THE DATA SHOULD NOT SPEAK FOR ITSELF: EPISTEMIC INJUSTICE AND DATA AS RHETORIC

Susan Haarman
Loyola University Chicago

Change the instruments, and you will change the entire social theory that goes with them.
- Bruno Latour

In the spring of 2020, I (along with nearly every other instructor at my university) moved my course online. While finishing out the semester felt at times like the equivalent of trying to do major surgery with a Band-Aid and a pocket flashlight, I did walk away from the experience having far better utilized our learning management system (LMS). From background information present in student profiles to click rates on articles and study guides, the LMS gave me a larger picture of my students and their participation. When I noticed click rates on the articles I posted flagging, I experimented with giving them podcasts instead. When I saw that students were only reading and responding to the first two student forum posts, I assigned conversation partners. Despite the transition, my teaching evaluations were stronger than ever, thanks in part to the data the LMS was giving me. A week after the semester ended, I received an email promotion from the LMS parent company offering a workshop to show me how to further maximize my teaching and predict student learning from the data gathered. They claimed that COVID-19's push to broader engagement with the platform and increased data had allowed them unprecedented levels of analysis and insight. I would be able to "know more about my students and how they learn than they know about themselves!"

As education becomes more digital in delivery in the wake of COVID-19, it generates a massive amount of data. The datafication of education is neither surprising nor new.[1] Collected data at all levels of educational systems from the individual student to national profiles can include everything from test scores, teaching choices, attendance, and management decisions.[2] Arguments for advancement and progress in education have conflated technology and datafication with improvement, and that false causality is used to justify further

[1] Juliane Jarke and Andreas Breiter, "The datafication of education," *Learning, Media and Technology* 44, no. 1 (2019): 1-6.
[2] Marko Teräs, Juha Suoranta, Hanna Teräs, and Mark Curcher, "Post-Covid-19 education and education technology 'solutionism': A seller's market," *Postdigital Science and Education* 2, no. 3 (2020): 863-878.

forms of data collection, turning schools into 'data platforms.'[3] These 'smart schools,' where data collection technology is firmly embedded, have tracking technologies that are "positioned to provide a constant stream of knowledge, in real time, about the activity and performance of every aspect of the institution, from facilities and administration to classroom pedagogy and student progress."[4] This data is used to determine everything from student tracking to institutional ranking.

These methods of datafication go beyond impacting the education process to reshaping central questions about how knowledge is created, how people engage with that knowledge, and even how they construct their world. Datafication has become central to routine knowledge acquisition and belief justification with massive epistemic consequences that are often overlooked.[5] Stefan Baack says that datafication alters "the conditions under which we can make sense of our world and our own actions shifting our capacity to act with agency."[6] In particular, metrics use data to provide an object around which value can be generated or extracted.[7] Datafication has epistemological consequences that reframe essential questions and assumptions about knowledge creation and truth claims.[8]

Through its specific way of framing, presenting, and activating information, datafication creates a system of knowledge that has its own rapidly changing epistemology. Often referred to as algorithmic culture, it is a subversion of human thought processes, behavior, and expression into the logic of large-scale computation – and capable of tremendous epistemic harm.[9] Datafication has the capacity to undermine confidence about self-knowledge: knowledge long believed to be essentially different from all other kinds of knowledge because of its infallibility. Origgi and Ciranna say that through datafication, "we are diminished as knowers, especially in the most intimate part of our epistemic competence."[10]

[3] Viktor Mayer-Schönberger and Kenneth Cukier, *Big data: A revolution that will transform how we live, work, and think* (Boston, MA: Houghton Mifflin Harcourt, 2013).
[4] Jarke and Breiter, "The datafication of education," 5.
[5] Boaz Miller and Isaac Record, "Justified belief in a digital age: On the epistemic implications of secret Internet technologies," *Episteme* 10, no. 2 (2013):117 - 134.
[6] Stefan Baack, "Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism," *Big Data & Society* 2, no. 2 (2015): 1.
[7] David Beer, *Metric power* (London: Palgrave Macmillan, 2016).
[8] Stefania Milan and Lonneke Van der Velden, "The alternative epistemologies of data activism," *Digital Culture & Society* 2, no. 2 (2016): 62.
[9] Ted Striphas, "Algorithmic culture," E*uropean Journal of Cultural Studies* 18, no. 4-5 (2015): 395-412.
[10] Gloria Orrigi and Serena Ciranna, "The case of digital environments," *The Routledge Handbook of Epistemic Injustice* (2017), 303.

