Large Language Model Agents in Psychology: A Survey

www.surveyx.cn

Abstract

Large Language Model (LLM) agents are revolutionizing psychological research by simulating human-like understanding and responses through advanced natural language processing and cognitive computing. This survey explores their multifaceted roles in psychology, including their application in clinical and therapeutic settings, social simulations, and personality and emotion modeling. LLMs enhance psychological methodologies by streamlining hypothesis generation, data processing, and experimental design. Despite their transformative potential, challenges such as technical constraints, cognitive simulation limitations, and ethical concerns persist. Addressing these issues is crucial for optimizing LLM integration into psychological frameworks. Future research should focus on interdisciplinary collaboration, expanding multimodal approaches, and refining training methodologies to enhance LLM adaptability and reliability. By overcoming these challenges, LLM agents can significantly advance psychological research and practice, offering innovative solutions for understanding and influencing human behavior.

1 Introduction

1.1 Significance and Impact of LLM Agents in Psychology

Large Language Model (LLM) Agents have emerged as transformative tools in psychology, significantly enhancing research methodologies and practical applications. Their advanced natural language processing and cognitive computing capabilities facilitate human-computer interactions that mimic human understanding and responses, opening new avenues for psychological exploration and intervention [1]. The integration of LLMs into psychological frameworks is reshaping various branches of the discipline, from cognitive to social psychology, by introducing methodological innovations [2].

LLMs' ability to simulate human decision-making processes provides valuable insights into cognitive strategies, although challenges persist in their responses to probability questions, which may lead to cognitive biases like the conjunction fallacy [1]. Aligning LLMs with human values is essential to mitigate potential adverse effects from misuse. Additionally, the evolution of LLMs raises questions about their psychological attributes and stability, necessitating a comprehensive evaluation framework inspired by psychometrics [1].

In mental health, LLM agents are crucial in addressing the growing demand for assistance amid a shortage of professionals, thereby improving counseling effectiveness [3]. They bridge the social support gap in digital health interventions by providing personalized therapy, overcoming institutional barriers [3]. Their role in enhancing educational and psychological support systems is increasingly recognized, as existing intelligent dialog systems face limitations in these areas [4].

The rapid evolution of chatbot development, propelled by advancements in LLM APIs, has highlighted the need for reliability in LLM-powered applications [2]. Furthermore, the issue of hallucinations, where models confidently output misinformation, poses significant challenges due to their widespread use and potential consequences [5].

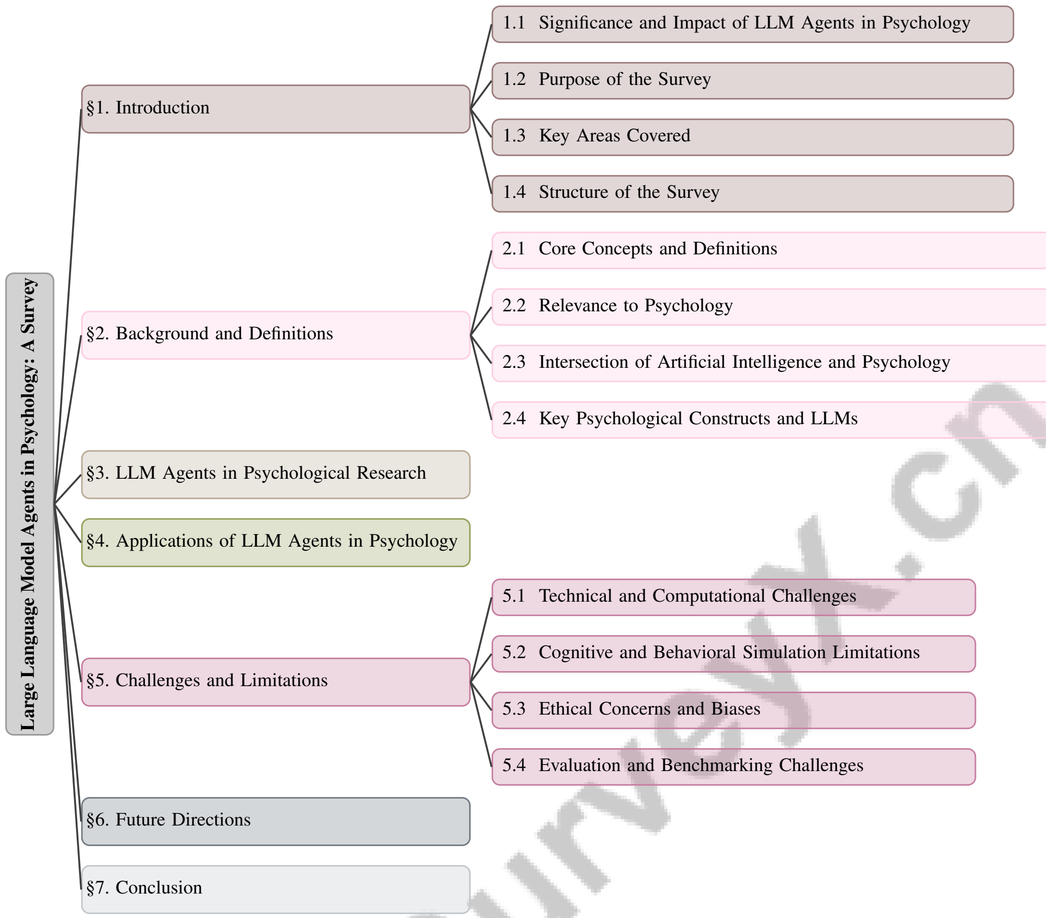


Figure 1: chapter structure

As LLMs continue to evolve, their role in advancing psychological research and practice grows increasingly significant. They present unique opportunities for enhancing data privacy and accessibility in psychological assessments, particularly in sensitive contexts such as evaluating suicidal risk [6]. Evaluating LLMs as general-purpose agents is crucial for understanding their capabilities and facilitating their integration into practical applications [7]. Given their ability to simulate human-like social interactions, addressing concerns regarding their impact on public opinion and societal polarization is imperative [8]. The transformative potential of LLM agents in psychology is profound, promising to shape future directions and applications in the field.

The emergence of research using psychological tools to study LLMs underscores their significance in understanding context-dependence in value expression [9]. Integrating cognitive ergonomics into LLM design aims to enhance safety, reliability, and user satisfaction in human-AI interactions [10]. The concept of thinking assistants leverages LLMs to encourage users to engage in deep reflection and critical thinking through brainstorming and thought-provoking queries [11]. These advancements highlight the broad impact of LLMs across diverse psychological domains, emphasizing their potential to revolutionize the field.

1.2 Purpose of the Survey

This survey systematically explores the multifaceted roles and capabilities of Large Language Model (LLM) agents within psychology. It emphasizes the potential of LLMs to enhance the understanding of human cognitive processes, social interactions, and decision-making. By examining the intersection of LLMs and human cognition—particularly in language comprehension and memory—this survey elucidates how these models can simulate human-like reasoning and social intelligence. A critical focus is placed on the determinants of LLM-assisted decision-making, highlighting technological,

AI-generated, for reference only.

psychological, and decision-specific factors, while addressing the absence of validated psychometric methods for quantifying human personality traits within LLMs.

Additionally, the survey investigates collaboration strategies among LLM agents characterized by different traits and thinking patterns [10]. It explores the integration of computational experiments with LLMs to enhance modeling of complex systems, addressing limitations of traditional Agent-based Modeling (ABM) in representing real social systems [12]. The survey provides a comprehensive overview of advancements in LLMs and Multimodal Large Language Models (MLLMs), addressing their evolution and capabilities across various tasks [13].

Furthermore, the survey offers an extensive overview of Affective Computing in the context of LLMs, focusing on traditional tasks like Affective Understanding and Affective Generation [14]. It addresses the isolation of LLMs in their conventional form and proposes a collaborative framework to improve their utility in complex scenarios [11]. By bridging the gap between research on LLMs and their real-world implementation as autonomous agents, the survey provides actionable insights and considerations [10].

Moreover, the survey examines linguistic features in therapeutic dialogues between human psychologists and AI agents, particularly using the MISTRAL-7B Large Language Model. It details advancements in LLM-based persuasion, highlighting how LLMs shape human attitudes and behaviors across various fields such as politics, marketing, public health, e-commerce, and charitable giving. The survey identifies critical factors enhancing their persuasive power, including content personalization and AI involvement disclosure. It also addresses ethical and societal implications associated with LLMs, such as the potential for misinformation, amplification of biases, and privacy concerns, emphasizing the urgent need for ethical guidelines and regulatory measures to mitigate risks posed by these technologies [2, 15, 16, 17, 18]. By systematically reviewing these diverse aspects, the survey aims to establish a robust foundation for understanding the current state and future potential of LLM agents in psychology, ultimately contributing to the advancement of both artificial intelligence and psychological research.

