TINKER-ING WITH MACHINE LEARNING: THE LEGALITY AND CONSEQUENCES OF ONLINE SURVEILLANCE OF STUDENTS

Amy B. Cyphert∗

All across the nation, high schools and middle schools are quietly entering into contracts with software companies to monitor the online activity of their students, attempting to predict the next school shooter or to intervene with a student who might be contemplating suicide. Systems using algorithms powered by machine learning trawl the Facebook posts of fifteen–year–olds and weed through the Twitter feeds of seventeen–year–olds. When certain keywords or features are flagged, the posts are forwarded to school administrators, who can decide whether the post requires an intervention and whether the student requires discipline. Who (or what) decides what these keywords are? What protections are given to the massive amounts of student data these third parties are collecting? Do parents and students even realize such online surveillance is happening?

Too often, the answers to these questions are unclear. This Article explores the legal and policy questions related to this new era of surveillance, which is fueled by machine learning. Although this technology is relatively new to schools, it has been used for decades now in the criminal justice system, which has embraced sentencing algorithms and predictive policing. As is true with so many things in the criminal justice system, there is evidence that these technologies have had a disproportionate impact on people of color. In much the same way, evidence is emerging that the online monitoring of students is having a disproportionate impact on students of color. Despite having an aura of neutrality, at each stage in the machine learning process, there is a possibility for bias to creep in.

The legality of schools entering into contracts for third-party surveillance of their students is uncertain, as courts have not ruled on it specifically and have just begun to rule on the legality of schools regulating student internet speech at

∗ Lecturer in Law and Director, ASPIRE, West Virginia University. B.A., 2001, Carnegie Mellon University; J.D., 2005, Harvard Law School. Devan Simmons provided excellent research assistance and keen insights on drafts of this Article. The staff of the Nevada Law Journal worked hard to improve this Article, and I thank them for it. My colleagues Valarie Blake, Amber Brugnoli, Jena Martin, Alison Peck, Kirsha Trychta, and Elaine Wilson provided invaluable motivation, encouragement, and feedback. The Hodges Fund provided support for the submission of this Article. As always, thank you to Bethany, Josh, and Violet. I am grateful to Sam Perl for so many things; for this Article, I thank him especially for patiently explaining machine learning and being a sounding board for my ideas.

all. The fact that every state has a cyberbullying law that arguably requires schools to police their students' online speech complicates the legality question. This Article explores what legal challenges to third-party surveillance under the First and Fourth Amendments and the Equal Protection Clause might look like, and the likelihood of success of those arguments. Because the legal challenges are hypothetical at best, and perhaps years away, the Article concludes with some policy recommendations aimed at ensuring safety and fairness for all students.

TABLE OF CONTENTS

INTRODUCTION 459
I. MACHINE LEARNING AND BIAS 461
A. Defining Machine Learning 461
B. Use in Criminal Justice System 464
1. Sentencing Algorithms 465
2. Predictive Policing 468
C. Current Cyber-Monitoring of High School Students 469
D. The Myth of Objectivity and Neutrality: How Bias Can Exist in Each Stage of Machine Learning 473
1. Specifying the Outputs 474
2. Constructing the Training Data Set 475
3. Feature Engineering 476
4. Training the Model 477
5. Testing and Validating the Model 477
6. Interpreting the Outputs 478
7. A Black Box? 478
II. POTENTIAL LEGAL CHALLENGES 479
A. Cyberbullying Laws 480
B. First Amendment Challenges 481
1. Challenge by Disciplined Students 481
2. Prior Restraint Challenge to Online Monitoring 487
C. Fourth Amendment Challenges 489
D. Equal Protection 495
III. POLICY RECOMMENDATIONS 497
A. Invest in High Quality High School Counselors 498
B. Provide Students and Families with Transparency and Privacy Protections 499
C. Take a More Intentional and Multidisciplinary Approach to the Use of Machine Learning 500
CONCLUSION 500

INTRODUCTION

In 2011, a seventeen–year–old named Mishka,[1] angry that his friends had recently been jumped in a fight, penned a Facebook post full of violence, including saying that his high school was "asking for a [expletive] shooting, or something."[2] Friends saw the post and alerted school officials, who contacted the police.[3] By the time psychologist Dr. John Van Dreal, who ran the Safety and Risk Management Program for Mishka's Oregon public school system, arrived, Mishka was in handcuffs.[4] Mishka and his classmates were lucky: their school system employed a risk management program, and Dr. Van Dreal was able to help talk with Mishka about what caused him to write the post.[5] Realizing that Mishka had no intention of harming anyone, Dr. Van Dreal helped Mishka avoid being charged with a criminal offense.[6] Dr. Van Dreal also arranged for him to attend a smaller school, where he found mentors, graduated on time, and is today a twenty–five–year–old working for a security firm.[7]

Had Mishka's story happened today, just eight short years later, it might have looked very different. First, instead of his friends noticing his troubled Facebook post and alerting his school, it might have been flagged by a machine learning algorithm developed by a software company that Mishka's school paid tens of thousands of dollars to per year. Although Mishka's post was clearly alarming and made obvious mention of possible violence, a post flagged by the algorithm might be seemingly innocuous and yet still contain terms or features that the algorithm had determined are statistically correlated with a higher likelihood of violence. An alert would be sent to school officials, though the algorithm would not necessarily explain what features about the post triggered it.[8] Dr. Van Dreal and the risk management program? They might have been cut in order to pay for the third-party monitoring conducted by the software company.[9] A school official would be left to decide whether Mishka's post warranted some form of school discipline, or even a referral to the authorities.[10]

---

[1] Rhitu Chatterjee & Rebecca Davis, *When Teens Threaten Violence, A Community Responds with Compassion*, NPR (Feb. 13, 2019, 7:30 PM), https://www.npr.org/sections/health-shots/2019/02/13/693136117/when-teens-threaten-violence-a-community-responds-with-compassion [https://perma.cc/QF69-PURS]. NPR did not provide Mishka's full name, in order to protect his privacy. *Id.*

[2] *Id.*

[3] *Id.*

[4] *Id.*

[5] *Id.*

[6] *Id.*

[7] *Id.*

[8] *See* Edward C. Baig, *Can Artificial Intelligence Prevent the Next Parkland Shooting?*, USA TODAY (Feb. 14, 2019, 2:30 PM), https://www.usatoday.com/story/tech/2019/02/13/preventing-next-parkland-artificial-intelligence-may-help/2801369002/ [https://perma.cc/UTS7-Y23A].

[9] Schools across the nation are facing budget shortfalls. Amanda Litvinov & Mary Ellen Flannery, *The High Cost of Education Budget Cuts*, NEA TODAY (July 16, 2018),

Predictive machine learning algorithms are currently being used to monitor the online activity of school students across the country.[11] These algorithms are familiar to many defense attorneys and criminal justice advocates, as the same technology underlies predictive policing and sentencing algorithms.[12] In the criminal justice system, these technological advances have come with alarming stories of bias against certain marginalized groups. What role should these algorithms have in schools monitoring their students? What are the implications, both legally and practically, of such a surveillance scheme?

Part I of this Article defines machine learning, attempting to go beyond the "black box" metaphor and to break down and explore the discrete stages of machine learning. At each stage, there is potential for bias, despite the sticky notion that algorithmic decision-making is somehow more objective and unbiased than that of humans. Examples of bias in predictive algorithms in the criminal justice system are examined, specifically sentencing algorithms and predictive policing. When the ACLU and others decried the use of these algorithms in the criminal justice sector, software companies shifted their focus to schools, and Part I also explores how surprisingly widespread the practice of schools using third parties to conduct online surveillance of their students has become.

Because the practice of using a third-party company's algorithm to monitor school students is still a relatively new practice, the legality of it is unclear at this point. Accordingly, Part II examines various legal challenges that impacted students might bring. All fifty states now have cyberbullying laws, some of which seemingly require that schools monitor their students' online activities, which complicates legal challenges.[13] Section II.B lays out two distinct First Amendment challenges that can be made. Section II.B lays out two distinct First Amendment challenges that can be made. First, disciplined students can challenge their suspensions or other discipline for their online speech *a la* the seminal *Tinker* case. Second, students can challenge the very act of hiring a third party to monitor student speech as a prior restraint. Section II.C lays out Fourth Amendment arguments, which would be premised under the theory that students have a privacy interest in their online speech. Finally, Section II.D looks at why disciplined students who argue they were unfairly targeted for discipline because of their race or other status are unlikely to succeed on an equal protection claim.

---

http://neatoday.org/2018/07/16/the-high-cost-of-education-budget-cuts [https://perma.cc/8JTC-9BKJ] ("In 2015, [twenty-nine] states provided less school funding than in 2008. Since state funding fuels nearly half of the nation's K–12 spending, these cuts have huge implications.").

[10] *See* Baig, *supra* note 8 (describing one product that uses machine learning to scan students' social media posts, and then alerts school administrators, parents and possibly law enforcement officials to potential problems).

[11] *See infra* Section I.C.

[12] *See infra* Section I.B.

[13] *Is Cyberbullying Illegal in Your State?*, GAGGLE, https://www.gaggle.net/speaks/is-cyberbullying-illegal-in-your-state [https://perma.cc/93CJ-ZNWF] (last visited Dec. 30, 2019).

