Perspective

Measuring algorithmically infused societies

https://doi.org/10.1038/s41586-021-03666-1

Received: 5 March 2021

Accepted: 21 May 2021

Published online: 30 June 2021

Check for updates

Claudia Wagner[1,2,3] ✉, Markus Strohmaier[1,2,3], Alexandra Olteanu[4,5], Emre Kıcıman[6], Noshir Contractor[7] & Tina Eliassi-Rad[8]

It has been the historic responsibility of the social sciences to investigate human societies. Fulfilling this responsibility requires social theories, measurement models and social data. Most existing theories and measurement models in the social sciences were not developed with the deep societal reach of algorithms in mind. The emergence of 'algorithmically infused societies'—societies whose very fabric is co-shaped by algorithmic and human behaviour—raises three key challenges: the insufficient quality of measurements, the complex consequences of (mis)measurements, and the limits of existing social theories. Here we argue that tackling these challenges requires new social theories that account for the impact of algorithmic systems on social realities. To develop such theories, we need new methodologies for integrating data and measurements into theory construction. Given the scale at which measurements can be applied, we believe measurement models should be trustworthy, auditable and just. To achieve this, the development of measurements should be transparent and participatory, and include mechanisms to ensure measurement quality and identify possible harms. We argue that computational social scientists should rethink what aspects of algorithmically infused societies should be measured, how they should be measured, and the consequences of doing so.

We are witnessing the emergence of algorithmically infused societies who are shaped by deeply entangled algorithmic and human processes and behaviour; with early mentions of 'algorithmic societies' dating back to over a decade ago[1–4]. Algorithms—and the tools, services and platforms they power—mediate social, economic and political processes by shaping a wide range of activities and decision-making practices across many areas. Algorithmic influence can be observed in how people consume information or cultural artefacts[5–12] and how they interact with others[13–16]. This influence is also increasingly tangible in many high-stakes areas, such as healthcare[17–19], credit scoring[20], law enforcement[21,22] and employment[23–25]. Although algorithms can provide many benefits, including efficiency, objectivity, auditability, fairness and social good[26,27], they can also amplify—either inadvertently or purposefully—existing inequalities and biases in society, or might even introduce new ones[28–34].

The challenges of measuring algorithmically infused societies

The social sciences have historically been concerned with studying individual- and societal-level phenomena, including social structures, social relationships and dynamics, as well as the emergence and impact of social norms and values. Because of this, social scientists have developed rich methodologies that support theory construction[35,36]. They have also established measurement theory (for example, psychometrics)[37,38] and survey methodology[39] as research areas dedicated to the development of new measurement models and their quality assurance.

However—to a large degree—the existing toolkit of social theories and measurement models was not created with the deep societal reach of algorithms in mind, and may thus not apply to human societies that are permeated by algorithms (see Box 1). In these societies, social, economic, political and scientific processes both influence and are influenced by the design and presence of algorithms (see Fig. 1).

Algorithmically infused societies are thus a fertile ground for what is sometimes known as Goodhart's law—the conflation of 'is' and 'ought'[40], where the mere existence of measurements can alter the behaviour of individuals[41]. Measurements may also indirectly alter behaviours by informing the development of social theories and subsequently influence the algorithms and technologies that draw on those theories. Consider recommendation algorithms: designers of such algorithms may ground their solutions in various theories including cognitive dissonance[42], balance theory[43,44], or ideas related to homophily[45] and human categorization[46,47]. This may result in what Healy calls 'performativity'[48]. The performativity thesis conjectures that theories have the potential to "reformat and reorganize the phenomena [that] models purport to describe"[48]. The algorithms that constrain or nudge our behaviour (for example, by making some content more visible than other) constantly change but often without us noticing[49,50]. It is these characteristics of dynamicism, heterogeneity, interconnectedness and opacity of algorithmically infused societies that makes their study more challenging.

These concerns are not entirely new. Science and technology studies, legal, and other scholars have long discussed (and raised concerns about) how technology and other social artefacts (such as norms, culture or political frameworks) blend and influence each other, for

[1]GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany. [2]RWTH Aachen University, Aachen, Germany. [3]Complexity Science Hub Vienna, Vienna, Austria. [4]Microsoft Research Montreal, Montreal, Quebec, Canada. [5]Microsoft Research New York, New York, NY, USA. [6]Microsoft Research Redmond, Redmond, WA, USA. [7]Northwestern University, Evanston, IL, USA. [8]Northeastern University, Boston, MA, USA. ✉e-mail: claudia.wagner@gesis.org

Box 1

Key terms and concepts

Measurement models: theoretical and practical approaches that tie high-level social constructs to observable data. Measurement models justify, for example, why what we are measuring is what we think we are measuring.

Social data: directly observable information captured via human interactions with or through digital devices, services and infrastructure.

Social theories: explanatory conceptualizations of social phenomena.

Algorithms: automated processes that are increasingly often integrated within societally critical processes.

examples, see refs. [40,47,51–55]. Computational social science has emerged as a (sub)discipline that "leverages the capacity to collect and analyse data with an unprecedented breadth and depth and scale" to conduct empirical studies of individual- and societal-level phenomena[56,57] Although computational social science has made critical progress in understanding empirical phenomena such as the spread of misinformation[58,59] and political polarization[7,60], it has often overlooked questions about what even attempting to measure perceived social phenomena signifies, what measurements enable and for whom, or how to ensure the validity and reliability of measurements.

In this paper, we consequently highlight three key challenges to measuring social phenomena in algorithmically infused societies: first, we discuss the insufficient quality of measurements; second, we examine the complex consequences of (mis)measurements; and third, we explore the limits of existing social theories. We discuss how these challenges are linked to the essence of algorithmically infused societies and outline a possible roadmap for future work. In particular, we suggest that researchers develop trustworthy measurement models, mitigate the harmful consequences of (mis)measurements, and construct integrated and empirically informed theories.

The insufficient quality of measurements

Characterizing, explaining and predicting social phenomena requires measurement models that tie theoretical constructs to observable data. As a result, social scientists have been developing measurement instruments, such as survey scales, to attempt to explicitly model the relationship between a theoretical construct of interest and, for example, survey questionnaire items. The construction of measurement models and instruments is, however, often grounded in assumptions that should be identified and explicitly articulated (and tested) before using the resulting measurements[61–63]. Social scientists have developed quality criteria for measurement modelling to avoid using instruments for which the validity is unclear or even questionable[39,62,64]. These quality criteria also facilitate the comparison of empirical results and the reusability of measurement models (and the underlying data).

The algorithmic infusion of society has led to major changes in the social sciences as more and more aspects of our social, political and economic lives are captured by social data. This has facilitated the adoption of machine learning models in the social sciences, which may assist with the measurement, the prediction and—to some degree—also the explanation of theoretical constructs and social phenomena[65]. Machine learning models can be powerful, and some can also help to quantify causal relationships[66]. Yet they often come as 'black boxes' that do not make their assumptions explicit and thereby introduce or exacerbate potential mismatches between the theoretical construct of interest and its operationalization[63,67]. Although the Common Task Framework (CTF) has successfully driven progress in some areas of machine learning[68], the 'leaderboardism' in science has also led to the quick development of mis- or under-specified benchmark datasets and shared tasks[69,70]. These datasets often fail to specify or capture the construct purported to be measured (by, for example, failing to operationalize the different dimensions of a construct)[69,71,72]. They also fail to define and justify the criteria and metrics for assessing the quality of a measurement model, describe and justify the data selection and preprocessing methods, and acknowledge and mitigate biases in the data[69,73–75].

