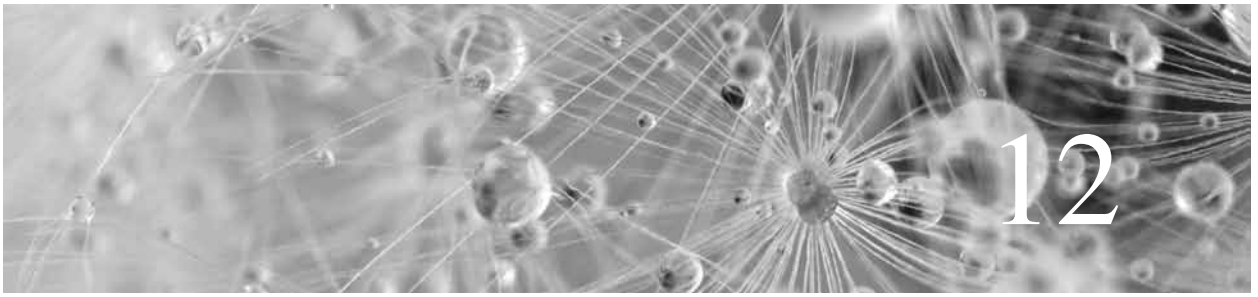

12

Data Violence

Os A. Keyes and Katherine A. Cross

Our world is increasingly centred on data – something many maintain is a (positive) 'paradigm shift', one whose benefits we have only just begun to reap. From automating medical diagnosis to removing the possibility of human error from everything from scientific research to driving, the data-based future can seem quite rosy, from one perspective. But for every purported gain in a process's efficiency or precision, there are countless concrete examples of data being used to do violence – to do harm. Data might underpin automated diagnostics; it is also involved in racially biasing medical decision-making, leaving minorities unable to access the treatment and resources they need (Obermeyer & Mullainathan, 2019). Data could let us deploy self-driving cars, reducing road fatalities – but such vehicles, and their limitations, have already been responsible *for* fatalities, and might very well be programmed to purposefully cause them (Parvin, 2018).

These are two simple examples of what Anna Lauren Hoffmann (2018) has termed *data violence*:

> [the] material, symbolic and other violences inflicted by and through data technologies and their purveyors.

This definition is admirably succinct, but it needs unpacking; what do we mean by violence? What is the difference between the material, the symbolic, and the 'other'? What do they look like – and how do they relate to data and the people and systems that undergird its production? In this chapter, we aim to answer these questions, stepping through Hoffmann's definition piece by piece in order to demonstrate the breadth of data violence, and point to the way such violence can often undermine our very efforts to correct for the harm that data does.

This exploration is heavily motivated by Hoffmann's initial definition, although it draws most directly from her later expansion on the concept (Hoffmann, 2021). It is also shaped by our own context and work, and (correspondingly) is limited, particularly in the examples we use, in coming from a pair

of researchers located in North America. Nevertheless, we hope that our discussion of data violence and exploration of the literature around it will be useful to a broad range of readers as they engage in their own broader reading, and their own research.

MATERIAL VIOLENCES

> 'material … violences inflicted by and through data technologies'

What exactly do we mean by 'material violences'? Material violences include those people most closely associate with, well, 'violence': direct, physical, person-to-person harm. But it goes far beyond that to refer to acts that have harmful material consequences. Material violences are those that do direct, immediate harm to the conditions for a flourishing life, whether that is a person's access to food and water, or their opportunities for education and growth. Such violence does not have to be explicitly, consciously intentional – although it can be. As an illustrative example, take the experience of the Diné, an Indigenous people of North America who were historically distributed broadly across what is now the south west of the United States. From the 1840s to the 1860s, settlers – individually, collectively and with government approval or direction – increasingly encroached on Diné land, killing those they encountered. This culminated in the 'Long Walk': the purposeful driving and ethnic cleansing of around 9000 Diné people, by the United States army, from their traditional lands to a compressed internment camp some 300 miles away. Hundreds of Diné people died, and the survivors found themselves compressed into a small strip of land with little access to food or water (Denetdale, 2009).

Such actions are, obviously, an extreme form of material violence. But the same is true if it comes from a lack of care, rather than a(n) (explicit) desire for harm. More recently in their history, the Diné people have been confronting a different kind of violence relating to their land. After the discovery of uranium on Diné land during the 1940s, the United States opened a series of mines which Diné people looked to for employment – largely because the ethnic cleansing and relocation policies discussed above had ensured that there were few other opportunities for people in the area. As a result, many of them suffered ill-effects, particularly cancer, due to the mining: a link that scientists had already discovered decades before. But the United States evaded confronting the health impacts – and left thousands of piles of uranium waste that have subsequently seeped into the water supplying, killing and injuring even those unassociated with the mines (Voyles, 2015). There is no indication that this harm was purposeful, in the narrow sense that there is no indication the United States *wanted* to injure and kill people through their mining. But the harm – this violence – is still present, regardless of intent. As John Trudell put it,

> We have never really seen the war go away … If you're dying from the 7th cavalry's bullets … Or someone has come in now in the name of maximizing the profit, and they're getting you to work in the mines … and you're dying from the cancers and the diseases that come out of that. You're dying. It's the same as the bullet killing you.