In this paper, I will discuss the ways in which datafication technologies such as Big Data and algorithms have the potential to either challenge or exacerbate what Miranda Fricker calls epistemic injustice.[11] I will briefly define epistemic injustice using Fricker's subsets of testimonial and hermeneutical injustice before moving to the potential ways in which datafication contributes to and may prevent epistemic injustice. In the face of this dual threat and opportunity, I will advocate for methods of data activism, including intentional participation and storytelling, before closing by reorienting data usage as a form of rhetoric.

DATAFICATION AND EPISTEMIC IMPLICATIONS

Within this paper, the phrase "datafication" will refer to the modern method and frameworks for large-scale data collection including, but not limited to, Big Data and algorithms. Datafication is shaping and shifting epistemic culture: the ways in which people relate to knowledge and filter the world through experiences.[12] These epistemic cultures become the nexus point where power, privilege, and oppression intersect with society's understanding of what is "truth" and what is recognized as important knowledge. When these understandings are tainted and impacted by systems of oppression, the epistemic cultures risk perpetuating further injustice.

If the body of knowledge in question cannot include the arguments or perspectives of individuals who argue differently or who claim other aspects of the world as salient, then it may be an example of epistemic injustice in action.[13] Some privileged groups may claim unearned authority due to supposed truthfulness of their arguments, even though the understanding of what is true would need to be revised if outside testimony were included. Datafication alters epistemic cultures and creates giant volumes of data that have the capacity to surface and make visible elements of society heretofore not recognized. It has the capacity to both amplify the voices of those previously ignored and shut out of epistemic cultures or to further the epistemic injustice they face. Before examining the implications that datafication has for epistemic injustice, this paper will define the term using Miranda Fricker's definition and subsets.

FRICKER'S CONCEPT OF EPISTEMIC INJUSTICE

Fricker's work *Epistemic Injustice: Power & the Ethics of Knowing* focuses on both the level of representation and participation in knowledge production, and how that perpetuates systems of injustice. Fricker claims that unjust epistemic practices can lead to silencing and causes larger society to lose epistemic resources that can create a more accurate picture of the world than is

---

[11] Miranda Fricker, *Epistemic injustice: Power and the ethics of knowing* (Oxford University Press, 2007).
[12] Milan and Van der Velden, "The alternative epistemologies of data activism," 63.
[13] Jeff Frank, "Mitigating against epistemic injustice in educational research," *Educational Researcher* 42, no. 7 (2013): 363-370.

currently held. Fricker specifies two types of epistemic injustice, one of which is concerned with the testimony of individuals and the other is concerned with how knowledge is generated and validated. The first one, testimonial injustice, is when an individual is treated unjustly in their capacity as a knower, usually having their experiences ignored or dismissed. In the second case, hermeneutical injustice, an individual experiences the injustice of having some significant area of one's social experience obscured from collective understanding owing to hermeneutical marginalization.[14]

The knowledge commonly available or recognized as valid in society may not give a full and accurate picture of the subject at hand or the world at large. Fricker believes the viewpoint of the disempowered is usually invisible to those who are in power, leaving the powerful with an unhelpfully truncated view of the world they live in. However, the powerful are still typically able to determine what perspectives are important and worth listening to. A person or group that has traditionally been disempowered and marginalized will often have different perspectives on that system than those benefiting from it or in positions of power.

Testimonial injustice is often the easier concept to grasp and identify. The #MeToo Movement included the testimony and stories of many women sharing about their experiences of sexual assault. For some, it was the first time these experiences were shared, but far more had shared these experiences with someone else and were discredited or dismissed. The tipping point seemed to occur when these stories came out *en mass.* Overwhelmed with new data points, society responded by examining its own tendency to reject or not give credence to stories of survivors. Allies of the movement went so far as to popularize the slogan "Believe Survivors," a direct admonition of previous instances of testimonial injustice.

Hermeneutical injustice differs in that it has less clear culpability for discrediting another person's capacity to know. Hermeneutical injustice arises when an individual cannot make sense of her experience because of a lack of available collective epistemic resources. Fricker uses the example of sexual harassment in her book. For many women in the workplace, they could not articulate why what they were experiencing was wrong until more and more stories began to arise and the term sexual harassment was coined. With the articulation of the experience and naming it, women were able to come forward and advocate for themselves.