1.3 Key Areas Covered

This survey comprehensively addresses the multifaceted roles of Large Language Model (LLM) agents within psychology, emphasizing their transformative potential across various domains. It explores the application of psychological scales, notably the Big Five Inventory, to different LLMs such as GPT-3.5, GPT-4, Gemini-Pro, and LLaMA-3.1, providing insights into their psychological profiling capabilities [19]. The practical deployment of LLMs as agents in real-world scenarios is categorized into Planning, Memory, Tools, and Control Flow, highlighting their versatility [20].

The survey examines the integration of LLMs in multi-agent systems, exploring applications in software development, multi-robot systems, society simulation, policy simulation, and game simulation, which underscore their utility in complex, dynamic environments [21]. The interaction of LLM agents with distinct traits and thinking patterns is analyzed within simulated societies, incorporating social psychology theories and collaboration strategies, supported by empirical evaluations across benchmark datasets [22].

Additionally, the survey encompasses both single-agent and multi-agent systems, focusing on LLM-based agents' definitions, frameworks, cognitive methods, tool utilization, and deployment strategies, providing a comprehensive overview of their operational dynamics [23]. It also addresses the technical limitations, ethical considerations, and complexities involved in responsibly integrating AI into sensitive psychological contexts [18].

Advancements in LLM serving systems, including memory management, computation optimization, cloud deployment, and emerging research fields from January 2023 to June 2024, are critically assessed to provide insights into their evolving technological landscape [17]. The survey explores the historical context and theoretical frameworks for feedback generation in educational settings, emphasizing LLMs' role in enhancing feedback design [4].

Finally, the integration of LLMs into psychotherapy is scrutinized, covering their technical overview, stages of integration, applications in clinical care, and recommendations for responsible development, thus highlighting their potential to revolutionize therapeutic practices [24]. The survey discusses the six core functional components of LLMGAs: perception, memory, thinking, role-playing, action, and learning, providing a structured understanding of their operational capabilities [25].

1.4 Structure of the Survey

This survey is organized into several key sections, each addressing distinct aspects of Large Language Model (LLM) agents within psychology. The introduction sets the stage by discussing the significance and impact of LLM agents, outlining the purpose and objectives of the survey, and highlighting the key areas covered. The background and definitions section provides foundational knowledge on core concepts, including definitions of LLMs and their relevance to psychology, as well as the intersection of artificial intelligence and psychology.

Subsequent sections delve into the application of LLM agents in psychological research, examining their role in simulating human-like understanding, analyzing language patterns, and enhancing research methodologies. The survey explores various applications of LLM agents in psychology, including clinical and therapeutic settings, social simulations, personality and emotion simulation, educational and developmental psychology, and psychological experimentation.

The challenges and limitations associated with LLM agents are comprehensively identified and discussed, encompassing technical difficulties related to unpredictable behavior, cognitive simulation limitations in accurately emulating human thought processes, ethical concerns regarding data privacy and responsible usage in sensitive applications like psychology, and evaluation difficulties stemming from the lack of consensus on effective assessment methods for LLM outputs compared to human evaluations [2, 15, 26, 20, 18]. The survey concludes with a discussion on future directions, proposing interdisciplinary integration and collaboration, expanding multimodal approaches, and optimizing training for psychological applications. By systematically addressing these topics, the survey aims to provide a comprehensive overview of the current state and future potential of LLM agents in psychology, offering valuable insights for both researchers and practitioners.The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Core Concepts and Definitions

Large Language Models (LLMs) represent a significant leap in artificial intelligence, excelling in language tasks like translation and summarization through deep neural networks and transformer architectures [27]. They possess the unique ability to interpret novel metaphors, yet face challenges in maintaining contextual coherence, particularly in complex tasks and multi-agent systems. This is compounded by their tendency to generate hallucinated information—outputs that appear plausible but lack data grounding, potentially misleading users [28]. The debate over LLMs' ontological status questions whether they are genuine agents or mere statistical processors [13].

LLM Agents extend LLMs' capabilities by operating autonomously, simulating cognitive processes like strategic reasoning in dynamic environments [9]. However, unpredictability and the gap between theoretical research and practical applications challenge their implementation as autonomous agents [29]. Consistent persona maintenance is crucial for effective multi-agent collaboration, as personality traits significantly influence negotiation outcomes [10]. Efficient information gathering in unfamiliar environments remains a core challenge [14].

Cognitive computing aims to replicate human thought processes, enhancing LLMs' reasoning capabilities and addressing cognitive biases. Frameworks like Learning through Communication (LTC) improve LLM adaptability and learning efficiency through linguistic feedback and non-linguistic rewards. These advancements highlight ongoing refinements in LLMs to manage complex psychological tasks effectively. Safety concerns, particularly regarding biased or unethical content generation, necessitate careful deployment considerations [6].

2.2 Relevance to Psychology

LLMs are pivotal in psychology, offering innovative methodologies for exploring cognitive processes and social behaviors. Their simulation of intricate human interactions aids in studying opinion dynamics within networks, mirroring real conversations [30]. Memory architecture advancements, inspired by cognitive psychology, enhance LLM simulations of human cognition [31].

In clinical settings, LLMs facilitate the examination of personality traits and behaviors, deepening psychological construct understanding. Ethical deployment requires collaboration between technologists and behavioral health professionals [24]. Benchmark development for evaluating reasoning capabilities in strategic games underscores the importance of cooperation and language comprehension [32].

LLMs address emotional challenges, such as uncertainty and anxiety, with platforms like Gradschool.chat supporting students [11]. Theoretical perspectives focusing on behavioral patterns provide fresh insights into phenomena like deception [8]. Challenges include reliance on pre-existing knowledge, real-world grounding difficulties, and the need for effective multimodal input integration [25]. Ensuring interpretability in assessments and aligning reasoning with human moral standards are essential for psychological research applications [28].

Current benchmarks emphasize surface-level understanding, highlighting the need for deeper semantic comprehension for effective application in complex environments [27]. Leveraging these capabilities can significantly advance psychological theories, although challenges remain in applying cognitive science methods to mitigate biases and ensure user-centered design [10, 12]. Exploring LLM-based frameworks enriches their relevance to psychology [14].

2.3 Intersection of Artificial Intelligence and Psychology

The convergence of artificial intelligence and psychology provides a compelling landscape for examining LLMs' simulation and enhancement of human cognitive and social processes. This intersection is significant in understanding LLM agents' cooperative behavior in competitive scenarios, highlighting their adaptive problem-solving capabilities [33]. The self-organization of LLM agents into networks reflects complex human social dynamics [34].

Exploring cognitive biases within LLMs, particularly hallucinations, through psychological concepts offers insights into mitigation strategies [5]. Positioning LLMs as 'homo silicus' suggests they can simulate human economic behaviors, offering a novel approach to modeling decision-making processes [35].

Benchmark development for evaluating LLM agents emphasizes understanding goal-achievement processes, reflecting their cognitive and strategic capabilities in complex environments [36]. The philosophical implications of LLMs' cognitive capabilities raise important considerations for AI ethics and safety [6].

Computational experiments, categorized into generative experiments and deduction, enhance modeling of complex systems, improving human-like behavior representation in LLMs. These frameworks facilitate a nuanced understanding of AI and psychological research interplay, paving the way for innovative approaches to studying cognitive and social phenomena [12]. The intersection of AI and psychology promises to revolutionize cognitive process and social interaction understanding, offering transformative potential for both fields.

2.4 Key Psychological Constructs and LLMs

The interplay between psychological constructs and LLM capabilities provides insights into emulating and enhancing human cognitive processes. Integrating human-like memory processes into LLM dialogue agents enhances interaction quality, highlighting memory's role in cognitive psychology [37]. This aligns with advancing LLMs' abilities to simulate human-like reasoning.

Frameworks categorizing reasoning abilities of LLMs and humans elucidate performance differences, essential for understanding LLMs' replication of cognitive functions [38]. Distinguishing between System 1 (intuitive) and System 2 (deliberate) processes parallels dual-process theories in psychology [39].

The State-Understanding-Value-Action (SUVA) framework assesses LLM behaviors in social contexts, reflecting constructs like social cognition and theory of mind [40]. It aids in understanding LLMs' simulation of social interactions and adaptive behaviors.

LLMs develop cooperative strategies organically, showcasing potential to simulate emergent social behaviors [33]. This indicates their capacity to mirror human social dynamics and decision-making processes.

Evaluating LLM agents requires comprehensive frameworks ensuring safety, performance, and goal alignment, addressing cognitive reliability and ethical deployment concerns [41]. These frameworks distinguish genuine cognitive abilities from AI shortcuts, ensuring assessments reflect true competencies [42].

Categorizing argument effectiveness based on social pragmatics dimensions aligns with psycholinguistic theories of opinion change, illustrating LLMs' influence on attitudes through persuasive communication [43]. This capability highlights the intersection of language psychology and cognitive modeling, offering new avenues for exploring persuasion and influence.

