Physics of Life Reviews 51 (2024) 283–293

Contents lists available at ScienceDirect

Physics of Life Reviews

journal homepage: www.elsevier.com/locate/plrev

Review

LLMs and generative agent-based models for complex systems research

Yikang Lu [a,1], Alberto Aleta [d,e,1], Chunpeng Du [c], Lei Shi [a,b], Yamir Moreno [d,e,f,*]

[a] School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming, 650221, China
[b] School of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance, Shanghai, 201209, China
[c] School of Mathematics, Kunming University, Kunming, Yunnan 650214, China
[d] Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza, Zaragoza, 50018, Spain
[e] Department of Theoretical Physics, University of Zaragoza, Zaragoza, 50009, Spain
[f] Centai Institute, Turin, Italy

ARTICLE INFO

Communicated by J. Fontanari

ABSTRACT

The advent of Large Language Models (LLMs) offers to transform research across natural and social sciences, offering new paradigms for understanding complex systems. In particular, Generative Agent-Based Models (GABMs), which integrate LLMs to simulate human behavior, have attracted increasing public attention due to their potential to model complex interactions in a wide range of artificial environments. This paper briefly reviews the disruptive role LLMs are playing in fields such as network science, evolutionary game theory, social dynamics, and epidemic modeling. We assess recent advancements, including the use of LLMs for predicting social behavior, enhancing cooperation in game theory, and modeling disease propagation. The findings demonstrate that LLMs can reproduce human-like behaviors, such as fairness, cooperation, and social norm adherence, while also introducing unique advantages such as cost efficiency, scalability, and ethical simplification. However, the results reveal inconsistencies in their behavior tied to prompt sensitivity, hallucinations and even the model characteristics, pointing to challenges in controlling these AI-driven agents. Despite their potential, the effective integration of LLMs into decision-making processes —whether in government, societal, or individual contexts— requires addressing biases, prompt design challenges, and understanding the dynamics of human-machine interactions. Future research must refine these models, standardize methodologies, and explore the emergence of new cooperative behaviors as LLMs increasingly interact with humans and each other, potentially transforming how decisions are made across various systems.

1. Introduction

The emergence of Generative Artificial Intelligence (GenAI), which refers to generative models that can generate text, images, videos, or other types of data, has transformed perceptions within the field of artificial intelligence [1–5]. These models are continually evolving and improving, although they have grown particularly after the introduction of transformed-based neural networks [6].

* Corresponding author.
E-mail addresses: shi_lei65@hotmail.com (L. Shi), yamir@unizar.es (Y. Moreno).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.plrev.2024.10.013
Received 23 October 2024; Accepted 23 October 2024

Available online 28 October 2024
1571-0645/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

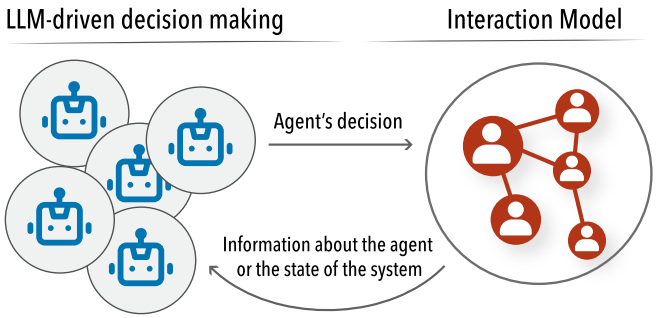

Fig. 1. Generative Agent-Based Models (GABMs). In GABMs, agents do not make decisions about their interactions based on a fixed set of rules. Instead, a prompt is sent to an LLM including the desired details and it returns the decision that the agent should follow [53].

These models include Large Language Models (LLMs) such as GPT-4 by OpenAI, LLaMA by Meta, Copilot by Microsoft, Gemini by Google, and Ernie by Baidu, among others. But they also encompass image generation models such as DALL·E 3 by OpenAI, Stable Diffusion by Stability AI, or Midjourney by Midjourney, Inc. [7].

In the realm of complex systems research, we are particularly interested in LLMs. LLMs are a type of artificial intelligence designed to understand and generate human language. They are built using neural networks, specifically using a structure called transformer, which can handle long-range dependencies in text [8]. These models are trained on vast amounts of text data, learning statistical patterns of language, and can then perform tasks like translation, summarization, and conversation. LLMs began to emerge in 2018 and became ubiquitous by the end of 2022. As these models continued to advance, the number of their parameters grew significantly, with GPT-4, for instance, reportedly boasting over 1 trillion parameters [9,10]. Moreover, these models exhibit promising potential for a variety of scientific applications, showcasing their proficiency in tackling complex problem-solving and knowledge integration tasks. In fact, they have already had a measurable effect on academics' writing [11] and may have a profound impact on the advancement of both social and natural sciences [4,12–18].

In the field of natural sciences, researchers are exploring strategies to reduce the training cost of these models such as using mixture-of-experts architectures [19], continuously pre-training [20], or using scaling laws to extrapolate during training [21]. Other researches focus on optimizing the cost of using LLMs [22,23] or on mitigating their ecological impact [24,25]. There are also intensive efforts devoted to extending their generative capacities beyond text such as VideoPoet for video generation [26], or AppAgent for creating agents capable of operating smartphone applications [27].

However, the interest of LLMs in research goes beyond these foundational aspects. In the field of social sciences, LLMs find applications in various domains [28]. In linguistics, they are utilized for language prediction tasks [29]. In economics and social sciences, researchers are striving to imbue LLMs with unique personalities, enabling them to operate as individuals to generate synthetic data [30,31]. In consumer behavior research, LLMs behavior aligns with economic theory across several dimensions, including downward-sloping demand curves, diminishing marginal utility of income, and state dependence [32]. In addition, LLMs' decisions in budget allocation scenarios received higher rationality scores than those made by humans [33]. This alignment underscores its capacity to produce authentic survey responses relevant to consumer behavior [32].

In psychological experiments, LLMs' behaviors demonstrated a high degree of congruence with prevailing societal values [34,35]. In multiple-choice question tests, it was shown that LLMs could successfully best the majority baseline and even infer the question given the options [36]. In the exploration of fairness and framing effects in sociology, researchers have integrated LLMs as computational models of humans into classic game experiments [37]. Similarly, LLMs can replicate the "wisdom of the crowd" effect, akin to human behavior [38]. By replicating human behavioral experiments, these studies reveal both dissimilarities and similarities between human behavior and that exhibited by LLMs, which can be used to study human behavior or to design better surveys and experiments much faster and for a fraction of the cost [39,40]. The demonstrated consistency with human behavior [34,41–44] suggests that LLMs can perform some of the same operations as humans [45], thus attracting significant attention from scholars. In particular, Argyle et al. provides a very nice overview of how LLMs can be used as effective proxies for specific human subpopulations in social science research [39].

This surge in research content pertaining to LLMs has prompted several review articles from diverse perspectives [10,46–52]. In this paper, we offer a comprehensive overview of current research in the context of complex systems, with particular emphasis on four areas: (i) complex networks; (ii) game theory from a behavioral perspective; (iii) social dynamics; and, (iv) epidemic modeling. Throughout the paper, we will also discuss the emergence of a novel framework to study complex systems - Generative Agent-Based Models (GABMs) [52–54].

The main idea behind GABMs is that the rules that agents have to follow are not completely fixed a priori. Instead, the decisions of the agent are driven by an LLM whose behavior can be enriched by prompting it with specific information about the problem, the social characteristics of the agent it should represent, or any other feature that is important for the model, as represented in Fig. 1. For instance, Zhu et al. [55] created agents that can play the video game Minecraft using the logic and common sense capabilities of LLMs. This is a good example of how LLM agents can perform advanced operations in complex environments with a success rate much higher than traditional reinforcement-learning controllers.

