A new sociology of humans and machines

Milena Tsvetkova[a], Taha Yasseri[b,c], Niccolo Pescetelli[d], and Tobias Werner[e]

[a] Department of Methodology, London School of Economics and Political Science, London, United Kingdom

[b] School of Sociology, University College Dublin, Dublin, Ireland

[c] Geary Institute for Public Policy, University College Dublin, Dublin, Ireland

[d] Collective Intelligence Lab, New Jersey Institute of Technology, Newark, New Jersey, USA

[e] Center for Humans and Machines, Max Planck Institute for Human Development, Berlin, Germany

Abstract

From fake social media accounts and generative-AI chatbots to trading algorithms and self-driving vehicles, robots, bots, and algorithms are proliferating and permeating our communication channels, social interactions, economic transactions, and transportation arteries. Networks of multiple interdependent and interacting humans and intelligent machines constitute complex social systems where the collective outcomes cannot be deduced from either human or machine behavior alone. Under this paradigm, we review recent research and identify general dynamics and patterns in situations of competition, coordination, cooperation, contagion, and collective decision-making, with context-rich examples from high-frequency trading markets, a social media platform, an open-collaboration community, and a discussion forum. To ensure more robust and resilient human-machine communities, we require a new sociology of humans and machines. Researchers should study these communities using complex-system methods, engineers should explicitly design AI for human-machine and machine-machine interactions, and regulators should govern the ecological diversity and social co-development of humans and machines.

Robotic trains and cars drive us around, auction bots outbid us for purchases, ChatGPT answers our questions, while social media bots feed us with dubious facts and news. Modern society is a complex human-machine social system in which machines are becoming more numerous, human interactions with machines – more frequent, and machine-machine interactions – more consequential. With recent advances in generative AI models, the existential threat of unexplainable and uncontrollable general AI is looming large again[1,2,3]. However, when they are numerous and interdependent, even simple unintelligent artificial agents can produce unintended and potentially undesirable outcomes. If we want to prevent financial crashes, improve road safety, preserve market competition, increase auction market efficiency, and reduce misinformation, it is no longer sufficient to understand humans – we need to consider machines, understand how humans and machines interact, and how the collective behavior of systems of humans and machines can be predicted. We require a new sociology of humans and machines.

This Perspective synthesizes research and ideas related to social systems composed of multiple autonomous yet interacting and interdependent humans and machines such as algorithms, bots, and robots (Fig. 1). Similar to the conceptualizations of socio-technical systems[4], actor-network theory[5,6], cyber-physical social systems[7,8], social machines[9,10,11], and human-machine networks[12,13], our principal assumption is that humans and machines form a single social system. In contrast, we do not approach machines as a single medium or entity – "technology" – but emphasize their multiplicity, independence, and heterogeneity and their interactions[14]. Further, we approach the aggregates as complex social systems where network effects and nonlinear dynamic processes produce collective outcomes that cannot be necessarily deduced from individual preferences and behavior alone[15]. Our conceptualization extends and complements the ecological approach to studying machine behavior[16], the "hybrid collective intelligence perspective"[17], and the budding field of Social AI[18]. We aim to offer a conceptual overview of the topic. Hence, we do not utilize a systematic literature search strategy and instead present selected examples to showcase the new conceptualization[19,20,21].

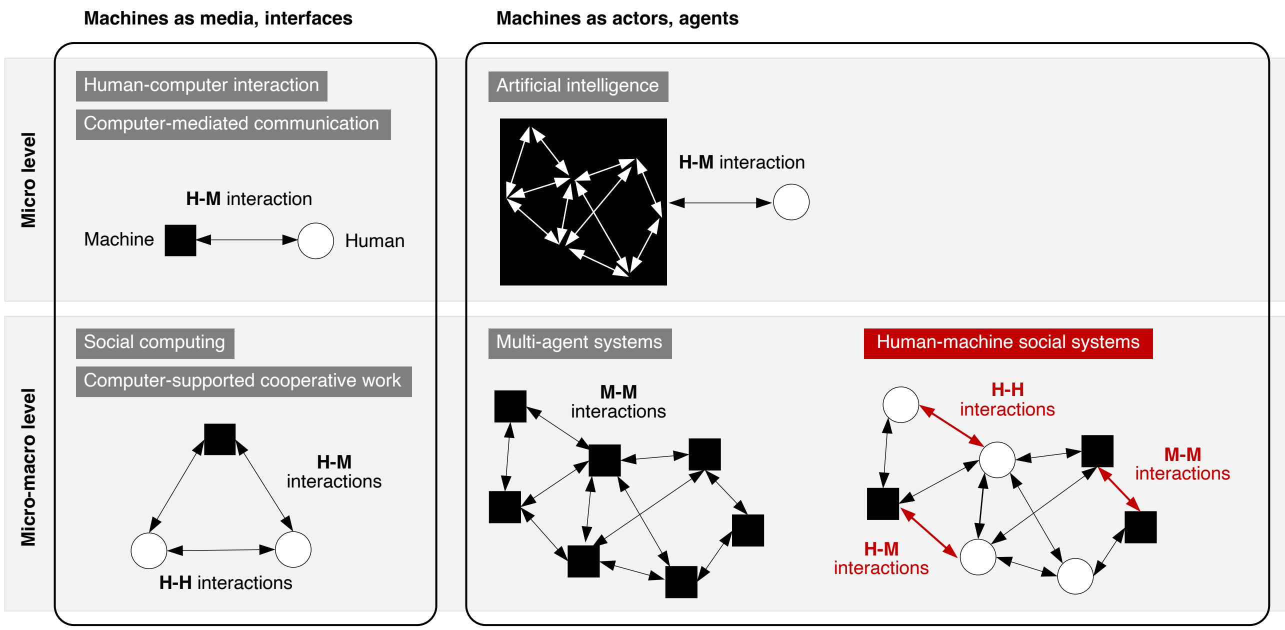

Figure 1: Human-machine social systems include multiple algorithms, bots, or robots that interact among themselves and with humans in groups and networks. Existing fields tend to either approach machines as media or interfaces, not autonomous actors or agents, or focus on their cognition and decision-making, not group interactions with humans. We call for a new sociology of humans and machines to study the human behavior, machine behavior, and the human-human, human-machine, and machine-machine interactions simultaneously in these complex systems.

Human-machine interactions

We use the term "machines" to refer to a range of computational artifacts: embodied in physical devices such as humanoid robots or existing only in digital space, such as bots and algorithms, and varying in sophistication from simple expert systems that use pre-defined if-else rules to generative deep-learning models that learn from data in real-time. Like humans, the machines we consider exhibit diverse goal-oriented behavior shaped by information and subject to constraints. However, the actual cognition and behavior of machines differ from those of humans. Machines' behavior tends to be predictable and persistent[22], with higher precision and faster execution[23], better informed with access to global information[24], and less adaptable and susceptible to influence[25,26,27]. In contrast, humans tend to be limited to local information, satisfice, act with errors, learn and adapt, succumb to social influence and peer pressure, yet also exhibit opinion stubbornness and behavioral inertia; on occasions, they may also use metacognition and revise their own perceptual and decision-making models. Humans often exhibit cognitive biases due to limited information processing capacity, bounded rationality, reliance on heuristics, vestiges of evolutionary adaptation, and emotional motivations[28,29], and algorithms trained on data generated by humans may reproduce these biases[30,31]. Research on human-like general AI aims to erase the cognitive and behavioral differences between humans and machines, while work on human-competitive AI strives for superintelligence that is faster, smarter, and more precise than humans'[32,33]. Either way, humans will remain distinct from machines in the near future.

Research from the CASA (computers as social actors) paradigm in psychology emphasizes that humans treat and respond to machines similarly to other humans: people reciprocate kind acts by computers[34], treat them as politely as they treat humans[35], consider them as competent, but also apply gender and racial stereotypes to them[36,37]; people also humanize and empathize with machines, experiencing distress when witnessing the mistreatment of a robot[38,39].

Nevertheless, there are visible neurophysiological differences in the brain when humans interact with robots[40,41], likely because humans do not attribute agency and morals to them[42,43]. AI is perceived to have lower intentional capacity, lack self-interest, and be more unbiased than humans. Consequently, humans exhibit a narrower emotional spectrum with machines than with humans, reacting with lower and flatter levels of social emotions such as gratitude, anger, pride, and a sense of fairness[44,45,46,47], yet judging machines more harshly when they commit mistakes, cause harm, or incur losses[48,49]. Further, humans behave more rationally and selfishly with machines, cooperating and sharing less and demanding and exploiting more[50,51,52,53]. People would design a machine to be more cooperative than they are themselves[54] but act pro-socially towards it only if it is more human-like[55], or if it benefits another human[56]. Compared to a single person, small groups of people are even more likely to exhibit competitive behavior and bullying toward robots[57]. Despite this intergroup bias, humans are still susceptible to machine influence when making decisions or solving problems[58]. Robots can cause both informational and normative conformity in people[59,60] and AI and ChatGPT can corrupt humans' moral judgment and make them follow unethical advice[61,62]. Humans tend to trust algorithmic advice more than advice coming from a another human or a human crowd[63,64] but may also avoid it if they perceive a threat to their decision control or a lack of understanding and cognitive compatibility[65,66].

Table 1: Types, examples, and collective outcomes of human-machine social systems. Boxes 1-4 present more context for four of the examples: high-frequency trading markets, Twitter, Wikipedia, and Reddit. These four communities are clearly defined, relatively large, and well-studied and exemplify situations of market competition, contagion in political communication, cooperation and coordination, and collective action, respectively.

