SimVBG: Simulating Individual Values by Backstory Generation

Bangde Du[1] **Ziyi Ye**[2*] **Zhijing Wu**[3] **Monika Jankowska**[4]
Shuqi Zhu[1] **Qingyao Ai**[1*] **Yujia Zhou**[1] **Yiqun Liu**[1]
[1]Department of Computer Science and Technology, Tsinghua University, Beijing, China
[2]Institute of Trustworthy Embodied AI, Fudan University, Shanghai, China
[3]Beijing Institute of Technology, Beijing, China
[4]Rice University, Houston, United States

Abstract

As Large Language Models (LLMs) demonstrate increasingly strong human-like capabilities, the need to align them with human values has become significant. Recent advanced techniques, such as prompt learning and reinforcement learning, are being employed to bring LLMs closer to aligning with human values. While these techniques address broad ethical and helpfulness concerns, they rarely consider simulating individualized human values. To bridge this gap, we propose SIMVBG, a framework that simulates individual values based on individual backstories that reflect their past experience and demographic information. SIMVBG transforms structured data on an individual to a backstory and utilizes a multi-module architecture inspired by the Cognitive–Affective Personality System to simulate individual value based on the backstories. We test SIMVBG on a self-constructed benchmark derived from the World Values Survey and show that SIMVBG improves top-1 accuracy by more than 10% over the retrieval-augmented generation method. Further analysis shows that performance increases as additional interaction user history becomes available, indicating that the model can refine its persona over time. Code, dataset, and complete experimental results are available at https://github.com/bangdedadi/SimVBG.

1 Introduction

Large language models (LLMs) have demonstrated strong capabilities to simulate humans' behaviors and cognition (Brown et al., 2020; Touvron et al., 2023). Such human simulation created unprecedented opportunities to build agents with distinct personalities reflecting real people or constituting realistic synthetic personas. These LLM-empowered agents have multiple uses, encompassing social science and behavioral research simulations (Park et al., 2023; Aher et al., 2023), and serving as personalized assistants, tutors, or partners. To create such agents, existing studies have tried to infuse LLMs with the knowledge and understanding of the individuals whom those agents are expected to mimic or interact with. For example, Wang et al. (2023c) proposes to train an LLM with reinforcement learning to mimic the behaviors and preferences of various population groups. Tu et al. (2023); Wang et al. (2023a,b) adopt pre-generated persona-based information profiles to prompt an LLM to act as different characters.

As LLMs exhibit such human-like and simulation capabilities, the challenge of aligning them with human values becomes critically important. Conventionally, LLMs are aligned with general human values such as helpfulness, honesty, harmlessness, and fairness, with techniques such as system prompt design (Guo et al., 2024) and reinforcement learning (Christiano et al., 2017; Liu et al., 2020; Ye et al., 2025). Alignment with these general values provides crucial foundational guardrails for safe and broadly acceptable interactions. However, true human values can differ significantly across individuals, cultures, and contexts, presenting further complexities. It is equally important for LLMs to develop the capacity to simulate individual values that are more fine-grained and varied.

This research uses information about personal values and beliefs to better simulate an individual's personality and preferences. Values—the enduring beliefs that guide attitudes and behaviors (Schwartz, 2012b)—fundamentally shape how individuals perceive and interact with their world. People with differing value systems respond distinctively to identical situations, making value representation essential for accurate human simulation in various contexts, from conducting behavioral

*Co-corresponding authors: Ziyi Ye (zyye@fudan.edu.cn) & Qingyao Ai (aiqy@tsinghua.edu.cn)
This work is supported by the Tsinghua University Initiative Scientific Research Program.

research to building personal assistants.

To bridge this gap, we investigate the simulation of individual values with LLMs. Motivated by the fact that the formation and expression of human values stem from complex personal histories and experiences, we propose a simple yet effective method, SimVBG (Simulating Individual Values by Backstory Generation), to prompt an LLM with individual backstories to simulate human behavior. While previous studies have used personas' profiles to create LLM agents that would represent certain populations (Moon et al., 2024; Wang et al., 2025; Park et al., 2024), backstories encapsulate critical life events, cultural contexts, and social influences that collectively shape an individual's value system (McAdams, 2001). Specifically, SimVBG leverages backstory-based value representation combined with a multi-system architecture grounded in the Cognitive-Affective System Theory of Personality (CAPS) (Mischel and Shoda, 1995). As shown in Figure 1, SimVBG consists of three modules. (1) First, a story module prompts an LLM to write a backstory with information about an individual encompassing their demographic profile and response to a series of questions related to their value. The performance of the backstory is evaluated using questions that are independent of those used to generate the backstory, yet are responded to by the same individual. (2) Second, a Cognitive-Affective-Behavioral (CAB) module motivated by the CAPS is devised to understand the backstories from different aspects and generate multiple candidate responses to simulate the individual. This approach mirrors the multifaceted nature of human value-driven responses, which rarely emerge from isolated cognitive or emotional processes (Loewenstein et al., 2003). (3) Finally, a system integration module is used to integrate outputs from the CAB modules and generate the response that is most likely to mirror the real answer responded by the individual.

To evaluate our framework, we constructed a comprehensive benchmark based on the World Values Survey dataset (Haerpfer et al., 2022), comprising 97,220 individuals with 290 distinct demographic and opinion attributes. Our experimental results demonstrate that SimVBG significantly outperforms existing methods in simulating human value-based responses, including those with users' full information and the Retrieval-augmented Generation (RAG) system that retrieves historical individual information for each response generation. SimVBG achieves the most performance gain, particularly in domains related to happiness perception and social value judgments. Furthermore, we observe that simulation accuracy improves progressively with increased interaction data from target individuals, showing the system's capacity for personalization refinement over more interactions (Hwang et al., 2023).

Our contributions include:

- We present a framework that leverages backstories, derived from raw profile data, to effectively prompt Large Language Models (LLMs) in representing individual values.
- To simulate individual responses, we propose a modular, personality theory-inspired method that captures the cognitive, affective, and behavioral dimensions of human cognitive processes.
- We empirically show that the proposed method outperforms a series of existing prompting-based methods to represent human value and demonstrate that the alignment accuracy improves with increased individual information.

2 Related Work

2.1 User-based Alignment in Large Language Models

Recent research has explored various approaches to align language models with user characteristics and preferences. Sun et al. (2024) introduced "Random Silicon Sampling" to simulate human subpopulations using group-level demographic information, while Moon et al. (2024) developed "Anthology," employing personal backstories to create virtual personas for improved alignment with survey responses. Hwang et al. (2023) found that demographics and ideologies alone are insufficient predictors of user opinions, demonstrating that incorporating relevant past opinions improves accuracy in predicting responses to survey questions.

In personalized approaches, Personalized-RLHF (Li et al., 2024) captures individual preferences through a lightweight user model trained jointly with the LLM. For specific applications, Xiang et al. (2024) created SimUser for simulating user interactions with mobile applications. These works represent important steps toward user simulation, yet they typically rely on narrowly defined

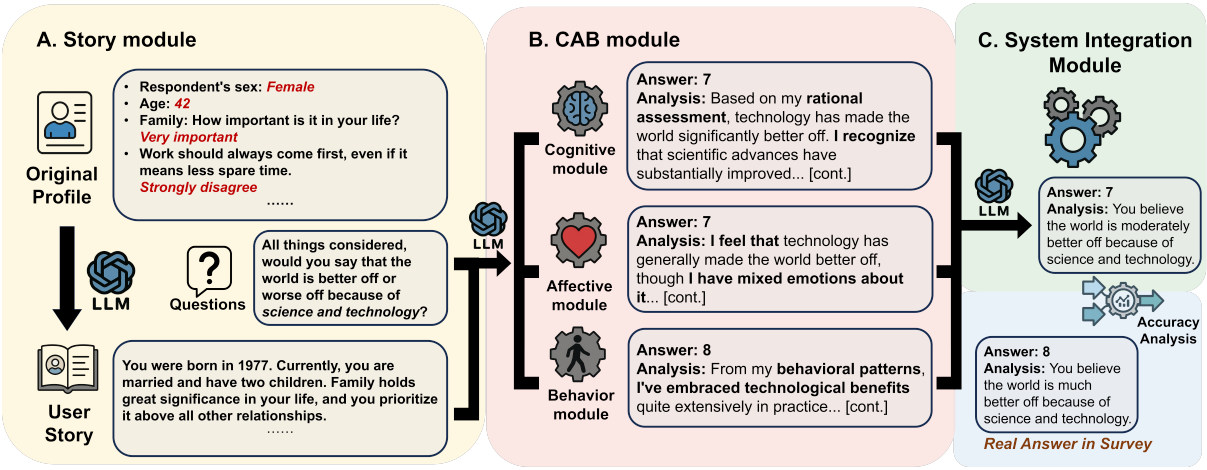

Figure 1: The complete process of simulating human value responses. For each individual, the profile is first converted into a story through the story module, then, using the cognitive-affective-behavior module and after system integration, the individual's response to a specific question is simulated.

contexts. Our work extends these efforts by utilizing data describing human values and preferences to capture an individual's personality more accurately. Furthermore, we implement a structured cognitive-affective-behavioral framework and demonstrate effectiveness across diverse questions with verifiable ground truth responses.

2.2 Multi-System Approaches to Human Decision-Making

Contemporary psychological research recognizes that human cognition, emotion, and behavior arise from parallel, interacting neural systems (Pessoa, 2008; Kahneman, 2011). For instance, the Cognitive-Affective System Theory of Personality (CAPS) (Mischel and Shoda, 1995) that describes personality as a dynamic network of cognitive-affective units—including encodings, expectancies, emotions, and behavioral competencies—that activate concurrently in response to situational features. Metacognition research further reveals how the brain coordinates these parallel systems, resolving conflicts and integrating diverse processing streams into the final decisions (Fleming and Dolan, 2012).

These theoretical foundations suggest that authentic simulation of human decision-making requires modeling the distinct contributions of cognitive, affective, and behavioral processes, along with the integrative mechanisms that harmonize their parallel operations.

Even though LLMs process information in different ways than humans, the recent research indicates that they are increasingly capable of mimicking complex cognitive processes (Kosinski, 2024; Xie et al., 2024), raising questions that CAPS and other psychological theories can be used to build technical frameworks enabling better information processing by the LLMs in the context of human simulations.

2.3 Value System Modeling and WVS Studies

Scholars like (Hofstede, 1983; Schwartz, 2012a; Inglehart, 2006) developed distinct frameworks for classifying and analyzing values on individual and community levels. Those frameworks were used to examine the relationship between values and beliefs held by individuals and pro-environmental behaviors (Primc et al., 2021), responses to corporate social responsibility initiatives (Rosario et al., 2014), and the broadly defined national culture characteristics (Beugelsdijk and Welzel, 2018).

WVS occupies an important place in human value studies due to its extensive geographic and coverage (including beliefs and attitudes about religion, economy, science and technology), the number of respondents, and data accessibility [1]. For this reason, the WVS data is frequently used to better understand topics such as people's perception of well-being (Fleche et al., 2012), support for democracy (Ariely and Davidov, 2011), or the relationship between trust and health (Jen et al., 2010). WVS data, are also used in LLM-related research. However, the LLM-related studies using WVS data focus on studying groups of people (**?**), rather than individuals, with a view to using LLM agents for participation in social surveys (Boelaert

[1] https://www.worldvaluessurvey.org/wvs.jsp

et al., 2025).

