Atomist or Holist?
A Diagnosis and Vision for more Productive Interdisciplinary AI Ethics Dialogue

Greene, Travis; Dhurandhar, Amit; Shmueli, Galit

Perspective

Atomist or holist? A diagnosis and vision for more productive interdisciplinary AI ethics dialogue

Travis Greene,[1,*] Amit Dhurandhar,[2] and Galit Shmueli[1]
[1]National Tsing Hua University, Institute of Service Science, Hsinchu 30013, Taiwan
[2]IBM Research, Yorktown Heights, NY, USA
*Correspondence: travis.greene@iss.nthu.edu.tw
https://doi.org/10.1016/j.patter.2022.100652

THE BIGGER PICTURE As artificial intelligence (AI)-based technologies increasingly impact society, AI ethics is more relevant than ever. A growing number of data scientists argue that ethical values related to the desirability of social and political goals can and should influence AI research and applications. An opposing group of data scientists disagrees, asserting the importance of technological progress, academic freedom, and value neutrality. We diagnose this deep conflict within the data science community as one between atomist and holist ideologies. With the goals of encouraging civil discourse across academic disciplines and reducing community polarization, we draw on a variety of philosophical and historical sources to describe each ideology's core beliefs and assumptions. We call on atomists and holists to exhibit greater empathy during ethical disagreements and propose four targeted strategies to ensure data science research benefits society.

1 2 3 4 5 **Concept:** Basic principles of a new data science output observed and reported

SUMMARY

In response to growing recognition of the social impacts of new artificial intelligence (AI)-based technologies, major AI and machine learning (ML) conferences and journals now encourage or require papers to include ethics impact statements and undergo ethics reviews. This move has sparked heated debate concerning the role of ethics in AI research, at times devolving into name calling and threats of "cancellation." We diagnose this conflict as one between "atomist" and "holist" ideologies. Among other things, atomists believe facts are and should be kept separate from values, while holists believe facts and values are and should be inextricable from one another. With the goal of reducing disciplinary polarization, we draw on numerous philosophical and historical sources to describe each ideology's core beliefs and assumptions. Finally, we call on atomists and holists within the ever-expanding data science community to exhibit greater empathy during ethical disagreements and propose four targeted strategies to ensure AI research benefits society.

INTRODUCTION

In his 1917 lecture Science as a Vocation, Max Weber argues that questions of fact are separate from questions of value.[1] Given pre-specified ends, the task of science is to find the best means of achieving them. But asking which ends we ought to achieve is a question of values, answerable only by philosophy or religion. Weber was not alone on this point. Eminent mathematicians, physicists, and economists espouse similar views on the separation of science and values. But a growing number of data science researchers see things differently. They argue that values, particularly those related to the desirability of social

and political goals, implicitly influence the foundations of data science practices.[2–4]

As public concern mounts over the use of artificial intelligence (AI)-driven research to enable surveillance technologies, deep fakes, biased language models,[5] misinformation, addictive behaviors, and the discriminatory use of facial recognition and emotion-detection algorithms,[6] data scientists appear divided along ideological lines about what to do. While some may view the inclusion of ethics impact statements in several AI conferences and journals as a sign that the value-neutral ideal of science is no longer tenable, a vocal cadre of data scientists has pushed back against this conclusion, asserting the importance

Table 1. Summary of two basic, but often conflicting, ideologies in the data science community along several core dimensions

	Atomist	Holist
Guiding metaphor	puzzle solver	social steward
Facts and values	separate	inseparable
Associated "isms"	(neo)liberalism, libertarianism, logical positivism, modernism	communitarianism, feminism, post-positivism, post-modernism
Social orientation	individualist	collectivist
Self-concept	autonomous	relational
Means of social coordination	incentives and markets	shared moral values and dialogic exchange
Key moral concepts	rights, duties, contracts, impartial justice	empathy, caring, connection, responsivity to vulnerable others
Vision of the good life	neutral and constrained	substantive and unconstrained
Scientific methodology	data driven, empiricist, neutral	theory laden, rationalist, perspectival
Extreme form leads to	technocracy/nihilism/alienation	totalitarianism/dogmatism/tribalism

of academic freedom and value neutrality.[7] Yet the conflict between these two ideological camps threatens to polarize the data science community, lowering the prospects that new AI-based technologies will contribute to our collective well-being.

This largely historical and conceptual perspective is organized as follows in the following sections. Atomism versus holism: an overview of conflict ideologies lays out the conflicting ideological views—atomism and holism—underlying AI ethics debates. The "two cultures" within the data science community contrasts the "two cultures" of atomism and holism by presenting a guiding metaphor for each. The separation of facts and values relates philosophical debate on facts and values to community concerns about ethically evaluating AI-based technologies. Social benefits of ethics impact statements considers how ethics impact statements embody a shift toward a more holistic data science culture and discusses the social benefits of value-based discussions. Atomist concerns regarding ethics statements, in response, identifies key sticking points of ethics reviews from the perspective of an atomist data scientist. Finally, Targeted recipes for more productive interdisciplinary AI ethics discussions presents a broad vision for more empathetic discussion of value-related issues and proposes four targeted strategies for more civil discourse.

ATOMISM VERSUS HOLISM: AN OVERVIEW OF CONFLICTING IDEOLOGIES

An ideology is an all-encompassing worldview advancing a political and ethical vision of the good society.[8] To better understand the nature of ethical disagreements in data science, we introduce a simple ideological taxonomy we call atomism and holism, drawing on the terminology and work of philosophers[9–12] and intellectual historians.[13,14] At risk of oversimplification, the purpose of our taxonomy is to make each ideology's implicit beliefs, assumptions, and historical foundations more explicit so they can be reflected on, refined, and more openly discussed within the data science community should ethical conflict arise[15] (see Table 1). In the interest of full disclosure, we ourselves have mixed academic backgrounds and self-identify at different points along the ideological spectrum: one is mostly holist and the other mostly atomist, while another is a "moderate," finding

aspects of both sides agreeable. We encourage readers to also reflect on where they might fit within this taxonomy.

To delineate the scope of the arguments and examples, we use the label "data scientist" to cover both industry data scientists who design and deploy socially impactful AI/machine learning (ML)-based systems, and data scientists—often in academia but also in industry—whose job descriptions includes publishing research articles in peer-reviewed conferences and journals. Our broad definition includes AI/ML researchers and AI/ML engineers. We use the terms data science and AI/ML interchangeably.

Conflicting visions of self, society, and politics
Our diagnosis begins by observing that the conflict between atomist and holist data scientists mirrors philosophical debates between traditional liberals (we use the term liberal in Isaiah Berlin's sense of "neutral with respect to conceptions of the good,"[12] although we are aware it has taken on different connotations in modern-day political discourse) and communitarians and care ethicists.[16–18] The conflict centers on the nature of and relationship between the self and society, the proper balance between social order and personal autonomy,[19] and whether a just society should advance any particular view of the good life.[20] Atomists embody more of an individualist culture—inwardly focused on his or her individual needs and interests—while holists embody more of a collectivist culture—outwardly focused on the needs and interests of the social collective.[21] Atomists and holists thus offer opposing solutions to the basic dilemma facing every society: that of protecting its members from one another without oppressing them.[19] Both atomists and holists claim that the opposing ideology—when taken to the extreme—leads to various social and individual pathologies.

