

Who Do They Think You Are? Categories, Classification, and Profiling

By: Zara Rahman

SURVEILLANCE TECHNOLOGY

The following passage is from Zara Rahman's new book Machine Readable Me: The Hidden Ways Tech Shapes Our Identities. Rahman is an author and researcher investigating the intersection of power, technology and justice, and her book builds upon over a decade of research on how data about who we are is used to shape the paths available to us in life. Machine Readable Me is published by 404 Ink and available to purchase now (DRM-free ebook or paperback, shipped from the UK).

As an immigrant twice over, first through my parents who moved to the UK from Bangladesh half a century ago, then through my own decision to move to Germany in the early 2010s, the feeling of not quite fitting into people's expected boxes, or the given categories on bureaucratic forms, is quite normal for me.

Sometimes it's left me unsure what the most accurate answer is for seemingly simple questions. Drop-down menus with the question "Where are you from?" leave me baffled. An imprecise question, usually with a specific and hidden intention behind it. If it's in a digital system, there's no way of asking what that specific intention actually is.

Borders change, but society's insistence on classifications and categories rarely keeps up.

My answer? The UK, I suppose, though I haven't lived there for over a decade. It could be Germany, as it's the place I've spent most of my adult life. If this interaction was to happen in person, at least in Germany, I rarely get to say either of those options without some gesture towards my skin color, a single eyebrow raised with "no, where are you *really* from?" exposing a stunning ignorance of colonialism and

migration. Borders change, but society's insistence on classifications and categories rarely keeps up.

But these problems are somewhat trivial when we consider what category deviation can do to people's lives. In India, 2010, the government introduced the Aadhaar card, a mandatory digital identification card which provides every Indian citizen with a unique ID number which can be used for proof of identity, address and more. The Aadhaar database is currently the largest database of biometric data in the entire world, with over a billion people represented in it. This type of digital ID system, one unique number to be used for multiple government and private sector services, served as an inspiration to Kenya's proposed Huduma Namba card, which I discussed earlier.

The general idea doesn't sound too bad, as many of us struggle with remembering logins and passwords, or being unable to transfer data from one government service to another. When my mother moved house in her 70s, it astonished me how many different places she needed to update her address. If she had had one unique number for everything related to government services (and, increasingly, also private sector services too), it's easy to imagine that those systems would be much more streamlined and efficient for everyone involved.



But of course, it's not that simple. Friction is (and, perhaps, should be more often) purposefully built into digital systems—making it harder to easily transfer data from one system to another as a security feature, or adding in a moment of pause to allow the user to double check that they're doing what they intended, for example. Researcher Anja Kovacs writes about the myriad of harms that have arisen from Aadhaar, including its connection with digital wallets leading to the identities of sex workers being revealed, which is a problem not just for general stigma but also because both 'soliciting and living off the income of sex work are crimes' in India.^[1]

Using just one system for everything can have problems, too. In 2017, Indian media outlet *The Wire* reported that there were at least 25 hunger-related deaths directly linked to issues with the Aadhaar card.^[2] Even though 64-year-old Premani Kunwar had “authenticated” herself using the Aadhaar system, her social security pension was paid into someone else’s bank account and she couldn’t get it changed. Eighty-year-old Khetrabasi Pradhan died because his age was incorrectly recorded on his voter card, so the system classified him as not being eligible for a pension when in fact he was.^[3]

The way that we are categorized matters—deciding what gets measured or counted and what gets ignored are all ways of exerting power over populations. Philosopher Kwame Anthony Appiah says, “once labels are applied to people, ideas about people who fit the label come to have social and psychological effects. In particular, these ideas shape the ways people conceive of themselves and their projects.”^[4] And in this way, the power that comes from deciding on labels and categories goes beyond simply making populations legible or controlling how populations are treated, but it also affects how we think about each other and ourselves. That’s a huge power to hand over to anyone.

For the Rohingya, an ethnic minority population who are indigenous to Myanmar, the right to self-determine their classifications has been a question of life and death. The Myanmar government has tried over decades to erase the Rohingya from Myanmar, starting by changing citizenship laws to no longer recognize the Rohingya as one of the 135 “national races” of Burma in 1982.

