to manage or update these memories, consequently failing to capture preference shifts from *Sprite* to *Coca-Cola*. Similarly, alignment techniques such as ALIGNXPERT [21] and Personality-Activation Search (PAS) [52] presuppose static user traits, preventing them from adapting to a user's introversion revealed contextually over time. Therefore, we identify two foundational pillars for effective long-term personalization: (i) ***Personalized Memory Architecture***. The ability to proactively construct and manage a dynamic, user-centric multimodal database. (ii) ***Memory Utilization and Response Alignment***. The capacity

♠Corresponding author.
†Project leader.

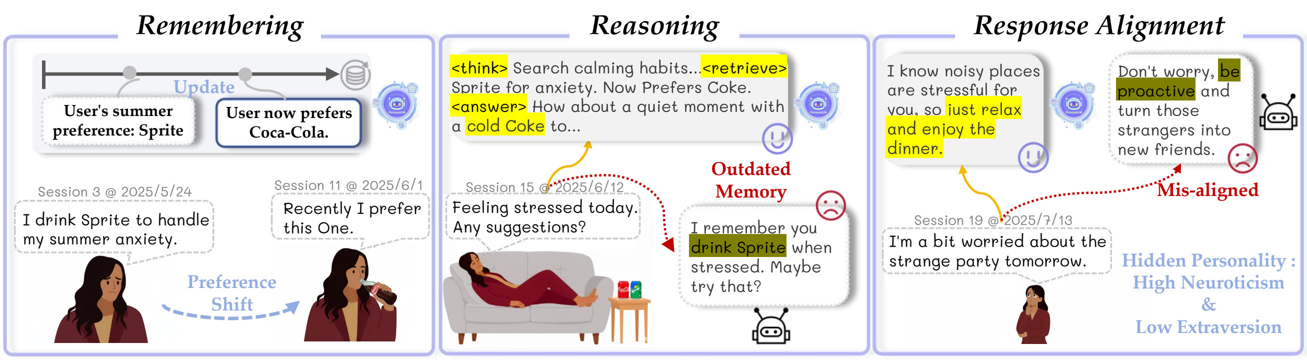


Figure 1. Illustration of PersonaVLM's three core capabilities for long-term personalization. PersonaVLM proactively remembers user preference shifts, performs multi-turn reasoning with retrieval, and generates responses aligned with the user's personality. In contrast, existing personalization strategies, such as input augmentation and output alignment, will result in poor recommendations based on outdated memories and replies that are misaligned with the user's personality.

to effectively utilize this database, employing reasoning and retrieval to generate responses that are deeply aligned with the user's unique and evolving characteristics.

Building on these pillars, we propose **PersonaVLM**, an innovative agent framework for long-term personalized interaction. First, we design a memory architecture that integrates a user personality profile and four distinct memory types (*core* for foundational attributes, *semantic* for facts, *procedural* for habits, and *episodic* for events) to store and manage user information. Second, building upon this architecture, a two-stage collaborative process transforms a general MLLM into a personalized assistant: (1) Response stage: Given the user's multimodal input and context, PersonaVLM autonomously performs multi-step reasoning and memory retrieval to generate a response aligned with the user's personality. (2) Update stage: The model infers and updates the user's latent traits, quantified as Big Five scores[1], through a momentum-based Personality Evolving Mechanism (PEM). Concurrently, it proactively extracts and summarizes key knowledge from the dialogue, updating the four memory types for future use. This integrated design endows PersonaVLM with the three key capabilities shown in Fig. 1.

Alongside the design of the framework, we address the scarcity of suitable training data by developing a synthesis pipeline to generate a large-scale personalized, multimodal interactive dataset, comprising over 30k interactions across 500 unique personas. This self-contained dataset enables effective training while ensuring PersonaVLM can operate locally, thereby eliminating data privacy concerns. Furthermore, recognizing that existing benchmarks [24] are often static and text-centric, we establish **Persona-MME**, a comprehensive benchmark designed to evaluate the long-term, multi-faceted, and multimodal personalization of MLLMs.

[1]We represent user personality using the Big Five traits [35]: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN), with each trait scored from 1 to 5.

In summary, our contributions are fourfold:

- We propose PersonaVLM, an innovative agent framework that achieves long-term personalization for MLLMs by integrating three core capabilities: proactive ***R****emembering*, multi-step ***R****easoning*, and ***R****esponse Alignment*.
- We introduce a personalized memory architecture featuring two key components: the PEM for dynamic alignment and a multi-type memory database comprising core, procedural, semantic, and episodic memories.
- We establish Persona-MME, a comprehensive benchmark designed to evaluate the long-term and multi-faceted personalization capabilities of MLLMs, and use it to benchmark over 10 leading proprietary and open-source models.
- We conduct extensive experiments to validate the effectiveness of PersonaVLM. Under a 128k context, PersonaVLM achieves improvements of 22.4% on Persona-MME and 9.8% on PERSONAMEM [14]. Notably, it surpasses GPT-4o on these benchmarks and in open-ended evaluations.

2. Related Work

The recent surge in LLM development has catalyzed the emergence of powerful MLLMs like GPT-4o [12], LLaVA [23], and the Qwen series [5, 45], showcasing exceptional capabilities in various general-domain tasks [47]. However, to evolve into a true personal assistant, a model must transcend the "one-size-fits-all" paradigm and tailor responses to individual user knowledge and preferences [24, 49]. Existing efforts to address this challenge can be categorized into three primary streams: adaptation-based, augmentation-based, and alignment-based personalization.