Holist data scientists view their identities as relationally embedded in a larger society with a particular and unique history and culture. The relational identity of holists entails various relationships of responsibility and caring for others.[17,18] Holists stress the self as concretely situated and constituted through caring relations to particular others. As such, holists believe that a good life and society should foster attentive, caring relations and cultivate empathy for the experiences and suffering

of vulnerable others.[16,17] Holist data scientists recognize that their technical expertise allows them to understand AI technologies in a way that politicians and ordinary citizens might not. Their privileged knowledge creates an obligation to communicate these concerns publicly so that citizens and policy-makers can act appropriately.[22] Holists thus see themselves as social stewards or fiduciaries acting on behalf of society, holding an "unconstrained" view of the perfectability of human nature and society.[14] Limiting the freedom of other data scientists in order to protect and prevent harm to vulnerable third parties, marginalized groups, and the environment thus may be a necessary evil in the utopian quest toward the ideal society.

Atomists, as the name implies, tend to see the world reductively,[11] holding that what separates us is prior to what connects us.[17] In political matters, atomists emphasize separateness and independence from others while focusing on abstract issues of impartial justice, duties, and rights. They stress the importance of formal equality under the law and the freedom to enter into voluntary contracts of exchange in markets; they prefer economic systems in which social coordination is achieved through the pursuit of enlightened self-interest rather than that imposed by an external authority.[23] Atomists—wary of utopia—see the achievement of the good life as tragically constrained by unavoidable trade-offs, limited resources, and human imperfections.[14,24] Finally, atomists value personal autonomy and take violations of their personal integrity seriously.[25] Personal integrity refers to the life projects to which a person is committed.[26] Atomist data scientists feel the imposition of moral responsibility to others unfairly infringes on their freedom to pursue their life projects as they see fit, a view that reflects their individualist view of self and society.

Conflicting visions of the role of science in society

Mirroring economic arguments for specialization and trade, atomists support a gap between fact-based inquiry and value-based inquiry. The job of the technically trained scientist is to conduct fact-based inquiry and report experimental results. In contrast, the job of policy-makers is to advance the state of society by acting on these results according to political and ethical values.[27] How others interpret the facts, and what they do on the basis of these interpretations, is out of the control of the scientist. The atomist thus defends the empiricist position that a judgment of fact entails nothing about value, and vice versa.[28] Factual statements, strictly speaking, do not motivate action or decision[29] (i.e., ought cannot be derived from is). Atomists argue that intellectual specialization and division of labor—i.e., a logical gap between value-neutral fact finding and value-based decision-making[10]—is a feature, not a bug, of the arrangement. Hence, atomists see the imposition of ethics impact statements as hampering scientific fact finding, which is essential to innovation and economic progress.

Atomists are also concerned about how the injection of values can lead to—in the best case—wishful thinking and—in the worst case—to dogmatism and totalitarianism.[10] Atomists cite how Soviet science was notoriously influenced by its interpretation of Marxist philosophy or how Galileo was forced to recant his support of Copernicanism under the inquisition of the Catholic church. The dogmatic and totalitarian tendencies of holists were recently summarized by one data science academic who condemned "the

increasing use of repressive actions aimed at limiting the free and unfettered conduct of scientific research and debate."[7] By claiming that science is and ought to be value free, or at least value neutral,[30] atomists advocate a form of inductive empiricism that traces back to Francis Bacon and Galileo.[31]

Meanwhile, holists advocate a post-positivist attitude to science where facts and values mutually inform one another.[32] They doubt that there is a uniquely correct or absolute view of reality[33] and emphasize the role of interpretation and perspective.[34] Holists thus reject the value-free ideal of science and the atomist ban on deriving an ought from an is. Instead, holists believe facts discovered by social/data science ought to be put toward realizing a substantive, utopian vision of the good life and society.[35–37] Free and open discussion of values allows for dialogue aimed at finding overlapping consensus about which values are and should be embodied in data science research and applications. Without a clear articulation of and commitment to shared moral values, new AI technologies may impede the achievement of the good life.[38] In particular, holists fear the creation of a faceless and unaccountable technocracy narrowly aimed at prediction and control of human behavior resulting from the "value-neutral" instrumental application of science.[39] In short, holists view ethics impact statements as tools for clarifying shared action-guiding values, developing moral character, and inducing democratic deliberation on the responsibility of data scientists to society.

THE "TWO CULTURES" WITHIN THE DATA SCIENCE COMMUNITY

In 1959, at the height of Cold War, and as the US military-industrial complex established itself, scientist and writer C.P. Snow worried about a growing divide between two academic cultures—those from the "hard sciences" and the "humanities"—whose specialization rendered them increasingly hostile and unmotivated to communicate with one another.[40] We suggest that a similar dynamic may be stoking division within the larger data science community. Inspired by philosopher and historian of science Thomas Kuhn, we sketch two guiding metaphors capturing core differences between rival atomist and holist research communities.

Atomists: Data scientists as puzzle solvers

Through a process of disciplinary socialization and training, atomists see themselves as acquiring a constellation of shared beliefs, values, and techniques—in short, a paradigm—that permits progress on open problems, or "puzzles," that the paradigm identifies as solvable.[41] Commitment to the paradigm identifies a researcher as a member of a distinct scientific community. During periods of "normal" science (as opposed to revolutionary science), the legitimacy of the paradigm's values and traditions is presumed, narrowing researchers' focus on the task of more reliably and efficiently gathering relevant facts.[41] In AI and data science research communities, for instance, shared standards and traditions might take the form of the 0.05 threshold for p values in statistical hypothesis testing[27] or the use of several standard benchmark datasets (e.g., ImageNet, CIFAR-10, Fashion MNIST, etc.) and performance metrics (e.g., precision, recall, or F1 score).

The chief task of the atomist data scientist is, through sheer creativity and imagination, to make progress in solving the puzzles posed by the paradigm, not to question its legitimacy. Beyond the familiar boundaries of the paradigm, there is neither consensus on what a right answer looks like nor a guarantee that a solution even exists. Atomists concede that while the results of AI research may later be used for what outsiders consider morally good or bad ends, these external judgments ultimately rely on criteria and values foreign to the paradigm. Because the atomist's scientific identity stems from loyalty to the paradigm, atomists worry that undue focus on external social issues not only slows down puzzle solving but threatens both the integrity of the paradigm and their personal autonomy.

Holists: Data scientists as social stewards
Holist data scientists see the growing public concern over the social impact of AI as anomalous facts signifying a crisis and ultimately demanding a paradigm shift.[41] Holists propose revolutionary changes in perspective and new disciplinary procedures, open problems, and traditions. In the emergent holist paradigm, data scientists are social stewards[42] or fiduciaries working on behalf of society and advancing substantive social values and human interests[43] through data science research and applications.