It is probably unlikely that any court will rule that schools are fully foreclosed from hiring a third party to monitor their students' online activity, notwithstanding serious potential legal problems with that practice. Thus, Part III makes a series of policy recommendations to help guide schools and policymakers as they move through this unchartered terrain. The first recommendation is to invest in high-quality counselors, a proven way to address both self-harm and violence against other students in schools. Second, schools need to work with students and their parents to be transparent about the kinds of monitoring they are engaging in, and to invite feedback and input. Finally, the programmers who are developing these monitoring algorithms need to be sure to form teams that are multidisciplinary in order to produce the least-biased products, and thus the least-biased results.

I. MACHINE LEARNING AND BIAS

When many lawyers think of "artificial intelligence," they conjure up images of Orwellian robots or Bladerunner. Artificial intelligence is actually a sub-field of computer science, one that focuses on how to train computers to do "intelligent" tasks that have traditionally been done by humans.[14] The commands that programmers use to get computers to accomplish these tasks are algorithms.[15] At the most basic level, an algorithm is simply a series of commands that will solve a problem or accomplish a goal.[16] In that sense, a recipe is an algorithm, as is a step-by-step set of driving directions. When most people use the term algorithm, they are referring to computer codes and to algorithms that are a series of commands telling a computer what to do.[17]

A. Defining Machine Learning

Machine learning is an umbrella term to describe a special subset of algorithms wherein a computer is programmed to revise the code it is using as it works, based on the results it is generating.[18] Frequently in machine learning, the algorithm works to identify patterns in the data it is examining, develop cer-

---

[14] Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 88–89 (2014).

[15] *See* David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 671 (2017); Surden, *supra* note 14, at 89.

[16] Lehr & Ohm, *supra* note 15, at 671.

[17] Paul Ford, *What is Code?*, BLOOMBERG BUSINESSWEEK (July 11, 2015), https://www.bloomberg.com/graphics/2015-paul-ford-what-is-code [https://perma.cc/7K3U-MF3T] (noting that most people say "algorithm" when they actually mean "code" or "software").

[18] Surden, *supra* note 14, at 89. For a comprehensive exploration of the various stages of machine learning, *see* Lehr & Ohm, *supra* note 15, at 655 (concluding that there are eight "key" steps to machine learning: (1) Problem Definition, (2) Data Collection, (3) Data Cleaning, (4) Summary Statistics Review, (5) Data Partitioning, (6) Model Selection, (7) Model Training, and (8) Model Deployment). Part I of this Article discusses certain steps in the production of a machine learning algorithm and how bias is possible in each. *See infra* Section I.D.

tain rules from those patterns (or "learns" from them), and then uses those rules to categorize the next set of data it looks at.[19] Most of us are familiar with common examples of machine learning, even if we do not recognize them as such. For example, the spam filter on your email is an example of machine learning.[20] It has been programmed to pay attention to the characteristics of emails you commonly delete—details such as keywords in the subject line or the identity of a sender—and to learn over time which you mark as spam and to adjust accordingly.[21] Machine learning is simply programming a computer to incorporate results into the next round of whatever you've asked it to do, with an aim toward making each round better.[22] Put another way, "[m]achine [l]earning focuses on the question of how to get computers to program themselves (from experience plus some initial structure)."[23]

Scholars have begun in recent years to examine the various ways that algorithms and machine learning can be inadvertent tools for deepening inequality. The argument is not that the algorithms themselves have been intentionally coded to disadvantage certain groups, such as an algorithm that automatically awards poorer credit scores to applicants who are black or disabled. Rather, the concern is that algorithms "can reproduce existing patterns of discrimination, inherit the prejudice of prior decision makers, or simply reflect the widespread biases that persist in society," even when "they have not been manually programmed to do so."[24]

---

[19] Vera Eidelman, *The First Amendment Case for Public Access to Secret Algorithms Used in Criminal Trials*, 34 GA. ST. U. L. REV. 915, 920 (2018) ("Machine learning occurs when a computer identifies patterns from a preexisting or training set of data, learns from those patterns, and incorporates the lessons into the algorithm.").

[20] Surden, *supra* note 14, at 90.

[21] *See id.* at 89–90 (defining machine learning as "a subfield of computer science concerned with computer programs that are able to learn from experience and thus improve their performance over time," and explaining that spam filters are a common example of this technology).

[22] *See* Chris Meserole, *What is Machine Learning?*, BROOKINGS: REPORT (Oct. 4, 2018), https://www.brookings.edu/research/what-is-machine-learning [https://perma.cc/A46M-NDF9] (describing the process of building an algorithm to detect human faces and concluding "[t]he magic of deep learning is that the algorithm learns to do all this on its own. The only thing a researcher does is feed the algorithm a bunch of images and specify a few key parameters, like how many layers to use and how many neurons should be in each layer, and the algorithm does the rest. At each pass through the data, the algorithm makes an educated guess about what type of information each neuron should look for, and then updates each guess based on how well it works.").

[23] Tom M. Mitchell, *The Discipline of Machine Learning*, *in* MACHINE LEARNING TECHNICAL REPORTS 1 (2006), http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf [https://perma.cc/8DFG-KWNX].

[24] Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671, 674 (2016).

For example, facial recognition software is notoriously bad at correctly identifying the faces of people of color.[25] When an MIT researcher examined three apps that purport to identify gender based on photographs, she found "that darker-skinned females are the most misclassified group (with error rates of up to 34.7%)" while "[t]he maximum error rate for lighter-skinned males is 0.8%."[26] Google had to apologize in 2015 when its image-recognition photo software—"Google Photos"—labeled photos of black people as "gorillas."[27] The negative impacts of bias in facial recognition software extend beyond embarrassing corporate gaffes. This software can be used for purposes ranging from identifying criminal subjects from security video footage to identifying melanoma from an image.[28] That the software has such racially inconsistent outcomes has very serious implications and has led the ACLU to decry its use in policing and its "potentially devastating outcomes."[29] The President of Microsoft even took the extraordinary step of publishing a blog post urging governments to regulate the technology—*technology that his company makes*—because "certain uses of facial recognition technology increase the risk of decisions, outcomes and experiences that are biased and even in violation of discrimination laws."[30]

Why does facial recognition software have so much trouble with non-white faces? The answer is complicated and contested. One reason may be that so many software engineers and coders are themselves white men, and as they build the algorithms that underlie facial recognition software, "they focus on

---

[25] Steve Lohr, *Facial Recognition is Accurate, if You're a White Guy*, N.Y. TIMES (Feb. 9, 2018), https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html [https://perma.cc/79WA-QS9Q].

[26] Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACHINE LEARNING RES. 1, 1 (2018).

[27] Conor Dougherty, *Google Photos Mistakenly Labels Black People 'Gorillas'*, N.Y. TIMES (July 1, 2015, 7:01 PM), https://bits.blogs.nytimes.com/2015/07/01/google-photos-mistakenly-labels-black-people-gorillas [https://perma.cc/R3X3-VYCU].

[28] Buolamwini & Gebru, *supra* note 26, at 1–2; *see also* Dhruv Khullar, *A.I. Could Worsen Health Disparities*, N.Y. TIMES (Jan. 31, 2019), https://www.nytimes.com/2019/01/31/opinion/ai-bias-healthcare.html [https://perma.cc/78SE-JR27] (listing the variety of ways artificial intelligence is now used as a diagnostic tool in medicine and pointing out the risks of bias: "A recent study found that some facial recognition programs incorrectly classify less than [one] percent of light-skinned men but more than one-third of dark-skinned women. What happens when we rely on such algorithms to diagnose melanoma on light versus dark skin?").

[29] Natasha Singer, *Amazon's Facial Recognition Wrongly Identifies 28 Lawmakers, A.C.L.U. Says*, N.Y. TIMES (July 26, 2018), https://www.nytimes.com/2018/07/26/technology/amazon-aclu-facial-recognition-congress.html [https://perma.cc/LEW6-A55D].

[30] Brad Smith, *Facial Recognition: It's Time for Action*, MICROSOFT: MICROSOFT ON THE ISSUES (Dec. 6, 2018), https://blogs.microsoft.com/on-the-issues/2018/12/06/facial-recognition-its-time-for-action [https://perma.cc/FV7V-YYEE] ("We believe it's important for governments in 2019 to start adopting laws to regulate this technology.").

facial features that may be more visible in one race, but not another."[31] Programmers, like most of us, will tend to focus on what they know and to extrapolate from their own experiences.[32] That initial programming bias is then further reinforced through machine learning.[33] The training data sets that data scientists use to teach the model how to recognize faces disproportionately feature white faces,[34] as the majority of image data sets in use now have a Western, Eurocentric focus.[35] Put another way, "[t]he code 'learns' by looking at more white people—which doesn't help it improve with a diverse array of races."[36] These issues—a lack of diversity in programmers, a lack of diversity in a training dataset, and others—can lead to serious bias problems in systems that use machine learning. That bias, and the results of it, are especially dangerous when machine learning systems are used in the criminal justice system.

B. Use in Criminal Justice System

The use of artificial intelligence and especially machine learning has been embraced in important ways by the criminal justice system, with results that have at times been disturbingly disproportionate for people of color. It is important to ground our analysis of online monitoring of students with the ways that machine learning technology has been used in the criminal justice system for three important reasons. First, the historical trend is that criminal justice tactics frequently creep down into school discipline, in a sort of reversion of the traditional school-to-prison pipeline, a phenomenon discussed below in Section I.C. Second, the software companies that created this technology and marketed it to police departments and probation officers have expanded their scope and now market it to school districts.[37] Third, because federal courts have not yet ruled on the constitutionality of schools outsourcing online monitoring of their

---

[31] Ali Breland, *How White Engineers Built Racist Code–and Why it's Dangerous for Black People*, GUARDIAN (Dec. 4, 2017), https://www.theguardian.com/technology/2017/dec/04/racist-facial-recognition-white-coders-black-people-police [https://perma.cc/UY43-5PNE].

