The Weaponization of Education Data

Audrey Watters

on 11 Dec 2017

13 min read

This is part five of my annual look at the year's "top ed-tech stories"

As in previous years, it would be quite easy to fill a whole article in this series on "data insecurity," on the data breaches and cyberattacks that continue to plague education – both schools and software. The issue extends well beyond education technology, of course, and in 2017 we witnessed yet again a number of high profile incidents (including some corporate admissions of breaches that had happened in years past): that over 140 million Social Security Numbers and other personal data had been stolen in a data breach at Equifax, for starters; that every single account at Yahoo – some 3 billion in all – had been affected in its 2013 breach.

In education, there were breaches at colleges and universities, breaches at K–12 schools, breaches at the Department of Education, breaches at education technology companies, and breaches with software schools commonly use. 77 million users accounts stolen from Edmodo. A file configuration error at Schoolzilla that exposed the data of some 1.3 million students. A ransomware attack at a school system in Maine. A ransomware attack at a community college in Texas. Computers affected by the WannaCry virus at the Massachusetts Institute of Technology, Trinity College, the University of Washington, North Dakota State University, the University of Maine, and elsewhere. 14 million college email username and passwords for sale on "the dark Web." W2 phishing scams at a school district in Texas. W2 phishing scams at a school district in Connecticut. W2 phishing scams at a school district in Minnesota. Phishing emails posing as Ofsted. Phishing emails posing as the University of California student health plan. $11.8 million scammed from MacEwan University. Keyloggers at the University of Iowa. Keyloggers at the University of Kansas. A hacked school Twitter account in Florida. A privacy breach at Stanford. Data stolen from a community college's health clinic. A data breach at a school board in Ontario. A data breach at the Chicago Public

Schools. A malware attack at the University of Alberta. And then there was the ominously named "Dark Overlord," who held the data of multiple school districts for ransom, in one case sending parents text messages threatening to kill their children if they did not pay up.

And that's not even *remotely* close to the complete list of hacks, scams, breaches, and thefts this year.

EdTech Strategies' Doug Levin launched a new project in 2017 – "The K–12 Cyber Incident Map" – that visualizes the breaches, ransomware attacks, DDOS attacks, phishing attacks, and so on that have been reported at US public schools. *Some 280 incidents since January 2016*. Levin's message has been consistent this year and in the past: schools are simply not prepared to address the cybersecurity threats they're facing – they're unprepared for the attacks on their IT infrastructure(including vulnerabilities in the various technology products they've adopted), and they're unprepared for attacks on individuals associated with schools (more on this in a forthcoming article about "free speech" on campus).

And yet schools continue to invest millions and millions of dollars in more and more technology. Investing in hardware and software. Investing, as I chronicled in the previous article in this series, in "the platform economy." Investing, that is, in technology companies and *in the ideology that underpins them*: that schools and their vendors must collect an ever-increasing amount of data.

We heard plenty of stories this year, arguing for precisely that: Every student is a potential data point, we were told. "Super users" of certain ed-tech products demand more data collection. More data collection will help schools gauge student well-being. More data collection will improve course design. More data collection will improve student learning and student outcomes. More data collection will mean more diversity at schools. More data collection will mean more customization – a "Netflix-and-Amazon-like experience." It will enable "personalization." "The Higher Ed Learning Revolution," as NPR put it: "Tracking Each Student's Every Move." And students are totally on board, we were told: they want even more of their data to be collected. (I call bullshit.) And perhaps my favorite: "Want Your Students to Remember You in 20 Years? Start Holding Weekly Data Conferences."

These justifications for more data collection aren't new; nor is an opposition to education's data regime. The research organization Data & Society published a report this year on "the Legacy of inBloom," a proposed data infrastructure initiative that shut down three years ago but that has surely left its mark on how schools and ed-tech companies frame discussions about data and "personalization." And yet, even in the face of ongoing pushback against data collection and concerns about data insecurity – from parents, educators, students, and others – even with the hundreds of hacks and breaches, some industry groups do still try to argue that that "Your Concerns About Student Privacy Are Being Exploited."

But 2017 made it clear, I'd like to think, that the dangers education technology and its penchant for data collection aren't simply a matter of a potential loss of privacy or a potential loss of data. The stakes now are much, much higher.

Education Technology in a Time of Trump

Immediately following the 2016 elections, I tweeted that "Under a Trump administration: I very much want ed-tech companies and schools to reconsider collecting so much data about students." I'd embed the tweet here in this article, but I deleted it. I delete all my old social media now on a regular and ongoing basis. I do so because I am uncomfortable about the ways in which our personal data is so easily used against us.