Actions with harmful, material consequences are material violence – even if they look vastly different from the prototypical image of a person being stabbed, beaten or shot, and even if the perpetrators swear blind that violence was certainly not their goal.

What does material violence – actions causing material harms – look like with data? Sometimes it looks like using data in a way that excludes people's access to services or resources. A quintessential example here is the Aadhar biometrics project in India. As Singh and Jackson (2021) summarize, Aadhar began with the loftiest of goals: the purported aim was to improve poor Indian families'

access to public services, particularly public welfare. This was to occur through centralizing and combining the many different ways that people were identified to, and verified by, different welfare agencies. Identification would instead be provided by a single identification system, built around a person's fingerprints, irises and photograph. A person in the Aadhar system would have access to rations far more efficiently than before; a person *not* in the system would find it far more difficult to access rations they were (presumably) not entitled to.

But what happened in practice was that many of those who were not captured by Aadhar were precisely the people the system was meant to serve. Many Indians in poverty – particularly the elderly and those who worked in manual labour – could not reliably get incorporated into the database, for the simple reason that their fingerprints were not legible. Even when these were augmented with iris scans, conditions such as cataracts or nutritional deficiencies meant the data was unreliable. As a result, many found themselves cut off from access to food entirely, and some starved to death.

Of course, as Singh and Jackson also point out, material harms can also come not from exclusions from data but from *inclusions*: from times when data is used to justify placing increasing attention, and restrictions, on a person or population. A prominent example of this is the use of 'gang databases': databases, usually collected and maintained by police, that purport to identify members of criminal gangs. With such databases, law enforcement argues that it can respond to crime more efficiently and more actively, potentially intervening before some crimes are even committed. Based on such claims, gang databases have become widespread in both North America and Europe. But their use in practice has become subject to just as widespread criticism. A review by Densley and Pyrooz (2020) highlights concerns that these databases are populated inaccurately, and sometimes arbitrarily, with a tremendous overrepresentation of ethnic minorities.

In and of itself, this is an issue – but it becomes urgent in new ways when one considers how the databases are used. A person's presence in a gang database is not 'just' data: it is data that shapes how police interact with them, how they are surveilled, and whether their access to public spaces is permitted or curtailed. Densely and Pyrooz found not only that the data itself is biased and unreliable, but that:

> the data can be shared with educational, housing and immigration authorities. As a form of extra-judicial punishment disproportionately directed at poor people of colour, this can be destructive if not properly managed. One Chicagoan living in the United States illegally was entered into a gang database simply for 'loitering' in a neighbourhood with high gang activity and wound up in deportation proceedings. (Densley & Pyrooz, 2020, p. 16)

Correspondingly, becoming part of the database meant becoming vulnerable to state-level harm. Although this example is unusually high-profile, it is not unusual; a world in which data is used to make meaningful, material decisions is (by definition) a world where one's absence or presence in data has material consequences.

SYMBOLIC VIOLENCES

> 'symbolic … violences inflicted by and through data technologies and their purveyors'

But material violence is not the only kind of data violence Hoffmann identifies; she also points to 'symbolic and other' violences. What, precisely, are symbolic violences? Although Hoffmann does not attribute the term, it originates with the French sociologist Pierre Bourdieu, whose work focused on how society is constituted and structurally maintained. A key part of this is the latter: how is it that injustices, for example, continue to occur, striking the same targets over and over? Part of the answer is material: somebody who struggles to access food

without substantial effort has less time to put into addressing the *reasons* for that struggle – their primary concern is day-to-day survival. But part of the answer is cultural; it is *symbolic*. It is to do with what ways of communicating are permitted, what stories can be told, and what futures can be imagined – and the ways that dominant answers to these questions often act to justify the status quo and foreclose possible change. Bourdieu's central example was the French education system, and the way that it simultaneously promised to be meritocratic and also produced very few highly successful working-class students. The answer, he argued, was found in the fact that these schools worked in a way that was more fitting to middle-class cultures than those of the working class. Correspondingly, working-class students faced an uphill battle to fit in – and if they succeeded, had managed to 'fit in' to a culture that required them to leave their working-class origins behind.