[32] *Id.*

[33] *See* Paul Teich, *Artificial Intelligence Can Reinforce Bias, Cloud Giants Announce Tools for AI Fairness*, FORBES (Sept. 24, 2018), https://www.forbes.com/sites/paulteich/2018/09/24/artificial-intelligence-can-reinforce-bias-cloud-giants-announce-tools-for-ai-fairness [https://perma.cc/8STB-HWSK] (citing to a white paper cautioning that machine learning can exacerbate bias in a data training set, given the repeated cycles used in machine learning, which can "[create] a vicious cycle.").

[34] *Id.*; *see also* Breland, supra note 31. Training data sets are explained more fully *infra* Section I.D.2.

[35] Shreya Shankar et al., *No Classification Without Representation: Assessing Geodiversity Issues in Open Data Sets for the Developing World*, GOOGLE RES. 2–3 (Nov. 22, 2017), https://arxiv.org/pdf/1711.08536.pdf [https://perma.cc/6WX7-ECLF] (concluding that in one open image data set with more than 9 million images, 60 percent of the images came from North America and Europe, with only 3 percent combined from China and India, the two most populous countries in the world).

[36] Breland, *supra* note 31.

[37] *See infra* Section I.C.

students to third parties, it is helpful to examine the corollary of how courts have responded to challenges from defendants when probation departments or police departments have embraced predictive technologies.[38] Accordingly, as we attempt to predict the outcomes of the expanded use of algorithms to monitor students, the use of algorithms in the criminal justice system offers important insights.

1. Sentencing Algorithms

A judge's determination of a defendant's likelihood of recidivism—how likely the judge believes it to be that the defendant will commit a future crime—is an important factor in determining that defendant's sentence, especially any carceral sentence.[39] If a judge believes it is less likely that the defendant will reoffend, she is more likely to issue a non-carceral sentence or a shorter carceral sentence.[40] The federal sentencing guidelines bake in a view on recidivism by including a defendant's criminal history in the recommended sentence calculation: under the guidelines, those who have offended before are viewed as more likely to offend again, and therefore require a lengthier sentence to deter them and protect the public.[41] Some scholars argue that the increase in judicial discretion brought forward by decisions such as *United States v. Booker*, which made the sentencing guidelines advisory rather than mandatory, has led to a greater judicial and legislative reliance on recidivism prediction instruments.[42]

Although scholars have been debating the use of predictive algorithms in criminal sentencing for several decades now,[43] much of the public at large

---

[38] *See infra* Section I.B.

[39] *See, e.g.*, United States v. Gayle, 389 F.3d 406, 409–10 (2d Cir. 2004) (noting that a defendant's criminal history category, one of two important calculations made under the Federal Sentencing Guidelines, "serves as a proxy for his likelihood of recidivism.").

[40] Sara Chodosh, *Courts Use Algorithms to Help Determine Sentencing, but Random People Get the Same Results*, POPULAR SCI. (Jan. 18, 2018), https://www.popsci.com/g00/recidivism-algorithm-random-bias [https://perma.cc/YRK5-6T52] ("A person who's unlikely to commit another crime is less of a threat to society, so a judge will generally give them a shorter sentence.").

[41] *See* U.S. SENTENCING GUIDELINES MANUAL ch. 4, pt. A, introductory cmt. (U.S. SENTENCING COMM'N 2018). These guidelines are structured to account for both the offense level and the criminal history background. *Id.* § 5A cmt. n.1. Thus, the higher the criminal history background, the longer the possible sentence. *See id.* § 5A.

[42] *See, e.g.*, Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 809–11 (2014) (tracing the use of recidivism prediction tools from the earliest use and noting an acceleration post-*United States v. Booker*, and noting "[t]ight sentencing guidelines leave little room to consider the defendant's individual risk, but in discretionary systems, judges are expected to assess it.").

[43] *See* Robert García, *"Garbage in, Gospel Out": Criminal Discovery, Computer Reliability, and the Constitution*, 38 UCLA L. REV. 1043, 1049, 1142–43 (1991) (discussing back in 1991 the "promise but also [] the threat of computers in the criminal justice system," and urging defense counsel to "aggressively seek discovery concerning computerized information from the government, including access to the underlying information, programs,

learned about them through an explosive ProPublica report in May 2016.[44] People who were arrested in Broward County, Florida were assigned a "risk assessment score" from an algorithm created by software company Northpointe.[45] The county paid about $22,000 a month for Northpointe's software—called COMPAS ("Correctional Offender Management Profiling for Alternative Sanctions")—in the hopes that it would help judges and probation officers more accurately predict the likelihood of recidivism for criminal defendants and "to help identify which defendants were low risk enough to be released on bail pending trial."[46] The ProPublica reporters examined the risk scores of 7,000 people arrested in 2013 and 2014 and compared those scores to the actual rates of recidivism.[47] Black defendants were 77% more likely to be labeled as "at higher risk of committing a future violent crime" than white defendants,[48] despite the fact that "race" was not one of the fields that COMPAS included in producing its risk scores.[49]

Despite the proliferation of stories like the ProPublica one,[50] proponents of sentencing algorithms claim that they are somehow less biased than judges or juries. Indeed, the newly revised Model Penal Code section encourages the use such actuarial predictions of risk, noting that "well-designed actuarial risk-assessment tools offer better predictions of future behavior than the clinical judgments of treatment professionals such as psychiatrists and psychologists, or the intuitions of criminal-justice professionals such as judges and probation officers."[51] However, a recent study concluded that the widely used commercial risk assessment software COMPAS—the subject of the ProPublica report men-

computers, manuals, procedures, tests, and personnel, in order to protect a client against the use of unreliable information during plea discussions, at pretrial hearings, at trial, and at sentencing.").

[44] Chodosh, *supra* note 40 ("Algorithms sold to courts across the United States have been crunching those numbers since 2000. And they did so without much oversight or criticism, until *ProPublica* released an investigation showing the bias of one particular system against black defendants.").

[45] Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/Y5QY-YBGX].

[46] *Id.* Broward County used the COMPAS software for pretrial bail decisions, not for sentencing purposes. *Id.*

[47] *Id.*

[48] *Id.* It is important to note that Northpointe sent ProPublica a letter, wherein it "criticized ProPublica's methodology and defended the accuracy of its test: 'Northpointe does not agree that the results of your analysis, or the claims being made based upon that analysis, are correct or that they accurately reflect the outcomes from the application of the model.' " *Id.*

[49] *Id.*

[50] *See, e.g.*, Stephanie Wykstra, *Just How Transparent Can a Criminal Justice Algorithm Be?*, SLATE (July 3, 2018), https://slate.com/technology/2018/07/pennsylvania-commission-on-sentencing-is-trying-to-make-its-algorithm-transparent.html [https://perma.cc/V9LD-KSPT] (discussing activists who spoke out against Pennsylvania's proposed predictive sentencing algorithm, noting that they were worried that it would "increase racial disparities").

[51] MODEL PENAL CODE § 6.03, cmt. f. (AM. LAW INST., Tentative Draft No. 3, 2014).

tioned above—was "no more accurate or fair than the predictions of people with little or no criminal justice expertise."[52]

To date, the Wisconsin Supreme Court is the tribunal that has performed the most substantive review of a challenge to the use of sentencing algorithms.[53] In *State v. Loomis*,[54] the defendant Eric Loomis challenged the use of his COMPAS score in his sentencing, which "indicated that he presented a high risk of recidivism on all three bar charts [representing pretrial recidivism risk, general recidivism risk, and violent recidivism risk.]"[55] Loomis argued that the use of the score violated his due process rights, in part because the algorithm used by COMPAS is a trade secret and so "the proprietary nature of COMPAS prevent[ed] him from assessing its accuracy . . . . "[56] The court rejected that argument, holding that "if used properly with an awareness of the limitations and cautions, [use] of a COMPAS risk assessment at sentencing does not violate a defendant's right to due process."[57] In so holding, the court noted that COMPAS used only publicly available information in formulating a risk score, and so "to the extent that Loomis's risk assessment [was] based upon his answers to questions and publicly available data about his criminal history, Loomis had the opportunity to verify that the questions and answers listed on the COMPAS report were accurate."[58]

The court did, however, provide some guidelines to lower courts about the proper way to use COMPAS scores. First, the court held that risks scores could not be used: "(1) to determine whether an offender is incarcerated; or (2) to determine the severity of the sentence," or (3) "as the determinative factor in deciding whether an offender can be supervised safely and effectively in the community."[59] Further, the scores must come with certain written cautions warning judges about their proper use.[60] Those cautions include that "[a] COMPAS risk assessment compares defendants to a national sample, but no cross-validation study for a Wisconsin population has yet been completed. Risk assessment tools must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations[,]" and that "COMPAS was not developed for use at sentencing, but was intended for use by the Department of

---

[52] Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, 4 SCI. ADVANCES 1, 3 (2018); *see also* Chodosh, *supra* note 40 (when "a group of 462 people were simply asked whether they thought [a] defendant was likely to commit another crime in the next two years[] [t]hey did so with almost exactly the same accuracy—and bias—as the COMPAS algorithm.").