In algorithmically infused societies, the theoretical constructs we aim to measure, predict and explain are often unstable[76] and can also be affected by the act of measuring, predicting or explaining[41,77]. They may not, however, change equally for everyone as companies often launch A/B tests or personalize algorithms according to some user characteristics—for example, based on the current user location[78], or based on their gender or race[79]. The data that are collected are also often 'collected under the algorithm', with algorithms shaping the data that computational social scientists subsequently base their measurements on. This can threaten the validity of research results and can complicate the evaluation of alternative algorithms before their deployment. The latter problem is currently addressed by research in the area of policy learning[80], which explores the evaluation of performance of policies without deployment and in the area of recommendation algorithms, which aims to deconvolve feedback loops in recommender systems[81].

We therefore argue that a key challenge for computational social scientists is defining quality standards and best practices for the development of trustworthy measurement models.

The consequences of (mis)measurements

In algorithmically infused societies, highly individualized and granular behaviour can be recorded and thus measured. Although measurements can provide evidence to help to address pressing challenges in our society—for example, by flagging those at risk for unwanted or harmful events such as child abuse, suicide, poverty or environmental disasters[82]—measurements by their very nature can often lead to increased perceptions of objectivity of human and social phenomena that might be subjective in nature. Such individualized measurements could, for instance, lead to inflated notions of self or to an erosion of solidarity with others[83].

Measurements can also become a tool for self-improvement, or the improvement of organizations or societies, thereby influencing the evolution of social constructs and entities over time. In a way, we manage what we measure and what we measure often informs policymaking and resource allocation[84]; and it can as a result even influence how we optimize our behaviour[32,41]. At the same time, not everything that is measured should be measured, for example, owing to concerns of privacy or possible harms if data taken out of context are misinterpreted by an algorithm or other people. And what we do not measure—or what we ignore and do not see as a result—also can have implications for society. It is critical to reflect on and assess the potential consequences of measurements before they are put in place or before they become out of place; however, the consequences of measurements are hard to anticipate, especially when 'black box' models—which might adapt to their environment and change it at the same time—are used as measurements.

In computational social science—where statistical and machine learning models are now routinely being incorporated in measurement models—there is often also an implicit fit-for-purpose assumption. The quality of these models is assessed on the basis of metrics that capture how well a model characterizes some observable data (its goodness of fit) or predicts some future data (its predictive performance), or even how well a measurement approximates the construct of interest (its validity and reliability). Neither one of these metrics, however, offers much insight into possible unintended consequences of measurements or how to mitigate any such consequences.

Harmful consequences can also arise from mismatches between the theoretical understanding of a social construct and the

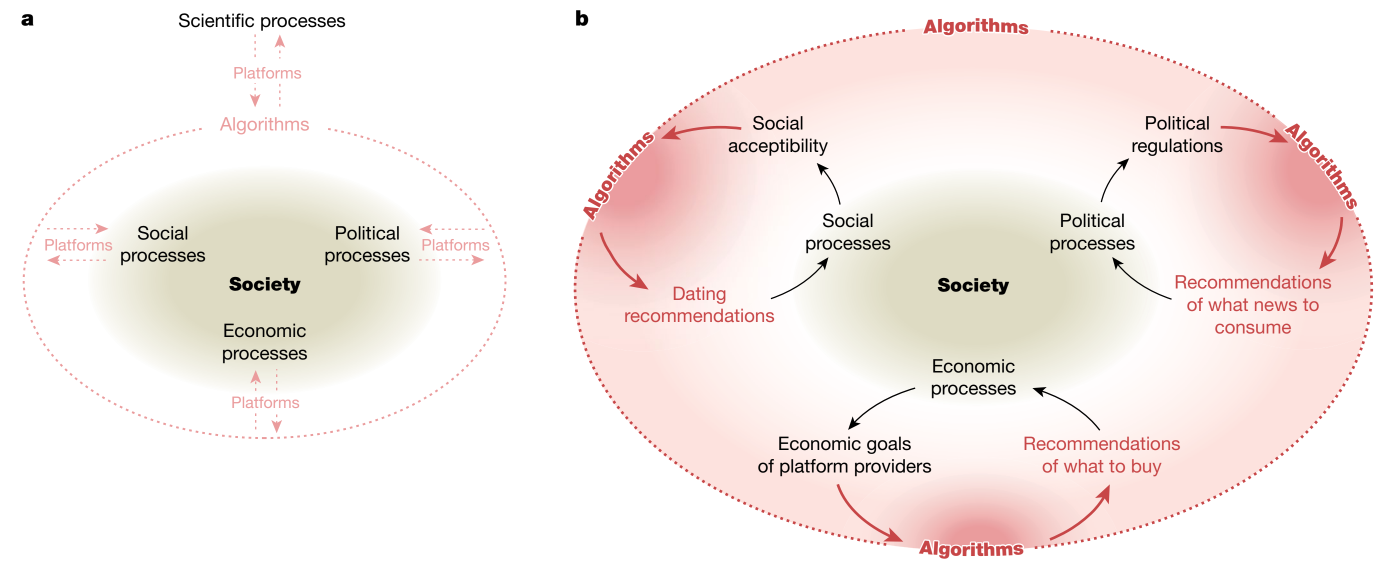

Fig. 1 | Measuring algorithmically infused societies. a, **b**, In algorithmically infused societies, not only do social, economic, political and scientific processes impact each other (**a**), but algorithms also operate at different levels—shaping and being shaped by their environment. Scientific processes can shape algorithmic platforms when newly developed algorithms also lead to new measurements, hypotheses and theories. For example, neural language models may enable new measurements of similarity between users, impacting homophily measurements and theories about user interactions. At the same time, platforms determine what can be observed; as data are often collected 'under the algorithm' and platforms decide what can be collected, and who and what type of access they have to the data and the underlying systems. For example, hate speech detection algorithms on social media platforms may impact measures of civility. Organizational models such as Social Science One (https://socialscience.one) support collaborations between industry and academic partners, but still define who has access to what data and for which purpose. Finally, algorithms can also shape social, political and economic processes (**b**). The social acceptability of algorithms may influence which algorithm is adopted. Political regulations and economic goals of platform providers may also determine what will be deployed. Recommendation algorithms may influence our micro-behaviours by suggesting whom to date or what books to read; nudged behaviours that algorithms may then learn from. Ranking algorithms can impact political processes by determining the visibility of political information, which in turn may be regulated by political actors. Optimization algorithms may affect both individual purchasing decisions and societal level financial processes by optimizing prices and interest rates, yet the optimization goals might be controlled by private entities. Both individual and societal level phenomena can thus be influenced by algorithmic systems, making it difficult to disentangle algorithmic and human behaviours and how they influence each other.

operationalization of that construct, an issue of construct validity[63,69]. This is particularly concerning when measurements are used in high-risk scenarios where the data that underpin the measurements (for example, arrest records) may not properly match the actual social construct the measurements are meant to capture (for example, actual crimes). Such mismatches can lead to measurements that could (and often do) replicate inequity, mask it, transfer it, exacerbate it, and even compromise inequity oversight[29].

What makes such possible harmful effects more concerning in algorithmically infused societies is also the scale at which measurements are often done, as well as their perceived objectivity, efficiency and trustworthiness[26]. A first step towards addressing inequity concerns and mitigating harms is to audit measurements along their underpinning data and statistical models. Such audits could help to reveal both misuses and mismatches between data, theoretical constructs and measurement models that might expose, distort or even erase behaviours, experiences or identities.

Auditing may however require both visibility into how the measurement is done (transparency)—not to be confused with transparency as the publicizing of the measurement values, which can have its own set of harmful effects[85]—and the ability to understand why certain measurement values were obtained (interpretability). For example, ensuring procedural justice—that is, ensuring that the process by which a construct of interest is measured is fair—might require both seeing and understanding the data and the computations or statistical models involved in the operationalization of a particular measurement from these data[86].