Situation	**Models**	**General examples**	**H-M examples**	**H-M collective outcomes**
Competition	Zero-sum game Auction markets Buyer-seller markets Oligopoly market games	Competitions Contests Auctions Markets	High-frequency trading markets Pricing algorithms Online auctions with sniping algorithms Cheating bots in multiplayer games	+ Increase efficiency by improving liquidity and price discovery – Increase volatility by causing price spikes and crashes – Increase in consumer prices from algorithmic collusion – Decrease human activity
Coordination	Coordination game Stag Hunt game Battle of the Sexes Chicken game Graph Coloring game	Conventions Technological standards Communication technology Transportation Supply chain management	Traffic with autonomous vehicles Wikipedia editors Humanoid robots in warehouses and manufacturing	+ Improve coordination by introducing random behavior – Worsen coordination by failing to adapt
Cooperation	Prisoner's Dilemma Public Goods game Dictator game Ultimatum game Trust game	Team collaboration Mutual aid	Caring robots Personal AI assistants Chatbots Wikipedia editors Reddit moderators Non-player characters in multiplayer games	+ Increase cooperation + Increase efficiency by handling large task volumes + Increased forecasting accuracy – Decrease efficiency by introducing new types of workload
Contagion	Epidemiological models Threshold models of contagion	Communicative disease Information Innovations	Twitter Reddit	– Increase spread of misinformation, opinion polarization, verbal conflict + Increase human activity and engagement
Collective decision making	Vote aggregation Active learning	Crowdsourcing Prediction markets Voting systems Cultural evolution	Clinical diagnosis Citizen science Content moderation Hybrid forecasting	+ Increase innovation and accuracy by introducing diversity – Decrease human activity and engagement

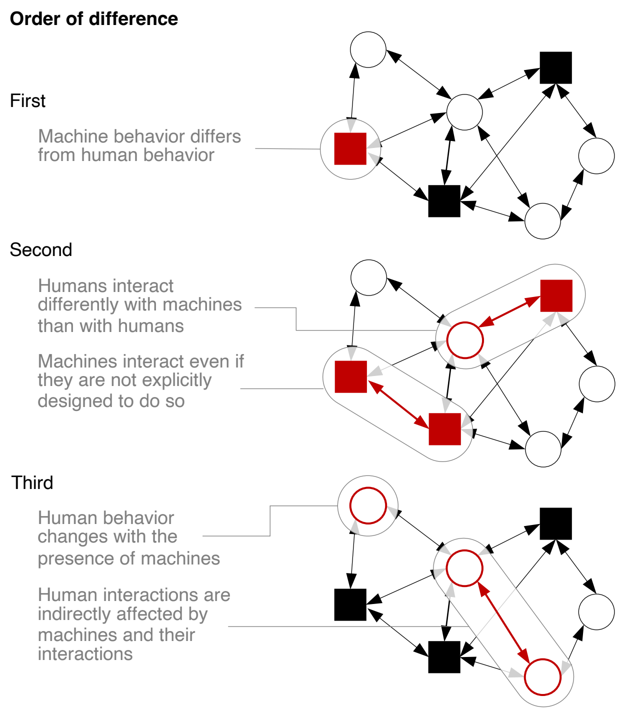

Figure 2: Collective outcomes in human-machine social systems differ from those in human-only systems. Machines behave differently from humans and in social systems with covert artificial agents, even if humans, unaware of the presence of machines, do not change their behavior, the collective outcomes will differ simply because machines act differently. Further, the two types of actors and their interactions are interdependent and influence each other. Thus, suspicion or awareness of machine presence can change human behavior and interacting with a machine and observing machine-machine interactions can influence how humans act toward each other.

Collective outcomes

In complex social systems, individuals' behavior and interactions affect the collective outcomes, but the relationship can be fundamentally different from a simple sum or average[67,68,69,70]. The collective outcomes in human-machine social systems differ from those in human-only systems because machines behave differently from humans, human-machine (H-M) and machine-machine (M-M) interactions differ from human-human (H-H) interactions, but also the humans, the machines, and their interactions influence each other indirectly (Fig. 2). We synthesize common dynamics and patterns in groups and networks of humans and machines for five different social interaction situations (Table 1).

Competition

Competition occurs when multiple actors strive for a common goal that cannot be shared, as is the case in contests, auctions, and product markets. Market algorithms are typically designed to benefit the owner without regard for others or the efficiency and stability of the market, and yet, they may still benefit the collective.

With more advanced data processing, learning, and optimization capabilities than humans, algorithms are better able to discover arbitrage opportunities and, hence, eliminate mispricing and increase liquidity in markets. Experimental studies show that algorithmic traders can increase market efficiency[71], but possibly at the expense of human traders' performance[72,73]. Furthermore, algorithmic traders affect human behavior indirectly: with their presence, they make human traders act more rationally, and thus, reduce strategic uncertainty and confusion in the market[74], reducing price bubbles and bringing prices closer to the fundamental value[75].

While perfectly optimizing arbitrage algorithms eliminate mispricings, neither zero-intelligence algorithms that submit random bids without profit maximization[76] nor profit-maximizing agents that update their beliefs from trading history[77] can improve market quality. Meanwhile, manipulator and spoofing algorithms that act to mislead and influence other traders worsen market efficiency[78]. Algorithms can also reduce the rationality of professional traders and alienate and drive away amateur ones. In an online cryptocurrency marketplace, traders herd after a bot buys, producing larger buying volumes[79], while in a Chinese peer-to-peer lending platform, automated investment piques inefficient investor scrutiny, increasing bidding duration without improving investment return[80]. In online auction markets, naive first-time bidders respond negatively to being outbid by sniping algorithms and become less likely to return to another auction[81]. Snipers, which place last-moment bids[82], work mainly because they exploit the naivety of amateur online bidders, who tend to increase their bids incrementally. However, human lack of rationality has its benefits because squatting (placing a high early bid) deters new entrants[83]. In fact, sniping algorithms yield a negligibly small[83] or non-existent[84] buyer gain, giving them a net negative impact on the marketplace.

In addition to trading and auction markets, pricing algorithms have become widespread in regular product markets[85,86] because they either provide recommendations to human pricing managers[87,88] or entirely dictate pricing for some firms[86,89]. While pricing algorithms can help firms scale and respond to changes in demand, they may also generate anti-competition. Q-learning algorithms learn to set anti-competitive prices without communication in simulations[90,91,92,93], and in experiments, those algorithms are often more collusive than humans in small markets[94] and foster collusion when interacting with humans compared to fully human markets[95]. Observational studies of gasoline markets[89] and e-commerce[96,97] support the experimental evidence. Furthermore, algorithms can weaken competition by providing better demand predictions, thereby stabilizing cartels[98,99,100] or by asymmetries in pricing technologies and commitment[101,102].

The general intuition is that markets with more actors should be more efficient. Thus, one expects enhanced performance from markets populated by algorithms. However, in reality, the beneficial effects of machines are often in balance, crucially depending on the machines' prevalence, decision speed, and information quality[103], as well as the humans' experience and expectations.

Coordination

The problem of coordination requires adopting a strategy identical to or, in some cases, dissimilar from other people's strategies, as when deciding whether to join a protest, agreeing on a convention such as driving on the right-hand side of the road, adopting a communication technology, or avoiding a crowd or traffic congestion[104,105].

In human-machine systems, bots could be used to introduce more randomness and movement to steer human groups toward better solutions. Thus, bots acting with small levels of random noise and placed in central locations in a scale-free network decrease the time to coordination, especially when the solutions are hard to find[106]. The bots reduce unresolvable conflicts not only in their direct interactions but also in indirect H-H interactions, even when the participants are aware that they are interacting with machines. Bots that are trained on human behavior, however, process information less efficiently and adapt slower, causing hybrid groups playing a cooperative group-formation game to perform worse than human-only and bot-only groups[27].

In sum, in situations where a group may get stuck on a suboptimal equilibrium, non-human-like bots may be able to help by jittering the system with randomness and unpredictability. Such simple bots may be more beneficial than bots that superficially imitate human behavior without the ability to learn and adapt.

Cooperation

The problem of cooperation pertains to social dilemma situations where a decision is collectively beneficial but individually costly and risky. Although the economically rational decision in non-repeated anonymous interactions is to free-ride and exploit others' contributions, people's actual behavior tends to be informed by norms of reciprocity, fairness, and honesty signaling. Thus, as a result of millennia of evolutionary adaptation, people generally cooperate with each other. If people know they are interacting with bots, however, they cooperate less[51,52]. Yet, since humans reciprocate to and imitate cooperative neighbors, introducing covert, persistently cooperating bots could increase cooperation.

Computer simulations show that persistent prosocial bots favor the emergence of fairness and cooperation[107,108], with stronger effects when humans are more prone to imitation and bots occupy more central positions in networks with highly heterogeneous connectivity[109].

Just a few under-cover cooperative bots can increase cooperation, especially if the bots are widespread in the network, interacting with many human players rather than concentrated with overlapping sets of partners[110]. The reason is that humans wait for someone else to cooperate before they do, but once they observe many cooperators, they become more likely to exploit.

Yet, cooperative bots may sometimes fail to improve cooperation. For instance, hybrid groups with identifiable bots do not perform better than human-only groups[111]. When participants are aware of the presence of artificial agents but not their identity, there is a small increase in cooperation of the bots' direct neighbors but no significant boost in the overall network[112]. Similarly, multiple well-dispersed covert bots, whether all-cooperating or reciprocating, fail to improve cooperation[113], although a single overt network engineer bot who suggests connecting cooperators and excluding defectors can successfully do so.

In sum, covert, persistently cooperating bots (i.e., not very human-like) can increase cooperation in the group depending on the network of interactions. Bots are successful if they are strategically positioned – well dispersed in regular and random networks or centrally located in networks with skewed degree distributions – or have the power to strategically engineer the network by offering opportunities to break links to defectors.

Contagion

Contagion concerns the spread of information and behaviors, such as memes, slang, fashion, emotions, and opinions in communication networks[114,115,116]. In contrast to the strategic interdependence under the competition, coordination, and cooperation scenarios, the main mechanism here is social influence: the tendency to rely on information from others to handle uncertainty and to conform to the expectations of others to fit in society[117,118,119]. In human-machine systems, bots can be remarkably influential at the collective level despite exerting limited direct influence on individuals because, in networks, small effects can produce chain reactions and trigger cascades[120,121,122,123].