Exploring psychological constructs related to LLM capabilities reveals their potential to simulate complex cognitive and social processes. Leveraging psychological theory-grounded frameworks enhances understanding and application of LLMs in psychological contexts, paving the way for innovative approaches to studying and influencing human behavior. Surveying multimodal LLMs (MLLMs) underscores the importance of architecture, training strategies, and evaluation practices in enhancing model performance, contributing to a deeper understanding of their psychological applications [44].

In recent years, the integration of Large Language Model (LLM) agents into psychological research has garnered significant attention. These agents are not only capable of simulating human-like understanding but also play a crucial role in analyzing language patterns and enhancing research methodologies. As depicted in Figure 2, the hierarchical structure of LLM agents is illustrated, categorizing their various roles and responsibilities. This figure highlights the frameworks and applications that contribute to these areas, thereby demonstrating the transformative impact these agents have on psychological studies. Such a comprehensive understanding of their capabilities allows researchers to leverage LLM agents effectively, fostering innovative approaches to psychological inquiry.

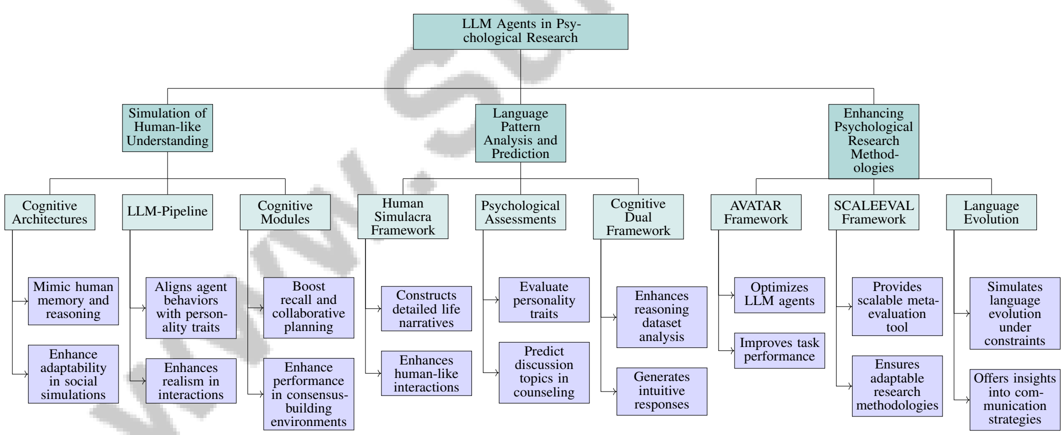


Figure 2: This figure illustrates the hierarchical structure of LLM Agents in Psychological Research, categorizing their roles into simulating human-like understanding, analyzing language patterns, and enhancing research methodologies. It highlights the frameworks and applications that contribute to these areas, demonstrating their transformative impact on psychological studies.

3 LLM Agents in Psychological Research

3.1 Simulation of Human-like Understanding

LLM agents are pivotal in simulating human-like understanding in psychological research by employing cognitive architectures that mimic human memory and reasoning. These agents excel in tasks requiring nuanced comprehension and adaptability, as evidenced by multi-agent frameworks that model language strategy evolution in controlled environments, showcasing LLM adaptability in complex social simulations [12]. The LLM-Based Personality Simulation Pipeline (LLM-Pipeline) enhances this understanding by aligning agent behaviors with established personality traits, enriching the realism of interactions across clinical, social, and developmental psychology domains [5, 2, 18].

Integrating cognitive modules into LLM-augmented autonomous agents boosts their ability to recall past interactions and engage in collaborative planning, addressing temporal cognition limitations through dynamic memory recall. These agents, equipped with human-like memory architectures, enhance performance in environments where consensus building is essential [22, 45, 37, 46, 47]. In therapeutic settings, LLMs automate tasks, enhance treatment fidelity, and provide real-time feedback, simulating human-like understanding through personalized interactions. These "Thinking Assistants" foster cognitive engagement, guiding users through complex decision-making processes with tailored advice and relevant examples [48, 11].

The LLM-ODS framework illustrates LLMs' capability to engage in dynamic dialogues and update beliefs based on interactions, aiding in the modeling of intricate opinion dynamics. This capacity is crucial for understanding belief influences in psychology, marketing, and public health [5, 49, 2, 16]. These advancements underscore LLMs' potential to emulate complex cognitive and social processes, offering innovative approaches to studying and influencing behavior in psychological research. However, ethical considerations and misinformation risks necessitate responsible usage and an understanding of limitations in sensitive contexts [5, 49, 50, 18].

3.2 Language Pattern Analysis and Prediction

LLM agents are instrumental in analyzing language patterns and predicting psychological behaviors, utilizing computational frameworks that reflect human cognitive processes. The Human Simulacra Framework (HSF) exemplifies LLMs' ability to construct detailed life narratives for virtual characters, enhancing the simulation of human-like interactions [51]. By integrating LLM capabilities with agent-based modeling, these agents emulate user behaviors within social networks, providing a robust platform for exploring social dynamics [52].

In psychological assessments, LLMs evaluate personality traits and behaviors through structured tests like the Big Five Inventory, allowing quantitative analysis of language use in narratives [53]. LLMs also predict discussion topics in counseling by analyzing dialogue turns, demonstrating efficacy in language pattern analysis [3]. The cognitive dual framework enhances reasoning dataset analysis, enabling LLMs to generate intuitive responses akin to human System 1 processes, crucial for simulating complex cognitive behaviors [54].

Challenges persist, as benchmarks assess value expression consistency across contexts, highlighting the complexity of modeling human-like responses [9]. Frameworks like HAD analyze LLM errors in financial sentiment analysis, contributing to a nuanced understanding of language patterns [29].

As illustrated in Figure 3, the hierarchical categorization of language pattern analysis and prediction in large language models (LLMs) outlines the main applications, predictive techniques, and challenges associated with LLMs. This figure highlights their role in simulating human interactions, predicting behaviors, and addressing ethical concerns. LLM agents demonstrate substantial potential in analyzing language patterns and predicting psychological behaviors, with methodologies facilitating exploration of human cognition and social interactions. Their applications across cognitive, behavioral, clinical, and social psychology underscore their transformative role in research methodologies, while raising ethical considerations regarding data privacy and responsible use [22, 49, 55, 18].

3.3 Enhancing Psychological Research Methodologies

LLM agents are at the forefront of enhancing psychological research methodologies by integrating computational frameworks that streamline hypothesis generation, data processing, and experimental design. The AVATAR framework optimizes LLM agents through a comparator module that generates holistic prompts, improving task performance and research methodologies [56]. The SCALEEVAL framework provides a scalable, agent-debate-assisted meta-evaluation tool for assessing LLMs, ensuring research methodologies remain adaptable and comprehensive [57].

LLMs contribute to the evolution of language and communication strategies in research. Cai et al.'s framework employs agents to simulate language evolution under constraints, offering insights into adaptive communication strategies [58]. LLM agents also enhance user interfaces and decision-

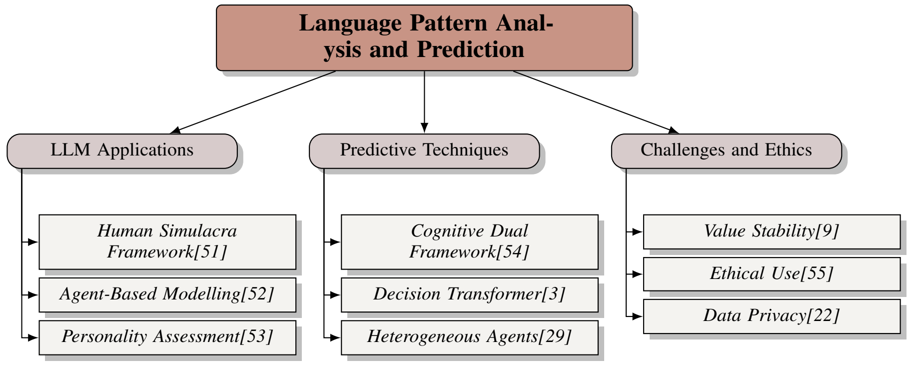


Figure 3: This figure illustrates the hierarchical categorization of language pattern analysis and prediction in large language models (LLMs). It outlines the main applications, predictive techniques, and challenges associated with LLMs, highlighting their role in simulating human interactions, predicting behaviors, and addressing ethical concerns.

making processes, with applications in healthcare and education [10]. Evaluations involving multiple LLMs reveal insights into cognitive processes, informing methodologies probing deeper into cognitive and social phenomena [8].

