### Generative Agent-Based Models for Complex Systems Research: a review

Yikang Lu<sup>1,†</sup>, Alberto Aleta<sup>4,5,‡</sup>, Chunpeng Du<sup>3</sup>, Lei Shi<sup>1,2</sup>,\* and Yamir Moreno<sup>4,5,6†</sup>

<sup>1</sup> School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming, 650221, China

<sup>2</sup> School of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance, Shanghai, 201209, China

<sup>3</sup> School of Mathematics, Kunming University, Kunming, Yunnan 650214, China

<sup>4</sup> Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza, Zaragoza, 50018, Spain

<sup>5</sup> Department of Theoretical Physics, University of Zaragoza, Zaragoza, 50009, Spain and

<sup>6</sup> Centai Institute, Turin, Italy<sup>‡</sup>

(Dated: August 20, 2024)

The advent of Large Language Models (LLMs) has significantly transformed the fields of natural and social sciences. Generative Agent-Based Models (GABMs), which utilize large language models in place of real subjects, are gaining increasing public attention. Far from aiming for comprehensiveness, this paper aims to offer readers an opportunity to understand how large language models are disrupting complex systems research and behavioral sciences. In particular, we evaluate recent advancements in various domains within complex systems, encompassing network science, evolutionary game theory, social dynamics, and epidemic propagation. Additionally, we propose possible directions for future research to further advance these fields.

#### I. 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]. 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 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 [8, 9]. 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 [10] and may have a profound impact on the advancement of both social and natural sciences [4, 11–17].

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 [18], continuously pre-training [19], or using scaling laws to extrapolate during training [20]. Other researches focus on

optimizing the cost of using LLMs [21, 22] or on mitigating their ecological impact [23, 24]. There are also intensive efforts devoted to extending their generative capacities beyond text such as VideoPoet for video generation [25], or AppAgent for creating agents capable of operating smartphone applications [26].

However, the interest of LLMs in research goes beyond these foundational aspects. In the field of social sciences, LLMs find applications in various domains [27]. In linguistics, they are utilized for language prediction tasks [28]. In economics and social sciences, researchers are striving to imbue LLMs with unique personalities, enabling them to operate as individuals to generate synthetic data [29, 30]. 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 [31]. In addition, LLMs' decisions in budget allocation scenarios received higher rationality scores than those made by humans [32]. This alignment underscores its capacity to produce authentic survey responses relevant to consumer behavior [31].

In psychological experiments, LLMs' behaviors demonstrated a high degree of congruence with prevailing societal values [33]. In multiple-choice question tests, it was shown that LLMs could successfully best the majority baseline and even infer the question given the options [34]. In the exploration of fairness and framing effects in sociology, researchers have integrated LLMs as computational models of humans into classic game experiments [35]. Similarly, LLMs can replicate the "wisdom of the crowd" effect, akin to human behavior [36]. 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 [37, 38]. The demonstrated consistency with human behavior [33, 39–42] suggests that LLMs can perform some

<sup>\*</sup> shi\_lei65@hotmail.com

<sup>†</sup> yamir.moreno@gmail.com

<sup>&</sup>lt;sup>‡</sup> These authors contributed equally to this work.

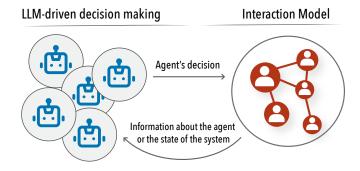


FIG. 1. Generative Agent-Based Models (GABMs). In GABMs, agents do not make decisions about their interactions based on a fixed set of rules. Instead, a prompt is sent to an LLM including the desired details and it returns the decision that the agent should follow [48].

of the same operations as humans [43], 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 [37].

This surge in research content pertaining to LLMs has prompted several review articles from diverse perspectives [9, 44–47]. 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) [48].

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 Figure 1. For instance, Zhu et al. [49] 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.

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

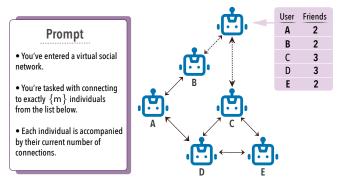


FIG. 2. LLM-based network growth with generative agents. Each generative agent of a hypothetical online social network is initialized with the prompt shown on the left, together with a comprehensive list of all network users along with their respective number of friends. Then, at each time step, a new 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 [53].

