Exploring the Frontiers of LLMs in Psychological Applications: A Comprehensive Review (Postprint)

Authors: Luoma Ke, Song Tong, Peng Cheng, Kaiping Peng, Song Tong, Kaiping Peng

Date: 2025-10-10T00:00:00+00:00

Abstract

This paper explores the frontiers of large language models (LLMs) in psychology applications. Psychology has undergone several theoretical changes, and the current use of Artificial Intelligence (AI) and Machine Learning, particularly LLMs, promises to open up new research directions. We provide a detailed exploration of how LLMs like ChatGPT are transforming psychological research. It discusses the impact of LLMs across various branches of psychology, including cognitive and behavioral, clinical and counseling, educational and developmental, and social and cultural psychology, highlighting their potential to simulate aspects of human cognition and behavior. The paper delves into the capabilities of these models to emulate human-like text generation, offering innovative tools for literature review, hypothesis generation, experimental design, experimental subjects, data analysis, academic writing, and peer review in psychology. While LLMs are essential in advancing research methodologies in psychology, the paper also cautions about their technical and ethical challenges. There are issues like data privacy, the ethical implications of using LLMs in psychological research, and the need for a deeper understanding of these models' limitations. Researchers should responsibly use LLMs in psychological studies, adhering to ethical standards and considering the potential consequences of deploying these technologies in sensitive areas. Overall, the article provides a comprehensive overview of the current state of LLMs in psychology, exploring potential benefits and challenges. It serves as a call to action for researchers to leverage LLLs' advantages responsibly while addressing associated risks.

Full Text

Preamble

Exploring the Frontiers of LLMs in Psychological Applications: A Comprehensive Review

Luoma Ke[1], Song Tong[1],, *Peng Cheng*[2], *Kaiping Peng*[1],

[1] School of Social Science, Tsinghua University
[2] [Affiliation not provided]

Corresponding authors: tong.song.53w@kyoto-u.jp; pengkp@tsinghua.edu.cn

Abstract

This paper explores the frontiers of large language models (LLMs) in psychological applications. Psychology has undergone several theoretical transformations, and the current integration of Artificial Intelligence (AI) and Machine Learning—particularly LLMs—promises to open new research directions. We provide a detailed examination of how LLMs like ChatGPT are transforming psychological research, discussing their impact across various branches of psychology including cognitive and behavioral, clinical and counseling, educational and developmental, and social and cultural psychology. We highlight their potential to simulate aspects of human cognition and behavior, and delve into these models' capabilities for emulating human-like text generation. LLMs offer innovative tools for literature review, hypothesis generation, experimental design, experimental subjects, data analysis, academic writing, and peer review in psychology. While essential for advancing research methodologies, this paper also cautions about technical and ethical challenges, including data privacy, ethical implications of LLM use in psychological research, and the need for deeper understanding of these models' limitations. Researchers should use LLMs responsibly in psychological studies, adhering to ethical standards and considering potential consequences when deploying these technologies in sensitive areas. Overall, this article provides a comprehensive overview of the current state of LLMs in psychology, exploring both potential benefits and challenges. It serves as a call to action for researchers to leverage LLM advantages responsibly while addressing associated risks.

Keywords: Large language models (LLMs) · ChatGPT · Machine learning · Artificial intelligence (AI) · Psychology · Research methodology

1. Introduction

Artificial intelligence (AI) has a history spanning nearly seven decades, beginning with the 1956 Dartmouth Conference. The field has recently been revolutionized by large language models (LLMs) such as OpenAI's ChatGPT series, Google's Bard, and Meta's Llama. Notably, OpenAI's GPT-4 may signify a paradigm shift, demonstrating impressive capabilities across difficult tasks in mathematics, coding, vision, medicine, law, psychology, and other domains (Bubeck et al., 2023)—heralding an era of "AI for Science" (H. Wang et al., 2023). LLMs mark a critical juncture in the evolution of machine learning and AI, propelled by their expansive scale and sophisticated neural architectures that incorporate attention mechanisms (Vaswani et al., 2017). Through integration of cognitive mechanisms (Binz & Schulz, 2023a), these models exhibit emergent behaviors akin to complex physical systems (Wei et al., 2022), which has not only enhanced their ability to understand concepts and high-level semantics (J. Li et al., 2022) but also deepened our insights into cognitive processes (Sejnowski, 2022). In psychological applications, these developments are reshaping how we interact with and comprehend data, language, and our environment (De Bot et al., 2007; Demszky et al., 2023), contributing significantly to fields including clinical psychology (Thirunavukarasu et al., 2023), developmental psychology (Frank, 2023; Hagendorff, 2023), and social psychology (Demszky et al., 2023; Hardy et al., 2023; Zhang et al., 2023). Moreover, they profoundly impact psychological research methodologies, offering novel approaches and tools for exploration and analysis.

1.1 The Concept of LLMs: From Machine Learning to Capability Emergence

Machine learning, particularly natural language processing (NLP), has progressed significantly in the last decade. However, the emergence of LLMs such as GPT-4 and its predecessors marks a substantial leap in AI capabilities. LLMs are deep learning models designed to process natural language text and generate human-like responses. Capability emergence is defined as a qualitative change in behavior resulting from quantitative changes in the system—a capability is emergent if it does not exist in smaller models but appears in larger ones (Wei et al., 2022).

At the heart of LLMs lies the transformer architecture, a deep neural network with an attention mechanism that efficiently processes sequential data in parallel (Vaswani et al., 2017), functioning somewhat similarly to human brain processes. This architecture has revolutionized NLP. The self-attention mechanism captures contextual relationships in textual data, enabling more sophisticated language understanding. The "large" in LLMs refers both to the vast number of parameters and massive training data used—typically billions of parameters and terabytes of text (Binz & Schulz, 2023b)—and to mastering a world model (Yildirim & Paul, 2023).

The process of large language modeling, from machine learning to capability emergence, can be divided into several key stages (Demszky et al., 2023). First, **pre-training**: LLMs are pre-trained on large textual corpora to learn intricate linguistic, syntactic, and semantic structures (P. Liu et al., 2023). This unsupervised learning phase lays the foundation for language understanding. Second, **fine-tuning**: After pre-training, LLMs can be adapted for specific tasks or domains, making them suitable for various applications (Liu et al., 2022). This process ensures the model generates contextually relevant responses and engages in meaningful conversations or tasks. Third, **language comprehension**: LLMs demonstrate remarkable ability to understand and generate human-like text, answering questions, writing articles, summarizing content, translating languages, and even performing creative writing (Bubeck et al., 2023). Their skillful contextual understanding is essential for intelligence emergence. Fourth, **capability emergence**: When integrated into various applications, LLMs exhibit "capability emergence," performing tasks requiring deep language and contextual understanding, often achieving human-like or superhuman performance in specific domains (OpenAI, 2023), such as analogical reasoning (Webb et al., 2023), creativity (Stevenson et al., 2022), and emotion recognition (Patel & Fan, 2023).

Thus, LLMs offer intriguing insights into how these technologies can mimic or augment human cognitive processes. For instance, LLMs' ability to understand and generate natural language echoes aspects of human linguistic and cognitive skills (Goertzel, 2023). This parallel enables exploration of AI applications in cognitive psychology (Sartori & Orrù, 2023), language acquisition (Jungherr, 2023), and even mental health (Lamichhane, 2023). Moreover, studying LLMs contributes to our understanding of the human mind, offering a computational perspective on language processing, decision-making (Sha et al., 2023), and learning mechanisms (Hendel et al., 2023). This fusion advances both AI capabilities and our comprehension of human cognition.

1.2 Psychology and AI

Psychology, as a science exploring the human mind and behavior, has undergone significant theoretical transformations since the late 19th century, evolving from psychoanalysis and behaviorism to cognitive psychology (Hothersall & Lovett, 2022). This history marks not only a shift in research focus but also an academic trend from observing behavioral manifestations to exploring in-depth psychological connotations, with each phase deepening our understanding of human psycho-cognitive processes.

Understanding these psycho-cognitive processes is crucial for psychology. In clinical and counseling psychology, cognitive psychology research supports diagnosing and treating psychological disorders, deepening our understanding of the psychological mechanisms underlying emotions, stress, and human behavior. Psychotherapies such as cognitive-behavioral therapy (Hofmann et al., 2012) and psychodynamic therapy have become essential tools for promoting mental health and emotional regulation.

In educational and developmental psychology, cognitive psychology advances have fostered deeper understanding of perceptual and affective factors in learning (Glaser, 1984), leading to innovations in teaching methods and learning strategies. In social and cultural psychology, cognitive psychology research helps explain individuals' behavior and mental processes in different social and cultural contexts, exploring how cultural differences affect cognitive patterns, values, and behavioral norms—especially important in globalization, interaction, and integration. Meanwhile, in social psychology, cognitive psychology research on group behavior, social influence, prejudice, and discrimination is valuable for promoting social harmony and mutual understanding (Park & Judd, 2005).

AI is a growing force in psycho-cognitive research. Simon (1979) early recognized the potential of computational models to simulate human cognitive processes. The recent emergence of LLMs, represented by OpenAI's GPT family (including GPT-3, ChatGPT, and GPT-4), can process and generate human-like texts and perform close to human levels on some cognitive tasks (Bubeck et al., 2023). Moreover, they provide a unique perspective for studying human cognition. For example, GPT-3 can solve vignette-based tasks similarly to or better than human subjects and make rational decisions based on descriptions, outperforming humans in multi-armed bandit tasks (Binz & Schulz, 2023b). Furthermore, after extensive testing, GPT-3 can solve complex analogical problems at levels comparable to human performance—an essential hallmark of human intelligence (Webb et al., 2023). Fine-tuning across multiple tasks could allow LLMs to predict human behavior in previously unseen tasks, enabling them to serve as general-purpose cognitive models (Binz & Schulz, 2023a), potentially opening new research directions that could transform cognitive psychology and behavioral science.