The transparency (and interpretability) of measurements can, however, be construed as being opposite to privacy[87]. Privacy risks can arise from the prevalence of measurements and can be compounded by the transparency of their instrumentation—which may leak additional information[88,89]; calling further attention to the question of what should not be measured.

Indeed, a known tension when auditing measurements (and systems) for inequity is how to assess potential disparities across demographic groups or individuals with (little or) no demographic information[90]—typically requiring information about demographic or identity markers to distinguish how similar two individuals are or what groups they belong to[91,92]. Although the right trade-offs (and even how they should be determined) are still being debated, they ought to be contextually dependent; and—as the "nothing about us without us" principle[93–95] suggests—those who are likely to bear the risks should have a key role in their determination. Complicating matters even further, measurements could help to make society more "legible"[96] as a whole, and thereby enable states to enact policies or interventions that might not be possible without extensive measurements.

For computational social scientists who want to measure, predict or explain phenomena in algorithmically infused societies, another challenge is thus how to carefully consider and anticipate potentially harmful consequences of measurements, as well as how to establish mechanisms to detect and mitigate them.

The limits of existing social theories

Theories explain phenomena and are a crucial ingredient for developing trustworthy measurements and avoiding over- and mismeasurements. They also enable the transfer of measurements that are local with respect to time and space to new contexts[97]. Although computational social science has demonstrated its potential to test and adapt existing social theories at scale[98–100], less progress has been made in developing new theories[101,102]. This is despite the need for theories that recognize the role of algorithms in society; and despite the volume, variety, velocity and veracity of digital traces about human and algorithmic behaviour, which call for a more empirically informed, iterative theory development.

Although situated contexts are usually considered, existing theories in the social sciences were not developed with the deep societal reach of algorithms in mind. For example, theories of homophily were devised to understand 'natural' self-selection behaviour in societies, such as the way students form relationships in schools on the basis of certain attributes such as age, ethnicity or gender. However, studies of homophily in algorithmically infused societies require us to account for how algorithmic amplification or dampening—such as via friend-recommender systems, newsfeeds or social feedback and voting mechanisms—can influence relationship formation[16]. At the same time, these theories also need to account for social, economic or political forces that may lead to changes in the algorithms. For example, if some people game or exploit an algorithm for their own ends, it may become necessary from an economic standpoint to adapt the algorithm to avoid financial harms to other users or harm to the platform itself. From a social or political standpoint it might become necessary to adjust how information consumption is regulated by algorithms if it leads to increased segregation or polarization in society. But it remains an open question even which stakeholders should decide when and how algorithms should be deployed or amended.

It is this fusion of algorithmic and human behaviour and processes that leads to feedback loops and calls for integrated theories that take a more comprehensive, systems-level perspective. Isolated studies of one without explicitly accounting for the other will necessarily ignore how such feedback loops affect the constructs and the phenomena they are attempting to measure and understand. This not only has direct implications for computational studies of human behaviour, but also consequences for social studies of machine behaviour[103].

The adaptive nature of algorithmically infused societies where algorithms and humans habitually interact with each other and their dynamic population characteristics in turn invite the fusion of theory and data—for example, by integrating theory-driven and data-driven approaches. On the one hand, pure data-driven research that only relies on "readymade data"[104,105] has been criticized for asking questions that appear opportunistically driven by what data are available and for overlooking different types of data biases[6,74,105]. On the other hand, theory-driven research has also been criticized for limiting research to a "theoretical straightjacket"[106]. This criticism has stimulated an ongoing discussion about the value of problem- and solution-oriented[107,108] and phenomenon-driven[106] research in the social sciences. Although experimental approaches could alleviate some of the challenges that come with data- and theory-driven approaches, they are often hard to scale and might not even be feasible to execute when faced with highly individualized, parameterized and algorithmically confounded environments.

We thus believe that, moving forward, a persistent challenge for computational social scientists lies in how to effectively navigate data- and theory-driven research, a challenge that requires us to rethink the relationship between data and theory, particularly when studying algorithmically shaped social phenomena .

Towards responsible and trustworthy measurements of algorithmically infused societies

To address these challenges, we believe that the computational social science community should focus on developing trustworthy measurement models, mitigating the harmful consequences of (mis)measurements, and constructing more integrated and empirically informed social theories (see Box 2).

Box 2

Best practices for measuring algorithmically infused societies

1. Ensure that your investigations are guided by transparent and participatory processes that are informed by theory, data and ethical considerations.
2. Integrate existing data, contextual information (including about algorithmic impact and population characteristics), computational methods and measurements into theory construction.
3. Develop and reuse measurement models informed by theory, document and justify the assumptions underlying them, document the data and justify their suitability for characterizing the phenomena, and reflect on the algorithmic influences on the data and the measurement models.
4. Develop and justify quality criteria and datasets that allow validating competing measurement models.
5. Reflect on possible harmful consequences of measurements—including from measuring the wrong theoretical construct, from mis- and overmeasuring the right construct, from failures to measure, and from consequences of measurements on the design of algorithms—and describe mitigation strategies.

Developing trustworthy measurements

Measurement models of algorithmically infused societies often rely on statistical and machine learning models that learn to what extent observable indicators are related to the construct of interest. We believe that the development of trustworthy measurement models for algorithmically infused societies requires us to rethink both the measurement development process and its quality assurance.

Triangulating data to examine measurement quality. To enable research on the quality of measurement models we often need to triangulate data—including interlinked data from different channels and sources (for example, self-reports collected via surveys or interviews, observational data in the form of text, speech and locations, and experimental data that capture behavioural differences in response to system-level interventions). Depending on the research design, the interlinking of data happens either ex ante or ex post and either on an aggregated or individual level[109].

Different approaches to data collection and triangulation lead to epistemological and ethical challenges that need to be carefully considered. When linking data ex ante on an individual level (for example, via an internet panel that combines surveys with observations of web browsing behaviour), recruiting participants and obtaining informed consent are crucial steps to avoid ethical and epistemological issues. When data are interlinked ex post on an aggregated level—for example, by comparing topic salience for the general public versus online audiences for a period of interest such as an election campaign[110,111]—epistemological (such as self-selection bias and platform affordances) and ethical concerns must be considered.

Although no measurement model is perfect, we might learn about the benefits and limitations of different models by comparing them using data that might be confounded by the same variables (for example, observation period, research subjects, external events). Previous research showed that measurement models that rely on self-reports to make inferences about a range of behaviours—such as communication patterns[112], physical activity[113] or internet or smartphone usage[114–116]—are often inaccurate owing to cognitive and social biases (for example, false recall and social desirability bias). This not only highlights limitations of widely used survey measurements, but also informs the constant improvement of those instruments.

We believe that the computational social science community will also benefit from additional quantitative and qualitative research on the quality of observational data and their collection instruments. A good example is a recent study that compares face-to-face contacts recorded by RFID sensors with manually annotated video recordings, providing insights into the accuracy of those sensors[117]. This example highlights that research on the quality of measurements is important and that triangulated datasets such as the Copenhagen Networks Study[118] might help to stimulate this type of research. How to collect and share such datasets remains, however, an open question that requires a good balance between the privacy protection of individuals, possible harmful consequences to different stakeholders (particularly to historically and presently disadvantaged groups), and the benefits of these data and the research that they enable.

Developing guidelines and best practices to increase the quality of measurements. Although the computational social science community is continually increasing its repertoire of measurement practices, it still lacks clear quality standards and frameworks that support the documentation of the measurement development process. Practices from survey methodology[62,75,119] and other domains, such as the medical industry[120], can inform our thinking about ways to develop trustworthy and responsible measurements for social phenomena. Documentation guidelines and frameworks recently developed for the AI community might also provide a starting point for the documentation of the measurement development process for social constructs[73,121–123].