However, hermeneutical injustice is not simply the absence of a name or basic knowledge of existence,[15] but rather it is connected to the prevailing current of social power and privilege and how they influence social understandings. Unequal power and experiences of oppression can leave the powerful and dominant groups to have easy access to the tools and language to make sense of their social experiences. In contrast, Fricker says "the powerless

[14] Fricker, *Epistemic injustice*, 158.

[15] Fricker simply calls this bad luck.

are more likely to find themselves having some experiences through a glass darkly, with at best ill-fitting meanings to draw on in the effort to render them intelligible."[16] This lack of resources to make an experience understandable to them leads to experiences of concrete harm such as a denial of services or disenfranchisement.

DATAFICATION AND TESTIMONIAL INJUSTICE

In both of the examples of epistemic injustice above, data played a significant role in addressing the injustice. Data rendered epistemic sources visible in a way that could not be patently ignored. Datafication is in a precarious place to either contribute to or combat this injustice. One of the primary functions of data is to render things visible. This capacity is especially crucial for people suffering from testimonial injustice who are invisible,[17] having been made objects of knowledge formation rather than taken seriously as knowledge creators in possession of epistemic resources. Data can surface and make visible the suppressed testimony of a previously dismissed group. Big Data becomes an essential tool in this pursuit because its sheer size tends to give it a level of assumed authenticity, allowing it to become a sort of virtual "foot in the door" for the voices of those who have experienced testimonial injustice. Students at my own university have demanded access to previous years' data around campus safety officer stops and detainment. They claimed for years that students and men of color are disproportionately questioned and removed from campus for trespassing, only to have that dismissed as anecdotal evidence. They believe that access to a large enough amount of the raw data will allow them to substantiate their experiences in a way the administration can no longer ignore.

While it may have been collected for entirely different (or even oppositional) reasons, Big Data's variety and multiplicity of discrete fields often amasses the informational haystack from which evidence validating a suppressed testimony may be surfaced. Big Data has a high level of variability, and its meaning is constantly shifting in relation to the context in which it is generated and analyzed.[18] As a result, Big Data initially collected by those with epistemic privilege may potentially be repurposed by those without it to redress their experiences of testimonial injustice. Larger scale data sets, previously the domain of social scientists with specialized knowledge and ability to manipulate them, are now often easy to access.

Datafication can amplify the suppressed testimony, but it can also perpetuate further epistemic injustice. Although it can make things visible, datafication cannot transform bad or incorrect inputs into correct and good

[16] Fricker, *Epistemic injustice*, 148.
[17] Franziska Dübgen, "Epistemic injustice in practice," *Wagadu: a Journal of Transnational Women's and Gender Studies* 15 (2016): 1.
[18] Eileen McNulty, "Understanding Big Data: The seven V's," May 22, 2014, https://dataconomy.com/2014/05/seven-vs-big-data/.

outputs. The constructed nature of data makes it quite possible for injustices to be embedded in the data itself. Regardless of whether this programmed bias was malicious or a result of ignorance, the process of constructing data builds social values and patterns of privilege into the data. Jeffrey Johnson says "where those values and privileges are unjust, the injustice is then a characteristic of the data itself. 'Garbage in, garbage out' is a central concept in data ethics."[19]

This is first seen in instances of exclusivity. Certain testimonies – the perspectives, experiences, values, and needs of individuals or groups – are left out of the body of data by the collection process and operation of the system as a whole. Sometimes this is done through the claim of a lack or relevance or a claim that they are outliers when they differ too much from what is conceived of as the norm. Those representations often mirror existing structures of privilege, with the marginalized less likely to be part of data producing actions.[20] Additionally algorithms cannot participate in what Fricker said was necessary to prevent or repair the harm done by testimonial injustice – reflection upon and revision of prejudiced attitudes over time. They are predictive and forward-looking. The algorithm has to be able to identify that the information it receives is oppressive or prejudicial based upon the larger reality around it, not the volume or amount of the information given. Contradicting data may be simply dismissed if it is in the minority. One single significant piece of dissenting information (e.g., that a study on the scholastic aptitude of students of color was funded by and run by known white supremacists) does not radically reverse the way an algorithm and machine learning self-modifies in response to the given data set.