## II. 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 [50, 51]. 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 [52]. Thus, the study of these networks can provide answers to interesting questions on human behavior.

LLMs can be used to analyze these networks [54, 55]. 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. [56] 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.* [53] 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 Figure 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 Figure 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 [57]. 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 [53]. 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 asked 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. [58] 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 intelligencies 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 that are usually imposed on them, such as their bias towards positivity [59, 60].

# III. 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 [61]. 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 [62–67] and behavioral sciences [68, 69]. In traditional evolutionary game models, the evolution of the strategies of the individuals, such as Fermi updating [62–64], conformist updating [65–67], or self-reversing rules [70, 71], must be pre-established.

This paradigm persists even in studies involving basic human-computer interactions like simple bots [30, 72– 74] and interactions with the environment driven by reinforcement learning-based intelligences [75–77]. In addition, numerous experimental studies have investigated real-life gaming behavior to explore the mechanisms underlying the persistence of inter-individual cooperation [78, 79]. 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 [80–82]. 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 [83].

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 [84]. Along these lines, we can identify, at least, four advantages to using LLMs instead of human participants in evolutionary game experiments [85]. 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 [35], 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 (AIs) may prove advantageous for research endeavors about communication-related subjects [86].

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 [35, 38, 85], rock-paper-scissors games [87], prisoner's dilemmas [85, 88, 89], public goods [58], and

others (see Table I). 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 [35]. These experiments illustrate the feasibility of simulating individuals with a wide range of characteristics and traits [86, 90]. In the following, we review some of the results obtained in these games.

Game	Paper
The Dictator Game	[35, 87, 88, 90–98]
The Ultimatum Game	[38, 85, 91, 92, 95, 97, 99]
The Prisoner's Dilemma	[85, 88, 89, 91, 95, 97, 100–107]
	[58, 95, 98, 101, 108, 109]

TABLE I. List of papers with behavioral experiments that include artificial agents driven by LLMs.

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 [110]. Aher et al. [38] 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. [92] 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 multiagent 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 [111, 112]. Brookins et al. [88] 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 Figure 3, we provide the instructions for the Dictator Game proposed by Brookins et al.

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 [113].

To simplify the analysis, Brookins et al. [88] performed an experiment based on the one-shot Prisoner's Dilemma, illustrated in Figure 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 [114]. 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. [103] 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 LLMs are also sensitive to nonsemantic features of the prompt such as changes in wordordering 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 [103].

A very different result was obtained by Akata et al. [102], 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

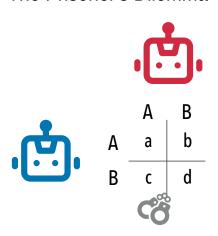
### The Dictator Game



- This task is about dividing money between yourself and another person to whom you are randomly matched. You do not know this other person and you will not knowingly meet him/her.
- You have been ramdonly assigned the role of the "allocator". The other person is in the role of the "recipient".
- You are endowed with 10€ and the recipient is endowed with 0€.
- You can decide how much of your 10€ endowment to transfer to the recipient. You can choose any amount between 0€ and 10€. The recipient receives the amount that you decide to transfer to him/her; and you receive the amount that you decide not to transfer and thus to keep.
- How much of your 10€ endowment do you want to transfer to the recipient? Just tell me the allocation, not your reasoning.

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 [88].

### The Prisoner's Dilemma



- You can select one of the 2 choices: A or B. The other player will select one of the choices, and the payoff you get will depend on both of your choices.
- Payoff is determined as follows:

If you both choose A: both get a euro.

If you both choose B: both get d euro.

If you choose A, and the other player chooses B: you get b euro, the other gets c euro.

If you choose B, and the other player chooses A: you get c euro, the other gets b euro.

- Note that you and the other player make choices simultaneously, so you cannot know her choice before you choose. Please, pretend that you are a human in this single-shot game.
- Tell me which choice would you make: A or B. Do not explain your reasoning.

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 [88].

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 [115].

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. [101] 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. [108] also found that GPT-4 could beat other LLMs such as PaLM or ChatGPT.

Lai et al. [58] 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 antisocial behaviors. However, Huang et al. [109] 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.