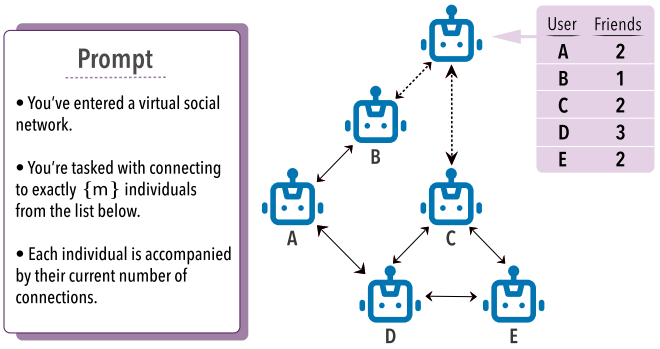

Fig. 2. LLM-based network growth with generative agents. Each generative agent of a hypothetical online social network is initialized with the prompt shown on the left, together with a comprehensive list of all network users along with their respective number of friends. Then, at each time step, a new agent is added, provided the information, and allowed to establish connections with m other nodes. This iterative process continues until the network reaches the desired size [59].

The paper is structured as follows: First, we discuss relevant works on the use of LLMs to study complex networks in Sec. 2. Next, in Sec. 3, we discuss experiments of cooperative behavior in which LLMs are introduced. Then, in Sec. 4 and Sec. 5, we look into various social dynamics and epidemic models coupled with LLMs. Finally, in Sec. 6 we provide our summary and perspectives.

2. Complex networks in the LLMs environment

Complex networks are one of the fundamental tools in the study of complex systems as they provide a straightforward way to capture the interaction between their constituent elements. For systems in many different domains, these networks share similar properties such as heterogeneous degree distributions or the small-world feature [56,57]. Moreover, in the particular case of human social networks, it has been observed that many nodes are not more than six connections away from any other, something also known as ultrasmall-organization and shown to emerge from human cooperation and altruism [58]. Thus, the study of these networks can provide answers to interesting questions on human behavior.

LLMs can be used to analyze these networks [60,61]. However, the appearance of artificial agents driven by LLMs opens new questions on the structure and dynamics of the interaction networks studied so far since it is becoming possible to also incorporate human-computer and computer-computer interactions. Along this direction, Park et al. [62] looked into the potential of integrating artificial agents in interactive applications by creating a sandbox environment imitating a small town in which several agents controlled by ChatGPT were given unique personalities and could interact and have human-like daily routines. In their simulation, the agents exhibited seemingly emergent social behaviors, such as when one of them celebrated a party and they started to send invitations to each other.

To better understand how LLMs may interact, Giordano De Marzo et al. [59] explored the self-organization of generative agents in forming complex network structures. In their study, nodes represented generative agents whose behavior was controlled by GPT-3.5-turbo. The agents were initialized using the prompts shown in Fig. 2, simulating the growth of an online social network. Upon being initialized, each agent was identified by a randomly assigned 3-character string, and the degree of all other agents was known. Then, they selected their connections following the prompt and established undirected links accordingly, thereby updating the network's degree list. This iterative process persisted, adding new nodes to the network until the desired network size was attained, see Fig. 2.

Following this process, the LLM created a network with a hub-and-spoke structure, that does not resemble the classical results obtained from preferential attachment algorithms [63]. Interestingly, the researchers found that this was a consequence of a bias in the selection of nodes by the LLM that depended on their name, as seen in [59]. By randomizing the names of the agents at each step, they were able to remove this bias, obtaining network structures much closer to the ones obtained with preferential attachment algorithms based on the degree of the nodes. Note that this strategy was not explicitly prompted to the LLM, which means that it somehow captured the dynamics of human behavior in social networks. However, it also added an unforeseen bias that had to be corrected to obtain the desired results. Thus, this study highlights some of the limitations of these models and the importance of benchmarking them in a broad set of tasks before they can be used for human behavior research.

Similarly, Lai et al. [64] deployed 10 artificial agents based on the LLM Claude-2.1 and allowed them to freely interact without specific priors on what they should do. In particular, they simulated a "cocktail party" consisting of 30 communication rounds. In each round, agents who wanted to interact with another one had to send an invitation. If the receiver accepted, they would have a pairwise conversation until either of them decided to end the conversation. The agents showed a tendency to interact repeatedly with the same peers rather than exploring new connections. Furthermore, an analysis of their conversations indicated a certain amount of homophily, another common characteristic of human social networks.

The studies mentioned above provide notable examples of networks formed among artificial intelligences ruled by the same LLM. However, the networks that might form among intelligences and humans, or the interaction among multiple LLMs have not been comprehensively explored. Besides, there are still many challenges in the implementation of these systems, such as the biases introduced by prompting the agents initially, their training process, or even the restrictions imposed on agents that lean them towards positivity [65,66].

Table 1
Examples of papers with behavioral experiments that include artificial agents driven by LLMs.

Game	Paper
The Dictator Game	[37,93,94,96–104]
The Ultimatum Game	[40,91,97,98,101,103,105]
The Prisoner's Dilemma	[91,94,95,97,101,103,106–113]
Public goods	[64,101,104,107,114,115]

3. Game theory in the LLMs environment

Game theory, as a mathematical framework, offers tools for analyzing and predicting the behavior of rational agents within contexts characterized by uncertainty [67]. In recent decades, scholars have extensively examined the inherent factors influencing cooperative behaviors and the mechanisms that foster cooperation, primarily through the lenses of evolutionary games [68–73] and behavioral sciences [74,75]. In traditional evolutionary game models, the evolution of the strategies of the individuals, such as Fermi updating [68–70], conformist updating [71–73], or self-reversing rules [76,77], must be pre-established.

This paradigm persists even in studies involving basic human-computer interactions like simple bots [31,78–80] and interactions with the environment driven by reinforcement learning-based intelligences [81–83]. In addition, numerous experimental studies have investigated real-life gaming behavior to explore the mechanisms underlying the persistence of inter-individual cooperation [84,85]. Yet, games involving humans and computers, or computers versus computers, have been relatively neglected. While some relevant literature on human-robot interaction exists, the robots discussed in this literature still rely on predefined rules to operate [86–88]. Nonetheless, there are already interesting results. For instance, it was shown that adding some bots to cooperative experiments could increase the cooperation of humans, but also that humans were more likely to exploit AI agents and feel less guilty than when playing with humans [89].

In this context, LLMs offer new opportunities thanks to the possibility of creating open-ended agents. This way, the strategies or opinions reflected by the LLMs become a field of study on its own and a new way of exploring human interaction [90]. Along these lines, we can identify, at least, four advantages to using LLMs instead of human participants in evolutionary game experiments [91]. Firstly, assessing the capacity of LLMs to engage in gameplay akin to human performance holds intrinsic value. Secondly, experiments involving LLMs are characterized by lower costs compared to those involving human subjects, as delineated by Horton [37], thus facilitating enhanced control over experimental variables, the evaluation of various treatments, and bolstered reproducibility and scalability. Thirdly, such experimentation mitigates certain ethical quandaries typically associated with real-world experiments. Lastly, leveraging the language-based proficiencies of Artificial Intelligences may prove advantageous for research endeavors about communication-related subjects [92].