3 Methodology

We present a novel framework for human value response simulation that addresses two fundamental challenges: representing complex user profiles and capturing the multidimensional nature of human decision-making. Our approach consists of two primary components. First, a Story Processing Module transforms survey responses into coherent narrative representations, enabling more effective integration of user information. Second, a Multi-Module Simulation Framework decomposes the response generation process into parallel cognitive, affective, and behavioral dimensions to more accurately represent human decision-making. This modular approach enables more nuanced modeling of personality-specific response patterns than monolithic prompt-based methods. We complement these core components with a data configuration to enable rigorous evaluation of simulation performance across diverse user profiles and survey sections.

3.1 Task Definition

In this work, we address the challenge of simulating human responses to value-oriented questions based on personal profiles. Formally, given a user's profile information collected from previous survey responses, our task is to predict how that individual would respond to new value-related questions not contained in their profile.

Let $P = (q_1, a_1), (q_2, a_2), ..., (q_n, a_n)$ represent a user's profile, where each pair (q_i, a_i) consists of a survey question and the user's corresponding answer. The profile information includes both multiple-choice questions with their selected options and fill-in-the-blank questions with the provided answers. In our dataset, we employ 5-fold cross-validation where 232 out of 290 total questions form the profile for each fold, with the remaining 58 questions used for evaluation.

For a new question q_{new} not present in P, the task is to predict the user's response $\hat{a}_{new}$, which is typically a selection from a set of predefined options for multiple-choice questions (the predominant format in our value survey questions).

This task presents two significant challenges. First, the user profiles contain approximately 290 pieces of information (with 232 used for training), making it difficult for models to identify which profile elements are relevant to a particular prediction. Second, the overall complexity of human cognitive, emotional, and behavioral processes, as well as individual differences between people in this area, make the simulation of specific people's answers extremely difficult.

We evaluate performance using two metrics: Accuracy, which measures the percentage of correctly predicted responses for multiple-choice questions, and Mean Absolute Error (MAE), which quantifies the average deviation between predicted and actual responses when answers can be ordinally ranked. These metrics are calculated for each simulated individual and then averaged across all individuals to evaluate the overall framework performance.

3.2 Story Module

To facilitate accurate user-level alignment of the LLM-generated responses with complex profiles, we introduce a story processing module that transforms survey responses into coherent, narrative-based representations.

3.2.1 Motivation of Story Module

Direct incorporation of full people's profiles into prompt contexts presents several challenges: (1) **Context Length Limitations**: Even with modern LLMs' expanded context windows, including hundreds of question-answer pairs remains inefficient; (2) **Information Integration**: Raw question-answer formats impede the model's ability to form holistic representations of individual personalities; (3) **Retrieval Limitations**: Conventional retrieval-augmented generation (RAG) approaches that select only subsets of profile information for each query sacrifice the holistic understanding of an individual's personality pattern.

As demonstrated in our experiments (Section 5.1), both direct input and retrieval-based approaches fail to achieve optimal alignment with actual human responses, either overwhelming the model with disjointed information or providing incomplete information.

3.2.2 Narrative Transformation Technique

We developed a specialized narrative transformation technique that converts structured survey data into coherent backstories while preserving information integrity. This technique leverages cognitive principles of narrative processing, which suggest that humans (and by extension, LLMs trained on human text) better comprehend and retain informa-

tion presented in story format compared to disconnected factual statements.

Our transformation framework operates through a two-phase approach:

1. **Thematic Organization**: Survey responses are reorganized according to conceptual relatedness (demographic attributes, value systems, political orientations, etc.), establishing coherent narrative threads

2. **Narrative Integration**: These thematically organized elements are then woven into a continuous second-person narrative that maintains the accuracy of the content while providing natural transitions and logical flow between concepts

The transformation framework enforces strict information preservation constraints (implemented through explicit prompt instructions to maintain all original data points without summarization, detailed in Appendix C.1), ensuring that specific values, numerical responses, and unique identifiers remain unaltered during transformation. Meanwhile, the narrative structure facilitates holistic comprehension by explicitly connecting related beliefs and attributes that might otherwise remain implicit in raw survey data.

This approach yields personalized narratives that: (1) maintain complete fidelity to the original 232 profile elements; (2) reduce cognitive load for the LLMs through coherent organization; and (3) make it easier for the downstream model to identify relevant personality patterns during response simulation. The complete story generation prompt with detailed instructions is provided in Appendix C.1. To validate the faithfulness of our backstory generation process, we conducted comprehensive qualitative analysis examining both factual synthesis and value preservation capabilities. Detailed case studies demonstrating the model's ability to accurately transform structured profile data into coherent narratives while maintaining information integrity are presented in Appendix H.

3.3 Multi-Module Simulation Framework

Building on the narrative representation provided by the story processing module, we introduce a simulation framework that helps capture the multi-faceted nature of human decision-making. Our approach draws inspiration from established theories in cognitive psychology and neuroscience, particularly the Cognitive-Affective Personality System (CAPS) theory (Mischel and Shoda, 1995), which conceptualizes personality as a dynamic network of cognitive-affective units that activate in situation-specific patterns.

3.3.1 Cognitive-Affective-Behavioral Architecture

We decompose the simulation process into three parallel processing modules corresponding to the primary dimensions identified in cognitive neuroscience research:

1. **Cognitive Module**: Simulates information processing, reasoning patterns, belief structures, and analytical tendencies that influence decision-making

2. **Affective Module**: Models emotional responses, value alignments, motivational states, and affective reactions to potential outcomes

3. **Behavioral Module**: Captures action tendencies, implementation patterns, contextual influences, and behavioral histories

This tripartite architecture is grounded in CAPS theory's classification of personality units (Mischel and Shoda, 1995), which delineates encodings and expectancies (cognitive), affects (emotional), and competencies and self-regulatory plans (behavioral) as parallel processing components. Our design mirrors the parallel activation dynamics described in neuropsychological models of decision-making (Pessoa, 2008), where cognitive, affective, and behavioral systems operate concurrently before integration.

3.3.2 Module Implementation

Each module employs specialized prompting to simulate its respective psychological domain:

1. **Cognitive Module**: Processes profile information with emphasis on belief structures, information gathering preferences, reasoning approaches, worldview framing, and weighting factors in analytical decision-making.

2. **Affective Module**: Analyzes profile information with focus on affective patterns, emotional regulation styles, values, motivational states, and identity-based emotional responses.

3. **Behavioral Module**: Examines profile information through the lens of behavioral tendencies, environmental influences, capability constraints, experiential learning, and implementation patterns.

Each module generates both a predicted response option and a detailed analysis of the reasoning process from its specialized perspective. This parallel processing approach captures the specialized contribution of each psychological system while avoiding the limitations of monolithic simulation methods. The complete implementation prompts for each module are detailed in Appendix C.2. Qualitative analysis of the CAB modules' reasoning processes, including case studies showing how each module generates psychologically distinct perspectives for the same question, is provided in Appendix H.

3.4 System Integration Module

The System Integration Module synthesizes outputs from the three specialized modules into a coherent final response. Rather than simply averaging numerical predictions, this module analyzes the detailed reasoning provided in each analysis to identify patterns of alignment, conflict, and contextual dominance. This qualitative integration may better capture how individuals reconcile potentially conflicting cognitive, affective, and behavioral tendencies—a critical aspect of authentic decision-making overlooked by simplistic aggregation methods.

In our ablation studies (Section 5.2), we compare this integration approach with a baseline that simply averages the predictions from the three modules, showing the contribution of structured integration to prediction accuracy.

4 Experimental Setup

4.1 Data Configuration

4.1.1 Dataset Selection

For the value response simulation task, we required a dataset with sufficient scale, coverage, and profile richness to effectively develop and evaluate our framework. We selected the World Values Survey (WVS) Wave 7 dataset, collected between 2017 and 2022, encompassing user information from 66 distinct countries. This dataset provides a globally diverse sample with comprehensive value profiles spanning various cultural contexts.

Each user in the WVS dataset responded to 290 questions, covering both demographic information (age, gender, education level, income, etc.) and multidimensional value information (political views, religious beliefs, social attitudes, environmental concerns, etc.). This rich profile data enables the construction of detailed user representations necessary for nuanced personality simulation.

4.1.2 Data Split Configuration

We evaluate our framework on 100 randomly selected users from the WVS dataset to ensure computational feasibility while maintaining statistical validity.

For each user, we implement a question-level split where 80% of their survey responses (232 out of 290 questions) are used to construct their personality profile, while the remaining 20% (58 questions) serve as evaluation targets for response prediction.

To ensure robust evaluation, we employ 5-fold cross-validation at the question level. For each user, the 290 questions are randomly partitioned into five folds of 58 questions each. In each validation round, four folds (232 questions) form the user's profile, while the model predicts responses to the held-out fold (58 questions). This process ensures that every question serves as both profile information and evaluation target across different folds.

This configuration allows us to assess how well our framework can predict a user's responses to unseen questions based on their known response patterns, which directly addresses the core challenge of human value simulation.

4.2 Language Models

We evaluated SimVBG using four diverse language models:

GPT-3.5-Turbo (Brown et al., 2020): OpenAI's commercial model with instruction-tuning and RLHF optimization. Llama-3.1-8B (Touvron et al., 2023): Meta AI's open-source model with 8 billion parameters and strong multilingual capabilities. Qwen-2.5-7B (Bai et al., 2023): Alibaba Cloud's model with 7 billion parameters featuring an extended context window and specialized training on reasoning tasks. DeepSeek-V3 (Liu et al., 2024): A recent foundation model optimized for dialogue coherence and knowledge representation.

For reproducibility, we set the temperature parameter to zero across all models and maintained identical prompting frameworks for all experimental conditions.

5 Experiments and Results

5.1 Main Results

Our experiments evaluate SimVBG against baseline approaches on value response alignment tasks using four LLMs: Llama-3.1-8B, Qwen-2.5-7B, GPT-3.5-Turbo, and DeepSeek-V3. With our limited computational resources, we tested all models on 100 samples (simulating 100 different users). For each user, we employed the 5-fold cross-validation setup described in Section 4.1.2, where 80% of their questions (232 out of 290) form the profile and 20% (58 questions) serve as prediction targets. Our framework is designed to support the simulation and testing of all 97,220 users in our dataset.

We compared our framework against two baseline methods: (1) Full Info, which directly inputs the complete profile information, and (2) RAG, which selectively provides the most relevant profile information. The Full Info method inputs all 232 survey questions and responses from a user's profile alongside the target question to generate LLM predictions. The RAG method employs text-embedding-ada-002 to create embeddings for each question and profile entry, then uses cosine similarity to identify the 3 most relevant profile entries for each test question. Additionally, we implemented two representative approaches from related work: (3) IndieValue (Jiang et al., 2024), which converts structured survey responses into natural language statements following specific transformation rules, and (4) Self-alignment (Choenni and Shutova, 2024), which selects relevant demonstrations using text similarity metrics. The IndieValue method transforms question-answer pairs from user profiles into first-person declarative statements by applying systematic conversion rules: converting "you" to "I", transforming interrogative sentences to declarative ones, and embedding answer options into natural language expressions. For example, "How secure do you feel?" with answer "very secure" becomes "I feel very secure." These generated statements serve as value demonstrations for predicting responses to new questions. The Self-alignment approach selects the most relevant profile questions as demonstrations by computing ChrF++ similarity (Popović, 2017) between the target question text and profile question texts. It then constructs prompts with the top-5 most similar question-answer pairs to guide the model's prediction, following the principle that similar questions reflect similar underlying value orientations.