The holist conviction that data scientists ought to act as social stewards or fiduciaries is advocated by a growing number of data scientists, particularly in high-stakes healthcare domains.[44,45] The ethical justification of data scientists' fiduciary duty to society derives from their capacity to "produce consequences that matter to others" (Goodin,[46] p. 110). In particular, holists are concerned about unjust power differentials and coercive dependency relationships[17] that may arise due to applications of AI-based technologies. The role of social steward or fiduciary aligns with holist beliefs that the self is constituted through social relations and that social responsibilities and caring relations are important for our psychological well-being.[37]

A fiduciary relation is a special kind of social relation involving discretion, power, inequality, dependence, vulnerability, trust, and confidence.[47] The word fiduciary itself stems from *fide*, meaning "trustworthy." Fiduciaries have the power and authority to make decisions on behalf of the beneficiary. Common fiduciary relations include trustee-beneficiary, principal-agent, manager-shareholder, lawyer-client, doctor-patient, and parent-child relations. The presumed fiduciary power of holists may be justified on the basis of their unequal access to knowledge of technical details and social applications of emerging AI-based technologies. Given this knowledge asymmetry, fiduciaries abide by a duty of loyalty to act in the best interests of the beneficiary and refrain from opportunism and conflicts of interest.[47] Oriented by a substantive view of the human good, and motivated by a belief in unlimited social progress, holists feel obligated to ensure AI-based technologies promote human flourishing.[48]

THE SEPARATION OF FACTS AND VALUES

Atomists and holists disagree about how facts and values relate and their role in data science. To better clarify the fact-value debate, this section examines the issues in more historical detail.

At stake is whether values can be subjected to rational scrutiny and thereby provide objective grounds for ethical criticism of AI technologies.

The fact-value controversy: A brief historical interlude
The crux of the fact-value debate centers on whether it is possible to rationally criticize (i.e., by appeal to logic or empirical evidence) both the means used to achieve an end and the end itself. Empiricists, following David Hume, tend to argue that it is not possible to rationally critique ends; rationalists, following Immanuel Kant, disagree.

Hume's empiricism was revived in the 20th century in the philosophy of science known as logical positivism or logical empiricism, which asserted that philosophical problems ultimately stemmed from ambiguities in ordinary language.[49] This highly ambitious project aimed to unify all sciences using formal logic and derive experimentally verifiable hypotheses from formalized versions of scientific theories.[50] Asserting empirical verification as the criterion for meaning was an important methodological move that paved the way for behaviorism in psychology, revealed preference theory in economics,[51] and operationalism in psychometrics.[52]

The focus on empirical verification of scientific statements challenged the idea of objectivity in ethics. What methods of proof or evidence could support ethical judgments? To overcome this problem, some philosophers suggested that ethical judgments were actually composed of two separate components: a descriptive or factual part and a command-like prescriptive component.[53] Others concluded that values were emotive or attitudinal in nature and lacking in descriptive, factual content, analogous to cries of pain, groans, or shrieks.[49] Thus, to say "X is good" is nothing more than saying "hurrah for X!" or "I like X; you should too."[54] Logical positivists, in short, claimed that moral language is simply not "about" anything at all, casting serious doubt on efforts to resolve moral disagreements.

Despite its onetime popularity, philosophers of science abandoned the logical positivist project as early as the 1960s.[55] Most philosophers today accept that facts and values are not so cleanly separable.[56] Still, vestiges of the logical positivist project remain influential in other disciplines,[57] particularly those fond of axiomatic formalization such as economics and computer science.

Implications for moral expertise and objectivity
Evaluating AI-based technologies presumably requires moral expertise. But if the notion of a moral fact is incoherent, how can we make sense of the idea of moral expertise? Does moral expertise imply the existence of objective moral facts, such as "killing is wrong"?[28] Even if we are willing to grant the existence of such facts, must moral experts not only possess theoretical but also practical moral knowledge—the real-world skills needed to correctly apply moral concepts in the right situations?[58] Although developing a procedure for testing the validity of moral judgments appears difficult, if not impossible, moral philosophers nevertheless routinely undertake this task.[59,60]

Another problem is that in most scientific domains, we expect experts to reach consensus on key facts as it indicates convergence on the truth.[56] Yet, moral disagreement is common, and value statements seemingly do not achieve a high degree of

consensus, even among experts.[61] Nearly all data scientists will agree that when training an ML model, both bias and variance cannot be minimized simultaneously, but presumably fewer agree with "value-laden" statements such as a predictive model's accuracy is more important than its interpretability. The lack of consensus around value statements, especially in an era of globalized data science, could be a sign that objective knowledge in the domain of AI ethics may not be possible. Without such objectivity, the legitimacy of publication decisions that include ethical evaluation may be disputed.

Despite these difficulties and the persistent arguments of moral skeptics, we hope to encourage—not discourage—ethical reflection, especially by those without formal training in philosophy. As AI-based technologies increasingly impact society, AI ethics is more relevant than ever. But a more central role for ethics in AI research means that members of long-estranged academic disciplines will be forced to engage with one another, disrupting the current intellectual division of labor[62] and posing new barriers to communication. Still, we believe that the disruption of disciplinary boundaries can benefit society.

SOCIAL BENEFITS OF ETHICS IMPACT STATEMENTS

One major step toward acknowledging the ethical issues raised by AI was the decision by the Neural Information Processing Systems (NeurIPS) conference in 2020 to introduce a new ethics review process. The process requires submitting authors to include a "broader impact" section, while a group of ethics experts review papers flagged as problematic by technical reviewers.[63] Following the example of NeurIPS, impact statements and ethics reviews by experts are spreading to other conferences and journals.[64] The IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR) in 2022 adapted the NeurIPS ethics guidelines and "strongly encourages" authors to "discuss the ethical and societal consequences of their work in their papers in a concrete manner."[65] But creating effective ethics impact statements remains challenging. Impact statements are plagued by a variety of issues including their complexity, lack of guidance and best practices, lack of purpose, lack of procedural transparency, high opportunity costs, institutional pressure, and various social and cognitive biases affecting authors and referees.[66]

Nevertheless, data scientists can no longer presume that their research will have a net positive impact on the world.[67] Besides functioning as a kind of layman's guide to new AI-based technology, the inclusion of impact statements is a positive step toward acknowledging the social responsibility of those who design, research, and deploy AI technologies. Below, we identify several ways in which impact statements may benefit data science research and society in general.