The first public talk I gave this year was on 2 February at the University of Richmond – "Ed-Tech in a Time of Trump." I repeated my call: education institutions and education companies must rethink their collection of data. I pointed to several historical examples of how the collection, categorization, and analysis of data led to discriminatory and even deadly political practices – racism and the US Census, for example, and the history of IBM and how its statistical analysis helped the Nazis identify Jews.

Even in the earliest days of the Trump administration – even in the campaign itself – it seemed obvious to me that immigrants to the US would targeted, *that immigration data, whether overtly collected or algorithmically inferred, would be weaponized*. Just one week after taking office President Trump signed an executive order banning all refugees from entering the US, as well as barring entry for citizens from seven majority-Muslim countries. The order had an immediate effect on scholars and students, many of whom had returned home over the holidays and were stuck outside the country – some even stranded mid-transit. There were immediately protests at airports and objections in the courts, the latter led in part by public universities who claimed they had legal standing to challenge the travel ban as it harmed their mission as research and teaching institutions. The ban was blocked, altered by the administration, blocked again, altered, blocked, altered, blocked… (The latest, almost one year later: the US Supreme Court has allowed the ban to take effect while legal challenges to it continue.)

No surprise, colleges and universities have expressed some concern about how the travel ban – and attitudes in the US towards foreigners more broadly – will affect their ability to recruit and retain students and scholars. There were other high profile incidents as well: the refusal, for example, to give visas to the all-girls robotics team from Afghanistan.

The actions of the administration should not come as a surprise. President Trump ran on a white ethno-nationalist platform and from his opening campaign speech promised that he would strengthen the country's borders and boost immigration enforcement. ICE, the agency responsible for the latter, has dramatically increased the number of arrests of immigrants this year, many of whom did not fall into categories previously targeted by law enforcement – they did not have criminal records, for example – and many of whom were detained at places not previously targeted by ICE either, including schools. Parents in LA were picked up as they dropped their children off at school. School officials in Pasadena were accused of threatening to call ICE on students or on their family members. Washington University in St. Louis threatened to report international students to ICE if they unionized. Some students were arrested during their regular check-ins with immigration. One teenager was arrested hours before his prom. A fourth grader in Queens was allegedly interrogated about his family's immigration status. After an ICE raid outside of Las Cruces, New Mexico, over 2000 students in the district missed school, fearing more ICE activity.

Students and families are afraid. They're afraid to go to school. They're afraid to collect benefits that, as citizens they have every right to receive. They're afraid to seek medical care.

According to data from Pew Research Center, there are approximately 11 million undocumented immigrants in the US. How this number affects schools is a bit harder to calculate, as students might be citizens and live with family members who are undocumented or students might be undocumented and live with families who are citizens and so on. Education Trust-West said this year that it believed one in eight students in California schools had at least one parent that was undocumented, for example, and the Los Angeles Unified School Board announced this year that it was committed to protecting its students and their families from ICE. Other campuses and cities also reaffirmed they would act as "sanctuaries," a move some Republican lawmakers tried to outlaw.

In September, the Trump administration announced its plans to end the Deferred Action for Childhood Arrivals (DACA) program. "Unprecedented," Vox's Dara Lind called it: "There's never really been a time when a generation of people, raised and rooted in the United States, has been stripped of official recognition and pushed back into the precarity of unauthorized-immigrant life."

DACA was established by the previous administration, providing a protected status to those who were brought to the country illegally when they were children. Some 800,000 undocumented immigrants qualified for the program, which had enabled these DREAMers to legally pursue work and education opportunities. An estimated 20,000 are educators – some 5000 in California, 2000 in New York, 2000 in Texas. They are in limbo, along with many of the students in their classrooms. Again, universities have said they would protect their students who are DREAMers, and some have sued the Trump administration for violating these students' rights. Meanwhile, DREAMers are already being arrested and deported.

Some technology companies have joined some of these lawsuits challenging the Trump administration's immigration policies. The tech industry has also expressed its own frustrations over curbs to visas for foreign technical workers and foreign entrepreneurs – "startup visas" – as well as threats to green cards.

But technology companies – *some of the very same technology companies* – are also working with the Trump administration to build the software for its "extreme vetting" programs. This software would track and analyze the social media and digital activity of visa holders in order to identify those who might be "high risk."
It's a process incredibly similar to what's marketed in education as "learning analytics" – tracking the social media and digital activity of students to identify those who might be "at risk."

What happens to all the DREAMers' data – data that they willingly handed over to the federal government? Will it be weaponized, as The Daily Beast argued in September it might?

What happens to all the data that schools and their software vendors have collected about students? Can that data be used to glean their immigration status (or their religion)? Can their status be deduced even if the specific data points about nationality or immigration status are not collected? Because this is, of course, the promise of algorithms and analytics – making inferences based on the data that's available. Profiling. Grouping. Predicting.

What education technology practices and products have schools already adopted that might be putting their students at risk – adding geo-tracking devices, for example, to laptops given to students in migrant education programs.