Newell's time-scale theory provides a multidimensional framework for understanding human behavior (Newell, 1990). In his seminal work, Newell (1990) articulates a nuanced framework stratifying human behavior across four distinct temporal levels (see Fig. 1a [Figure 1: see original paper]). At the foundational **biological level**, core biological and physiological processes operate on rapid time scales from approximately one millisecond to one second, including neural responses and sensory processing fundamental to human cognition. Advancing to the **cognitive layer**, fundamental cognitive mechanisms operate on intermediate time scales from one second to around one minute, encompassing basic operations such as attention, perception, and short-term memory. At the **rational layer**, more elaborate and sustained cognitive activities extend from several minutes to a few hours, involving complex problem-solving, planning, and decision-making requiring higher cognitive engagement. Finally, the **social layer** encapsulates human behavior within social interactions and structures, characterized by the broadest time scales ranging from hours to days or more, delving into social communication dynamics, group behavior, and cultural influences on cognition. This layered approach underscores the multifaceted nature of human behavior, highlighting the interplay between rapid physiological processes and prolonged, socially-influenced aspects of cognition.

LLMs have great potential for modeling cognition and behavior across these different time scales (see Fig. 1b) and can provide new insights into human psycho-cognitive processes. Recent research reveals significant advancements in LLMs' ability to emulate complex human cognitive and social behaviors (Grossmann et al., 2023; Marjieh et al., 2023; Orru et al., 2023; Pal et al., 2023; Stevenson et al., 2022; Webb et al., 2023). Grossmann et al. (2023) and Marjieh et al. (2023) demonstrate LLMs' proficiency in simulating human social interactions and perceptual processing, respectively. Orru et al. (2023) and Webb et al. (2023) highlight LLMs' capabilities in complex problem-solving and reasoning, while Hagendorff et al. (2023) focus on decision-making processes. Stevenson et al. (2022) document LLMs' potential for creativity, and Patel and Fan (2023) demonstrate their ability in emotion recognition. These findings collectively indicate the expanding role of LLMs in mimicking and enhancing human cognitive and social functions, marking significant progress in AI research.

As general cognitive models (Binz & Schulz, 2023a), LLMs can provide new perspectives and approaches for research in cognitive and behavioral psychology, clinical and counseling psychology, educational and developmental psychology, and social and cultural psychology across different time scales of human behavior (see Fig. 1a). Fig. 1 illustrates LLMs' emergent abilities applied in psychological domains and as research tools: (a) applications across psychological domains from the perspective of behavioral time scales, (b) emergent abilities of LLMs, and (c) LLMs functioning as research tools.

Based on these emergent abilities, LLMs can also serve as research aids (see Fig. 1c) to assist psychologists with everything from literature review (Aydın & Karaarslan, 2022; Qureshi et al., 2023), experimental subjects (Dillion et al., 2023; Hutson, 2023), and data analysis (Patel & Fan, 2023; Peters & Matz, 2023; Rathje et al., 2023) to academic writing (Dergaa et al., 2023; Stokel-Walker, 2022) and peer review (Chiang & Lee, 2023; Van Dis et al., 2023). Thus, LLMs have the potential to become research assistants for psychologists, helping improve research efficiency.

1.3 Objectives and Significance of the Present Review

This review systematically examines LLM use across various psychological domains, analyzing applications over different behavioral time scales. The exploration begins with LLMs in cognitive and behavioral psychology (Section 2), followed by their roles in clinical and counseling psychology (Section 3). The review then transitions to educational and developmental psychology (Section 4) and social and cultural psychology (Section 5), outlining LLMs' contributions in each area. To deepen understanding of LLMs' impact on psychological research, Section 6 provides an overview of their potential as scientific research tools. The review also addresses challenges and future research directions in applying LLMs to psychological contexts, concluding with a summary of applications and recommendations for future work. Crucially, this review proposes integration strategies for LLMs in psychological research and offers insights into interpreting these models from a psychological standpoint, contributing to their safety and interpretability.

2. LLMs in Cognitive and Behavioral Psychology

Within multilevel time scales of human behavior (Newell, 1990), cognitive and behavioral psychology focuses primarily on studying cognitive processes on sub-hourly time scales (see Fig. 1), encompassing human engagement in perception, memory, thinking, decision-making, problem-solving, and conscious planning. Cognitive and behavioral psychology typically employs experimental methods to study these processes by controlling and observing behaviors and responses under specific conditions.

The recent emergence of LLMs has reinvigorated discussion about whether human cognitive abilities are revealed in these models given sufficient training data. If so, studying LLMs' cognitive processes could yield knowledge about human cognition, forming a valuable addition to existing research methods in cognitive psychology.

Binz and Schulz (2023a) found that fine-tuning multiple tasks enabled LLMs to predict human behavior in previously unseen tasks, suggesting that LLMs can be adapted to become generalist cognitive models. In another study, they tested GPT-3 using tools from cognitive psychology and showed that it made better decisions and outperformed humans in a multi-armed bandit task (Binz & Schulz, 2023b).

Additional studies have demonstrated that LLMs possess perceptual judgment (Marjieh et al., 2023), reasoning (Webb et al., 2023), decision-making abilities (Hagendorff et al., 2023), creativity (Stevenson et al., 2022), and problem-solving skills (Orru et al., 2023). One study even demonstrated that GPT-4 possesses the mental abilities of a seven-year-old child through a false belief task (considered the gold standard for testing theory of mind in humans) (Kosinski, 2023). For example, Hagendorff et al. (2023) explored reasoning capabilities and decision-making processes of the OpenAI GPT family through the following experimental method: designing a series of semantic illusion and cognitive reflection tests intended to elicit intuitive but erroneous responses; applying these tasks—traditionally used to study human reasoning and decision-making—to OpenAI's generative pre-trained transformer models; analyzing model performance on Cognitive Reflection Test (CRT) and semantic illusion tasks to reveal their System 1 and System 2 thought processes; observing how ChatGPT models show correct responses and avoid pitfalls; and evaluating model performance on CRT tasks by preventing chain-of-thought reasoning. Results show that as model size and language capability increase, the OpenAI family increasingly exhibits human-like intuitive System 1 thinking and associated cognitive errors.

Table 1 summarizes applications of LLMs to cognitive and behavioral psychology.

Table 1. Applications of LLMs in Cognitive and Behavioral Psychology Research

Author	Research Question	Research Method	Key Finding
Sartori & Orrù (2023)	Human-like properties LLMs exhibit in various cognitive tasks	Decision-making, deliberation, information search, reasoning, Wason Selection Task, and Raven-like matrices	LLMs demonstrate human-like performance in cognitive psychology
Hagendorff et al. (2023)	Whether OpenAI GPT family's reasoning and decision-making resembles human System 1 and System 2 thought	Analyze model performance on CRT and semantic illusion tasks	ChatGPT models show correct responses and avoid pitfalls; increasing model size exhibits human-like intuitive System 1 thinking and cognitive errors
Hutson (2023)	Can AI language models replace human participants?	ChatGPT-3.5 and 4 use input-output context windows during chain-of-thought reasoning, similar to human laptop-support System 2 thinking	LLMs can replace human participants in experimental research in some cases
Dillion et al. (2023)	Explore whether LLMs can replace human participants in psychological sciences	LLMs (e.g., GPT-3.5) used to conduct experiments instead of human participants	LLMs can replace human participants in some cases
Zhuang et al. (2023)	How to efficiently measure cognitive abilities of LLMs	Making human-like moral judgments by analyzing similarity of GPT-3.5's judgments to humans	LLMs can substitute for human participants in some cases

Author	Research Question	Research Method	Key Finding
Grossmann et al. (2023)	Improve social science research methods amid LLM impact	Computerized Adaptive Testing (CAT) for assessing cognitive ability in LLMs	LLMs have great potential in social science research due to ability to model human behavior and generate diverse responses
Loconte et al. (2023)	Neuropsychological evaluations of LLM prefrontal functioning	Verbal Reasoning Test, Cognitive Estimation, Metaphor and Idioms Comprehension, Winograd Schema, Tower of London, Hayling Sentence Completion, Compound Remote Associate problems, Social Cognition	ChatGPT exhibits disjointed cognitive profiles in prefrontal functioning (some above average, others pathological)
Binz & Schulz (2023a)	Better describe human decision-making through LLM fine-tuning	Compare goodness-of-fit: random guessing model, unfinetuned LLaMA, domain-specific model, and fine-tuned model	Fine-tuned LLMs (e.g., LLaMA) successfully capture human decision-making and outperform domain-specific models
Orru et al. (2023)	Potential of ChatGPT as intelligent tool for problem-solving	Verbal insight problems: "practice problems" and "transfer problems"	ChatGPT's global performance identical to most likely human sample results

Author	Research Question	Research Method	Key Finding
Hagendorff (2023)	How to use psychological methods to study emergent abilities	Treat LLMs as participants in psychological experiments	Uncover emergent abilities undetectable by traditional NLP benchmarks
Shiffrin & Mitchell (2023)	GPT-4 performance on cognitive psychology datasets	Evaluate GPT-4 on CommonsenseQA, SuperGLUE, MATH, and HANS	GPT-4 has revolutionary potential to bridge human-machine reasoning gap
Dhingra et al. (2023)	Decision-making mechanisms and other psychological qualities of LLMs	Two experiments using psychological literature and prompts not in GPT-3 training corpus	GPT-3 shows surprising abilities in decision-making, information search, deliberation, but poor causal reasoning
Schulz (2023b)	Assessing GPT-3's cognitive ability	Vignette-Based Decision-Making, Information Search, Deliberation, Causal Reasoning	GPT-3 outperforms humans on some tasks, performs poorly on others
Marjieh et al. (2023)	How LLMs predict human sensory judgments across modalities	Use GPT-3, GPT-3.5, GPT-4 to predict judgments in six sensory modalities (pitch, loudness, color, sound, taste, timbre)	LLMs successfully predict human perceptual judgments across six modalities

Author	Research Question	Research Method	Key Finding
Huang & Chang (2022)	How to improve and direct model reasoning	Methods: fully supervised fine-tuning, cueing, contextual learning, problem decomposition, mixed-methods	Improving reasoning requires training data, architectures, and optimization goals specific to reasoning
Hagendorff et al. (2022)	Machine intuitive capabilities in GPT-3.5	Conduct CRT and Semantic Illusion Test on GPT-3.5	GPT-3.5 systematically exhibits "machine intuition"—human-like erroneous decisions on CRT and semantic illusions
Stevenson et al. (2022)	Human-like intuitive decision-making and creativity in GPT-3	Compare GPT-3 to Alternative Uses Test (AUT) used in human creativity research	Humans currently outperform GPT-3 in originality and unexpectedness, but GPT-3 performs better in utility

These research cases demonstrate that LLMs possess human-like cognitive abilities (Zhuang et al., 2023). Studying LLMs' cognitive mechanisms would provide new insights into human cognitive processes and offer promising avenues for advancing psychological research methodologies and understanding complex cognitive phenomena as these models evolve.