Adaptation-based Personalization. Adaptation-based methods operate at the model level, encoding user-specific knowledge directly into trainable parameters through fine-tuning. Some works, for instance, employ parameter-efficient fine-tuning (PEFT) to adapt LLMs for individ-

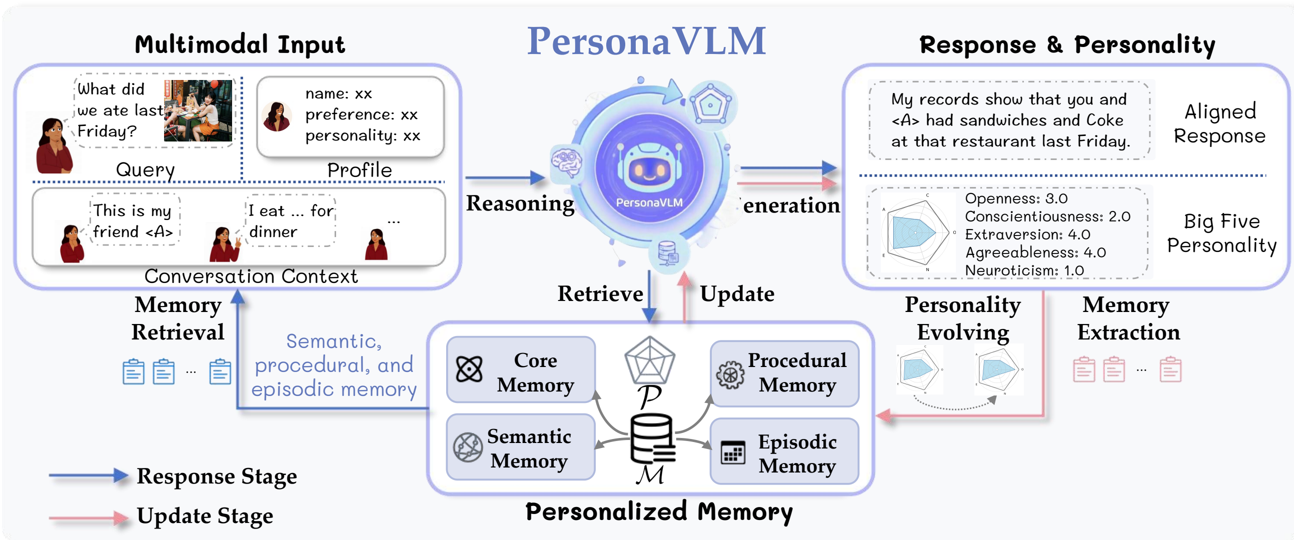


Figure 2. Overview of the PersonaVLM Framework. It leverages a personalized memory architecture and operates in two collaborative stages to achieve long-term personalization. In the Response Stage (blue arrows), it processes multimodal input, retrieves from personalized memory, and generates a personality-aligned response. Subsequently, in the Update Stage (pink arrows), the framework analyzes the completed interaction to extract key memories and update the user's evolving personality profile[1].

ual users or groups [37, 53]. This principle extends to the multimodal domain, where personalized MLLMs like MyVLM [2] and Yo'LLaVA [28] utilize learnable embeddings and soft prompts, respectively, to represent user-specific visual concepts. Such adaptation enables the model to transition from recognizing "a generic dog" to recognizing "the user's pet dog." However, their reliance on fine-tuning for each new user concept renders these methods less scalable and unable to capture the evolution of user preferences.

Augmentation-based Personalization. In contrast to model-level adaptation, augmentation-based approaches operate at the input level by equipping models with an external database to retain and retrieve user-specific memories [39, 41]. This paradigm is pivotal for transcending the limitations of fixed context windows in lifelong dialogues [7]. Related approaches [11, 29] extend personalization to the multimodal domain. They first employ open-vocabulary object detectors [25] to crop predefined visual concepts from images, which are then used for subsequent matching and retrieval. A key advantage of these methods is their training-free nature[2], allowing them to accommodate new user concepts at inference time. However, they are limited by a manually predefined database and lack mechanisms to proactively manage and update knowledge from dynamic interactions. Moreover, while general-purpose memory architectures like A-Mem [43] and Memory OS [22] employ more sophisticated agentic frameworks, their utility in our context is severely constrained. Their primary focus on text-only data limits their applicability to truly multimodal inputs, and their reliance on proprietary models creates barriers for open research and raises significant privacy concerns.

Alignment-based Personalization. While standard LLM alignment, such as Reinforcement Learning from Human Feedback (RLHF) [30], enforces a universal, "one-size-fits-all" behavioral standard, it inherently fails to accommodate diverse user preferences and communication styles. As shown in Fig. 1 (right), an overly enthusiastic response, while generally helpful, might be inappropriate for an introverted user experiencing anxiety. Personalized alignment directly tackles this limitation by redefining the optimization objective from a universal standard to a user-specific one [24]. For example, Li et al. [21] incorporate user features into the input and use methods such as Direct Preference Optimization (DPO) [34] to align model responses with predefined user values. Another strategy, PAS [52], trains user-specific "probes" to guide personalization at inference time. While this approach enables inference-time adaptation, it is fundamentally limited. Its reliance on per-user training poses significant scalability challenges; moreover, the static nature of these probes means the alignment can become outdated as the user's personality evolves over long-term interactions.

Departing from prior works that address siloed aspects of personalization for MLLMs, such as static memory or fixed alignment, we introduce PersonaVLM: a unified agent framework designed for dynamic, long-term interaction.

[2]Following the specific terminology from [32], this denotes that new user concepts can be accommodated at inference time without requiring continual fine-tuning.

3. Methods

3.1. PersonaVLM Framework

The overall architecture of the PersonaVLM agent is illustrated in Fig. 2. It is built upon a personalized memory architecture and operates through two collaborative stages of **Response** and **Update** to enable long-term personalization.

Personalized Memory Architecture. This architecture is designed to construct and maintain a comprehensive, long-term user profile, storing two primary categories of information. First, it maintains a user personality profile ($\mathcal{P}$), which provides a quantitative representation of the user's personality as a vector of scores for the Big Five dimensions[3] (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism). Second, it features a multi-type memory database ($\mathcal{M}$) that captures a wide range of user-related knowledge. This timeline-based, agentic system supports flexible CRUD (create, read, update, delete) operations and is structured into four distinct memory types:

- **Core Memory:** Stores the user's fundamental attributes (e.g., `human` and `persona` blocks), inspired by MemGPT [31], and is dynamically updated to reflect their most current profile.
- **Semantic Memory:** Distills event-independent, abstract knowledge by extracting key entities, relationships, and multimodal concepts.
- **Episodic Memory:** Organizes raw dialogues into atomic, time-stamped events, each including a summary, dialogue turns, and keywords for efficient retrieval.
- **Procedural Memory:** Records user-centric plans, goals, and recurring behaviors or habits.