This bureaucratic violence^[5] was combined with physical attempts at genocide, forced labor, and destruction of property and homes, with the largest genocide taking place as recently as 2017. For decades, the Myanmar government has refused to use the term “Rohingya,” instead describing them as “Bengali.” Even when defending her country’s government against charges of genocide at the International Court of Justice in 2019, then-head of state Aung San Suu Kyi spoke for thirty minutes and did not use the word “Rohingya” once, preferring instead to describe them as “Muslims” or “members of Rakhine communities.” Rohingya activist Wai Wai Nu told Al Jazeera at the time “refusing to use the term Rohingya means she still doesn’t acknowledge the root cause of the genocide allegation.”^[6]

When there’s a sudden influx of people on the move or refugees to a new country, head counts are essential, so that humanitarian and governmental agencies know what kind of support the community might need.

After the genocide, hundreds of thousands of Rohingya fled Myanmar and ended up in neighboring Bangladesh, where one of the world’s largest refugee camps stands today. When there’s a sudden influx of people on the move or refugees to a new country, head counts are essential, so that humanitarian and governmental agencies know what kind of support the community might need. When the Rohingya arrived in Bangladesh, the UNHCR agreed with the government of Bangladesh to carry out the counting

exercise together, then issue identification cards to the Rohingya.^[7] But once again, the cards did not acknowledge their Rohingya identity. Instead, they were identified as being from Myanmar. On some identification documents from Myanmar, they were 29 labelled as being from Bangladesh. No official documents acknowledged their ethnic identity as what they wanted it to be: Rohingya.

In November 2018, Rohingya communities in refugee camps at Cox's Bazar in Bangladesh went on strike, with their demands shared on social media:^[8]

We are doing this to demand:

1. Must stop forcing Rohingya refugees to take the Smart Card
2. Must stop barricading Rohingya inside Camp 21 due to their refusing to take the Smart Card
3. Must include our ethnic name "Rohingya" on the Smart Card, not just "Forcibly Displaced Myanmar National"
4. Stop the collecting of our biodata and do not share biodata already collected with the Myanmar Government [sic]

Their strike was even more notable within the broader context of collecting or labelling people according to ethnicity because there's a strong precedent in many countries against collecting or documenting ethnicity data—for good reasons, as we'll see later on in this chapter. In contrast to that precedent, this strike, and the Rohingya's demands, were so clearly connected to their context, given that they had been forced to flee their home country and 30 had their identity and personhood denied for so long, both bureaucratic and explicitly violent.

Intentionally refusing to allow members of marginalized communities to identify themselves as they see fit is a technique used by states all over the world to oppress and silence groups. Speaking to Maya Ch'orti' and Zapotec environmental scientist Dr. Jessica Hernandez, Maya Mam activist Juanita Cabrera López notes that "our Indigeneity as Maya people does not appear or is reflected in our birth certificates. It does not appear in our IDs... It is really in the interest of states to deny our existence and to deny our identity."^[9] But when context differs, the very same type of data—ethnicity data—can play a very different role. This is where acts of protest and resistance can include actions like destroying data.

In the Netherlands, prior to Nazi occupation, a "comprehensive population registration system for administrative and statistical purposes"^[10] had already been completed, which included the collection of ethnicity data, notably flagging who was Jewish and who was not. This structure, aimed at following people "from cradle to grave,"^[11] was then adapted to create special registration systems covering Jewish and Roma populations within the Netherlands, and "played an important role in the[ir] apprehension."^[12]

In 1943, the Dutch resistance, including members of the PBC, carried out a complex and daring operation to bomb the Amsterdam civil registry office to prevent Nazis from continuing to use the registry to identify and persecute Jews.

Forging identification cards became a vital act of resistance against the Nazis, many created by the 31 *Persoonsbewijzen* centrale (Identity Card Center, PBC), which created an estimated 65,000 forged cards.^[13] But as the identification cards were matched against records in the civil registries, it wasn't enough to carry a forged card as it could easily be proven fake. So, in 1943, the Dutch resistance, including members of the PBC, carried out a complex and daring operation to bomb the Amsterdam civil registry office to prevent Nazis from continuing to use the registry to identify and persecute Jews. Though they weren't entirely successful in their efforts, they managed to destroy 15 percent of the records (800,000 cards), likely saving the lives of many people. Resistance members were even helped by sympathetic fire fighters who intentionally delayed deploying fire engines to give the fire as much time as possible to spread—and they sprayed water liberally to destroy the records, too.^[14] Twelve members of the group who carried out the bombing were executed in 1944.

Despite these incredibly courageous efforts, Dutch Jews had the highest death rate among all other occupied western European countries, at 73 percent of its population. As comparison, the death rate among the Jewish population in Belgium was 40 percent and in France, 25 percent. Dutch digital rights activist Hans de Zwart notes there are many, many lessons about privacy rights that we can learn from this tragic period in history—including that personal data can obviously be used for nefarious means.^[15]

These examples demonstrate how deeply context-specific identity-focused data is. The broader context around *who* is doing the data collection, *how* the data is collected, and *who* decides its use can be the difference between life and death, without exaggeration.