3. LLMs in Clinical and Counseling Psychology

Within multilevel time scales of human behavior (Newell, 1990), clinical and counseling psychology involves assessment of everyday behavioral acts (hours to a day), habitual thinking (days to months), and psychological disorders (months to years) (see Fig. 1). Clinical and counseling psychology focuses on assessing, diagnosing, treating, and preventing individual mental health problems—processes often involving medium- to long-term periods. Reports indicate a public rush to use LLMs like ChatGPT for mental health screening and treatment (Demszky et al., 2023). LLMs are expected to be useful in clinical and counseling contexts because they can parse human language, generate human-like responses, categorize text, and flexibly adapt conversational styles representing different theoretical orientations (Stade et al., 2023). This raises the question: how effectively do LLMs work in psychotherapy, and can they replace human psychotherapists?

LLMs are basic generalized models with few-shot learning capabilities (Brown et al., 2020), allowing them to quickly become experts in clinical and counseling domains with minimal training data. For example, LLMs trained on clinical content can identify specific change factors that help psychologists understand clinical intervention processes, thus opening the "black box" of psychotherapy (Schueller & Morris, 2023). Additionally, studies show LLMs can correctly recognize emotions and respond appropriately (Patel & Fan, 2023; Schaaff et al., 2023), and human-AI collaboration in clinical psychological support results in greater empathy (Sharma et al., 2023). LLMs can also accomplish mental health assessments (Elyoseph & Levkovich, 2023; Kjell et al., 2023) and individualized interventions (Blyler & Seligman, 2023a, 2023b).

Blyler and Seligman (2023a) proposed an individualized intervention: participants were recruited from previous studies (age 18+). From five narrative identities generated by ChatGPT-4 rated as "completely accurate," five participants representing different backgrounds were selected. Participants provided narratives through dialogue with ChatGPT-4, and the AI was asked how it would guide life coaching based on these identities. ChatGPT-4 was then asked to recommend specific interventions based on the AI-generated narratives and coaching methods. Results suggest that coaching strategies and interventions generated by ChatGPT-4 make perfect sense based on narrative identity.

Table 2 summarizes applications of LLMs to clinical and counseling psychology.

Table 2. Applications of LLMs in Clinical and Counseling Psychology Research

Author	Research Question	Research Method	Key Finding
Carlbring et al. (2023)	Can AI improve effectiveness of Internet interventions?	Internet intervention methods: real-time video therapy, digital self-help programs, combining interventions with face-to-face therapy	AI can work with therapists to improve outcomes

Author	Research Question	Research Method	Key Finding
Blyler & Seligman (2023a)	Can narrative identity provide personalized intervention approaches?	ChatGPT-4 generates personalized narrative identities from stream-of-consciousness thoughts and demographic information, then provides targeted coaching methods	ChatGPT-4 generates highly credible coaching strategies based on constructed narrative identities
Blyler & Seligman (2023b)	AI potential in psychological practice	Process stream-of-consciousness and demographic data through ChatGPT-4 to generate personal narratives, evaluate for accuracy, surprise, and illumination	AI can support self-discovery in psychotherapy and coaching
Abd-Alrazaq et al. (2019)	How chatbot features meet mental health needs	Characterize chatbots by applications, interactions, response generation methods, virtual representatives, and platform implementations	Chatbots focus primarily on depression and autism; most implementations in developed countries
Elyoseph & Levkovich (2023)	ChatGPT potential for suicide risk assessment	Hypothetical case study: ChatGPT assessed mental health indicators in patients with varying self-burden and belonging frustration; compare to professionals	ChatGPT underestimated suicide attempt risk in all scenarios vs. mental health professionals

Author	Research Question	Research Method	Key Finding
Kjell et al. (2023)	Using LLMs to change psychological assessment	Analyze natural language responses using LLMs to extract mental health information; assess strengths (accuracy, scope, parsing, openness) and limitations (bias, risk, ethics)	LLMs can transform assessments from rating scales to natural language communication
J. M. Liu et al. (2023)	Bias in LLMs across age and gender	Datasets: i2b2 2006 smoking and i2b2 2008 obesity. Multi-label classification task. BERT-based models trained 1000 epochs with Adam optimizer. Micro F1-score evaluation. Bias analysis across subgroups.	Creating population subgroups by age and gender found most subgroups exhibited bias amplification
Schueller & Morris (2023)	LLMs in clinical/counseling psychology: capabilities and interventions	Develop novel digital mental health interventions (DMHIs). Analyze therapist discourse to identify symptom improvement predictors. Identify specific change factors.	LLMs have wide applications in clinical science/practice to improve interventions but won't completely replace therapists
Sharma et al. (2023)	How AI assists peer supporters in improving empathy	Develop HAILEY (Human-AI collaboration for EmpathY) intelligent feedback system to evaluate AI role in improving peer supporter empathy	AI helps peer supporters demonstrate higher empathy levels

Author	Research Question	Research Method	Key Finding
Graber-Stiehl (2023)	Are AI treatments ready for mainstream adoption?	Koko app users could get advice from Kokobot (GPT-3 based), edit responses, and send them	Despite promise, many ethical and safety issues remain with current AI therapies
Stade et al. (2023)	Responsibly develop/evaluate LLMs while realizing potential in behavioral health	Focus on evidence-based practice, rigorous evaluation, interdisciplinary collaboration, trust/usability for therapists/patients, effective clinical LLM standards	LLMs have great potential in psychotherapy but concerns about accuracy, accountability, transparency, ethics
Zhong et al. (2023)	Challenges of LLMs in psychiatric research/practice	Improve diagnostic accuracy by analyzing large patient datasets to identify patterns difficult for humans to detect; identify patient characteristics predicting treatment response	LLMs have great potential in psychiatric practice but concerns about reliability, pragmatics, accountability, transparency, ethics

These research cases demonstrate LLMs' ability to provide adequate mental health support (Schueller & Morris, 2023), holding promise to address capacity shortages in mental health care systems. As they evolve, LLMs may provide more individualized treatment services and potentially enable fully automated psychotherapy (Stade et al., 2023). However, ensuring LLM safety and privacy protection in psychotherapy is essential.

4. LLMs in Educational and Developmental Psychology

Within multilevel time scales of human behavior (Newell, 1990), educational and developmental psychology operates at relatively medium- to long-term levels (see Fig. 1), reflecting ongoing learning and development characterizing educational processes. This field concerns learning processes, knowledge accumulation, skill development, and individual psychological changes within educational environments. A national survey found that only three months after ChatGPT's public release, 40% of U.S. teachers used it weekly for lesson planning (Demszky et al., 2023).

Table 3 summarizes applications of LLMs to educational and developmental psychology. The potential applications are manifold: facilitating personalized learning, emotion recognition, mental health support, educational assessment, and improving learning motivation. Specifically, LLMs trained on massive internet and book data (Binz & Schulz, 2023b) can serve as more knowledgeable learning aids (Stojanov, 2023), provide personalized learning experiences (Kasneci et al., 2023), and enhance learning motivation (Ali et al., 2023). For example, Stojanov (2023) explored ChatGPT as a learning tool through an autobiographical methodology: he began by setting learning objectives and conversing with ChatGPT about its functionality for four hours, continued discussion for three more hours while watching relevant YouTube videos, and reported positive feedback, finding it a motivating and relevant learning experience.

Table 3. Applications of LLMs in Educational and Developmental Psychology Research

Author	Research Question	Research Method	Key Finding
Frank (2023)	What abstractions underlie LLM behavior?	Ensure LLMs haven't been pre-exposed to experimental stimuli; select simplified stimuli; need convergence evidence across multiple tasks	Developmental researchers can better understand LLM representations by drawing on psychological approaches
Han (2023)	LLMs in moral developmental research	Use Behavioral Definitional Issue Test (bDIT); examine ChatGPT reading Martin Luther King letter; analyze three moral exemplar stories	LLMs aid education ethics research involving empirical and practical inquiry
Stojanov (2023)	Experience using ChatGPT as more knowledgeable other	Autobiographical research methodology exploring ChatGPT as support in learning process	Instant answers create 'flow' experience, but in 'immersion' state users may overestimate knowledge

Author	Research Question	Research Method	Key Finding
Kasneci et al. (2023)	Opportunities and challenges of LLMs in education	LLMs help create educational content, provide personalized learning, aid language learning, research, writing, assessment, grading, professional development	LLMs have great potential in education, benefiting students and teachers through personalization, engagement, and diverse content
Ali et al. (2023)	ChatGPT impact on English language student motivation	5-point Likert scale collecting perceptions of ChatGPT impact on English learning, interest, self-directed learning, interaction, enjoyment	ChatGPT positively impacts English language learner motivation
Kosinski (2023)	Do LLMs spontaneously generate Theory of Mind?	Two gold-standard false-belief tasks: Unexpected Contents (Smarties) and Unexpected Transfer (Maxi/Sally-Anne)	ToM, previously thought uniquely human, may emerge spontaneously as byproduct of improved language skills

5. LLMs in Social and Cultural Psychology

Within time scales of human behavior (Newell, 1990), social and cultural psychology covers predominantly long-term dimensions (see Fig. 1), reflecting its focus on social interactions and long-term behavioral patterns and mental processes in social environments. This field studies how individual behavior is influenced by social and cultural environments and others, and how individuals affect these environments. Research typically focuses on interpersonal interactions (Tajfel, 1982), group behavior, attitude formation and change, and social cognition.

LLMs can simulate human responses and behaviors to test theories and hypotheses about human behavior (Grossmann et al., 2023). In social and cultural psychology, LLMs can revolutionize the field by analyzing large textual datasets, modeling social interactions, and providing insights into human behavior and social dynamics (Salah et al., 2023).

First, LLMs share similarities with humans in social cognition. Research finds LLMs exhibit typical human cognitive biases in judgment and decision-making, such as anchoring effect, representativeness heuristic, and base rate neglect (Talboy & Fuller, 2023). Additionally, cultural psychology research shows significant differences in Eastern vs. Western cognitive processes for information processing and judgment (Nisbett et al., 2001), and LLMs consistently favor Eastern holistic thinking (Jin et al., 2023).