A messy solution is often proposed in the form of simply saying that more data is needed. This can lead to an indecipherable methodology that does not actually surface or illuminate suppressed or excluded testimonies, but rather leads to confusion and deeper levels of false confidence in the veracity of the knowledge presented. Will Oremus calls this overfitting, saying, "the more sophisticated your model becomes, the more perfectly it seems to match up with all your past observations, and the more faith you place in it, the greater the danger that it will eventually fail you in a dramatic way."[21]

This false certainty created by the exhaustive amount of data inputs can lead to a blithe and uncritical approach that encourages casual attitudes towards claims of causality. If production of knowledge is deferred more and more to algorithms, mathematical data predictions may be trusted more than an individual's own version of events.[22] These algorithms may even deprive people

---

[19] Jeffrey Alan Johnson, "From open data to information justice," *Ethics and Information Technology* 16, no. 4 (2014): 265.
[20] Jathan Sadowski, "When data is capital: Datafication, accumulation, and extraction," *Big Data & Society* 6, no. 1 (2019).
[21] Will Oremus, "How Big Data Went Bust," October 10, 2017, https://slate.com/technology/2017/10/what-happened-to-big-data.html.
[22] Dan McQuillan, "Algorithmic states of exception," *European Journal of Cultural Studies* 18, no. 4-5 (2015): 564-576.

of their credibility about themselves. The amount of data about individuals' identities and actions has created an asymmetry of powers between algorithms and humans. "Algorithms are perceived today as being better knowers of ourselves than we are, thus weakening our entitlement to be credible about ourselves."[23]

Data itself does not have a meaning, and different conclusions occur from the same data set depending on what questions or framing is being utilized and by whom. Using the context of environmental research, Gwen Ottinger highlights that communities who live in an area where data collection occurs (and are even themselves helping generate the data) often experience a "narrative mismatch" with the experts who are collecting or analyzing the data.[24] Ottinger said community members "ask questions that aren't being asked by regulatory scientists; they assert the relevance of factors that aren't represented in standard scientific paradigms; they call for different standards of proof."[25] As a community-based learning practitioner, I have often seen community members try to communicate the complexity of their lived reality to academic researchers (e.g., pointing out nuance of specific historical events that impact a community's current history or indicating a strong generational difference in the opinions of a refugee community that may not be reflected in surveys), only to have it ignored when it did not fit a specific data category or dismissed as anecdotes. Communities often cannot share the information that could help to demonstrate what they are experiencing because it is not coming in the recognized and epistemically "correct" forms.

DATAFICATION AND HERMENEUTIC INJUSTICE

Just as datafication makes the unseen or obscured visible, it also aids in surfacing and, to an extent, creating realities. Datafication often creates the thing it measures the moment it measures it, becoming an essential for understanding the world it has itself helped to construct.[26] Statistical categories created for census measurements have now become a way that communities form and define their collective identity. Algorithms can also mold people's self-understandings by enabling specific representations of the world they find themselves in – and inviting them to internalize them.[27] In light of datafication's ability to unearth evidence and frames for reality, it can become a great aid for those experiencing hermeneutical injustice. They do not have the language to share their unique

[23] Orrigi and Ciranna, "The case of digital environments," 303.
[24] Gwen Ottinger, "Making sense of citizen science: stories as a hermeneutic resource," *Energy research & social science* 31 (2017): 41-49.
[25] Ottinger, "Making sense of citizen science," 41.
[26] Theodore Porter, *Trust in numbers: the pursuit of objectivity in science and public life* (Princeton University Press, 2020): 84.
[27] Alessandra Renzi and Garrett Langlois, "Data activism," in *Compromised Data: New Paradigms in Social Media Theory and Methods,* eds. G. Elmer, G. Langlois, and J. Redden (London: Bloomsbury, 2020), 202-225.

perspective, often because it does not exist. Fricker describes saying that there is a "lacuna where the name of a distinctive social experience should be" in the larger body of hermeneutical resources.[28] However this is not just a simple gap in language, but rather a product of social structures and constructed meaning essentially designed to keep those obscured experiences in the dark.

Data then becomes a way to surface and provide evidence for these obscured experiences so that people may see and better understand themselves. Data can show the flaws in assumptions of what is "normal" by providing evidence that points towards other perspectives that need to be articulated and recognized. Data may give marginalized communities the raw material from which to create their alternative epistemology. The massive variety in Big Data is especially helpful for this work and is remarkably different from an older and more traditional principle of quantification which built universal categories that make sense across national, class, religious and other strata. Big Data rarely needs to be commensurated and can instead just hold the disparate information points, making it more likely that someone whose distinctive experience has been buried or would be otherwise ignored as atypical can find themselves within it. When given access to Big Data that allows counter-expertise to emerge, marginalized individuals and groups can better advocate for themselves and are often more likely to be taken seriously. Milan and Van der Velden cited the example of HIV+ patients whose experiences were being ignored by their doctors and did not have the necessary epistemic tools to better articulate what they needed. After inundating themselves with articles from scientific journals and studies, they formed opinions on what was necessary for their treatment, advocated for themselves, and eventually convinced their doctors as well.