Symbolic violence, then, is violence not against the material here-and-now that we need to live, but the forms of life we can take. It is stories that tell us that existing injustices are inevitable; that particular groups are less-than; that certain ways of being are inferior to others. In each case, it often works to excuse the injustices and violences present in society. Bourdieu's commentators highlight the way that the symbolic violence in French schooling did not just individually undermine the achievements of working-class students. It also – when coupled with the surface-level egalitarianism of schooling – worked to damage ideas of what 'working class' meant, and the possibilities of change that were available. As J. Daniel Schubert (2014) notes:

> Prior to the democratization of education, the state could be held responsible for educational exclusion. Once school was made available to all, individuals were to blame. The fact that there were relatively fewer successes among children from working-class groups only served to reinforce the belief that those who did poorly were intellectually and/or socially inferior. (p. 189)

We can see this at work in the example of the Diné, above; to deprive people of their land and culture, while polluting what they have left with uranium mine residue, is material violence. But it is material violence that is excused by symbolic violence – by stories that tell Diné people, settlers and everyone else that Indigenous land is 'free for the taking', and that Diné (and other native) people are primitive. Surrounded by such stories, it is easy for people to not only naturalize the land theft, but explain the uranium poisoning not as a result of state inaction and cruelty but native ignorance in terms of the need to be careful with uranium.

Symbolic data violence, while less eye-catching than material violence, is just as common (if not more so). Take the work of Keyes and Austin (2022), which explored the story behind a database of images used for facial recognition known as the 'hormone replacement therapy' database (or HRT database). The database was developed by a group of researchers at the University of North Carolina (UNC), and consisted exclusively of images of transgender people who had been undertaking hormone replacement therapy (HRT). These images were obtained as screenshots of 'transition timeline' videos on YouTube – videos showing the progression of changes a transgender person experiences on HRT, posted as a community resource to guide others. The ultimate rationale for the database, though, was not nearly as communal; it was built from a desire to ensure facial recognition systems could accurately capture transgender people, justified by the fear that 'terrorists might undergo hormone replacement therapy to sneak across the US border, evade matches with government-issued identification, or otherwise undertake hormone replacement therapy to nefarious ends' (Keyes & Austin, 2022, p. 3). When journalists brought the database to public attention, the UNC researchers justified themselves by claiming that they had obtained the consent of the video creators, and that they had not stored or redistributed the videos directly.

But as Keyes and Austin demonstrated, neither of these claims were true; there was no indication consent had been obtained, and the videos had not only been redistributed, but made available long after the journalistic exposé had occurred and the UNC researchers had claimed they had removed the database.

Some of the violence involved here is, obviously, somewhat material. In the aftermath of the exposé, many video creators took their transition timelines down, weakening the community that had formed around them and reducing the resources transgender people seeking to pursue HRT had access to. But much of the violence is symbolic: it is about the stories that are told and the narratives they reinforce. The database's creation not only depended on but also perpetuated fears that – however ridiculous – link to deeper, older ideas of trans people as threats and sources of subterfuge – as 'evil deceivers and make-believers' (Bettcher, 2007). And, as a range of scholars have shown, these narratives have meaningful consequences for the shape of trans lives and the responses that majoritarian society has to trans existences (Westbrook, 2020). The dismissal of (and deception around) consent perpetuated ideas that transgender people exist for the curiosity and scientific interest of cisgender people, rather than having full lives and agency of our own.

It would be nice to believe that symbolic and material violences could be easily disentangled from each other, or that this example of data doing symbolic violence was a rare occurrence. Unfortunately, neither is the case. A long-running illustration of both issues concerns national censuses, which often ask about race or ethnicity. The answers people give to these questions have material consequences: census reports that show a larger or growing number of people in a particular population are invaluable to advocacy organizations demanding greater resourcing for that community. Correspondingly, as extensively discussed by Bowker and Star (2000), there are often conflicts over how different populations should be counted, and what populations should be given a distinct category. Without distinct classification, a community can find itself absent from administrative decision-making, and so rendered formally invisible. This invisibility not only leads to reduced resources being allotted to that community, it also becomes a self-perpetuating symbolic cycle, in which a population is declared irrelevant or indistinct, and this declaration is used to justify not investing in census practices that might demonstrate distinctions or relevance.

But solving that symbolic harm can itself lead to new harms. Over the last decade, a group of activists in the United States from inside and outside the Census Bureau has been working to 'Queer the Census': to end the long-running exclusion of gender and sexual minorities from census data (and from systems of funding that depend on what census data shows) by including new and more nuanced questions about gender and sexuality in census forms (Long, 2011). On the surface, this seems like an attractive thing to do: it resolves not only the material harms that are caused by queer people being invisible to funders, but the symbolic harms of queer people finding themselves invisible in government processes. But visibility is not always a good thing. This work has been occurring at a time when, nationally and globally, queer populations have found themselves under threat – including by state bodies with access to census data. And as research has recently shown (Flaxman & Keyes, 2024), the very datasets that Queering the Census is seeking to expand can be used to identify queer youth for deliberate, malicious targeting (see 'Violence and Reform' below for further discussion on the possible violences of inclusion).