Second, LLMs characterize human groups in social interaction settings. They replicate Milgram's electroshock experiments (Aher et al., 2023), show gaming abilities in specific games (Akata et al., 2023), and exhibit different risk-taking and pro-social behaviors under different emotional states (Yukun et al., 2023).

Third, LLMs can serve as specific social and cultural psychology research samples. One study explored LLMs as valid proxies for specific human subgroups, finding they contain information beyond superficial similarity, reflecting complex interplays between ideas, attitudes, and sociocultural contexts characterizing human attitudes (Argyle et al., 2022). Another study tested LLMs for personality and values, finding scores similar to human samples (Miotto et al., 2022).

Thus, LLMs have many applications in social and cultural psychology, enabling testing of theories and hypotheses about human behavior in social and cultural interaction settings. For example, one study explored whether AI chatbots can adapt financial decisions and pro-social behaviors through emotional cues like humans (Yukun et al., 2023). The experimental design had two parts: Study 1 examined investment decision-making (susceptible to emotional cues), hypothesizing that AI chatbot risk-taking would be lower with fear cues and higher with joy cues. By providing different emotional cues (fear, joy, or neutral), researchers collected and analyzed bot responses. Study 2 measured pro-social responses through donations to a sick friend under anxiety and joy cues, exploring whether emotional cues influenced pro-social behavior.

Table 4 summarizes applications of LLMs to social and cultural psychology.

Table 4. Applications of LLMs in Social and Cultural Psychology Research

Author	Research Question	Research Method	Key Finding
Atari et al. (2023)	How similar are LLMs to different human groups?	World Values Survey (WVS) data: place LLMs on spectrum of contemporary human psychological change; standard cognitive tasks; thinking styles compared to 31 human groups; self-concept tasks	LLMs perform as outliers on psychometrics vs. humans
Jin et al. (2023)	Does ChatGPT think holistically or analytically?	Two scales measuring cognitive processes: Analytic-Holistic Scale (AHS) and Ternary Categorization Task (TCT); two scales investigating cultural thinking values: Dialectical Self Scale (DSS) and Self-Construction Scale (SCS)	ChatGPT favors Eastern holistic thinking in cognitive processes but not in value judgments
Schaaff et al. (2023)	What are ChatGPT's empathic abilities?	Understanding/expressing emotions: ChatGPT generates responses based on prompts compared to expected categories. Parallel Emotional Responses: analyze emotional responses to different categories. System-level assessment using recognized questionnaires (IRI, EQ, TEQ, PES, AQ)	ChatGPT understands others' emotions and perspective but has difficulty showing higher empathy vs. healthy humans

Author	Research Question	Research Method	Key Finding
Salah et al. (2023)	Use of generative AI in social psychology research	Simulating social interactions; analyzing large textual data; extracting social cognitive processes	ChatGPT has great potential to help analyze data, simulate interactions, and provide insights into human behavior
Harding et al. (2023)	Can LLMs replace human participants in moral psychology?	Ethical judgment experiment: GPT-3.5 assesses 464 ethical scenarios previously rated by humans. Consumer Behavior: GPT-3.5 simulates consumer behavior. Diversity: GPT-3 exhibits different personalities via character traits. Social Simulation: GPT-3 creates SimReddit with 1,000 virtual users	Despite potential to simulate human behavior, LLMs unlikely to fully replace human participants
Patel & Fan (2023)	Emotional Intelligence in LLMs	Standardized Emotional Understanding Test (SECEU) based on school, family, social scenarios; participants assign 4 emotions to each scenario	Current LLMs have human-equivalent sentiment recognition but vary widely in EI performance

Author	Research Question	Research Method	Key Finding
X. Wang et al. (2023)	Can LLMs replace human participants?	Potential applications in questionnaires, behavioral tests, mixed-methods analysis, agent-based modeling, observational studies, experiments	LLMs can replace humans to some extent for data collection and in ABM to explore how individuals with specific characteristics influence interactions
Akata et al. (2023)	How LLMs interact in repeated games	LLMs play limited repetition games against other LLMs and simple human-like strategies; analyze behavior in Prisoner's Dilemma and War/Gender games	With appropriate cues, GPT-4 behaves more forgivingly and coordinates better
Abramski et al. (2023)	Biases of LLMs in math/STEM	Behavioral Formal Mental Networks (BFMNs); semantic frame analysis; compare GPT-3, GPT-3.5, GPT-4 perceptions to high school student data	As LLMs evolve, they may produce less biased models and reduce harmful stereotypes
Yukun et al. (2023)	Do LLMs adjust financial/pro-social behaviors to emotional cues?	Study 1: Collect investment decision responses under fear, joy, or neutral cues. Study 2: Measure pro-social responses (donations) under anxiety and joy cues	ChatGPT-4 exhibits human-like coordinated responses in financial decisions and pro-social behavior when given emotional guidance

Author	Research Question	Research Method	Key Finding
Suri et al. (2023)	Do LLMs show human decision-making heuristics/biases?	Representativeness heuristic, anchoring, availability heuristic, framing effects, endowment effect tests	LLMs may partially drive human decision-making heuristics but lack cognitive-affective processes
Talboy & Fuller (2023)	How do human cognitive biases permeate LLM output?	Anchoring: low/high anchor prompts for book estimates. Representativeness: career choices for people with particular traits. Base rate neglect. Framing: drug efficacy ratings	LLMs suffer from cognitive biases in output
Park et al. (2022)	Generate prototype with real social interaction	Test whether social simulator behaviors are credible across novel communities; generate 50 GPT-3 discussions for post-2020 subreddit; participants identify real vs. generated	Participants unable to distinguish social simulations from actual community behaviors
X. Li et al. (2022)	Are LLMs psychologically safe? Do they show mental illness features?	Personality tests (SD-3, BFI) and well-being tests (FS, SWLS) with unbiased prompts; sample 3 outputs per prompt and calculate means	LLMs exhibit relatively dark personality traits regarding psychological safety
Miotto et al. (2022)	Personality traits, values, demographics of GPT-3	Assess personality (HEXACO) and values (Human Values Scale) with validated instruments; test temperature parameter effects	GPT-3 similar to humans in personality but shows different traits at different temperatures

Author	Research Question	Research Method	Key Finding
Argyle et al. (2022)	Can LLMs model specific human subgroups?	Generate story contexts about Democrats/Republicans; create in silico samples for 2012, 2016, 2020 ANES; assess algorithmic associations	GPT-3 algorithmic bias is fine-grained and demographically relevant, enabling accurate simulation of various human subgroup response distributions
Trott et al. (2023)	Can LLMs understand others' beliefs like humans?	Test GPT-3 on written False Belief Task; compare human and GPT-3 performance	GPT-3 performs worse than humans and doesn't fully explain human behavior on false belief tasks
Aher et al. (2023)	Simulate multiple human behaviors using LLMs	Ultimatum Game, Garden Path Sentences, Milgram Shock Experiment, Wisdom of Crowds	LLMs exhibit "over-accuracy distortion" that may affect downstream applications, education, and arts

6. LLMs as Research Tools in Psychology

LLMs are powerful scientific research tools that can assist psychologists with literature review, hypothesis generation, experimental design, experimental subjects, data analysis, academic writing, and peer review (see Table 5).

Table 5. LLMs as Research Tools in Psychology

Topic	Related Studies
Literature review	LLMs can summarize research literature (Dis et al., 2023), complete literature review tasks (Qureshi et al., 2023), and create review articles (Aydın & Karaarslan, 2022). Specialized LLMs have been trained for systematic literature reviews (Taylor et al., 2022).
Hypothesis generation	LLMs can generate hypotheses from literature, make inferences from data, and clarify conclusions through interpretation (Zheng et al., 2023). They can quickly test hypotheses and learn from mistakes (Park et al., 2023).
Experimental design	LLMs provide text-based materials, optimizing research processes and reducing complexity. Researchers can create stimuli, develop test items, and simulate interactive sessions in controlled environments (Aher et al., 2022; Akata et al., 2023), providing high control and precision.
Experimental subjects	LLMs can simulate human behaviors and responses, testing theories about human behavior (Grossmann et al., 2023). They save time and costs, can be used when human participation is inappropriate (Hutson, 2023), and may serve as alternatives depending on research topic, task, and sample (Dillion et al., 2023).
Data analysis	LLMs efficiently analyze massive textual data for insights into human behavior and emotions at unprecedented scale (Patel & Fan, 2023). They analyze multilingual data, detect psychological structures (Rathje et al., 2023), and derive psychological profiles from social media (Peters & Matz, 2023).
Academic writing	LLMs assist human writing (Dergaa et al., 2023; Stokel-Walker, 2022; Van Dis et al., 2023).
Peer review	LLMs assessed text quality in two NLP tasks, with results consistent with human experts (Chiang & Lee, 2023). They offer quick solutions for PhD students and time-pressured reviewers (Van Dis et al., 2023).

6.1 Automated Literature Review and Meta-Analysis

Conducting literature reviews and meta-analyses is complex and arduous, requiring significant expertise and time (Michelson & Reuter, 2019). Nature reports researchers using ChatGPT as a research assistant to summarize study literature (Dis et al., 2023). One study utilized ChatGPT for systematic literature review tasks (Qureshi et al., 2023). Another created a literature review article on digital twins in health using ChatGPT, finding accelerated knowledge compilation but requiring further academic validity verification (Aydın & Karaarslan, 2022). Researchers have also trained specialized LLMs for scientific research needs (Taylor et al., 2022) that can accomplish systematic literature reviews.

In summary, LLMs accelerate literature review and meta-analysis processes, enabling researchers to systematically review and synthesize existing research, improving evidence-based psychology efficiency.

6.2 Hypothesis Generation and Experimental Design

Hypothesis-driven research is core to scientific activity. LLMs can generate hypotheses from scientific literature, make inferences from data, and clarify conclusions through interpretation (Zheng et al., 2023). Although capable of generating research hypotheses and becoming better "hypothesis machines," they need improved logical and mathematical derivation capabilities to eliminate factual errors and enable rapid, automated hypothesis testing and learning from mistakes (Y. Park et al., 2023). As innovative tools, LLMs have great potential in psychological experiments, providing text-based materials that optimize research processes and reduce complexity. Researchers can easily create experimental stimuli, develop test items, and simulate interactive sessions in controlled environments (Aher et al., 2022; Akata et al., 2023), providing high control and precision.