[53] The United States Supreme Court declined to review the decision. Loomis v. Wisconsin, 137 S. Ct. 2290 (2017).

[54] State v. Loomis, 881 N.W.2d 749 (Wis. 2016), *cert. denied*, 137 S. Ct. 2290 (2017).

[55] *Id.* at 753–55.

[56] *Id.* at 757.

[57] *Id.* at 770.

[58] *Id.* at 761.

[59] *Id.* at 769.

[60] *Id.*

Corrections in making determinations regarding treatment, supervision, and parole."[61]

It is unclear what impact these admonitions from the Wisconsin Supreme Court will have on lower courts who continue to use the COMPAS scores as part of the sentencing process. As Dartmouth computer scientist Hany Farid said when he studied COMPAS:

> Our concern is that when you have software like COMPAS that's a black box, that sounds complicated and fancy, that the judges may not be applying the proportional amount of confidence as they would if we said '[twelve] people online think this person is high risk' . . . Maybe we should be a little concerned that we have multiple commercial entities selling algorithms to courts that haven't been analyzed. Maybe someone like the Department of Justice should be in the business of putting these algorithms through a vetting process. That seems like a reasonable thing to do.[62]

2. Predictive Policing

Algorithms are not just used to help sentence people once they have been arrested; they are also used to determine who is most likely to commit a crime in the first place. In Chicago, almost 400,000 citizens have an "official police risk score[,]" ranging from 1 to 500-plus.[63] The algorithm that police used to develop these scores is not publicly available, but it influences a stunning array of police decisions, from who receives a home visit by police officers to who receives additional police surveillance.[64] In 2013, the Chicago Police Department placed approximately 400 people on a "heat list," a list of people who "had all been forecast to be potentially involved in violent crime, based on an analysis of geographic location and arrest data."[65] The heat list included people like Robert McDaniel, a then twenty–two–year–old black man who received an unannounced visit from the police warning him "not to commit any further crimes[,]" despite the fact that he had committed no crimes and had no violent criminal record.[66]

The use of predictive policing like that in Chicago is on the rise, and "[t]he technology has far outpaced any legal or political accountability and has largely escaped academic scrutiny."[67] For over a year in Boston, police identified po-

---

[61] *Id.* at 769–70.

[62] Chodosh, *supra* note 40.

[63] Andrew G. Ferguson, *The Police are Using Computer Algorithms to Tell If You're a Threat*, TIME (Oct. 3, 2017), https://time.com/4966125/police-departments-algorithms-chicago [https://perma.cc/NCL6-3XD4].

[64] *Id.*

[65] Kristian Lum & William Isaac, *To Predict and Serve?*, 13 SIGNIFICANCE 14, 15 (2016).

[66] *Id.*

[67] Andrew G. Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1109, 1109 (2017).

tential threats by monitoring people's social media accounts.[68] But the tool they used has been decried as unfair, given that it "swept up the posts of people using the hashtag #MuslimLivesMatter and a lawmaker's Facebook update about racial inequality" in its attempt to predict violence.[69] The police department planned to spend $1.4 million for the software but "dropped those plans amid backlash from groups like the ACLU," who argued that the social media monitoring "appear[ed] to have had little benefit to public safety while unfairly focusing on groups such as Muslims."[70]

The Brennan Center for Justice has sued the New York City Police Department over its use of predictive policing, especially its refusal to provide certain information about the algorithm at use.[71] That case is presently winding its way through discovery, and the New York state judge presiding over the case has ordered the city to turn over large amounts of information about the software it uses.[72] Similar lawsuits have been filed by activists in Chicago[73] and Los Angeles.[74] The outcomes of these cases may well be instructive for predicting how courts may treat challenges to the online monitoring of students.

C. Current Cyber-Monitoring of High School Students

As the companies that provide social-media-monitoring services came under fire for their contracts with police districts, some quietly turned their attention to servicing schools instead.[75] All across America, school districts have entered into agreements with software companies to use artificial intelligence to monitor students. In Billerica, Massachusetts, high schools monitor students' social-media accounts, both to prevent school violence and also to flag those at

---

[68] Alanna D. Richer, *Boston Police's Social Media Surveillance Unfairly Targeted Muslims, ACLU Says*, BOSTON GLOBE (Feb. 7, 2018), https://www.bostonglobe.com/metro/2018/02/07/boston-police-social-media-surveillance-unfairly-targeted-muslims-aclu-says/9JUpzPmy8Tsr5RLxvCm61M/story.html [https://perma.cc/E9R5-35V9].

[69] *Id.*

[70] *Id.*

[71] Rachel Levinson-Waldman & Erica Posey, *Predictive Policing Goes to Court*, BRENNAN CTR. FOR JUST. (Sept. 5, 2017), https://www.brennancenter.org/our-work/analysis-opinion/predictive-policing-goes-court [https://perma.cc/G95J-6C9A].

[72] Brennan Ctr. for Justice v. N.Y.C. Police Dep't., No. 160541/2016, 2017 WL 6610414, at *8 (N.Y. Sup. Ct. Dec. 27, 2017).

[73] Dave Collins, *Police Departments Sued Over Predictive Policing Programs*, POLICEONE (Jul. 5, 2018), https://www.policeone.com/legal/articles/police-departments-sued-over-predictive-policing-programs-oEOyziTEEgyMrLNv [https://perma.cc/FS7U-B5J9].

[74] Brenda Gazzar, *Activists File Lawsuit Over LAPD's Predictive Policing Program*, GOV'T TECH. (Feb. 14, 2018), https://www.govtech.com/public-safety/Activists-File-Lawsuit-Over-LAPDs-Predictive-Policing-Program.html [https://perma.cc/RT5W-J9AX].

[75] *See* Aaron Liebowitz, *Could Monitoring Students on Social Media Stop the Next School Shooting?*, N.Y. TIMES (Sept. 6, 2018), https://www.nytimes.com/2018/09/06/us/social-media-monitoring-school-shootings.html [https://perma.cc/T3XX-JM9W].

risk of suicide.[76] In Huntsville, Alabama, students have been expelled for tweets.[77] The Glendale Unified School District in California hired a company to monitor online bullying.[78] One company, Social Sentinel, claims to have contracts with school districts in thirty states.[79]

Machine learning features prominently in the monitoring services these companies provide to the school districts they contract with. The algorithms that are being developed are predicated on the theory that we may be able to accurately predict which students are going to become violent by looking for patterns in their online activities.[80] One CEO explained that when his company developed its social media monitoring algorithm:

> "We went back . . . and looked at the language that school shooters, as one example, have used in the past in various manifestos—what's been published or that they've shared on social media . . . . And we went to understand similarities and patterns. And we can teach computers, to an extent, how to identify some of that nuance."[81]

The companies that offer the social-media monitoring to school districts commonly use a method called "geofencing," wherein software is programmed to trawl various social-media sites within a specified geographic area and to flag posts containing certain keywords for school administrators to review.[82] Despite the serious problems with facial recognition software and bias discussed earlier, that technology is also being marketed to school districts for use on students.[83]

How effective is this monitoring? It's difficult to assess the efficacy of social-media monitoring to disrupt something like a school shooting, since those

---

[76] Lynn Jolicoeur & Lisa Mullins, *To Detect Threats and Prevent Suicides, Schools Pay Company to Scan Social Media Posts*, WBUR (Mar. 22, 2018), https://www.wbur.org/news/2018/03/22/school-threats-suicide-prevention-tech [https://perma.cc/N9U5-M74U].

[77] Liebowitz, *supra* note 75.

[78] *Id.*

[79] *Id.* Even those school districts that have not hired third parties to monitor their students through algorithms still monitor their students in the more "old-fashioned" way—by "relying on students or parents as whistleblowers who bring alarming circumstances to the school administration's attention." Catherine E. Mendola, Note, *Big Brother as Parent: Using Surveillance to Patrol Students' Internet Speech*, 35 B.C. J.L. & SOC. JUST. 153, 168 (2015).

[80] *See, e.g.*, Randy Rieland, *Can Artificial Intelligence Help Stop School Shootings?*, SMITHSONIAN (June 22, 2018), https://www.smithsonianmag.com/innovation/can-artificial-intelligence-help-stop-school-shootings [https://perma.cc/AXJ5-8NBP] ("The idea is that algorithms might be able to better analyze data related to school shootings, and perhaps even identify patterns in student language or behavior that could foreshadow school violence.").

[81] Jolicoeur & Mullins, *supra* note 76.

[82] Liebowitz, *supra* note 75 (noting that geofencing technology allows companies to "sweep up posts within a given geographic area and use keywords to narrow the pool."); *see also* Margaret Hu, *Cybersurveillance Intrusions and an Evolving* Katz *Privacy Test*, 55 AM. CRIM. L. REV. 127, 133 (2018) ("Geofencing builds a 'virtual fence' around a designated physical location and permits social media posts from that defined area to be identified and stored.").