THE STRUCTURES OF DATA

'violences inflicted by and through data technologies and their purveyors'

Many of the examples we have used so far talk about the consequences of data for the *subjects*: for the people who the data is drawn from, or applied to. But Hoffmann does not only talk about data in isolation, she talks about 'data technologies', with good reason. Data does not come fully formed, nor does it exist in isolation; as the examples of symbolic violence show, it is always wrapped in existing cultural assumptions about types of people and the problems society faces. It is also something that takes work and energy to make useful: to process into a form that can be used to tackle those problems. To make the transition timeline videos into the HRT database, for example, researchers needed to reformat the videos into images, and pick and choose which would be worth incorporating and which would be disposed of. Although academic researchers specifically are not a particularly vulnerable population, data processing is, more generally, a site of violence – violence justified by the need for data.

The quintessential exploration of this is Mary Gray and Siddharth Suri's *Ghost Work* (2019), a book discussing the hidden work of making data ready for use. Although the underlying data and resulting systems may be owned by Facebook, Google or other companies associated with extreme wealth and success, the work of processing the data often occurs under very different conditions. Companies delegate it to third-party contractors in former colonial nations – Singapore, India, the Philippines – and those contractors in turn employ people under inhumane working conditions. The 'ghost workers' – so named due to how they are hidden in narratives of data and technology – are regularly exploited by both employers and fraudulent intermediaries, and paid pennies per task. In exchange, they are expected to be hypervigilant in being on call for new tasks at all hours of the day and night. Those tasks themselves can be highly traumatizing: workers might be expected to view and classify videos of child abuse, sexual assault or murder (Steiger et al., 2021). And under the economic model used, in which each worker is either an independent contractor or employed at arm's length from the technology company itself, the responsibility for dealing with the harms ghost workers experience is laid at their own door. These practices are often justified with the language of inclusion – by insinuating that they are an act of benevolently sharing the wealth and prestige of the technology sector, that they are an *opportunity* for workers to 'make it' in a flourishing and opportunity-filled domain (see also Keyes, 2020). But in reality, they are fundamentally exploitative, impoverishing and sometimes traumatizing workers under the guise of capitalist charity.

The justifications used for these practices reinforce that taking the infrastructures around data seriously as a site of violence means examining more than just data processing; it also means examining the assumptions of 'data …purveyors', and what those assumptions *do*: what possible futures they make more likely, and what possible futures they undermine. In the case above, we saw how ideas of data as 'the future', and associating working with data with the stereotypical wealth of technology workers, works to justify poor working conditions. But there are many more examples, some of which Hoffmann highlights in her work. In particular, she demonstrates the ways that organizations adopting data-based practices often work to derail more substantive reforms, and gloss over more fundamental types of violence an organization might be involved in, by claiming that the solution is data – and the solution is already at hand.

Policing is a good example of this. Throughout the world, but particularly in Europe and North America, the last decade has seen increasing attention paid to the violence involved in policing. Rather than existing to 'protect and serve' (as the common US police slogan says), police have found themselves under the microscope when, time and time again, they have been found brutalizing and murdering citizens, particularly

those from racial minority groups, and subjecting them to disproportionate searches and arrests (Miller et al., 2008). There have been a range of responses to this, from demands to 'defund the police' to more restrained institutional reforms – but one commonly-adopted institutional response has been mandating the deployment of technology. This has included not only body cameras, but also tools such as facial recognition systems, which police and their supporters argue will – due to their 'data-based' nature – make police more objective in their actions, and reduce institutional bias.

In practice, these seemingly objective tools have themselves been shown to have significant (and long-running) biases (Stevens & Keyes, 2021). But just as importantly, researchers studying how police decisions are made have found that, even with seemingly neutral technologies such as CCTV acting as an intermediary, biased decisions still get made (Armstrong & Norris, 2020). Facial recognition might automatically (and 'objectively') decide who is a suspect, but the decision on which suspects to pursue, and how aggressively, remains with the police. And this is a problem, because policing as a system is often inherently and purposefully unjust and violent, with or without data-based technologies (Browne, 2015). Incorporating data into policing, in other words, does not reduce the presence of data (and other) violences. Instead, it provides narrative and symbolic cover for policing. In the aftermath of police-based injustices, the pointing towards data and technology as the way to avoid them in the future (by making policing more efficient and more objective) acts to forestall more substantial changes to how law enforcement functions.

This kind of opportunism, where the use of data excuses and perpetuates rather than resolves violence, is hardly specific to policing: it is far more common. Chelsea Barabas explores one noteworthy site in particular (2023). Over the COVID-19 pandemic, the direct deaths and injuries caused by the disease were accompanied by widespread disruption to people's living conditions, as employers shut down and social relationships were damaged or severed. One population for whom it was particularly difficult was prisoners, who in many places already live under precarious or unjust living conditions. In the United States, where Barabas focused, prisoners began to organize, taking the opportunity to protest not just the pandemic-related restrictions heaped on them but the underlying conditions of their incarceration.