What might be a fight for self-determination for one community, and a fight that they're willing to forego food and water for, might be a target on the backs for another community, one that they want to get rid of at all costs. Solidarity is needed in all such situations, recognition of the differing forms of oppression at play, and support in resistance, even if that support manifests itself in drastically different ways in different contexts.

During the Covid-19 pandemic of 2020, analysis by English government body Public Health England entitled "Disparities in the Risk and Outcomes of Covid-19" revealed that people of Bangladeshi ethnicity were twice as likely to die of Covid than people of white British ethnicity.^[16] And indeed, throughout the three waves of the pandemic (thus far) in the United Kingdom, the rate of death involving Covid-19 has been highest for Bangladeshi people than any other ethnic group, while all ethnic minority groups face higher risks than white British people.

Learning this hit me hard, as my entire family is of Bangladeshi origin. To know that they were more likely to be at risk was gutting to read. Even worse was the realization that nobody knew precisely why. Was it embedded inequities within the healthcare system itself? I knew of research in the United States that revealed Black people's pain is generally taken less seriously and that there's significant and documented racial bias in pain treatment recommendations.^[17] Or was it due to other factors?

In their research they found that Black and ethnic minority people were more likely to be classed as “key workers” in the UK...

Report after report confirmed this pattern. The Office for National Statistics released corroborating data,^[18] and *The Lancet* called for ethnicity and Covid-19 to be an “urgent public health research priority.”^[19] But British race equality and civil rights think tank The Runnymede Trust noted that it wasn't just that health and ethnicity were deeply intertwined. “Covid-19 is not just a health crisis; it is also a social and economic crisis...and the ability to cope...is vastly different for people from different ethnic and socioeconomic backgrounds.”^[20] In their research they found that Black and ethnic minority people were more likely to be classed as “key workers” in the UK; to be living in a household with more people; to be working outside their house during the pandemic; and to have experienced financial strife due to the pandemic.^[21]

Being able to see these patterns was only possible because of disaggregated data—and that was only possible because of the ethnicity categories that were available for people to choose from in the Census in England. Those categories and the way they're presented differs greatly across different countries and even across different governmental services offered in the United Kingdom.

I was deeply grateful, perhaps naively, that this awful pattern could be identified, though I remained largely unconvinced that this discovery would lead to the systemic change necessary to properly address it. But here in Germany, the site of some of the worst offenses of historical data misuse, ethnicity data is now rarely collected. This lack of data means that inequities based on race are hard to prove, assess and understand, and even harder to address as a result. In 2020, a Berlin-based NGO based called Each One Teach One took it upon themselves to address this dearth of data. They set up the Afrozensus,^[22] a project that asked people of African origin living in Germany for more information about themselves, so that this often-overlooked community had some quantitative data to use when advocating for their needs and rights in policy debates.

It's because of data collected about ethnicity that important insights demonstrating race-based inequities can be proven. Without that kind of data, it would be impossible to say how people of different ethnicities were affected by Covid, for example. So, perhaps counterintuitively, the desire to protect ethnic minorities by not collecting ethnicity data can actually end up harming them, leaving minority groups in countries like Germany unable to prove that they are being systemically discriminated against. That is

starting to change among some communities though, as science journalist Hristio Boytchev notes in *Nature*, “cultural unease [about collecting data on race or ethnicity] is starting to shift.”^[23]



It means that collecting this kind of data is, if done right, an act of empowerment, somewhat similar in goals and motivation to the Rohingya. Until race- or ethnicity-based discrimination has been entirely eliminated, there’s a clear need for this kind of data to be collected—to be able to highlight where those inequities are, and, ideally, to suggest interventions and systemic changes that can be made to address them.

If and when this kind of data is collected, though, the structures of accountability and trust around it really matter, as well as *how* it is collected and under what terms. Databases of such sensitive information can easily fall into the wrong hands and be used to persecute instead of empowering. We’ve seen this throughout history, from the 1939 Census in Germany including a supplementary card inquiring if a person’s grandparents are or were a Jew, along with the assurance that “anonymity would be respected,”^[24] which in fact went straight to a department run by Adolf Eichmann to “close gaps on in their existing card index on Jews,”^[25] breaking that anonymity immediately, to biometric data collected by Western governments falling into the hands of the Taliban in Afghanistan in 2021.^[26]

In short, respecting people’s self-determination and autonomy when it comes to sensitive data about who we are is complex and hard to do well. But ignoring that kind of data is also not an option.