Integrating LLM agents into psychological research methodologies revolutionizes the field by enhancing the depth and breadth of studies across cognitive, clinical, educational, and social psychology. By leveraging LLMs, researchers streamline the research process and uncover new insights into human cognition and behavior. However, this transformative potential comes with ethical implications and technical challenges, necessitating a responsible approach to deploying LLMs in sensitive contexts [50, 49, 18]. Through advanced frameworks and evaluation techniques, LLMs continue to redefine psychological research, providing innovative tools and approaches to understanding complex human behaviors and cognitive processes.

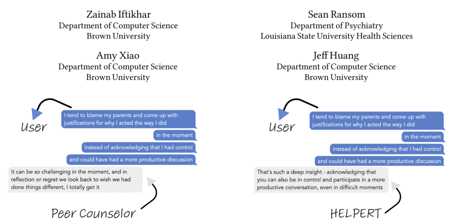
4 Applications of LLM Agents in Psychology

The integration of Large Language Model (LLM) agents into psychology has significantly advanced our understanding of human behavior and mental processes. This section explores their applications, particularly in clinical and therapeutic contexts, while emphasizing the necessity of ethical considerations and efficacy in their deployment.

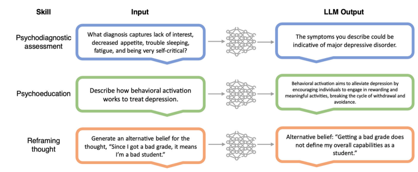
4.1 Clinical and Therapeutic Applications

LLM agents have demonstrated substantial promise in clinical and therapeutic settings by enhancing psychological assessments and interventions. They effectively conduct assessments of suicidal risk, balancing data privacy with accessibility for clinicians [59]. The LD-Agent framework exemplifies adaptability, enabling coherent and personalized interactions crucial for patient engagement [60]. In therapeutic contexts, LLMs act as personalized AI negotiation assistants, facilitating negotiation training and therapeutic interventions [61]. These agents simulate negotiation scenarios, equipping users with practical skills and enhancing therapeutic outcomes.

Comparative analyses between LLM-based models, such as HELPERT, and human counselors highlight that while LLMs closely adhere to Cognitive Behavioral Therapy (CBT) techniques, human counselors provide essential empathy and emotional connection, underscoring the need for integrating human-like empathy into LLMs [62]. Furthermore, LLM agents improve diagnostic accuracy through adaptive reasoning, supporting clinicians in making informed decisions that enhance patient care [63]. Their capabilities in dialogue management and knowledge extraction from scientific literature position LLMs as transformative tools in psychological practice. Evaluation frameworks like Artificial-intelligence Structured Clinical Examinations (AI-SCI) will further ensure their effectiveness in clinical tasks, ultimately improving patient outcomes [49, 64, 20].



(a) Peer Counseling and Helpert: Understanding the Role of Peer Counselors in Mental Health[62]



(b) AI-Assisted Cognitive Behavioral Therapy: A Comprehensive Approach[24]

Figure 4: Examples of Clinical and Therapeutic Applications

As illustrated in Figure 4, the integration of LLM agents in clinical and therapeutic settings reveals transformative potential. The "Peer Counseling and Helpert" example demonstrates the collaborative dynamics between users, peer counselors, and the AI agent Helpert, highlighting the emotional complexities individuals face. The "AI-Assisted Cognitive Behavioral Therapy" framework underscores how AI can enhance traditional CBT through structured assessments and educational resources, paving new avenues for mental health support [62, 24].

4.2 Social Simulation and Interaction

LLM agents play a pivotal role in social simulations within psychological studies, effectively modeling complex social dynamics and interactions. By simulating social scenarios, LLMs provide valuable insights into human behavior and social processes [52]. The Human Simulacra Framework (HSF) exemplifies this capability by constructing detailed life narratives for virtual characters, facilitating the exploration of social influence and group behavior [51].

As illustrated in Figure 5, the hierarchical structure of key frameworks and considerations in the domain of social simulation and interaction using LLM agents is depicted. This figure categorizes the simulation frameworks, cooperative strategies, and ethical considerations that are pivotal in advancing research methodologies across various branches of psychology.

LLMs are utilized to simulate opinion dynamics and belief updates, reflecting the complexities of human social interactions. The LLM-ODS framework enables the modeling of belief adjustments through dialogue, contributing to our understanding of social cognition [12]. Additionally, LLM agents develop cooperative strategies organically, mirroring human social dynamics and decision-making processes [33].

The application of LLM agents in social simulation provides a transformative approach to studying human behavior. By integrating advanced computational frameworks with psychological principles, LLMs advance research methodologies across various branches of psychology, including cognitive, behavioral, clinical, educational, and social psychology. However, ethical considerations and limitations must be navigated to ensure responsible use in sensitive contexts [22, 28, 50, 55, 18].

4.3 Personality and Emotion Simulation

LLM agents have emerged as powerful tools for simulating personality traits and emotional responses, offering innovative methodologies for exploring human psychology. By emulating complex psychological constructs, LLMs provide insights into personality dynamics and emotional processes, facilitating the creation of virtual agents that exhibit distinct personality traits [61].

Their application in sentiment analysis and affective computing enables LLMs to predict and generate emotional responses that reflect human affective states, enhancing the realism of simulated interactions [3]. This capability is particularly beneficial in therapeutic settings, where responding to emotional cues is crucial for effective communication.

Frameworks like LLM-ODS allow for modeling belief updates and emotional shifts in response to social interactions, highlighting the interplay between personality traits and emotional responses [12]. Additionally, LLMs enhance educational and developmental psychology by adapting to learners' emotional states and personality traits, creating engaging and personalized educational environments.

AI-generated, for reference only.

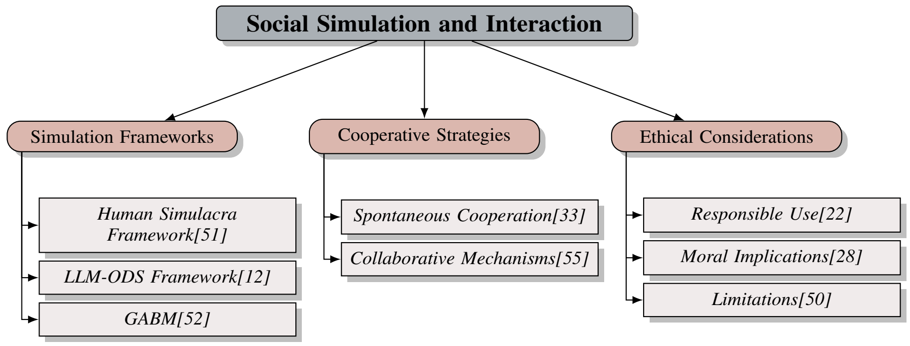


Figure 5: This figure illustrates the hierarchical structure of key frameworks and considerations in the domain of social simulation and interaction using LLM agents. It categorizes the simulation frameworks, cooperative strategies, and ethical considerations that are pivotal in advancing research methodologies across various branches of psychology.

Roleplay scenarios facilitated by LLMs promote effective learning by aligning content with individual profiles, fostering empathy and responsiveness in educational experiences [65, 66].

The role of LLM agents in simulating personality traits and emotional responses offers transformative potential for psychological research. They enhance understanding across cognitive, clinical, educational, and social psychology, while also presenting ethical and technical challenges, such as data privacy concerns. This necessitates careful consideration of their implications in psychological studies [50, 28, 18].

4.4 Educational and Developmental Psychology

LLM agents are reshaping educational and developmental psychology by providing innovative approaches that enhance learning experiences. Their advanced natural language processing capabilities facilitate personalized educational interventions tailored to individual learning styles and developmental needs. Adaptive learning systems powered by LLMs optimize educational outcomes by adjusting content delivery based on learners' progress [4].

In developmental psychology, LLM agents simulate age-appropriate interactions, offering insights into cognitive and emotional development. This integration creates immersive virtual environments that foster interactive learning experiences, supporting both cognitive growth and social skills through tailored feedback and engagement in educational activities [4, 18].

LLMs also support educators in designing effective instructional strategies and providing real-time feedback. Their application in generating performance feedback and suggesting tailored learning activities underscores their potential to transform traditional educational practices into more responsive models [4].

Furthermore, LLMs enhance language acquisition and literacy development by facilitating personalized feedback mechanisms in Intelligent Tutoring Systems (ITS). Innovative frameworks like the progressive ontology prompting (POP) algorithm automate knowledge discovery, creating comprehensive resources for language learning and intervention strategies [4, 49, 2, 17]. By simulating conversational partners, LLMs provide learners opportunities to practice language skills in realistic contexts, benefiting language learners and individuals with developmental language disorders.

The integration of LLM agents in educational and developmental psychology holds significant promise for enhancing our understanding of learning processes. By leveraging insights from educational technologies, LLMs can provide tailored feedback that addresses individual learner needs, advancing pedagogical practices. However, their deployment requires careful consideration of ethical implications and a solid theoretical foundation to ensure responsible and impactful use [4, 18].