[83] *Id.* at 133–34.

remain (fortunately) very rare events.[84] Assuming for the sake of argument that this kind of surveillance was successful at reducing school shootings, how would we know? It would be difficult to state with any confidence that online surveillance had definitively stopped any one particular school shooting, as it would be hard to authoritatively state that such a shooting absolutely would have occurred but for the intervention. A New York Times investigation concluded that there was "little evidence" that social-media monitoring of students has "helped ferret out brewing threats of violence, bullying or self-harm . . . ."[85]

"False negatives," wherein the software would fail to alert school officials to a student who is planning to engage in violent acts, is not the only way the software could fail. There is also great potential for "false positives"—students who might be flagged for some reason and subject to intervention but who would never have engaged in violence. These are students who might fit the stereotypical (and inaccurate) "profile" of a school shooter that has taken root in the public imagination despite there being no empirical evidence of any such particular profile.[86] They might be students who are "fascinated with guns, violent video games and dark song lyrics—but would never turn violent."[87] It cannot be overstated that false positives could be potentially devastating for the students who are inaccurately identified as being at risk for violence. Despite the instinct by some to downplay the potential impact,[88] the student who is falsely labeled at risk faces stigma as well as potential disciplinary actions.[89] One scholar has concluded that schools engaging in professional surveillance of their students are not only engaging in actions that are "ineffective," but ones

---

[84] *See, e.g.*, Daniel P. Mears et al., *Columbine Revisited: Myths and Realities About the Bullying-School Shootings Connection*, 12 VICTIMS & OFFENDERS 939, 942 (2017) (noting that "although school shootings receive widespread media coverage, it is not clear that school shootings have dramatically increased" in the past decades); *see also* Rhitu Chatterjee, *School Shooters: What's Their Path to Violence?*, NPR (Feb. 10, 2019), https://www.npr.org/sections/health-shots/2019/02/10/690372199/school-shooters-whats-their-path-to-violence [https://perma.cc/LCP3-7TYT] (reporting that "there have been [eleven] mass shootings (where four or more people died) in schools since the Columbine High School shooting in . . . 1999").

[85] Liebowitz, *supra* note 75.

[86] Mears et al., *supra* note 84, at 943 (noting that researchers have investigated a number of possible "causal factors" for school shootings, including: "a history of being bullied; mental illness; . . . being a 'loner'; dressing and acting 'Goth'; . . . exposure to violent video games and graphic violence; listening to violent music; . . . and an interest in weapons," but concluding that "none of these factors, or any others, have been shown to exert an effect on the probability of school shootings or of individuals becoming school shooters . . . .").

[87] Jolicoeur & Mullins, *supra* note 76.

[88] *See, e.g.*, Germain Chastel, *Predictive Algorithms Are Infiltrating Schools—Here's Why That's a Good Thing*, NEXT WEB (May 27, 2018), https://thenextweb.com/contributors/2018/05/27/predictive-algorithms-are-infiltrating-schools-heres-why-thats-a-good-thing [https://perma.cc/3WBY-ESCD] ("The worst case scenario [of a false positive] is a child feeling upset at being placed in an intervention program when it's not necessarily needed.").

[89] Mears et al., *supra* note 84, at 949 (the "high false positive rate [for predicting school shooters] would result in many individuals being labeled, and possibly harmed from the labels . . . .").

that are "corrosive to a trusting relationship between students and their schools."[90]

Although the efficacy of programs to monitor students' online activity is questionable at best, the biased impact they have is painfully clear. According to a Southern Poverty Law Center report, there has been a disproportionate impact as students of color are more likely to be suspended or expelled for their social media posts.[91] For example, the Huntsville, Alabama school district began monitoring its students' social-media posts, going so far as to hire a former FBI investigator (it appears this monitoring was more "old-fashioned" and relied on the investigator to flag activity, rather than an algorithm).[92] In 2013–2014, twelve of the fourteen students who were expelled for the content of their social media were African American, despite the fact that African American students make up only 40% of the district.[93] One black student was suspended for five days after a school resource officer assumed the young woman was in a gang, because she posted a picture to her Instagram account of her wearing a sweatshirt that featured an airbrushed image of her father, who had been violently murdered.[94] The school resource officer was suspicious that the colors of her sweatshirt indicated gang involvement.[95]

The racial disparity in the expulsions in Huntsville are not a statistical anomaly. Nationwide, students of color face school discipline that is harsher and more frequent than their white counterparts. A 2018 Government Accountability Office report concluded that black students comprised approximately 15% of all public-school students but represented almost 39% of school suspensions.[96] Despite the fact that activists and scholars have been decrying this racial discrepancy for decades now, Department of Education data show that the disparity is only continuing to grow.[97] The consequences of this disparity could not be more serious. Study after study concludes that students who are expelled or suspended are much more likely to end up incarcerated as adults,

---

[90] Mendola, *supra* note 79, at 158.

[91] Sharada Jambulapati, *Story from the Field: Children of Color Pushed Out of Alabama Schools Over Social Media Posts*, S. POVERTY L. CTR. (July 9, 2015), https://www.splcenter.org/news/2015/07/09/story-field-children-color-pushed-out-alabama-schools-over-social-media-posts-0 [https://perma.cc/8Y2V-DFYB].

[92] *Id.*

[93] *Id.*

[94] *Id.*

[95] *Id.*

[96] U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-18-258, K-12 EDUCATION: DISCIPLINE DISPARITIES FOR BLACK STUDENTS, BOYS, AND STUDENTS WITH DISABILITIES 12–13 (2018).

[97] Moriah Balingit, *Racial Disparities in School Discipline Are Growing, Federal Data Show*, WASH. POST (Apr. 24, 2018), https://www.washingtonpost.com/local/education/racial-disparities-in-school-discipline-are-growing-federal-data-shows/2018/04/24/67b5d2b8-47e4-11e8-827e-190efaf1f1ee_story.html [https://perma.cc/D26D-4NK3] ("Black students faced greater rates of suspension, expulsion and arrest than their white classmates, according to federal data released [in April 2018], disparities that have widened despite efforts to fix them.").

something at times referred to as the "school-to-prison pipeline."[98] Therefore the frightening possibility emerges that the same technology that targets a student for discipline as a high schooler can then be used to label him as "high risk" by his local police department and ultimately tell a judge that he is more likely to reoffend when he is being sentenced.

D. The Myth of Objectivity and Neutrality: How Bias Can Exist in Each Stage of Machine Learning

It is tempting to think of any artificial intelligence, including an algorithm, as neutral and objective.[99] Laypeople without technical expertise can be especially vulnerable to placing too much faith in algorithmic outcomes. "Computers have an aura of reliability that may be unwarranted, but nevertheless hard to dispel. This is because many lawyers and judges do not adequately understand information technology."[100] If an algorithm is developed to monitor students' social media accounts for language that is violent or that threatens self-harm, it is tempting to conclude that the algorithm itself cannot possibly be biased, as it is simply a computer code. As one robotics researcher said in a TEDx Talk, "in [artificial intelligence], we have Milgram's ultimate authority figure," and many laypeople are tempted to place blind faith in it.[101] But, "algorithms are not infallible oracles,"[102] and, as artificial intelligence researchers themselves

---

[98] Amy B. Cyphert, *Addressing Racial Disparities in Preschool Suspension and Expulsion Rates*, 82 TENN. L. REV. 893, 902–03 (2015).

> Zero tolerance policies contribute negatively to what has been termed the 'school to prison pipeline,' wherein disciplinary policies and practices 'push our nation's schoolchildren, especially our most at-risk children, out of classrooms and into the juvenile and criminal justice systems.' Children who are suspended or expelled may be left unsupervised and therefore more likely to get into legal trouble, and they certainly miss critical time in their classes, raising their risk for dropping out.

Id. (citing *Locating the School-to-Prison Pipeline,* ACLU, http.www.aclu.org/images/asset_upload_file966_35553.pdf).

[99] Eidelman, *supra* note 19, at 924 (noting that "mechanized data analysis . . . offers an additional sheen of objectivity, neutrality, and complexity.").

[100] García, *supra* note 43, at 1049–50; *see also* Eidelman, *supra* note 19, at 923 (noting that, "[n]otwithstanding the complexity of computerized algorithms, when their results are introduced in court, legal experts and prosecutors generally suggest that they are infallible and that their results are foolproof, 'overstat[ing] the probative value of their evidence, going far beyond what the relevant science can justify.' And juries, frequently deprived of the source code or any countervailing testimony that could expose the algorithm's potential pitfalls, generally do not question the prosecution's results.").

[101] Tedx Talks, *The Real Reason to Be Afraid of Artificial Intelligence*, YOUTUBE (Dec. 15, 2017), https://www.youtube.com/watch?v=TRzBk_KuIaM [https://perma.cc/3KTJ-G3A5] (referring to social psychologist Stanley Milgram's famous authority experiments where participants believed they were providing painful electrical shocks to other people and kept doing so because an authority figure told them to).

[102] Eidelman, *supra* note 19, at 923.

concede, "algorithms (and the complex systems they are a part of) can make mistakes" and those mistakes can often involve bias.[103]

Arguments that outputs resulting from machine learning are somehow devoid of bias harken back to those that critical race theorists had to confront regarding the objectivity and neutrality of law.[104] "Critical race theory expresses skepticism toward dominant legal claims of neutrality,"[105] and "[p]ioneering theorists . . . lead the charge to expose the structural effects of racism embedded in the law and to rebut the notion that the law is neutral and color-blind."[106] These algorithms do not appear as oracles in the sky, however. Rather, they are the products of humans, with our imperfect biases, and they often capture and magnify those biases.[107] As shown below, there is potential for bias at every step of the process of using an algorithm to monitor the online activity of students,[108] and if programmers and data scientists are "not careful, the process can result in disproportionately adverse outcomes concentrated within historically disadvantaged groups in ways that look a lot like discrimination."[109]

1. Specifying the Outputs

In an early stage of machine learning, the programmers ask, "What do we wish to accomplish?" and determine the results they want their model to produce. Put another way, "programmers must specify an algorithm's output vari-

---

[103] Rachel Thomas, *Five Things That Scare Me About AI*, FAST AI (Jan. 29, 2019), https://www.fast.ai/2019/01/29/five-scary-things [https://perma.cc/7G9P-M7VW].