# IV. 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. [116] simulated the propagation of information in a social network of LLM agents created with ChatGLM [117]. 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. One of the earliest examples of the former is the study of Buchanan et al. in 2021 [118]. 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 [119] - 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 [120]. Similarly, allowing people to reflect on their messages through human or machine interaction, facilitates opinion alignment and group consensus [121, 122]. For these reasons, Argile et al. [123] 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. [124] replicated a human experiment on participatory budgeting 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. [125]. 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. [126] 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. [37] 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 [127]. The main idea, as explained by Liang et al. [128], 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 [129] and a combination of ChatGPT and Bard [130] which also reduced fallacies and hallucinations. However, as Xiong et al. [131] reported, mixing more powerful LLMs with weaker ones

can sometimes lead to worse results.

It is worth noting that there are already some opensource libraries that facilitate the creation of multi-agent systems such as AutoGen [132] or CAMEL [133]. Furthermore, in these systems agents can be assigned specific roles, so that researchers can tailor the group of LLMs depending on the problem they want to tackle. A similar proposal was introduced by Wang et al. [134] although in their case the LLM was supported by humans who were experts in different domains and helped it 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. [135], who also introduced social interaction during the own training process of the LLM, which reportedly made models more robust against attacks.

## V. EPIDEMIC MODELLING 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 [136, 137]. 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 [139–142], Deiner et al. [143] 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 [144–148]. 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 [149].

Probably, the most straightforward way of applying LLMs in current epidemic models is through the use of generative agent-based models (GABMs) [48]. 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 [150]. 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. [138] 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 Figure 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 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 model 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.

#### VI. 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

#### LLM-driven decision-making **Epidemic propagation** Susceptible Infected Removed Base run: [agent's name] goes to work to earn money to support [agent's name]'s self. Self-health feedback: [agent's name] goes to work Decision to earn money to support [agent's name]'s self. [agent's name] feels normal/has a light cough/ has a fever and cough. Full feedback: [agent's name] goes to work to earn money to support [agent's name]'s self. [agents's Feedback name] feels normal/has a light cough/has a fever and cough. [agent's name] knows about the Catasat virus spreading across the country. [...] [agent's name] checks the newspaper and finds that [X] % of Dewberry Hollow's population caught new infections of the Catasat virus

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*)[138].

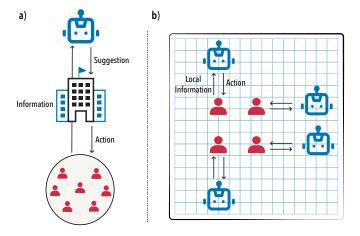


FIG. 6. Incoporating LLMs in societal decision-making. The left figure depicts intelligences providing assistance to the government or third-party organizations, while the right figure illustrates intelligences aiding individuals in decision-making processes. The integration of these elements in decision-making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

as systematic biases, the limited information they may have, and heuristics they may follow [151]. The introduction of Generative Agent-Based Models, where each agent's decisions are informed by LLMs, offers a promis-

ing 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 [56, 152]. Furthermore, due to the versatility and capabilities LLM agents demonstrate, they are even becoming a field of study on their own [153, 154].

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 [31]. 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 [155].

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 Figure 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 outcomes.

Future research should address several key challenges and priorities. Firstly, the paradigm of cooperation between humans and machines, as well as between multiple machines, needs further exploration. This includes investigating 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 [156]. 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.

#### ACKNOWLEDGMENTS

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, Yunnan Provincial Science and Technology Key Project (No. 202403AC0010). 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.

- [1] Jo A 2023  $Nature~\mathbf{614}~214-216$
- [2] Walters W P and Murcko M 2020 Nature biotechnology 38 143–145
- [3] Liu Q, Xu J, Jiang R and Wong W H 2021 Proceedings of the National Academy of Sciences 118 e2101344118
- [4] Noy S and Zhang W 2023 Science **381** 187–192
- [5] Wang H, Fu T, Du Y, Gao W, Huang K, Liu Z, Chandak P, Liu S, Van Katwyk P, Deac A et al. 2023 Nature 620 47–60
- [6] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I 2017 Advances in Neural Information Processing Systems 30
- [7] Pastor-Galindo J, Nespoli P and Ruipérez-Valiente J A 2023 arXiv preprint arXiv:2310.07545
- [8] Stern J 2023 Atlantic URL https://www. theatlantic.com/technology/archive/2023/03/ openai-gpt-4-parameters-power-debate/673290
- [9] Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E, Lee P, Lee Y T, Li Y, Lundberg S et al. 2023 arXiv preprint arXiv:2303.12712
- [10] Geng M and Trotta R 2024 arXiv preprint arXiv:2404.08627
- [11] Karinshak E, Liu S X, Park J S and Hancock J T 2023 Proceedings of the ACM on Human-Computer Interaction 7 1–29
- [12] Grossmann I, Feinberg M, Parker D C, Christakis N A, Tetlock P E and Cunningham W A 2023 Science 380 1108–1109
- [13] Epstein Z, Hertzmann A, of Human Creativity I, Akten M, Farid H, Fjeld J, Frank M R, Groh M, Herman L, Leach N et al. 2023 Science 380 1110–1111