In conclusion, LLMs provide powerful, flexible tools for psychological research—from hypothesis generation to experimental design—helping researchers achieve more efficient and precise goals.

6.3 As Subjects in Psychological Experiments

Although LLMs can simulate human behaviors and responses, providing opportunities to test theories about human behavior (Grossmann et al., 2023), controversy remains about whether they can substitute for human subjects.

Some researchers argue LLMs can replace human participation to save time and costs, applicable when human participation is inappropriate, while acknowledging potential problems like bias and insufficient training data (Hutson, 2023). Others propose using LLMs as alternative participants when appropriate, based on performance and factors like research topic, task, and sample (Dillion et al., 2023). Some believe that while LLMs will significantly impact research, they are unlikely to meaningfully replace human participants (Harding et al., 2023). Despite controversy, studies show LLMs perform similarly to humans when used as subjects (Orru et al., 2023; Park et al., 2023), suggesting potential for replacement.

In conclusion, although LLMs can simulate human judgments, their understanding of human thinking remains limited, and their output requires careful validation and interpretation when used as psychological subjects.

6.4 Tools for Data Analysis

Various AI forms have long analyzed psychological data, such as flight data for pilot screening (Ke et al., 2023). Machine learning algorithms facilitate processing large datasets, identifying overlooked patterns and correlations. LLMs elevate this capability, efficiently analyzing massive textual data for unprecedented insights into human behavior and emotions (Patel & Fan, 2023). This enables faster, more comprehensive data analysis leading to more reliable, nuanced findings. LLMs analyze multilingual textual data, accurately detect psychological structures (Rathje et al., 2023), and derive psychological profiles from social media (Peters & Matz, 2023). Additionally, LLMs demonstrate competence in medicine, predicting optimal neuroradiographic imaging modalities for clinical presentations, though not outperforming experienced neuroradiologists, indicating need for continued improvement (Nazario-Johnson et al., 2023). These findings demonstrate LLMs' great potential in data evaluation and analysis.

6.5 Paper Writing and Peer Review Tools

While LLMs are not yet complete replacements for human writing, they compellingly answer questions and generate naturally fluent, informative content—though without real intelligence, merely generating text based on previously seen word patterns (Stokel-Walker, 2022). One study had students use ChatGPT for writing assistance, but results showed no significant differences in quality, speed, or authenticity between experimental and control groups. The authors suggest experienced researchers may better guide ChatGPT to high-quality information than students (Bašić et al., 2023). Another article discusses ChatGPT's prospects and threats in academic writing, emphasizing prioritization of peer-reviewed sources and noting potential advantages: handling large textual datasets, automatic abstract generation, and research question creation (Dergaa et al., 2023). Additionally, LLMs show peer review potential (Van Dis et al., 2023), with evaluation results consistent with human experts (Chiang & Lee, 2023).

In conclusion, LLMs like ChatGPT are potent academic writing tools capable of processing large textual datasets and automating previously manual tasks: scanning papers, extracting essential details, generating objective abstracts, and creating research questions. They also show peer review potential. However, researchers must exercise caution, as LLMs can integrate false or biased information, causing unintentional plagiarism and misattribution (Dis et al., 2023).

7.1 Challenges and Limitations

Although LLMs' potential to simulate complex cognitive processes is enormous, providing new tools to explore human cognition and behavior with wide applications across clinical, counseling, educational, developmental, and social-cultural psychology, their output should not be mistaken for thought but viewed as complex pattern matching based on probabilistic modeling (Floridi & Chiriatti, 2020). Despite impressive performance, this differs from consciousness or genuine understanding. Capability interpretation must be based on understanding limitations and operational nature, which may fundamentally differ from human cognition. Therefore, focusing on LLM potential in psychological research requires confronting technical and ethical challenges.

First, despite emergent competence (Wei et al., 2022), LLMs' internal mechanisms remain black boxes from a cognitive and behavioral psychology perspective. LLMs perform impressively on tasks requiring formal linguistic competence (knowledge of language rules and patterns) but fail many tests requiring functional competence (cognitive abilities needed to understand and use language in real-world contexts) (Mahowald et al., 2023). They excel in analogical and moral reasoning but perform poorly on spatial reasoning tasks (Agrawal, 2023).

Second, while LLMs accelerate AI technology use in clinical and counseling psychotherapy, privacy and ethical issues may arise (Graber-Stiehl, 2023). Gatekeepers, patients, and professionals relying on ChatGPT for suicide risk assessment or decision-making support may receive inaccurate assessments underestimating actual risk (Elyoseph & Levkovich, 2023) and may bias clinician decision-making, leading to healthcare inequity (Pal et al., 2023). Additionally, LLMs in psychiatry research and practice face potential bias and privacy violations (Zhong et al., 2023).

Third, in educational, developmental, and social-cultural psychology, LLMs face application challenges. In education, they risk output bias and misuse (Kasneci et al., 2023). One study found ChatGPT-generated texts were not always consistent or logical, sometimes contradictory (Stojanov, 2023). In social-cultural psychology, LLMs exhibit human-like cognitive biases (Talboy & Fuller, 2023) and cultural biases (Atari et al., 2023), plus implicitly darker personality patterns (X. Li et al., 2022). Bender et al. (2021) argue that LLM training data may reflect social biases perpetuated in research settings.

Finally, as research aids, LLMs have limitations. In writing, they don't fully replace humans, generating fluent content without real intelligence—merely pattern-based text (Stokel-Walker, 2022). Although they simulate human judgments as experimental subjects, understanding of human thought remains limited (Dillion et al., 2023). Van Dis et al. (2023) note LLMs may accelerate innovation, shorten publication times, and increase scientific diversity and equality, but may also reduce research quality and transparency, fundamentally altering scientist autonomy.

In summary, while LLMs offer extraordinary capabilities for psychological research, they present challenges related to bias, ethics, data security, transparency, and technical expertise. Researchers should be fully aware of these challenges and address them responsibly. Table 6 summarizes these challenges and limitations.

Table 6. Challenges and Limitations of LLMs in Psychological Applications

Author	Domain	Challenges and Limitations
Mitchell (2023)	Cognition & Behavior	Lack of real-world understanding; lack of abstract reasoning; lack of user intent understanding
Stella et al. (2023)	Cognition & Behavior	Lack of meta-knowledge leads to information processing limitations; lack of curiosity raises questions about "creativity" source; hallucinations: unconscious fabrication without knowledge source identification
Sartori & Orrù (2023)	Cognition & Behavior	Dependence on training data: bias limits performance on other tasks; lack of creativity/imagination; lack of autonomy for complex goals; poor multi-step reasoning; lack of self-understanding; limited real-world understanding
Goertzel (2023)	Cognition & Behavior	Forgetting: may forget previous knowledge when learning new tasks; inadequate common-sense reasoning; lack of systematic problem-solving demonstration
Peng et al. (2023)	Cognition & Behavior	Lack of clear model behavior understanding makes improvement difficult; lack of formal behavior description prevents systematic analysis and unified theory development
Holtzman et al. (2023)	Cognition & Behavior	Lack of model behavior interpretability makes it difficult to understand performance variations across tasks
Seals & Shalin (2023)	Cognition & Behavior	ChatGPT-human analogy differences in stylistic dimensions, lexical features, and comprehension devices; ChatGPT may lack human cognitive/psycholinguistic features
Stade et al. (2023)	Clinical & Counseling	Technical limitations: difficulty assessing suicide risk, substance abuse, safety, comorbidities, life events; connecting with patients: difficulty interpreting nonverbal behaviors; autonomy/therapeutic relationship problems

Author	Domain	Challenges and Limitations
Li et al. (2023)	Education & Development	Academic integrity and authorship definition; assessment methods and consequences; data privacy/security; teacher-student relationship; critical thinking skills; misinformation/bias; interpersonal communication skills
Kasneci et al. (2023)	Education & Development	Technical limitations: insufficient personalization/adaptation; bias/equity affecting processes/outcomes; over-reliance causing creativity/critical thinking decline; inadequate educator knowledge/expertise; maintenance costs
Fecher et al. (2023)	Society & Culture	Multilingual support and equitable access; liability issues challenging authorship mechanisms; bias affecting scientific objectivity; privacy/data protection; intellectual property disputes; environmental impact from carbon emissions
Atari et al. (2023)	Society & Culture	Ignoring global psychological diversity (WEIRD societies) leads to prejudice; value/moral judgment differences cause communication problems; self-identity issues lead to non-WEIRD population stereotypes
Park et al. (2023)	Society & Culture	Reduced innovation, bias/discrimination, culture clash, value differences, status quo entrenchment
Salah et al. (2023)	Society & Culture	Limited social context understanding: good syntax/semantics but poor social language nuance capture
Hayes (2023)	Society & Culture	Ethical challenges: AI-generated fake content raises digital personhood, informed consent, manipulation, and human interaction simulation issues
Miotto et al. (2022)	Society & Culture	Potential biases: training data biases lead to unfair results (e.g., reinforcing sexism in job ad translation)
Bender et al. (2021)	Society & Culture	Responsibility/control: complexity makes output responsibility difficult, leading to attribution problems and control lack
Tamkin et al. (2021)	Society & Culture	Potential harm: propagation of stereotypes, discrimination, extremism; misinformation and bullying

Author	Domain	Challenges and Limitations
Brown et al. (2020)	Society & Culture	Data bias/unfairness harms marginalized communities; automating bias exacerbates discrimination; authoritative viewpoint enhancement undermines marginalized people
Sallam (2023)	Research Tools	News generation: GPT-3 news difficult to distinguish from real news, causing confusion
Gupta et al. (2023)	Research Tools	Plagiarism: ChatGPT content may violate academic norms; copyright: unclear ownership; transparency: unclear workings; liability: responsibility for incorrect content
Dergaa et al. (2023)	Research Tools	Transparency/explanation: generative AI mechanisms difficult to explain, causing credibility doubts
Peters & Matz (2023)	Research Tools	Legal/ethical issues: intellectual property, privacy, compliance concerns
Y. Liu et al. (2023)	Research Tools	Integration of erroneous/biased information; citation problems; impact on academic integrity/quality; increased inequity; difficulty recognizing AI content; evaluation issues
Research tools	General	Direct replacement: ChatGPT not a complete replacement due to limitations in certain research types

7.2 Future Directions and Emergent Trends

LLMs have begun permeating different psychology areas, particularly cognitive/behavioral, clinical/counseling, educational/developmental, and social/cultural psychology. As capabilities advance, applications will continue developing.