Will this data be used to punish students – to refuse them admission or aid?

How will all the education data and analysis that's gathered be used? Education technology companies and big data proponents always have the sunniest futures to sell. But what are the implications of algorithmic decision-making in a time of Trump?

Algorithmic Discrimination

"Should big data be used to discourage poor students from university?" ZDNet asked this summer, describing an algorithm that could help predict whether or not low-income students would be successful at school – but not so more resources could be directed their way to help them succeed. Not so that they received more help, more money, more support. Nope. Rather, the big data would be utilized so these students could be discouraged from going to school in the first place.

Typically, stories about predictive analytics in education aren't framed that way, no surprise. Education technology proponents like to say that big data will be used to *encourage* low-income students – or at least to send them nudges and notifications when students appear, algorithmically at least, to be struggling. These algorithmic products are marketed as helping students succeed and – no surprise – helping schools make more money by retaining their tuition dollars.

There are major ethical implications of these sorts of analytics in education. If, for example, a school doesn't have the resources to *help* struggling students, perhaps as that ZDNet article suggests, it would rather discourage them from attending.

There seemed to be much more discussion in the media this year about ethics and the discriminatory tendencies in algorithmic decision-making, as more mainstream attention was brought to the topic through a combination of investigative journalism and academic scholarship. (Through the work, for example, of Cathy O'Neil, Zeynep Tufekci, Julia Angwin, Frank Pasquale, and many others.)

That attention came in part because of the ongoing concerns about the role technology companies played in the 2016 election – something I discussed briefly in the first article in this series. How do algorithms shape news and information sharing? What did

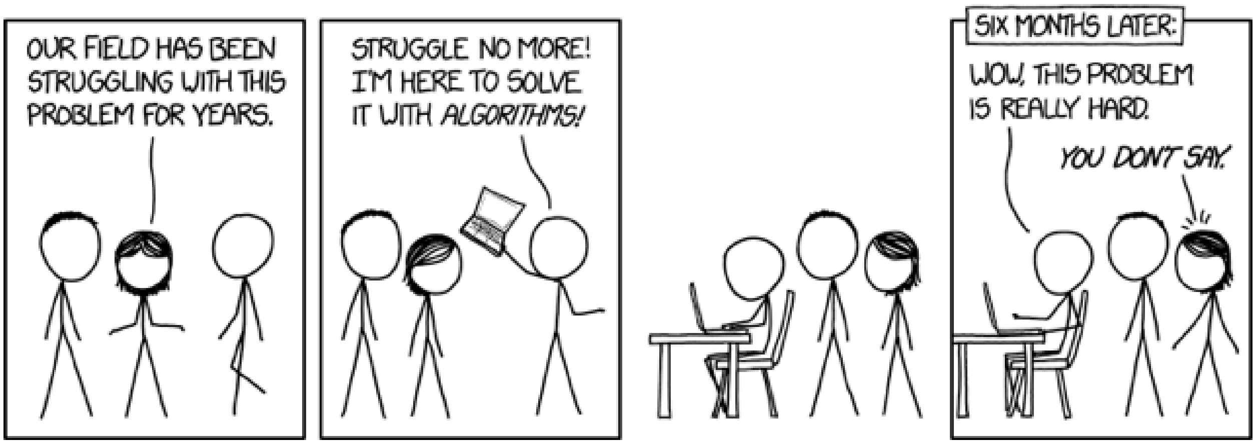
Facebook and Google's algorithms show people in their news feeds and in their searches during the election? How were specific groups targeted for certain kinds of advertising, messaging, and "promoted content"? ProPublica found, for example, that "Facebook Enabled Advertisers to Reach 'Jew Haters'" during the presidential campaign. Google similarly allowed advertisers to target people who were searching for racist phrases. To reiterate: algorithms on these platforms dictate what you see and *what you don't see*. And that view isn't simply a matter of political party affiliation. ProPublica had discovered last year that Facebook let advertisers target housing ads to white audiences only – a violation of the federal Fair Housing Act – and in follow-up reporting this fall the publication found that it was still able to place discriminatory ads, despite Facebook's insistence that it had fixed the problem.

The algorithms of these platform companies are increasingly inscrutable – a "black box," as Frank Pasquale has described them – and that's certainly part of the problem. Journalists and scholars are not allowed to peer "under the hood," if you will (although there has been some talk of requiring companies to "open" their algorithms to scrutiny.) So how do we know their algorithms "work" – or rather, what sorts of work do these algorithms do? Again, it's not just about the news feed. Facebook, for its part, has also promised that its algorithms will identify terrorists and individuals who might be suicidal. *Its algorithms are not merely informative; they are extra-judicial.*

And clearly the use of this sort of algorithmic decision-making extends well beyond what happens in high profile technology companies and what happens on social media. Algorithms are being used to determine prison sentences. They're being used to predict criminal activity. They're being used to identify children who might be at risk of abuse. In all these cases, algorithms raise serious questions about the potential for discrimination. In the latter, it's also simply a question of the algorithm failing altogether – that is, failing to accurately identify children at risk as the Illinois Department of Child and Family Services recently discovered after spending $366,000 on data-mining software.