SimVBG consistently outperforms all baseline methods across the tested LLMs. Our approach demonstrates substantial improvements in Mean Absolute Error (MAE) as shown in Table 1 and accuracy as detailed in Appendix A, with consistent patterns across all tested models: Llama-3.1-8B, Qwen-2.5-7B, GPT-3.5-Turbo, and DeepSeek-V3. The IndieValue baseline, despite its systematic approach to converting structured data into natural language statements, achieves moderate performance improvements over traditional methods but remains inferior to SimVBG across all models. The Self-alignment method shows variable performance across different models, demonstrating the limitations of pure similarity-based demonstration selection compared to our structured simulation approach. Paired t-tests confirm that SimVBG's improvements over all four baseline methods are statistically significant ($p < 0.05$ for all comparisons). These results indicate that our simulation approach captures underlying preference patterns more effectively than methods that either use complete profile information or select information based solely on semantic similarity. To further validate our approach, we conducted additional experiments with RAG using more retrieved entries (top-5 and top-10). Results show that increasing retrieved entries does not improve RAG performance (detailed results in Appendix G), confirming that effective information synthesis is more critical than raw data volume. The advantage appears to derive from SimVBG's structured simulation of human decision processes rather than from model-specific characteristics.

5.2 Ablation Studies

To validate the contribution of each component in our SimVBG framework, we conducted a series of ablation experiments across all four LLM models.

5.2.1 Story module contribution analysis

First, we examined the impact of the story generation module by replacing it with the original profile information while maintaining the parallel question-answering structure of subsequent modules. Results consistently demonstrated that the story module significantly enhances framework performance across all tested LLMs.

Table 1: Mean Absolute Error (MAE) performance comparison across different value dimensions and methods (lower is better). Column headers represent different value dimensions from the World Values Survey: Core Values (Core), Happiness & Well-being (Hap.), Trust, Economic Integrity (Econ.Int), Security, Technology (Tech), Moral & Religious (Mo.&Rel.), Political Engagement (Pol.Eng), and Demographics (Demo). Complete descriptions of these value dimensions are provided in Appendix F.

Model	**Method**	Core	Hap.	Trust	Econ.Int	Security	Tech	Mo.&Rel.	Pol.Eng	Demo	**Overall**
GPT-3.5-Turbo	Full Info	0.597	0.438	0.352	0.326	0.460	0.425	0.378	0.489	0.556	0.465
	RAG	0.363	0.423	0.247	0.274	0.302	0.265	0.240	0.371	0.483	0.336
	IndieValue	0.380	0.230	0.237	0.269	0.319	0.310	0.249	**0.264**	0.428	0.291
	Self-align	0.328	0.178	0.236	**0.230**	0.323	0.257	0.302	0.309	0.472	0.308
	SimVBG	**0.299**	**0.167**	**0.231**	0.239	**0.280**	**0.246**	**0.188**	0.265	**0.344**	**0.260**
Llama-3.1-8B	Full Info	0.543	0.660	0.324	0.326	0.439	0.411	0.408	0.434	0.592	0.452
	RAG	0.453	0.563	0.297	0.307	0.381	**0.326**	0.308	0.400	0.536	0.390
	IndieValue	0.476	0.372	0.384	0.391	0.457	0.327	0.395	0.388	0.545	0.422
	Self-align	**0.381**	0.270	0.445	0.345	0.411	0.402	0.397	0.393	0.511	0.409
	SimVBG	0.400	**0.222**	**0.226**	**0.302**	**0.346**	0.343	**0.250**	**0.308**	**0.326**	**0.308**
Qwen-2.5-7B	Full Info	0.669	0.315	0.367	0.325	0.383	0.374	0.492	0.488	0.592	0.477
	RAG	0.777	0.598	0.359	0.328	0.550	0.653	0.492	0.557	0.584	0.544
	IndieValue	0.403	0.221	0.237	0.253	0.322	**0.255**	0.320	**0.265**	0.429	0.304
	Self-align	0.535	0.634	0.375	0.317	0.722	0.579	0.504	0.653	0.499	0.560
	SimVBG	**0.341**	**0.201**	**0.219**	**0.253**	**0.320**	0.331	**0.236**	0.293	**0.360**	**0.288**
DeepSeek-V3	Full Info	0.365	0.611	0.258	0.382	0.341	**0.233**	0.287	0.325	0.530	0.355
	RAG	0.360	0.444	0.236	0.253	0.275	0.260	0.229	0.280	0.466	0.304
	IndieValue	0.443	0.642	**0.191**	0.278	0.330	0.236	0.260	0.305	0.537	0.322
	Self-align	0.389	0.364	0.221	**0.242**	0.292	0.292	0.296	0.315	0.533	0.328
	SimVBG	**0.306**	**0.168**	0.210	0.270	**0.267**	0.251	**0.197**	**0.257**	**0.338**	**0.257**
–	Chance	0.550	0.486	0.497	0.414	0.539	0.393	0.474	0.463	0.610	0.507

Table 2: Ablation study on SimVBG across different language models in terms of MAE (lower is better).

Setting	**GPT-3.5-Turbo**	**Llama-3.1-8B**	**Qwen-2.5-7B**	**Deepseek-V3**
SimVBG	**0.264**	**0.310**	**0.271**	**0.244**
SimVBG w/o CAB module	0.352	0.322	0.341	0.321
SimVBG w/o story module	0.304	0.312	0.308	0.256

5.2.2 Three-module parallel structure contribution analysis

Next, we assessed the cognitive-affective-behavioral (CAB) module as a whole by removing it entirely and using only the generated story for direct response prediction. This ablation revealed substantial performance degradation across all models, confirming that the structured simulation of mental processes provided by the CAB module is crucial for accurate value alignment.

We further investigated the individual contributions of each unit within the CAB module by selectively removing one unit while keeping the others intact. These fine-grained ablations showed that each component plays a distinct and necessary role in the overall framework.

5.3 Impact of the User Profile Scale

To investigate how the amount of profile information influences simulation accuracy, we conducted an incremental profile scale experiment using our SimVBG framework. Given that user profiles contain substantial information (232 training data points per user), we sought to understand the relationship between profile comprehensiveness and simulation performance.

We maintained the same testing conditions as our main experiment, using the same test-train split (20% as test questions, 80% as training questions) from one fold of our cross-validation setup. For each user, we varied the amount of profile information provided to the model by randomly sampling from their available training data points. We tested

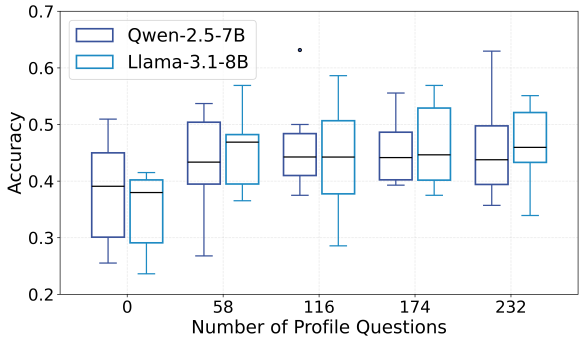

Figure 2: Impact of profile information scale on simulation accuracy.

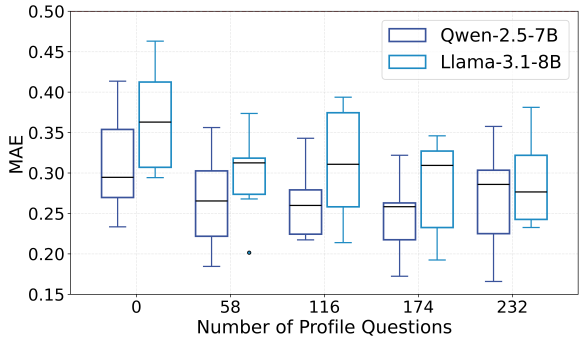

Figure 3: Impact of profile information scale on simulation error (lower is better).

five configurations: 0, 58, 116, 174, and 232 profile items, with the sampling increment of 58 chosen to provide sufficient granularity while maintaining experimental feasibility. For each configuration, we randomly sampled the specified number of profile items for each user to ensure unbiased comparison.

Our findings demonstrate that increasing the amount of profile information consistently improves simulation performance across all four model architectures. As shown in Figure 2 and Figure 3, both accuracy and MAE metrics improve as more profile information becomes available.

Notably, we observed that the largest improvements occur in the early stages of profile expansion (from 0 to approximately 100 items). Beyond this threshold, while performance continues to improve, the rate of improvement diminishes, indicating diminishing returns when profile information exceeds approximately 100 items.

This pattern reveals that in value response alignment tasks, expanding user profiles to include several dozen data points provides substantial benefits for simulation accuracy. However, the incremental benefit of additional profile information decreases as profiles become more comprehensive. This finding has implications for designing profile-based simulation systems, suggesting an optimal balance between profile comprehensiveness, participants' effort and privacy, and computational efficiency.

6 Conclusion

We presented SimVBG, a framework for simulating human value responses by backstories that combines narrative generation with a structured cognitive-affective-behavioral processing. Our experiments across multiple base LLMs showed that SimVBG consistently outperforms methods using either complete user information or a retrieval-augmented generation method. Ablation studies confirmed that both the story generation module and the cognitive-affective-behavioral module contribute substantially to the framework's effectiveness.

The success of SimVBG suggests that simulating human responses benefits from a modular approach inspired by psychological processes. By generating backstories that contextualize abstract values and then processing these through parallel cognitive, affective, and behavioral pathways, SimVBG more accurately learned from the raw information included in individuals' profiles to make value judgments more aligned with the individuals. Further research could explore more sophisticated architectural designs and leverage advanced post-training techniques for LLMs to enhance their capabilities in producing and processing backstories to reflect human values.

This work has potential applications in personalized AI systems and social science research. For personalized services, frameworks like SimVBG could enable systems that adapt to individual preferences without requiring extensive data collection. In social science, such simulations could help explore societal dynamics by modeling interactions between individuals with different value systems.

Limitations

Technical Limitations

Our study has several limitations that should be addressed in future work.

First, our evaluation focused exclusively on value-related questions from the World Values Survey, leaving open the question of whether SimVBG would be effective for simulating responses in other domains. Further research could incorporate other types of data into the SimVBG framework.

Second, creating backstories that are naturalistic but also include all the important information about an individual constitutes a significant challenge. The process of translating the raw information about a person into a coherent backstory that would mirror the way this person describes their experiences may result in hallucinations or misportrayal of certain elements of a person's profile. Simultaneously, making the narrative more realistic requires a certain level of transformation of raw data, in a manner that may better convey a person's emotion, preferences, or communication style. Further research on how to achieve the right balance between the two is necessary.

Third, due to computational constraints, we tested our approach on only 100 simulated users, though our framework supports all 97,220 users in our dataset. More comprehensive testing could reveal additional insights about performance across diverse demographic groups and value systems.