[104] *See, e.g.*, Mari J. Matsuda, *Looking to the Bottom: Critical Legal Studies and Reparations*, 22 HARV. CIV. RTS.-CIV. LIBERTIES L. REV. 323, 323 (1987) (noting that "black people" understand that any "claim to neutral application of legal principles is false.").

[105] MARI J. MATSUDA ET AL., WORDS THAT WOUND: CRITICAL RACE THEORY, ASSAULTIVE SPEECH, AND THE FIRST AMENDMENT 6 (1993).

[106] Alex M. Johnson, Jr., *What the Tea Party Movement Means for Contemporary Race Relations: A Historical and Contextual Analysis*, 7 GEO. J.L. & MOD. CRITICAL RACE PERSP. 201, 240–41 (2015).

[107] Lehr & Ohm, *supra* note 15, at 717 ("From the moment these humans conceptualize a predictive task to the moment the running model is deployed, they exert significant and articulable influence over everything from how the data are cleaned to how simple or complex the algorithm's learning process is.").

[108] As noted in *supra* note 15, Lehr & Ohm suggest that there are eight discrete steps in the machine learning process. Lehr & Ohm, *supra* note 15, at 670. Their excellent article discusses an earlier article by Barocas & Selbst, *supra* note 24, which delineates three steps in machine learning and the possibility for bias therein. Lehr & Ohm, *supra* note 15, at 670; *see also* Barocas & Selbst, *supra* note 24, at 678, 684, 688. Lehr and Ohm argue that Barocas and Selbst's article "fails to consider all of the stages of machine learning between input variable selection and the deployment of the running model," and that in this failure "they neglect how the stages of machine learning that occur 'inside' the black box can provide opportunities to remedy" certain harms. *Id.* at 666 (emphasis omitted). Both of these articles are excellent and have informed the analysis here. This Article generally examines the three steps in the Barocas and Selbst article but deepens that analysis by selecting additional insights from the Lehr and Ohm article as well. *See infra* Section I.D.

[109] Barocas & Selbst, *supra* note 24, at 673.

able—what is to be estimated or predicted."[110] For an algorithm that has been developed to monitor students' online activity to predict some form of violence—either violence against other students or self-harm—one output variable (also called the dependent variable) is just that—the likelihood that the student will engage in acts of violence. Programmers "must translate some amorphous problem"—here, how to predict school shootings or student suicides—"into a question that can be expressed in more formal terms that computers can parse."[111] There are many decisions that the programmers and data scientists will have to make as they translate the data, and accordingly many opportunities for bias to creep in.

2. Constructing the Training Data Set

In the second step of developing a machine learning system, programmers collect the training data, "quite literally, the data that train the model to behave in a certain way."[112] For an algorithm that attempts to predict school violence, that training data might include looking back over the language of the manifestos that school shooters had posted, as the CEO discussed above acknowledged his programmers did. They might look for certain words to flag or patterns in posting. Of course, "biased training data leads to discriminatory models,"[113] and there are at least two distinct possibilities of bias in this stage of machine learning.

First, the selection of the words that the algorithm will flag as potentially "dangerous" is a fraught endeavor, and one that could lead to a "garbage in, garbage out" problem.[114] The programmers could "rely[] on data that reflect existing human biases,"[115] as in the case of the Boston Police Department's decision to use a predictive policing algorithm that flagged search terms like "Muslim Lives Matter." When machine learning models have been used for the process of "word embedding, a popular framework" where words are converted to word vectors so that the algorithm can identify relationships between words, researchers have found the resulting outputs "exhibit [gender] stereotypes to a

---

110 Lehr & Ohm, *supra* note 15, at 665.

111 Barocas & Selbst, *supra* note 24, at 678.

112 *Id.* at 680; *see also* Eidelman, *supra* note 19, at 926 ("On the machine learning side, humans also impact the algorithm's design by, for example, choosing the training data—another decision that can have significant effects on the algorithm's output and in ways that differentially affect suspects of different races, ethnicities, or ancestral backgrounds.").

113 Barocas & Selbst, *supra* note 24, at 680 (internal citations omitted).

114 *See, e.g.*, Govind Chandrasekhar, *The GIGO Principle in Machine Learning*, SEMANTICS3 BLOG (July 4, 2017), https://www.semantics3.com/blog/thoughts-on-the-gigo-principle-in-machine-learning-4fbd3af43dc4 [https://perma.cc/4XBG-5F9S] ("*Garbage-In-Garbage-Out* is the idea that the output of an algorithm, or any computer function for that matter, is only as good as the quality of the input that it receives.").

115 Lehr & Ohm, *supra* note 15, at 665.

disturbing extent."[116] For example, one algorithm completed the analogy " 'man is to computer programmer as woman is to x' with x=homemaker."[117]

Second, because the data set of school shooters who have left online manifestos is (fortunately) relatively small, the ability to cull enough meaningful data from it to train an algorithm is questionable. School shootings are rare events,[118] and attempting to predict them is a difficult task, with a heightened risk of "false positives." Although machine learning can be used to help predict rare events, in order to improve the accuracy in a rare event prediction scenario, a common method is to obtain as large a data set as possible to train and test the algorithm.[119] Even then accuracy is not, of course, guaranteed, only theoretically improved.[120] "Although there is no technical bar to running machine-learning algorithms on small data sets, doing so is, in practice, pointless," and the smaller the dataset, the less accurate the machine-learning algorithm.[121]

3. Feature Engineering

In the stage of machine learning known as feature selection or feature engineering, programmers "make choices about what attributes they observe and subsequently fold into their analyses."[122] In the feature engineering stage, the goal is to transform the raw data from the training data set into formats that will better facilitate the machine learning process.[123] For a model attempting to pre-

---

[116] Tolga Bolukbasi et al., *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*, CORNELL, 1 (July 21, 2016), https://arxiv.org/abs/1607.06520 [https://perma.cc/63YG-E5QS].

[117] *Id.* at 3.

[118] David Ropeik, *School Shootings Are Extraordinarily Rare. Why is Fear of Them Driving Policy?*, WASH. POST (Mar. 8, 2018), https://www.washingtonpost.com/outlook/school-shootings-are-extraordinarily-rare-why-is-fear-of-them-driving-policy/2018/03/08/f4ead9f2-2247-11e8-94da-ebf9d112159c_story.html [https://perma.cc/VPC6-ZJNV] ("[T]he statistical likelihood of any given public school student being killed by a gun, in school, on any given day since 1999 was roughly 1 in 614,000,000. And since the 1990s, shootings at schools have been getting less common.").

[119] *See* Gary King & Langche Zeng, *Logistic Regression in Rare Events Data*, 9 POL. ANALYSIS 137, 137 (2001) (noting that "commonly used data collection strategies are grossly inefficient for rare events data. The fear of collecting data with too few events has led to data collections with huge numbers of observations but relatively few, and poorly measured, explanatory variables").

[120] Lehr & Ohm, *supra* note 15, at 678–79, 687 ("Often, machine learning is applied to predict exactly these kinds of rare events, but evenly splitting a dataset into training and test data could risk few of these observations ending up in the training data.").

[121] *Id.* at 678 ("To reap the predictive benefits of machine learning, a sufficiently large number of observations is required.").

[122] Barocas & Selbst, *supra* note 24, at 688.

[123] *See, e.g.*, Warren E. Agin, *Using Machine Learning to Predict Success or Failure in Chapter 13 Bankruptcy Cases*, *in* 2018 NORTON ANN. SURV. OF BANKR. L. 369, 390 (William L. Norton & Richard Lieb eds., 2018) (noting that feature engineering involves "[t]ransform[ing] some of the data . . . to create new data fields (or features) that might better suit the machine learning program's processes.").

dict school violence from social-media posts, the feature engineering stage might involve data scientists augmenting the raw data of words in school shooter manifestos by identifying synonyms for certain violent words. For example, if the manifestos frequently include the word "die," data scientists may choose to also include terms like "death" or "dead" or "dying." That same data scientist may choose to exclude the term "killer," recognizing that it is often used by teenagers slangily as an adjective ("that was a killer party!"). Obviously, there is an art to decisions like these, and the possibility for bias with these kinds of decisions.

4. Training the Model

In the next stage of machine learning, the transformed data is used to train the model. The goal of the model at this point is to figure out what features within the data training set are predictive and which are statistical white noise.[124] In our example, the model would run sophisticated statistical algorithms to answer questions such as whether the time of day a social media post was made is significant with respect to predicting our output data, school violence. If the model determines that feature is highly predictive of school violence, it will update itself accordingly by placing higher weight on it in its subsequent calculations. This process may repeat itself millions of times as the model focuses on hundreds or even thousands of features. The model might ultimately conclude that features that seem innocuous: tweeting in the early morning hours, quoting lyrics from certain artists, etc., are statistically correlated with becoming violent. It is critical to note that, depending on the technology used to create it, the model would not necessarily be able to communicate to the programmers—or the assistant principal who receives a notification that a post was flagged—what features or feature combinations caused it to flag the post.