DATA ACTIVISM AND EPISTEMIC RESPONSIBILITY

Datafication is making certain unknowable things knowable and obfuscates others, which in turn changes existing standards of epistemic responsibility.[29] Epistemic resources are being located in technological infrastructures, many of which individuals do not have the capacity to manipulate. While this puts constraints on what is practicable for us to know and do, it does not take away our epistemic responsibility. Miller and Record say that individuals must take every action they can (even if it is just awareness) to ensure that when people come to beliefs, they do so by interrogating them and asking how they can be justified. Doing nothing in the face of epistemic concealment and obfuscation is unacceptable, and instead epistemic resistance must counteract intellectual laziness.

Datafication has both the power to reinforce epistemic injustice, but also provide tools to diffuse it and instead engage people for action. Milan and

[28] Fricker, *Epistemic injustice,* 150.
[29] Boaz Miller and Isaac Record, "Responsible epistemic technologies: A social-epistemological analysis of autocompleted web search," *New Media & Society* 19, no. 12 (2017): 1945-1963.

Van der Velden say that "data activism" is a form of action and activism in today's information-based society that specifically tangles with the new forms of information and knowledge production. It attempts to challenge the dominant approach and understanding of datafication, and it can be seen "as an exercise in creating alternative ways of seeing the world."[30]

Resistance to this dispossession and taking a more active role in the face of the potential fragmentation and commodification of the self is another important part of data activism. One of the biggest hallmarks – and threats – of datafication is the high level of granularity of the information it can surface and produce. The effect this has on individuals is particularly sinister as a life that is continuously tracked and processed for Big Data is one that is also dispossessed from its own distinctness.[31] Datafication's dismantling of the 'human subject' can transform it from a singular entity deserving ethical treatment and turn it into discrete categories as means of targeting.[32] Data activism calls for an individual to demand that datafication continues to understand them as more than the sum of their data points.

boyd & Crawford highlight three types of people who interact with Big Data.[33] There are those who create data, whether intentionally or through their digital footprint; those who collect it; and the individuals with the skills to analyze it. Especially in the cases of learning management systems and wider education data, Anna Kruse and Rob Pongsajapan propose a "student-centric" approach that would merge all three roles.[34] Students would be recognized as a co-interpreter of their data and potentially identify important variables within it and even track additional data that they believe is important. Not only would students have more agency within the system, but the process may also lead to more self-awareness among students.

Taking a more critically participatory stance around data can reduce the opportunities for increased epistemic injustice. Levels of deeper participation could lead to a higher value placed upon the testimony of marginalized individuals and allow them to analyze their own situation.[35] This would lead to more accurate information and inputs for algorithms. However, this deeper participation of the subject could also lead to a push for more qualitative data to

---

[30] Milan and Van der Velden, "The alternative epistemologies of data activism," 62.

[31] Nick Couldry and Ulises A. Mejias, "Data colonialism: Rethinking big data's relation to the contemporary subject," *Television & New Media* 20, no. 4 (2019): 336-349.

[32] Dan McQuillan, "Data science as machanic neoplatonism," *Philosophy & Technology* 31, no. 2 (2018): 253-272.

[33] danah boyd and Kate Crawford, "Critical Questions for Big Data in Information," *Communication and Society* 15, no. 5 (2012): 662-679.

[34] Anna Kruse and Rob Pongsajapan, "Student-centered learning analytics," *CNDLS Thought Papers* 1, no. 9 (2012).

[35] Gottfried Schweiger, "Epistemic Injustice and Powerlessness in the Context of Global Justice. An Argument for 'Thick' and 'Small' Knowledge," *Wagadu: A Journal of Transnational Women's and Gender Studies* 15 (2016): 104-115.

be incorporated into data sets, leading to a richer (though more complex) product. More incorporation of rich qualitative data would encourage more attitudes of information pluralism. This perspective adopts an enthusiasm for messy data and sees conflicting information not as obviously incorrect, but rather an indication for further points of learning and epistemic opportunity.