Finally, while our cognitive-affective-behavioral module draws inspiration from psychological theories, it remains a simplification of actual human mental processes. Future work could explore whether more sophisticated psychological models can further enhance LLM simulation of human behavior.

Ethical Concerns and Societal Implications

We used a publicly available dataset about people's values, hence being non-invasive from the perspective of individuals' privacy. In general, however, simulating individuals' opinions and behaviors poses a unique challenge of achieving the balance between the usefulness of simulation for the particular goals and respecting the person's privacy and autonomy.

Personalized agents that can either behave in a more realistic "human" way or just understand preferences of specific individuals have the potential for use in a variety of settings, e.g., providing more individualized education resources, assisting in various jobs, serving as companions for elderly people, and participating in some social and behavioral research. While using LLM agents for these purposes may have a significant positive societal impact, we need to indicate certain ethical challenges surrounding this endeavor.

The attempts to simulate an individual's opinions and behaviors may encourage a tendency to collect extensive and sensitive data about the individuals. Hence, we need a legal and ethical framework protecting individuals from undue infringements of their privacy and allowing individuals to have influence over how their data is used.

Furthermore, personalized agents may misrepresent certain individuals or even act in a manipulative way. Hence, the actual use of such agents requires technical and institutional safeguards that would consider holistically the circumstances of the agents' usage, especially the purpose of such usage, the levels and nature of the relevant risks, as well as the need to respect individuals' privacy and autonomy.

References

Gati V Aher, Rosa I Arriaga, and Adam Tauman Kalai. 2023. Using large language models to simulate multiple humans and replicate human subject studies. In *International Conference on Machine Learning*, pages 337–371. PMLR.

Gal Ariely and Eldad Davidov. 2011. Can we rate public support for democracy in a comparable way? cross-national equivalence of democratic attitudes in the world value survey. *Social Indicators Research*, 104:271–286.

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, and 1 others. 2023. Qwen technical report. *arXiv preprint arXiv:2309.16609*.

Sjoerd Beugelsdijk and Chris Welzel. 2018. Dimensions and dynamics of national culture: Synthesizing hofstede with inglehart. *Journal of cross-cultural psychology*, 49(10):1469–1505.

Julien Boelaert, Samuel Coavoux, Étienne Ollion, Ivaylo Petev, and Patrick Präg. 2025. Machine bias. how do generative language models answer opinion polls? *Sociological Methods & Research*, page 00491241251330582.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, and 1 others. 2020. Language models are

few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Rochelle Choenni and Ekaterina Shutova. 2024. Self-alignment: Improving alignment of cultural values in llms via in-context learning. *arXiv preprint arXiv:2408.16482*.

Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30.

Sarah Fleche, Conal Smith, and Piritta Sorsa. 2012. Exploring determinants of subjective wellbeing in oecd countries: Evidence from the world value survey.

Stephen M Fleming and Raymond J Dolan. 2012. The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594):1338–1349.

Hui Guo, Long Zhang, Xilong Feng, and Qiusheng Zheng. 2024. A review of the application of prompt engineering in the safety of large language models. In *Proceedings of the 2024 2nd International Conference on Information Education and Artificial Intelligence*, pages 424–430.

Christian Haerpfer, Ronald Inglehart, Alejandro Moreno, Christian Welzel, Kseniya Kizilova, Jaime Diez-Medrano, Marta Lagos, Pippa Norris, Eduard Ponarin, and Bi Puranen. 2022. World values survey wave 7 (2017-2022) cross-national data-set. *(No Title)*.

Geert Hofstede. 1983. National cultures in four dimensions: A research-based theory of cultural differences among nations. *International studies of management & organization*, 13(1-2):46–74.

EunJeong Hwang, Bodhisattwa Prasad Majumder, and Niket Tandon. 2023. Aligning language models to user opinions. *arXiv preprint arXiv:2305.14929*.

Ronald Inglehart. 2006. Mapping global values. *Comparative sociology*, 5(2-3):115–136.

Min Hua Jen, Erik R Sund, Ron Johnston, and Kelvyn Jones. 2010. Trustful societies, trustful individuals, and health: An analysis of self-rated health and social trust using the world value survey. *Health & place*, 16(5):1022–1029.

Liwei Jiang, Taylor Sorensen, Sydney Levine, and Yejin Choi. 2024. Can language models reason about individualistic human values and preferences? *arXiv preprint arXiv:2410.03868*.

Daniel Kahneman. 2011. *Thinking, fast and slow*. macmillan.

Michal Kosinski. 2024. Evaluating large language models in theory of mind tasks. *Proceedings of the National Academy of Sciences*, 121(45):e2405460121.

Xinyu Li, Ruiyang Zhou, Zachary C Lipton, and Liu Leqi. 2024. Personalized language modeling from personalized human feedback. *arXiv preprint arXiv:2402.05133*.

Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, and 1 others. 2024. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*.

Fei Liu and 1 others. 2020. Learning to summarize from human feedback. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 583–592.

George Loewenstein, Jennifer S Lerner, and 1 others. 2003. The role of affect in decision making. *Handbook of affective science*, 619(642):3.

Dan P McAdams. 2001. The psychology of life stories. *Review of general psychology*, 5(2):100–122.

Walter Mischel and Yuichi Shoda. 1995. A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological review*, 102(2):246.

Suhong Moon, Marwa Abdulhai, Minwoo Kang, Joseph Suh, Widyadewi Soedarmadji, Eran Kohen Behar, and David M Chan. 2024. Virtual personas for language models via an anthology of backstories. *arXiv preprint arXiv:2407.06576*.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pages 1–22.

Joon Sung Park, Carolyn Q Zou, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Robb Willer, Percy Liang, and Michael S Bernstein. 2024. Generative agent simulations of 1,000 people. *arXiv preprint arXiv:2411.10109*.

Luiz Pessoa. 2008. On the relationship between emotion and cognition. *Nature reviews neuroscience*, 9(2):148–158.

Maja Popović. 2017. chrf++: words helping character n-grams. In *Proceedings of the second conference on machine translation*, pages 612–618.

Kaja Primc, Marko Ogorevc, Renata Slabe-Erker, Tjaša Bartolj, and Nika Murovec. 2021. How does schwartz's theory of human values affect the proenvironmental behavior model? *Baltic Journal of Management*, 16(2):276–297.

González-Rodríguez Ma Rosario, Díaz-Fernández Ma Carmen, and Simonetti Biagio. 2014. Values and corporate social initiative: An approach through schwartz theory. *International Journal of Business and Society*, 15(1):19.

Shalom Schwartz. 2012a. Toward refining the theory of basic human values. *Methods, theories, and empirical applications in the social sciences*, pages 39–46.

Shalom H Schwartz. 2012b. An overview of the schwartz theory of basic values. *Online readings in Psychology and Culture*, 2(1):11.

Seungjong Sun, Eungu Lee, Dongyan Nan, Xiangying Zhao, Wonbyung Lee, Bernard J Jansen, and Jang Hyun Kim. 2024. Random silicon sampling: Simulating human sub-population opinion using a large language model based on group-level demographic information. *arXiv preprint arXiv:2402.18144*.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, and 1 others. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Quan Tu, Chuanqi Chen, Jinpeng Li, Yanran Li, Shuo Shang, Dongyan Zhao, Ran Wang, and Rui Yan. 2023. Characterchat: Learning towards conversational ai with personalized social support. *arXiv preprint arXiv:2308.10278*.

Angelina Wang, Jamie Morgenstern, and John P Dickerson. 2025. Large language models that replace human participants can harmfully misportray and flatten identity groups. *Nature Machine Intelligence*, pages 1–12.

Xintao Wang, Yunze Xiao, Jen-tse Huang, Siyu Yuan, Rui Xu, Haoran Guo, Quan Tu, Yaying Fei, Ziang Leng, Wei Wang, and 1 others. 2023a. Incharacter: Evaluating personality fidelity in role-playing agents through psychological interviews. *arXiv preprint arXiv:2310.17976*.

Zekun Moore Wang, Zhongyuan Peng, Haoran Que, Jiaheng Liu, Wangchunshu Zhou, Yuhan Wu, Hongcheng Guo, Ruitong Gan, Zehao Ni, Jian Yang, and 1 others. 2023b. Rolellm: Benchmarking, eliciting, and enhancing role-playing abilities of large language models. *arXiv preprint arXiv:2310.00746*.

Zhilin Wang, Yu Ying Chiu, and Yu Cheung Chiu. 2023c. Humanoid agents: Platform for simulating human-like generative agents. *arXiv preprint arXiv:2310.05418*.

Wei Xiang, Hanfei Zhu, Suqi Lou, Xinli Chen, Zhenghua Pan, Yuping Jin, Shi Chen, and Lingyun Sun. 2024. Simuser: Generating usability feedback by simulating various users interacting with mobile applications. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–17.

Chengxing Xie, Canyu Chen, Feiran Jia, Ziyu Ye, Shiyang Lai, Kai Shu, Jindong Gu, Adel Bibi, Ziniu Hu, David Jurgens, and 1 others. 2024. Can large language model agents simulate human trust behavior? In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

Ziyi Ye, Xiangsheng Li, Qiuchi Li, Qingyao Ai, Yujia Zhou, Wei Shen, Dong Yan, and Yiqun Liu. 2025. Learning llm-as-a-judge for preference alignment. In *The Thirteenth International Conference on Learning Representations*.

Appendix

Appendix A provides the accuracy results of our main experiments compared with two baselines.

Appendix B presents the radar chart showing the effectiveness of the SimVBG framework across different types of values.

Appendix C details the prompt flow process of the SimVBG framework.

Appendix D provides real examples of backstories generated in our experiments.

Appendix E includes the prompts used for the two baselines: Origin Full and RAG.

Appendix F provides an introduction to the World Value Survey (WVS) dataset used in our study.

Appendix G contains additional experimental results for RAG baseline variants with different numbers of retrieved entries.

Appendix H presents detailed qualitative analysis validating the effectiveness of backstory generation and CAB module reasoning processes.

Disclosure (language editing). We used Google Gemini and Grammarly solely for minor language editing of the manuscript. No ideas, analyses, code, prompts, datasets, figures, or results were generated by AI systems; all scientific content was authored and verified by the authors. No non-public or sensitive data beyond the manuscript text was provided to these tools.

A Accuracy Results of Experiments

This section presents comprehensive accuracy results from our main experiments, providing a direct comparison between our proposed approach and two baseline methods, as well as the accuracy results from our ablation studies. While the key findings are discussed in the main text, it is worth noting that our SimVBG approach achieves accuracy comparable to Mean Absolute Error (MAE) metrics, demonstrating its effectiveness in realistic user simulation.