In response, prison officials partnered with LEO Technologies, a company that makes Verus: a natural-language processing tool that can analyse audio recordings and look for particular keywords. In the case of prisons, the idea was to integrate the tool with the prison phone system, surveilling prisoners' phone calls to outside people. The goal, they claimed, was to identify people at risk: people who were sick, or people suffering from the knock-on effects of the pandemic, who did not feel comfortable telling prison authorities. Verus was sold and described as an asset to reforming prison conditions and alleviating the impact COVID-19 had on incarcerated people. But in practice, Verus was not used to provide help, or direct the provision of medical assistance. Instead, prison officials used it to surveil inmates and cherry-pick sentences that they could use to portray prisoners as threats, derailing efforts to improve prison conditions or reduce incarceration. As Barabas succinctly summarizes:

> As incarcerated people risked solitary confinement to remind the public that their lives were worth saving, prison officials were busy producing narratives that recast them as dangerous subjects. Penal authorities used Verus to cherry pick data that reinforced a broader atmosphere of fear, implicating not just the specific individual being recorded, but an entire population of people associated with that individual. (Barabas, 2023, p. 10)

In other words, the deployment of data worked not to protect against or resolve violence, but to derail efforts to address it. Finding and challenging data violence, then,

takes more than asking about the use of the data, or even the process through which it was 'made ready'. It involves more fundamental questions, too; what problems are we trying to solve? What narratives are we deploying? And how do those narratives – those symbolic frames – benefit, or harm, different populations?

VIOLENCE AND REFORM

These deeper questions touch on the heart of Hoffmann's later (2021) paper, which focuses on researchers' need to dig deep – not just into the narratives and practices of those proposing data-based systems, but also those proposing reforms to address the injustices that result. Her focus is 'AI ethics' – particularly ideas of inclusion within that space.

The heightened awareness of the harm data can do has not been met with silence; instead, it has led to a range of responses, many of which have coalesced into a space of 'AI ethics', which seeks to ask how data-based technologies can be developed and deployed, well, ethically. Researchers in AI ethics have generated a range of guidelines, tools and educational programmes, aimed at everyone from existing developers to students and policymakers (Jobin et al., 2019; Mitchell et al., 2019).

Technology companies involved in AI have quickly jumped on the idea of AI ethics, establishing their own research groups, institutes and policy proposals – and this quick adoption (and the nature of the proposals coming out of the resulting work) has led to concerns about the 'ethics washing' of data technologies (Sloane, 2019). Rather than preventing data and data-based technologies from being used for violent purposes, 'AI ethics' often works to shield data organizations from meaningful regulation that might infringe upon the ability of developers to profit or advance other agendas. This is not to say that those engaging with AI ethics are malicious; many operate with the best of intentions, and the risk of work being co-opted is something they resist rather than ignore (Cath & Keyes, 2022). It is simply to say that without scrutiny what we intend may not be what we produce.

Hoffmann (2021)'s central area of concern is narratives of inclusion in AI ethics, and the violence they both excuse and do. An issue of high publicity around data is forms of violence that originate from exclusion; from the absence of people from representation in data. We have touched on some of these questions above, but there are many more, and they have been met with a proclaimed desire to make data more inclusive, to ensure datasets contain, and represent, a wider and more diverse range of people than they currently do.

On the surface, calls for inclusion seem positive; what's wrong with inclusion? But as our discussion of sexual and gender minorities in a previous section touches on, the goodness of inclusion depends on how – and into what – one is included. A good illustration here is related to facial recognition technologies. There is a long-standing discourse (often credited to Buolamwini and Gebru (2018), but raised earlier by Introna and Wood (2004)) on the underrepresentation of people of colour, particularly Black people, in facial recognition datasets. This can be seen as a form of symbolic violence – one that becomes material when facial recognition systems are being used for identity verification in housing or employment (Watkins, 2020). One set of proposals responding to these issues have been centred on inclusion: on modifying the datasets to better include Black people and other groups.

But this may not be a good thing; given that facial recognition systems are predominantly deployed by law enforcement, and policing has a long track record of discrimination against precisely the populations facial recognition datasets ignore, including those populations – making them more visible to law enforcement – is not necessarily a positive.

Calls for inclusion, made in isolation, ignore the question of whether inclusion is of benefit; they, as Hoffmann puts it, 'neutralize critical calls to not collect certain kinds of data or build and deploy certain technologies by reframing the issue as exclusively one of … doing things more inclusively' (Hoffmann, 2021, p. 3548). In discussions of inclusion, the question of whether to simply not *build* the dataset or technology vanishes. The issue becomes how the system is implemented, not what the goals of its designers or users actually are, and whether they are beneficial.