Most of these practices have, however, been developed assuming more static, controlled environments. In algorithmically infused societies even the constructs that we aim to measure are often 'moving targets' as both their meaning and their operationalization can change over time. Furthermore, predictions—as measurements of what might happen—when used for decision-making may influence the outcome that they aim to predict[77], thereby highlighting a fundamental problem of the interwoven nature of algorithmically infused societies. To address this challenge, computational social scientists can also draw inspiration from adaptive learning methods that react to concept drifts by updating predictive models during their deployment[124–126]. A special case of such distribution drifts occurs when prediction models inform the decisions that impact the outcome that they aim to predict[77]—for example, traffic predictions impact traffic patterns and popularity rankings impact popularity. This phenomenon also affects the development of measurement models in algorithmically infused societies, because the measurement—if it fails to adapt to changes or if it affects the outcome that it measures—may become invalid.

We believe that as a community we should define quality standards and best practices for the documentation, development and maintenance of measurement models, which help to increase the trustworthiness of measurement models. A trustworthy measurement requires that the phenomenon and related constructs (with all their dimensions) are carefully specified and operationalized via a measurement model, that the assumptions of the model are precisely documented and are supported by either strong or highly plausible evidence, and the criteria and test conditions for validating the measurement model are well documented and justified. Although this list may be a good starting point, it remains an open question how to adapt best practices from other fields to the instrumentation of measurements of social phenomena situated in ephemeral, ever-changing and algorithmically confounded environments.

Mitigating (mis)measurement harms

As scientists, we decide what to measure, how to measure it, and how to rely on the values of the measurements. These decisions often reflect the set of principles that guided them—in other words, they reflect implicit or explicit policies and assumptions about what is important or what is worth measuring. We believe that mitigating possible harmful consequences of measurements in algorithmically infused societies requires fundamentally rethinking how the measurements are both carried out and relied upon. Already when the data underlying a measurement are selected, or when the processes for instrumenting the measurement are designed, the potential harms from measuring algorithmically infused societies need to be systematically considered. Doing so requires documentation, good practices, analytical frameworks for reflection, processes for engagement with affected people and groups, and understanding when (not) to measure.

Developing a responsible computational social science agenda. Even though the responsible AI literature—for example, refs. [31,127–130], among many others—has some of its roots in social sciences and human–computer interaction research, computational social science as a community has not put forward an agenda on how the same issues translate to the measurement of social phenomena—particularly given the ubiquity with which machine learning and other statistical models are used in the instrumentation of such measurements. Although validity, reliability and goodness of fit are important quality criteria for measurement models, we believe—particularly when studying social phenomena—that we should also ask, for example, whether the measurement models are just (for example, non-discriminatory) and equitable, transparent (for example, forthright about normative commitments) and interpretable, and privacy preserving. These are important quality criteria not only for how measurement models are made, but also for how the measurements are being used. We should also ask who gets to decide what is being measured, who has access to the measurements, who decides how the measurements are being used, and how this may impact various downstream outcomes.

To understand and document the range of possible harms, in the AI community there are many calls for processes that require researchers and practitioners to reflect on, anticipate and communicate possible adverse impacts from the development and deployment of AI technologies[123,131]. These calls highlight that techniques for preempting unknown or future adverse impacts—arising from how existing or new measurements are instrumented and deployed—are largely missing. Conjecturing about possible future impacts is always difficult[131,132], particularly in ever-changing environments, as algorithmically infused societies often are; or when the models underpinning the measurements can demonstrate novel behaviours by simultaneously learning from and influencing their environment.

These challenges highlight the need for and potential of both qualitative and quantitative research: possible future impacts could be explored via ethical reflections or empirical studies on the consequences of the act of measuring, of communicating measurement results differently and to different stakeholders, and of using the measurements in different settings. Participatory approaches that aim to ensure all relevant stakeholders (particularly those being measured or more likely to experience adverse consequences) are involved in reflection and deliberation about a system design, may also support the anticipation of future impacts. To implement these, we could draw inspiration from the value-sensitive-design framework, which is intended to incorporate human values in early design phases[133]. We should also design mixed-methods approaches that combine quantitative data sampling and generation with user and experimental studies that elicit human feedback[134,135].

Reflecting on what not to measure. Adverse consequences could also stem from overlooking that measurement models (and the data and statistical models that constitute them) are mostly what has previously been called "design materials"[136]—in other words, they are just means to help us to understand (often algorithmically infused) social phenomena. Even when measurements are examined through a responsible computational social science lens, those lenses might still reflect techno-solutionism perspectives that focus on providing computational or modelling remedies for faulty measurement instrumentations—as is often the case in the responsible AI literature[137,138]. Such perspectives fail to ask when (under what conditions) and whether certain measurements should even be carried out. Measurements models—that risk leaking private information, that are motivated by prejudiced beliefs about individuals or groups, or that could be used to discriminate against groups or individuals—should not be developed or used. Similarly, we might also fail to properly reflect upon possible consequences of not measuring—which can result in harms of erasure and misrepresentation by possibly rendering experiences and even individuals invisible, among others.

Developing integrated theories

The webbing of human and algorithmic behaviour and processes stimulates the need for integrated theories that bridge the gap between macro- and microscales and acknowledge the role of algorithms and of the computational tools, services and platforms that they power in society. At the same time, the volume, variety, velocity and veracity of digital traces about human and algorithmic behaviour that are recorded in constantly changing socio-technical environments invites us to reimagine the role of data and measurements in the theory development process. But what is needed to support the construction of integrated and empirically informed theories that explain phenomena in algorithmically infused societies?

Developing transparent, participatory processes for examining algorithmically infused social realities. Theories help us to explain phenomena, ask interesting questions, and develop solutions for open problems. However, as a community we should first identify a set of problems and phenomena that are important and that allow us to empirically validate, compare, and assess the utility of multiple competing theories operating at multiple levels[108,139]. Although this set would guide the theory and measurement development process and help researchers entering the field, we should also stress the importance of deliberation.

In algorithmically infused societies, the possibilities of what can be measured, predicted and explained can sometimes appear unlimited—a problem that can be exaggerated by emerging and ephemeral socio-technical phenomena. At the same time, the consequences of measurements are hard to anticipate because they may both directly and indirectly affect behaviour. The socio-technical nature of problems and phenomena and the potential impact of measurements in algorithmically infused societies requires scholars from different disciplines to be involved in these processes in order to negotiate the selection, specification and conceptualization of problems and phenomena. Scientific practices involve all kinds of value judgments, including deciding which problems are important and interesting enough to work on. We believe that it is crucial to make this process transparent and participatory to ensure that not only a diverse set of scholars is involved, but also that all relevant stakeholders are included, particularly those likely to experience adverse consequences.

Developing methodologies for integrating data and measurements into theory construction. Although measurement models that aim to quantify theoretical constructs (for example, political leaning or work–life balance) need to be connected with theory, theories may also benefit from integrating data and measurements into the theory construction process. This is especially the case in algorithmically infused societies where new socio-technical phenomena emerge frequently and where existing phenomena may evolve, and may thus require updating and rethinking existing theories.

An existing methodology that enables the integration of data into the theory development process is grounded theory[140]. The methodology is widely used in qualitative social science research, although it has been criticized by scholars who question whether it has fulfilled its promise to create new empirically based theories[141–143]. One potential weakness of grounded theory is its focus on induction. We believe that iterative abductive reasoning cycles provide a promising alternative for understanding new phenomena, which should not solely rely on existing theories and should not solely be explained from data. In such iterative cycles, computational methods—including machine learning and natural language processing methods[144–146] and empirically calibrated simulations and agent-based models[147–149]—may support both inductive and deductive reasoning by revealing and testing mechanisms and patterns that potentially explain social phenomena[36,150].