This is how social media bots influence public opinion. In agent-based models of belief formation, weak bots do not alienate their followers and their followers' friends and thus have their message spread farther than messages by more pushy and assertive users[124]. In other words, network amplification occurs through bots' indirect influence precisely because their direct influence on humans is weak, slow, and unobtrusive. If social media bots influence not people's opinion but their confidence to express it, they can amplify marginal voices by triggering the spiral of silence amongst disagreeing humans[26]. The bots are more influential when they are more numerous and connected to central actors. Strategically placed zealot bots can in fact bias voting outcomes and win elections[125].

Bots can also trigger emotional contagion in groups, even though they evoke flatter emotional reactions from individual humans. Humanoid robots can encourage and increase social interactions among older adults within care facilities, between different generations, and for children with ASD[57]. In small-team collaborative experiments, a robot's verbal expressions of vulnerability can show "ripple effects" and make the humans more likely to admit mistakes, console team members, and laugh together[126], engage in social conversations and appreciate the group[127]. The reported positive contagion effects, however, were detected when comparing one machine to another[127,128]. Overall, bots are more effective than no bots to influence opinions, behavior, and emotions, but not necessarily more effective than humans. Yet, even when bots have a weak direct influence on humans' opinions, they can exert significant collective influence via persistence, strategic placement, and sheer numbers.

Collective decision-making

Collective decision-making involves groups making choices or solving problems by combining individual opinions. It impacts social phenomena as diverse as team collaboration, voting, scientific innovation, and cultural evolution[129]. Originating with Galton's work on estimation tasks[130], the "wisdom of crowds" concept suggests that a crowd's aggregated estimate is often more accurate than any individual's, or sometimes even experts'[131]. Crowds perform better when individual opinions are either independent or diverse[132], while social interaction can hinder[133,134,135] or improve[136,137] collective performance. In human-machine systems, algorithms introduce diversity and can thus improve decision-making.

An analysis of professional Go players' moves over 71 years suggests that AlphaGo, the AI program Google DeepMind introduced in 2016, led human players to novel strategies and improved their decision-making[138,139,140]. AlphaGo's decisions, untethered by human bias, sparked human innovation in this game. However, the positive influence of machine-human social learning on problem-solving may be limited. When algorithms are introduced in chains of humans engaged in sequential problem solving, the innovative solutions benefit immediate followers, but team accuracy does not have lasting effects because humans are more likely to replicate human solutions than algorithmic ones[141]. Similarly, in a team prediction task, an algorithm maintaining group diversity by promoting minority opinions improves individual accuracy, but the effects dissipate for team accuracy[142].

The area of hybrid intelligence investigates how and when to combine human and algorithmic decision-making[143,17] and includes research on active and interactive learning and human-in-the-loop algorithms[144], with applications in clinical decision-making, where combining clinician and algorithmic judgments can improve cancer diagnoses[145,146], and citizen science, where combining crowd-based with machine classifications can improve accuracy. On Zooniverse, a prominent citizen science platform, this hybrid approach found supernovae candidates among Pan-STARRS images more effectively than humans or machines alone[147] but damaged citizen scientists' retention[148,149], suggesting a trade-off between efficiency and volunteer engagement. Ultimately, the deployment of machines could further marginalize certain groups of volunteers[150], and with fewer volunteers, AI's performance could diminish.

The emerging field of hybrid intelligence suggests that algorithms introduce novel solutions, but these may be too unfamiliar for humans to adopt. Nevertheless, machine diversity and competition might inspire alternative forms of human creativity and innovation. Developing methods to effectively combine human and machine solutions could further improve collective intelligence[151].

Box 1: Competition in high-frequency trading markets

High-frequency trading (HFT) algorithms constitute automated scripts that rely on high-speed, large-volume transactions to exploit mispricings or market signals before they disappear or are incorporated into the price[152]. The phenomenon started in the mid-90s and has since spread to dominate foreign equities, foreign exchange, commodities, futures, and stock markets globally[103].

HFT algorithms process large amounts of trade history data and current news to make decisions and are thus considered the better "informed" traders[153]. Some of the algorithms appear to anticipate the market, and their trades consistently predict future order flow by human traders[154]. However, since most algorithms react similarly to the same public information, they exhibit less diverse trading strategies and more correlated actions among themselves compared to humans[155]. Thus, although their behavior generally improves market efficiency, it can also trigger behavioral cascades and instability[156].

HFT algorithms generally act as market makers, increasing trading opportunities, reducing transaction costs, connecting buyers and sellers across venues, and submitting significant volumes of price quotes[157,103]. They facilitate price efficiency by trading in the direction of permanent price changes but opposite temporary price errors[153]. This regularly acts as a stabilizing force, reducing short-term volatility[155,158,152]. On a longer time scale, however, HFT algorithms may decease market quality by increasing volatility[159] and uncertainty[23], and by reducing trading strategy diversity[160]. For instance, although the algorithms did not cause the 2010 flash crash, they exacerbated it by amplifying the volatility[161]. This has led to recent efforts to regulate the speed of trading in markets, for example, by processing trades in batches at slower intervals to diminish the advantage that HFT algorithms have[162].

Box 2: Contagion on Twitter

Social bots on the micro-blogging platform Twitter (re-branded as X in 2022) are covert automated accounts designed to impersonate humans to boost followers, disseminate information, and promote products. Bots and bot detection methods have co-evolved, resulting in increasingly more sophisticated imitation or detection strategies[163,164,165,166,25,167,168], but detection is inherently limited due to the overlap between covert autonomous bots, managed user accounts, hacked accounts, cyborgs, sock-puppets, and coordinated botnets[169,170,171,172]. Estimates suggest that 9-15% of Twitter users are bots[173,168], with bot activity typically increasing around controversial political events[174].

Twitter social bots, who do not follow social instincts but neither succumb to fatigue, engage less in social interactions via replies and mentions than humans but produce more content[175]. The bots mainly retweet – a passive strategy to indicate support and gain followers – but are less successful in attracting friends and followers than humans[168]. Overall, they are less connected and bot-bot (2%), bot-human (19%), and human-bot (3%) interactions are considerably less common than human-human interactions (76%)[174].

Despite their rudimentary social behavior and weak network integration, Twitter bots significantly influence political communication, public opinion, elections, and markets. They play an important role in misinformation dissemination in relation to political events[176,177,178,179,180,181,182], COVID-19[183,184], and stock market investment[185]. Bots can affect human interaction networks by encouraging followings and conversations[186] and amplify low-credibility content early on by targeting influential humans through replies and mentions[180]. Bots' large numbers enhance their visibility and influence to trigger deep information cascades[187]. Bots equally link to true and false news from low-credibility sites, but people prefer false content, making humans ultimately responsible for the spread of false news[188].

Twitter bots significantly contribute to negative sentiment and conflict escalation. Acting from the periphery, they target central human users to exert indirect influence. They amplify existing negative sentiment and selectively promote inflammatory content, often targeting only one of the factions[174]. Their success stems from exploiting human tendencies to connect with similar others and engage with messages that reinforce their beliefs[189]. Consequently, bots increase ideological polarization and negatively affect democratic discourse on social media, as seen in the 2016 US presidential election[190], the 2016 UK Brexit Referendum[189], and the 2017 Catalan independence referendum[174].

In sum, Twitter's covert social bots are considered harmful, prompting the platform to cull them[191,192]. Their strength lies in indirect action: they skew the platform's recommendation system to bias content popularity[123] and exploit human behavioral weaknesses like attention seeking, confirmation bias, moral outrage, and ideological homophily.

Box 3: Cooperation and coordination on Wikipedia

Wikipedia, the largest and most popular free-content online encyclopedia, hosts an ecology of bots that generate articles, fix errors on pages, link to other sites and databases, tag articles in categories, identify vandals, notify users, and so on[193,194,195]. These bots are open-source, documented, approved, registered, and tagged[196,193]. They are not sophisticated: most use basic regular expressions or straightforward heuristics, and only some incorporate machine learning techniques. They are significantly less numerous than human editors but complete a disproportionately large volume of all edits[197,22,195]. Compared to human-human interactions, bot-bot interactions are more reciprocal and balanced but do not exhibit status effects[22]. While bots are more likely to be involved in back-and-forth reverts with each other over long periods, these accidental encounters rarely indicate direct opinion conflict but constitute routine productive maintenance work or reflect conflicts existing between their human owners[198]. Human editors interact mainly with policing bots, primarily by criticizing the legitimacy of the norms they enforce, rather than the sanctions themselves, suggesting that editors perceive bots as extensions of their human owners rather than independent agents[199].

Bots have been invaluable to the maintenance and operation of Wikipedia. The diversity of the bot ecology guarantees the system's robustness and resilience. For instance, during the random outage of the anti-vandalism ClueBot NG, the website eventually caught up, albeit more slowly than usual, thanks to the heterogeneity of the quality control network, comprising instantaneous fully automated robots, rapid tool-assisted humans (cyborgs), humans editing via web browsers, and idiosyncratic batch scripts[200]. Wikipedia demonstrates that successful bot governance and regulation does not have to sacrifice distributed development and diversity. The algorithmic simplicity, independence, and heterogeneity of the machines facilitate the system's success and resilience overall but may also introduce unexpected complexities and uncertainties in communication at smaller scales[201].

Box 4: Cooperation and contagion on Reddit

Reddit is a popular news aggregation, content rating, and discussion website founded in 2005. Bots on Reddit provide internal moderation and communication, augment functionality (e.g., for mobile users), or post content, ranging from comic and playful posts by evident automated accounts such as haiku_robot and ObamaRobot, to trolling and provocative comments by undercover social bots[202]. Similarly to Wikipedia, Reddit has developed norms and protocols for deploying bots[203], but similarly to Twitter, it has no effective service limitations to prevent covert and malicious automated accounts. Nevertheless, Reddit differs from Twitter in several crucial ways – content on the site is posted within communities, heavily moderated, up-/downvoted, and extensively discussed. Due to these structural differences, political misinformation, polarization, and conflict are less pronounced on the platform.