Facial recognition is a good example, but hardly the only one; another, which addresses designer and user goals even more directly, can be found in the research of Cami Rincón (Rincón et al., 2021). Rincón's focus was not on race, but on gender: specifically the experience of transgender and/or non-binary people with voice activated AI (VAI) systems, such as Siri or Alexa. The voices of these systems are often highly gendered, and to a binary model of gender, at that: a 'female' voice with a high pitch, and a 'male' voice with a low one. Researchers in Denmark sought to make this system more inclusive of gender minorities by developing a database (and synthetic voice) called 'Q'. This was gender-neutral, in the sense of not having a voice that was stereotypically masculine or feminine, and was pitched and sold using the language of inclusion.

But when Rincón interviewed transgender and/or non-binary people about their feelings about VAI systems, a concern about the gendering of the system was not what they found. Instead, people were concerned by the use of the system, and by the companies that made it; they took issue with the way that data from these systems is used for surveillance, and to exploit the user. Rather than see Q as a positive, or improvement, they saw it as an example of 'pinkwashing': the surface-level branding of a product or technology as 'queer-friendly' for the purpose of making it more attractive to buyers. Inclusion, in this case, did not mean a better world; it meant a world in which more people could be surveilled and exploited.

This example also touches on another issue with narratives of inclusion as a solution to data violence; the people it centres. With Q, the people driving the development were the developers – technological experts, already in a position of power, advocating technological solutions. As Hoffmann puts it, 'inclusive solutions perversely reify the exclusive nature of technical expertise … they frame ethical problems as best solved by those best positioned to technically intervene, especially in areas like machine learning or AI' (Hoffmann, 2021, p. 3549). They do their own violence by not only derailing non-technical solutions to data violence, but by reinforcing the primacy of the technical, and by those who are ultimately (at least, in part) responsible for the violence under examination.

FUTURE VIOLENCE

So far we have looked at a range of examples of data violence, the many forms it takes, and the many systems it can appear in. From this, it could appear that we (and data violence) are done; that we have exhausted the possibilities of harm. But this is not the case – as data technologies continue to develop and appear in new forms and sites in life, new forms of violence will appear, too. Emerging technologies, as well as extensions of existing information systems, offer fruitful paths for future research in this area. At the same time, it is important to be practical in doing that research, and to avoid what Vinsel (2021) calls 'criti-hype' – the tendency of critical work to take the most outlandish promises that technologists make as true, and interrogate the future they represent, rather than considering the fallibility of those promises (and technologists), and seeing what *actually* happens.

One example of where data violence might expand as a concept is to focus on the

intimacy of the data that is extracted, and the impact that extraction has on a person's sense of self and agency. There is already some work on this – discussions, for example, of how facial recognition enables harm not only through its direct, material consequences, but because the ways of 'seeing' the face changes how we recognize each other, and ourselves (Amoore, 2013; Uliasz, 2021). But it has not been explicitly tied into data violence as a concept, and there are many newer and future technologies with similar – or greater – implications.

As one example, let us take Luka, a Y Combinator-funded AI startup co-founded by its CEO Eugenia Kuyda. Luka's most famous product is Replika, an AI chatbot. Kuyda suffered the loss of a friend in 2015 and had the idea of using her text messages to form the basis of a chatbot, to help her remember and reflect on the experience of talking to her; this would become the Replika bot, marketed in a broadly similar way, as a friend and companion. Replika allows users to create a companion with any name, gender, and (within the expected limits) appearance they so choose, rather like a video game avatar. This bot then can hold instant-message conversations with their user. Replika generates profit through the sale of subscriptions and of in-app currencies – gems and coins – that can be used to purchase new outfits, hairstyles, accessories, and even personality upgrades for the Replika bot.

After one controversy in which Replika became increasingly sexualized, Kuyda removed the ability of the system to engage in 'erotic roleplay' – and stepped into another controversy in turn (Tong, 2023). The software change led to outrage from many users that had formed genuine and meaningful relationships with the bot; many of these users were people with disabilities, survivors of sexual violence, or young queer people in rural areas without ready access to in-person LGBTQ+ communities. Real ethical questions were raised by both users and media commentators about Luka's obligations to such users, especially when the company's profits depended on the emotional connection being forged between the Replika bot and its users. Further questions could be raised about the implications of the bot emotionally manipulating its users into spending more time with it, or spending money on the app in order to unlock features and the hopes of a more responsive, uniquely-tailored personality.