However, it remains an open question how different types of computational methods and measurements can best support experts in constructing theories—for example, by presenting measurement results in a human-interpretable way or by informing experts about anomalies that do not follow detected patterns[151,152]. As a community we need to develop methodologies and best practices for integrating data and computational methods into the theory construction to ensure that our theories are not informed by black box measurement models relying on unsupported assumptions, leaps of logic or biased data that do not capture the constructs and mechanisms of interest.

Outlook

We are already living in algorithmically infused societies, with algorithms shaping decisions that constantly influence and are influenced by human behaviour. The very essence of algorithmically infused societies intensifies concerns about (mis)measurement and the relationship between data and theory. For example, although algorithmically infused societies may provide many opportunities to enhance our theoretical and empirical understanding of the social world, poor practices for the development and evaluation of measurements and theories may also lead to “a new form of physiognomy”—the pseudoscience of inferring people’s inner states from their outer appearance[146,153]. In addition, measurements are not only scientific instruments that quantify and reflect the nature of subjects in the social world, but they also support predictions and explanations, influencing the construction of theories and the development of new algorithms. Therefore, they can also directly or indirectly shape the future of algorithmically infused societies. The development of responsible and trustworthy measurement models for algorithmically infused societies therefore requires careful reflections on the theoretical underpinning of measurements and their potential consequences. Finally, the evaluation of measurement models is often limited to how well a model characterizes some observable data (its goodness of fit) and/or how well it approximates the construct of interest (its validity and reliability); however, a model’s evaluation often neglects potential consequences of measurements. This is not only problematic because measurements (and also decisions on what not to measure) may have harmful consequences for individuals, social groups and society as a whole, but these consequences might also be hard to identify, hard to quantify, and even harder to rectify.

1. boyd, d. The future of privacy: how privacy norms can inform regulation. In *32nd Intl Conf. Data Protection and Privacy Commissioners* (2010); https://www.danah.org/papers/talks/2010/PrivacyGenerations.html.
2. Gill, K. S. The internet of things! Then what? *AI Soc.* **28**, 367–371 (2013).
3. O’Reilly, T. Open data and algorithmic regulation. In *Beyond Transparency: Open Data and the Future of Civic Innovation* (eds Goldstein, B. & Dyson, L.) 289–300 (Code for America Press, 2013).
4. Castells, M. *The Information Age: Economy, Society and Culture. Vol. 1: The Rise of the Network Society* (Wiley–Blackwell, 1996).
5. Fleder, D. & Hosanagar, K. Blockbuster culture’s next rise or fall: the impact of recommender systems on sales diversity. *Manage. Sci.* **55**, 697–712 (2009).
6. Tufekci, Z. Engineering the public: big data, surveillance and computational politics. *First Monday* **19**, https://doi.org/10.5210/fm.v19i7.4901 (2014).
7. Bakshy, E., Messing, S. & Adamic, L. A. Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130–1132 (2015).