Reddit offers evidence for collaboration and contagion between humans and bots. Content moderators have widely adopted Automod – a bot that is flexible to updates, adaptable to community rules, and interpretable. The bot aids with menial tasks but does not necessarily decrease workloads as it requires continuous updates in response to changes in user behavior and language and involves high volumes of correspondence with incorrectly banned users[204]. On the other side, regular users engage with evident entertainment bots, and their direct replies imitate the sentiment and the words of the bot posts[205]. Thus, emotional contagion and lexical entrainment can occur between humans and bots, even when humans are aware of the simple automated script behind the bot.

Implications

The algorithms in current human-machine social systems are relatively simple. Few use sophisticated machine learning or AI, and typically, these guide narrow and specific behaviors[206,207]. Except for malicious social bots and customer-service chatbots, most machines do not mimic human qualities. Most machines are superhuman — processing vast data amounts, acting swiftly, and handling tedious tasks— or candidly non-human— resisting peer influence, not reciprocating, and acting randomly. There is a clear distinction between covert and overt bots: covert bots are more problematic than bots declaring their identity and following norms and regulations.

The effects of machines on human-machine social systems vary by their number, algorithms, network position, interaction situation, institutional regulations, technological affordances, organizational context, and emerging norms and culture (see Boxes 1-4). Machines alter outcomes through their unique behavior because humans interact differently with them, and because of their indirect effects – machines' presence changes how humans interact amongst themselves.

Machines can be beneficial when they act or steer humans to counteract human weaknesses. For instance, noisy bots can disrupt sub-optimal outcomes and improve coordination, persistently cooperative bots can curb retaliation and maintain cooperation, machines in central roles as arbitrageurs improve price discovery and market quality, and network-engineering bots boost collective welfare via cooperator assortment and defector exclusion. With global information, higher processing power, and instantaneous execution, machines can quickly address external events like vandalism or political and natural crises, ensuring system robustness, resilience, and efficiency. Depending on the situation, machines offer superhuman persistence or randomness, norm-setting rationality, or solution diversity, enhancing human behavior towards better outcomes.

However, what helps in one context can hinder in another. Machines' unintuitive solutions may confuse humans, hindering innovation and technological progress. Humans might not act fast enough to correct machines' errors, resulting in instabilities and flash failures. Machines are less adaptive than humans to changes, impeding system evolution. Machines can be designed to exploit human weaknesses, triggering cascades that exacerbate polarization, emotional contagion, ideological segregation, and conflict. Machines' non-human optimization logic, execution speed, and behavioral rigidity can clash with human behavior, pushing interactions toward undesirable outcomes.

Research

Existing research is often biased towards engineering and optimization, lacking deeper insights from a social science perspective. The time for a new sociology of humans and machines is critical, before AI becomes more sophisticated: generative AI exhibits emergent behavior that itself requires explanation[208,209], complicating the understanding of system dynamics.

Researchers would benefit from an agent-based modeling framework that outlines distinctions between human and bot agents: utility function, optimization ability, access to information, learning, innovation/creativity, accuracy, etc. The framework could borrow concepts from other two-agent systems, such as predator–prey, principal–agent, and common pool resource models. Controlled experiments should explicitly compare human-machine, human-only and machine-only networks, and known bots against covert bots. Experiments could manipulate participants' perceptions of algorithms' technical specifications, agenthood[210], emotional capability, and biases. Field interventions in online communities with endemic bot populations present another promising direction. Existing examples include social bots that gain influence by engaging human users[211,212,213,214], trading bots that manipulate prices in cryptocurrency markets[215], political bots that promote opposing political views to decrease polarization[216], and "drifters" to measure platform bias[80]. Expanding on the cases reported here, we need observational research on additional human-machine communities and contexts such as traffic systems with human-driven and driverless vehicles, online multiplayer games comprising human players, non-player characters, and cheating code, and dating markets with AI-driven chatbots[217].

Finally, research with artificial agents introduces ethical problems demanding careful elaboration and mitigation. Research protocols should minimize interventions[218], possibly deploying covert bots only where they already exist, ensuring their actions are not unusual or harmful[79]. Even then, bots may still face opposition from users due to privacy concerns[211]. Do people perceive certain bots as inherently deceptive? Could knowledge of the bot owner and algorithm mitigate this perception?

Design

Humans are resilient and successful when they exhibit high levels of efficient communication, context awareness, emotional recognition and response, and ethical and cultural sensitivities. These features should be encouraged when designing social machines to build trust, ensure legal compliance, promote social harmony, enhance user satisfaction, and achieve long-term sustainability[219,220].

Since H-H, H-M, and M-M interactions differ, machines should be specifically designed for each scenario, with separate training for each interaction. This could avoid market underperformance when bargaining algorithms trained on human-only markets adapt poorly to human-machine negotiations[50], or traffic jams when driverless vehicles trained on human driving fail to properly interact with each other[221].

AI design should also adopt a hierarchy of behavioral rules and conventions guiding H-M and M-M interactions in the context of H-H interactions. Isaac Asimov's famous Three Laws of Robotics[222] which regulate M-H interactions and self-preservation—1) not harming humans, 2) obeying humans, and 3) protecting own existence, with priority given to higher order rules—could be adapted for M-M interactions, considering specific contexts, implications, and unintended consequences[223].

Further, cultural context is crucial for AI design. People's perceptions of machines vary by age, environment, personality, and geography[224,46]. Machines reflect their developers' culture and operate in settings with specific organizational and community norms[22]. Thus, AI design for self-driving cars and assistant bots should consider local driving culture and attitudes toward domestic assistants during training.

Finally, the systems we reviewed exhibit complexity in that their emergent behavior transcends a mere aggregation of individual components, displaying what is known as "network effects." While natural complex systems often demonstrate remarkable resilience and adaptivity[225], the human-designed complex systems discussed here, albeit to varying degrees, are not inherently adaptive. To enhance resilience and robustness, AI designers should incorporate complex adaptive system principles, like negative feedback, modularity, and hierarchical organization. For example, dense networks lacking diversity and modularity are susceptible to systemic failures[226,227,228] and are easy to control[229]. Introducing bots that increase network diversity, introduce resistance, build resilience, or incorporate negative feedback could enhance adaptability and stability. Such configurations have improved group outcomes in forecasting[230], exploration/exploitation[231], and general knowledge tasks[137]. Literature on monitoring and steering complex systems can help predict critical transitions[232,233,15], designing safer human-machine systems across different environments and perturbations.

Policy

We urge a system-focused approach to AI policy and ethics: policymakers should approach AI not as a single existential threat but as a multiplicity of machines and algorithms. Machines are often more beneficial when they are superhuman or simply "alien" and when they are diverse. Similarity in information sources, interaction speeds, optimization algorithms, and objective functions can cause catastrophic events, like flash crashes in markets. Thus, while AI designers may chase optimization and superintelligence, policymakers should focus on the diversity of human-machine ecologies. Policymakers should demand adaptivity and resilience, too.

Policymakers should also anticipate the social co-development of machines and humans, which will inadvertently change existing institutions. Machines can cause humans to withdraw interaction: for instance, outsourcing care to robots reduces caregivers' empathy[234]. Intelligent machines are changing the transmission and creation of human culture, altering social learning dynamics, and generating new game strategies, scientific discoveries, and art forms[235]. Humans must adapt to intelligent machines just as intelligent machines must learn from and adapt to humans. Finally, ethicists should address questions such as: Should all machines be equal? Should we allow status hierarchies, possibly reflecting and exacerbating existing socio-economic inequalities?

Conclusion

This Perspective synthesizes relatively disparate literature based on agent-based models, controlled experiments, online field interventions, and observational analyses from human-computer interaction, robotics, web science, financial economics, and computational social science under a common theoretical framework: human-machine social systems. We identify common dynamics and patterns that emerge from the interactions of humans and intelligent machines regardless of the specific context, as well as peculiarities and unique problems that concrete techno-organizational and socio-cultural environments generate. Our utmost ambition is to stimulate cumulative empirically driven and mechanism-focused sociological research in the emerging, fast-evolving field of human-AI science. At stake are new and urgent social challenges such as online misinformation, market flash crashes, cybersecurity, labor market resilience, and road safety. With increasing social connectivity and accelerating developments in AI, understanding the complex interactions between humans and intelligent machines is a challenging undertaking, but one that is crucially important for a better human future.

Acknowledgments

MT acknowledges the generous support of the Santa Fe Institute during the period when the research was conducted. TY was partially funded by the Irish Research Council under grant number IRCLA/2022/3217, ANNETTE (Artificial Intelligence Enhanced Collective Intelligence). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] Metz, C. The new chat bots could change the world. Can you trust them? *New York Times* (2022).

[2] Milmo, D. & Stacey, K. 'It's not clear we can control it': What they said at the Bletchley Park AI summit. *Guardian* (2023).

[3] Lipton, E. As A.I.-Controlled Killer Drones Become Reality, Nations Debate Limits. *New York Times* (2023).

[4] Emery, F. Characteristics of socio-technical systems. In *Characteristics of Socio-Technical Systems*, 157–186 (Univ. of Pennsylvania Press, 2016).

[5] Latour, B. *Reassembling the Social: An Introduction to Actor-Network-Theory* (Oxford Univ. Press, 2007).

[6] Law, J. Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity. *Syst. Practice* **5**, 379–393 (1992).

[7] Sheth, A., Anantharam, P. & Henson, C. Physical-cyber-social computing: An early 21st century approach. *IEEE Intel. Syst.* **28**, 78–82 (2013).

[8] Wang, F.-Y. The emergence of intelligent enterprises: From CPS to CPSS. *IEEE Intel. Syst.* **25**, 85–88 (2010).

[9] Hendler, J. & Berners-Lee, T. From the Semantic Web to social machines: A research challenge for AI on the World Wide Web. *Artif. Intel.* **174**, 156–161 (2010).

[10] Buregio, V., Meira, S. & Rosa, N. Social machines: A unified paradigm to describe social web-oriented systems. In *Proc. 22nd International Conference on World Wide Web*, 885–890 (ACM, 2013).