5. Testing and Validating the Model

At this point, after training is complete, the model will be tested on a test set of data. The test set of data for our example would include both innocuous social media posts from students who did not go on to commit any violence, as well as posts from students who did. How accurately does the model flag language and predict that its author will go on to commit violence? Through this validation process, a data scientist can determine the model's theoretical error

[124] *See* Hyunjong Ryan Jin, *Think Big! The Need for Patent Rights in the Era of Big Data and Machine Learning*, 7 N.Y.U. J. INTELL. PROP. & ENT. L. 78, 91–92 (2018) ("The objective of machine learning models is to identify and quantify 'features' from a given data set . . . . Through this process, the machine learning algorithm selects the model that best describes the characteristics and trends of the target features from the test and validation sets.").

rate.[125] How often did it correctly predict that a post in the data test set was made by a student who was about to become violent? Of course, the theoretical error rate is just that: performance on a test set is not a guarantee of similar performance in real life on actual new data.

6. Interpreting the Outputs

Once the training, testing, and validating is complete, the algorithm will be ready to produce outputs. For an algorithm designed to predict the likelihood of future violence from a student's social-media posts, that output would presumably send some sort of alert to school officials warning them that a post has been flagged. Bias is possible at this stage as well, given that "at the output stage, people interpret the algorithm's results and translate them into terms that others can understand."[126] Someone will have to decide what threshold is worthy of a flag being sent. If the algorithm is a probabilistic model and determines that a post is 34% likely to indicate that the student is contemplating violence, is that a high enough percentage? What about 18%? Ultimately, an algorithm will not discipline a high school student for a tweet. Some school official will have to review the tweet once it has been flagged and make a decision about whether to discipline the student or not. In light of the well-documented racial disparities present in school discipline decisions, there is good reason to worry about bias.

7. A Black Box?

At the end of the machine learning process, we have a machine learning model. That model has been developed through data inputs, data that was refined through feature engineering. The model has performed statistical analyses—machine learning algorithms—on this data in an attempt to predict an output. After testing it on our data test set, we have an error rate, one that, if well-designed, should theoretically tell us how the model will perform in the real world. What we usually do *not* have is an easy, human-understandable explanation of how a given output was achieved by the model based upon a given input. We do *not* have an easy explanation of exactly how the model weighs every feature, complicated by the fact that the model may weigh countless combinations of features, and may weigh feature combinations differently with each prediction it makes. Hence, the popular narrative that machine learning models are "black boxes" where it is impossible to know why they make the prediction/produce the output they do.[127] Despite the persistence of this narra-

[125] *See id.* at 92 ("The test set is then used to calculate the generalized prediction error, which is reported to the end user for proper assessment of the predictive power of the model.").

[126] Eidelman, *supra* note 19, at 927.

[127] *See, e.g.*, H. James Wilson, et al., *The Future of AI Will Be About Less Data, Not More*, HARV. BUS. REV. (Jan. 14, 2019), https://hbr.org/2019/01/the-future-of-ai-will-be-about-less-

tive, as is shown above, there are several stages in the machine learning process where humans are making decisions and where safeguards can help control against bias.[128] Part III of this Article includes several policy recommendations for doing just that.

II. POTENTIAL LEGAL CHALLENGES

As journalists continue to write about the online surveillance many schools are conducting of their students, the issue is turning into a "nationwide controversy."[129] Many scholars are beginning their analysis with the question, "is this even legal?" According to the director of legal advocacy for the National School Boards Association, "[s]chool lawyers are advising administrators to be 'very cautious,'" when it comes to online monitoring of their students.[130] The question of legality is made more complicated because the Supreme Court has declined to rule on cases that deal with online student speech, leaving the question of what kind of monitoring and discipline is appropriate to the lower courts for now.[131] "Given that the Supreme Court has not ruled on off-campus Internet speech, the legality of student Internet surveillance carried out by schools remains uncertain."[132] Of course, to the extent that this Article examines constitutional challenges to schools' online monitoring of their students, such a challenge is only available to students at public schools (which are state actors) or to students in public and private schools in states that have provided a state law corollary to the constitutional right at issue.[133]

---

data-not-more [https://perma.cc/AVU7-ZETH] ("[T]hese [machine learning] systems are black boxes—it's not clear how they use input data to arrive at outputs like actions or decisions.").

[128] *See, e.g.*, Lehr and Ohm, *supra* note 15, at 657 ("As many have documented, a running model is often viewed as an inscrutable black box, but there are opportunities for auditing (record-keeping requirements, keystroke loggers, etc.) and mandated interpretability during playing with the data . . . . Regulation skeptics . . . often rely on descriptions of machine learning as more art than science," which runs the risk of "inappropriately assum[ing] that black-box algorithms have black-box workflows; [despite the fact that] the steps of playing with the data are actually quite articulable.").

[129] *See* Maiya Dempsey, *Easy to Say, Easy to See: Social Media and the Constitutional Rights of Public School Students*, 17 WHITTIER J. CHILD & FAM. ADVOC. 82, 90–91 (2018) ("As technology has developed and students have growing access to the Internet and social media, the issue of whether a school can monitor its students' social media accounts has become a nationwide controversy.").

[130] Liebowitz, *supra* note 75.

[131] *See* Mendola, *supra* note 79, at 157 ("In January 2012, the Court denied certiorari for three cases involving student Internet speech, any of which could have established necessary guidelines for school action in response to student Internet speech.").

[132] *Id.* at 168.

[133] *See* Michael K. Park, *Restricting Anonymous "Yik Yak": The Constitutionality of Regulating Students' Off-Campus Online Speech in the Age of Social Media*, 52 WILLAMETTE L. REV. 405, 414–15 (2016) (explaining that the First Amendment applies only to government action, not to private action, and so only to public schools, but also noting that "a few states

A. Cyberbullying Laws

In order to assess the legality of the use of machine-learning algorithms to monitor high school students, it is important to understand the legal landscape of cyberbullying laws that school districts operate under. "Cyberbullying laws, mostly passed as part of states' education codes, prohibit cyberbullying, or bullying by electronic means, and provide schools with the authority to discipline students for it."[134] These laws exist in some form in every state,[135] though only a handful of states have laws that explicitly define *cyber*bullying as something distinct from traditional bullying that happens in an online forum.[136] The laws vary from state to state, but in general they prohibit students from taking actions that put other students in reasonable fear of harm, or otherwise harassing fellow students.[137] One scholar has categorized the degree of surveillance authority that the laws grant to schools in three ways:

> (1) a grant of authority with no nexus to school or school-related activity; (2) a grant of authority with a limited nexus to school or school-related activity; and (3) a grant requiring a relatively substantial nexus to school or school-related activity. The vast majority of cyberbullying laws provide schools with surveillance authority that falls into one of the first two categories. In those states, the schools have nearly unlimited or unlimited surveillance authority over students' online and electronic activity.[138]

Although none of the cyberbullying laws *explicitly* direct schools to monitor students' social media or other online activity, the laws *implicitly* do so.[139] Further, fourteen states extend the scope of school authorities to regulate off-campus student speech[140] (a decision with consequences under the First and Fourth Amendments' precedents discussed below). Although schools can conduct this surveillance themselves, as noted above, many are increasingly turn-

. . . have tried to provide students at private educational institutions with free speech protections parallel to those in the First Amendment.").

[134] Emily F. Suski, *Beyond the Schoolhouse Gates: The Unprecedented Expansion of School Surveillance Authority Under Cyberbullying Laws*, 65 CASE W. RES. L. REV. 63, 65 (2014).

[135] Elizabeth A. Shaver, *Denying Certiorari in* Bell v. Itawamba County School Board*: A Missed Opportunity to Clarify Students' First Amendment Rights in the Digital Age*, 82 BROOK. L. REV. 1539, 1590 n.404 (2017) ("In 2015, Montana became the last state in the nation to enact anti-bullying legislation.") (internal citations omitted).

[136] Suski, *supra* note 134, at 73 (discussing in 2014 that twenty-five states had laws that were "simply additions to or variations of . . . general definitions of bullying," but that sixteen states had "separate statutory or regulatory definitions of cyberbullying.").

[137] *Id.* at 71.

[138] *Id.* at 70–71.

[139] *Id.* at 74 ("[A]ll of the states with cyberbullying laws authorize schools to monitor students' online and electronic activity. None, however, do so explicitly. Instead, they implicitly allow schools to engage in surveillance of students' online and electronic activity by authorizing or requiring that schools discipline students for electronic acts that constitute bullying.").

[140] Philip Lee, *Expanding the Schoolhouse Gate: Public Schools (K-12) and the Regulation of Cyberbullying*, 2016 UTAH L. REV. 831, 848 (2016).

ing to software companies to comply with their monitoring obligations under cyberbullying laws.[141]

B. First Amendment Challenges

There are two different legal challenges that could be made under the First Amendment to the online monitoring of students. First, students who are actually disciplined by their public school for social media posts they made outside of school property and outside of school hours may have a viable First Amendment claim against their school. As Section II.B.1 below explains, the success of any such claim will be more dependent on the content of the speech than on the fact that it was made off campus. Second, the very practice of monitoring students' online speech—whether with machine learning or through more "old-fashioned" methods—could be challenged as an unlawful prior restraint, even where schools are not disciplining students on the basis of what they post.[142] This strategy has been examined by scholars but is presently hypothetical, as no court has ever ruled on the issue.[143]

1. Challenge by Disciplined Students

As a threshold matter, the legality of free speech regulation often turns on a "forum analysis," wherein the level of restriction that the government can impose on speech depends on the location of the speaker and where—the more traditionally public the space is, the less the government may regulate speech there.[144] For example—for a traditionally public forum, like a street corner, content-based restriction of individual speech violates the First Amendment "unless the government can pass strict scrutiny, showing the restriction is narrowly tailored to further a compelling government interest."[145] Conversely, for a traditionally "nonpublic" forum, like a jail, the government need only establish that a content-based restriction is reasonable.[146]

---

[141] Suski, *supra* note 134, at 63, 76–77 ("As explained below, most of the states have this implicit authority, and some schools are therefore starting to employ companies, such as Geo Listening and Safe Outlook Corporation, to conduct comprehensive surveillance of students.").