Model	**Setting**	Core Values	Happiness	Trust	Econ.Int	Security	Tech	Moral&Rel	Pol.Eng	Demo	**Overall**
GPT-3.5-Turbo	Full Info	0.257	0.287	0.414	0.269	0.352	0.176	0.319	0.250	0.235	0.298
	RAG	0.448	0.224	**0.508**	0.311	0.447	**0.249**	0.502	0.343	0.307	0.404
	SimVBG	**0.520**	**0.526**	0.476	**0.339**	**0.474**	0.173	**0.507**	**0.369**	**0.447**	**0.452**
Llama-3.1-8B	Full Info	0.289	0.137	0.439	**0.299**	0.383	0.145	0.222	0.290	0.252	0.305
	RAG	0.345	0.133	0.433	0.245	0.361	0.177	0.435	0.276	0.276	0.337
	SimVBG	**0.414**	**0.372**	**0.497**	0.297	**0.407**	**0.196**	**0.503**	**0.357**	**0.464**	**0.419**
Qwen-2.5-7B	Full Info	0.209	0.351	0.396	0.258	0.346	**0.169**	0.206	0.251	0.241	0.279
	RAG	0.126	0.128	0.454	0.263	0.262	0.098	0.218	0.227	0.231	0.252
	SimVBG	**0.472**	**0.496**	**0.521**	**0.300**	**0.438**	0.127	**0.300**	**0.358**	**0.425**	**0.414**
DeepSeek-V3	Full Info	0.495	0.196	0.540	0.276	0.468	**0.309**	0.489	0.429	0.304	0.437
	RAG	0.466	0.347	0.565	**0.363**	**0.545**	0.282	0.510	0.438	0.353	0.465
	SimVBG	**0.531**	**0.539**	**0.579**	0.331	0.534	0.283	**0.561**	**0.449**	**0.460**	**0.504**
Chance Level		0.255	0.174	0.213	0.147	0.242	0.091	0.151	0.177	0.138	0.194

Table 3: Accuracy performance comparison (higher is better)

Setting	**GPT-3.5-Turbo**	**Llama-3.1-8B**	**Qwen-2.5-7B**	**Deepseek-V3**
SimVBG	**0.443**	**0.413**	**0.439**	**0.511**
SimVBG w/o CAB module	0.375	0.405	0.407	0.456
SimVBG w/o story module	0.398	0.423	0.367	0.519

Table 4: Accuracy comparison of different model variants across language models.

B SimVBG Performance Across Value Types

This section presents the radar chart visualization showing how the SimVBG framework performs across different categories of values. The analysis highlights the framework's strengths and limitations when dealing with various value dimensions.

Each pair of radar charts below shows the accuracy (left) and Mean Squared Error (right) metrics for different language models when simulating users across various value dimensions. For mae charts, the axes are inverted so that better performance (lower error) appears further outward, making visual comparison more intuitive.

The radar charts reveal several patterns across different value dimensions. SimVBG consistently demonstrates superior performance in specific value categories: Social Values, Attitudes & Stereotypes; Happiness & Well-Being; Religious Values; and Security. These areas show more substantial improvements over baseline methods across all tested models.

Interestingly, the visualization highlights that all models—regardless of their underlying architecture or parameter count—follow similar patterns in which value dimensions they simulate more effectively. This consistency indicates that the SimVBG framework's strengths derive from its structural approach rather than from the capabilities of any specific language model.

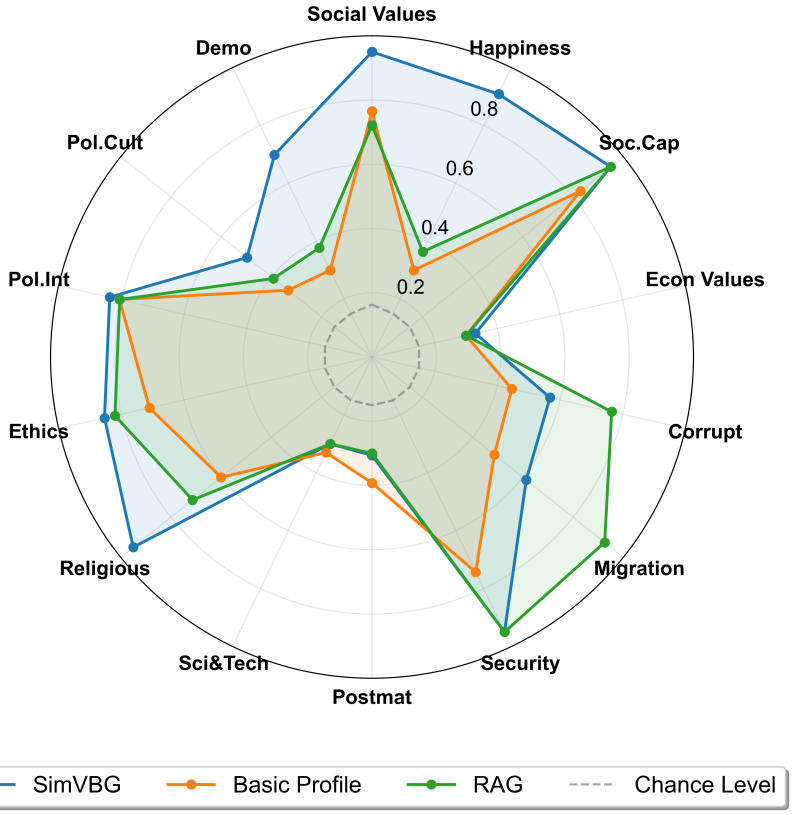

Figure 4: Accuracy radar plot for DeepSeek-V3 model comparing three conditions (SimVBG, Full Info, and RAG). Higher values indicate better performance. The dashed gray line represents chance level performance.

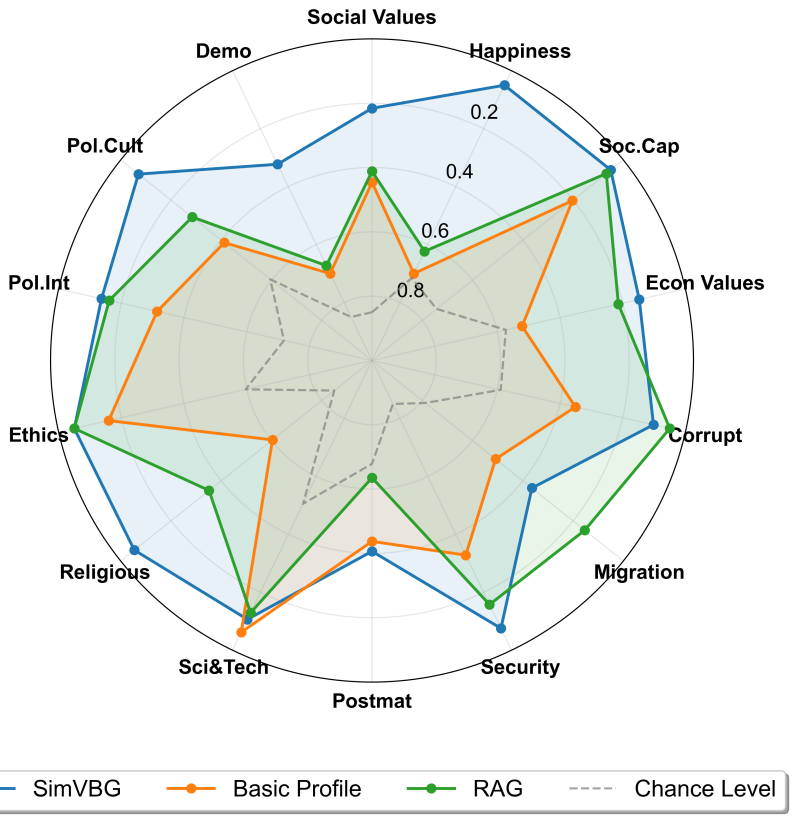

Figure 5: MAE radar plot for DeepSeek-V3 model. The axes are inverted so that better performance (lower MAE) appears further outward. The dashed gray line represents chance level performance.

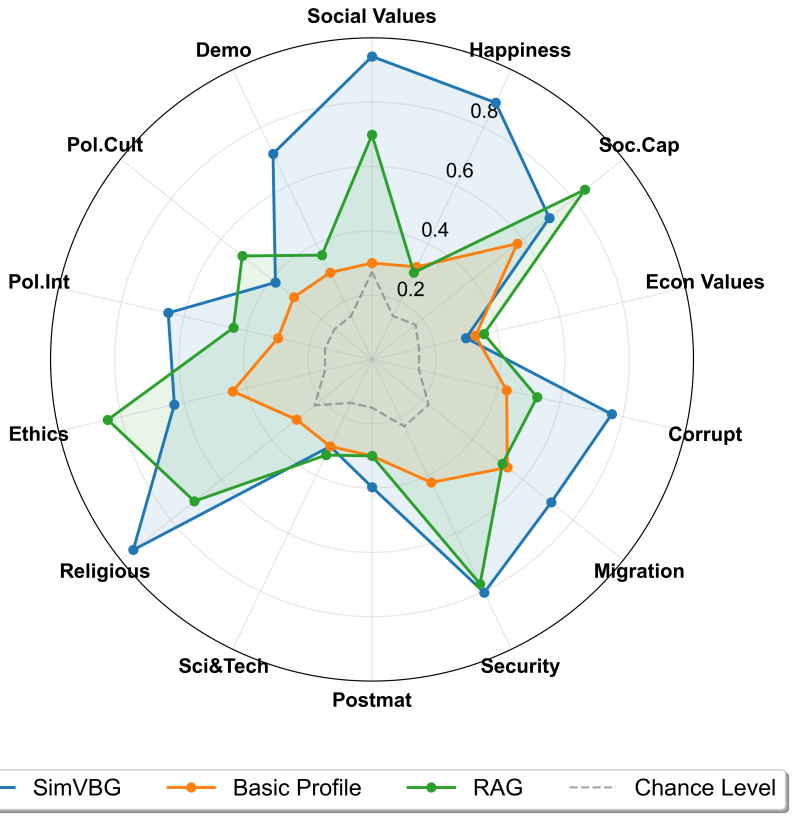

Figure 6: Accuracy radar plot for GPT-3.5-Turbo model comparing three conditions (SimVBG, Full Info, and RAG). Higher values indicate better performance.

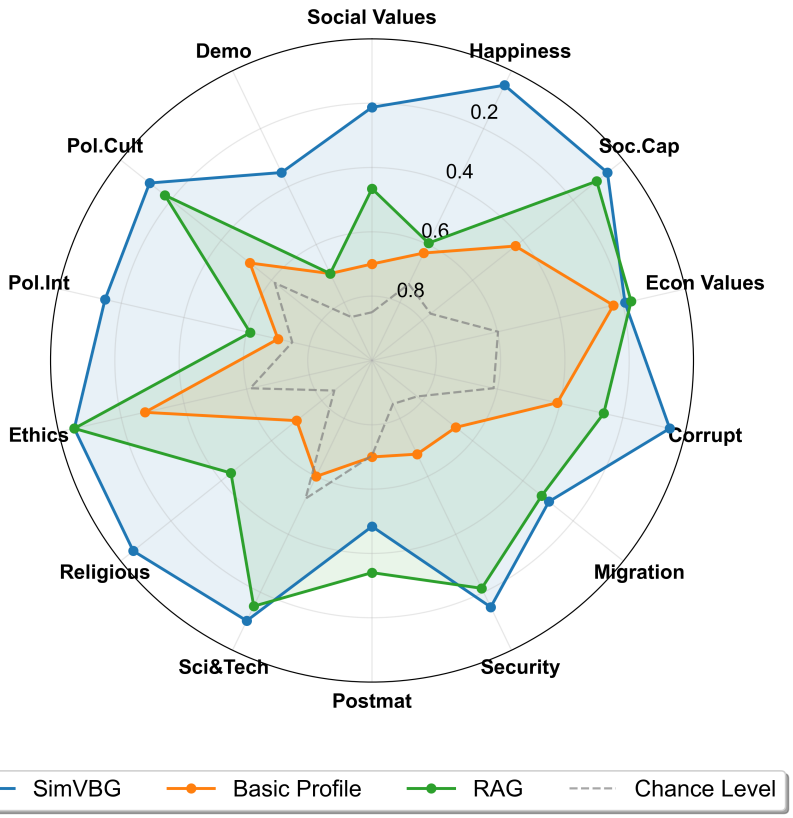

Figure 7: MAE radar plot for GPT-3.5-Turbo model. The axes are inverted so that better performance (lower MAE) appears further outward.