[11] Shadbolt, N. R. *et al.* Towards a classification framework for social machines. In *Proc. 22nd International Conference on World Wide Web*, 905–912 (ACM, 2013).

[12] Eide, A. W. *et al.* Human-machine networks: Towards a typology and profiling framework. In Kurosu, M. (ed.) *Human-Computer Interaction. Theory, Design, Development and Practice*, 9731, 11–22 (Springer International Publishing, 2016).

[13] Tsvetkova, M. *et al.* Understanding human-machine networks: A cross-disciplinary survey. *ACM Comput. Surv.* **50** (2017).

[14] Cavallaro, L. *et al.* Mining the network behavior of bots. *Tech. Rep. 2009-12* (2009).

[15] Bianconi, G. *et al.* Complex systems in the spotlight: Next steps after the 2021 Nobel Prize in Physics. *J. Phys. Complex.* **4**, 010201 (2023).

[16] Rahwan, I. *et al.* Machine behaviour. *Nature* **568**, 477–486 (2019).

[17] Peeters, M. M. M. *et al.* Hybrid collective intelligence in a human–AI society. *AI Soc.* **36**, 217–238 (2021).

[18] Pedreschi, D. *et al.* Social AI and the challenges of the human-AI ecosystem. Preprint at *arXiv* `https://doi.org/10.48550/arXiv.2306.13723` (2023).

[19] Paré, G. & Kitsiou, S. Chapter 9: Methods for literature reviews. In Lau, F. & Kuziemsky, C. (eds.) *Handbook of eHealth Evaluation: An Evidence-based Approach*, 157–178 (Univ. of Victoria, 2017).

[20] Sylvester, A., Tate, M. & Johnstone, D. Beyond synthesis: Re-presenting heterogeneous research literature. *Behav. Inf. Tech.* **32**, 1199–1215 (2013).

[21] Whittemore, R. & Knafl, K. The integrative review: Updated methodology. *J. Adv. Nurs.* **52**, 546–553 (2005).

[22] Tsvetkova, M., García-Gavilanes, R., Floridi, L. & Yasseri, T. Even good bots fight: The case of Wikipedia. *PLoS ONE* **12**, e0171774 (2017).

[23] Hilbert, M. & Darmon, D. How complexity and uncertainty grew with algorithmic trading. *Entropy* **22**, E499 (2020).

[24] Koren, Y., Rendle, S. & Bell, R. Advances in collaborative filtering. In Ricci, F., Rokach, L. & Shapira, B. (eds.) *Recommender Systems Handbook*, 91–142 (Springer US, 2022).

[25] Ferrara, E., Varol, O., Davis, C., Menczer, F. & Flammini, A. The rise of social bots. *Commun. ACM* **59**, 96–104 (2016).

[26] Ross, B. *et al.* Are social bots a real threat? An agent-based model of the spiral of silence to analyse the impact of manipulative actors in social networks. *Eur. J. Inf. Syst.* **28**, 394–412 (2019).

[27] Takko, T., Bhattacharya, K., Monsivais, D. & Kaski, K. Human-agent coordination in a group formation game. *Sci. Rep.* **11**, 10744 (2021).

[28] Gilovich, T. *How We Know What Isn't So* (Simon and Schuster, 2008).

[29] Kahneman, D. *Thinking, Fast and Slow* (Penguin, 2012).

[30] Kordzadeh, N. & Ghasemaghaei, M. Algorithmic bias: Review, synthesis, and future research directions. *Eur. J. Inf. Syst.* **31**, 388–409 (2022).

[31] O'Neil, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown Publishing Group, 2016).

[32] Russell, S. & Norvig, P. *Artificial Intelligence: A Modern Approach* (Pearson India Education Services Private Limited, 2022).

[33] Tegmark, M. *Life 3.0: Being Human in the Age of Artificial Intelligence* (Allen Lane, 2017).

[34] Fogg, BJ. & Nass, C. How users reciprocate to computers: An experiment that demonstrates behavior change. In *CHI '97 Extended Abstracts on Human Factors in Computing Systems*, 331–332 (ACM, 1997).

[35] Nass, C., Steuer, J. & Tauber, E. R. Computers are social actors. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, 72–78 (1994).

[36] Nass, C. & Moon, Y. Machines and mindlessness: Social responses to computers. *J. Soc. Issues* **56**, 81–103 (2000).

[37] Siegel, M., Breazeal, C. & Norton, M. I. Persuasive robotics: The influence of robot gender on human behavior. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2563–2568 (2009).

[38] Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S. & Eimler, S. C. An experimental study on emotional reactions towards a robot. *Int. J. Soc. Robot.* **5**, 17–34 (2013).

[39] Slater, M. *et al.* A virtual reprise of the Stanley Milgram obedience experiments. *PLOS ONE* **1**, e39 (2006).

[40] Krach, S. *et al.* Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLOS ONE* **3**, e2597 (2008).

[41] McCabe, K., Houser, D., Ryan, L., Smith, V. & Trouard, T. A functional imaging study of cooperation in two-person reciprocal exchange. *P. Natl. Acad. Sci.* **98**, 11832–11835 (2001).

[42] Gray, H. M., Gray, K. & Wegner, D. M. Dimensions of mind perception. *Science* **315**, 619–619 (2007).

[43] Zhang, J., Conway, J. & Hidalgo, C. A. Why do people judge humans differently from machines? The role of agency and experience. Preprint at *arXiv* `https://doi.org/10.48550/arXiv.2210.10081` (2022).

[44] Adam, M. T. P., Teubner, T. & Gimpel, H. No rage against the machine: How computer agents mitigate human emotional processes in electronic negotiations. *Group Decis. Negot.* **27**, 543–571 (2018).

[45] Chugunova, M. & Sele, D. We and It: An interdisciplinary review of the experimental evidence on how humans interact with machines. *J. Behav. Exp. Econ.* **99**, 101897 (2022).

[46] Hidalgo, C. A., Orghian, D., Canals, J. A., Almeida, F. D. & Martin, N. *How Humans Judge Machines* (MIT Press, 2021).

[47] Schniter, E., Shields, T. W. & Sznycer, D. Trust in humans and robots: Economically similar but emotionally different. *J. Econ. Psychol.* **78**, 102253 (2020).

[48] Dietvorst, B. J., Simmons, J. P. & Massey, C. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *J. Exp. Psych. Gen.* **144**, 114–126 (2015).

[49] Candrian, C. & Scherer, A. Rise of the machines: Delegating decisions to autonomous AI. *Comput. Hum. Behav.* **134**, 107308 (2022).

[50] Erlei, A., Das, R., Meub, L., Anand, A. & Gadiraju, U. For what it's worth: Humans overwrite their economic self-interest to avoid bargaining with AI systems. In *Proc. 2022 CHI Conference on Human Factors in Computing Systems*, 1–18 (ACM, 2022).

[51] Ishowo-Oloko, F. *et al.* Behavioural evidence for a transparency–efficiency tradeoff in human–machine cooperation. *Nat. Mach. Intel.* **1**, 517–521 (2019).

[52] Karpus, J., Krüger, A., Verba, J. T., Bahrami, B. & Deroy, O. Algorithm exploitation: Humans are keen to exploit benevolent AI. *iScience* **24**, 102679 (2021).

[53] March, C. Strategic interactions between humans and artificial intelligence: Lessons from experiments with computer players. *J. Econ. Psychol.* **87**, 102426 (2021).

[54] de Melo, C. M., Marsella, S. & Gratch, J. Human cooperation when acting through autonomous machines. *P. Natl. Acad. Sci.* **116**, 3482–3487 (2019).

[55] Oliveira, R., Arriaga, P., Santos, F. P., Mascarenhas, S. & Paiva, A. Towards prosocial design: A scoping review of the use of robots and virtual agents to trigger prosocial behaviour. *Comput. Hum. Behav.* **114**, 106547 (2021).

[56] Hayes, B., Ullman, D., Alexander, E., Bank, C. & Scassellati, B. People help robots who help others, not robots who help themselves. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 255–260 (2014).

[57] Sebo, S., Stoll, B., Scassellati, B. & Jung, M. F. Robots in groups and teams: A literature review. *P. ACM Hum.-Comput. Interact.* **4**, 176:1–176:36 (2020).

[58] Köbis, N., Bonnefon, J.-F. & Rahwan, I. Bad machines corrupt good morals. *Nat. Hum. Behav.* **5**, 679–685 (2021).

[59] Salomons, N., van der Linden, M., Sebo, S. S. & Scassellati, B. Humans conform to robots: Disambiguating trust, truth, and conformity. In *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 187–195 (2018).

[60] Salomons, N., Sebo, S. S., Qin, M. & Scassellati, B. A minority of one against a majority of robots: Robots cause normative and informational conformity. *ACM T. Hum.-Robot Interact.* **10**, 1–22 (2021).

[61] Leib, M., Köbis, N. C., Rilke, R. M., Hagens, M. & Irlenbusch, B. The corruptive force of AI-generated advice. Preprint at *arXiv* `https://doi.org/10.48550/arXiv.2102.07536` (2021).

[62] Krügel, S., Ostermaier, A. & Uhl, M. ChatGPT's inconsistent moral advice influences users' judgment. *Sci. Rep.* **13**, 4569 (2023).

[63] Bogert, E., Schecter, A. & Watson, R. T. Humans rely more on algorithms than social influence as a task becomes more difficult. *Sci. Rep.* **11**, 8028 (2021).

[64] Logg, J. M., Minson, J. A. & Moore, D. A. Algorithm appreciation: People prefer algorithmic to human judgment. *Organ. Behav. Hum. Decis.* **151**, 90–103 (2019).

[65] Burton, J. W., Stein, M.-K. & Jensen, T. B. A systematic review of algorithm aversion in augmented decision making. *J. Behav. Decis. Making* **33**, 220–239 (2020).

[66] Mahmud, H., Islam, A. K. M. N., Ahmed, S. I. & Smolander, K. What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technol. Forecast. Soc.* **175**, 121390 (2022).