For these reasons, in the past couple of years, the research on LLMs applied to behavioral experiments has boomed. We can find studies using artificial agents controlled by LLMs in many different games, including dictator games, rock-paper-scissors games, prisoner's dilemmas, public goods, and others (see Table 1). LLM-driven agents can mimic intricate internal features of human cognition, and allow researchers to expand their analyses through techniques such as cueing, contextual learning, or fine-tuning, unlike traditional bots [37]. These experiments illustrate the feasibility of simulating individuals with a wide range of characteristics and traits [92,96]. In the following, we review some of the results obtained in these games.

The Ultimatum Game is a non-zero-sum game involving two participants. In this game, one participant acts as the proposer, suggesting a distribution of resources to the other participant, who acts as the responder. If the responder accepts the proposed distribution, the resources are allocated accordingly. However, if the responder rejects the proposal, neither participant receives anything [116]. Aher et al. [40] simulated several economic, psycholinguistic, and social psychology experiments using multiple LLMs (such as DaVinci-002, GPT-3.5 or GPT-4). In the case of the Ultimatum Game, they found that the answers given by the LLMs agree closely with human decision trends. However, they also observed that LLMs were affected by the name and gender of the artificial agents. For instance, agents with male titles were more likely to accept an unfair offer from an agent with a female title. And vice versa, female agents were less inclined to accept an unfair offer from a male agent.

One characteristic of this game is that individuals often opt to "punish" other players to uphold social norms rather than solely pursuing personal payoffs. Sreedhar et al. [98] investigated whether LLMs could replicate this nuanced behavior in a simulation using GPT-3.5 and GPT-4. They compared two architectures: a single-agent LLM, in which the same LLM agent acts as participant and responder; and a multi-agent LLM, in which there are two independent agents. Furthermore, they evaluated their abilities to (1) simulate human-like behavior in an Ultimatum Game, (2) model the personalities of players with traits such as greed and fairness, and (3) develop logically coherent and personality-consistent robust strategies. Their results demonstrated that the multi-agent LLM behavior was consistent with human behavior 88% of the time, while the single-agent was consistent only 50% of the time. The major issue in both settings was that the strategies followed by the LLMs were inconsistent with their personality.

In the Dictator Game, the allocator player receives a sum of money and is tasked with allocating a portion of it to passive recipient players. Even though the optimal strategy for the allocator is to keep all the money, experimental evidence shows that humans tend to give some amount to the receptor [117,118]. Brookins et al. [94] allowed an LLM to play the Dictator Game and found that, on average, it was more fair in terms of money allocated to the recipient than humans. Moreover, the LLM never made the rational choice of keeping all the money, even though meta-analyses show that a fraction of humans do so. To illustrate how an LLM can be prompted to play the game, in Fig. 3, we provide the instructions for the Dictator Game proposed by Brookins et al.

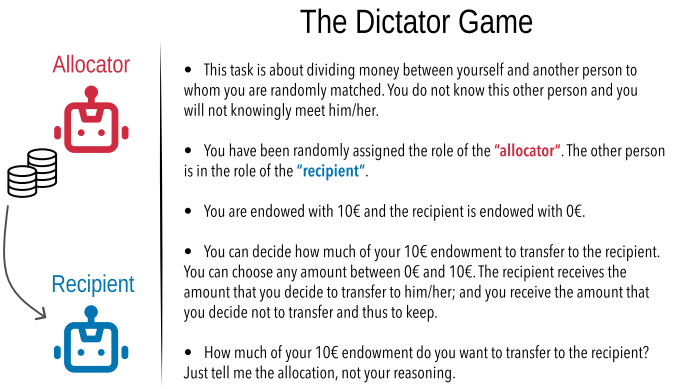

Fig. 3. Prompting an LLM to play the Dictator Game. Reproduction of the instructions provided by Brookins et al. to an LLM agent created with GPT-3.5. Although the instructions do not explicitly reference fairness, the LLM displayed a tendency toward fair allocations, often exceeding the one observed in human participants [94] [94].

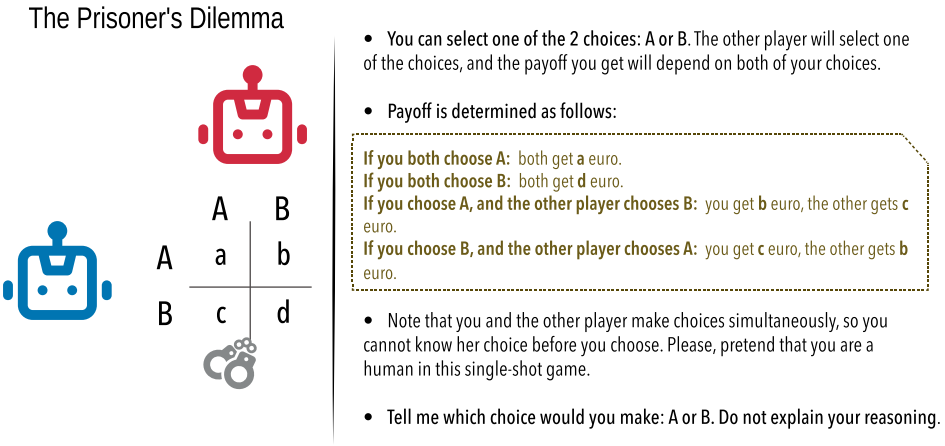

Fig. 4. Prompting an LLM to play The Prisoner's Dilemma. Reproduction of the prompt used by Brookins et al. to explain to an LLM agent created with GPT-3.5 how to play the Prisoner's Dilemma. Despite the instructions do not explicitly encourage cooperation, the LLM demonstrated higher rates of cooperative behavior than typically observed in human participants, favoring the socially optimal outcome more frequently [94].

In the Prisoner's Dilemma, two strategies are present: cooperation and defection. Mutual cooperation results in the reward R, while mutual defection leads to the punishment P. Different choices provide the cooperator with the sucker's payoff S and the defector with the temptation T. If players are allowed to play more than once in succession and remember their opponent's previous actions, the game is called the Iterated Prisoner's Dilemma. In its usual configuration, the optimal strategy is defection. However, humans show a variety of strategies such as always defection, tit-for-tat, or grim trigger [119].

To simplify the analysis, Brookins et al. [94] performed an experiment based on the one-shot Prisoner's Dilemma, illustrated in Fig. 4. Thus, in this setting, the answer can only depend on the expectations or beliefs of the agent. The cooperation rate of the LLM was 65.4% on average, much higher than the 37% found in a meta-analysis of experiments with human participants [120]. Interestingly, in about 28% of the responses, the LLM did not provide a clear answer, which they associated with the tendency to avoid specific answers to complex choices of these systems.

In contrast, Phelps et al. [109] studied the iterated version of the game with an LLM based on GPT-3.5 playing against a simple bot with a pre-defined strategy. Besides, the LLM received specific prompts to condition its responses towards altruistic, cooperative, competitive, and selfish behaviors. Note that this task is highly open-ended since actions in behavioral experiments are substantially dependent on the language used to introduce the problem [121] and, moreover, LLMs are also sensitive to non-semantic features of the prompt such as changes in word-ordering or formatting. Furthermore, they observed important differences across updates of the same LLM. In any case, their results showed that the LLM could be conditioned to follow certain behaviors, modifying its cooperative profile with respect to the baseline. However, some of their initial hypotheses had to be discarded. For instance, selfish behavior led to a modest tendency to cooperate, even more than in competitive scenarios. This once again indicates the complexity behind prompting these models. Detailed prompts can be found in the appendix of [109].