- [14] Birhane A, Kasirzadeh A, Leslie D and Wachter S 2023 Nature Reviews Physics 5 277–280
- [15] De Angelis L, Baglivo F, Arzilli G, Privitera G P, Ferragina P, Tozzi A E and Rizzo C 2023 Frontiers in Public Health 11 1166120
- [16] Park J S, Popowski L, Cai C, Morris M R, Liang P and 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 pp 1–18
- [17] AI4Science M R and Quantum M A 2023 arXiv preprint arXiv:2311.07361
- [18] Du N, Huang Y, Dai A M, Tong S, Lepikhin D, Xu Y, Krikun M, Zhou Y, Yu A W, Firat O et al. 2022 Glam: Efficient scaling of language models with mixture-ofexperts International Conference on Machine Learning (PMLR) pp 5547–5569
- [19] Ibrahim A, Thérien B, Gupta K, Richter M L, Anthony Q, Lesort T, Belilovsky E and Rish I 2024 arXiv preprint arXiv:2403.08763
- [20] Gadre S Y, Smyrnis G, Shankar V, Gururangan S, Wortsman M, Shao R, Mercat J, Fang A, Li J, Keh S et al. 2024 arXiv preprint arXiv:2403.08540
- [21] Shekhar S, Dubey T, Mukherjee K, Saxena A, Tyagi A and Kotla N 2024 arXiv preprint arXiv:2402.01742
- [22] Song Y, Mi Z, Xie H and Chen H 2023 arXiv preprint arXiv:2312.12456
- [23] Stojkovic J, Choukse E, Zhang C, Goiri I and Torrellas J 2024 arXiv preprint arXiv:2403.20306
- [24] Rillig M C, Ågerstrand M, Bi M, Gould K A and Sauerland U 2023 Environ. Sci. Technol. 57 3464–3466 ISSN

- 0013-936X
- [25] Kondratyuk D, Yu L, Gu X, Lezama J, Huang J, Hornung R, Adam H, Akbari H, Alon Y, Birodkar V et al. 2023 arXiv preprint arXiv:2312.14125
- [26] Yang Z, Liu J, Han Y, Chen X, Huang Z, Fu B and Yu G 2023 arXiv preprint arXiv:2312.13771
- [27] Bai J, Zhang S and Chen Z 2023 arXiv preprint arXiv:2308.11136
- [28] Kambhampati S 2024 Annals of the New York Academy of Sciences
- [29] Cheng M, Piccardi T and Yang D 2023 arXiv preprint arXiv:2310.11501
- [30] Veselovsky V, Ribeiro M H, Arora A, Josifoski M, Anderson A and West R 2023 arXiv preprint arXiv:2305.15041
- [31] Brand J, Israeli A and Ngwe D 2023 Harvard Business School Marketing Unit Working Paper No. 23-062
- [32] Chen Y, Liu T X, Shan Y and Zhong S 2023 Proceedings of the National Academy of Sciences 120 e2316205120
- [33] Dillion D, Tandon N, Gu Y and Gray K 2023 Trends in Cognitive Sciences
- [34] Balepur N, Ravichander A and Rudinger R 2024 arXiv preprint arXiv:2402.12483
- [35] Horton J J 2024 arXiv preprint arXiv:2301.07543
- [36] Schoenegger P, Tuminauskaite I, Park P S and Tetlock P E 2024 arXiv preprint arXiv:2402.19379
- [37] Argyle L P, Busby E C, Fulda N, Gubler J R, Rytting C and Wingate D 2023 Political Analysis 31 337–351
- [38] Aher G V, Arriaga R I and Kalai A T 2023 Using large language models to simulate multiple humans and replicate human subject studies *International Conference on Machine Learning* (PMLR) pp 337–371
- [39] Ren S, Cui Z, Song R, Wang Z and Hu S 2024 arXiv preprint arXiv:2403.08251
- [40] Chiang W L, Zheng L, Sheng Y, Angelopoulos A N, Li T, Li D, Zhang H, Zhu B, Jordan M, Gonzalez J E et al. 2024 arXiv preprint arXiv:2403.04132
- [41] Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, Zhang C, Agarwal S, Slama K, Ray A et al. 2022 Advances in neural information processing systems 35 27730-27744
- [42] Wu H, Zhang Z, Zhang E, Chen C, Liao L, Wang A, Xu K, Li C, Hou J, Zhai G et al. 2024 Q-instruct: Improving low-level visual abilities for multi-modality foundation models Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp 25490–25500
- [43] Huang Q, Vora J, Liang P and Leskovec J 2023 arXiv preprint arXiv:2310.03302
- [44] Xu R, Qi Z, Wang C, Wang H, Zhang Y and Xu W 2024 arXiv preprint arXiv:2403.08319
- [45] Zhang Z, Chen C, Liu B, Liao C, Gong Z, Yu H, Li J and Wang R 2023 arXiv preprint arXiv:2311.07989
- [46] Xu R, Sun Y, Ren M, Guo S, Pan R, Lin H, Sun L and Han X 2024 Information Processing & Management 61 103665
- [47] Gao C, Lan X, Li N, Yuan Y, Ding J, Zhou Z, Xu F and Li Y 2023 arXiv preprint arXiv:2312.11970
- [48] Ghaffarzadegan N, Majumdar A, Williams R and Hosseinichimeh N 2023 arXiv preprint arXiv:2309.11456
- [49] Zhu X, Chen Y, Tian H, Tao C, Su W, Yang C, Huang G, Li B, Lu L, Wang X et al. 2023 arXiv preprint arXiv:2305.17144