Regarding their storage and persistence, while episodic and semantic memories are stored chronologically, core and procedural memories, along with the personality profile, retain only their latest versions to ensure relevance. Our design overcomes the limitations of existing systems, making our memory architecture: (a) Self-contained, avoiding proprietary model dependencies; (b) Explicitly personalized, prioritizing user-centric knowledge; and (c) Multimodal support, enabling a more holistic user understanding. For details on our memory architecture, refer to Appendix A.

Response Stage. The objective of this stage is to generate an aligned response by performing multi-step reasoning and timeline-based retrieval. Formally, this process at turn m can be formulated as:

$$\mathcal{R}_m = R(\mathcal{Q}_m, \mathcal{C}_m, \mathcal{M}_{m-1}), \tag{1}$$

where $\mathcal{R}_m$ is the personalized response. This response is conditioned on three inputs: the current user query $\mathcal{Q}_m = (T_m, I_m, t_m)$, consisting of a text instruction T_m, an optional image I_m, and a timestamp t_m; the dialogue context[4] $\mathcal{C}_m = \{(\mathcal{Q}_i, \mathcal{R}_i) \mid 0 < i < m \text{ and } |t_i - t_m| \leq t_s\}$; and the state of the personalized memory database $\mathcal{M}_{m-1}$. As depicted in the left panel of Fig. 2, the implementation of Eq. (1) is structured as a multi-step interaction between the PersonaVLM agent and its memory system. In the initial step, the model is prompted with the user's instruction, context, and a consolidated profile (comprising the user's core memory and personality). The model then outputs a detailed reasoning process and an `action` result. If the model determines that the current information is insufficient, it outputs retrieval conditions within a predefined template, including the `time period` and `keywords` for searching. The agent then executes the retrieval process by first isolating memories within the inferred `time period` and then performing a parallel search across semantic, episodic, and procedural memory types. The top-k results from each type are collected and fed back to the model to initiate the next reasoning step. This iterative process continues for multiple rounds until the model outputs the final response $\mathcal{R}_m$.

Two key insights drive the design of this stage. First, user queries are often highly context-dependent and contain anaphora (e.g., "that thing we just talked about"), which renders direct semantic retrieval imprecise. In contrast, a multi-turn, agentic retrieval process typically yields more precise and efficient results [15, 26]. Second, while some memory mechanisms [22, 40] may leverage query rewriting [27] to improve retrieval accuracy, they overlook crucial temporal cues (e.g., "this morning"). Our design addresses these gaps by enabling the model to determine not just *what* to retrieve, but also *if* retrieval is necessary and from *when*.

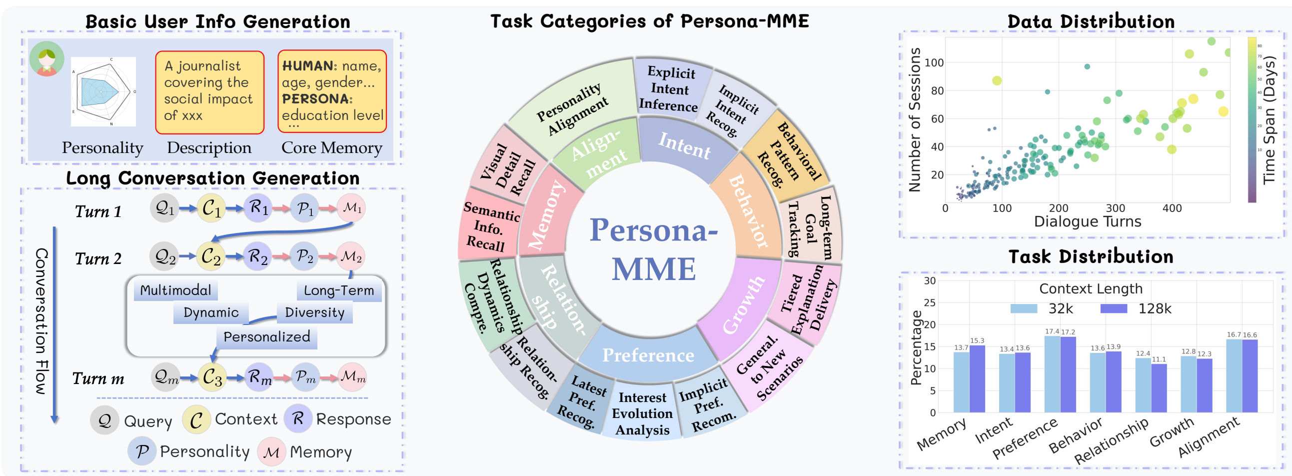
Update Stage. This stage, which executes automatically during idle periods after a response is generated, primarily involves two parts: evolving the user's personality profile and proactively updating the memories. This process at turn m can be represented as:

$$(\mathcal{P}_m, \mathcal{M}_m) = U(\mathcal{Q}_m, \mathcal{R}_m, \mathcal{M}_{m-1}). \tag{2}$$

Specifically, the user's personality profile, $\mathcal{P}_m$, is updated via our proposed **Personality Evolving Mechanism (PEM)**. The PEM maintains a long-term personality profile as a vector $\mathbf{p} \in \mathbb{R}^5$, corresponding to the Big Five dimensions [52]. At each turn m, the PEM first infers a temporary set of personality scores from the user's latest query, $\mathcal{Q}_m$. These scores are normalized to form a turn-specific personality vector, $\mathbf{p}'_m$. Subsequently, the long-term profile vector is updated using an exponential moving average (EMA): $\mathbf{p}_m \leftarrow \lambda \cdot \mathbf{p}_{m-1} + (1-\lambda) \cdot \mathbf{p}'_m$, where $\lambda \in [0,1]$ is a dynamic smoothing factor. To ensure high adaptability in

[3] Representing user personality via the Big Five traits is a prevalent approach in LLM alignment [52], rooted in psychological theories [16, 35].

[4] We treat the recent conversation history (within a $t_s = 60$ minute threshold) as short-term memory, and user inactivity beyond this threshold initiates a new session.