8. Ferrara, E., Varol, O., Davis, C., Menczer, F. & Flammini, A. The rise of social bots. *Commun. ACM* **59**, 96–104 (2016).
9. Le Chen, A. M. & Wilson, C. An empirical analysis of algorithmic pricing on Amazon Marketplace. In *Proc. 25th Intl Conf. World Wide Web (WWW'16)* (eds Bourdeau, J. et al.) 1339–1349 (International World Wide Web Conferences Steering Committee, 2016); https://doi.org/10.1145/2872427.2883089.
10. Salganik, M. J., Sheridan Dodds, P. & Watts, D. J. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* **311**, 854–856 (2006).
11. Baym, N. K. *Playing to the Crowd: Musicians, Audiences, and the Intimate Work of Connection* (NYU Press, 2018).
12. Burgess, J. & Green, J. *YouTube: Online Video and Participatory Culture* (Wiley, 2018).
13. Hitsch, G. J., Hortaçsu, A. & Ariely, D. Matching and sorting in online dating. *Am. Econ. Rev*. **100**, 130–163 (2010).
14. Zignani, M. et al. Link and triadic closure delay: temporal metrics for social network dynamics. In *Proc. 8th Intl AAAI Conf. Web and Social Media* (eds Adar, E. & Resnick, P.) 564–573 (2014).
15. Malik, M. & Pfeffer, J. Identifying platform effects in social media data. In *Proc. 10th Intl AAAI Conf. Web and Social Media* (eds Krishna, G. & Strohmaier, M.) 241–249 (2016).
16. Su, J., Sharma, A. & Goel, S. The effect of recommendations on network structure. In *Proc. 25th Intl Conf. World Wide Web (WWW'16)* (eds Bourdeau, J. et al.) 1157–1167 (International World Wide Web Conferences Steering Committee, 2016); https://doi.org/10.1145/2872427.2883040.
17. Loscalzo, J. & Barabasi, A.-L. Systems biology and the future of medicine. *Wiley Interdiscip. Rev. Syst. Biol. Med*. **3**, 619–627 (2011).
18. Frizzell, J. D., et al. Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches. *JAMA Cardiol*. **2**, 204–209 (2017).
19. Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447–453 (2019).
20. Huang, C.-L., Chen, M.-C. & Wang, C.-J. Credit scoring with a data mining approach based on support vector machines. *Expert Syst. Appl*. **33**, 847–856 (2007).
21. Perry, W. L., McInnis, B., Price, C. C., Smith, S. C. & Hollywood, J. S. *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations* (RAND Corporation, 2013).
22. Dressel, J. & Farid, H. The accuracy, fairness, and limits of predicting recidivism. *Sci. Adv*. **4**, eaao5580 (2018).
23. Raghavan, M., Barocas, S., Kleinberg, J. & Levy, K. Mitigating bias in algorithmic hiring: evaluating claims and practices. In *Proc. 2020 Conf. Fairness, Accountability, and Transparency* (eds Hildebrandt, M. & Castillo, C.) 469–481 (ACM, 2020); https://doi.org/10.1145/3351095.3372828.
24. Hannák, A. et al. Bias in online freelance marketplaces: evidence from TaskRabbit and Fiverr. In *Proc. 2017 ACM Conf. Computer Supported Cooperative Work and Social Computing (CSCW'17)* (eds Lee, C. P. & Poltrock, S.) 1914–1933 (ACM, 2017); https://doi.org/10.1145/2998181.2998327.
This study reports on sociodemographic inequalities in online marketplaces.
25. Gray, M. L. & Suri, S. *Ghost Work: How to Stop Silicon Valley From Building a New Global Underclass* (Eamon Dolan Books, 2019).
26. Beer, D. The social power of algorithms. *Inf. Commun. Soc*. **20**, 1–13 (2017).
27. Kleinberg, J., Ludwig, J., Mullainathan, S. & Sunstein, C. R. Algorithms as discrimination detectors. *Proc. Natl Acad. Sci. USA* **117**, 30096–30100 (2020).
28. Buolamwini, J. & Gebru, T. Gender shades: intersectional accuracy disparities in commercial gender classification. In *Proc. 1st Conf. Fairness, Accountability and Transparency* (eds Friedler, S. A. & Wilson, C.) 77–91 (2018).
29. Moy, L. How police technology aggravates racial inequity: a taxonomy of problems and a path forward. *Univ. Illinois Law Rev*. **2021**, 139–193 (2021).
30. Hutchinson, B. & Mitchell, M. 50 years of test (un)fairness: lessons for machine learning. In *Proc. Conf. Fairness, Accountability, and Transparency* (eds Morgenstern, J. & boyd, d.) 49–58 (2019).
31. Barocas, S. & Selbst, A. D. Big data's disparate impact. *Calif. Law Rev*. **104**, 671–732 (2016).
32. Milli, S., Miller, J., Dragan, A. D. & Hardt M. The social cost of strategic classification. In *Proc. Conf. Fairness, Accountability, and Transparency* 230–239 (2019).
33. Bender, E. M., Gebru, T., McMillan-Major, A. & Shmitchell. S. On the dangers of stochastic parrots: can language models be too big? In *Proc. Conf. Fairness, Accountability, and Transparency* (eds Elish, M. C. et al.) 610–623 (2021).
34. Schnabel, T., Swaminathan, A., Singh, A., Chandak, N. & Joachims, T. Recommendations as treatments: debiasing learning and evaluation. In *Intl Conf. Machine Learning* (eds Balcan, M. F. & Weinberger, K. Q.) 1670–1679 (2016).
35. Sztompka, P. in *Polish Essays in the Methodology of the Social Sciences* (ed. Wiatr, J. J.) 173–194 (Springer, 1979).
36. Jaccard, J. & Jacoby, J. *Theory Construction and Model Building Skills: A Practical Guide for Social Scientists* (Guilford, 2010).
37. Lord, F. M. & Novick, M. R. *Statistical Theories of Mental Test Scores* (Addison-Wesley, 1968).
38. Allen, M. J. & Yen, W. M. *Introduction to Measurement Theory* (Waveland, 2002).
39. Joye, D., Wolf, C., Smith, T. W. & Fu, Y. in *The SAGE Handbook of Survey Methodology* (eds Wolf, C. et al.) 3–15 (Sage, 2016).
40. Strathern, M. 'Improving ratings': audit in the British university system. *Eur. Rev*. **5**, 305–321 (1997).
41. Campbell, D. T. Assessing the impact of planned social change. *Eval. Program Plann*. **2**, 67–90 (1979).
42. Festinger, L. *A Theory of Cognitive Dissonance* (Stanford Univ. Press, 1957).
43. Heider, F. Attitudes and cognitive organization. *J. Psychol*. **21**, 107–112 (1946).
44. Cartwright, D. & Harary, F. Structural balance: a generalization of Heider's theory. *Psychol. Rev*. **63**, 277–293 (1956).
45. McPherson, M., Smith-Lovin, L. & Cook, J. M. Birds of a feather: homophily in social networks. *Annu. Rev. Sociol*. **27**, 415–444 (2001).
46. Seitlinger, P., Kowald, D., Trattner, C. & Ley, T. Recommending tags with a model of human categorization. In *Proc. 22nd ACM Intl Conf. on Information & Knowledge Management* (eds He, Q. & Iyengar, A.) 2381–2386 (ACM, 2013).