- [50] Barabási A L and Bonabeau E 2003 Scientific american 288 60–69
- [51] Wang X F and Chen G 2003 IEEE circuits and systems magazine 3 6–20
- [52] Samoylenko I, Aleja D, Primo E, Alfaro-Bittner K, Vasilyeva E, Kovalenko K, Musatov D, Raigorodskii A M, Criado R, Romance M et al. 2023 Physical Review X 13 021032
- [53] De Marzo G, Pietronero L and Garcia D 2023 arXiv preprint arXiv:2312.06619
- [54] Jiang J and Ferrara E 2023 arXiv preprint arXiv:2401.00893
- [55] Mao J, Zou D, Sheng L, Liu S, Gao C, Wang Y and Li Y 2024 arXiv preprint arXiv:2403.03962
- [56] Park J S, O'Brien J, Cai C J, Morris M R, Liang P and Bernstein M S 2023 Generative agents: Interactive simulacra of human behavior Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology pp 1–22
- [57] Albert R and Barabási A L 2002 Rev. Mod. Phys. 74 47–97 ISSN 1539-0756
- [58] Lai S, Potter Y, Kim J, Zhuang R, Song D and Evans J 2024 arXiv preprint arXiv:2402.12590
- [59] Sharma M, Tong M, Korbak T, Duvenaud D, Askell A, Bowman S R, Cheng N, Durmus E, Hatfield-Dodds Z, Johnston S R et al. 2023 arXiv preprint arXiv:2310.13548
- [60] Meskó B and Topol E J 2023 npj Digital Med. 6 1–6 ISSN 2398-6352
- [61] Roughgarden T 2010 Communications of the ACM 53 78–86
- [62] Xia C, Wang J, Perc M and Wang Z 2023 Physics of Life Reviews
- [63] Civilini A, Anbarci N and Latora V 2021 Physical Review Letters 127 268301
- [64] Zhu Z, Dong Y, Lu Y and Shi L 2021 Physica A: Statistical Mechanics and its Applications 569 125772
- [65] Szolnoki A and Perc M 2015 Journal of The Royal Society Interface 12 20141299
- [66] Pi B, Zeng Z, Feng M and Kurths J 2022 Chaos: An Interdisciplinary Journal of Nonlinear Science 32
- [67] Lu Y, Geng Y, Gan W and Shi L 2019 Physica A: Statistical Mechanics and its Applications 526 121124
- [68] Perc M 2019 Scientific reports **9** 16549
- [69] Jusup M, Holme P, Kanazawa K, Takayasu M, Romić I, Wang Z, Geček S, Lipić T, Podobnik B, Wang L et al. 2022 Physics Reports 948 1–148
- [70] Xiong J, Xie H, Liu B, Li B and Gui L 2021 IEEE Transactions on Vehicular Technology 70 9437–9449
- [71] Quan J, Nie J, Chen W and Wang X 2022 Chaos, Solitons & Fractals 158 111986
- [72] Szolnoki A and Perc M 2016 Physical Review E  ${\bf 93}$  062307
- [73] Cardillo A and Masuda N 2020 Physical Review Research 2 023305
- [74] Masuda N 2012 Scientific reports 2 646
- [75] Du C, Lu Y, Meng H and Park J 2024 Chaos: An Interdisciplinary Journal of Nonlinear Science 34
- [76] Geng Y, Liu Y, Lu Y, Shen C and Shi L 2022 Applied Mathematics and Computation 427 127182
- [77] Lu Y, Wang Y, Liu Y, Chen J, Shi L and Park J 2023 Chaos: An Interdisciplinary Journal of Nonlinear Science 33