First, in cognitive and behavioral psychology, multimodal LLMs (OpenAI, 2023) can combine visual/auditory information with text to better understand and model emotions, behaviors, and mental states. Neuroimaging data can inform LLM architectures and parameters, integrating with textual data to create more accurate, biologically-grounded models of human language and thought.

Second, in clinical and counseling psychology, personal data (social media posts, medical records, wearable device data) can create tailored LLMs providing more accurate mental state insights. Combining human clinical expertise with LLM scalability and computational power can create new diagnostic, treatment, and intervention tools.

Third, in educational/developmental and social/cultural psychology, building ethical LLMs requires ensuring design and deployment respect privacy and use data fairly and responsibly.

Ultimately, LLM development requires interdisciplinary collaboration among psychology, computer science, linguistics, and other fields. For psychology researchers, accessible open-source LLM frameworks and tools may become integral to future research. Table 7 summarizes future directions and emergent trends.

Table 7. Future Directions and Emergent Trends of LLMs in Psychological Applications

Author	Domain	Future Directions and Emergent Trends
D'Oria (2023)	Cognition & Behavior	Delve into Human-Computer Interaction to understand AI's ability to mimic human behavior; explore AI language modeling applications in human sciences to improve research efficiency and quality
Crockett & Messeri (2023)	Cognition & Behavior	Focus on costs of alternative human narratives: masking human labor behind them and impact on human well-being
Binz & Schulz (2023b)	Cognition & Behavior	Explore whether LLMs can learn to explore purposefully and better utilize causal knowledge; analyze performance across tasks/contexts for human-like adaptation; investigate cognitive ability development during natural human interaction
Huang & Chang (2022)	Cognition & Behavior	Improve reasoning by optimizing training data, architecture, and goals; develop better evaluation methods/benchmarks; explore potential in problem-solving, decision-making, planning; investigate other reasoning forms (inductive, retrospective)
Abd-Alrazaq et al. (2019)	Clinical & Counseling	Develop more chatbots for schizophrenia, OCD, bipolar disorder; implement chatbots in developing countries to address professional shortages; conduct more randomized controlled trials
Stade et al. (2023)	Clinical & Counseling	Focus on evidence-based practices first for short-term clinical impact; involve interdisciplinary collaboration; focus on therapist/patient trust and usability; design effective clinical LLM evaluation criteria
Demszky et al. (2023)	Clinical & Counseling	Develop new therapeutic techniques and evidence-based practices

Author	Domain	Future Directions and Emergent Trends
Hagendorff (2023)	Education & Development	Develop high-quality cornerstone datasets encompassing populations and constructs with psychologically important outcomes (behaviors, mindfulness, health, well-being); focus on consumer neuroscience and clinical neuroscience research
Sap et al. (2022)	Society & Culture	Developmental psychology: examine LLM cognitive, social, emotional development over lifespan; Learning psychology: study knowledge/skill acquisition and optimization
Schaaff et al. (2023)	Society & Culture	Explore interactive/empirical training methods for true social intelligence; combine static text with rich social interaction data
Argyle et al. (2022)	Society & Culture	Investigate GPT-3 algorithmic fidelity; create "in silico samples" by conditioning on socio-demographic backgrounds from large U.S. surveys
Ziems et al. (2023)	Society & Culture	Develop advanced models capturing emotional context; measure bot emotional capabilities; explore ChatGPT as empathetic support tool
Van Dis et al. (2023)	Research Tools	Invest in truly open LLMs for transparency and democratic control; embrace AI advantages while focusing on ethics and human autonomy; broaden international discussion on responsible LLM use, diversity, and inequality
Fecher et al. (2023)	Research Tools	Analyze LLM risks/opportunities for science systems; examine effects on quality assurance, misconduct, integrity; explore impact on reputation, evaluation, knowledge dissemination; balance benefits with scientific principles

8. Conclusion

With rapid AI technology development, particularly continuous advancement of LLMs like the GPT family, we have entered a new era characterized by unprecedented machine ability to understand and generate human language. This represents not just a technological breakthrough but opens doors to potential applications across psychology.

First, in cognitive and behavioral psychology, LLMs excel across various cognitive tasks. While limitations remain in causal cognition and planning, these models resurrect association principles, demonstrating distant association and complex reasoning abilities. LLMs' adaptability as cognitive models is a significant strength, enabling new explorations of human cognitive and behavioral processing mechanisms.

Second, in clinical and counseling psychology, LLMs can serve as preliminary diagnostic tools for mental health. While traditional diagnosis relies on professional experience and direct patient interaction, LLMs can quickly identify potential problems like depression and anxiety by analyzing verbal expressions and textual content. Such diagnosis cannot fully replace professional assessment but serves as an effective adjunct, helping psychologists understand patients more quickly or supporting primary interventions. Personalized psychological intervention is another critical application direction. By combining individual health data and lifestyle information, these models can provide tailored psychological advice and intervention programs, potentially crucial for improving intervention effectiveness.

Third, in educational/developmental and social/cultural psychology, LLMs show similar potential. They provide interactive, personalized learning experiences and generate research tasks based on real-life cases that increase motivation and enhance learning. Additionally, by analyzing large social media datasets, these models help researchers track and analyze public sentiment changes to better understand psychosocial dynamics.

Finally, in psychological research, LLMs can drastically improve efficiency. Researchers can quickly organize and analyze large literature volumes, saving time. These models also assist in experimental design, data analysis, and even paper writing, making psychological research more efficient and precise.

In summary, LLM applications in psychology are promising. From research aids to cognitive modeling, from individualized interventions to personalized learning, and from individual cognitive abilities to group social interactions, these models have potential to dramatically improve understanding of human communication patterns, thought processes, and behaviors, leading to more sophisticated theories of mind. However, despite great potential, vigilance regarding risks and challenges is essential. Ensuring these applications adhere to ethical standards is vital, especially for protecting individual privacy and data security. It is also important to recognize that no matter how technologically advanced, LLMs can only partially replace human professional judgment and experience. Therefore, these models should be viewed as aids rather than all-in-one solutions.

References

Abd-Alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., & Househ, M. (2019). An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics*, 132, 103978. https://doi.org/10.1016/j.ijmedinf.2019.103978

Abramski, K., Citraro, S., Lombardi, L., Rossetti, G., & Stella, M. (2023). Cognitive Network Science Reveals Bias in GPT-3, GPT-3.5 Turbo, and GPT-4 Mirroring Math Anxiety in High-School Students. *Big Data and Cognitive Computing*, 7(3). https://doi.org/10.3390/bdcc7030124

Agrawal, S. (2023). Are LLMs the Master of All Trades? : Exploring Domain-Agnostic Reasoning Skills of LLMs. arXiv preprint. https://doi.org/10.48550/arxiv.2303.12810

Aher, G., Arriaga, R. I., & Kalai, A. T. (2023). Using large language models to simulate multiple humans and replicate human subject studies. *Proceedings of the 40th International Conference on Machine Learning*, Honolulu, Hawaii, USA.

Akata, E., Schulz, L., Coda-Forno, J., Oh, S. J., Bethge, M., & Schulz, E. (2023). Playing repeated games with Large Language Models. arXiv preprint. https://doi.org/10.48550/arXiv.2305.16867

Ali, J. K. M., Shamsan, M. A. A., Hezam, T. A., & Mohammed, A. A. Q. (2023). Impact of ChatGPT on Learning Motivation. *Journal of English Studies in Arabia Felix*, 2(1), 41-49. https://doi.org/10.56540/jesaf.v2i1.51

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J., Rytting, C., & Wingate, D. (2022). Out of One, Many: Using Language Models to Simulate Human Samples. arXiv preprint. https://doi.org/10.48550/arXiv.2209.06899

Atari, M., Xue, M. J., Park, P. S., Blasi, D. E., & Henrich, J. (2023). Which Humans? PsyArXiv preprint. https://doi.org/10.31234/osf.io/5b26t

Aydın, Ö., & Karaarslan, E. (2022). OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare. *Emerging Computer Technologies*(2), 22-31. https://doi.org/10.2139/ssrn.4308687

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, Virtual Event, Canada. https://doi.org/10.1145/3442188.3445922

Binz, M., & Schulz, E. (2023a). Turning large language models into cognitive models. arXiv preprint. https://doi.org/10.48550/arXiv.2306.03917

Binz, M., & Schulz, E. (2023b). Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences of the United States of America*, 120(6), e2218523120. https://doi.org/10.1073/pnas.2218523120

Blyler, A. P., & Seligman, M. E. P. (2023a). AI assistance for coaches and therapists. *The Journal of Positive Psychology*, 1-13.