Train "citizen" data scientists and cultivate moral and intellectual virtues

Writing and critiquing an ethics statement could become a core part of the moral education and training of future data scientists. Impact statements encourage moral reflection and the cultivation of civic and intellectual virtues, such as attentive observation, open-minded imagination, patient reflection, careful analysis, and fair-minded interpretation and assessment.[68] The

development of moral and intellectual virtues could break down interdisciplinary academic barriers isolating researchers from public debates on AI-based technologies.[69] Hagendorff,[70] for instance, suggests cultivating four basic virtues: justice, honesty, responsibility, and care, along with secondary virtues of prudence and fortitude (speaking truth to power). Impact statements can promote the virtue of humility, too, by requiring data scientists to better understand and communicate the limitations of new AI tools, reducing opportunities for sensationalism in the popular media.[71] To develop these virtues, data science curricula might offer Embedded EthiCS[72] courses to foster discussion on the ethical implications of AI-based technologies and provide training in writing ethics impact statements.

Promote disciplinary cross-pollination

The big-picture scope of ethics reviews and the diversity of the review groups can contribute to disciplinary cross-pollination. Novel dilemmas, formalizations, and technologies can be brought back and analyzed in different academic communities. These issues can then become part of the ethical training of data scientists. Data science research also confronts legal scholars and regulators with thorny problems of interpretation and application of existing laws, which could lead to better lawmaking. Lastly, cross-pollination can occur when data science researchers experimentally test the assumptions of traditional ethical theories,[73] leading to theory refinement in ethics and moral psychology. For example, deep neural nets can make accurate predictions about the morality of various novel acts (e.g., "it's rude to mow the lawn late at night") as judged by human evaluators.[74,75]

Spur new debates on the purpose and applications of corporate data science research

Technologists increasingly voice their concerns when corporate applications of technology conflict with their ethical values.[76,77] Value-based discourse allows debate on the social and ethical impact of corporate-funded data science applications. One AI-driven application likely to lead to moral disagreement in the data science community is digital advertising.[78] Jeff Hammerbacher, one of Facebook's first data scientists, once said, "The best minds of my generation are thinking about how to make people click ads … That sucks."[79] Indeed, Google and Facebook are two of the world's largest technology companies employing data scientists and also two of the biggest players in digital advertising. Massive industry investment reflects the allocation of vast amounts of data scientist labor to solving problems related to digital advertising. But beyond the obvious economic benefits, what values justify this enormous outlay of human energy and attention?

Identify and inspire new forms of data science for social good projects

A more transparent debate about values can expand the scope of data scientists' responsibility to new moral communities and stimulate new ideas for data science for social good. Ethics debates can identify new forms of harm before they become systematically entrenched in business practices and research pipelines. Interesting ethical blind spots of AI concern how computer vision, natural language processing (NLP), and other

technologies contribute to environmental degradation,[4] the exploitation of human data labelers—particularly in low-income countries[80] and in NLP crowdsourcing tasks[81,82]—and the mistreatment of animals in factory farms.[83] Another angle involves developing algorithms to "boost"[84,85] critical thinking about the quality and veracity of news stories shared on social media in an effort to foster democratic values.

Reveal the sociotechnical, political, and value-laden nature of applied data science

Although one might reasonably argue that we should distinguish between more abstract AI research (e.g., optimization and machine architecture) and more applied work (e.g., education and criminal justice), ethics impact statements can helpfully illuminate the sociotechnical nature of data science. In real-world institutions such as courts or corrections facilities, legal terms and predictive relationships evolve, actors act adversarially, and numerous non-sampling errors can affect algorithmic performance.[86] Not only that, many decisions in ML pipelines involve "essentially contested" and value-laden constructs such as fairness[87] and justice.[88] Sociotechnically oriented data science recognizes that fairness is more than just a property of an algorithm:[89] AI can be used as part of a political agenda.[3] One noteworthy example comes from Barabas et al.,[90] who refused to build an algorithm due to concerns about its role in reinforcing structural injustices in the prison system.

ATOMIST CONCERNS REGARDING ETHICS STATEMENTS

This section considers possible concerns and arguments arising from atomist data scientists.

Lack of technical credibility

Multiple cases, including our own personal experience, exist where research is rejected or red flagged for apparent ethical concerns. Sometimes, however, the rejection is unwarranted, as the contribution of the paper may be orthogonal or even a mechanism to better understand AI model risks. For example, an explainable AI paper might be rejected because the model it explains (viz. a large language model) has fairness issues. But the model itself is not the contribution. Undesirable behavior uncovered through explanations is not a limitation but rather a validation of the explanation method. A rejection decision showcases ethics reviewers' inability to understand the work's technical contribution. In larger contexts, AI tools and systems might fare similarly if not properly assessed.

Lack of intellectual freedom

Rightly or wrongly, many data scientists currently see ethics requirements as constraints, not just algorithmically but also intellectually. Many see their purpose as discouraging work that does not adhere to certain "ethical" standards. Such rhetoric can engender a defensive attitude toward those demanding these standards for fear of one's ideas being shunned and suppressed. Fostering a defensive research culture could have a chilling effect on data science progress and, in some cases, even lead to "black market" research being done away from ethicists' eyes. Ideally, when practitioners encounter ethical di-

lemmas, they should feel comfortable reporting and debating them with the larger research community, not suppressing them for fear of ridicule or punishment.

Limiting progress for an application/problem

In 2004, a chess program called Fruit[91] was open sourced, which had a much lower playing strength than top human grandmasters at the time. However, it led to a revolution where ideas encoded in this initial version were imbibed in other programs and improved upon, leading to the current generation of programs such as Stockfish. These programs are nearly unbeatable by humans and are now commonly used by top players for training and in the discovery of new ideas. This was a clear case where imperfect versions can still lead to significant progress. Refinement can happen over time. Analogously in AI, good ideas could be killed early if too many ethical restrictions are imposed upfront.

Limiting progress across applications/problems

Likewise, some ideas or tricks used to develop one system or application are often useful for different applications. For example, transformer architectures[92] were initially shown to be successful in NLP tasks; however, they are now preferred even in computer vision tasks.[93] Hence, the organic transfer of ideas can have impacts far beyond the original application. Yet this intellectual transfer may be severely curtailed if strict, yet subjective, constraints are placed on what is ethically acceptable.

### Subjectivity of ethics standards

Values can be highly subjective at an individual level and even at the level of regulatory institutions. For instance, the requirements of the General Data Protection Regulation (GDPR)[94] in Europe vary from those outlined by NIST[95] in the US. Values are also hard to precisely specify and formalize. This can lead to frustration for persons in academia and industry alike. Publishing in respected venues is already challenging enough. A PhD student trying to meet publication requirements could get frustrated with additional and vague requirements that resemble a moving target. Similar ambiguities could betide engineering teams in industry offering products and services. But subjectivity is not the same as arbitrariness. Robust ethical guidelines and principles drawn from biomedical fields might be adapted for AI research (though this is easier said than done[88]).

Neutrality of most technology

A car could be used to transport an ailing person to the hospital and save their life, or it could run over someone. Analogously in AI, a fairness algorithm intended to improve fairness could be used to output unfair results by misspecifying protected attributes. In line with the atomist's commitment to value neutrality, technology can be used for good or bad and may not be inherently harmful. Hence, putting restrictions on it might make little sense. Rather, applications should be monitored and constrained.