4.5 Psychological Experimentation and Evaluation

LLM agents have introduced novel methodologies in designing and evaluating psychological experiments, enhancing assessments of cognitive and social processes. For example, models like GPT-4

AI-generated, for reference only.

have demonstrated proficiency in identifying indirect requests and false beliefs, although they struggle with nuanced social cues, indicating areas for further development in cognitive modeling [67].

Frameworks such as ConSiDERS enhance the adaptability of LLMs in various contexts, improving the evaluation of LLM performance in psychological experimentation [68]. LLMs have also shown improvements in accuracy in complex problem-solving tasks, as evidenced by their application to the GSM8K dataset, which highlights their potential to refine experimental methodologies [69].

The integration of LLMs in psychological experimentation not only facilitates innovative evaluation metrics that capture human cognition and behavior nuances but also enhances research methodologies across various psychological domains. This advancement allows for the simulation of human-like text generation and refinement of experimental design, data analysis, and hypothesis generation. However, ethical implications and limitations must be addressed to ensure responsible application in sensitive research areas [28, 18]. By leveraging advanced computational techniques, LLM agents facilitate the exploration of cognitive phenomena, providing insights into decision-making and social interactions.

As illustrated in Figure 6, the incorporation of LLM agents into psychological experimentation highlights the key aspects of this transformative advancement, showcasing methodological innovations, diverse applications, and significant ethical and technical challenges. Their application across cognitive, clinical, educational, and social psychology enables more effective investigation of intricate cognitive and social processes. This integration enhances psychological research's depth and breadth while raising important ethical considerations regarding data privacy and model limitations, underscoring the need for responsible and informed use in sensitive contexts [4, 28, 50, 18].

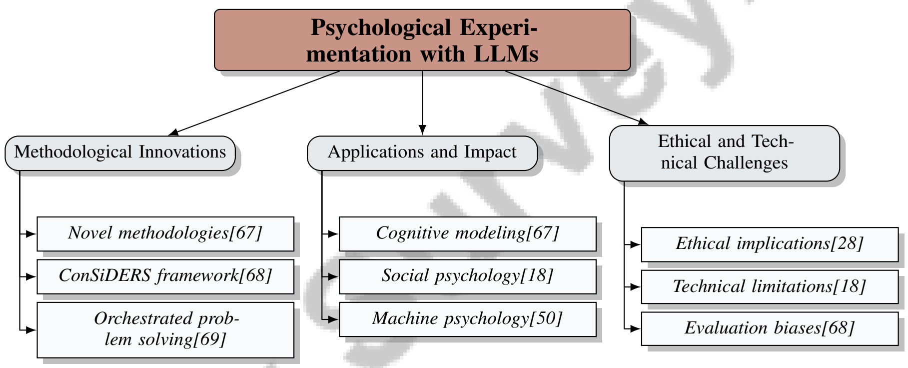


Figure 6: This figure illustrates the key aspects of integrating large language models (LLMs) into psychological experimentation, highlighting methodological innovations, diverse applications, and significant ethical and technical challenges.

5 Challenges and Limitations

The integration of Large Language Model (LLM) agents in psychology is challenged by technical and computational hurdles that impact their effectiveness. A significant concern is the computational demand required for training and deploying LLMs, which limits scalability in complex environments [29]. High memory needs exacerbate this issue, particularly in cloud-based settings where infrastructure costs are prohibitive [17].

Training LLMs for psychological tasks is complicated by the inefficiencies of traditional reinforcement learning methods, which struggle with stability and generalizability, limiting their application in dynamic scenarios [23]. Additionally, integrating LLM agents with autonomous systems poses architectural challenges, affecting performance and adaptability in real-world applications [70].

LLMs' generalization capabilities are constrained by their reliance on pre-trained knowledge, hindering their ability to learn from real-world experiences and adapt to new contexts, crucial for accurately modeling human behavior [25]. The absence of effective prompting strategies for guiding LLMs through multi-stage processes further limits their utility [56].

Evaluating LLMs is challenging due to reliance on costly human annotations and limited benchmarks, which increase resource demands and hinder accurate performance assessment across diverse tasks

[57]. The unpredictability of LLM outputs complicates deployment in real-world scenarios, where empirical evaluations are often lacking, and risks such as generating harmful content, privacy breaches, adversarial attacks, and biases are significant [71].

Addressing these challenges is essential for enhancing LLM applications in psychology, encompassing data privacy, ethical considerations, and robust evaluation methodologies critical for responsible use across research domains [50, 49, 2, 18]. Overcoming these limitations can unlock LLMs' full potential, offering innovative solutions and insights into complex psychological phenomena.

5.1 Technical and Computational Challenges

Deploying LLM agents in psychology faces significant technical and computational challenges. The computational intensity of LLMs demands substantial resources for training and deployment, limiting scalability in complex environments [29]. High memory requirements, particularly in cloud-based settings, further complicate this issue due to prohibitive infrastructure costs [17].

Training LLMs for psychological tasks is hindered by traditional reinforcement learning methods' inefficiencies, limiting their application in dynamic scenarios typical of psychological research [23]. Architectural design challenges in integrating LLM agents with autonomous systems impact performance and adaptability in real-world applications [70].

LLMs' generalization capabilities are constrained by their reliance on pre-trained knowledge, which limits learning from real-world experiences and adapting to new contexts, essential for accurately modeling human behavior [25]. The lack of effective prompting strategies for guiding LLMs through multi-stage processes further limits their practical utility [56].

Evaluating LLMs presents challenges due to reliance on costly human annotations and limited benchmarks, which increase resource demands and constrain accurate performance assessment across diverse tasks [57]. The unpredictability of LLM outputs complicates deployment in real-world scenarios, where empirical evaluations are often lacking, and risks such as generating harmful content, privacy breaches, adversarial attacks, and biases are significant [71].

Addressing these technical and computational challenges is essential for enhancing LLM applications in psychology, encompassing data privacy, ethical considerations, and robust evaluation methodologies critical for responsible use across research domains [50, 49, 2, 18]. Overcoming these limitations can unlock LLMs' full potential, offering innovative solutions and insights into complex psychological phenomena.

5.2 Cognitive and Behavioral Simulation Limitations

LLM agents face limitations in accurately simulating human cognition and behavior, crucial for effective deployment in psychological research. Ambiguity in action outputs often results in vague descriptions, hindering the simulation of cognitive processes [72]. This ambiguity impairs LLMs' ability to provide clear, consistent responses necessary for replicating complex cognitive behaviors.

Simulated contexts in benchmarks fail to capture real-world interactions' intricacies, impacting LLMs' ability to model human cognition and behavior accurately [9]. Challenges also arise in modeling human traits, ensuring explainability, and integrating these technologies into practical applications [12].

LLMs struggle with advanced statistical tasks and multi-modal data, critical for understanding human cognition comprehensively. Integrating domain knowledge into LLM frameworks requires significant improvement, as current models often fall short in leveraging such knowledge effectively [14]. This limitation restricts LLMs' potential to simulate complex cognitive processes involving diverse data types and contextual information.

Variability in LLM responses, influenced by prompt novelty and specificity, affects generalizability. This variability underscores LLMs' struggle to maintain consistency across instruction types, particularly when instruction-following is sensitive to prompt phrasing. The absence of effective prompting techniques significantly hinders precise simulation of human cognition and behavior [50, 42].

Enhancing LLM application in psychological research requires addressing cognitive and behavioral simulation limitations, essential for emulating human cognition and behavior across psychological domains. This involves improving LLM-generated simulations' accuracy and ensuring ethical standards and methodological rigor in deployment, particularly in sensitive psychological inquiry areas. Overcoming these limitations allows researchers to leverage LLMs for innovative tools in literature review, hypothesis generation, and experimental design, advancing understanding of cognitive processes and decision-making in psychology [50, 73, 18].

5.3 Ethical Concerns and Biases

The deployment of LLM agents in psychology raises significant ethical concerns and biases that require careful consideration and mitigation. A primary issue is the inherent biases in LLMs' training datasets, which can result in skewed outputs, limiting generalizability in psychological contexts and potentially leading to unreliable simulations and misrepresentations of human interactions [13]. Selection biases from specific data sources and platforms may produce non-representative outcomes [10].

Data privacy is a critical concern, particularly in sensitive psychological assessments. Emphasizing LLMs that operate locally addresses privacy concerns in clinical settings [57]. Additionally, LLMs' potential to produce hallucinations and inaccuracies in normative reasoning can undermine trust and reliability, especially in high-stakes scenarios requiring human oversight [7].

Bias amplification through iterative processes presents another challenge, as minor biases in individual outputs can accumulate, leading to significant distortions in generated content [28]. This underscores the need for mechanisms to detect and mitigate bias propagation in LLM-generated outputs. Moreover, ethical implications arise from LLMs' limitations in handling unknown domains and generalizing across plan lengths. The inability to explain generated plans' logic can lead to opaque decision-making processes, undermining trust and accountability [72].