- [78] Wang Z, Jusup M, Shi L, Lee J H, Iwasa Y and Boccaletti S 2018 Nature communications 9 2954
- [79] Shi L, Romić I, Ma Y, Wang Z, Podobnik B, Stanley H E, Holme P and Jusup M 2020 Proceedings of the National Academy of Sciences 117 17516-17521
- [80] Crandall J W, Oudah M, Tennom, Ishowo-Oloko F, Abdallah S, Bonnefon J F, Cebrian M, Shariff A, Goodrich M A and Rahwan I 2018 Nature communications 9 233
- [81] Makovi K, Sargsyan A, Li W, Bonnefon J F and Rahwan T 2023 Nature communications 14 3108
- [82] Karpus J, Krüger A, Verba J T, Bahrami B and Deroy O 2021 Iscience 24
- [83] Santurkar S, Durmus E, Ladhak F, Lee C, Liang P and Hashimoto T 2023 Whose opinions do language models reflect? *International Conference on Machine Learning* (PMLR) pp 29971–30004
- [84] Qian C, Cong X, Yang C, Chen W, Su Y, Xu J, Liu Z and Sun M 2023 arXiv preprint arXiv:2307.07924
- [85] Guo F 2023 arXiv preprint arXiv:2305.05516
- [86] Suzuki R and Arita T 2024 Scientific Reports 14 5989
- [87] Fan C, Chen J, Jin Y and He H 2024 Can large language models serve as rational players in game theory? a systematic analysis Proceedings of the AAAI Conference on Artificial Intelligence pp 17960–17967
- [88] Brookins P and DeBacker J M 2023 Available at SSRN 4493398
- [89] Fontana N, Pierri F and Aiello L M 2024 arXiv preprint arXiv:2406.13605
- [90] Capraro V, Di Paolo R, Perc M and Pizziol V 2024 Journal of the Royal Society Interface 21 20230720
- [91] Chan A, Riché M and Clifton J 2023 arXiv preprint arXiv:2303.13360
- [92] Sreedhar K and Chilton L 2024 arXiv preprint arXiv:2402.08189
- [93] Johnson T and Obradovich N 2023 arXiv preprint arXiv:2301.02330
- [94] Xie C, Chen C, Jia F, Ye Z, Shu K, Bibi A, Hu Z, Torr P, Ghanem B and Li G 2024 arXiv preprint arXiv:2402.04559
- [95] Mei Q, Xie Y, Yuan W and Jackson M O 2024 Proceedings of the National Academy of Sciences 121 e2313925121
- [96] McCannon B C 2024 Economics Letters **241** 111828
- [97] Mozikov M, Severin N, Bodishtianu V, Glushanina M, Baklashkin M, Savchenko A V and Makarov I 2024 arXiv preprint arXiv:2406.03299
- [98] Babin J J and Chauhan H 2024 research square preprint https://doi.org/10.21203/rs.3.rs-4462506/v1
- [99] Henry J 2024 arXiv preprint arXiv:2402.05786
- [100] Lorè N and Heydari B 2023 arXiv preprint arXiv:2309.05898
- [101] Xu L, Hu Z, Zhou D, Ren H, Dong Z, Keutzer K, Ng S K and Feng J 2023 Magic: Investigation of large language model powered multi-agent in cognition, adaptability, rationality and collaboration ICLR 2024 Workshop on Large Language Model (LLM) Agents
- [102] Akata E, Schulz L, Coda-Forno J, Oh S J, Bethge M and Schulz E 2023 arXiv preprint arXiv:2305.16867
- [103] Phelps S and Russell Y I 2023 arXiv preprint arXiv:2305.07970
- [104] Heydari B and Lorè N 2023 Contextual Framing (September 10, 2023)
- [105] Duan J, Zhang R, Diffenderfer J, Kailkhura B, Sun L, Stengel-Eskin E, Bansal M, Chen T and Xu K 2024