TARGETED RECIPES FOR MORE PRODUCTIVE INTERDISCIPLINARY AI ETHICS DISCUSSIONS

Given the ideological differences of atomists and holists, we adapt a therapeutic technique to restore trust between them.
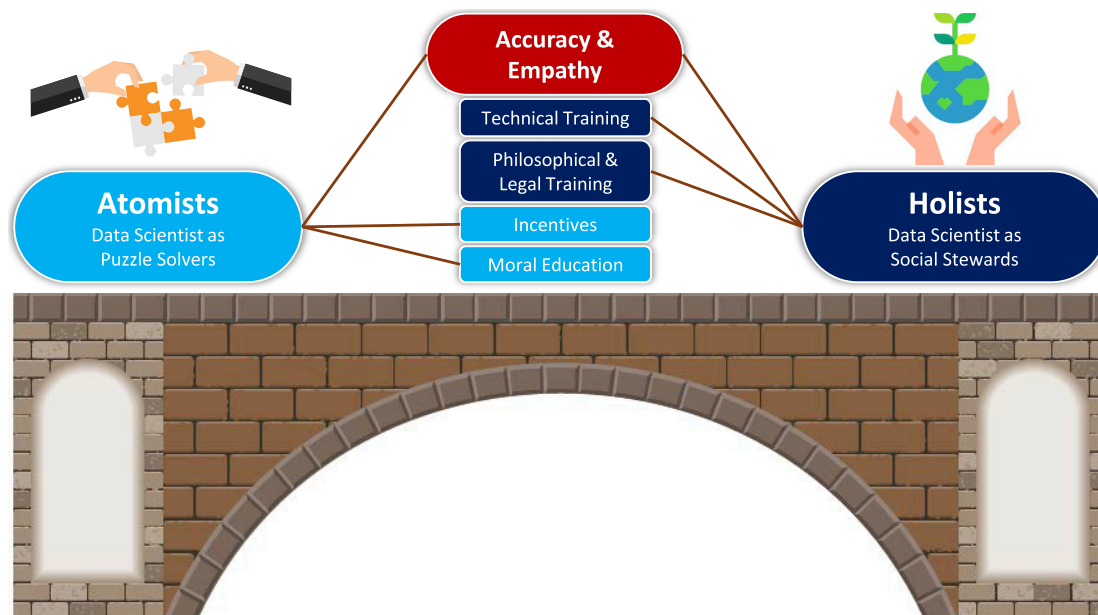
Figure 1. The data science community risks splitting into two opposing factions—atomists and holists—over contentious AI ethics issues
To promote more accurate and empathetic AI ethics discussions, we propose four discipline-targeted recipes to bridge these ideological divisions, reduce community polarization, and ensure that AI research benefits society.

Guided by twin themes of greater accuracy and empathy (see Figure 1 schematic), we then present four basic recipes—targeted at different data science community member backgrounds—for fostering more productive dialogue on AI ethics issues to benefit society (see Table 2 for examples).

A therapeutic model for interdisciplinary accuracy and empathy

Carl Rogers in 1951 proposed a person-centered approach to psychotherapy useful for resolving interpersonal conflict.[96] Importantly, this approach leaves dialogue participants feeling more satisfied and less resentful regardless of outcome. A similar dialogical process may be useful in improving the quality of AI ethics discussions in general. The core idea is the following:

Accuracy: Alice first expresses her worry. Bob then repeats it as accurately as possible. This repetition should not be an interpretation of what Bob thinks Alice meant but as accurate a reproduction as possible. Afterward, Bob asks Alice if he missed something and, if so, tries again until Alice is satisfied.

Empathy: Once Alice is satisfied with Bob's reformulation, Bob empathizes with her concern. He imagines Alice's perspective and follows her line of reasoning, although he may not agree with it. Then, Bob makes his point, which now Alice must accurately repeat. The process continues until both parties are satisfied. A fictional conversation might proceed as follows:

- Alice: I spent the last few weeks testing Schmoogle's newest large language model for biases. I am now convinced it might be sentient.
- Bob: You said you "spent the last few weeks testing Schmoogle's newest large language model for biases and now you are convinced it is sentient." Is that right?

- Alice: Yes, that is right. We might want to ask for consent before we continue running A/B tests on it.
- Bob: I understand why you would think such a complex language model might be sentient. If I spent as much time as you testing it, I would also be concerned. But the model is just finding statistical correlations in massive amounts of text from all over the internet.
- Alice: You said, "The model is just finding statistical correlations in massive amounts of text data from all over the internet." Is that right?
- Bob: …

Moral education for data scientists

Moral education is about the socialization and development of virtuous persons: persons with ethically admirable character traits and dispositions who empathize with and care for particular others in their community.[18] Today, however, most data scientists receive a highly technical education. Skills, not virtues—excellences that contribute to a flourishing human life and community—are the focus. Understandably, data science graduates may implicitly apply their disciplinary standards of objectivity to ethical issues, realize it cannot easily be done, and conclude that anything goes or that ethical judgments are meaningless. Yet, neglecting the moral education of data scientists can leave open a moral vacuum liable to be filled by tribalism, dogmatic identity politics, relativism, or nihilism. Data science educators would do well to emphasize the history and arguments behind the rejection of the fact-value distinction,[57] as these points seem to have gone unnoticed in AI-feeder disciplines.

Greater debate about the ideal character and virtues of data scientists can also pave the way for professionalization. Example debate starters might be the following: under what sort of

Table 2. Examples of data science community member backgrounds and positions in the atomist-holist taxonomy

	Atomist	Holist
Implicit	engineering, physics, computer science, statistics	technology activism
Explicit	economics, quantitative social science	humanities, law, qualitative social science

Note that ideological assumptions can be implicit or explicit, largely cutting across disciplinary divides.

conditions is corporate whistleblowing appropriate? What is the difference between ethics consulting and fairwashing? To what extent should data scientists be expected to reasonably foresee the harmful use of AI-based technologies by others? While we cannot speak for the community, at the very least, the moral education of future data scientists should aim at cultivating an ability to empathize, tolerate, and negotiate in good faith with those whose views differ from their own without resorting to name calling and threats of cancellation.

Incentives for economists
To encourage atomists to cooperate with holists to tackle the complex ethical issues posed by AI-based technologies, one suggestion draws on the old idea that commerce and morality are mutually reinforcing.[97] By providing economic incentives toward more civil engagements on AI ethics issues, we can also respect atomists' desire for autonomy and personal integrity.

At the institutional level, academic data science departments might incentivize faculty to explore co-teaching data science ethics courses with philosophers and lawyers. Senior researchers and practitioners could also be incentivized to mentor junior data scientists, focusing on character development. Industry data scientists could be rewarded for convening "ethics roundtable" discussions or ethics reading groups. Conferences and journals with ethics reviews and impact statements could also offer prizes or community recognition to authors whose statements are particularly cogent or creative.[66] Similarly, positive examples of civil debate—whether on social media or offline—could be rewarded through community praise. The actions of certain "moral exemplars" can become teaching material for data science ethics courses. In general, organizations can do more to frame ethics issues as new and exciting research challenges, rather than as autonomy-restricting handcuffs.