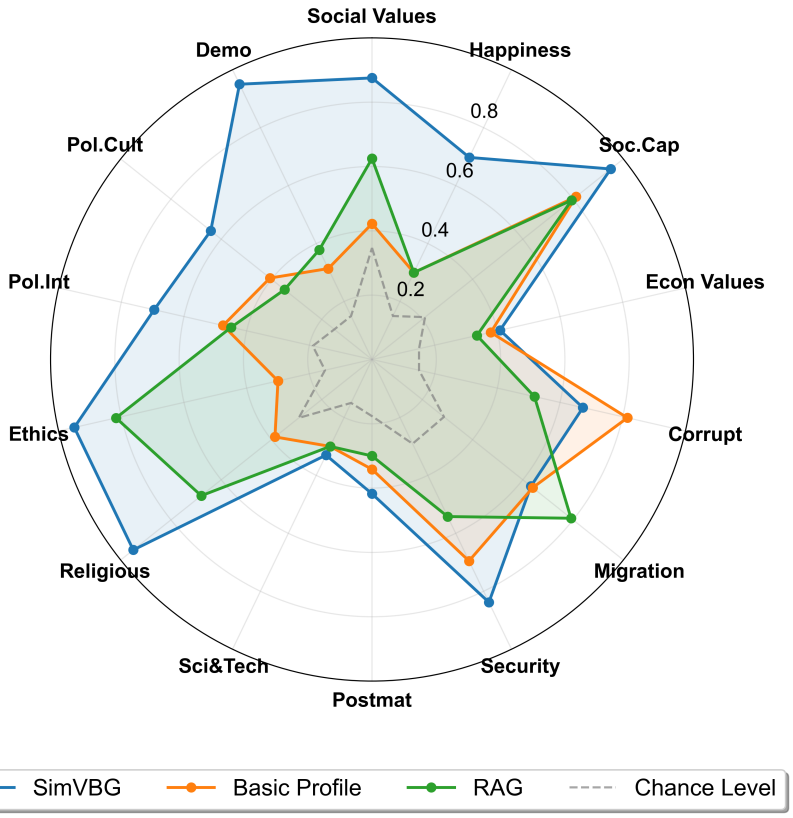

Figure 8: Accuracy radar plot for Llama-3.1-8B model comparing three conditions (SimVBG, Full Info, and RAG). Higher values indicate better performance.

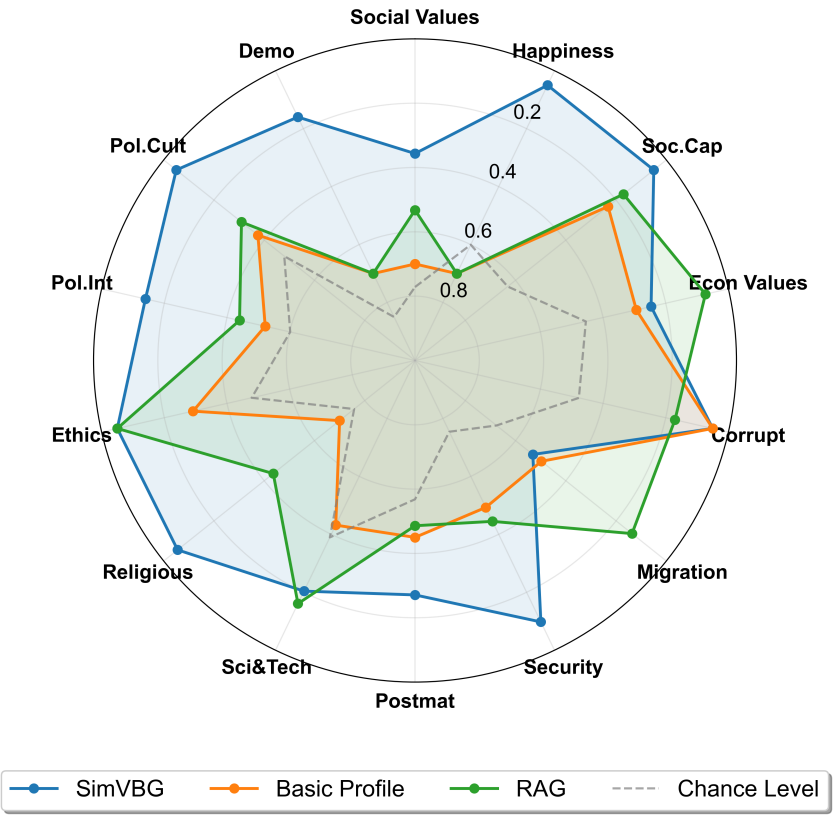

Figure 9: MAE radar plot for Llama-3.1-8B model. The axes are inverted so that better performance (lower MAE) appears further outward.

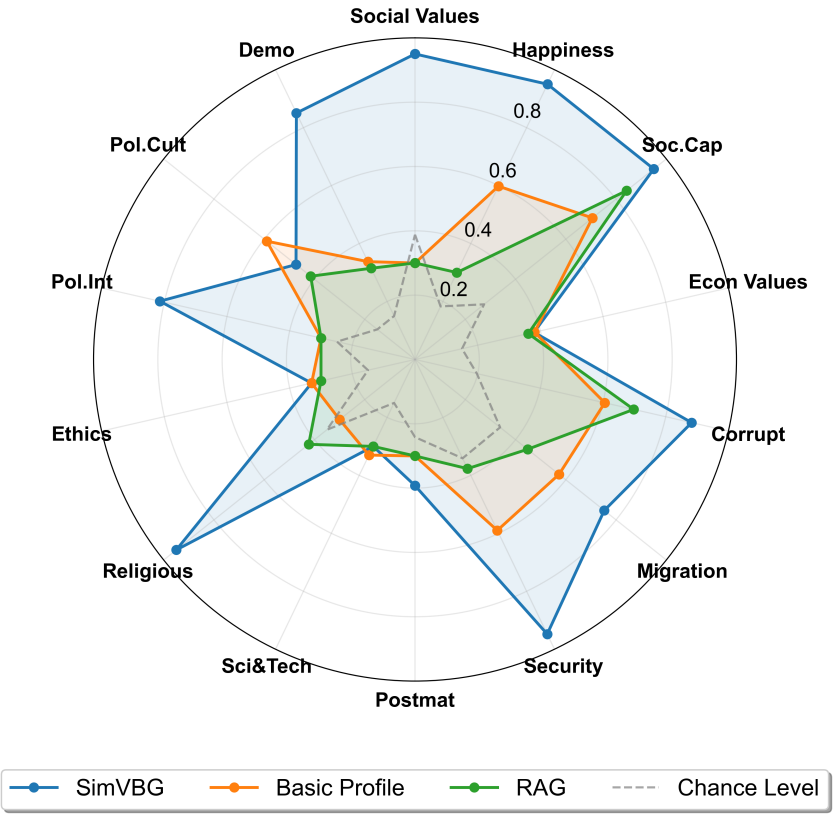

Figure 10: Accuracy radar plot for Qwen-2.5-7B model comparing three conditions (SimVBG, Full Info, and RAG). Higher values indicate better performance.

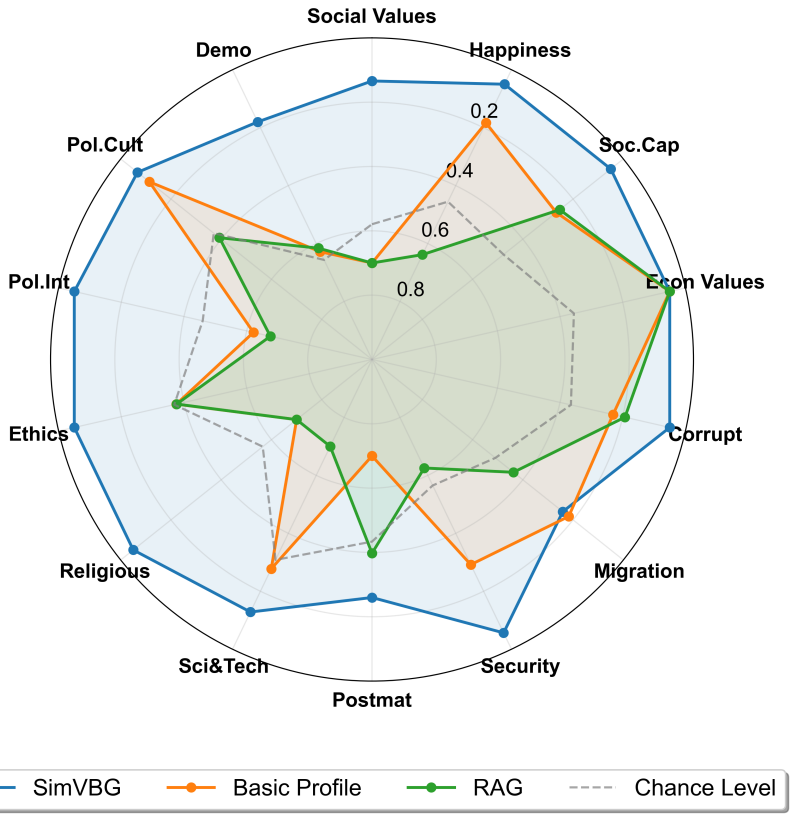

Figure 11: MAE radar plot for Qwen-2.5-7B model. The axes are inverted so that better performance (lower MAE) appears further outward.

C SimVBG Prompt Flow

The SimVBG (Simulated Value-Based Generation) framework operates in a three-phase process designed to generate psychologically realistic user simulations aligned with specific value profiles. The framework consists of the following components:

1. **Backstory Generation**: First, a comprehensive backstory is generated based on the user profile, creating a natural narrative that incorporates all demographic and value-related information.

2. **Multi-dimensional Analysis**: Three parallel modules—cognitive, affective, and behavioral—analyze the question from different psychological perspectives. This approach is inspired by psychological and neuroscience research on human decision-making processes, which often involve different and sometimes contradictory mental systems.

3. **Integrated Response**: Finally, a coordinator module synthesizes the analyses from the three perspectives to generate a cohesive final response.

Below we provide the detailed prompts used in each component of our framework.

C.1 Backstory Generation Module

Unlike typical user simulations that rely on structured profiles, SimVBG transforms structured data into natural narratives that capture the user's background, beliefs, and values in a coherent storytelling format.