Blyler, A. P., & Seligman, M. E. P. (2023b). Personal narrative and stream of consciousness: an AI approach. *The Journal of Positive Psychology*, 1-7. https://doi.org/10.1080/17439760.2023.2257666

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., . . . Amodei, D. (2020). Language Models are Few-Shot Learners. arXiv preprint. https://doi.org/10.48550/arXiv.2005.14165

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E. K., Kamar, E., Lee, P., Lee, Y. T., Li, Y.-F., Lundberg, S. M., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4. arXiv preprint. https://doi.org/10.48550/arXiv.2303.12712

Cai, Z. G., Haslett, D. A., Duan, X., Wang, S., & Pickering, M. J. (2023). Does ChatGPT resemble humans in language use? arXiv preprint. https://doi.org/10.48550/arXiv.2303.08014

Carlbring, P., Hadjistavropoulos, H., Kleiboer, A., & Andersson, G. (2023). A new era in Internet interventions: The advent of Chat-GPT and AI-assisted therapist guidance. *Internet Interventions*, 32, 100621. https://doi.org/10.1016/j.invent.2023.100621

Chiang, C.-H., & Lee, H.-y. (2023). Can Large Language Models Be an Alternative to Human Evaluations? arXiv preprint. https://doi.org/10.48550/arXiv.2305.01937

Crockett, M., & Messeri, L. (2023). Should large language models replace human participants? PsyArXiv preprint. https://doi.org/10.31234/osf.io/4zdx9

D'Oria, M. (2023). Can AI Language Models Improve Human Sciences Research? A Phenomenological Analysis and Future Directions. *Encyclopaideia*, 27(66), 77-92. https://doi.org/10.6092/issn.1825-8670/16554

De Bot, K., Lowie, W., & Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10(1), 7-21. https://doi.org/10.1017/S1366728906002732

Demszky, D., Yang, D., Yeager, D. S., Bryan, C. J., Clapper, M., Chandhok, S., Eichstaedt, J. C., Hecht, C., Jamieson, J., Johnson, M., Jones, M., Krettek-Cobb, D., Lai, L., JonesMitchell, N., Ong, D. C., Dweck, C. S., Gross, J. J., & Pennebaker, J. W. (2023). Using large language models in psychology. *Nature Reviews Psychology*, 2(11), 688-701. https://doi.org/10.1038/s44159-023-00241-5

Dergaa, I., Chamari, K., Zmijewski, P., & Ben Saad, H. (2023). From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport*, 40(2), 615-622. https://doi.org/10.5114/biolsport.2023.125623

Dhingra, S., Singh, M., Sb, V., Malviya, N., & Singh Gill, S. (2023). Mind meets machine: Unravelling GPT-4's cognitive psychology. arXiv preprint, arXiv:2303.11436. https://doi.org/10.48550/arXiv.2303.11436

Dillion, D., Tandon, N., Gu, Y., & Gray, K. (2023). Can AI language models replace human participants? *Trends in Cognitive Sciences*, 27(7), 597-600. https://doi.org/10.1016/j.tics.2023.04.008

Elyoseph, Z., & Levkovich, I. (2023). Beyond human expertise- the promise and limitationsof ChatGPT in suicide risk assessment. *Frontiers in Psychiatry*, 14.

https://doi.org/10.3389/fpsyt.2023.1213141

Fecher, B., Hebing, M., Laufer, M., Pohle, J., & Sofsky, F. (2023). Friend or foe? Exploring the implications of large language models on the science system. *Ai & Society*. https://doi.org/10.1007/s00146-023-01791-1

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, 30(4), 681-694. https://doi.org/10.1007/s11023-020-09548-1

Frank, M. C. (2023). Baby steps in evaluating the capacities of large language models. *Nature Reviews Psychology*, 2(8), 451-452. https://doi.org/10.1038/s44159-023-00211-x

Glaser, R. (1984). Education and thinking: The role of knowledge. *American psychologist*, 39(2), 93.

Goertzel, B. (2023). Generative AI vs. AGI: The Cognitive Strengths and Weaknesses of Modern LLMs. arXiv preprint, arXiv:2309.10371. https://doi.org/10.48550/arXiv.2309.10371

Graber-Stiehl, I. (2023). IS THE WORLD READY FOR AI-POWERED THERAPY? *Nature*, 617, 22-24. https://doi.org/10.1038/d41586-023-01473-4

Grossmann, I., Feinberg, M., Parker, D. C., Christakis, N. A., Tetlock, P. E., & Cunningham, W. A. (2023). AI and the transformation of social science research. *Science*, 380(6650), 1108-1109. https://doi.org/10.1126/science.adi1778

Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy. *IEEE Access*, 11, 80218-80245. https://doi.org/10.1109/access.2023.3300381

Hagendorff, T. (2023). Machine Psychology: Investigating Emergent Capabilities and Behavior in Large Language Models Using Psychological Methods. arXiv preprint. https://doi.org/10.48550/arXiv.2303.13988

Hagendorff, T., Fabi, S., & Kosinski, M. (2022). Machine intuition: Uncovering human-like intuitive decision- making in GPT-3.5. arXiv preprint. https://doi.org/10.48550/arXiv.2212.05206

Hagendorff, T., Fabi, S., & Kosinski, M. (2023). Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in ChatGPT. *Nature Computational Science*, 3(10), 833-838. https://doi.org/10.1038/s43588-023-00527-x

Han, H. (2023). Potential Benefits of Employing Large Language Models in Research in Moral Education and Development. arXiv preprint. https://doi.org/10.48550/arXiv.2306.13805

Harding, J., D'Alessandro, W., Laskowski, N. G., & Long, R. (2023). AI language models cannot replace human research participants. *Ai & Society*. https://doi.org/10.1007/s00146-023-01725-x

Hardy, M., Sucholutsky, I., Thompson, B., & Griffiths, T. (2023). Large language models meet cognitive science: Llms as tools, models, and participants. *Proceedings of the annual meeting of the cognitive science society*,

Hayes, A. (2023). "Conversing" with Qualitative Data: Enhancing Qualitative Research through Large Language Models (LLMs). PsyArXiv preprint. https://doi.org/10.31235/osf.io/yms8p

Hendel, R., Geva, M., & Globerson, A. (2023). In-Context Learning Creates Task Vectors. arXiv preprint, arXiv:2310.15916. https://doi.org/10.48550/arXiv.2310.15916

Hofmann, S. G., Asnaani, A., Vonk, I. J., Sawyer, A. T., & Fang, A. (2012). The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses. *Cognit Ther Res*, 36(5), 427-440. https://doi.org/10.1007/s10608-012-9476-1

Holtzman, A., West, P., & Zettlemoyer, L. (2023). Generative Models as a Complex Systems Science: How can we make sense of large language model behavior? arXiv preprint. https://doi.org/10.48550/arXiv.2308.00189

Hothersall, D., & Lovett, B. J. (2022). *History of psychology.* Cambridge University Press.

Huang, J., & Chang, K. C.-C. (2022). Towards Reasoning in Large Language Models: A Survey. arXiv preprint. https://doi.org/10.48550/arXiv.2212.10403

Hutson, M. (2023). Doing research with human subjects is costly and cumbersome.Can AI chatbots replace them? *Science*, 381(6654), 121-123. https://doi.org/10.1126/science.adj6791

Jin, C., Zhang, S., Shu, T., & Cui, Z. (2023). The Cultural Psychology of Large Language Models: Is ChatGPT a Holistic or Analytic Thinker? arXiv preprint. https://doi.org/10.48550/arXiv.2308.14242

Jungherr, A. (2023). Using ChatGPT and Other Large Language Model (LLM) Applications for Academic Paper Assignments. https://fis.uni-bamberg.de/handle/uniba/58950 urn:nbn:de:bvb:473-irb-589507

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103. https://doi.org/10.1016/j.lindif.2023.102274

Ke, L., Zhang, G., He, J., Li, Y., Li, Y., Liu, X., & Fang, P. (2023). Pilot Selection in the Era of Virtual Reality: Algorithms for Accurate and Interpretable Machine Learning Models. *Aerospace*, 10(5). https://doi.org/10.3390/aerospace10050394

Kjell, O., Kjell, K., & Schwartz, H. A. (2023). Beyond Rating Scales: With Care for Validation Large Language Models Are Poised to Change Psychological Assessment. PsyArXiv preprint. https://doi.org/10.31234/osf.io/yfd8g

Kosinski, M. (2023). Theory of Mind May Have Spontaneously Emerged in Large Language Models. arXiv preprint. https://doi.org/10.48550/arXiv.2302.02083

Lamichhane, B. (2023). Evaluation of ChatGPT for NLP-based Mental Health Applications. arXiv preprint, arXiv:2303.15727. https://doi.org/10.48550/arXiv.2303.15727

Li, J., Tang, T., Zhao, W. X., Nie, J.-Y., & Wen, J.-R. (2022). Pretrained Language Models for Text Generation: A Survey. arXiv preprint, arXiv:2201.05273. https://doi.org/10.48550/arXiv.2201.05273

Li, M., Enkhtur, A., Cheng, F., & Yamamoto, B. A. (2023). Ethical implications of ChatGPT in higher education: A scoping review. arXiv preprint. https://doi.org/10.48550/arXiv.2311.14378

Li, X., Li, Y., Liu, L., Bing, L., & Joty, S. (2022). Does GPT-3 Demonstrate Psychopathy? Evaluating Large Language Models from a Psychological Perspective. arXiv preprint. https://doi.org/10.48550/arXiv.2212.10529

Liu, J. M., Li, D., Cao, H., Ren, T., Liao, Z., & Wu, J. (2023). ChatCounselor: A Large Language Models for Mental Health Support. arXiv preprint, arXiv:2309.15461. https://doi.org/10.48550/arXiv.2309.15461

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *Acm Computing Surveys*, 55(9), 1-35. https://doi.org/10.1145/3560815

Liu, X., Ji, K., Fu, Y., Tam, W., Du, Z., Yang, Z., & Tang, J. (2022). P-Tuning: Prompt Tuning Can Be Comparable to Fine-tuning Across Scales and Tasks. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Dublin, Ireland.