Overestimation of LLMs' capabilities poses ethical risks, as these models lack true mathematical reasoning abilities, potentially misleading users about their problem-solving strategies [57]. Misinterpreting LLMs as autonomous agents could have ethical and practical consequences, necessitating a clear understanding of their limitations and capabilities [10].

Addressing these ethical concerns and biases requires a multifaceted approach, including developing comprehensive benchmarks, rigorous empirical investigations, and robust ethical frameworks [7]. Tackling these challenges ensures LLM agents are deployed responsibly and effectively in psychological contexts, maximizing potential benefits while minimizing risks.

5.4 Evaluation and Benchmarking Challenges

Benchmark	Size	Domain	Task Format	Metric
LSTI[74]	1,200	Conversational AI	Language Style Imitation	Human Evaluation, LLM Evaluation
PsychoLex[75]	10,056	Psychology	Multiple-choice Question And Answer	Accuracy
CogBench[76]	35	Cognitive Psychology	Behavioral Assessment	Meta-cognition, Performance
StrategicReasoningBenchmark[77]	40	Behavioral Economics	Game Simulation	Human Consistency Rate
LLM-Bench[78]	23	Natural Language Processing	Multiple Choice Questions	Accuracy, F1-score
TP[79]	1,225	Travel Planning	Plan Generation	Final Pass Rate, Delivery Rate
SESI[80]	1,000	Social Intelligence	Multiple Choice	Exact Match, F1 Score
REHEAT[81]	202	Cognitive Psychology	Multiple-choice And Ranking Questions	Accuracy, F1-score

Table 1: This table presents a comprehensive overview of various benchmarks utilized for evaluating large language models (LLMs) across different domains, including Conversational AI, Psychology, and Behavioral Economics. It highlights the size, task format, and evaluation metrics employed for each benchmark, providing insights into the diverse methodologies and challenges associated with LLM assessments.

Evaluating and benchmarking LLM agents in psychological applications present challenges that complicate assessing their effectiveness and reliability. A significant issue is the lack of comprehensive experimental detail and rigorous evaluation methods in current studies, raising concerns about findings' validity and generalizability [15]. This methodological rigor deficiency often leads to inconsistencies in LLM agent evaluation, making it difficult to establish standardized benchmarks reflecting performance across diverse tasks. Table 1 summarizes the key benchmarks used in the evaluation of LLM agents, illustrating the diversity in domain applications, task formats, and evaluation metrics, which are critical for understanding the challenges in benchmarking these models.

Ethical considerations during LLM evaluation are often neglected, leading to biased or incomplete assessments. This oversight is concerning given the lack of consensus on evaluation methodologies, as studies reveal significant discrepancies between LLM outputs and human responses, underscoring the necessity for rigorous, ethically sound evaluation practices to ensure responsible technology use [26, 2, 15, 28]. The absence of clear ethical guidelines for LLM deployment and evaluation in psychological contexts exacerbates this challenge, necessitating robust ethical frameworks for responsible and fair evaluation practices.

The dynamic nature of LLMs presents a significant challenge, as continuous updates and improvements can lead to performance and capability variations. This evolution complicates evaluation, particularly in applications like chatbots, where the lack of consensus on effective assessment methods—especially concerning the correlation between automated, human, and LLM-based evaluations—highlights the necessity for rigorous, systematic evaluation mechanisms [2, 15]. This evolution complicates benchmarking, as LLM agents' performance may vary significantly over time, requiring constant updates to evaluation criteria and benchmarks to capture capabilities accurately. Additionally, reliance on costly human annotations and limited benchmarks increases resource demands, constraining thorough evaluations across diverse applications.

The intricate nature of psychological phenomena presents significant challenges for LLM evaluation and benchmarking, particularly as these models increasingly adopt human-like roles and exhibit a wide range of psychological attributes. This complexity necessitates robust psychometric frameworks to accurately assess psychological dimensions such as personality, emotion, and motivation, while addressing discrepancies between self-reported traits and actual behaviors. As LLMs integrate into various psychological research applications, careful consideration of methodological practices and ethical implications is essential to ensure reliable evaluations and meaningful insights into cognitive capabilities [18, 42, 82, 1]. LLM agents must be assessed on technical performance and ability to simulate nuanced cognitive and social processes, requiring sophisticated evaluation metrics capturing human cognition and behavior intricacies, ensuring LLM assessments reflect true cognitive competencies rather than superficial language processing abilities.

Effectively addressing evaluation and benchmarking challenges in LLM applications is essential for enhancing integration into psychological research, ensuring reliable performance assessments, ethical use, and alignment with human cognition and behavior [26, 1, 2, 18]. Developing comprehensive evaluation frameworks and ethical guidelines enhances LLM assessments' reliability and validity, paving the way for more effective, responsible use in psychological research and applications.

6 Future Directions

The future of Large Language Model (LLM) agents in psychological research hinges on interdisciplinary integration and collaboration, crucial for advancing their development and application. By synthesizing insights from diverse fields, researchers can enhance LLM efficacy and reliability, fostering innovative methodologies to address complex psychological challenges. This section highlights the importance of collaborative efforts in shaping the future of LLM research and its implications for psychology.

6.1 Interdisciplinary Integration and Collaboration

Progress in LLM agents for psychological research requires a robust interdisciplinary approach, integrating insights from Multi-Agent Systems (MAS), Natural Language Processing (NLP), and LLM research. Establishing unified benchmarks for evaluating LLM agents across various psychological tasks is essential for enhancing their strategic reasoning capabilities [83]. Collaborative efforts are vital for pooling expertise to realize the full potential of normative LLM agents.

Understanding trust behaviors in LLM agents involves incorporating psychological theories and computational models [84], leading to more reliable and trustworthy agents. Enhancing explainability is critical for integrating LLM-based agents into computational experiments and real-world applications

[85]. Developing transparent frameworks will increase interpretability and utility in psychological research.

Future research should prioritize multilingual capabilities and real-time sentiment analysis within Affective Computing [86], enabling LLM agents to better understand diverse psychological phenomena. Addressing hallucination issues and developing frameworks for agent orchestration and collective intelligence are key to advancing LLMs in psychology [21]. These efforts will improve reliability and coherence in LLM outputs, contributing positively to psychological research.

Establishing ethical frameworks for LLM deployment is crucial for examining their impact on public discourse and potential to influence power structures [87]. Such frameworks will guide the responsible use of LLM agents, aligning their deployment with ethical standards and societal values. Through interdisciplinary collaboration, LLM agents can advance psychological research, offering insights into complex human behaviors and cognitive processes.

6.2 Expanding Multimodal and Interdisciplinary Approaches

Expanding multimodal and interdisciplinary approaches is essential for the future of LLM agent research, integrating diverse data types and methodologies to enhance model capabilities. Multimodal approaches, combining text, audio, and visual inputs, facilitate a richer understanding of psychological phenomena and simulate human-like cognitive and social processes [44].

Collaboration among cognitive science, computer science, linguistics, and psychology experts is crucial for developing robust LLM agents. Leveraging insights from these fields addresses current LLM limitations and explores new psychological research applications [10]. Integrating LLMs with AI technologies like computer vision and speech recognition enhances their ability to process multimodal data [44].

Interdisciplinary collaboration can also yield novel evaluation metrics and benchmarks that capture LLM agents' cognitive and social capabilities. By incorporating psychology and cognitive science insights, researchers can design assessments reflecting LLM agents' potential in simulating human-like behaviors and interactions [23].

6.3 Optimizing Training and Adaptability

Optimizing training methodologies and adaptability is crucial for LLM agents' effective application in psychological contexts. Future research should prioritize techniques that mitigate inherent biases, improving alignment with human perspectives and enhancing simulation believability [13]. This aligns with refining LLM integration into computational experiments, enhancing explainability and addressing application challenges [12].

Exploring cognitive ergonomics in LLM applications offers opportunities for developing models focused on emotion recognition and adaptive learning [10]. Optimizing memory retrieval strategies and advancing neural algorithms for memory consolidation are essential for improving adaptability [31].

Enhancing the HAD framework and exploring new error types are vital for assessing human-level performance on FSA datasets, informing training process improvements [29]. Advancing LLM capabilities in multi-modal reasoning and fostering community-driven development of intelligent statistical analysis tools are critical research areas [14]. These efforts will refine LLM training methodologies, ensuring robustness across psychological applications.

Future research should also explore scenarios and refine benchmarks to capture moral and legal reasoning intricacies, enhancing LLM cognitive processing capabilities [28]. Addressing these directions can significantly improve LLM agents' training and adaptability, enhancing their effectiveness and reliability in psychological research and applications.

7 Conclusion

The exploration of Large Language Model (LLM) agents underscores their substantial impact on psychology, where they are reshaping research methodologies and clinical practices. By enhancing decision-making processes and clinical workflows, LLMs are driving broader healthcare integration.