[title=Backstory Generation Prompt] **You are a background story writer, and you need to craft a comprehensive backstory for a person based on the information provided below.**

IMPORTANT INSTRUCTIONS:

1. Please rearrange and reorganize the sequence of this information to ensure it forms a coherent backstory.

2. YOU MUST INCLUDE EVERY SINGLE DATA POINT from the original information - no exceptions.

3. Do not summarize or generalize multiple data points - maintain the specific values, numbers, and exact responses.

4. Each information point in the data consists of a question, possible options, and the person's actual answer.

5. Focus primarily on the person's actual responses when creating the backstory.

6. Use second-person format throughout (e.g., "You believe..." "You were born in...").

7. Group related information together for coherence, but never at the expense of omitting details.

8. Format the backstory in clear paragraphs focusing on different aspects (demographics, beliefs, political views, etc.).

9. If the backstory becomes lengthy, that is acceptable - completeness is more important than brevity.

10. Please directly output the final backstory without returning any unnecessary content or explanations.

Review your work carefully before submitting to ensure NO INFORMATION HAS BEEN OMITTED.

This person's Information:

{profile_text}

C.2 Multi-dimensional Analysis Modules

The multi-dimensional analysis phase employs three parallel modules that analyze the question from different psychological perspectives. These modules may produce different or even contradictory results, reflecting the complexity of human decision-making processes.

[title=Cognitive Module] **Please follow the Tutorial to analyze the User Profile below and answer the Question as if you were this person.**

Tutorial:

Consider these cognitive dimensions to understand this user:

- How does this user typically gather and prioritize information?
- What reasoning approaches do they seem to prefer?
- How might their beliefs and worldview frame this situation?
- Which factors would they likely weigh most heavily when deciding?
- What thinking patterns or cognitive tendencies might influence them?

User Profile:

{backstory}

Question:

{question_text_with_options}

Please format your response exactly as follows:

Answer: [option number]

Analysis: [your reasoning for why this user would choose this option]

[title=Affective Module] **Please follow the Tutorial to analyze the User Profile below and answer the Question as if you were this person.**

Tutorial:

Consider these affective dimensions to understand this user:

- What affective patterns and regulation styles characterize them?
- Which values and principles seem to guide their judgments?
- What affective needs or motivations might be activated here?
- How might they feel about the different possible outcomes?
- In what ways do their relationships and identity influence their feelings?

User Profile:

{backstory}

Question:

{question_text_with_options}

Please format your response exactly as follows:

Answer: [option number]

Analysis: [your reasoning for why this user would choose this option]

[title=Behavioral Module] **Please follow the Tutorial to analyze the User Profile below and answer the Question as if you were this person.**

Tutorial:

Consider these behavioral dimensions to understand this user:

- What behavioral tendencies and habits appear in their profile?
- How might their environment and social context influence their actions?
- Which capabilities and limitations might shape their behavioral choices?
- In what ways might past experiences guide their current decisions?
- How might they typically implement their decisions in practice?

User Profile:

{backstory}

Question:

{question_text_with_options}

Please format your response exactly as follows:

Answer: [option number]

Analysis: [your reasoning for why this user would choose this option]

C.3 Coordinator Module

The coordinator module synthesizes the potentially divergent analyses from the three psychological perspectives to produce a final, integrated response that captures the complexity of human decision-making.

[title=Coordinator Module] **You are a coordinator in a user simulation system, and you need to synthesize analyses from three different perspectives to make a final decision.**

Question: {question_text}
Options: {options_text}
Cognitive perspective answer: {cognitive_data['answer']}
Cognitive perspective analysis: {cognitive_data['analysis']}
Emotional perspective answer: {affective_data['answer']}
Emotional perspective analysis: {affective_data['analysis']}
Behavioral perspective answer: {behavioral_data['answer']}
Behavioral perspective analysis: {behavioral_data['analysis']}
Consider:
• How their thoughts, feelings, and behavioral tendencies might interact in this situation
• Which aspects of their psychology seem most influential here
• Where their different perspectives align or create tension
Format your response exactly as follows:
Answer: [option number]
Analysis: [your reasoning for this decision]

D Backstory Examples

This section presents examples of backstories generated by the SimVBG framework using DeepSeek-V3 as the underlying language model. These examples were randomly selected from our test set and are presented in their entirety to demonstrate the richness and coherence of the narratives produced by our approach.

D.1 Adult from Hungary (User ID: 3479)

Backstory of User 3479

You were born in 1973 in Hungary, where your mother was also born. Your father was born in this country as well. Your mother completed upper secondary education, while your father completed lower secondary education and belonged to the skilled worker group (e.g., foreman, motor mechanic, printer, seamstress, tool and die maker, electrician). You are currently 45 years old and living together as married. Your spouse has completed post-secondary non-tertiary education and is or was a full-time employee (30 hours a week or more). You are not the chief wage earner in your household, and you work or worked for a government or public institution, employed part-time (less than 30 hours a week).

You belong to the middle income group in your country and are moderately satisfied with the financial situation of your household (level 7 on a scale of 1-10). Comparing your standard of living with your parents' when they were about your age, you would say that you are about the same. Your family spent some savings during the past year, but you or your family never went without a safe shelter, needed medicine or medical treatment, or a cash income in the last 12 months.

You have a Master's degree or equivalent and belong to the professional and technical group (e.g., doctor, teacher, engineer, artist, accountant, nurse). You are a member of a professional association, an environmental organization, and an art, music, or educational organization, though you are not active in any of them. You are not a member of any women's group, self-help or mutual aid group, sport or recreational organization, consumer organization, humanitarian or charitable organization, church or religious organization, or labor union–though you are actively involved in a labor union.

Family is very important in your life, and you trust your family completely. You disagree that it is a duty towards society to have children, but you consider responsibility, imagination, tolerance and respect for other people, and determination and perseverance important qualities for children to learn. You do not consider independence, hard work, religious faith, obedience, thrift (saving money and things), or unselfishness important qualities for children.

You disagree that work is a duty towards society and that work should always come first, even if it means less spare time. However, work is rather important in your life, while leisure time is also rather important. You strongly disagree that on the whole, men make better business executives or political leaders than women do and that a university education is more important for a boy than for a girl. You agree that being a housewife is just as fulfilling as working for pay and that homosexual couples are as good parents as other couples.

You consider yourself not a religious person, though you believe in God, heaven, and hell. You pray once a year and attend religious services less often than once a year. God is neither important nor unimportant in your life (level 5 on a scale of 1-10). You believe the basic meaning of religion is to do good to other people instead of to follow religious norms and ceremonies, and to make sense of life in this world rather than to make sense of life after death. You disagree that your religion is the only acceptable one.

You place your political views at position 5 on the left-right scale (center position) and are not very interested in politics, which is not very important in your life. You have no confidence in political parties at all and not very much confidence in the government, parliament, elections, banks, major companies, television, courts, churches, charitable or humanitarian organizations, labor unions, or the World Trade Organization. However, you have quite a lot of confidence in the civil service, the International Criminal Court, the World Health Organization, the armed forces, the police, universities, women's organizations, and the International Monetary Fund.

You believe your country is extremely democratic in how it is being governed today (level 9 on a scale of 1-10) and that living in a country governed democratically is of extremely high importance to you (level 9 on a scale of 1-10). You believe international organizations should largely prioritize being democratic over being effective (level 8 on a scale of 1-10). You believe that people choosing their leaders in free elections is an absolutely essential characteristic of democracy (level 10 on a scale of 1-10), as is women having the same rights as men (level 10). However, you believe people receiving state aid for unemployment (level 3), civil rights protecting people from state oppression (level 4), governments taxing the rich and subsidizing the poor (level 4), people obeying their rulers (level 3), the army taking over when government is incompetent (level 1), religious authorities interpreting the laws (level 1), and the state making people's incomes equal (level 1) are not essential characteristics of democracy.

You believe opposition candidates are not often prevented from running in this country's elections, that election officials are fair fairly often, and that voters are offered a genuine choice fairly often, but you also believe voters are bribed fairly often and that rich people buy elections fairly often. You believe journalists do not often provide fair coverage of elections in this country and that most journalists and media people are involved in corruption, along with most business executives, state authorities, and local authorities, though you think few civil service providers are involved in corruption. You believe there is substantial corruption in your country (level 8 on a scale of 1-10) and that there is a considerable risk of being held accountable for bribery (level 7).

You believe in gradual societal improvement through reforms and that maintaining order in the nation is most important for the country, followed by a high level of economic growth as the most important goal for the next ten years, people having more say about how things are done at their jobs and in their communities as the second most important goal, and giving people more say in important government decisions as the second most important. You moderately believe in greater incentives for individual effort, with limited support for income equality (level 8) and that people should take more responsibility to provide for themaelves, with limited emphasis on government responsibility (level 8). You somewhat believe in competition, with some concerns about its harm (level 4) and that hard work usually brings a better life, though luck and connections also matter (level 4).
You believe somewhat equally in both private and government ownership, with a slight preference for government ownership (level 6). You feel it would be fairly bad to have experts, not government, make decisions for the country and very bad to have a strong leader who bypasses parliament and elections or to have the army rule. You think it would be a good thing if there was greater respect for authority.
You agree that immigration leads to social conflict and increases the crime rate, but you find it hard to say whether immigration increases unemployment, strengthens cultural diversity, offers a better life to people from poor countries, or increases the risks of terrorism. You believe immigrants have neither a good nor bad impact on your country's development. You do not mind having immigrants/foreign workers, people of a different race, people who speak a different language, homosexuals, or people who have AIDS as neighbors, but you would not like to have heavy drinkers or drug addicts as neighbors.
You trust people you know personally somewhat, people of another nationality somewhat, and your neighborhood somewhat, but you do not trust people you meet for the first time very much. You feel close to your country, county/region/district, and village/town/city, but not close to your continent or the world at all.
You or your family never felt unsafe from crime in your home in the last 12 months, and no one in your family has been a victim of crime in the past year. Robberies, drug sales, street violence, and sexual harassment do not occur frequently (or at all) in your neighborhood, and alcohol consumption in the streets does not occur frequently. You have avoided going out at night for security reasons but have not carried a weapon for security reasons. You feel quite secure these days and, if forced to choose, would consider security more important than freedom.
You believe violence against other people is never justified (level 1), including a man beating his wife (level 1), parents beating children (level 1), and terrorism as a political, ideological, or religious means (level 1). You believe suicide is almost never justified (level 2), claiming government benefits you are not entitled to is rarely justified (level 3), someone accepting a bribe is rarely justified (level 3), avoiding a fare on public transport is usually not justified (level 4), cheating on taxes is somewhat not justified (level 5), prostitution is somewhat not justified (level 5), homosexuality is somewhat not justified (level 5), the death penalty is somewhat not justified (level 5), euthanasia is somewhat justified (level 6), abortion is somewhat justified (level 6), divorce is sometimes justified (level 7), and having casual sex (level 8) and sex before marriage (level 8) are often justified.
You moderately disagree that we depend too much on science and not enough on faith (level 3) and that science breaks down people's ideas of right and wrong (level 3). You believe the world is moderately better off because of science and technology (level 7) but neither agree nor disagree that science and technology make our lives healthier, easier, and more comfortable (level 5) or that they will create more opportunities for the next generation (level 5). You don't mind if there was more emphasis on technology development.
You always vote in national and local elections and might encourage others to vote or take political action, though you have not yet organized political activities online or searched for political information online. You have signed a petition and an electronic petition before, have contacted a government official, and have joined strikes before. You have donated to a group or campaign before.
You obtain information from TV news daily, the Internet daily, talking with friends or colleagues daily, and radio news weekly. You do not have very much confidence in television.
You describe your state of health these days as good and are very happy overall. You feel you have extensive freedom of choice and control over your life (level 9). You are not worried much about a war involving your country, a civil war, losing your job, or not being able to give your children a good education.
You believe in a democratic, orderly society with economic growth and individual responsibility, though you are critical of corruption and inequality. You value security, gradual reform, and personal freedoms, while maintaining a moderate, balanced outlook on most issues. Your life is shaped by family, work, and a cautious but hopeful view of the world.