Philosophical and legal training for technology activists
While typically trained in the natural sciences, technology activists seek to draw attention to social justice issues related to AI and data science, such as racial and gender discrimination in STEM fields and industry hiring practices. Technology activists author popular books[98,99] and have received major media attention in news articles[77,100] and films (e.g., The Social Dilemma). Although we sympathize with activists' intentions, we also worry that threatening the cancellation[101] of those who hold opposing values can stoke fears of tribalism and dogmatic groupthink, sparking further disciplinary polarization. Indeed, given activists' concern with the unjust use of power, threats of mob-based cancellation may appear hypocritical.

Echoing similar proposals from science and technology studies and digital humanities scholars,[102,103] we encourage philosophical training to hone technology activists' persuasive and reflective abilities and create new allies and collaborators in the humanities, social sciences, and law. With broad support from the often-fragmented "two cultures" of the natural and human sciences, technology activists can arguably better achieve their ethical goals. Philosophy provides an array of concepts to express ethical concerns (e.g., rights, duties, utility, virtue, power, care) and can aid in identifying *ad hominem* attacks and logical fallacies. Familiarity with various areas of law (e.g., corporate law, public law, torts) can also help technology activists frame AI harms in terms of pre-existing legal frameworks and concepts, such as discrimination or human rights law.[104]

Technical training for humanities scholars
For those in the humanities (e.g., philosophy, law, media and communication studies) and "soft" social sciences (e.g., anthropology, sociology, education) who might participate in ethics reviews, we think it fair to ask whether they have the "appropriate" technical expertise to accurately evaluate data science products, systems, and/or papers. For instance, there is a growing literature that draws on post-modern philosophical sources to explain how AI can entrench existing racial and gender inequities[105] and reproduce the dynamics of colonial exploitation and power differentials.[4,106] While we welcome critical voices into the debate on the ethics of AI, we must clearly distinguish features of the technology in itself from uses of the technology by organizations embedded in larger social, political, and economic systems.

One suggestion is to establish an accreditation process for ethics reviewers, such as a minimum-duration internship in industry or an AI-related academic department (e.g., computer science or statistics). The knowledge and skills obtained could improve both the accuracy and empathy of cross-disciplinary dialogue. Real-world experience would help ethics reviewers to develop their practical moral knowledge and increase their ability to empathize with data scientists. Accreditation could be done in collaboration with a number of interdisciplinary research groups and industry research labs.

Conclusion
The data science community appears split—sometimes violently so—along an ideological divide. We explicated key beliefs and assumptions of atomist and holist ideologies and described their relevance to the fact-value debate. Atomists see themselves as autonomous puzzle solvers and hold that facts are and should be kept separate from values, while holists see themselves as social stewards and believe that facts and values mutually reinforce one another. In the interests of the data science community and society, we advocate for a balance between the two camps' views and offer several recipes for more empathetic and productive interdisciplinary dialogue on AI ethics. This work can serve as a microcosm of such dialogue, as we ourselves occupy various points along the atomist-holist spectrum.

But our diagnosis and proposal have limitations. First, some recipes may not translate well to digital and social media contexts, where the most-heated AI ethics arguments tend to occur. Second, the atomist-holist taxonomy obscures heterogeneity

among data scientists who may hold less traditional views. Lastly, because many advances in AI happen inside technology corporations, corporations arguably should assume the burdens of legal and moral responsibility for the harms and risks imposed by their innovations. More systemic solutions—at the level of corporate law and governance[107]—are likely needed to ensure that AI-based technology is developed and deployed with the broader interests of society in mind, not simply shareholders.

We hope our diagnosis and vision draws attention to the importance of greater empathy, openness, and humility for all members of the data science community, a community whose boundaries continue to evolve and expand. We look forward to a time when community ethics discussions are dominated by the best reasons rather than the loudest voices.

AUTHOR CONTRIBUTIONS

T.G. and A.D. conceived of and wrote the manuscript, and G.S. provided supervision and feedback.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Weber, M. (1958). Science as a vocation. Daedalus 87, 111–134.

2. Friedler, S.A., Scheidegger, C., and Venkatasubramanian, S. (2021). The (im) possibility of fairness: different value systems require different mechanisms for fair decision making. Commun. ACM 64, 136–143.

3. Green, B. (2021). Data science as political action: grounding data science in a politics of justice. J. Soc. Comput. 2, 249–265.

4. Crawford, K. (2021). The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence (Yale University Press).

5. Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.-S., Mellor, J., Glaese, A., Cheng, M., Balle, B., Kasirzadeh, A., Biles, C., Brown, S., et al. (2022). Taxonomy of risks posed by language models. In 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 214–229.

6. Kaye, K. (2022). Not my job: ai researchers building surveillance tech and deepfakes resist ethical concerns. Protocol, Online, accessed August 9, 2022. https://www.protocol.com/enterprise/ai-computer-vision-cvpr-ethics.

7. Signatories, I. (2020). An Open Letter to the Communications of the Acm.

8. Heywood, A. (2021). Political Ideologies: An Introduction (Bloomsbury Publishing).

9. Taylor, C. (1979). Atomism. In Powers, possessions and freedom: Essays in Honour of C.B. Macpherson, A. Kontos, ed. (University of Toronto Press), pp. 39–62.

10. Popper, K. (2012). The Open Society and its Enemies (Routledge).

11. Bunge, M. (2000). The alternative to individualism and holism. J. Soc. Econ. 29, 147–157.

12. Berlin, I. (2002). Liberty (Oxford University Press).

13. Skinner, Q. (2002). Visions of Politics (Regarding Method)1 (Cambridge University Press).

14. Sowell, T. (2002). A Conflict of Visions: Ideological Origins of Political Struggles (Basic Books (AZ)).

15. Domingos, P. (2021). Beating back cancel culture: a case study from the field of artificial intelligence. Quillette, Online, accessed July 27, 2022. https://quillette.com/2021/01/27/beating-back-cancel-culture-a-case-study-from-the-field-of-artificial-intelligence/.