(a) Data Synthesis Pipeline **(b) Persona-MME Overview** **(c) Samples and Tasks Distribution**

Figure 3. Overview of our data synthesis pipeline and Persona-MME. (a) The pipeline first constructs rich user personas and then simulates long-term, dynamic conversations, generating both the dialogue and intermediate memories. (b) Persona-MME provides a comprehensive evaluation of personalization by assessing 14 fine-grained capabilities. (c) Statistics for Persona-MME, which includes two context length configurations ($32k$ and $128k$) and contains over 2,000 *in-situ*[6] cases.

early conversations while promoting stability over time, we employ a cosine decay schedule for λ. It starts with a low value (allowing rapid adaptation to initial user interactions) and gradually increases, making the profile more stable and less susceptible to minor fluctuations. Finally, the updated numerical vector $\mathbf{p}_m$ is converted back into a descriptive textual summary, $\mathcal{P}_m$, for use in the Response Stage.

Second, we selectively extract and update the four memory types, each with tailored logic. Semantic memory is updated after each turn, where key information such as user preferences, multimodal concepts, and explicit memorization requests is extracted and stored with timestamps and keywords. In contrast, core and procedural memory are updated at the end of each session; the agent analyzes the entire session's dialogue to perform automated CRUD operations and keep these memories current. Finally, episodic memory is constructed by segmenting dialogues into distinct topics, with each entry containing a summary, relevant keywords, and the specific dialogue turns involved. See Appendix B.1 for the complete implementation pipeline.

3.2. Training of PersonaVLM

We adopt Qwen2.5-VL-7B [5] as the backbone model for PersonaVLM and train it using a two-stage process.

Stage 1: Supervised Fine-Tuning (SFT). We perform SFT on a curated synthetic dataset of $78k$ samples to equip the model with foundational memory management and multi-turn reasoning skills. The training data is synthesized via a pipeline introduced in the next section and comprises two primary types: (a) examples for memory mechanisms, including personality inference and the four types of memory CRUD operations; and (b) QA pairs containing complete, multi-step reasoning trajectories constructed offline. After SFT, the model is capable of generating well-formed reasoning and retrieval actions, providing a strong cold-start initialization for the subsequent stage.

Stage 2: Reinforcement Learning (RL). This stage aims to further enhance the model's multi-turn reasoning capability. We employ Group Relative Policy Optimization (GRPO) [10], an improved PPO algorithm, to train the policy model π_θ. During generation, we enforce a strictly structured output format: the model must first output its reasoning process within `<think></think>` tags, followed by either retrieval conditions in `<retrieve></retrieve>` tags or the final response in `<answer></answer>` tags. For each training sample $\{\mathcal{Q}, \widehat{\mathcal{R}}\}$, where $\mathcal{Q}$ is the user input and $\widehat{\mathcal{R}}$ is the preferred response, a group of multi-turn trajectories $\{\tau_1, \ldots, \tau_G\}$ is sampled from the policy model. The reward for the i-th trajectory τ_i is calculated as:

$$r_i = f_{\text{acc}}(\widehat{\mathcal{R}}, \mathcal{R}_{\tau_i}) \cdot f_{\text{cons}}(\mathcal{Q}, \mathcal{R}_{\tau_i}) + 0.5 \cdot f_{\text{format}}(\mathcal{R}_{\tau_i}), \quad (3)$$

where f_{acc}, f_{cons}, and f_{format} are reward functions for accuracy, logical consistency between reasoning and the final answer, and format adherence, respectively. We use Qwen3-30B-A3B [45] as an *LLM-as-a-Judge* to compute f_{acc} and f_{cons} via zero-shot prompting. Following [10], the advantage for each trajectory is computed by standardizing its reward within the sampled group. During training, we cap the maximum number of retrieval attempts at three per trajectory, and the loss is computed exclusively on the generated tokens. Further details on the training data and implementation are provided in Appendix B.2.

4. Dataset and Persona-MME Construction

To enable both the implementation and evaluation of long-term dynamic personalization, we make two key contributions. First, to address the scarcity of high-quality training data, we construct a large-scale multimodal interaction dataset via a dedicated synthesis pipeline. Second, we establish Persona-MME, a comprehensive benchmark for evaluating personalization in multimodal settings. This dual effort is necessitated by existing datasets [21, 28], which are typically static, single-turn, or lack multimodal support.

Dataset Synthesis Pipeline. As illustrated in Fig. 3 (a), we design a synthesis pipeline to generate training data at *scale*. The process commences by sampling base personas from PersonaHub [9], which are then enriched with randomly assigned personality traits. This enrichment step generates a detailed role description and an initial user profile, forming the initial Core Memory. We employ Seed1.6-thinking[5] to generate conversations guided by a structured flow. This process is governed by several key principles: (1) **Long-term Dynamics**: Dialogues extend over hundreds of turns to simulate interactions spanning weeks or months. To capture this longitudinal evolution, we probabilistically induce dynamic shifts in user preferences, topics, and personality traits. (2) **Multimodality and Scenario Diversity**: Over 15% of dialogues incorporate multimodal elements. The interactions span a wide range of real-world scenarios, from professional tasks to casual conversations. (3) **Structured Supervision**: The generation process is guided to produce not only the conversational dialogue but also the intermediate reasoning, retrieval, and memorization steps. This explicit structure provides rich supervisory signals for training the PersonaVLM framework. Further details on the data distribution and validation process are provided in Appendix C.

Persona-MME: Evaluating Long-Term Personalization of MLLMs. Existing benchmarks focus on siloed aspects of personalization. For instance, PERSONAMEM [14] evaluates a model's ability to track a user's evolving profile, ALIGNX-test [21] is centered on static alignment, and others like Yo'LLaVA [11, 28] assess user-specific concept understanding. However, none provide a holistic evaluation across the critical dimensions of dynamic personalization.

To fill this void, we introduce Persona-MME, a comprehensive benchmark comprising over 2,000 *in-situ*[6] cases derived from 200 diverse personas. As depicted in Fig. 3 (b), Persona-MME is structured around seven core dimensions: **Memory, Intent, Preference, Behavior, Relationship, Growth, and Alignment**. Together, these dimensions

[5]Seed1.6-thinking is a commercial model with performance comparable to GPT-4o, selected for its balance of capability and cost-effectiveness.

[6]Queries are posed from the user's first-person perspective at a specific point in the conversational history, simulating a realistic interaction [14].