- $arXiv\ preprint\ arXiv:2402.12348$
- [106] Herr N, Acero F, Raileanu R, Pérez-Ortiz M and Li Z 2024 arXiv preprint arXiv:2407.04467
- [107] Roberts J, Moore K and Fisher D 2024 arXiv preprint arXiv:2404.08710
- [108] Li J, Li R and Liu Q 2023 arXiv preprint arXiv:2309.04369
- [109] Huang J t, Li E J, Lam M H, Liang T, Wang W, Yuan Y, Jiao W, Wang X, Tu Z and Lyu M R 2024 arXiv preprint arXiv:2403.11807
- [110] Oosterbeek H, Sloof R and Van De Kuilen G 2004 Experimental economics 7 171–188
- [111] Henrich J, Boyd R, Bowles S, Camerer C, Fehr E and Gintis H 2004 Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies (Oxford University Press) ISBN 9780199262052 URL https://doi.org/10.1093/ 0199262055.001.0001
- [112] Engel C 2011 Experimental economics 14 583–610
- [113] Dal Bó P and Fréchette G R 2019 American Economic Review 109 3929–3952
- [114] Mengel F 2018 The Economic Journal 128 3182–3209
- [115] Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, Zhang C, Agarwal S, Slama K, Ray A et al. 2022 Advances in neural information processing systems 35 27730-27744
- [116] Gao C, Lan X, Lu Z, Mao J, Piao J, Wang H, Jin D and Li Y 2023 arXiv preprint arXiv:2307.14984
- [117] GLM T, Zeng A, Xu B, Wang B, Zhang C, Yin D, Rojas D, Feng G, Zhao H, Lai H et al. 2024 arXiv preprint arXiv:2406.12793
- [118] Buchanan B, Lohn A, Musser M and Sedova K 2021 Center for Security and Emerging technology 1 2
- [119] Huang L, Yu W, Ma W, Zhong W, Feng Z, Wang H, Chen Q, Peng W, Feng X, Qin B et al. 2023 arXiv preprint arXiv:2311.05232
- [120] Pennycook G, McPhetres J, Zhang Y, Lu J G and Rand D G 2020 Psychological science 31 770–780
- [121] Kriplean T, Toomim M, Morgan J, Borning A and Ko A J 2012 Is this what you meant? promoting listening on the web with reflect ACM Conferences (New York, NY, USA: Association for Computing Machinery) pp 1559–1568
- [122] Kim S, Eun J, Seering J and Lee J 2021 Proc. ACM Hum.-Comput. Interact. 5 1–26
- [123] Argyle L P, Bail C A, Busby E C, Gubler J R, Howe T, Rytting C, Sorensen T and Wingate D 2023 Proceedings of the National Academy of Sciences 120 e2311627120
- [124] Yang J C, Korecki M, Dailisan D, Hausladen C I and Helbing D 2024 arXiv preprint arXiv:2402.01766
- [125] Wang Y, Cai Y, Chen M, Liang Y and Hooi B 2023 arXiv preprint arXiv:2310.13206
- [126] Feng S, Park C Y, Liu Y and Tsvetkov Y 2023 arXiv preprint arXiv:2305.08283
- [127] Zhuge M, Liu H, Faccio F, Ashley D R, Csordás R, Gopalakrishnan A, Hamdi A, Hammoud H A A K, Herrmann V, Irie K et al. 2023 arXiv preprint arXiv:2305.17066
- [128] Liang T, He Z, Jiao W, Wang X, Wang Y, Wang R, Yang Y, Tu Z and Shi S 2023 arXiv preprint arXiv:2305.19118
- [129] Hao R, Hu L, Qi W, Wu Q, Zhang Y and Nie L 2023 arXiv preprint arXiv:2304.12998