As activist and professor Ibram X. Kendi explains, “terminating racial categories is potentially the last, not the first, step in the anti-racist struggle.”^[27] It really matters who decides what the categories are, how that data is gathered, who holds it, and if and when it’s deleted. The data we need is shaped by the

society we live in, which have racial injustices built into their very core. We need data about ethnicity to make those injustices visible. But as ever, just having the data is not enough.

Footnotes

- 1 Anja Kovacs, ["When Our Bodies Become Data, Where Does that Leave Us?"](#) *DeepDives*, May 28, 2020.
- 2 ["Of 42 'Hunger-Related' Deaths Since 2017, 25 'Linked to Aadhaar Issues',"](#) *The Wire*, September 21, 2018.
- 3 ["Of 42 'Hunger-Related' Deaths Since 2017, 25 'Linked to Aadhaar Issues',"](#) *The Wire*.
- 4 Kwame Anthony Appiah, *The Ethics of Identity* (Princeton, NJ: Princeton University Press, 2005), 66.
- 5 David Graeber, *The Utopia of Rules: On Technology, Stupidity, and the Secret Joys of Bureaucracy* (Brooklyn, NY: Melville House, 2016).
- 6 Anealla Safdar and Usaid Siddiqui, ["ICJ Speech: Suu Kyi Fails to Use 'Rohingya' to Describe Minority,"](#) *Al Jazeera*, December 13, 2019.
- 7 UNHCR, ["UNHCR Refugee Population Factsheet,"](#) July 2018.
- 8 John Quinley III (@john_hq3), ["Full press release about the #Rohingya refugee strike in the camps,"](#) X, November 25, 2018, 11:39 p.m.
- 9 Jessica Hernandez, *Fresh Banana Leaves: Healing Indigenous Landscapes through Indigenous Science* (New York: North Atlantic Books, 2022), 233.
- 10 William Seltzer and Margo Anderson, ["The Dark Side of Numbers: The Role of Population Data Systems in Human Rights Abuses,"](#) *Social Research*, 68, no. 2 (Summer 2001): 481-513.
- 11 Seltzer and Anderson, "The Dark Side of Numbers."
- 12 Seltzer and Anderson, "The Dark Side of Numbers."
- 13 ["Forged ID Documents,"](#) Gedenkstätte Stille Helden, accessed September 21, 2023.
- 14 Hans de Zwart, ["During World War II, We Did Have Something to Hide,"](#) Medium, April 30, 2015.
- 15 de Zwart, "During World War II, We Did Have Something to Hide."
- 16 Public Health England, *Disparities in the Risk and Outcomes of COVID-19* (London: August 2020).
- 17 Kelly M. Hoffman et al., ["Racial Bias in Pain Assessment and Treatment Recommendations, and False Beliefs about Biological Differences between Blacks and Whites,"](#) *Proceedings of the National Academy of Sciences of the United States of America* 113, no. 16 (2016): 4296-4301.
- 18 ["Updating Ethnic Contrasts in Deaths Involving the Coronavirus \(COVID-19\), England: 8 December 2020 to 1 December 2021,"](#) Office for National Statistics, January 26, 2022.
- 19 Manish Pareek et al., ["Ethnicity and COVID-19: An Urgent Public Health Research Priority,"](#) *The Lancet* 395, no. 10234: (2020): 1421-1422.

- 20 Zubaida Haque, Laia Becares, and Nick Treloar, *Over-Exposed and Under-Protected: The Devastating Impact of COVID-19 on Black and Minority Ethnic Communities in Great Britain* (London: The Runnymede Trust, August 2020).
- 21 Haque, Becares, and Treloar, *Over-Exposed and Under-Protected*.
- 22 Afrozensus, accessed September 21, 2023.
- 23 Hristio Boytchev, "Diversity in German Science: Researchers Push for Missing Ethnicity Data," *Nature*, April 5, 2023.
- 24 Michael Burleigh and Wolfgang Wippermann, *The Racial State: Germany 1933-1945* (Cambridge University Press, 1991), 59.
- 25 Burleigh and Wippermann, *The Racial State*, 59.
- 26 Human Rights Watch, "New Evidence that Biometric Data Systems Imperil Afghans," news release, March 30, 2022.
- 27 Ibram X. Kendi, *How To Be an Antiracist* (New York: One World, 2019), 54.