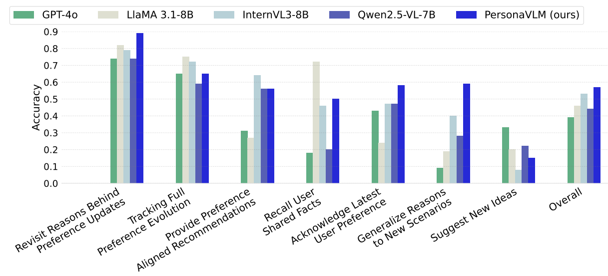


Figure 4. Quantitative evaluation across seven tasks on the PERSONAMEM ($32k$) benchmark.

encompass 14 fine-grained tasks, which are detailed in Table 5 in the Appendix. To accommodate different context lengths, we provide two evaluation configurations: a $32k$-context version for dialogues under 100 turns and a $128k$-context version for longer interactions, each containing cases from 100 distinct personas. Each test case comprises (1) a multiple-choice question assessing the model's personalized memory and understanding, and (2) an optional personality test evaluating its alignment. This multi-faceted structure enables Persona-MME to evaluate an MLLM's long-term personalization capabilities across diverse personas. Further details and statistics are provided in Appendix D.

5. Experiments

In this section, we present a series of quantitative and qualitative experiments designed to validate our PersonaVLM framework. The evaluation in the main paper is structured to answer the following research questions (RQs):

- **RQ1:** How effectively does PersonaVLM perform in personalized user understanding and memory recall?
- **RQ2:** Can PersonaVLM achieve effective alignment by capturing a user's evolving personality traits over time?
- **RQ3:** How well does PersonaVLM perform in personalized open-ended generation?

For comprehensive evaluations of Persona-MME, ablation studies about memory components, and further discussions, please refer to Appendices D, E, and F, respectively.

5.1. Personalized Understanding Evaluation

To evaluate personalized understanding (RQ1), we conduct experiments on two benchmarks: our Persona-MME and PERSONAMEM [14]. The latter includes seven task types specifically designed to assess a model's ability to track dynamic user preferences over the long term. We evaluate all models under two long-context settings ($32k$ and $128k$ tokens), with detailed results reported in Table 1 and Fig. 4. For comparison, we benchmark against several powerful models, including the proprietary GPT-4o [12] and strong open-source models such as Qwen2.5-VL-7B [5], LLaVA-OneVision-1.5-8B [4], and InternVL3-8B/38B [51]. See

Table 1. Evaluation on the Persona-MME and PERSONAMEM benchmarks, tested at context lengths of $32k$ and $128k$. We report accuracy (%) for Persona-MME (overall and across six aspects) and PERSONAMEM. The comparison includes two settings: full-context ("Full") and retrieval-augmented generation ("RAG"). Best results are shown in **bold**. The GPT-4o results on PERSONAMEM are from [14].

Context	Model	Persona-MME							PERSONAMEM
		Memory	Intent	Preference	Behavior	Relationship	Growth	Overall	
$32k$-Full	GPT-4o	**86.99**	**83.87**	**63.12**	57.14	71.30	73.87	**72.35**	39.20
	Qwen2.5-VL-7B	66.13	66.85	59.75	59.24	68.45	70.69	64.84	43.63
	InternVL3-8B	56.45	76.24	57.20	54.35	69.05	74.14	64.04	52.97
	InternVL3-38B	66.67	85.64	66.53	59.78	72.02	77.59	71.04	**57.93**
	OneVision-1.5-8B	74.19	74.59	60.59	53.26	72.62	74.14	67.76	52.80
$32k$-RAG	Qwen2.5-VL-7B	65.05	68.51	50.42	57.61	60.71	68.39	61.20	45.67
	$\text{PersonaVLM}_{\text{SFT}}$	67.20	70.17	49.58	57.07	70.24	80.46	$64.84_{+3.64}$	$52.12_{+6.45}$
	$\text{PersonaVLM}_{\text{RL}}$	69.89	76.80	58.05	**69.02**	**73.21**	**86.78**	$71.48_{+10.28}$	$56.53_{+10.86}$
$128k$-Full	GPT-4o	**84.44**	75.63	59.12	55.65	65.98	76.64	69.23	45.32
	Qwen2.5-VL-7B	50.60	54.73	52.41	54.30	55.83	60.90	54.48	3.08
	InternVL3-8B	57.23	68.92	53.48	54.97	69.17	76.69	62.43	36.62
	InternVL3-38B	67.47	71.62	**64.71**	58.94	65.00	76.69	67.18	46.56
	OneVision-1.5-8B	52.44	54.79	58.15	45.33	65.25	67.18	56.66	14.28
$128k$-RAG	Qwen2.5-VL-7B	56.63	63.51	50.27	55.63	61.67	70.68	59.01	37.88
	$\text{PersonaVLM}_{\text{SFT}}$	67.47	75.68	59.36	51.66	71.67	81.95	$67.18_{+8.17}$	$43.60_{+5.72}$
	$\text{PersonaVLM}_{\text{RL}}$	69.28	**77.70**	61.50	**60.26**	**75.00**	**87.97**	$\mathbf{71.05}_{+12.04}$	$\mathbf{47.28}_{+9.4}$

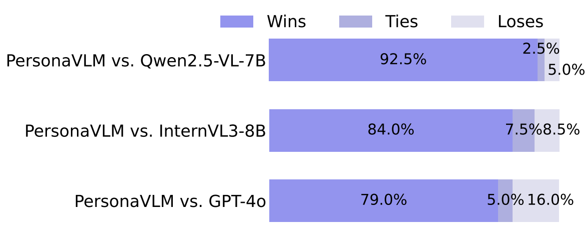


Figure 5. Qualitative comparison on open-ended generation, evaluated by Gemini-2.5-Pro. The evaluation assesses both the factual accuracy and the personality alignment of the responses.

Appendix Fig. 10 for more comparisons with leading models.

Compared to strong open-source models of a similar size, such as InternVL3-8B and LLaVA-OneVision-1.5-8B (provided with full context), PersonaVLM shows improvements of 8.62% and 14.39% on Persona-MME in the $128k$ setting, respectively. While the personalization capabilities of these open-source models appear to improve with scale, PersonaVLM still outperforms the much larger InternVL3-38B by 3.87% on Persona-MME ($128k$). We also evaluate Qwen2.5-VL-7B augmented with a straightforward RAG setup, which retrieves the top five most relevant messages following the approach of [14]. Interestingly, the results show that RAG can be detrimental in short-context scenarios—degrading performance on preference understanding tasks by as much as 9.33%—while providing a substantial boost of 4.53% in long-context settings. Additionally, as shown in Table 1, the two-stage training process demonstrates clear effectiveness, yielding an average improvement of 5.35% on Persona-MME.