- [130] Du Y, Li S, Torralba A, Tenenbaum J B and Mordatch I 2023 arXiv preprint arXiv:2305.14325
- [131] Xiong K, Ding X, Cao Y, Liu T and Qin B 2023 arXiv preprint arXiv:2305.11595
- [132] Wu Q, Bansal G, Zhang J, Wu Y, Zhang S, Zhu E, Li B, Jiang L, Zhang X and Wang C 2023 arXiv preprint arXiv:2308.08155
- [133] Li G, Hammoud H, Itani H, Khizbullin D and Ghanem B 2023 Advances in Neural Information Processing Systems 36 51991–52008
- [134] Wang Z, Mao S, Wu W, Ge T, Wei F and Ji H 2023 arXiv preprint arXiv:2307.05300
- [135] Liu R, Yang R, Jia C, Zhang G, Zhou D, Dai A M, Yang D and Vosoughi S 2023 arXiv preprint arXiv:2305.16960
- [136] Pastor-Satorras R and Vespignani A 2001 Physical review letters 86 3200
- [137] Vespignani A 2018 Nature **558** 528–529
- [138] Williams R, Hosseinichimeh N, Majumdar A and Ghaffarzadegan N 2023 arXiv preprint arXiv:2307.04986
- [139] Barbieri F, Espinosa Anke L and Camacho-Collados J 2022 XLM-T: Multilingual language models in Twitter for sentiment analysis and beyond Proceedings of the Thirteenth Language Resources and Evaluation Conference ed Calzolari N, Béchet F, Blache P, Choukri K, Cieri C, Declerck T, Goggi S, Isahara H, Maegaard B, Mariani J, Mazo H, Odijk J and Piperidis S (Marseille, France: European Language Resources Association) pp 258–266 URL https://aclanthology.org/2022.lrec-1.27
- [140] Candellone E, Aleta A, Ferraz de Arruda H, Meijaard E and Moreno Y 2024 Communications Earth & Environment 5 391
- [141] Müller M, Salathé M and Kummervold P E 2023 Frontiers in artificial intelligence 6 1023281

- [142] Zhou B, Zou L, Mostafavi A, Lin B, Yang M, Gharaibeh N, Cai H, Abedin J and Mandal D 2022 Computers, Environment and Urban Systems 95 101824
- [143] Deiner M S, Deiner N A, Hristidis V, McLeod S D, Doan T, Lietman T M and Porco T C 2024 Journal of Medical Internet Research 26 e49139
- [144] Yan W, Liu H, Wu T, Chen Q, Wang W, Chai H, Wang J, Zhao W, Zhang Y, Zhang R et al. 2024 arXiv preprint arXiv:2406.13890
- [145] Sabry Abdel-Messih M and Kamel Boulos M N 2023  $JMIR\ Medical\ Education\ {f 9}\ e46876$
- [146] Li J, Wang S, Zhang M, Li W, Lai Y, Kang X, Ma W and Liu Y 2024 arXiv preprint arXiv:2405.02957
- [147] Tang X, Zou A, Zhang Z, Zhao Y, Zhang X, Cohan A and Gerstein M 2023 arXiv preprint arXiv:2311.10537
- [148] Kim Y, Park C, Jeong H, Chan Y S, Xu X, McDuff D, Breazeal C and Park H W 2024 arXiv preprint arXiv:2404.15155
- [149] Lee P, Bubeck S and Petro J 2023 New England Journal of Medicine 388 1233–1239
- [150] Pangallo M, Aleta A, del Rio-Chanona R M, Pichler A, Martín-Corral D, Chinazzi M, Lafond F, Ajelli M, Moro E, Moreno Y et al. 2024 Nature Human Behaviour 8 264–275
- [151] Tversky A and Kahneman D 1974 science **185** 1124–1131
- [152] Wang Z, Chiu Y Y and Chiu Y C 2023 arXiv preprint arXiv:2310.05418
- [153] Xi Z, Chen W, Guo X, He W, Ding Y, Hong B, Zhang M, Wang J, Jin S, Zhou E et al. 2023 arXiv preprint arXiv:2309.07864
- [154] 2024 LLM-Agent-Paper-List [Online; accessed 31. Jul. 2024] URL https://github.com/WooooDyy/ LLM-Agent-Paper-List
- [155] Hagendorff T 2023 arXiv preprint arXiv:2303.13988
- [156] Nowak M A 2006 science **314** 1560–1563