And yet, despite the repeated concerns about the discriminatory potential for these algorithms and the ongoing questions about whether or not these systems are accurate or effective or just, many schools have plowed forward with adopting these sorts of tools. *Algorithmic decision-making is the basis for "personalized learning,"* a trend that venture capitalists and venture philanthropists and education reformers and technology companies want very much to happen (whether the research says "it works" or not). And whether they're "all in" on "personalized learning," many schools are adopting analytics and surveillance technologies to monitor and predict student behavior: to identify cyberbullying and suicide threats; to recommend what students should be reading; to recommend what lessons students should be working on; to ascertain if teachers should be awarded tenure; to determine school bus routes; to identify learning disabilities; to identify college students who are struggling with classes; to identify K–12 students who are struggling with classes; to recommend students enroll in certain courses or pursue certain degrees; to help colleges decide who to admit in the first place. "Can you predict your students' final grade at the start of the course?" Technológica de Monterrey asked on its website, "Yes, you can with Artificial Intelligence." "Will You Graduate?" asked an article in The New York Times. "Ask Big Data."

Education technology, as a field and as an industry, places an incredible amount of faith in data and algorithms to address social problems that are incredibly complex.


OUR FIELD HAS BEEN STRUGGLING WITH THIS PROBLEM FOR YEARS.
STRUGGLE NO MORE! I'M HERE TO SOLVE IT WITH *ALGORITHMS!*
SIX MONTHS LATER:
WOW, THIS PROBLEM IS REALLY HARD.
YOU DON'T SAY.

But that faith in data is just part of the problem. Just *collecting* data alters how decisions get made, some research suggests. And the types of data that are collected is facilitated by the types of technologies and systems already in place – the learning management system, most obviously.

Algorithms get layered on top of these existing structures. And artificial intelligence comes with deep, deep biases – biases at the very core of the discipline. Racism. Sexism. Biases in language and in image recognition. Biases based on the training data. Biases that comes from the engineers. Wrapped in the shine of science (and pseudoscience) and backed with billions of dollars of venture capital and PR, these biases are, as Blaise Agüera y Arcas, Margaret Mitchell and Alexander Todorov have argued, "Physiognomy's New Clothes."

If these algorithms make use of existing data and are layered on top of existing practices and systems, then it seems even more likely that they will reinforce the education system's existing biases rather than radically upend them. Racial biases in school discipline, for example. Biases in admissions decisions. Teachers' biases. Administrators' biases. Department of Education biases. Long-standing beliefs and practices about who students are and what students need. Legal precedentas to what rights students have and do not have while at school (and perhaps even while at home). Indeed, perhaps some of these beliefs and practices (and fantasies) are why surveillance technologies have such a powerful appeal to many in education.
"This will go down on your permanent record…"

Education Technology and School Surveillance

Schools surveil to prevent cheating, which we're told is now more pervasive because of new technologies. They install cameras. They install microphones. They monitor social media. They demand biometric datafrom students in order to prove their identity. They use proctoring software with facial recognition. They buy software that scans for plagiarism and software that monitors students' location while they're doing schoolwork. They adopt devices that monitor students at school and students at home. Schools surveil to prevent violence. They put body cameras on campus police. Schools surveil to track attendance. They install finger print scanners at the schoolhouse door. Schools surveil to monitor students while they're at school. They use facial recognition devices to make sure students are paying attention. They install finger print scanners in the lunch room. They install iris-scanners in the cafeteria.

Schools surveil to ensure student safety and well-being. They use heart-rate monitors to track students' physical activity. They use fitness trackers to monitor students' sleep. They monitor all sorts of activities – sleep, exercise, and more – of student athletes. They scan the license plates of those who come on campus. They install facial recognition devices in dormitories. They filter websites, blocking content deemed "inappropriate." They monitor students' ID cards to track their location – before, during, and after school.

Students can see how these systems work, you know – the decisions that are human-made and the decisions that are machine-made and the decisions that are historical and the decisions that are structural. They worry that they are being set up to fail. They worry that their data – their very identities – are being weaponized against them. It's not simply "the algorithm" that causes educational inequalities to persist. Students know that. Algorithms are just becoming an easier way to justify unjust decision-making.

Financial data on the major corporations and investors involved in this and all the trends I cover in this series can be found on funding.hackeducation.com.

WRITTEN BY

Audrey Watters

Published 11 Dec 2017

CREDITS

2018 · About the author

Header image credits

Back to Archives