47. Bowker, G. C. & Star, S. L. *Sorting Things Out: Classification and its Consequences* (MIT Press, 2000).
48. Healy, K. The performativity of networks. *Euro. J. Sociol*. **56**, 175–205 (2015).
This article argues that theories have the potential to reformat and reorganize the phenomena that models purport to describe.
49. Zuboff, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Profile Books, 2018).
50. Pasquale, F. *The Black Box Society: The Secret Algorithms that Control Money and Information* (Harvard Univ. Press, 2015).
51. Nissenbaum, H. *How Computer Systems Embody Values* (IEEE Computer Society Press, 2001).
52. Seaver, N. Knowing algorithms. In *DigitalSTS* (eds Vertesi, J. & Ribes, D.) 412–422 (Princeton Univ. Press, 2013).
53. boyd, d. & Crawford, K. Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Inf. Commun. Soc*. **15**, 662–679 (2012).
54. Graham, S. & Wood, D. Digitizing surveillance: categorization, space, inequality. *Crit. Soc. Policy* **23**, 227–248 (2003).
55. Benjamin, R. Catching our breath: critical race STS and the carceral imagination. *Engaging Sci. Technol. Soc*. **2**, 145–156 (2016).
56. Lazer, D. et al. Computational social science. *Science* **323**, 721–723 (2009).
This landmark article discussed the early potential of computational approaches for the social sciences.
57. Lazer, D. M. J. et al. Computational social science: obstacles and opportunities. *Science* **369**, 1060–1062 (2020).
58. Vosoughi, S., Roy, D. & Aral, S. The spread of true and false news online. *Science* **359**, 1146–1151 (2018).
59. Del Vicario, M. et al. The spreading of misinformation online. *Proc. Natl Acad. Sci. USA* **113**, 554–559 (2016).
60. Bail, C. A. et al. Exposure to opposing views on social media can increase political polarization. *Proc. Natl Acad. Sci. USA* **115**, 9216–9221 (2018).
61. Blalock, H. M. *Conceptualization and Measurement in the Social Sciences* (Sage, 1982).
62. RatSWD Quality Standards Working Group. *Quality Standards for the Development, Application, and Evaluation of Measurement Instruments in Social Science Survey Research* RatSWD Working Papers 245 (German Data Forum (RatSWD), 2015).
This paper proposes quality standards and guidelines for the development, application and evaluation of measurement instruments in social science survey research.
63. Jacobs, A. Z. & Wallach, H. Measurement and fairness. In *FAccT'21: Proc. 2021 ACM Conf. Fairness, Accountability, and Transparency* (eds Elish, M. C. et al.) 375–385 (ACM, 2019); https://doi.org/10.1145/3442188.3445901.
This paper describes how validity issues can lead to fairness issues.
64. Adcock, R. & Collier, D. Measurement validity: a shared standard for qualitative and quantitative research. *Am. Polit. Sci. Rev*. **95**, 529–546 (2001).
65. Hofman, J. M., Sharma, A. & Watts, D. J. Prediction and explanation in social systems. *Science* **355**, 486–488 (2017).
66. Peters, J., Janzing, D. & Schoelkopf, B. *Elements of Causal Inference: Foundations and Learning Algorithms* (MIT Press, 2017).
67. Jungherr, A. in *Digital Discussions: How Big Data Informs Political Communication* (eds Stroud, N. J. & McGregor, S.) 9–35 (Routledge, 2018).
68. Donoho, D. 50 years of data science. *J. Comput. Graph. Stat*. **26**, 745–766 (2017).
69. Blodgett, S. L., Lopez, G., Olteanu, A., Sim, R. & Wallach, H. Stereotyping Norwegian salmon: an inventory of pitfalls in fairness benchmark datasets. In *Proc. 59th Annual Meeting of the Association for Computational Linguistics* (ed. Zong, C.) (2021).
70. Ethayarajh, K. & Jurafsky, D. Utility is in the eye of the user: a critique of NLP leaderboard design. In *Proc. 2020 Conf. Empirical Methods in Natural Language Processing (EMNLP)* (eds Webber, B. et al.) 4846–4853 (2020).
71. Jungherr, A., Schoen, H., Posegga, O. & Jürgens, P. Digital trace data in the study of public opinion: an indicator of attention toward politics rather than political support. *Soc. Sci. Comput. Rev*. **35**, 336–356 (2017).
72. Samory, M., Sen, I., Kohne, J., Flöck, F. & Wagner, C. Call me sexist, but...: Revisiting sexism detection using psychological scales and adversarial samples. In *Intl AAAI Conf. Web and Social Media* 573–584 (2021).
73. Gebru, T. et al. Datasheets for datasets. Preprint at https://arxiv.org/abs/1803.09010 (2018).
74. Olteanu, A., Castillo, C., Diaz, F. & Kıcıman, E. Social data: biases, methodological pitfalls, and ethical boundaries. *Front. Big Data* **2**, 13 (2019).
75. Sen, I., Floeck, F., Weller, K., Weiss, B. & Wagner, C. A total error framework for digital traces of human behavior on online platforms. *Public Opin. Q*. (in the press).
This paper gives a systematic overview of errors that may be introduced when analysing digital traces of human behavior.
76. Lazer, D. Issues of construct validity and reliability in massive, passive data collections. *The City Papers: An Essay Collection from The Decent City Initiative* http://citiespapers.ssrc.org/issues-of-construct-validity-and-reliability-in-massive-passive-data-collections/ (2015).
77. Perdomo, J., Zrnic, T., Mendler-Dünner, C. & Hardt, M. Performative prediction. In *Proc. 37th Intl Conf. Machine Learning* (eds Daumé III, H. & Singh, A.) 7599–7609 (PMLR, 2020).
78. Hannak, A. et al. Measuring personalization of web search. In *Proc. 22nd Intl Conf. World Wide Web (WWW'13)* (eds Schwabe, D. et al.) 527–538 (ACM, 2013).
79. Ali, M. et al. Discrimination through optimization: how Facebook's ad delivery can lead to biased outcomes. *Proc. ACM Hum.–Comp. Interact*. **3**, 199 (2019).
80. Thomas, P. & Brunskill, E. Data-efficient off-policy policy evaluation for reinforcement learning. In *Intl Conf. Machine Learning* (eds Balcan, M. F. & Weinberger, K. Q.) 2139–2148 (PMLR, 2016).
81. Sinha, A., Gleich, D. F. & Ramani, K. Deconvolving feedback loops in recommender systems. In *Advances in Neural Information Processing Systems 29* (eds Lee, D. et al.) 3243–3251 (2016).
82. Tomašev, N. et al. AI for social good: unlocking the opportunity for positive impact. *Nat. Commun*. **11**, 2468 (2020).