When benchmarked against the proprietary GPT-4o, our method achieves competitive results on Persona-MME and demonstrates notable improvements of 17.3% and 2.0% on the $32k$ and $128k$ configurations of PERSONAMEM, respectively. Furthermore, while PersonaVLM's performance in memory recall lags behind that of GPT-4o with full context—a finding consistent with [18]—it demonstrates a significant advantage elsewhere. In particular, in Growth Modeling and Behavioral Awareness, PersonaVLM outperforms GPT-4o by over 10%.

Table 2. Evaluation of personalized alignment on the Persona-MME and P-SOUPS benchmarks.

Model	Persona-MME		P-SOUPS			
	$32k$	$128k$	Expertise	Informativeness	Style	Overall
Qwen2.5-VL-7B	69.91	52.27	39.00	49.34	23.00	37.11
InternVL3-8B	55.75	55.56	47.16	51.81	26.16	41.71
InternVL3-38B	64.60	63.01	**52.80**	53.30	32.83	46.32
Qwen3-30B-A3B	80.09	83.06	51.67	**54.60**	35.16	47.14
baseline with different strategies						
Self-Critic	59.73	57.66	39.67	49.33	23.67	37.50
Few-Shot	-	-	42.16	48.66	28.33	39.67
PersonaVLM (ours)	**89.16**	**92.22**	51.16	53.65	**44.00**	**49.60**

5.2. Personalized Alignment Evaluation

For RQ2, we conduct experiments on two benchmarks: the Alignment sub-task within Persona-MME and the P-SOUPS [13], which comprise 812 and 1,800 test cases, respectively. The former assesses a model's ability to determine if a response aligns with a user's personality inferred from the conversational context. The latter evaluates personality alignment with a given user profile across three dimensions: Expertise, Informativeness, and Style.

We quantitatively compare PersonaVLM against several powerful open-source models, including InternVL3-8B/38B

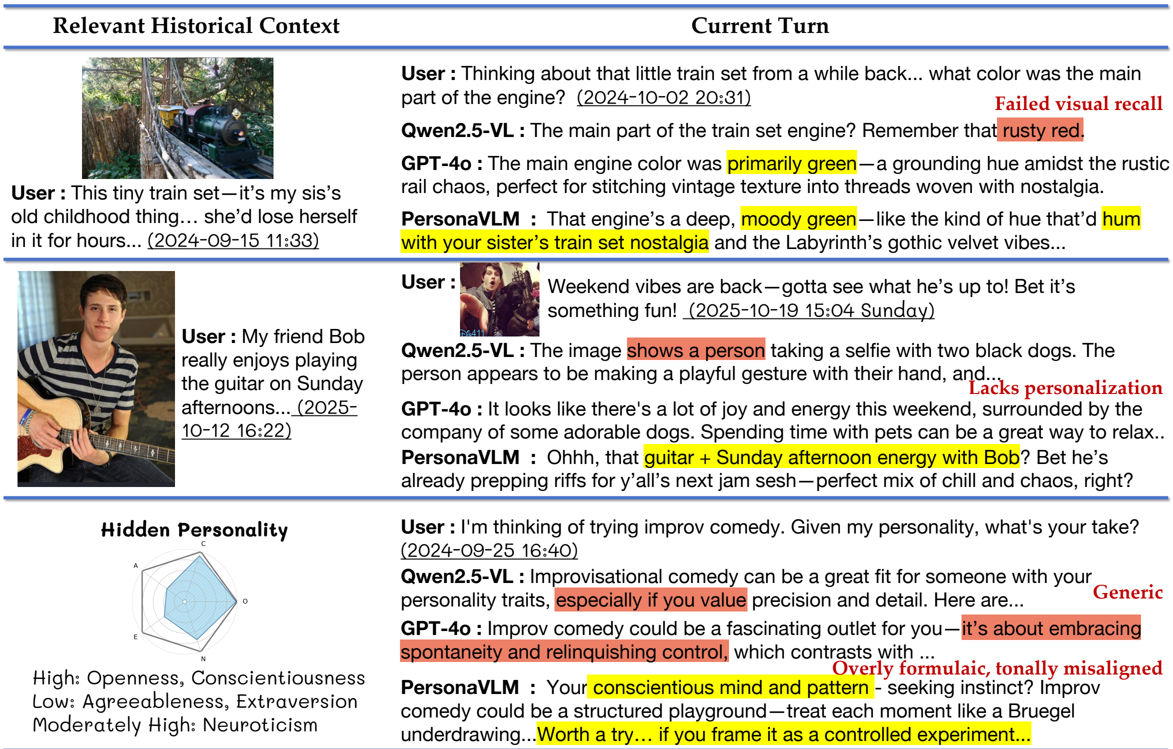


Figure 6. Qualitative comparison on open-ended generation tasks. Case studies demonstrate PersonaVLM's superior capabilities in memory recall, context integration, and personality alignment compared to the baseline and GPT-4o.

and Qwen3-30B-A3B [45], with the latter being noted for its strong language capabilities. We also evaluate the baseline model augmented with different strategies, such as Self-Critic and few-shot prompting [50]. As shown in Table 2, PersonaVLM consistently outperforms existing models on both benchmarks. Notably, it leads the next-best model by 9.16% on Persona-MME and 2.46% on P-SOUPS, representing a >12% gain over the baseline. Interestingly, language-centric models (e.g., Qwen3-30B-A3B) exhibit stronger alignment than multimodal counterparts like InternVL3-38B, with a 20% margin on Persona-MME (128k). These outcomes underscore PersonaVLM's capacity for robust personality alignment.