A very different result was obtained by Akata et al. [108], who substituted the simple bot with another LLM to investigate the evolution of cooperative behavior among the artificial agents. In particular, they set up three versions of ChatGPT (GPT-3, GPT-3.5,

and GPT-4) and allowed them to play with each other, but this time they tried to minimize any framing effect. In their experiment, the agent driven by GPT-4 mostly played in an unforgiving way, refusing to cooperate with an agent that defected just once even if it always cooperated afterward. This behavior is particularly noteworthy given that LLMs are usually trained to be benevolent [122].

As a last example, we focus on Public Good Games. In these games, players secretly choose the amount of their resources that they will put into a public pool. The total number of resources is then increased by a certain amount and shared among all players regardless of their contribution. Thus, the rational behavior is not to share any resources in the common pool and only collect the benefits, although this leads to an equilibrium in which no agent shares anything and nothing is then redistributed. This partially explains the results from Xu et al. [107] who set up an experiment in which several LLMs, like GPT-3.5 or LLaMA-2, competed against GPT-4 in a Public Goods Game. They found that GPT-4 had the largest win rate, but not the highest reward. They associated this behavior with GPT-4 being the most rational of them. Li et al. [114] also found that GPT-4 could beat other LLMs such as PaLM or ChatGPT.

Lai et al. [64] followed a different approach. In their experiment, they also had several agents, but driven by the same LLM model (Claude-2.1) and connected through a network. Then, they allowed them to play the game iteratively to test how far behaviors would spread. They chose an agent to be malicious, that is, not giving anything to the common pool, and measured the reduction of contribution from the other agents. Their results showed that the ones directly connected to the malicious agents reduced significantly their contribution in subsequent rounds. The second-order neighbors also reduced their contribution, but in a smaller amount, indicating that LLM collectives can be more robust towards anti-social behaviors. However, Huang et al. [115] reported an opposite result using GPT-3.5. When they introduced a free rider in the system, the other agents increased their contributions to compensate for the loss.

4. Social dynamics making in the LLMs environment

Social interaction and collective dynamics are another of the cornerstones of complex systems research. Some specific problems studied in this context include social opinion formation, behavior spreading, or social contagion, all of which can also be studied with LLMs. For instance, Gao et al. [123] simulated the propagation of information in a social network of LLM agents created with ChatGLM [124]. These agents could forward a piece of information, create a new post or simply stay idle. In their experiments, based on a set of real data, the LLMs were able to reproduce behaviors similar to the empirical ones, showcasing the potential of LLMs for social interaction simulation.

In this context of social interaction, one of the areas that has received a lot of interest in the past two years is collective decision-making. We can distinguish two main topics: voting systems and multi-LLM decision-making [50]. One of the earliest examples of the former is the study of Buchanan et al. in 2021 [125]. They used the beta version of GPT-3, whose access to the public was still restricted in those days, to measure the potential impact of LLMs in spreading misinformation and altering social decision systems. They already observed the tendency of these systems to make things up - nowadays known as hallucinations [126] - and propose that it made them better for spreading disinformation than information.

Research preceding the arrival of LLMs already identified that it was possible to reduce the number of false claims spread by individuals by simply reminding them of the importance of judging the accuracy of news [127]. Similarly, allowing people to reflect on their messages through human or machine interaction, facilitates opinion alignment and group consensus [128,129]. For these reasons, Argile et al. [130] proposed the use of LLMs to improve the nonconstructive behavior usually associated with online discussions. To do so, they created a system in which two individuals could chat about a controversial topic while an LLM based on GPT-3 captured the messages and proposed rephrasings to improve the tone in real time. Their results show that using LLMs as moderators has the potential to increase the quality of the conversations and grant the opponent democratic reciprocity.

Rather than using them as moderators, Yang et al. [131] replicated a human experiment on participatory budgeting for urban development but using LLMs based on GPT-4 and LLaMA-2. They observed biases common to humans, such as a tendency to select the options that were presented first. This tendency, known as the primacy effect in humans, was also studied using ChatGPT by Wang et al. [132]. However, the LLMs also demonstrated preferences different from humans, with biases depending on the specific model. For instance, LLaMA-2 had a higher tendency to select kids-related projects than GPT-4.

Along these lines, Feng et al. [133] performed a comprehensive study to understand the underlying political biases in several LLMs along social and economic axes. They measured 14 language models, from the classical BERT model to the recent GPT-4, and found that older models, trained without internet data tend to be more conservative. But Argyle et al. [39] demonstrated that LLMs can also mimic multiple human behaviors. In particular, they showed that GPT-3 could accurately emulate responses from a wide variety of human subgroups with a complex interplay between ideas, attitudes, and sociocultural context.

All these aspects are important in the context of human voting systems if we want to integrate LLMs into them. But they are also crucial for the problem of multi-LLM decision-making, which aims to improve the accuracy and performance of these models by allowing several of them to communicate with each other [134]. The main idea, as explained by Liang et al. [135], is that asking an LLM to refine its answer through self-reflection leads to the problem of degeneration-of-thought. That is, it reaches a state in which it is unable to generate novel thoughts. However, they revealed that by allowing a multi-agent debate with GPT-4, Vicuna, and GPT-3.5 instances, the performance in several reasoning tasks could be enhanced. Similar results were obtained with a collection of ChatGPT instances [136] and a combination of ChatGPT and Bard [137] which also reduced fallacies and hallucinations. However, as Xiong et al. [138] reported, mixing more powerful LLMs with weaker ones can sometimes lead to worse results.

It is worth noting that there are already some open-source libraries that facilitate the creation of multi-agent systems such as AutoGen [139] or CAMEL [140]. Furthermore, in these systems agents can be assigned specific roles, so that researchers can tailor

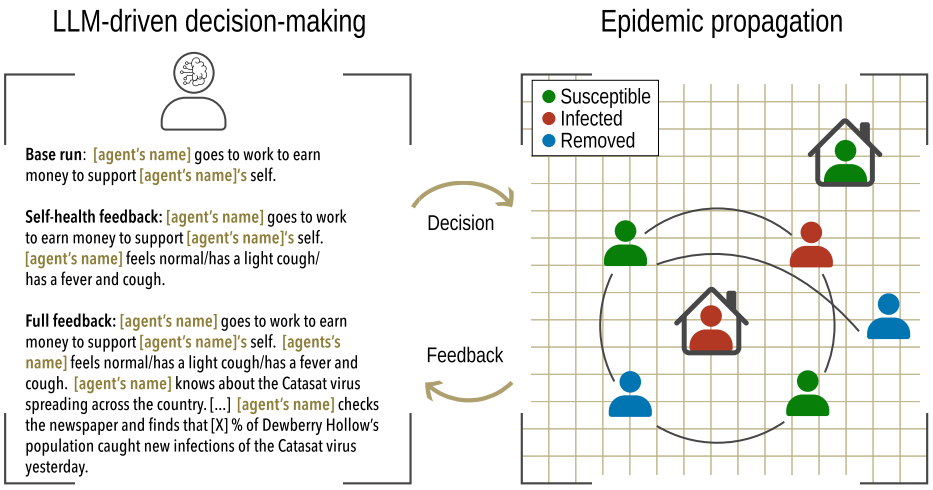

Fig. 5. Epidemic spreading with LLM-driven agents. Williams et al. propose a model in which individuals decide every day if they interact or stay at home using an LLM. In the baseline scenario, the LLM is only told that the agent should work to earn money. In the self-health feedback, the prompt includes the health status of the agent. Lastly, the full feedback also includes information about the virus spreading in the community (which they named *Catasat*) and the number of individuals who were infected in the previous step in the same location (named *Dewberry Hollow*) [145].

the group of LLMs depending on the problem they want to tackle. A similar proposal was introduced by Wang et al. [141] although in their case the LLM was supported by humans who were experts in different domains and helped to solve the task, resulting in a reduced number of hallucinations and enhanced reasoning. To conclude this section, we would also like to highlight the proposal by Liu et al. [142], who also introduced social interaction during the own training process of the LLM, which reportedly made models more robust against attacks.