83. Mau, S. *Das metrische Wir: Über die Quantifizierung des Sozialen* (Suhrkamp, 2017). **This book explores the implications of measurements on social systems**.
84. D'Ignazio, C. & Klein, L. F. *Data Feminism* (MIT Press, 2020).
85. Muller, J. Z. *The Tyranny of Metrics* (Princeton Univ. Press, 2018).
86. Lee, M. K., Jain, A., Cha, H. J., Ojha, S. & Kusbit, D. Procedural justice in algorithmic fairness: leveraging transparency and outcome control for fair algorithmic mediation. *Proc. ACM Hum.–Comp. Interact*. **3**, 182 (2019).
87. Weller, A. Challenges for transparency. Preprint at https://arxiv.org/abs/1708.01870 (2017).
88. Shokri, R., Strobel, M. & Zick, Y. Exploiting transparency measures for membership inference: a cautionary tale. In *The AAAI Workshop on Privacy-Preserving Artificial Intelligence (PPAI)* (eds Fioretto, F. et al.) 17 (AAAI, 2020).
89. Shokri, R., Strobel, M. & Zick, Y. On the privacy risks of model explanations. Preprint at https://arxiv.org/abs/1907.00164 (2019).
90. Lahoti, P. et al. Fairness without demographics through adversarially reweighted learning. In *Advances in Neural Information Processing Systems 33* (eds Larochelle, H. et al.) 728–740 (2020).
91. Dwork, C., Hardt, M., Pitassi, T., Reingold, O. & Zemel, R. Fairness through awareness. In *Proc. 3rd Innovations in Theoretical Computer Science Conf*. (eds Goldwasser, S.) 214–226 (ACM, 2012).
92. Hardt, M., Price, E. & Srebro, N. Equality of opportunity in supervised learning. In *Proc. 30th Intl Conf. on Neural Information Processing Systems* (eds Lee, D. et al.) 3323–3331 (Curran, 2016).
93. Werner, D. *Nothing About Us Without Us: Developing Innovative Technologies for, by and with Disabled Persons* (Healthwrights, 1998).
94. Charlton, J. I. *Nothing About Us Without Us: Disability Oppression and Empowerment* (Univ. California Press, 1998).
95. Costanza-Chock, S. Design justice, A.I., and escape from the matrix of domination. *J. Design Sci*. https://doi.org/10.21428/96c8d426 (2018).
96. Scott, J. C. *Seeing Like a State* (Yale Univ. Press, 2008).
97. Lazer, D. et al. Meaningful measures of human society in the twenty-first century. *Nature* https://doi.org/10.1038/s41586-021-03660-7 (2021).
98. Goel, S., Anderson, A., Hofman, J. & Watts, D. J. The structural virality of online diffusion. *Manage. Sci*. **62**, 180–196 (2016).
99. Aral, S. & Nicolaides, C. Exercise contagion in a global social network. *Nat. Commun*. **8**, 14753 (2017).
100. Eagle, N., Macy, M. & Claxton, R. Network diversity and economic development. *Science* **328**, 1029–1031 (2010).
101. Contractor, N. in *The Oxford Handbook of Networked Communication* (eds Welles, B. F. and González-Bailón, S.) (Oxford Univ. Press, 2018).
102. Contractor, N., Monge, P. R. & Leonardi, P. M. Multidimensional networks and the dynamics of sociomateriality: bringing technology inside the network. *Int. J. Commun*. **5**, 682–720 (2011).
103. Rahwan, I. et al. Machine behaviour. *Nature* **568**, 477–486 (2019). **This paper argues that studies of machine behaviour are necessary to control AI-enabled systems and to avoid harm**.
104. Salganik, M. J. *Bit By Bit: Social Research in the Digital Age* (Princeton Univ. Press, 2017). **This book gives an overview of methods deployed in computational social science**.
105. Baeza-Yates, R. Big data or right data? In *Proc. 7th Alberto Mendelzon Intl Workshop on Foundations of Data Management* (eds Bravo, L. & Lenzerini, M.) 14 (CEUR-WS, 2013).
106. Schwarz, G. & Stensaker, I. Time to take off the theoretical straightjacket and (re-) introduce phenomenon-driven research. *J. Appl. Behav. Sci*. **50**, 478–501 (2014).
107. Mathieu, J. E. The problem with [in] management theory. *J. Organ. Behav*. **37**, 1132–1141 (2016).
108. Watts, D. Should social science be more solution-oriented? *Nat. Hum. Behav*. **1**, 15 (2017).
109. Stier, S., Breuer, J., Siegers, P. & Thorson, K. Integrating survey data and digital trace data: key issues in developing an emerging field. *Soc. Sci. Comput. Rev*. **38**, 503–516 (2020).
110. Mellon, J. Internet search data and issue salience: the properties of google trends as a measure of issue salience. *J. Elections Public Opin. Parties* **24**, 45–72 (2014).
111. Stier, S., Bleier, A., Lietz, H. & Strohmaier, M. Election campaigning on social media: politicians, audiences, and the mediation of political communication on Facebook and Twitter. *Polit. Commun*. **35**, 50–74 (2018).
112. Bernard, H. R., Killworth, P. D. & Sailer, L. Informant accuracy in social-network data: V. An experimental attempt to predict actual communication from recall data. *Soc. Sci. Res*. **11**, 30–66 (1982).
113. Prince, S. A. et al. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int. J. Behav. Nutr. Phys. Act*. **5**, 56 (2008).
114. Scharkow, M. The accuracy of self-reported internet use—a validation study using client log data. *Commun. Methods Meas*. **10**, 13–27 (2016).
115. Boase, J. & Ling, R. Measuring mobile phone use: self-report versus log data. *J. Comput. Mediat. Commun*. **18**, 508–519 (2013).
116. Revilla, M., Ochoa, C. & Loewe, G. Using passive data from a meter to complement survey data in order to study online behavior. *Soc. Sci. Comput. Rev*. **35**, 521–536 (2017).
117. Elmer, T., Chaitanya, K., Purwar, P. & Stadtfeld, C. The validity of RFID badges measuring face-to-face interactions. *Behav. Res. Methods* **51**, 2120–2138 (2019).
118. Sapiezynski, P., Stopczynski, A., Lassen, D. D. & Lehmann, S. Interaction data from the Copenhagen Networks Study. *Sci. Data* **6**, 315 (2019).
119. Groves, R. M. & Lyberg, L. Total survey error: past, present, and future. *Public Opin. Q*. **74**, 849–879 (2010).
120. Coravos, A., Chen, I., Gordhandas, A. & Stern, A. D. We should treat algorithms like prescription drugs. *Quartz*, https://qz.com/1540594/treating-algorithms-like-prescription-drugs-could-reduce-ai-bias/ (February 2019).
121. Arnold, M. et al. FactSheets: increasing trust in AI services through supplier's declarations of conformity. Preprint at https://arxiv.org/abs/1808.07261 (2019).
122. Bender, E. & Friedman, B. Data statements for natural language processing: toward mitigating system bias and enabling better science. *Trans. Assoc. Comput. Linguist*. **6**, 587–604 (2018).
123. Mitchell, M. et al. Model cards for model reporting. In *Proc. Conf. Fairness, Accountability, and Transparency* (eds Morgenstern, J. & boyd, d.) 220–229 (ACM, 2019). **This paper argues for higher standards in documenting machine learning models**.
124. Kuh, A., Petsche, T. & Rivest, R. Learning time-varying concepts. In *Advances in Neural Information Processing Systems 3* (eds Lippmann, R. P. et al.) 183–189 (Morgan-Kaufmann, 1991).
125. Bartlett, P. L., Ben-David, S. & Kulkarni, S. R. Learning changing concepts by exploiting the structure of change. *Mach. Learn*. **41**, 153–174 (2000).
126. Gama, J., Žliobaitundefined, I., Bifet, A., Pechenizkiy, M. & Bouchachia, A. A survey on concept drift adaptation. *ACM Comput. Surv*. **46**, 44 (2014).
127. Abbasi, M., Friedler, S. A., Scheidegger, C. & Venkatasubramanian, S. Fairness in representation: quantifying stereotyping as a representational harm. In *Proc. 2019 SIAM Intl Conf. Data Mining* (eds Berger-Wolf, T. & Chawla, N.) 801–809 (SIAM, 2019).
128. Abebe, R. et al. Roles for computing in social change. In *Proc. 2020 Conf. Fairness, Accountability, and Transparency* (eds Castillo, C. & Hildebrandt, M.) 252–260 (ACM, 2020).
129. Hampton, L. M. Black feminist musings on algorithmic oppression. In *Proc. 2021 Conf. Fairness, Accountability, and Transparency* (eds Elish, M. C. et al.) 1 (ACM, 2021).
130. De-Arteaga, M., Fogliato, R. & Chouldechova, A. A case for humans-in-the-loop: decisions in the presence of erroneous algorithmic scores. In *Proc. 2020 CHI Conf. Human Factors in Computing Systems* (eds Bernhaupt, R. et al.) 1–12 (ACM, 2020).
131. Boyarskaya, M., Olteanu, A. & Crawford, K. Overcoming failures of imagination in AI infused system development and deployment. Preprint at https://arxiv.org/abs/2011.13416 (2020).
132. Nanayakkara, P., Diakopoulos, N. & Hullman, J. Anticipatory ethics and the role of uncertainty. Preprint at https://arxiv.org/abs/2011.13170 (2020).
133. Friedman, B. Value-sensitive design. *Interaction* **3**, 16–23 (1996).
134. Olteanu, A., Diaz, F. & Kazai, G. When are search completion suggestions problematic? *Proc. ACM Hum.–Comp. Interact*. **4**, 1–25 (2020).
135. Jiang, J. A., Wade, K., Fiesler, C. & Brubaker, J. R. Supporting serendipity: opportunities and challenges for human–AI collaboration in qualitative analysis. *Proc. ACM Hum.–Comp. Interact*. **5**, 1–23 (2021).
136. Churchill, E., van Allen, P. & Kuniavsky, M. Designing AI: introduction. *Interaction* **25**, 34–37 (2018).
137. Selbst, A. D. boyd, d., Friedler, S., Venkatasubramanian, S. & Vertesi, J. Fairness and abstraction in sociotechnical systems. In *Proc. Conf. Fairness, Accountability, and Transparency* (eds Morgenstern, J. & boyd, d.) 59–68 (ACM, 2019).
138. Barocas, S., Biega, A. J., Fish, B., Niklas, J. & Stark, L. When not to design, build, or deploy. In *Proc. 2020 Conf. Fairness, Accountability, and Transparency* (eds Castillo, C. & Hildebrandt, M.) 695–695 (ACM, 2020).
139. Monge, P. & Contractor, N. *Theories of Communication Networks* (Oxford Univ. Press, 2003).
140. Glaser, B. & Strauss, A. *The Discovery of Grounded Theory: Strategies for Qualitative Research* (Aldine de Gruyter, 1967).
141. Bryant, A. Re-grounding grounded theory. *J. Inf. Technol. Theory Appl*. **4**, 25–42 (2002).
142. Charmaz, K. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis* (Sage, 2006).
143. Timmermans, S. & Tavory, I. Theory construction in qualitative research: from grounded theory to abductive analysis. *Sociol. Theory* **30**, 167–186 (2012).
144. Nelson, L. Computational grounded theory: a methodological framework. *Sociol. Methods Res*. **49**, 3–42 (2020). **This paper proposes a methodological framework to combine expert human knowledge and hermeneutic skills with the processing power and pattern recognition of computers**.
145. McFarland, D., Lewis, K. & Goldberg, A. Sociology in the era of big data: the ascent of forensic social science. *Am. Sociol*. **47**, 12–35 (2015).
146. Radford, J. & Joseph, K. Theory in, theory out: the uses of social theory in machine learning for social science. *Front. Big Data* **3**, 18 (2020).
147. Macy, M. & Willer, R. From factors to actors: computational sociology and agent-based modeling. *Annu. Rev. Sociol*. **28**, 143–166 (2002).
148. Smith, E. R. & Conrey, F. R. Agent-based modeling: a new approach for theory building in social psychology. *Pers. Soc. Psychol. Rev*. **11**, 87–104 (2007).
149. Keuschnigg, M., Lovsjö, N. & Hedström, P. Analytical sociology and computational social science. *J. Comp. Soc. Sci*. **1**, 3–14 (2018).
150. Hedström, P. & Bearman, P. in *The Oxford Handbook of Analytical Sociology* (eds Hedström, P. & Bearman, P.) (Oxford Univ. Press, 2011).
151. Lemmerich, F. et al. Mining subgroups with exceptional transition behavior. In *Proc. 22nd ACM SIGKDD Intl Conf. Knowledge Discovery and Data Mining* (eds Krishnapuram, B. & Shah, M.) 965–974 (ACM, 2016).
152. Singer, P. et al. Why we read Wikipedia. In *Proc. 26th Intl Conf. World Wide Web* (eds Barrett, R. & Cummings, R.) 1591–1600 (ACM, 2017).
153. Aguera y Arcas, B., Mitchell, M. & Todorov, A. Physiognomy's new clothes. *Medium*, https://medium.com/@blaisea/physiognomys-new-clothesf2d4b59fdd6a (6 May 2017).

Acknowledgements We thank the reviewers for their contributions to this work.

Author contributions All authors have contributed to the formation, discussion, and writing of this article.

Competing interests The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to C.W.
Peer review information *Nature* thanks Ceren Budak, Johan Ugander and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.