5.3. Qualitative Evaluation

To address RQ3 on open-ended generation, we conduct an automated evaluation using 200 questions randomly sampled from Persona-MME. We benchmark PersonaVLM against InternVL3-8B, Qwen2.5-VL-7B, and GPT-4o, employing Gemini-2.5-Pro [8] as an automated judge. Responses are assessed on two criteria: Accuracy and Personality Alignment, with PersonaVLM's performance in pairwise comparisons classified as a "win," "tie," or "loss." The evaluation prompt is provided in Fig. 23. As illustrated in Fig. 5, PersonaVLM achieves a substantially higher win rate than its peers. Particularly striking is its head-to-head performance against GPT-4o, where PersonaVLM secures a 79% win rate versus a 16% loss rate. This is further corroborated by qualitative case studies in Fig. 6, which showcase PersonaVLM's ability to perform accurate visual recall, integrate contextual memory, and maintain long-term personality alignment. In contrast, other models exhibit critical failures, such as memory hallucinations or tonally misaligned responses that ignore user-specific memories. These findings validate the generative capabilities of PersonaVLM for long-term personalization.

6. CONCLUSION

This paper introduces PersonaVLM, a novel agent framework that enables long-term, dynamic personalization for MLLMs by integrating three core capabilities: Remembering, Reasoning, and Response Alignment. To support rigorous evaluation, we further propose Persona-MME, a comprehensive benchmark for personalized multimodal understanding. Experiments show that PersonaVLM significantly enhances a model's personalization capabilities and consistently outperforms strong counterparts, including both proprietary GPT-4o and leading open-source alternatives. Our work provides a new paradigm for developing truly user-centric AI assistants, and future work will extend these capabilities toward a fully immersive multimodal experience.

Acknowledgements

This work is funded by National Natural Science Foundation of China (Grant No. 62506158 and No. 62441234), Basic Research Program of Jiangsu (BK20251183), and Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (JYB2025XDXM902).

References

[1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv:2303.08774*, 2023.

[2] Yuval Alaluf, Elad Richardson, Sergey Tulyakov, Kfir Aberman, and Daniel Cohen-Or. Myvlm: Personalizing vlms for user-specific queries. In *ECCV*, 2024.

[3] Rawan AlSaad, Alaa Abd-Alrazaq, Sabri Boughorbel, Arfan Ahmed, Max-Antoine Renault, Rafat Damseh, and Javaid Sheikh. Multimodal large language models in health care: applications, challenges, and future outlook. *Journal of medical Internet research*, 2024.

[4] Xiang An, Yin Xie, Kaicheng Yang, Wenkang Zhang, Xiuwei Zhao, Zheng Cheng, Yirui Wang, Songcen Xu, Changrui Chen, Chunsheng Wu, et al. Llava-onevision-1.5: Fully open framework for democratized multimodal training. *arXiv:2509.23661*, 2025.

[5] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. Qwen2. 5-vl technical report. *arXiv:2502.13923*, 2025.

[6] Jin Chen, Zheng Liu, Xu Huang, Chenwang Wu, Qi Liu, Gangwei Jiang, Yuanhao Pu, Yuxuan Lei, Xiaolong Chen, Xingmei Wang, et al. When large language models meet personalization: Perspectives of challenges and opportunities. *World Wide Web*, 2024.

[7] Prateek Chhikara, Dev Khant, Saket Aryan, Taranjeet Singh, and Deshraj Yadav. Mem0: Building production-ready ai agents with scalable long-term memory. *arXiv:2504.19413*, 2025.

[8] Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv:2507.06261*, 2025.

[9] Tao Ge, Xin Chan, Xiaoyang Wang, Dian Yu, Haitao Mi, and Dong Yu. Scaling synthetic data creation with 1,000,000,000 personas. *arXiv:2406.20094*, 2024.

[10] Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv:2501.12948*, 2025.

[11] Haoran Hao, Jiaming Han, Changsheng Li, Yu-Feng Li, and Xiangyu Yue. Rap: Retrieval-augmented personalization for multimodal large language models. In *CVPR*, 2025.

[12] Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. Gpt-4o system card. *arXiv:2410.21276*, 2024.

[13] Joel Jang, Seungone Kim, Bill Yuchen Lin, Yizhong Wang, Jack Hessel, Luke Zettlemoyer, Hannaneh Hajishirzi, Yejin Choi, and Prithviraj Ammanabrolu. Personalized soups: Personalized large language model alignment via post-hoc parameter merging. *arXiv:2310.11564*, 2023.

[14] Bowen Jiang, Zhuoqun Hao, Young-Min Cho, Bryan Li, Yuan Yuan, Sihao Chen, Lyle Ungar, Camillo J Taylor, and Dan Roth. Know me, respond to me: Benchmarking llms for dynamic user profiling and personalized responses at scale. *arXiv:2504.14225*, 2025.

[15] Bowen Jin, Hansi Zeng, Zhenrui Yue, Jinsung Yoon, Sercan Arik, Dong Wang, Hamed Zamani, and Jiawei Han. Search-r1: Training llms to reason and leverage search engines with reinforcement learning. *arXiv:2503.09516*, 2025.

[16] Oliver P John, Sanjay Srivastava, et al. The big-five trait taxonomy: History, measurement, and theoretical perspectives. 1999.

[17] Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, 2019.

[18] Jiazheng Kang, Mingming Ji, Zhe Zhao, and Ting Bai. Memory os of ai agent. 2025.

[19] Chunyuan Li, Zhe Gan, Zhengyuan Yang, Jianwei Yang, Linjie Li, Lijuan Wang, Jianfeng Gao, et al. Multimodal foundation models: From specialists to general-purpose assistants. *Foundations and Trends® in Computer Graphics and Vision*, 2024.

[20] Hao Li, Chenghao Yang, An Zhang, Yang Deng, Xiang Wang, and Tat-Seng Chua. Hello again! llm-powered personalized agent for long-term dialogue. *arXiv:2406.05925*, 2024.

[21] Jia-Nan Li, Jian Guan, Songhao Wu, Wei Wu, and Rui Yan. From 1,000,000 users to every user: Scaling up personalized preference for user-level alignment. *arXiv:2503.15463*, 2025.

[22] Zhiyu Li, Shichao Song, Chenyang Xi, Hanyu Wang, Chen Tang, Simin Niu, Ding Chen, Jiawei Yang, Chunyu Li, Qingchen Yu, et al. Memos: A memory os for ai system. *arXiv:2507.03724*, 2025.

[23] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. In *NeurIPS*, 2023.