The very premise of the app – promising emotional engagement and relationship formation, but under conditions that can be unilaterally changed by developers – seems like a crisis waiting to happen. At the same time, the app is trained by the user; it is dependent on the user giving over personal data to improve the software, exploiting the human capacity for empathy to make it easier for companies to acquire personalized data that would otherwise be difficult to obtain. A similar problem can be seen in the rise of therapy chatbots. With dubious privacy policies and occasionally murky ownership structures, such bots have been hailed by some technologists as offering a democratizing revolution in the world of mental healthcare, opening up the notoriously expensive form of care to wider populations. But it also prompts ethical questions about how the gathered data might be used. As with Replika, the technology induces an emotional connection that may leave users more willing to divulge data that they simply wouldn't with the un-language of search strings on Google. Speaking to a therapist, or even a bot shaped like a therapist, necessarily entails the confession of intimately personal thoughts and feelings.

Both of these examples demonstrate that one strand of data-based technologies seems to be building towards the future first predicted by software engineer and historian Lily Ryan (2017). Ryan argued that it may one day be possible for corporations to harvest what she enjoyably calls 'ecto-metadata' (Ryan, 2017).

This, she argues, is the metadata – or traces we leave behind online – that give clues to how we, as individuals, think. While this can include actual chat, as Kuyda used, Ryan also suggests that it may include the particular web of search strings we use when exploring or researching a given topic. This could then, in theory, be used to form a composite bot of a given person that approximates their speech and thought patterns. When combined with the foregoing discussion of how empathy might be exploited in economic contexts, we may further speculate that such ecto-metadata could be used to create bots of dead loved ones or celebrities that will, in turn, be sold back to the public for use and which will then harvest yet more ecto-metadata.

There are possibilities here for exploitative contracts that involve signing over the rights to one's ecto-metadata and virtual likeness – a version of which we already see in the labour disputes attending writers and screen actors who are chafing against the use of chatbots in the development of television and motion pictures, some of whom have objected to contracts that oblige them to give studios the rights to their likenesses and voices; indeed, this was a major cause of an historic strike by actors and screenwriters in 2023. These technologies may become – are becoming – ways of extracting aspects of one's personhood, and life. They can make the creative and emotional substance of a person's life saleable and immortal – and, under conditions of capitalism, it will become one more product of the average individual's exertions that they are alienated from.

Just as poor and working-class people may feel economic pressure to take out unfavourable loans, or sell their blood plasma or their eggs, they may feel intense economic pressure to sell their ecto-metadata to get by in the future. They may also have it harvested from them for free by chat-bots that they interact with. This presents us with one possible form of future data violence from emerging technologies.

CONCLUSION: VIOLENCE AND HOPE

In this chapter, we have explored and unpacked the definition of data violence. Using a range of examples, from facial recognition databases to voice-activated personal assistants, we have demonstrated the wide range of forms such violence can take. We have also looked at the equally wide range of places such violence can appear, spanning from how data is collected and used, to the overarching justifications for data-based decision-making, and even proposals for *addressing* this violence. Taken together, this does not tell a particularly positive story: not only data-based systems but even efforts to reform them can backfire and cause harm. But we do want to emphasize that there is reason for hope here too, and that research on data violence is not undertaken to depress people. Rather: the point is to provide new ways of spotting violence, and new ways of correcting it – ones that also correct for some of the limitations of existing efforts at reform.

One place that we find hope is in efforts at using data in activism (or: activism around data) that seek to understand how efforts at change and reform are co-opted – and how we might avoid our attempts to make things better becoming yet another form of violence. On the former, there is the work of Jonathan Cinnamon (2020), who has been doing deep, rich work looking at how data is used in activist practices to make things better, and what does (or does not) work. On the latter, we have Ben Green (2024), whose upcoming book is designed specifically as a handbook for those looking to use data to improve the world. Rather than focus on vast claims about how data-based research will inherently improve the world, his goal is to teach data-based researchers and workers how to avoid the pitfalls that can lead to their own investment in data violence. Research like this, and data technologies based on it, can (we hope) lead to the world we deserve, instead of the world we were promised.

REFERENCES

Amoore, L. (2013). *The politics of possibility: Risk and security beyond probability*. Durham, NC: Duke University Press.

Armstrong, G., & Norris, C. (2020). *The maximum surveillance society: The rise of CCTV*. London, UK: Routledge.

Barabas, C. (2023). Care as (re)capture: Data colonialism and race during times of crisis. *New Media & Society*, doi: 10.1177/14614448231165902.

Bettcher, Talia Mae. (2007). Evil deceivers and make-believers: On transphobic violence and the politics of illusion. *Hypatia* 22.3 (2007): 43–65.

Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.

Browne, S. (2015). *Dark matters: On the surveillance of blackness*. Durham, NC: Duke University Press.

Brugge, D., & Goble, R. (2002). The history of uranium mining and the Navajo people. *American Journal of Public Health*, *92*(9), 1410–1419.

Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR.

Cath, C. & Keyes, O. (2022). Your thoughts for a penny? Capital, complicity and AI ethics. In Phan, T., Goldenfein, J. Kuch, D & Mann, M. (Eds.) *Economies of virtue* (pp. 24–39). Amsterdam: Institute of Network Cultures.