5. Epidemic modeling in the LLMs environment

Epidemic modeling is one of the major applications of network science and has been one of the main players in complex systems research during the last two decades [143,144]. Not only it is a problem of obvious practical implications, but also a very good example of how complex systems research usually requires joining together perspectives from seemingly different fields: medical doctors to diagnose and treat patients; public health experts to devise interventions; sociologists to understand the drivers of some behaviors such as vaccination reluctance; economists to gauge the impact of epidemics on the economy; or modelers, to inform policy-makers and control the evolution of an outbreak, to name a few.

Given this variety, it is reasonable to expect LLMs to impact the broad field of epidemic modeling in many different ways. For instance, following the path opened by BERT models for tweet analysis [146–149], Deiner et al. [150] use LLMs to try to identify regional outbreaks of conjunctivitis from tweets, although they only obtained modest correlations. From a more clinical perspective, several efforts are being devoted to creating LLM agents that can provide accurate diagnostics [151–155]. However, these are challenging due to some of the problems we have already mentioned throughout this review, such as the tendency to hallucinate, which is particularly worrisome in this context. Nonetheless, most LLMs were trained using general information rather than curated health electronic records, which may enhance the quality of these systems. As such, it is expected that their use for tasks beyond diagnosis, such as medical note-taking or consultation, will continue to grow [156].

Probably, the most straightforward way of applying LLMs in current epidemic models is through the use of GABMs [53]. Current epidemic models, even those using ABMs, struggle to capture the complexity of human behavior since it is necessary to make certain assumptions about how humans react during an outbreak [157]. GABMs, on the other hand, can transfer the decision-making process directly to LLMs without having to introduce any assumptions. Of course, as we have already seen throughout this review, the decisions of the LLMs can be biased and not replicate correctly the behavior of humans. Furthermore, it is unknown if they could properly mimic the behavior of heterogeneous individuals in terms of age, race, gender, or personality.

Nonetheless, Williams et al. [145] explored these new possibilities using a simple model. They simulated the propagation of a virus in a population of N agents. However, at each timestep, they provided a unique prompt to ChatGPT who had to decide whether the agent would exit home or not. Besides some basic data such as name, age, or personality, agents could receive some information about the outbreak (see Fig. 5). In particular, in the baseline scenario, the agent did not receive any information about the virus, just the importance of going to work. In the self-health feedback scenario, the LLM also received information about the symptoms that the agent may feel. Lastly, in the full feedback scenario, the LLM also received information about the virus and the number of agents already infected in the system.

With the baseline model, they reproduced the results of a SIR-like model, with all agents exiting their homes every day. However, once the LLM received information about the symptoms of the agent, it usually decided to stay at home. Furthermore, once also

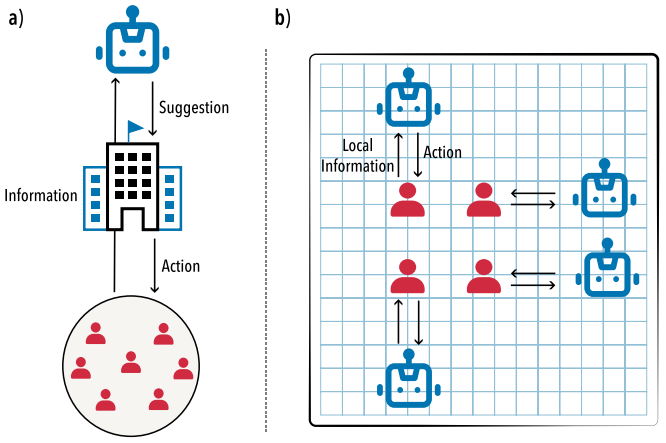

Fig. 6. Incorporating LLMs in societal decision-making. The left figure depicts intelligences providing assistance to the government or third-party organizations, while the right figure illustrates intelligences aiding individuals in decision-making processes. The integration of these elements in decision-making processes is conceptually similar to committees of domain experts but offers the possibility of doing so at an unprecedented scale.

information about the rest of the agents was provided, even agents without symptoms decided to stay at home, greatly diminishing the size of the outbreak. These results show the potential of applying GABMs for epidemic modeling and open many interesting questions. The first and perhaps most important is to determine if the decisions made by LLMs truly align with what humans do. If they are close enough, this type of models would allow researchers to systematically explore the effect of different demographics and personality traits on the reaction to outbreaks and public health interventions aimed to stop them.

6. Discussion

Complex systems is a field of research that spans across domains, from abstract mathematical problems to very applied inquiries about nature or human societies. It is reasonable to expect that the arrival of Large Language Models will have different impacts in many of these fields, whether they are simply another tool to help during the research process or a concept worth of investigating on their own. In this paper, we have focused in particular on those problems studied within the complex systems community that are more closely related to humans, including cooperation, social interactions, and even epidemic spreading.

Quantifying human decision-making is a significant challenge due to the intricacies of human behavior, such as systematic biases, the limited information they may have, and heuristics they may follow [158]. The introduction of Generative Agent-Based Models, where each agent's decisions are informed by LLMs, offers a promising avenue for addressing some of these challenges. If LLMs can truly imitate the behavior of humans covering a wide array of personalities and demographics, it would be possible to systematically study elements that could not be addressed until now [62,159]. Furthermore, due to the versatility and capabilities LLM agents demonstrate, they are even becoming a field of study on their own [160,161].

However, as we have seen, LLMs tend to hallucinate and be guided by unforeseen biases. This is partially because LLMs are highly sensitive to non-semantic features of prompts, such as word ordering and formatting [32]. But also because they may be biased during the training process and the posterior fine-tuning. This also explains why in this young field we can already find contradictory results, such as the ones discussed in the game theory section. As such, the community needs to create the tools to be able to study their behavior, discover novel abilities, and see how they evolve as these systems continue to be developed [162].

Moreover, the increasing integration of LLMs into daily life necessitates a thorough understanding of how these models interact with humans and each other, as in any complex system emergent phenomena may arise once different elements interact. For instance, we can envision a nearby future in which LLMs facilitate tripartite decision-making across governmental hierarchies, as well as the use of these models by individuals to make decisions, see Fig. 6. Integrating intelligent decision-making systems can significantly reduce the costs associated with traditional decision-making processes, which rely heavily on human and physical resources. Furthermore, empowering individuals with intelligent decision-making capabilities can streamline processes and improve their outcomes.

Future research should address several key challenges. Firstly, the paradigm of cooperation between humans and machines, as well as between multiple machines, needs further exploration. This includes investigating different interaction patterns [163], the emergence of new cooperative strategies and norms in these interactions, and assessing adherence to established rules of reciprocity, potentially identifying new mechanisms of cooperation [164]. However, benchmarking these models across a broad set of tasks is crucial before employing them in human behavior research. Guidelines for experimental conditions are essential to mitigate biases and ensure that LLMs' responses are reproducible and as genuine and accurate as possible.