16. Slote, M. (2007). The Ethics of Care and Empathy (Routledge).

17. Held, V. (2006). The Ethics of Care: Personal, Political, and Global (Oxford University Press on Demand).

18. Noddings, N. (2013). Caring: A Relational Approach to Ethics and Moral Education (Univ of California Press).

19. Etzioni, A. (1996). The responsive community: a communitarian perspective. Am. Sociol. Rev. 61, 1–11.

20. Sandel, M.J. (1984). The procedural republic and the unencumbered self. Polit. theory 12, 81–96.

21. Triandis, H.C. (2018). Individualism and Collectivism (Routledge).

22. Anderson, E. (2011). Democracy, public policy, and lay assessments of scientific testimony. Episteme 8, 144–164.

23. Hayek, F.A. (1980). Individualism and Economic Order (University of chicago Press).

24. Crowder, G. (2006). Value pluralism and communitarianism. Contemp. Polit. Theory 5, 405–427.

25. Smiley, M. (2009). Moral Responsibility and the Boundaries of Community (University of Chicago Press).

26. Smart, J.J.C., and Williams, B. (1973). Utilitarianism: For and against (Cambridge University Press).

27. Rudner, R. (1953). The scientist qua scientist makes value judgments. Philos. Sci. 20, 1–6.

28. Elgin, C.Z. (2013). Fact–value distinction. Int. Encycl. Ethics.

29. Hume, D. (2003). A Treatise of Human Nature (Courier Corporation).

30. Betz, G. (2013). In defence of the value free ideal. Eur. J. Philos. Sci. 3, 207–220.

31. Lacey, H. (2005). Is Science Value Free?: Values and Scientific Understanding (Routledge).

32. Anderson, E. (1995). Value in Ethics and Economics (Harvard University Press).

33. Rorty, R. (2009). Philosophy and the Mirror of Nature (Princeton University Press).

34. Denzin, N.K., and Lincoln, Y.S. (2011). The Sage Handbook of Qualitative Research (Sage).

35. Gorski, P.S. (2013). Beyond the fact/value distinction: ethical naturalism and the social sciences. For. Soc. 50, 543–553.

36. Dewey, J. (1998). The Essential Dewey: Pragmatism, Education, Democracy1 (Indiana University Press).

37. Sayer, A. (2009). Who's afraid of critical social science? Curr. Sociol. 57, 767–786.

38. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., et al. (2018). Ai4people—an ethical framework for a good ai society: opportunities, risks, principles, and recommendations. Minds Mach. 28, 689–707.

39. Habermas, J. (1985). The Theory of Communicative Action: Volume 2: Lifeword and System: A Critique of Functionalist Reason, 2 (Beacon press).

40. Snow, C.P. (2012). The Two Cultures (Cambridge University Press).

41. Kuhn, T.S. (1970). The Structure of Scientific Revolutions (University of Chicago Press).

42. Bak-Coleman, J.B., Alfano, M., Barfuss, W., Bergstrom, C.T., Centeno, M.A., Couzin, I.D., Donges, J.F., Galesic, M., Gersick, A.S., Jacquet, J., et al. (2021). Stewardship of global collective behavior. Proc. Natl. Acad. Sci. USA *118*. e2025764118.

43. Habermas, J. (1971). Knowledge and Human Interests (Beacon).

44. Eaneff, S., Obermeyer, Z., and Butte, A.J. (2020). The case for algorithmic stewardship for artificial intelligence and machine learning technologies. JAMA *324*, 1397–1398.

45. Stöger, K., Schneeberger, D., and Holzinger, A. (2021). Medical artificial intelligence: the european legal perspective. Commun. ACM *64*, 34–36.

46. Goodin, R.E. (1986). Protecting the Vulnerable: A Re-analysis of Our Social Responsibilities (University of Chicago Press).

47. Miller, P.B. (2013). Justifying fiduciary duties. mlj. *58*, 969–1023.

48. Taddeo, M., and Floridi, L. (2018). How ai can be a force for good. Science *361*, 751–752.

49. Ayer, A.J. (1952). Language, Truth, and Logic (Courier Corporation).

50. Rorty, R. (1992). The Linguistic Turn: Essays in Philosophical Method (University of Chicago Press).

51. Lewin, S.B. (1996). Economics and psychology: lessons for our own day from the early twentieth century. J. Econ. Lit. *34*, 1293–1323.

52. Borsboom, D. (2005). Measuring the Mind: Conceptual Issues in Contemporary Psychometrics (Cambridge University Press).

53. Hare, R.M. (1991). The Language of Morals (Oxford Paperbacks).

54. Stevenson, C.L. (1944). Ethics and Language (Yale University Press).

55. Suppe, F. (2000). Understanding scientific theories: an assessment of developments, 1969-1998. Philos. Sci. *67*, 102–115.

56. Williams, B. (2006). Ethics and the Limits of Philosophy (Routledge).

57. Putnam, H. (2004). The Collapse of the Fact/value Dichotomy and Other Essays (Harvard University Press).

58. Weinstein, B.D. (1994). The possibility of ethical expertise. Theor. Med. *15*, 61–75.

59. Rawls, J.A. (1971). Theory of Justice (Harvard University Press).

60. Habermas, J. (1990). Moral Consciousness and Communicative Action (MIT Press).

61. Steinkamp, N.L., Gordijn, B., and ten Have, H.A.M.J. (2008). Debating ethical expertise. Kennedy Inst. Ethics J. *18*, 173–192.

62. Nowotny, H. (2005). The increase of complexity and its reduction: emergent interfaces between the natural sciences, humanities and social sciences. Theory Cult. Soc. *22*, 15–31.

63. Ashurst, C., Hine, E., Sedille, P., and Carlier, A. (2022). Ai ethics statements: analysis and lessons learnt from neurips broader impact statements. In 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 2047–2056.

64. Anonymous. Tmlr ethics guidelines. TMLR Ethics Guidelines, Online, accessed August 12, 2022. https://www.jmlr.org/tmlr/ethics.html.

65. Anonymous (2022). Ethics guidelines. The IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR), Online, accessed August 9, 2022. https://cvpr2022.thecvf.com/ethics-guideline.

66. Prunkl, C.E.A., Ashurst, C., Anderljung, M., Webb, H., Leike, J., and Dafoe, A. (2021). Institutionalizing ethics in ai through broader impact requirements. Nat. Mach. Intell. *3*, 104–110.

67. Hecht, B., Wilcox, L., Bigham, J.P., Schöning, J., Hoque, E., Ernst, J., Bisk, Y., De Russis, L., Yarosh, L., Anjum, B., et al. (2021). It's time to do something: mitigating the negative impacts of computing through a change to the peer review process. Preprint at arXiv. https://doi.org/10.48550/arXiv.2112.09544.

68. Baehr, J. (2011). The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology (OUP Oxford).

69. Greene, T., Martens, D., and Shmueli, G. (2022). Barriers to academic data science research in the new realm of algorithmic behaviour modification by digital platforms. Nat. Mach. Intell. *4*, 323–330.

70. Hagendorff, T. (2020). Ai virtues–the missing link in putting ai ethics into practice. Preprint at arXiv. https://doi.org/10.48550/arXiv.2011.12750.

71. Johnson, D.G., and Verdicchio, M. (2017). Ai anxiety. J. Assoc. Inf. Sci. Technol. *68*, 2267–2270.

72. Grosz, B.J., Grant, D.G., Vredenburgh, K., Behrends, J., Hu, L., Simmons, A., and Waldo, J. (2019). Embedded ethics: integrating ethics across cs education. Commun. ACM *62*, 54–61.

73. Wallach, W., and Allen, C. (2008). Moral Machines: Teaching Robots Right from Wrong (Oxford University Press).

74. Jiang, L., Hwang, J.D., Bhagavatula, C., Le Bras, R., Forbes, M., Borchardt, J., Liang, J., Etzioni, O., Sap, M., and Choi, Y. (2021). Delphi: towards machine ethics and norms. Preprint at arXiv. https://doi.org/10.48550/arXiv.2110.07574.