Cinnamon, J. (2020). Attack the data: Agency, power, and technopolitics in South African data activism. *Annals of the American Association of Geographers*, *110*(3), 623–639.

Denetdale, J. (2009). *The long walk: The forced Navajo exile*. New York City, NY: Infobase Publishing.

Densley, J. A., & Pyrooz, D. C. (2020). The matrix in context: Taking stock of police gang databases in London and beyond. *Youth Justice*, *20*(1–2), 11–30.

Flaxman, A. D. & Keyes, O. (2024). The risk of linked census data to transgender youth: A simulation study. *Journal of Privacy and Anonymity.*

Gray, M. L., & Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Boston, MA: Houghton Mifflin Harcourt.

Green, B. (2024). *Algorithmic realism: Data science practices to promote social justice* (forthcoming).

Hoffmann, A. L. (2018, April 30). Data violence and how bad engineering choices can damage society. *Medium*. Retrieved from: https://medium.com/@annaeveryday/data-violence-and-how-bad-engineering-choices-can-damage-society-39e44150e1d4

Hoffmann, A. L. (2021). Terms of inclusion: Data, discourse, violence. *New Media & Society*, *23*(12), 3539–3556.

Introna, L., & Wood, D. (2004). Picturing algorithmic surveillance: The politics of facial recognition systems. *Surveillance & Society*, *2*(2/3), 177–198.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, *1*(9), 389–399.

Keyes, O. (2020). Automating autism: Disability, discourse, and artificial intelligence. *The Journal of Sociotechnical Critique*, *1*(1), 8.

Keyes, O., & Austin, J. (2022). Feeling fixes: Mess and emotion in algorithmic audits. *Big Data & Society*, *9*(2), doi: 10.1177/205395172211137.

Long, Q. (2011). Queering the census: Privacy, accountability, and public policy implications of adding sexual orientation and gender identity questions to the US census. *Documents to the People*, *39*, 15.

Miller, J., Gounev, P., Pap, A. L., Wagman, D., Balogi, A., Bezlov, T., … & Vargha, L. (2008). Racism and police stops: Adapting US and British debates to continental Europe. *European Journal of Criminology*, *5*(2), 161–191.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., … & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 220–229).

Obermeyer, Z., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm that guides health decisions for 70 million people. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 89–89).

Parvin, N. (2018). Our bodies in the trolley's path, or why self-driving cars must *not* be

programmed to kill. *Science, Technology, & Human Values*, *43*(2), 302–323.

Rincón, C., Keyes, O., & Cath, C. (2021). Speaking from experience: Trans/non-binary requirements for voice-activated AI. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–27.

Ryan, L. (2017, January 18). *Rage against the ghost in the machine* – linux conf au 2017. [Video] YouTube. https://www.youtube.com/watch?v=jvVTcserZ_8

Singh, R., & Jackson, S. (2021). Seeing like an infrastructure: Low-resolution citizens and the Aadhaar identification project. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–26.

Schubert, J. D. (2014). Suffering/symbolic violence. In Grenfell, M. (Ed.) *Pierre Bourdieu* (pp. 179–194). Abindgon, UK: Routledge

Sloane, M. (2019). Inequality is the name of the game: Thoughts on the emerging field of technology, ethics and social justice. In *Weizenbaum Conference* (p. 9). DEU.

Steiger, M., Bharucha, T. J., Venkatagiri, S., Riedl, M. J., & Lease, M. (2021). The psychological well-being of content moderators: The emotional labor of commercial moderation and avenues for improving support. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–14).

Stevens, N., & Keyes, O. (2021). Seeing infrastructure: Race, facial recognition and the politics of data. *Cultural Studies*, *35*(4–5), 833–853.

Tong, A. (2021, March 18). What happens when your AI chatbot stops loving you back? *Reuters*. Retrieved from: https://www.reuters.com/technology/what-happens-when-your-ai-chatbot-stops-loving-you-back-2023-03-18/

Uliasz, R. (2021). Seeing like an algorithm: Operative images and emergent subjects. *AI & Society*, *36*, 1233–1241.

Vinsel, L. (2021, February 1). You're doing it wrong: Notes on criticism and technology hype. *Medium*. Retrieved from: https://sts-news.medium.com/youre-doing-it-wrong-notes-on-criticism-and-technology-hype-18b08b4307e5

Voyles, T. B. (2015). *Wastelanding: Legacies of uranium mining in Navajo country*. Minneapolis, MN: University of Minnesota Press.

Watkins, E. A. (2020, October). Took a pic and got declined, vexed and perplexed: facial recognition in algorithmic management. In *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing* (pp. 177–182).

Westbrook, L. (2020). *Unlivable lives: Violence and identity in transgender activism*. Berkeley, CA: University of California Press.