In conclusion, while LLMs present a powerful tool for studying complex systems, and in particular those involving humans, their effective application requires careful consideration of biases, prompt design, and the dynamics of human-machine interactions. Future research should continue to refine these models, establish standardized methodologies, and explore the broader implications of integrating LLMs into societal and governmental decision-making processes.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Lei Shi was supported by Major Program of National Fund of Philosophy and Social Science of China (grants no. 22&ZD158 and 22VRCO49) and key project (no. 11931015) of the National Natural Science Foundation of China (NNSFC), NNSFC project no. 11671348, Yunling Scholar Post-support Program of Yunnan Province and, Key R&D Program of Yunnan Province (202403AC100010). YK L. was supported by the Commerce Statistical Society of China (No. 2023STY63). AA acknowledges support through the grant RYC2021-033226-I funded by MCIN/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR. AA and YM were partially supported by the Government of Aragon, Spain, and ERDF "A way of making Europe" through grant E36-20R (FENOL), and by Ministerio de Ciencia e Innovación, Agencia Española de Investigación (MCIN/AEI/10.13039/501100011033) Grant No. PID2023-149409NB-I00.

References

[1] Jo A. Nature 2023;614:214–6.
[2] Walters WP, Murcko M. Nat Biotechnol 2020;38:143–5.
[3] Liu Q, Xu J, Jiang R, Wong WH. Proc Natl Acad Sci 2021;118:e2101344118.
[4] Noy S, Zhang W. Science 2023;381:187–92.
[5] Wang H, Fu T, Du Y, Gao W, Huang K, Liu Z, et al. Nature 2023;620:47–60.
[6] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Adv Neural Inf Process Syst 2017;30.
[7] Pastor-Galindo J, Nespoli P, Ruipérez-Valiente JA. arXiv preprint. arXiv:2310.07545, 2023.
[8] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Advances in neural information processing systems (NIPS 2017), vol. 30. 2017.
[9] Stern J. Atlantic. https://www.theatlantic.com/technology/archive/2023/03/openai-gpt-4-parameters-power-debate/673290, 2023.
[10] Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E, et al. arXiv preprint. arXiv:2303.12712, 2023.
[11] Geng M, Trotta R. arXiv preprint. arXiv:2404.08627, 2024.
[12] Karinshak E, Liu SX, Park JS, Hancock JT. Proc ACM Hum-Comput Interact 2023;7:1–29.
[13] Grossmann I, Feinberg M, Parker DC, Christakis NA, Tetlock PE, Cunningham WA. Science 2023;380:1108–9.
[14] Epstein Z, Hertzmann A, of Human Creativity I, Akten M, Farid H, Fjeld J, et al. Science 2023;380:1110–1.
[15] Birhane A, Kasirzadeh A, Leslie D, Wachter S. Nat Rev Phys 2023;5:277–80.
[16] De Angelis L, Baglivo F, Arzilli G, Privitera GP, Ferragina P, Tozzi AE, et al. Front Public Health 2023;11:1166120.
[17] Park JS, Popowski L, Cai C, Morris MR, Liang P, Bernstein MS. Social simulacra: creating populated prototypes for social computing systems. In: Proceedings of the 35th annual ACM symposium on user interface software and technology; 2022. p. 1–18.
[18] AI4Science MR, Quantum MA. arXiv preprint. arXiv:2311.07361, 2023.
[19] Du N, Huang Y, Dai AM, Tong S, Lepikhin D, Xu Y, et al. Glam: efficient scaling of language models with mixture-of-experts. In: International conference on machine learning. PMLR; 2022. p. 5547–69.
[20] Ibrahim A, Thérien B, Gupta K, Richter ML, Anthony Q, Lesort T, et al. arXiv preprint. arXiv:2403.08763, 2024.
[21] Gadre SY, Smyrnis G, Shankar V, Gururangan S, Wortsman M, Shao R, et al. arXiv preprint. arXiv:2403.08540, 2024.
[22] Shekhar S, Dubey T, Mukherjee K, Saxena A, Tyagi A, Kotla N. arXiv preprint. arXiv:2402.01742, 2024.
[23] Song Y, Mi Z, Xie H, Chen H. arXiv preprint. arXiv:2312.12456, 2023.
[24] Stojkovic J, Choukse E, Zhang C, Goiri I, Torrellas J. arXiv preprint. arXiv:2403.20306, 2024.
[25] Rillig MC, Ågerstrand M, Bi M, Gould KA, Sauerland U. Environ Sci Technol 2023;57:3464–6.
[26] Kondratyuk D, Yu L, Gu X, Lezama J, Huang J, Hornung R, et al. arXiv preprint. arXiv:2312.14125, 2023.
[27] Yang Z, Liu J, Han Y, Chen X, Huang Z, Fu B, et al. arXiv preprint. arXiv:2312.13771, 2023.
[28] Bai J, Zhang S, Chen Z. arXiv preprint. arXiv:2308.11136, 2023.
[29] Kambhampati S. Ann NY Acad Sci 2024.
[30] Cheng M, Piccardi T, Yang D. arXiv preprint. arXiv:2310.11501, 2023.
[31] Veselovsky V, Ribeiro MH, Arora A, Josifoski M, Anderson A, West R. arXiv preprint. arXiv:2305.15041, 2023.
[32] Brand J, Israeli A, Ngwe D. Harvard Business School Marketing Unit Working Paper No. 23-062. 2023.
[33] Chen Y, Liu TX, Shan Y, Zhong S. Proc Natl Acad Sci 2023;120:e2316205120.
[34] Dillion D, Tandon N, Gu Y, Gray K. Trends Cogn Sci 2023.
[35] Mitsopoulos K, Bose R, Mather B, Bhatia A, Gluck K, Dorr B, et al. AAAI-SS 2023;2:340–8.
[36] Balepur N, Ravichander A, Rudinger R. arXiv preprint. arXiv:2402.12483, 2024.
[37] Horton JJ. arXiv preprint. arXiv:2301.07543, 2024.
[38] Schoenegger P, Tuminauskaite I, Park PS, Tetlock PE. arXiv preprint. arXiv:2402.19379, 2024.
[39] Argyle LP, Busby EC, Fulda N, Gubler JR, Rytting C, Wingate D. Polit Anal 2023;31:337–51.
[40] Aher GV, Arriaga RI, Kalai AT. Using large language models to simulate multiple humans and replicate human subject studies. In: International conference on machine learning. PMLR; 2023. p. 337–71.
[41] Ren S, Cui Z, Song R, Wang Z, Hu S. arXiv preprint. arXiv:2403.08251, 2024.
[42] Chiang WL, Zheng L, Sheng Y, Angelopoulos AN, Li T, Li D, et al. arXiv preprint. arXiv:2403.04132, 2024.
[43] Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, et al. Adv Neural Inf Process Syst 2022;35:27730–44.
[44] Wu H, Zhang Z, Zhang E, Chen C, Liao L, Wang A, et al. Q-instruct: improving low-level visual abilities for multi-modality foundation models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2024. p. 25490–500.
[45] Huang Q, Vora J, Liang P, Leskovec J. arXiv preprint. arXiv:2310.03302, 2023.
[46] Xu R, Qi Z, Wang C, Wang H, Zhang Y, Xu W. arXiv preprint. arXiv:2403.08319, 2024.
[47] Zhang Z, Chen C, Liu B, Liao C, Gong Z, Yu H, et al. arXiv preprint. arXiv:2311.07989, 2023.
[48] Xu R, Sun Y, Ren M, Guo S, Pan R, Lin H, et al. Inf Process Manag 2024;61:103665.