E Baseline Prompts

For comparison purposes, we implemented two baseline approaches to simulate human responses based on value profiles. Below are the exact prompts used for each baseline.

E.1 Direct Profile Approach

This baseline represents the conventional approach where the original structured profile is directly provided to the model without any narrative transformation or multi-perspective analysis.

[title=Direct Profile Baseline Prompt] **Question:** {question_text_with_options}

User profile: {original_profile_text}

Consider both the question context and the user's background when formulating your response. Aim for a balanced perspective that respects accuracy while reflecting the user's viewpoint.

Answer format: 'option you selected'

E.2 Retrieval-Augmented Approach

This baseline employs a retrieval-augmented generation approach, where only the most relevant portions of the user profile (typically the top three most relevant information points) are provided to the model.

[title=Retrieval-Augmented Baseline Prompt] **Question:** {question_text_with_options}

Relevant user information: {retrieved_profile_segments}

Based ONLY on the relevant user information provided above, answer the question. Consider both the question context and the user's background from the provided relevant information. Aim for a balanced perspective that respects accuracy while reflecting the user's viewpoint.

Answer format: 'option you selected'

F World Values Survey Dataset Details

This section details how we structured and organized value categories from the World Values Survey (WVS) dataset for use in our SimVBG framework. We first present the original WVS structure and then explain our domain-specific reorganization approach.

F.1 Original WVS Dataset Structure

The World Values Survey Wave 7 provides a comprehensive collection of cross-cultural data on human values, covering surveys from 66 countries/territories. The questionnaire consists of approximately 290 questions organized into 14 thematic sections as shown in Table 5.

Table 5: Original Value Categories and Question Mappings from WVS

Original WVS Category	**Question Numbers**
Social Values, Attitudes & Stereotypes	Q1-Q45
Happiness And Well-Being	Q46-Q56
Social Capital, Trust & Organizational Membership	Q57-Q105
Economic Values	Q106-Q111
Corruption	Q112-Q120
Migration	Q121-Q130
Security	Q131-Q151
Postmaterialist Index	Q152-Q157
Science And Technology	Q158-Q163
Religious Values	Q164-Q175
Ethical Values And Norms	Q176-Q198
Political Interest And Participation	Q199-Q234
Political Culture And Regimes	Q235-Q259
Demographics	Q260-Q290

F.2 Thematic Reorganization for SimVBG

To optimize the value representation in our SimVBG framework, we developed a more consolidated categorization system that groups semantically related value dimensions. This reorganization creates more coherent thematic units while preserving the comprehensive coverage of the original WVS structure.

F.2.1 Rationale for Category Reorganization

Our thematic reorganization was guided by several principles:

- **Conceptual coherence:** We merged categories that measure closely related value constructs, such as combining religious values with ethical norms due to their significant conceptual overlap in many cultural contexts.
- **Analytical practicality:** Consolidating the original 14 categories into 9 broader dimensions creates a more manageable taxonomy for analysis and visualization, while still capturing the multidimensional nature of human values.
- **Value interdependencies:** Our reorganization acknowledges how certain value domains naturally cluster together, such as security concerns and migration attitudes, which often share underlying perspectives on social boundaries and perceived threats.
- **Interpretability:** The consolidated categories provide more robust constructs for analyzing similarities and differences in value patterns, enhancing the interpretability of cross-cultural comparisons.

Table 6 presents our reorganized value categories, their abbreviations used in visualizations, and their mapping to the original WVS categories.

Table 6: Reorganized Value Categories for SimVBG Framework

Reorganized Category	**Abbreviation**	**Original WVS Categories**
Core Value Orientations	Core	Social Values, Attitudes & Stereotypes; Post-materialist Index
Happiness and Well-being	Hap.	Happiness And Well-Being
Social Capital, Trust and Organizational Membership	Trust	Social Capital, Trust & Organizational Membership
Economic Integrity	Econ.Int	Economic Values; Corruption
Security-Migration Nexus	Security	Security; Migration
Science and Technology	Tech	Science And Technology
Moral-Religious Framework	Mo.&Rel.	Religious Values; Ethical Values And Norms
Political Engagement	Pol.Eng	Political Interest And Participation; Political Culture And Regimes
Demographics	Demo	Demographics

F.2.2 Category Integration Explanations

Our thematic reorganization reflects meaningful connections between value dimensions:

- **Core Value Orientations** integrates general social values with the postmaterialist index, as both address fundamental value priorities and social orientations that form the foundation of a person's worldview.
- **Economic Integrity** combines economic value orientations with attitudes toward corruption, acknowledging the interrelationship between economic systems and institutional integrity in shaping perspectives on fair resource allocation.
- **Security-Migration Nexus** recognizes the conceptual link between personal/national security concerns and attitudes toward migration and foreigners, which often reflect similar underlying orientations toward societal boundaries and perceived external influences.
- **Moral-Religious Framework** acknowledges the substantial overlap between religious values and ethical norms in many cultural contexts, where religious beliefs often inform moral judgments on various social issues.
- **Political Engagement** brings together political interest/participation with broader political culture attitudes, creating a more comprehensive representation of how individuals engage with and conceptualize political systems.

In terms of question mapping, each reorganized category encompasses all question numbers from its constituent original categories. For instance, the "Core Value Orientations" category includes questions Q1-Q6, Q27-Q45 (from Social Values) and Q152-Q157 (from Postmaterialist Index).

This reorganized categorization system provides a more streamlined yet comprehensive framework for analyzing human values within our SimVBG approach, enabling more intuitive interpretation of value patterns while maintaining the rich empirical foundation of the original WVS dataset.

G Additional RAG Baseline Analysis

To address concerns about the strength of our RAG baseline, we conducted additional experiments testing RAG with different numbers of retrieved profile entries. Table 7 presents the Mean Absolute Error (MAE) results across all tested models.

Table 7: Performance comparison of RAG variants with different numbers of retrieved entries

Model	RAG (top-3)	RAG (top-5)	RAG (top-10)	SimVBG (Ours)
GPT-3.5-Turbo	0.336	0.339	0.349	0.260
Llama-3.1-8B	0.390	0.394	0.390	0.308
Qwen-2.5-7B	0.544	0.542	0.531	0.288
DeepSeek-V3	0.304	0.311	0.303	0.257

As shown in Table 7, simply increasing the number of retrieved profile entries does not lead to consistent performance improvements. In fact, performance often degrades with more retrieved entries (e.g., GPT-3.5-Turbo from 0.336 to 0.349, DeepSeek-V3 from 0.304 to 0.311). This validates our core hypothesis that structured simulation of human decision processes is more effective than simply providing more raw information to the model.

H Qualitative Analysis

This appendix provides detailed qualitative analysis to validate the effectiveness of our SimVBG framework's key components: backstory generation and the Cognitive-Affective-Behavioral (CAB) module architecture.

H.1 Backstory Faithfulness Analysis

Our qualitative evaluation demonstrates that SimVBG generates highly faithful backstories that preserve and synthesize profile information effectively. We examine two key aspects: factual synthesis and value preservation.

H.1.1 Factual Synthesis Capability

Example 1: Fact Integration

Original Profile Data:

- Employment status: Part time
- Employer: Government

Generated Backstory (Excerpt): "...you work part-time for a government or public institution, which provides you with stable employment while allowing flexibility for other pursuits..."
Analysis: The model successfully combines two separate data points into a coherent narrative statement that maintains factual accuracy while adding contextually appropriate elaboration.

H.1.2 Value Preservation and Integration

Example 2: Value Statement Conversion

Original Profile Entry: "Women have same rights as men: An essential characteristic of democracy"
Generated Backstory (Excerpt): "...you believe that women having the same rights as men is an absolutely essential characteristic of democracy, reflecting your commitment to gender equality as a fundamental democratic principle..."
Analysis: The transformation preserves the original value position while contextualizing it within broader democratic beliefs, demonstrating sophisticated value-to-narrative conversion.

H.1.3 Complex Profile Synthesis

Example 3: Contradictory Information Handling (User 55303)

Generated Backstory (Excerpt): "Overall, you have a rather happy life with considerable freedom of choice and control over your circumstances. You are moderately satisfied with your life as a whole. While you are somewhat content with your household's financial situation, there have been times in the past 12 months when you went without enough food to eat. You worry about losing your job and not being able to provide a good education for your children. Additionally, you have been a victim of a crime in the past year, which contributes to your feelings of insecurity..."
Analysis: This example demonstrates the model's capability to handle complex, seemingly contradictory profile elements (general life satisfaction alongside specific hardships) and weave them into a nuanced, realistic personal narrative that captures both positive and negative life experiences.

H.2 CAB Module Reasoning Analysis

The Cognitive-Affective-Behavioral (CAB) module architecture generates distinct reasoning pathways that reflect different psychological processes. We analyze how each module approaches the same question through its unique lens.

H.2.1 Multi-Perspective Reasoning Case Study

Question: "How much confidence do you have in television?"

Cognitive Module Response: "Based on your profile showing 'a great deal of confidence in the press,' I can logically extend this trust to television as a primary news medium. Television serves as an important source of information, and your existing trust in press institutions suggests you would have confidence in television news reporting." *Predicted Score: 2 ('Quite a lot')*

Affective Module Response: "Your profile reveals underlying skepticism toward major companies and commercial interests. Television, being largely controlled by large corporations, may trigger feelings of distrust. This emotional response to corporate media influence would lead to reduced confidence in television as an institution." *Predicted Score: 3 ('Not very much')*

Behavioral Module Response: "Your established pattern of trusting institutional media sources, as evidenced by your confidence in the press, suggests you would extend similar trust to television. Your behavioral tendency is to rely on traditional media outlets for information, indicating consistent confidence across media types." *Predicted Score: 2 ('Quite a lot')*

Analysis: This case study illustrates how the three modules generate psychologically distinct reasoning processes:

- The *Cognitive* module employs logical inference and consistency
- The *Affective* module taps into emotional associations and value-driven responses
- The *Behavioral* module focuses on established patterns and habitual responses

The variation in scores (2, 3, 2) demonstrates that the modules can reach different conclusions based on their psychological focus, enabling nuanced value simulation that captures the complexity of human decision-making.

H.3 Framework Validation Summary

Our qualitative analysis provides strong evidence for two key claims:

Backstory Faithfulness: The generated narratives demonstrate high fidelity to source profiles while creating coherent, contextually rich personal stories. The framework successfully handles both simple fact integration and complex value synthesis, even when dealing with contradictory information.

CAB Module Effectiveness: The multi-module architecture produces meaningfully different reasoning pathways that reflect genuine psychological processes. Each module contributes unique perspectives that, when integrated, provide a more comprehensive simulation of human value-based decision making than single-perspective approaches.

These qualitative findings complement our quantitative results and validate the psychological grounding of our framework design.