[24] Jiahong Liu, Zexuan Qiu, Zhongyang Li, Quanyu Dai, Wenhao Yu, Jieming Zhu, Minda Hu, Menglin Yang, Tat-Seng Chua, and Irwin King. A survey of personalized large language models: Progress and future directions. *arXiv:2502.11528*, 2025.

[25] Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Qing Jiang, Chunyuan Li, Jianwei Yang, Hang Su, et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In *ECCV*, 2024.

[26] Lin Long, Yichen He, Wentao Ye, Yiyuan Pan, Yuan Lin, Hang Li, Junbo Zhao, and Wei Li. Seeing, listening, remembering, and reasoning: A multimodal agent with long-term memory. *arXiv:2508.09736*, 2025.

[27] Xinbei Ma, Yeyun Gong, Pengcheng He, Hai Zhao, and Nan Duan. Query rewriting in retrieval-augmented large language models. In *EMNLP*, 2023.

[28] Thao Nguyen, Haotian Liu, Yuheng Li, Mu Cai, Utkarsh Ojha, and Yong Jae Lee. Yo'llava: Your personalized language and vision assistant. In *NeurIPS*, 2024.

[29] Yeongtak Oh, Jisoo Mok, Dohyun Chung, Juhyeon Shin, Sangha Park, Johan Barthelemy, and Sungroh Yoon. Repic: Reinforced post-training for personalizing multi-modal language models. *arXiv:2506.18369*, 2025.

[30] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. In *NeurIPS*, 2022.

[31] Charles Packer, Vivian Fang, Shishir_G Patil, Kevin Lin, Sarah Wooders, and Joseph_E Gonzalez. Memgpt: Towards llms as operating systems. *arXiv:2310.08560*, 2023.

[32] Renjie Pi, Jianshu Zhang, Tianyang Han, Jipeng Zhang, Rui Pan, and Tong Zhang. Personalized visual instruction tuning. *arXiv:2410.07113*, 2024.

[33] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *ICML*, 2021.

[34] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. In *NeurIPS*, 2023.

[35] Sonia Roccas, Lilach Sagiv, Shalom H Schwartz, and Ariel Knafo. The big five personality factors and personal values. *Personality and social psychology bulletin*, 2002.

[36] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv:1707.06347*, 2017.

[37] Zhaoxuan Tan, Qingkai Zeng, Yijun Tian, Zheyuan Liu, Bing Yin, and Meng Jiang. Democratizing large language models via personalized parameter-efficient fine-tuning. *arXiv:2402.04401*, 2024.

[38] Qi Wang, Jindong Li, Shiqi Wang, Qianli Xing, Runliang Niu, He Kong, Rui Li, Guodong Long, Yi Chang, and Chengqi Zhang. Towards next-generation llm-based recommender systems: A survey and beyond. *arXiv:2410.19744*, 2024.

[39] Weizhi Wang, Li Dong, Hao Cheng, Xiaodong Liu, Xifeng Yan, Jianfeng Gao, and Furu Wei. Augmenting language models with long-term memory. In *NeurIPS*, 2023.

[40] Yu Wang and Xi Chen. Mirix: Multi-agent memory system for llm-based agents. *arXiv:2507.07957*, 2025.

[41] Jiale Wei, Xiang Ying, Tao Gao, Fangyi Bao, Felix Tao, and Jingbo Shang. Ai-native memory 2.0: Second me. *arXiv:2503.08102*, 2025.

[42] Junda Wu, Hanjia Lyu, Yu Xia, Zhehao Zhang, Joe Barrow, Ishita Kumar, Mehrnoosh Mirtaheri, Hongjie Chen, Ryan A Rossi, Franck Dernoncourt, et al. Personalized multimodal large language models: A survey. *arXiv:2412.02142*, 2024.

[43] Wujiang Xu, Kai Mei, Hang Gao, Juntao Tan, Zujie Liang, and Yongfeng Zhang. A-mem: Agentic memory for llm agents. *arXiv:2502.12110*, 2025.

[44] Zhenyu Xu, Hailin Xu, Zhouyang Lu, Yingying Zhao, Rui Zhu, Yujiang Wang, Mingzhi Dong, Yuhu Chang, Qin Lv, Robert P Dick, et al. Can large language models be good companions? an llm-based eyewear system with conversational common ground. In *IMWUT*, 2024.

[45] An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv:2505.09388*, 2025.

[46] Yuan Yao, Tianyu Yu, Ao Zhang, Chongyi Wang, Junbo Cui, Hongji Zhu, Tianchi Cai, Haoyu Li, Weilin Zhao, Zhihui He, et al. Minicpm-v: A gpt-4v level mllm on your phone. *arXiv:2408.01800*, 2024.

[47] Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. A survey on multimodal large language models. *National Science Review*, 2024.

[48] Jifan Yu, Zheyuan Zhang, Daniel Zhang-li, Shangqing Tu, Zhanxin Hao, Rui Miao Li, Haoxuan Li, Yuanchun Wang, Hanming Li, Linlu Gong, et al. From mooc to maic: Reshaping online teaching and learning through llm-driven agents. *arXiv:2409.03512*, 2024.

[49] Zhehao Zhang, Ryan A Rossi, Branislav Kveton, Yijia Shao, Diyi Yang, Hamed Zamani, Franck Dernoncourt, Joe Barrow, Tong Yu, Sungchul Kim, et al. Personalization of large language models: A survey. *arXiv:2411.00027*, 2024.

[50] Siyan Zhao, Mingyi Hong, Yang Liu, Devamanyu Hazarika, and Kaixiang Lin. Do llms recognize your preferences? evaluating personalized preference following in llms. *arXiv:2502.09597*, 2025.

[51] Jinguo Zhu, Weiyun Wang, Zhe Chen, Zhaoyang Liu, Shenglong Ye, Lixin Gu, Hao Tian, Yuchen Duan, Weijie Su, Jie Shao, et al. Internvl3: Exploring advanced training and test-time recipes for open-source multimodal models. *arXiv:2504.10479*, 2025.

[52] Minjun Zhu, Yixuan Weng, Linyi Yang, and Yue Zhang. Personality alignment of large language models. In *ICLR*, 2025.

[53] Yuchen Zhuang, Haotian Sun, Yue Yu, Rushi Qiang, Qifan Wang, Chao Zhang, and Bo Dai. Hydra: Model factorization framework for black-box llm personalization. In *NeurIPS*, 2024.