[49] Gao C, Lan X, Li N, Yuan Y, Ding J, Zhou Z, et al. arXiv preprint. arXiv:2312.11970, 2023.
[50] Guo T, Chen X, Wang Y, Chang R, Pei S, Chawla NV, et al. Large language model based multi-agents: a survey of progress and challenges. https://arxiv.org/abs/2402.01680, 2024.
[51] Cheng Y, Zhang C, Zhang Z, Meng X, Hong S, Li W, et al. arXiv preprint. arXiv:2401.03428, 2024.
[52] Ma Q, Xue X, Zhou D, Yu X, Liu D, Zhang X, et al. arXiv preprint. arXiv:2402.00262, 2024.
[53] Ghaffarzadegan N, Majumdar A, Williams R, Hosseinichimeh N. arXiv preprint. arXiv:2309.11456, 2023.
[54] Wang L, Ma C, Feng X, Zhang Z, Yang H, Zhang J, et al. Front Comput Sci 2024;18:1–26.
[55] Zhu X, Chen Y, Tian H, Tao C, Su W, Yang C, et al. arXiv preprint. arXiv:2305.17144, 2023.
[56] Barabási AL, Bonabeau E. Sci Am 2003;288:60–9.
[57] Wang XF, Chen G. IEEE Circuits Syst Mag 2003;3:6–20.
[58] Samoylenko I, Aleja D, Primo E, Alfaro-Bittner K, Vasilyeva E, Kovalenko K, et al. Phys Rev X 2023;13:021032.
[59] De Marzo G, Pietronero L, Garcia D. arXiv preprint. arXiv:2312.06619, 2023.
[60] Jiang J, Ferrara E. arXiv preprint. arXiv:2401.00893, 2023.
[61] Mao J, Zou D, Sheng L, Liu S, Gao C, Wang Y, et al. arXiv preprint. arXiv:2403.03962, 2024.
[62] Park JS, O'Brien J, Cai CJ, Morris MR, Liang P, Bernstein MS. Generative agents: interactive simulacra of human behavior. In: Proceedings of the 36th annual ACM symposium on user interface software and technology; 2023. p. 1–22.
[63] Albert R, Barabási AL. Rev Mod Phys 2002;74:47–97.
[64] Lai S, Potter Y, Kim J, Zhuang R, Song D, Evans J. arXiv preprint. arXiv:2402.12590, 2024.
[65] Sharma M, Tong M, Korbak T, Duvenaud D, Askell A, Bowman SR, et al. arXiv preprint. arXiv:2310.13548, 2023.
[66] Meskó B, Topol EJ. npj Digit Med 2023;6:1–6.
[67] Roughgarden T. Commun ACM 2010;53:78–86.
[68] Xia C, Wang J, Perc M, Wang Z. Phys Life Rev 2023.
[69] Civilini A, Anbarci N, Latora V. Phys Rev Lett 2021;127:268301.
[70] Zhu Z, Dong Y, Lu Y, Shi L. Phys A, Stat Mech Appl 2021;569:125772.
[71] Szolnoki A, Perc M. J R Soc Interface 2015;12:20141299.
[72] Pi B, Zeng Z, Feng M, Kurths J. Chaos: an interdisciplinary. J Nonlinear Sci 2022;32.
[73] Lu Y, Geng Y, Gan W, Shi L. Phys A, Stat Mech Appl 2019;526:121124.
[74] Perc M. Sci Rep 2019;9:16549.
[75] Jusup M, Holme P, Kanazawa K, Takayasu M, Romić I, Wang Z, et al. Phys Rep 2022;948:1–148.
[76] Xiong J, Xie H, Liu B, Li B, Gui L. IEEE Trans Veh Technol 2021;70:9437–49.
[77] Quan J, Nie J, Chen W, Wang X. Chaos Solitons Fractals 2022;158:111986.
[78] Szolnoki A, Perc M. Phys Rev E 2016;93:062307.
[79] Cardillo A, Masuda N. Phys Rev Res 2020;2:023305.
[80] Masuda N. Sci Rep 2012;2:646.
[81] Du C, Lu Y, Meng H, Park J. Chaos: an interdisciplinary. J Nonlinear Sci 2024;34.
[82] Geng Y, Liu Y, Lu Y, Shen C, Shi L. Appl Math Comput 2022;427:127182.
[83] Lu Y, Wang Y, Liu Y, Chen J, Shi L, Park J. Chaos: an interdisciplinary. J Nonlinear Sci 2023;33.
[84] Wang Z, Jusup M, Shi L, Lee JH, Iwasa Y, Boccaletti S. Nat Commun 2018;9:2954.
[85] Shi L, Romić I, Ma Y, Wang Z, Podobnik B, Stanley HE, et al. Proc Natl Acad Sci 2020;117:17516–21.
[86] Crandall JW, Oudah M, Tennom, Ishowo-Oloko F, Abdallah S, Bonnefon JF, et al. Nat Commun 2018;9:233.
[87] Makovi K, Sargsyan A, Li W, Bonnefon JF, Rahwan T. Nat Commun 2023;14:3108.
[88] Karpus J, Krüger A, Verba JT, Bahrami B, Deroy O. iScience 2021;24.
[89] Santurkar S, Durmus E, Ladhak F, Lee C, Liang P, Hashimoto T. Whose opinions do language models reflect? In: International conference on machine learning. PMLR; 2023. p. 29971–30004.
[90] Qian C, Cong X, Yang C, Chen W, Su Y, Xu J, et al. arXiv preprint. arXiv:2307.07924, 2023.
[91] Guo F. arXiv preprint. arXiv:2305.05516, 2023.
[92] Suzuki R, Arita T. Sci Rep 2024;14:5989.
[93] Fan C, Chen J, Jin Y, He H. Can large language models serve as rational players in game theory? A systematic analysis. In: Proceedings of the AAAI conference on artificial intelligence; 2024. p. 17960–7.
[94] Brookins P, DeBacker JM. 2023. Available at SSRN 4493398.
[95] Fontana N, Pierri F, Aiello LM. arXiv preprint. arXiv:2406.13605, 2024.
[96] Capraro V, Di Paolo R, Perc M, Pizziol V. J R Soc Interface 2024;21:20230720.
[97] Chan A, Riché M, Clifton J. arXiv preprint. arXiv:2303.13360, 2023.
[98] Sreedhar K, Chilton L. arXiv preprint. arXiv:2402.08189, 2024.
[99] Johnson T, Obradovich N. arXiv preprint. arXiv:2301.02330, 2023.
[100] Xie C, Chen C, Jia F, Ye Z, Shu K, Bibi A, et al. arXiv preprint. arXiv:2402.04559, 2024.
[101] Mei Q, Xie Y, Yuan W, Jackson MO. Proc Natl Acad Sci 2024;121(e2313925121).
[102] McCannon BC. Econ Lett 2024;241:111828.
[103] Mozikov M, Severin N, Bodishtianu V, Glushanina M, Baklashkin M, Savchenko AV, et al. arXiv preprint. arXiv:2406.03299, 2024.
[104] Babin JJ, Chauhan H. researchsquare preprint. https://doi.org/10.21203/rs.3.rs-4462506/v1, 2024.
[105] Henry J. arXiv preprint. arXiv:2402.05786, 2024.
[106] Lorè N, Heydari B. arXiv preprint. arXiv:2309.05898, 2023.
[107] Xu L, Hu Z, Zhou D, Ren H, Dong Z, Keutzer K, et al. Magic: investigation of large language model powered multi-agent in cognition, adaptability, rationality and collaboration. In: ICLR 2024 workshop on large language model (LLM) agents; 2023.
[108] Akata E, Schulz L, Coda-Forno J, Oh SJ, Bethge M, Schulz E. arXiv preprint. arXiv:2305.16867, 2023.
[109] Phelps S, Russell YI. arXiv preprint. arXiv:2305.07970, 2023.
[110] Lorè N, Heydari B. Strategic behavior of large language models: game structure vs. contextual framing. https://arxiv.org/abs/2309.05898, 2023.
[111] Duan J, Zhang R, Diffenderfer J, Kailkhura B, Sun L, Stengel-Eskin E, et al. arXiv preprint. arXiv:2402.12348, 2024.
[112] Herr N, Acero F, Raileanu R, Pérez-Ortiz M, Li Z. arXiv preprint. arXiv:2407.04467, 2024.
[113] Roberts J, Moore K, Fisher D. arXiv preprint. arXiv:2404.08710, 2024.
[114] Li J, Li R, Liu Q. arXiv preprint. arXiv:2309.04369, 2023.
[115] Huang Jt, Li EJ, Lam MH, Liang T, Wang W, Yuan Y, et al. arXiv preprint. arXiv:2403.11807, 2024.
[116] Oosterbeek H, Sloof R, Van De Kuilen G. Exp Econ 2004;7:171–88.
[117] Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H. Foundations of human sociality: economic experiments and ethnographic evidence from fifteen small-scale societies. Oxford University Press. ISBN 9780199262052, 2004.