Personal narratives from those now more involved participants would also present an important opportunity to combat epistemic injustice and present a more precise view of the "data" that is important to that individual or community. Ottinger advocates for storytelling, saying that it "gives community groups a way to refuse dominant narratives about them, and advance their own alternative understandings of their communities, how they've been treated, what is owed to them by regulators and other dominant groups, and what their future should look like."[36] In amplifying their own testimony *in their own voice*, communities or individuals keep datafication from fragmenting and reassembling their experiences to fit a constructed model based on incorrect understandings of their context. It can address the narrative mismatch often experienced by communities who are the subject of research.

DATAFICATION AS RHETORIC AND ARGUMENTATIVE INJUSTICE

The need for storytelling when Big Data is available might seem unnecessary, especially given the commonly accepted wisdom that 'the numbers speak for themselves.' The numbers are indeed speaking, but not in the way commonly thought. McQuillan argues that just as storytelling is a form of rhetoric – persuasive argumentation – arguably so is datafication. "Datafication itself is a rhetorical move, because it is saying that the important aspects of reality are ones that can be expressed as data."[37] Rhetoric is a mode of conveying information. Although it seeks to convince the listener that what it is conveying is both true and worthy, there is no guarantee that what is being conveyed is either. Reframing datafication as one form of rhetoric amongst many pulls it away from the false trapping of scientism that guarantees its truth claims.

Although data cannot accurately claim to be *the* singular truth (or potentially even be sure of its own truth claims), it is still a form of knowledge creation for many and is becoming more and more decisive in its creation of the world and culture. Big Data and algorithms are making arguments to the public for how the world is to be understood. As those affected by the data may experience epistemic injustice, datafication is a prime example of argumentative injustice. Bondy's concept of argumentative justice is similar to testimonial injustice, but focused on argumentation, rather than testimony. It typically occurs when an arguer is given too much credibility or too little.[38] When datafication and its outputs are understood as rhetoric – presenting their argument for a view of the truth and reality – it is easy to see how often data benefits from

[36] Ottinger, "Making sense of citizen science," 40-41.
[37] McQuillan, "Data science as machinic neoplatonism," 266.
[38] Patrick Bondy, "Argumentative injustice," *Informal Logic* 30, no. 3 (2010).

argumentative injustice. Wrapped in the false objectivity granted by scientism, data is often automatically given the benefit of the doubt as presenting unbiased, proven information. On the contrary, all data is situated and contains biases.

With this in mind, it would be wise to adopt Bondy's concept of metadistrust when regarding arguments made by data. Bondy says that becoming aware of our own biases is a helpful first step, but that knowledge of them is not enough to counteract the possibility for committing argumentative injustice. Practicing metadistrust, or self-doubt regarding initial credibility judgements, slows the initial temptation to assume the infallibility of a claim because of its claimant. Metadistrust also helps expose Big Data's tendency toward apophenia, or seeing patterns where none actually exist, simply because enormous quantities of data can offer connections that radiate in all directions. Rather than assume that because there is so much data a finding must be true, an attitude of metadistrust would ask the "listener" of a datafied argument to examine whether or not the data (despite its bulk) is actually salient. Because epistemic justice does not simply mean including the epistemically marginalized, but rather opening spaces for additional discourse and different ways of knowledge formation, the practice of metadistrust becomes essential. Metadistrust becomes another essential practice of epistemic responsibility, guarding against the dangers of becoming complicit in structural injustice through an acceptance of 'conventional' truth claims that marginalize challenges from the oppressed.

COVID-19 brought a majority of our classrooms into the digital realm and provided heretofore unimagined amounts of data into easier access of tech companies and instructors alike. But even with all of this new information, the data does not, and should not, speak for itself. As other methods for determining causality and patterns are lost in the sheer volume of raw information, those working with and studying data should prevent numbers from speaking for themselves.

Because when those numbers come from a data setting with a power imbalance or misaligned collection, then they run the risk of being not only discriminatory, not only empirically wrong, but actually dangerous in their reinforcement of an unjust status quo.[39]

Lest it continue to perpetuate epistemic injustice, crushing valid and important testimony under its weight or fragmenting a student down to a simple roster of attributes before attempting to fit them into a predictive mold that does not consider their singularity of context, datafication must be considered one form of rhetoric amongst many ways to claim truth and to learn.

[39] D'Ignazio and Klein, *Data feminism* (MIT press, 2020) 103.