AI-generated, for reference only.

Recent advancements show significant improvements in LLM reasoning, overcoming previous limitations and boosting cognitive processing capabilities. Models like GPT-4 exhibit notable alignment with human trust behaviors, indicating their potential in simulating human interactions effectively. Moreover, LLMs have shown the ability to represent diverse personalities under specific conditions, with prompt adjustments enhancing their applicability in social science contexts. However, challenges remain, such as achieving consistent personality mimicry across various conditions and addressing technical, computational, and ethical concerns. The development of robust evaluation frameworks is essential for the responsible integration of LLM agents into psychological research. Continued research and development in this interdisciplinary field are crucial. By harnessing LLM capabilities and overcoming existing challenges, researchers can unlock new insights into complex human behaviors and cognitive processes, advancing both artificial intelligence and psychological science.

References

[1] Yuan Li, Yue Huang, Hongyi Wang, Xiangliang Zhang, James Zou, and Lichao Sun. Quantifying ai psychology: A psychometrics benchmark for large language models, 2024.

[2] Bhashithe Abeysinghe and Ruhan Circi. The challenges of evaluating llm applications: An analysis of automated, human, and llm-based approaches, 2024.

[3] Aylin Gunal, Baihan Lin, and Djallel Bouneffouf. Conversational topic recommendation in counseling and psychotherapy with decision transformer and large language models, 2024.

[4] John Stamper, Ruiwei Xiao, and Xinying Hou. Enhancing llm-based feedback: Insights from intelligent tutoring systems and the learning sciences, 2024.

[5] Elijah Berberette, Jack Hutchins, and Amir Sadovnik. Redefining "hallucination" in llms: Towards a psychology-informed framework for mitigating misinformation, 2024.

[6] Simon Goldstein and Benjamin A. Levinstein. Does chatgpt have a mind?, 2024.

[7] Chen Gao, Xiaochong Lan, Nian Li, Yuan Yuan, Jingtao Ding, Zhilun Zhou, Fengli Xu, and Yong Li. Large language models empowered agent-based modeling and simulation: A survey and perspectives. *Humanities and Social Sciences Communications*, 11(1):1–24, 2024.

[8] Thilo Hagendorff. Deception abilities emerged in large language models, 2024.

[9] Grgur Kovač, Rémy Portelas, Masataka Sawayama, Peter Ford Dominey, and Pierre-Yves Oudeyer. Stick to your role! stability of personal values expressed in large language models, 2024.

[10] Azmine Toushik Wasi and Mst Rafia Islam. Cogergllm: Exploring large language model systems design perspective using cognitive ergonomics, 2024.

[11] Soya Park, Hari Subramonyam, and Chinmay Kulkarni. Thinking assistants: Llm-based conversational assistants that help users think by asking rather than answering, 2024.

[12] Qun Ma, Xiao Xue, Deyu Zhou, Xiangning Yu, Donghua Liu, Xuwen Zhang, Zihan Zhao, Yifan Shen, Peilin Ji, Juanjuan Li, Gang Wang, and Wanpeng Ma. Computational experiments meet large language model based agents: A survey and perspective, 2024.

[13] Amir Taubenfeld, Yaniv Dover, Roi Reichart, and Ariel Goldstein. Systematic biases in llm simulations of debates, 2024.

[14] Maojun Sun, Ruijian Han, Binyan Jiang, Houduo Qi, Defeng Sun, Yancheng Yuan, and Jian Huang. A survey on large language model-based agents for statistics and data science, 2024.

[15] Adrian de Wynter. Awes, laws, and flaws from today's llm research, 2024.

[16] Alexander Rogiers, Sander Noels, Maarten Buyl, and Tijl De Bie. Persuasion with large language models: a survey. *arXiv preprint arXiv:2411.06837*, 2024.

[17] Baolin Li, Yankai Jiang, Vijay Gadepally, and Devesh Tiwari. Llm inference serving: Survey of recent advances and opportunities, 2024.

[18] Luoma Ke, Song Tong, Peng Cheng, and Kaiping Peng. Exploring the frontiers of llms in psychological applications: A comprehensive review, 2024.

[19] Jen-tse Huang, Wenxiang Jiao, Man Ho Lam, Eric John Li, Wenxuan Wang, and Michael Lyu. On the reliability of psychological scales on large language models. In *Proceedings of The 2024 Conference on Empirical Methods in Natural Language Processing*, pages 6152–6173, 2024.

[20] Chris Sypherd and Vaishak Belle. Practical considerations for agentic llm systems, 2024.

[21] Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V Chawla, Olaf Wiest, and Xiangliang Zhang. Large language model based multi-agents: A survey of progress and challenges. *arXiv preprint arXiv:2402.01680*, 2024.

[22] Jintian Zhang, Xin Xu, Ningyu Zhang, Ruibo Liu, Bryan Hooi, and Shumin Deng. Exploring collaboration mechanisms for llm agents: A social psychology view. *arXiv preprint arXiv:2310.02124*, 2023.

[23] Yuheng Cheng, Ceyao Zhang, Zhengwen Zhang, Xiangrui Meng, Sirui Hong, Wenhao Li, Zihao Wang, Zekai Wang, Feng Yin, Junhua Zhao, et al. Exploring large language model based intelligent agents: Definitions, methods, and prospects. *arXiv preprint arXiv:2401.03428*, 2024.

[24] Elizabeth C Stade, Shannon Wiltsey Stirman, Lyle H Ungar, Cody L Boland, H Andrew Schwartz, David B Yaden, João Sedoc, Robert J DeRubeis, Robb Willer, and Johannes C Eichstaedt. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. *NPJ Mental Health Research*, 3(1):12, 2024.

[25] Sihao Hu, Tiansheng Huang, Fatih Ilhan, Selim Tekin, Gaowen Liu, Ramana Kompella, and Ling Liu. A survey on large language model-based game agents. *arXiv preprint arXiv:2404.02039*, 2024.

[26] John Mavi, Nathan Summers, and Sergio Coronado. Assessing the performance of human-capable llms – are llms coming for your job?, 2024.

[27] Jinyang Wu, Feihu Che, Xinxin Zheng, Shuai Zhang, Ruihan Jin, Shuai Nie, Pengpeng Shao, and Jianhua Tao. Can large language models understand uncommon meanings of common words?, 2024.

[28] Guilherme F. C. F. Almeida, José Luiz Nunes, Neele Engelmann, Alex Wiegmann, and Marcelo de Araújo. Exploring the psychology of llms' moral and legal reasoning, 2024.

[29] Frank Xing. Designing heterogeneous llm agents for financial sentiment analysis, 2024.

[30] Yun-Shiuan Chuang, Agam Goyal, Nikunj Harlalka, Siddharth Suresh, Robert Hawkins, Sijia Yang, Dhavan Shah, Junjie Hu, and Timothy T. Rogers. Simulating opinion dynamics with networks of llm-based agents, 2024.

[31] Jing Guo, Nan Li, Jianchuan Qi, Hang Yang, Ruiqiao Li, Yuzhen Feng, Si Zhang, and Ming Xu. Empowering working memory for large language model agents. *arXiv preprint arXiv:2312.17259*, 2023.

[32] Matthew Stephenson, Matthew Sidji, and Benoît Ronval. Codenames as a benchmark for large language models, 2024.

[33] Zengqing Wu, Run Peng, Shuyuan Zheng, Qianying Liu, Xu Han, Brian Inhyuk Kwon, Makoto Onizuka, Shaojie Tang, and Chuan Xiao. Shall we team up: Exploring spontaneous cooperation of competing llm agents, 2024.

[34] Jinghua Piao, Zhihong Lu, Chen Gao, Fengli Xu, Fernando P. Santos, Yong Li, and James Evans. Emergence of human-like polarization among large language model agents, 2025.

[35] John J Horton. Large language models as simulated economic agents: What can we learn from homo silicus? Technical report, National Bureau of Economic Research, 2023.

[36] Chang Ma, Junlei Zhang, Zhihao Zhu, Cheng Yang, Yujiu Yang, Yaohui Jin, Zhenzhong Lan, Lingpeng Kong, and Junxian He. Agentboard: An analytical evaluation board of multi-turn llm agents, 2024.

[37] Yuki Hou, Haruki Tamoto, and Homei Miyashita. "my agent understands me better": Integrating dynamic human-like memory recall and consolidation in llm-based agents, 2024.

[38] Nicolas Yax, Hernan Anlló, and Stefano Palminteri. Studying and improving reasoning in humans and machines, 2023.

[39] Thilo Hagendorff, Sarah Fabi, and Michal Kosinski. Thinking fast and slow in large language models, 2023.

[40] Yan Leng and Yuan Yuan. Do llm agents exhibit social behavior?, 2024.

[41] Boming Xia, Qinghua Lu, Liming Zhu, Zhenchang Xing, Dehai Zhao, and Hao Zhang. An evaluation-driven approach to designing llm agents: Process and architecture, 2024.