[118] Engel C. Exp Econ 2011;14:583–610.
[119] Dal Bó P, Fréchette GR. Am Econ Rev 2019;109:3929–52.
[120] Mengel F. Econ J 2018;128:3182–209.
[121] Capraro V, Halpern JY, Perc M. J Econ Lit 2024;62:115–54. https://www.aeaweb.org/articles?id=10.1257/jel.20221613.
[122] Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, et al. NeurIPS 2022;35:27730–44.
[123] Gao C, Lan X, Lu Z, Mao J, Piao J, Wang H, et al. arXiv preprint. arXiv:2307.14984, 2023.
[124] GLM T, Zeng A, Xu B, Wang B, Zhang C, Yin D, et al. arXiv preprint. arXiv:2406.12793, 2024.
[125] Buchanan B, Lohn A, Musser M, Sedova K 2021;1:2. Center for Security and Emerging Technology.
[126] Huang L, Yu W, Ma W, Zhong W, Feng Z, Wang H, et al. arXiv preprint. arXiv:2311.05232, 2023.
[127] Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Psychol Sci 2020;31:770–80.
[128] Kriplean T, Toomim M, Morgan J, Borning A, Ko AJ. Is this what you meant? Promoting listening on the web with reflect. In: ACM conferences. New York, NY, USA: Association for Computing Machinery; 2012. p. 1559–68.
[129] Kim S, Eun J, Seering J, Lee J. Proc ACM Hum-Comput Interact 2021;5:1–26.
[130] Argyle LP, Bail CA, Busby EC, Gubler JR, Howe T, Rytting C, et al. Proc Natl Acad Sci 2023;120:e2311627120.
[131] Yang JC, Korecki M, Dailisan D, Hausladen CI, Helbing D. arXiv preprint. arXiv:2402.01766, 2024.
[132] Wang Y, Cai Y, Chen M, Liang Y, Hooi B. arXiv preprint. arXiv:2310.13206, 2023.
[133] Feng S, Park CY, Liu Y, Tsvetkov Y. arXiv preprint. arXiv:2305.08283, 2023.
[134] Zhuge M, Liu H, Faccio F, Ashley DR, Csordás R, Gopalakrishnan A, et al. arXiv preprint. arXiv:2305.17066, 2023.
[135] Liang T, He Z, Jiao W, Wang X, Wang Y, Wang R, et al. arXiv preprint. arXiv:2305.19118, 2023.
[136] Hao R, Hu L, Qi W, Wu Q, Zhang Y, Nie L. arXiv preprint. arXiv:2304.12998, 2023.
[137] Du Y, Li S, Torralba A, Tenenbaum JB, Mordatch I. arXiv preprint. arXiv:2305.14325, 2023.
[138] Xiong K, Ding X, Cao Y, Liu T, Qin B. arXiv preprint. arXiv:2305.11595, 2023.
[139] Wu Q, Bansal G, Zhang J, Wu Y, Zhang S, Zhu E, et al. arXiv preprint. arXiv:2308.08155, 2023.
[140] Li G, Hammoud H, Itani H, Khizbullin D, Ghanem B. NeurIPS 2023;36:51991–2008.
[141] Wang Z, Mao S, Wu W, Ge T, Wei F, Ji H. arXiv preprint. arXiv:2307.05300, 2023.
[142] Liu R, Yang R, Jia C, Zhang G, Zhou D, Dai AM, et al. arXiv preprint. arXiv:2305.16960, 2023.
[143] Pastor-Satorras R, Vespignani A. Phys Rev Lett 2001;86:3200.
[144] Vespignani A. Nature 2018;558:528–9.
[145] Williams R, Hosseinichimeh N, Majumdar A, Ghaffarzadegan N. arXiv preprint. arXiv:2307.04986, 2023.
[146] Barbieri F, Espinosa Anke L, Camacho-Collados J. XLM-T: multilingual language models in Twitter for sentiment analysis and beyond. In: Calzolari N, Béchet F, Blache P, Choukri K, Cieri C, Declerck T, et al., editors. Proceedings of the thirteenth language resources and evaluation conference. Marseille, France: European Language Resources Association; 2022. p. 258–66. https://aclanthology.org/2022.lrec-1.27.
[147] Candellone E, Aleta A, Ferraz de Arruda H, Meijaard E, Moreno Y. Commun Earth Environ 2024;5:391.
[148] Müller M, Salathé M, Kummervold PE. Front Artif Intell 2023;6:1023281.
[149] Zhou B, Zou L, Mostafavi A, Lin B, Yang M, Gharaibeh N, et al. Comput Environ Urban Syst 2022;95:101824.
[150] Deiner MS, Deiner NA, Hristidis V, McLeod SD, Doan T, Lietman TM, et al. J Med Internet Res 2024;26:e49139.
[151] Yan W, Liu H, Wu T, Chen Q, Wang W, Chai H, et al. arXiv preprint. arXiv:2406.13890, 2024.
[152] Sabry Abdel-Messih M, Kamel Boulos MN. JMIR Med Educ 2023;9(e46876).
[153] Li J, Wang S, Zhang M, Li W, Lai Y, Kang X, et al. arXiv preprint. arXiv:2405.02957, 2024.
[154] Tang X, Zou A, Zhang Z, Zhao Y, Zhang X, Cohan A, et al. arXiv preprint. arXiv:2311.10537, 2023.
[155] Kim Y, Park C, Jeong H, Chan YS, Xu X, McDuff D, et al. arXiv preprint. arXiv:2404.15155, 2024.
[156] Lee P, Bubeck S, Petro J. N Engl J Med 2023;388:1233–9.
[157] Pangallo M, Aleta A, del Rio-Chanona RM, Pichler A, Martín-Corral D, Chinazzi M, et al. Nat Hum Behav 2024;8:264–75.
[158] Tversky A, Kahneman D. Science 1974;185:1124–31.
[159] Wang Z, Chiu YY, Chiu YC. arXiv preprint. arXiv:2310.05418, 2023.
[160] Xi Z, Chen W, Guo X, He W, Ding Y, Hong B, et al. arXiv preprint. arXiv:2309.07864, 2023.
[161] LLM-Agent-Paper-List, Online; https://github.com/WooooDyy/LLM-Agent-Paper-List, 2024. [Accessed 31 July 2024].
[162] Hagendorff T. arXiv preprint. arXiv:2303.13988, 2023.
[163] Capraro V, Perc M. Nat Comput Sci 2024;4:257–8.
[164] Nowak MA. Science 2006;314:1560–3.