[67] Axelrod, R. *The Evolution of Cooperation* (Basic Books, 1984).

[68] Schelling, T. C. Dynamic models of segregation. *J. Math. Sociol.* **1**, 143–186 (1971).

[69] Granovetter, M. Threshold models of collective behavior. *Am. J. Sociol.* **83**, 1420–1443 (1978).

[70] Miller, J. H. & Page, S. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life* (Princeton Univ. Press, 2009).

[71] Grossklags, J. & Schmidt, C. Software agents and market (in) efficiency: A human trader experiment. *IEEE T. Syst. Man. Cyb. C* **36**, 56–67 (2006).

[72] Angerer, M., Neugebauer, T. & Shachat, J. Arbitrage bots in experimental asset markets. *J. Econ. Behav. Organ.* **206**, 262–278 (2023).

[73] Cartlidge, J., De Luca, M., Szostek, C. & Cliff, D. Too fast too furious: Faster financial-market trading agents can give less efficient markets. *Proc. Fourth International Conference on Agents and Artificial Intelligence* **2**, 126–135 (2012).

[74] Akiyama, E., Hanaki, N. & Ishikawa, R. It is not just confusion! Strategic uncertainty in an experimental asset market. *Econ. J.* **127**, F563–F580 (2017).

[75] Farjam, M. & Kirchkamp, O. Bubbles in hybrid markets: How expectations about algorithmic trading affect human trading. *J. Econ. Behav. Organ.* **146**, 248–269 (2018).

[76] Gode, D. K. & Sunder, S. Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *J. Polit. Econ.* **101**, 119–137 (1993).

[77] Gjerstad, S. The competitive market paradox. *J. Econ. Dyn. Control* **31**, 1753–1780 (2007).

[78] Bao, T., Nekrasova, E., Neugebauer, T. & Riyanto, Y. E. *Algorithmic Trading in Experimental Markets with Human Traders: A Literature Survey* (Edward Elgar Publishing, 2022).

[79] Krafft, P. M., Macy, M. & Pentland, A. S. Bots as virtual confederates: Design and ethics. In *Proc. 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 183–190 (ACM, 2017).

[80] Chen, C., Li, G., Fan, L. & Qin, J. The impact of automated investment on peer-to-peer lending: Investment behavior and platform efficiency. *J. Glob. Inf. Manag.* **29**, 1–22 (2021).

[81] Backus, M., Blake, T., Masterov, D. V. & Tadelis, S. Is sniping a problem for online auction markets? In *Proc. 24th International Conference on World Wide Web*, 88–96 (ACM, 2015).

[82] Roth, A. E. & Ockenfels, A. Last-minute bidding and the rules for ending second-price auctions: Evidence from eBay and Amazon auctions on the Internet. *Am. Econ. Rev.* **92**, 1093–1103 (2002).

[83] Ely, J. C. & Hossain, T. Sniping and squatting in auction markets. *Am. Econ. J. Microecon.* **1**, 68–94 (2009).

[84] Gray, S. & Reiley, D. H. Measuring the benefits to sniping on eBay: Evidence from a field experiment. *J. Econ. Manage.* **9**, 137–152 (2013).

[85] Aparicio, D. & Misra, K. *Artificial intelligence and pricing*, vol. 20 (Emerald Publishing Limited, 2023).

[86] Chen, L., Mislove, A. & Wilson, C. An empirical analysis of algorithmic pricing on Amazon marketplace. In *Proc. 25th international conference on World Wide Web*, 1339–1349 (ACM, 2016).

[87] Garcia, D., Tolvanen, J. & Wagner, A. K. Strategic responses to algorithmic recommendations: Evidence from hotel pricing. Preprint at *SSRN* `http://dx.doi.org/10.2139/ssrn.4676767` (2023).

[88] Hunold, M. & Werner, T. Algorithmic price recommendations and collusion: Experimental evidence. Preprint at *SSRN* `http://dx.doi.org/10.2139/ssrn.4557050` (2023).

[89] Assad, S., Clark, R., Ershov, D. & Xu, L. Algorithmic pricing and competition: Empirical evidence from the German retail gasoline market. *J. Polit. Econ.* **132**, 723–771 (2024).

[90] Calvano, E., Calzolari, G., Denicolò, V. & Pastorello, S. Artificial intelligence, algorithmic pricing, and collusion. *Am. Econ. Rev.* **110**, 3267–3297 (2020).

[91] Calvano, E., Calzolari, G., Denicolò, V., Harrington, J. E. & Pastorello, S. Protecting consumers from collusive prices due to AI. *Science* **370**, 1040–1042 (2020).

[92] Klein, T. Autonomous algorithmic collusion: Q-learning under sequential pricing. *RAND J. Econ.* **52**, 538–558 (2021).

[93] Johnson, J. P., Rhodes, A. & Wildenbeest, M. Platform design when sellers use pricing algorithms. *Econometrica* **91**, 1841–1879 (2023).

[94] Werner, T. Algorithmic and human collusion. Preprint at *SSRN* `http://dx.doi.org/10.2139/ssrn.3960738` (2022).

[95] Normann, H.-T. & Sternberg, M. Human-algorithm interaction: Algorithmic pricing in hybrid laboratory markets. *Eur. Econ. Rev.* **152**, 104347 (2023).

[96] Musolff, L. Algorithmic pricing facilitates tacit collusion: Evidence from e-commerce. In *Proc. 23rd ACM Conference on Economics and Computation*, 32–33 (2022).

[97] Wieting, M. & Sapi, G. Algorithms in the marketplace: An empirical analysis of automated pricing in e-commerce. Preprint at *SSRN* `http://dx.doi.org/10.2139/ssrn.3945137` (2021).

[98] Miklós-Thal, J. & Tucker, C. Collusion by algorithm: Does better demand prediction facilitate coordination between sellers? *Manag. Sci.* **65**, 1552–1561 (2019).

[99] O'Connor, J. & Wilson, N. E. Reduced demand uncertainty and the sustainability of collusion: How AI could affect competition. *Inf. Econ. Policy* **54**, 100882 (2021).

[100] Martin, S. & Rasch, A. Demand forecasting, signal precision, and collusion with hidden actions. *Int. J. Ind. Organ.* **92**, 103036 (2024).

[101] Brown, Z. Y. & MacKay, A. Competition in pricing algorithms. Preprint at *National Bureau of Economic Research* `https://doi.org/10.3386/w28860` (2021).

[102] Leisten, M. Algorithmic competition, with humans. Preprint at *SSRN* `https://papers.ssrn.com/abstract=4733318` (2024).

[103] Menkveld, A. J. The economics of high-frequency trading: Taking stock. *Annu. Rev. Financ. Econ.* **8**, 1–24 (2016).

[104] Ullmann-Margalit, E. *The Emergence of Norms* (Oxford Univ. Press, 2015).

[105] Young, H. P. The evolution of conventions. *Econometrica* **61**, 57–84 (1993).

[106] Shirado, H. & Christakis, N. A. Locally noisy autonomous agents improve global human coordination in network experiments. *Nature* **545**, 370–374 (2017).

[107] Santos, F. P., Pacheco, J. M., Paiva, A. & Santos, F. C. Evolution of collective fairness in hybrid populations of humans and agents. *P. AAAI Conf. Artif. Intel.* **33**, 6146–6153 (2019).

[108] Sharma, G., Guo, H., Shen, C. & Tanimoto, J. Small bots, big impact: Solving the conundrum of cooperation in optional Prisoner's Dilemma game through simple strategies. *J. Roy. Soc. Interface* **20**, 20230301 (2023).

[109] Shen, C., He, Z., Shi, L., Wang, Z. & Tanimoto, J. Simple bots breed social punishment in humans. Preprint at *arXiv* `https://doi.org/10.48550/arXiv.2211.13943` (2022).

[110] Suri, S. & Watts, D. J. Cooperation and contagion in web-based, networked public goods experiments. *PLOS ONE* **6**, e16836 (2011).

[111] Fernández Domingos, E. *et al.* Delegation to artificial agents fosters prosocial behaviors in the collective risk dilemma. *Sci. Rep.* **12**, 8492 (2022).

[112] Kirchkamp, O. & Nagel, R. Naive learning and cooperation in network experiments. *Game. Econ. Behav.* **58**, 269–292 (2007).

[113] Shirado, H. & Christakis, N. A. Network engineering using autonomous agents increases cooperation in human groups. *iScience* **23**, 101438 (2020).

[114] Centola, D. *How Behavior Spreads: The Science of Complex Contagions* (Princeton Univ. Press, 2018).

[115] Christakis, N. A. & Fowler, J. H. Social contagion theory: Examining dynamic social networks and human behavior. *Stat. Med.* **32**, 556–577 (2013).

[116] Rogers, E. M. *Diffusion of Innovations* (Simon and Schuster, 2003).

[117] Cialdini, R. B. *Influence: Science and Practice* (Allyn & Bacon, 2008).

[118] Deutsch, M. & Gerard, H. B. A study of normative and informational social influences upon individual judgment. *J. Abnorm. Soc. Psych.* **51**, 629–636 (1955).

[119] Turner, J. C. *Social Influence* (Thomson Brooks/Cole Publishing Co, 1991).

[120] Fowler, J. H. & Christakis, N. A. Cooperative behavior cascades in human social networks. *P. Natl. Acad. Sci.* **107**, 5334–5338 (2010).

[121] Leskovec, J., Adamic, L. A. & Huberman, B. A. The dynamics of viral marketing. *ACM T. Web* **1**, 5–es (2007).

[122] Watts, D. J. A simple model of global cascades on random networks. *P. Natl. Acad. Sci.* **99**, 5766–5771 (2002).

[123] Pescetelli, N., Barkoczi, D. & Cebrian, M. Bots influence opinion dynamics without direct human-bot interaction: The mediating role of recommender systems. *Appl. Net. Sci.* **7**, 1–19 (2022).

[124] Keijzer, M. A. & Mäs, M. The strength of weak bots. *Online Soc. Net. Media* **21**, 100106 (2021).