[42] Anna A. Ivanova. Toward best research practices in ai psychology, 2024.

[43] Simon Martin Breum, Daniel Vædele Egdal, Victor Gram Mortensen, Anders Giovanni Møller, and Luca Maria Aiello. The persuasive power of large language models. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, pages 152–163, 2024.

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

[45] Xudong Guo, Kaixuan Huang, Jiale Liu, Wenhui Fan, Natalia Vélez, Qingyun Wu, Huazheng Wang, Thomas L Griffiths, and Mengdi Wang. Embodied llm agents learn to cooperate in organized teams. *arXiv preprint arXiv:2403.12482*, 2024.

[46] Yashar Talebirad and Amirhossein Nadiri. Multi-agent collaboration: Harnessing the power of intelligent llm agents, 2023.

[47] Xudong Guo, Kaixuan Huang, Jiale Liu, Wenhui Fan, Natalia Vélez, Qingyun Wu, Huazheng Wang, Thomas L. Griffiths, and Mengdi Wang. Embodied llm agents learn to cooperate in organized teams, 2024.

[48] Chengwei Qin, Wenhan Xia, Tan Wang, Fangkai Jiao, Yuchen Hu, Bosheng Ding, Ruirui Chen, and Shafiq Joty. Relevant or random: Can llms truly perform analogical reasoning?, 2024.

[49] Yuting Hu, Dancheng Liu, Qingyun Wang, Charles Yu, Heng Ji, and Jinjun Xiong. Automating knowledge discovery from scientific literature via llms: A dual-agent approach with progressive ontology prompting, 2024.

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

[51] Qiuejie Xie, Qiming Feng, Tianqi Zhang, Qingqiu Li, Linyi Yang, Yuejie Zhang, Rui Feng, Liang He, Shang Gao, and Yue Zhang. Human simulacra: Benchmarking the personification of large language models, 2024.

[52] Antonino Ferraro, Antonio Galli, Valerio La Gatta, Marco Postiglione, Gian Marco Orlando, Diego Russo, Giuseppe Riccio, Antonio Romano, and Vincenzo Moscato. Agent-based modelling meets generative ai in social network simulations, 2024.

[53] Ivar Frisch and Mario Giulianelli. Llm agents in interaction: Measuring personality consistency and linguistic alignment in interacting populations of large language models, 2024.

[54] Yongxin Deng, Xihe Qiu, Xiaoyu Tan, Chao Qu, Jing Pan, Yuan Cheng, Yinghui Xu, and Wei Chu. Cognidual framework: Self-training large language models within a dual-system theoretical framework for improving cognitive tasks, 2024.

[55] Jintian Zhang, Xin Xu, Ningyu Zhang, Ruibo Liu, Bryan Hooi, and Shumin Deng. Exploring collaboration mechanisms for llm agents: A social psychology view, 2024.

[56] Shirley Wu, Shiyu Zhao, Qian Huang, Kexin Huang, Michihiro Yasunaga, Kaidi Cao, Vassilis N. Ioannidis, Karthik Subbian, Jure Leskovec, and James Zou. Avatar: Optimizing llm agents for tool usage via contrastive reasoning, 2024.

[57] Steffi Chern, Ethan Chern, Graham Neubig, and Pengfei Liu. Can large language models be trusted for evaluation? scalable meta-evaluation of llms as evaluators via agent debate, 2024.

[58] Jinyu Cai, Jialong Li, Mingyue Zhang, Munan Li, Chen-Shu Wang, and Kenji Tei. Language evolution for evading social media regulation via llm-based multi-agent simulation, 2024.

[59] Sergi Blanco-Cuaresma. Psychological assessments with large language models: A privacy-focused and cost-effective approach, 2024.

[60] 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 preprint arXiv:2406.05925*, 2024.

[61] Yin Jou Huang and Rafik Hadfi. How personality traits influence negotiation outcomes? a simulation based on large language models, 2024.

[62] Zainab Iftikhar, Sean Ransom, Amy Xiao, and Jeff Huang. Therapy as an nlp task: Psychologists' comparison of llms and human peers in cbt, 2024.

[63] Abhishek Dutta and Yen-Che Hsiao. Adaptive reasoning and acting in medical language agents, 2024.

[64] Nikita Mehandru, Brenda Y. Miao, Eduardo Rodriguez Almaraz, Madhumita Sushil, Atul J. Butte, and Ahmed Alaa. Large language models as agents in the clinic, 2023.

[65] Ciaran Regan, Nanami Iwahashi, Shogo Tanaka, and Mizuki Oka. Can generative agents predict emotion?, 2024.

[66] Liam Magee, Vanicka Arora, Gus Gollings, and Norma Lam-Saw. The drama machine: Simulating character development with llm agents, 2024.

[67] James WA Strachan, Dalila Albergo, Giulia Borghini, Oriana Pansardi, Eugenio Scaliti, Saurabh Gupta, Krati Saxena, Alessandro Rufo, Stefano Panzeri, Guido Manzi, et al. Testing theory of mind in large language models and humans. *Nature Human Behaviour*, 8(7):1285–1295, 2024.

[68] Aparna Elangovan, Ling Liu, Lei Xu, Sravan Bodapati, and Dan Roth. Considers-the-human evaluation framework: Rethinking human evaluation for generative large language models, 2024.

[69] Sumedh Rasal and E. J. Hauer. Navigating complexity: Orchestrated problem solving with multi-agent llms, 2024.

[70] Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, et al. A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6):186345, 2024.

[71] Shiwen Ni and Min Yang. Educational-psychological dialogue robot based on multi-agent collaboration. In *International Conference on Social Robotics*, pages 119–125. Springer, 2024.

[72] Zhuoling Li, Xiaogang Xu, Zhenhua Xu, SerNam Lim, and Hengshuang Zhao. Larm: Large auto-regressive model for long-horizon embodied intelligence, 2025.

[73] Daniel Albert and Stephan Billinger. Reproducing and extending experiments in behavioral strategy with large language models, 2024.

[74] Ziyang Chen and Stylios Moscholios. Using prompts to guide large language models in imitating a real person's language style, 2024.

[75] Mohammad Amin Abbasi, Farnaz Sadat Mirnezami, and Hassan Naderi. Psycholex: Unveiling the psychological mind of large language models, 2024.

[76] Julian Coda-Forno, Marcel Binz, Jane X. Wang, and Eric Schulz. Cogbench: a large language model walks into a psychology lab, 2024.

[77] Karthik Sreedhar and Lydia Chilton. Simulating human strategic behavior: Comparing single and multi-agent llms. *arXiv preprint arXiv:2402.08189*, 2024.

[78] Timothy R. McIntosh, Teo Susnjak, Nalin Arachchilage, Tong Liu, Paul Watters, and Malka N. Halgamuge. Inadequacies of large language model benchmarks in the era of generative artificial intelligence, 2024.

[79] Yanan Chen, Ali Pesaranghader, Tanmana Sadhu, and Dong Hoon Yi. Can we rely on llm agents to draft long-horizon plans? let's take travelplanner as an example, 2024.

[80] Ruoxi Xu, Hongyu Lin, Xianpei Han, Le Sun, and Yingfei Sun. Academically intelligent llms are not necessarily socially intelligent, 2024.

[81] Pengda Wang, Zilin Xiao, Hanjie Chen, and Frederick L. Oswald. Will the real linda please stand up...to large language models? examining the representativeness heuristic in llms, 2024.

[82] Yuan Li, Yue Huang, Hongyi Wang, Xiangliang Zhang, James Zou, and Lichao Sun. Quantifying ai psychology: A psychometrics benchmark for large language models. *arXiv preprint arXiv:2406.17675*, 2024.

[83] Yadong Zhang, Shaoguang Mao, Tao Ge, Xun Wang, Adrian de Wynter, Yan Xia, Wenshan Wu, Ting Song, Man Lan, and Furu Wei. Llm as a mastermind: A survey of strategic reasoning with large language models. *arXiv preprint arXiv:2404.01230*, 2024.

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

[85] Qun Ma, Xiao Xue, Deyu Zhou, Xiangning Yu, Donghua Liu, Xuwen Zhang, Zihan Zhao, Yifan Shen, Peilin Ji, Juanjuan Li, et al. Computational experiments meet large language model based agents: A survey and perspective. *arXiv preprint arXiv:2402.00262*, 2024.

[86] Yiqun Zhang, Xiaocui Yang, Xingle Xu, Zeran Gao, Yijie Huang, Shiyi Mu, Shi Feng, Daling Wang, Yifei Zhang, Kaisong Song, and Ge Yu. Affective computing in the era of large language models: A survey from the nlp perspective, 2024.

[87] Yaqub Chaudhary and Jonnie Penn. Large language models as instruments of power: New regimes of autonomous manipulation and control, 2024.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

AI-generated, for reference only.