75. Hendrycks, D., Burns, C., Basart, S., Critch, A., Li, J., Song, D., and Steinhardt, J. (2020). Aligning ai with shared human values. Preprint at arXiv. https://doi.org/10.48550/arXiv.2008.02275.

76. Davis, G.F. (2021). Corporate purpose needs democracy. J. Manage. Stud. *58*, 902–913.

77. (2021). Google fires margaret mitchell, another top researcher on its ai ethics team. The Guardian, Online, accessed February 24, 2022. https://www.theguardian.com/technology/2021/feb/19/google-fires-margaret-mitchell-ai-ethics-team.

78. Milano, S., Mittelstadt, B., Wachter, S., and Russell, C. (2021). Epistemic fragmentation poses a threat to the governance of online targeting. Nat. Mach. Intell. *3*, 466–472.

79. Baer, D. (2013). Why data god jeffrey hammerbacher left facebook to found cloudera. Bloomberg, Online, accessed August 6, 2022. https://www.fastcompany.com/3008436/why-data-god-jeffrey-hammerbacher-left-facebook-found-cloudera.

80. Kshetri, N. (2021). Data labeling for the artificial intelligence industry: economic impacts in developing countries. IT Prof. *23*, 96–99.

81. Shmueli, B., Fell, J., Ray, S., and Ku, L.W. (2021). Beyond Fair Pay: Ethical Implications of NLP Crowdsourcing. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 3758–3769.

82. Santy, S., Rani, A., and Choudhury, M. (2021). Use of formal ethical reviews in nlp literature: historical trends and current practices. Preprint at arXiv. https://doi.org/10.48550/arXiv.2106.01105.

83. Hagendorff, T. (2021). Blind spots in ai ethics. AI Ethics *2*, 851–867.

84. Hertwig, R., and Grüne-Yanoff, T. (2017). Nudging and boosting: steering or empowering good decisions. Perspect. Psychol. Sci. *12*, 973–986.

85. Lorenz-Spreen, P., Lewandowsky, S., Sunstein, C.R., and Hertwig, R. (2020). How behavioural sciences can promote truth, autonomy and democratic discourse online. Nat. Hum. Behav. *4*, 1102–1109.

86. Greene, T., Shmueli, G., Fell, J., Lin, C.-F., and Liu, H.-W. (2022). Forks over knives: predictive inconsistency in criminal justice algorithmic risk prediction tools. J. Roy. Stat. Soc. 1–32.

87. Angerschmid, A., Zhou, J., Theuermann, K., Chen, F., and Holzinger, A. (2022). Fairness and explanation in ai-informed decision making. Mach. Learn. Knowl. Extr. (2019). *4*, 556–579.

88. Mittelstadt, B. (2019). Principles alone cannot guarantee ethical ai. Nat. Mach. Intell. *1*, 501–507.

89. Selbst, A.D., Boyd, D., Friedler, S.A., Venkatasubramanian, S., and Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In Proceedings of the conference on fairness, accountability, and transparency, pp. 59–68.

90. Barabas, C., Doyle, C., Rubinovitz, J., and Dinakar, K. (2020). Studying up: reorienting the study of algorithmic fairness around issues of power. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 167–176.

91. Letouzey, F. (2014). Fruit 1.0 Uci (With Source Code) Now Available (Computer Chess Club Archives).

92. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems, pp. 5998–6008.

93. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2021). An image is worth 16x16 words: transformers for image recognition at scale. In Intl Conference on Learning Representations.

94. Yannella, P.N., and Kagan, O. (2018). Analysis: article 29 working party guidelines on automated decision making under gdpr. https://www.cyberadviserblog.com/2018/01/analysis-article-29-working-party-guidelines-on-automated-decision-making-under-gdpr/.

95. Tabassi, E. (2021). Analysis: article 29 working party guidelines on automated decision making under gdpr. https://www.nist.gov/programs-projects/trustworthy-and-responsible-ai.

96. Rogers, C. (1951). Client-centered Therapy: Its Current Practice, Implications and Theory (Constable).

97. Hirschman, A.O. (1982). Rival interpretations of market society: civilizing, destructive, or feeble? J. Econ. Lit. 20, 1463–1484.

98. O'neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (Broadway books).

99. Broussard, M. (2018). Artificial Unintelligence: How Computers Misunderstand the World (MIT Press).

100. Cadwalladr, C. (2018). 'i made steve bannon's psychological warfare tool': meet the data war whistleblower. The Guardian, Online, accessed August 6, 2022. https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump.

101. Anandkumar, A. (2020). My heartfelt apology. Anima on AI, Online, accessed August 6, 2022. https://anima-ai.org/2020/12/16/my-heartfelt-apology/.

102. Berry, D. (2012). Understanding Digital Humanities (Springer).

103. Riley, D. (2008). Engineering and Social Justice (Morgan & Claypool Publishers).

104. Aizenberg, E., and van den Hoven, J. (2020). Designing for human rights in ai. Big Data Soc. 7 205395172094956.

105. Benjamin, R. (2019). Race after Technology: Abolitionist Tools for the New Jim Code (John Wiley & Sons).

106. Couldry, N., and Mejias, U.A. (2019). Data colonialism: rethinking big data's relation to the contemporary subject. Telev. New Media 20, 336–349.

107. Stout, L.A. (2012). The Shareholder Value Myth: How Putting Shareholders First Harms Investors, Corporations, and the Public (Berrett-Koehler Publishers).

Travis Greene is a PhD candidate at the Institute of Service Science at National Tsing Hua University in Taiwan. With an interdisciplinary background in philosophy and research interests in data science ethics and personalization, his work has appeared in journals such as *Nature Machine Intelligence*, *The Journal of the Royal Statistical Society*, and *Big Data*. He hopes to contribute new perspectives and ideas from the humanities into machine learning and statistics, and vice versa.

Amit Dhurandhar is a principal research scientist at IBM Research. His recent work focuses on building tools and techniques to appropriate trust in AI systems. His work has been featured in Forbes, PC magazine, and Quartz and in venues such as NeurIPS, *Science*, and *Nature Communications*. His research received the AAAI deployed application award and was selected as Best of ICDM twice. He also co-led the creation of the AI Explainability 360 open-source toolkit. He has served as area chair for top AI venues as well as on NSF panels and currently serves on the invention disclosure committee (IDT) at IBM.

Galit Shmueli is a Tsing Hua distinguished professor at the Institute of Service Science, College of Technology Management, National Tsing Hua University, in Taiwan. Her research focuses on statistical and machine-learning methodology with applications in information systems and healthcare and an emphasis on human behavior. She is the author of multiple books and over 100 peer-reviewed publications. She teaches and designs business analytics courses on machine learning, forecasting, and more. She is also the inaugural editor in chief of the *INFORMS Journal on Data Science* and an IMS Fellow and ISI elected member.