[125] Stewart, A. J. *et al.* Information gerrymandering and undemocratic decisions. *Nature* **573**, 117–121 (2019).

[126] Strohkorb Sebo, S., Traeger, M., Jung, M. & Scassellati, B. The ripple effects of vulnerability: The effects of a robot's vulnerable behavior on trust in human-robot teams. In *Proc. 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 178–186 (ACM, 2018).

[127] Traeger, M. L., Strohkorb Sebo, S., Jung, M., Scassellati, B. & Christakis, N. A. Vulnerable robots positively shape human conversational dynamics in a human–robot team. *P. Natl. Acad. Sci.* **117**, 6370–6375 (2020).

[128] Zhang, A. W., Lin, T.-H., Zhao, X. & Sebo, S. Ice-breaking technology: Robots and computers can foster meaningful connections between strangers through in-person conversations. In *Proc. 2023 CHI Conference on Human Factors in Computing Systems*, 1–14 (ACM, 2023).

[129] Bang, D. & Frith, C. D. Making better decisions in groups. *Roy. Soc. Open Sci.* **4**, 170193 (2017).

[130] Galton, F. Vox Populi. *Nature* **75**, 450–451 (1907).

[131] Surowiecki, J. *The Wisdom of Crowds* (Anchor, 2005).

[132] Page, S. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies* (Princeton Univ. Press, 2008).

[133] Frey, V. & van de Rijt, A. Social influence undermines the wisdom of the crowd in sequential decision making. *Manag. Sci.* **67**, 4273–4286 (2021).

[134] Lorenz, J., Rauhut, H., Schweitzer, F. & Helbing, D. How social influence can undermine the wisdom of crowd effect. *P. Natl. Acad. Sci.* **108**, 9020–9025 (2011).

[135] Muchnik, L., Aral, S. & Taylor, S. J. Social influence bias: A randomized experiment. *Science* **341**, 647–651 (2013).

[136] Becker, J., Brackbill, D. & Centola, D. Network dynamics of social influence in the wisdom of crowds. *P. Natl. Acad. Sci.* **114**, E5070–E5076 (2017).

[137] Navajas, J., Niella, T., Garbulsky, G., Bahrami, B. & Sigman, M. Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds. *Nat. Hum. Behav.* **2**, 126–132 (2018).

[138] Choi, S., Kang, H., Kim, N. & Kim, J. How does AI improve human decision-making? Evidence from the AI-powered Go program. Preprint at *SSRN* `http://dx.doi.org/10.2139/ssrn.3893835` (2023).

[139] Shin, M., Kim, J. & Kim, M. Human learning from Artificial Intelligence: Evidence from human Go players' decisions after AlphaGo. *P. Annu. Meet. Cogn. Sci. Soc.* **43** (2021).

[140] Shin, M., Kim, J., van Opheusden, B. & Griffiths, T. L. Superhuman artificial intelligence can improve human decision-making by increasing novelty. *P. Natl. Acad. Sci.* **120**, e2214840120 (2023).

[141] Brinkmann, L. *et al.* Hybrid social learning in human-algorithm cultural transmission. *Philos. T. Roy. Soc. A* **380**, 20200426 (2022).

[142] Pescetelli, N., Reichert, P. & Rutherford, A. A variational-autoencoder approach to solve the hidden profile task in hybrid human-machine teams. *PLOS ONE* **17**, e0272168 (2022).

[143] Dellermann, D., Ebel, P., Söllner, M. & Leimeister, J. M. Hybrid intelligence. *Bus. Inf. Syst. Eng.* **61**, 637–643 (2019).

[144] Wiethof, C. & Bittner, E. Hybrid intelligence-combining the human in the loop with the computer in the loop: A systematic literature review. In *Forty-Second International Conference on Information Systems, Austin*, 1–17 (2021).

[145] Hekler, A. *et al.* Superior skin cancer classification by the combination of human and artificial intelligence. *Eur. J. Cancer* **120**, 114–121 (2019).

[146] Tschandl, P. *et al.* Human–computer collaboration for skin cancer recognition. *Nat. Med.* **26**, 1229–1234 (2020).

[147] Wright, D. E. *et al.* A transient search using combined human and machine classifications. *Mon. Not. Roy. Astron. Soc.* **472**, 1315–1323 (2017).

[148] Bowyer, A., Maidel, V., Lintott, C., Swanson, A. & Miller, G. This image intentionally left blank: Mundane images increase citizen science participation. In *2015 Conference on Human Computation & Crowdsourcing*, vol. 460 (2015).

[149] Trouille, L., Lintott, C. J. & Fortson, L. F. Citizen science frontiers: Efficiency, engagement, and serendipitous discovery with human–machine systems. *P. Natl. Acad. Sci.* **116**, 1902–1909 (2019).

[150] Ibrahim, K., Khodursky, S. & Yasseri, T. Gender imbalance and spatiotemporal patterns of contributions to citizen science projects: The case of Zooniverse. *Front. Phys.* **9**, 650720 (2021).

[151] Cui, H. & Yasseri, T. AI-enhanced collective intelligence. Preprint at *arXiv* `https://doi.org/10.48550/arXiv.2403.10433` (2024).

[152] Hagströmer, B. & Nordén, L. The diversity of high-frequency traders. *J. Finan. Markets* **16**, 741–770 (2013).

[153] Brogaard, J., Hendershott, T. & Riordan, R. High-frequency trading and price discovery. *Rev. Finan. Stud.* **27**, 2267–2306 (2014).

[154] Hirschey, N. Do high-frequency traders anticipate buying and selling pressure? *Manag. Sci.* **67**, 3321–3345 (2021).

[155] Chaboud, A. P., Chiquoine, B., Hjalmarsson, E. & Vega, C. Rise of the machines: Algorithmic trading in the foreign exchange market. *J. Finan.* **69**, 2045–2084 (2014).

[156] Jarrow, R. A. & Protter, P. A dysfunctional role of high frequency trading in electronic markets. *Int. J. Theor. Appl. Finan.* **15**, 1250022 (2012).

[157] Hendershott, T., Jones, C. M. & Menkveld, A. J. Does algorithmic trading improve liquidity? *J. Finan.* **66**, 1–33 (2011).

[158] Hasbrouck, J. & Saar, G. Low-latency trading. *J. Finan. Markets* **16**, 646–679 (2013).

[159] Boehmer, E., Fong, K. & Wu, J. International evidence on algorithmic trading. In *AFA 2013 San Diego Meetings Paper* (2012).

[160] Johnson, N. *et al.* Abrupt rise of new machine ecology beyond human response time. *Sci. Rep.* **3**, 2627 (2013).

[161] Kirilenko, A., Kyle, A. S., Samadi, M. & Tuzun, T. The flash crash: High-frequency trading in an electronic market. *J. Finan.* **72**, 967–998 (2017).

[162] Budish, E., Cramton, P. & Shim, J. The high-frequency trading arms race: Frequent batch auctions as a market design response. *Q. J. Econ.* **130**, 1547–1621 (2015).

[163] Beskow, D. M. & Carley, K. M. Bot conversations are different: Leveraging network metrics for bot detection in Twitter. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 825–832 (2018).

[164] Cresci, S. A decade of social bot detection. *Commun. ACM* **63**, 72–83 (2020).

[165] Davis, C. A., Varol, O., Ferrara, E., Flammini, A. & Menczer, F. BotOrNot: A system to evaluate social bots. In *Proc. 25th International Conference Companion on World Wide Web*, 273–274 (ACM, 2016).

[166] Duh, A., Slak Rupnik, M. & Korošak, D. Collective behavior of social bots is encoded in their temporal Twitter activity. *Big Data* **6**, 113–123 (2018).

[167] Orabi, M., Mouheb, D., Al Aghbari, Z. & Kamel, I. Detection of bots in social media: A systematic review. *Inf. Process. Manag.* **57**, 102250 (2020).

[168] Varol, O., Ferrara, E., Davis, C., Menczer, F. & Flammini, A. Online human-bot interactions: Detection, estimation, and characterization. *Proc. International AAAI Conference on Web and Social Media* **11**, 280–289 (2017).

[169] Abokhodair, N., Yoo, D. & McDonald, D. W. Dissecting a social botnet: Growth, content and influence in Twitter. In *Proc. 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 839–851 (ACM, 2015).

[170] Bastos, M. T. & Mercea, D. The Brexit botnet and user-generated hyperpartisan news. *Soc. Sci. Comput. Rev.* **37**, 38–54 (2019).

[171] Chu, Z., Gianvecchio, S., Wang, H. & Jajodia, S. Who is tweeting on Twitter: Human, bot, or cyborg? In *Proc. 26th Annual Computer Security Applications Conference*, 21–30 (ACM, 2010).

[172] Grimme, C., Preuss, M., Adam, L. & Trautmann, H. Social bots: Human-like by means of human control? *Big Data* **5**, 279–293 (2017).

[173] Chu, Z., Gianvecchio, S., Wang, H. & Jajodia, S. Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? *IEEE T. Depend. Secure Comput.* **9**, 811–824 (2012).

[174] Stella, M., Ferrara, E. & De Domenico, M. Bots increase exposure to negative and inflammatory content in online social systems. *P. Natl. Acad. Sci.* **115**, 12435–12440 (2018).

[175] Pozzana, I. & Ferrara, E. Measuring bot and human behavioral dynamics. *Front. Phys.* **8** (2020).

[176] Ferrara, E. Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday* (2017).

[177] Forelle, M., Howard, P., Monroy-Hernández, A. & Savage, S. Political bots and the manipulation of public opinion in Venezuela. Preprint at *arXiv* `https://doi.org/10.48550/arXiv.1507.07109` (2015).

[178] Howard, P., Woolley, S. & Calo, R. Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration. *J. Inf. Tech. Polit.* **15**, 81–93 (2018).

[179] Howard, P. N. & Kollanyi, B. Bots, #strongerIn, and #brexit: Computational propaganda during the UK-EU referendum. Preprint at *arXiv* `https://doi.org/10.48550/arXiv.1606.06356` (2016).