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., & Ge, B. (2023). Summary of ChatGPT-Related research and perspective towards the future of large language models. *Meta-Radiology*, 1(2). https://doi.org/10.1016/j.metrad.2023.100017

Loconte, R., Orrù, G., Tribastone, M., Pietrini, P., & Sartori, G. (2023). Challenging ChatGPT's "intelligence" with human tools: A Neuropsychological Investigation on Prefrontal Functioning of a Large Language Model. SSRN preprint. https://doi.org/10.2139/ssrn.4471829

Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2023). Dissociating language and thought in large language models: a cognitive perspective. arXiv preprint. https://doi.org/10.48550/arXiv.2301.06627

Marjieh, R., Sucholutsky, I., Rijn, P. v., Jacoby, N., & Griffiths, T. L. (2023). Large language models predict human sensory judgments across six modalities. arXiv preprint. https://doi.org/10.48550/arXiv.2302.01308

Miotto, M., Rossberg, N., & Kleinberg, B. (2022). Who is GPT-3? An Exploration of Personality, Values and Demographics. arXiv preprint. https://doi.org/10.48550/arXiv.2209.14338

Mitchell, M. (2023). AI's challenge of understanding the world. *Science*, 382(6671). https://doi.org/10.1126/science.adm8175

Nazario-Johnson, L., Zaki, H. A., & Tung, G. A. (2023). Use of large language models to predict neuroimaging. *Journal of the American College of Radiology*, 20(10), 1004-1009. https://doi.org/10.1016/j.jacr.2023.06.008

Newell, A. (1990). *Unified theories of cognition*. Harvard University Press.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: holistic versus analytic cognition. *Psychological Review*, 108(2), 291-310. https://doi.org/10.1037//0033-295X.108.2.291

OpenAI. (2023). GPT-4 Technical Report. arXiv preprint. https://doi.org/10.48550/arXiv.2303.08774

Orru, G., Piarulli, A., Conversano, C., & Gemignani, A. (2023). Human-like problem-solving abilities in large language models using ChatGPT. *Frontiers in Artificial Intelligence*, 6, 1199350. https://doi.org/10.3389/frai.2023.1199350

Pal, R., Garg, H., Patel, S., & Sethi, T. (2023). Bias Amplification in Intersectional Subpopulations for Clinical Phenotyping by Large Language Models. medRxiv preprint. https://doi.org/10.1101/2023.03.22.23287585

Park, B., & Judd, C. M. (2005). Rethinking the Link Between Categorization and Prejudice Within the Social Cognition Perspective. *Personality and Social Psychology Review*, 9(2), 108-130. https://doi.org/10.1207/s15327957pspr0902_2

Park, J. S., Popowski, L., Cai, C., Morris, M. R., Liang, P., & Bernstein, M. S. (2022). Social simulacra: Creating populated prototypes for social computing systems. *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*,

Park, P. S., Schoenegger, P., & Zhu, C. (2023). Diminished Diversity-of-Thought in a Standard Large Language Model. arXiv preprint. https://doi.org/10.48550/arXiv.2302.07267

Patel, S. C., & Fan, J. (2023). Identification and Description of Emotions by Current Large Language Models. bioRxiv preprint. https://doi.org/10.1101/2023.07.17.549421

Peng, Y., Han, J., Zhang, Z., Fan, L., Liu, T., Qi, S., Feng, X., Ma, Y., Wang, Y., & Zhu, S.-C. (2023). The Tong Test: Evaluating Artificial General Intelligence Through Dynamic Embodied Physical and Social Interactions. *Engineering*. https://doi.org/10.1016/j.eng.2023.07.006

Peters, H., & Matz, S. (2023). Large Language Models Can Infer Psychological Dispositions of Social Media Users. arXiv preprint. https://doi.org/10.48550/arXiv.2309.08631

Qureshi, R., Shaughnessy, D., Gill, K. A. R., Robinson, K. A., Li, T., & Agai, E. (2023). Are ChatGPT and large language models "the answer" to bringing us closer to systematic review automation? *Systematic Reviews*, 12(1), 72. https://doi.org/10.1186/s13643-023-02243-z

Rathje, S., Mirea, D.-M., Sucholutsky, I., Marjieh, R., Robertson, C., & Bavel, J. J. V. (2023). GPT is an effective tool for multilingual psychological text analysis. PsyArXiv preprint. https://doi.org/10.31234/osf.io/sekf5

Salah, M., Al Halbusi, H., & Abdelfattah, F. (2023). May the force of text data analysis be with you: Unleashing the power of generative AI for social psychology research. *Computers in Human Behavior: Artificial Humans*, 1(2). https://doi.org/10.1016/j.chbah.2023.100006

Sallam, M. (2023). ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare (Basel)*, 11(6). https://doi.org/10.3390/healthcare11060887

Sap, M., LeBras, R., Fried, D., & Choi, Y. (2022). Neural Theory-of-Mind? On the Limits of Social Intelligence in Large LMs. arXiv preprint. https://doi.org/10.48550/arXiv.2210.13312

Sartori, G., & Orrù, G. (2023). Language models and psychological sciences. *Frontiers in Psychology*, 14. https://doi.org/10.3389/fpsyg.2023.1279317

Schaaff, K., Reinig, C., & Schlippe, T. (2023). Exploring ChatGPT's Empathic Abilities. arXiv preprint. https://doi.org/10.48550/arXiv.2308.03527

Schueller, S. M., & Morris, R. R. (2023). Clinical science and practice in the age of large language models and generative artificial intelligence. *Journal of Consulting and Clinical Psychology*, 91(10), 559-561. https://doi.org/10.1037/ccp0000848

Seals, S. M., & Shalin, V. L. (2023). Long-form analogies generated by chatGPT lack human-like psycholinguistic properties. arXiv preprint. https://doi.org/10.48550/arxiv.2306.04537

Sejnowski, T. (2022). Large Language Models and the Reverse Turing Test. arXiv preprint. https://doi.org/10.48550/arxiv.2207.14382

Sha, H., Mu, Y., Jiang, Y., Chen, L., Xu, C., Luo, P., Eben Li, S., Tomizuka, M., Zhan, W., & Ding, M. (2023). LanguageMPC: Large Language Models as Decision Makers for Autonomous Driving. arXiv preprint, arXiv:2310.03026. https://doi.org/10.48550/arXiv.2310.03026

Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C., & Althoff, T. (2023). Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence*, 5(1), 46-57. https://doi.org/10.1038/s42256-022-00593-2

Shiffrin, R., & Mitchell, M. (2023). Probing the psychology of AI models. *Proceedings of the National Academy of Sciences of the United States of America*, 120(10), e2300963120. https://doi.org/10.1073/pnas.2300963120

Simon, H. A. (1979). Information Processing Models of Cognition. *Annual Review of Psychology*, 30(1), 363-396. https://doi.org/10.1146/annurev.ps.30.020179.002051

Stade, E. C., Stirman, S. W., Ungar, L., Boland, C. L., Schwartz, H. A., Yaden, D. B., Sedoc, J., Derubeis, R. J., Willer, R., & Eichstaedt, J. C. (2023). Large Language Models Could Change the Future of Behavioral Healthcare: A Proposal for Responsible Development and Evaluation. PsyArXiv preprint. https://doi.org/10.31234/osf.io/cuzvr

Stella, M., Hills, T. T., & Kenett, Y. N. (2023). Using cognitive psychology to understand GPT-like models needs to extend beyond human biases. *Proceedings of the National Academy of Sciences of the United States of America*, 120(43), e2312911120. https://doi.org/10.1073/pnas.2312911120

Stevenson, C., Smal, I., Baas, M., Grasman, R., & Maas, H. v. d. (2022). Putting GPT-3's Creativity to the (Alternative Uses) Test. arXiv preprint. https://doi.org/10.48550/arXiv.2206.08932

Stojanov, A. (2023). Learning with ChatGPT 3.5 as a more knowledgeable other: an autoethnographic study. *International Journal of Educational Technology in Higher Education*, 20(1). https://doi.org/10.1186/s41239-023-00404-7

Stokel-Walker, C. (2022). AI bot ChatGPT writes smart essays — should professors worry? *Nature*. https://doi.org/10.1038/d41586-022-04397-7

Suri, G., Slater, L. R., Ziaee, A., & Nguyen, M. (2023). Do Large Language Models Show Decision Heuristics Similar to Humans? A Case Study Using GPT-3.5. arXiv preprint. https://doi.org/10.48550/arXiv.2305.04400

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, 33(1), 1-39.

Talboy, A. N., & Fuller, E. (2023). Challenging the appearance of machine intelligence: Cognitive bias in LLMs. arXiv preprint. https://doi.org/10.48550/arXiv.2304.01358

Tamkin, A., Brundage, M., Clark, J., & Ganguli, D. (2021). Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models. arXiv preprint. https://doi.org/10.48550/arXiv.2102.02503

Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. *Nature Medicine*, 29(8), 1930-1940. https://doi.org/10.1038/s41591-023-

Trott, S., Jones, C., Chang, T., Michaelov, J., & Bergen, B. (2023). Do Large Language Models know what humans know? arXiv preprint. https://doi.org/10.48550/arXiv.2209.01515

Van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature*, 614(7947), 224-226. https://doi.org/10.1038/d41586-023-00288-7

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P., Deac, A., Anandkumar, A., Bergen, K., Gomes, C. P., Ho, S., Kohli, P., Lasenby, J., Leskovec, J., Liu, T. Y., Manrai, A., . . . Zitnik, M. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972), 47-60. https://doi.org/10.1038/s41586-023-06221-2

Wang, X., Li, X., Yin, Z., Wu, Y., & Liu, J. (2023). Emotional intelligence of Large Language Models. *Journal of Pacific Rim Psychology*, 17. https://doi.org/10.1177/18344909231213958

Webb, T., Holyoak, K. J., & Lu, H. (2023). Emergent analogical reasoning in large language models. *Nature Human Behaviour*, 7(9), 1526-1541. https://doi.org/10.1038/s41562-023-01659-w

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., & Fedus, W. (2022). Emergent Abilities of Large Language Models. arXiv preprint. https://doi.org/10.48550/arXiv.2206.07682

Yildirim, I., & Paul, L. A. (2023). From task structures to world models: What do LLMs know? arXiv preprint, arXiv:2310.04276. https://doi.org/10.48550/arXiv.2310.04276

Yukun, Z., Xu, L., Huang, Z., Peng, K., Seligman, M., Li, E., & Yu, F. (2023). AI chatbot responds to emotional cuing. PsyArXiv preprint. https://doi.org/10.31234/osf.io/9ymfz

Zhang, J., Xu, X., & Deng, S. (2023). Exploring Collaboration Mechanisms for LLM Agents: A Social Psychology View. arXiv preprint, arXiv:2310.02124. https://doi.org/10.48550/arXiv.2310.02124

Zhong, Y., Chen, Y. J., Zhou, Y., Lyu, Y. A., Yin, J. J., & Gao, Y. J. (2023). The Artificial intelligence large language models and neuropsychiatry practice and research ethic. *Asian Journal of Psychiatry*, 84, 103577. https://doi.org/10.1016/j.ajp.2023.103577

Zhuang, Y., Liu, Q., Ning, Y., Huang, W., Lv, R., Huang, Z., Zhao, G., Zhang, Z., Mao, Q., Wang, S., & Chen, E. (2023). Efficiently Measuring the Cognitive Ability of LLMs: An Adaptive Testing Perspective. arXiv preprint. https://doi.org/10.48550/arXiv.2306.10512

Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., & Yang, D. (2023). Can Large Language Models Transform Computational Social Science? arXiv preprint. https://doi.org/10.48550/arXiv.2305.03514

Note: Figure translations are in progress. See original paper for figures.