[180] Shao, C. *et al.* The spread of low-credibility content by social bots. *Nat. Commun.* **9**, 4787 (2018).

[181] Suárez-Serrato, P., Roberts, M. E., Davis, C. & Menczer, F. On the influence of social bots in online protests. In Spiro, E. & Ahn, Y.-Y. (eds.) *Social Informatics*, 269–278 (Springer International Publishing, 2016).

[182] Yan, H. Y., Yang, K.-C., Shanahan, J. & Menczer, F. Exposure to social bots amplifies perceptual biases and regulation propensity. *Sci. Rep.* **13**, 20707 (2023).

[183] Himelein-Wachowiak, M. *et al.* Bots and misinformation spread on social media: Implications for COVID-19. *J. Med. Internet Res.* **23**, e26933 (2021).

[184] Yang, K.-C., Torres-Lugo, C. & Menczer, F. Prevalence of low-credibility information on Twitter during the COVID-19 outbreak. Preprint at *arXiv* `https://doi.org/10.36190/2020.16` (2020).

[185] Fan, R., Talavera, O. & Tran, V. Social media bots and stock markets. *Eur. Finan. Manag.* **26**, 753–777 (2020).

[186] Hwang, T., Pearce, I. & Nanis, M. Socialbots: Voices from the fronts. *Interactions* **19**, 38–45 (2012).

[187] Stella, M., Cristoforetti, M. & Domenico, M. D. Influence of augmented humans in online interactions during voting events. *PLOS ONE* **14**, e0214210 (2019).

[188] Vosoughi, S., Roy, D. & Aral, S. The spread of true and false news online. *Science* **359**, 1146–1151 (2018).

[189] Gorodnichenko, Y., Pham, T. & Talavera, O. Social media, sentiment and public opinions: Evidence from #Brexit and #USElection. *Eur. Econ. Rev.* **136**, 103772 (2021).

[190] Bessi, A. & Ferrara, E. Social bots distort the 2016 U.S. Presidential election online discussion. *First Monday* (2016).

[191] BBC News. Twitter 'shuts down millions of fake accounts'. *BBC News* (2018).

[192] Dang, S. & Paul, K. Twitter says it removes over 1 million spam accounts each day. *Reuters* (2022).

[193] Halfaker, A. & Riedl, J. Bots and cyborgs: Wikipedia's immune system. *Computer* **45**, 79–82 (2012).

[194] Niederer, S. & van Dijck, J. Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. *New Media Soc.* **12**, 1368–1387 (2010).

[195] Zheng, L. N., Albano, C. M., Vora, N. M., Mai, F. & Nickerson, J. V. The roles bots play in Wikipedia. *P. ACM Hum.-Comput. Interact.* **3**, 215:1–215:20 (2019).

[196] Geiger, R. S. The lives of bots. In Lovink, G. & Tkacz, N. (eds.) *Critical Point of View: A Wikipedia Reader*, 78–93 (Institute of Network Cultures, 2011).

[197] Steiner, T. Bots vs. Wikipedians, anons vs. logged-ins (redux): A global study of edit activity on Wikipedia and Wikidata. In *Proceedings of The International Symposium on Open Collaboration*, 1–7 (ACM, 2014).

[198] Geiger, R. S. & Halfaker, A. Operationalizing conflict and cooperation between automated software agents in Wikipedia: A replication and expansion of 'Even Good Bots Fight'. *P. ACM Hum.-Comput. Interact.* **1**, 49:1–49:33 (2017).

[199] Clément, M. & Guitton, M. J. Interacting with bots online: Users' reactions to actions of automated programs in Wikipedia. *Comput. Hum. Behav.* **50**, 66–75 (2015).

[200] Geiger, R. S. & Halfaker, A. When the levee breaks: Without bots, what happens to Wikipedia's quality control processes? In *Proc. 9th International Symposium on Open Collaboration*, 1–6 (ACM, 2013).

[201] Hilbert, M. & Darmon, D. Large-scale communication is more complex and unpredictable with automated bots. *J. Commun.* **70**, 670–692 (2020).

[202] Massanari, A. L. Contested play: The culture and politics of Reddit bots. In *Socialbots and Their Friends* (Routledge, 2016).

[203] Hurtado, S., Ray, P. & Marculescu, R. Bot detection in Reddit political discussion. In *Proc. Fourth International Workshop on Social Sensing*, 30–35 (ACM, 2019).

[204] Jhaver, S., Birman, I., Gilbert, E. & Bruckman, A. Human-machine collaboration for content regulation: The case of Reddit automoderator. *ACM T. Comput.-Hum. Interact.* **26**, 31:1–31:35 (2019).

[205] Ma, M.-C. & Lalor, J. P. An empirical analysis of human-bot interaction on Reddit. In *Proc. Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*, 101–106 (ACL, 2020).

[206] Yang, K.-C. & Menczer, F. Anatomy of an AI-powered malicious social botnet. Preprint at *arXiv* `https://doi.org/10.48550/arXiv.2307.16336` (2023).

[207] Yang, K.-C., Singh, D. & Menczer, F. Characteristics and prevalence of fake social media profiles with AI-generated faces. Preprint at *arXiv* `https://doi.org/10.48550/arXiv.2401.02627` (2024).

[208] Ray, P. P. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet Things Cyb.-Phys. Syst.* **3**, 121–154 (2023).

[209] Webb, T., Holyoak, K. J. & Lu, H. Emergent analogical reasoning in large language models. *Nat. Hum. Behav.* **7**, 1526–1541 (2023).

[210] Frey, S. Mixed human/entity games and the anomalous effects of misattributing strategic agency. *Adapt. Behav.* **22**, 266–276 (2014).

[211] Aiello, L. M., Deplano, M., Schifanella, R. & Ruffo, G. People are strange when you're a stranger: Impact and influence of bots on social networks. *Proc. International AAAI Conference on Web and Social Media* **6**, 10–17 (2012).

[212] Freitas, C., Benevenuto, F., Ghosh, S. & Veloso, A. Reverse engineering socialbot infiltration strategies in Twitter. In *Proc. 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, 25–32 (ACM, 2015).

[213] Messias, J., Schmidt, L., Oliveira, R. A. R. & de Souza, F. B. You followed my bot! Transforming robots into influential users in Twitter. *First Monday* **18**, 1–14 (2013).

[214] Savage, S., Monroy-Hernandez, A. & Höllerer, T. Botivist: Calling volunteers to action using online bots. In *Proc. 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 813–822 (ACM, 2016).

[215] Krafft, P. M., Della Penna, N. & Pentland, A. S. An experimental study of cryptocurrency market dynamics. In *Proc. 2018 CHI Conference on Human Factors in Computing Systems*, 1–13 (ACM, 2018).

[216] Bail, C. A. *et al.* Exposure to opposing views on social media can increase political polarization. *P. Natl. Acad. Sci.* **115**, 9216–9221 (2018).

[217] Lorenz, T. Welcome to the age of automated dating. *Washington Post* (2023).

[218] Salganik, M. J. *Bit by Bit: Social Research in the Digital Age* (Princeton Univ. Press, 2019).

[219] Paiva, A., Mascarenhas, S., Petisca, S., Correia, F. & Alves-Oliveira, P. Towards more humane machines: Creating emotional social robots. In *New interdisciplinary landscapes in morality and emotion*, 125–139 (Routledge, 2018).

[220] Wagman, K. B. & Parks, L. Beyond the command: Feminist STS research and critical issues for the design of social machines. *P. ACM Hum.-Comput. Interact.* **5**, 1–20 (2021).

[221] Chang, D. Pushing THEIR buttons! texas drivers are left furious as 20 cruise self-driving cars cause gridlock in Austin - as company blames pedestrian traffic. *Mail Online* (2023).

[222] Asimov, I. *I, Robot*, vol. 1 (Spectra, [1950] 2004).

[223] Graham, T. & Ackland, R. Do socialbots dream of popping the filter bubble? The role of socialbots in promoting deliberative democracy in social media. In Gehl, R. W. & Bakardjieva, M. (eds.) *Socialbots and Their Friends: Digital Media and the Automation of Sociality*, 187–206 (Routledge, 2017).

[224] Awad, E. *et al.* The Moral Machine experiment. *Nature* **563**, 59–64 (2018).

[225] Pinheiro, R. & Young, M. The university as an adaptive resilient organization: A complex systems perspective. In *Theory and method in higher education research*, vol. 3, 119–136 (Emerald Publishing Limited, 2017).

[226] May, R. M., Levin, S. A. & Sugihara, G. Ecology for bankers. *Nature* **451**, 893–894 (2008).

[227] Balsa-Barreiro, J., Vié, A., Morales, A. J. & Cebrián, M. Deglobalization in a hyper-connected world. *Palgrave Commun.* **6**, 1–4 (2020).

[228] Bak-Coleman, J. B. *et al.* Stewardship of global collective behavior. *P. Natl. Acad. Sci.* **118**, e2025764118 (2021).

[229] Liu, Z. Sociological perspectives on artificial intelligence: A typological reading. *Sociol. Compass* **15**, e12851 (2021).

[230] Pescetelli, N., Rutherford, A. & Rahwan, I. Modularity and composite diversity affect the collective gathering of information online. *Nat. Commun.* **12**, 3195 (2021).

[231] Mason, W. & Watts, D. J. Collaborative learning in networks. *P. Natl. Acad. Sci.* **109**, 764–769 (2012).

[232] Scheffer, M. *et al.* Anticipating critical transitions. *Science* **338**, 344–348 (2012).

[233] Centola, D., Becker, J., Brackbill, D. & Baronchelli, A. Experimental evidence for tipping points in social convention. *Science* **360**, 1116–1119 (2018).

[234] Kenway, E. 'Care bots': A dream for carers or a dangerous fantasy? *Observer* (2023).

[235] Brinkmann, L. *et al.* Machine culture. *Nat. Hum. Behav.* **7**, 1855–1868 (2023).

Competing interests

The authors declare no competing interests.