[142] *See infra* Section II.B.2.

[143] *See infra* Section II.B.2.

[144] *See, e.g.*, Perry Educ. Ass'n v. Perry Local Educs' Ass'n, 460 U.S. 37, 45–46 (1983) ("In places which by long tradition or by government fiat have been devoted to assembly and debate, the rights of the State to limit expressive activity are sharply circumscribed . . . A second category consists of public property which the State has opened for use by the public as a place for expressive activity . . . . Public property which is not by tradition or designation a forum for public communication is governed by different standards."); *see also* Nisha Chandran, *Crossing the Line: When Cyberbullying Prevention Operates as a Prior Restraint on Student Speech*, 2016 U. ILL. J.L. TECH. & POL'Y 277, 290 (2016).

[145] *Id.*; *see also* Perry Educ. Ass'n, 460 U.S. at 45–46.

[146] Chandran, *supra* note 144, at 290.

In the school setting, the Supreme Court has held that schools have more latitude to regulate student speech than the government does to regulate the speech of adults,[147] especially if that speech is lewd,[148] encourages illegal activity like drug use,[149] or is somehow school-sponsored, such as in the case of a school newspaper.[150] The Supreme Court has also held that schools can sometimes regulate student speech made outside of school grounds where such speech occurs during a school-sponsored activity or during school hours.[151] However, the Supreme Court has not yet ruled on what latitude schools have in regulating student speech that is entirely off campus and not during school hours or school activities.[152]

In the seminal school-free-speech case of *Tinker v. Des Moines Independent Community School District*,[153] the Supreme Court famously noted that students do not "shed their constitutional rights to freedom of speech or expression at the schoolhouse gates."[154] The students in *Tinker* were suspended from their high school when they wore black armbands to protest the Vietnam War.[155] The Court noted that such speech "[did] not concern speech or action that intrude[d] upon the work of the schools or the rights of other students,"[156] and that the school's "undifferentiated fear or apprehension" that the students might cause a disturbance was not enough to justify the infringement on their First Amendment rights.[157] Although the conduct at issue occurred during school, the Court cautioned that "conduct by the student, in class *or out of it*, which for any reason . . . materially disrupts classwork or involves substantial disorder or invasion of the rights of others is, of course, not immunized . . . ."[158] Thus, the *Tinker* Court made clear that schools could regulate student speech made with-

147 *See, e.g.*, Bethel Sch. Dist. No. 403 v. Fraser, 478 U.S. 675, 682 (1986) ("[T]he constitutional rights of students in public school are not automatically coextensive with the rights of adults in other settings.").

148 *Id.* at 677–78, 685 (finding no First Amendment violation when a school disciplined a student who made a speech at a school assembly where he referred to another student "in terms of an elaborate, graphic, and explicit sexual metaphor.").

149 Morse v. Frederick, 551 U.S. 393, 397, 403 (2007) (finding no First Amendment violation when a school disciplined students for unfurling a banner reading "BONG HiTS 4 JESUS" at an off-campus, school-sanctioned, school-supervised event).

150 Hazelwood Sch. Dist. v. Kuhlmeier, 484 U.S. 260, 273, 276 (1988) (finding no First Amendment violation when a school stopped student journalists from publishing an article in the student newspaper).

151 *Morse*, 551 U.S. at 401.

152 *See, e.g.*, Mendola, *supra* note 79, at 156–57.

153 Tinker v. Des Moines Indep. Cmty. Sch. Dist., 393 U.S. 503 (1969).

154 *Id.* at 506.

155 *Id.* at 504.

156 *Id.* at 508.

157 *Id.*

158 *Id.* at 513 (emphasis added).

in the school or even outside it where it disrupted the school's work or invaded the rights of others, but not otherwise.[159]

Tinker itself dealt with student speech made *within* the school grounds *during* the school day. Of course, "[a] school's action in response to its students' Internet postings has several characteristics that make it distinct from a strict free speech issue," including that even when students are posting from their own homes, on their own time, and on their own machines, students can still sometimes post *about* school or can be communicating with their classmates.[160] So what about a student who is tweeting or posting to Instagram outside of school property and outside of school hours? What level of regulation, if any, may a school constitutionally exercise over such student speech? "The federal appellate courts have . . . been left to grapple with applying the *Tinker* standard to address discipline of students' online speech when it does not occur in the school building or at a school-related or sponsored event."[161] So far, seven circuits have examined the question of whether *Tinker* applies to student conduct that occurs outside of school grounds with six of them concluding it does and the seventh assuming it does without formally holding so.[162]

In one of the most recent decisions to hold that *Tinker* applies to off-campus, electronic speech by students, the Fifth Circuit examined how the school environment had changed since that holding. In *Bell v. Itawamba County School Board*,[163] the Fifth Circuit, on rehearing en banc, held that the *Tinker* rule applied to student speech made on the internet, "even when such speech originated, and was disseminated, off-campus without the use of school resources."[164] In *Bell*, a high school student recorded a rap video which he posted

---

[159] *Id.* at 512–13.

[160] Mendola, *supra* note 79, at 157–58.

[161] Suski, *supra* note 134, at 89.

[162] *See* Doe v. Valencia Coll., 903 F.3d 1220, 1231 (11th Cir. 2018) (citation omitted) ("We need not decide how far *Tinker's* 'in class or out of it' language extends. It is enough to hold, as we do, that *Tinker* does not foreclose a school from regulating all off-campus conduct."); C.R. v. Eugene Sch. Dist. 4J, 835 F.3d 1142, 1155 (9th Cir. 2016) (finding that a school district has authority to discipline students for off-campus, sexually harassing speech); Bell v. Itawamba Cty. Sch. Bd., 799 F.3d 379, 391 (5th Cir. 2015) (discussed at length below); S.J.W. v. Lee's Summit Sch. Dist., 696 F.3d 771, 773, 778 (8th Cir. 2012) (finding that school could discipline twin students for a website they created, even when made off campus, provided that it was "targeted at" the school community); Kowalski v. Berkeley Cty. Schs., 652 F.3d 565, 573–74 (4th Cir. 2011) (concluding a school district did not violate a student's First Amendment rights by suspending her for creating a website on which other students posted defamatory information about a classmate); Doninger v. Niehoff, 527 F.3d 41, 48, 50 (2d Cir. 2008) ("[A] student may be disciplined for expressive conduct, even conduct occurring off school grounds, when this conduct 'would foreseeably create a risk of substantial disruption within the school environment,' at least when it was similarly foreseeable that the off-campus expression might also reach campus.") (internal citation omitted). The remainder of the circuits (First, Sixth, Seventh, Tenth, D.C.) do not appear to have addressed the question of whether *Tinker* can be applied to off-campus speech by students.

[163] *Bell*, 799 F.3d at 379.

[164] *Id.* at 396.

to the internet, first on his publicly accessible Facebook page and then on YouTube.[165] The rap lyrics, which the Fifth Circuit described as "incredibly profane and vulgar," named two high school coaches, alleged that they were having sexual relationships with students, and made several threatening remarks about harming the teachers with guns.[166] Although there is no indication that the school was monitoring the student's Facebook page, the wife of one of the teachers heard about the recording from a friend and alerted her husband, who alerted the principal.[167] A disciplinary committee ultimately decided that the rap lyrics "constituted harassment and intimidation of two teachers" and recommended that the student be given a seven-day suspension and then "placed in the county's alternative school for the remainder of the nine-week grading period (approximately six weeks)."[168] The student filed an action alleging that the discipline violated his First Amendment rights.[169]

The Fifth Circuit noted *Tinker's* holding that students do not give up their First Amendment rights simply by being students.[170] But the court also noted that these rights are not absolute, and cited to Supreme Court precedent for the proposition that school students enjoy less expansive First Amendment rights than adults, concluding that "certain speech, which would be protected in other settings, might not be afforded First Amendment protection in the school setting."[171] The Fifth Circuit noted that "[t]he pervasive and omnipresent nature of the Internet has obfuscated the on-campus/off-campus distinction" with respect to school regulation of student speech.[172] The court held that *Tinker* applied to the conduct at issue and also acknowledged that the passage of time, horrors of school shootings, and advances in technology were necessary to keep in mind when applying the *Tinker* test.[173] "Over [forty-five] years ago, when *Tinker* was decided, the Internet, cellphones, smartphones, and digital social media did not exist. The advent of these technologies and their sweeping adoption by students present new and evolving challenges for school administrators, confounding previously delineated boundaries of permissible regulations."[174] The court emphasized in its analysis that schools were duty bound to be vigilant and pay special attention to threats against teachers in light of recent school shootings.[175] "This now-tragically common violence increases the importance of

[165] *Id.* at 383.
[166] *Id.* at 384–85.
[167] *Id.* at 385.
[168] *Id.* at 386.
[169] *Id.* at 387.
[170] *Id.* at 389 (citing to *Tinker*, 393 U.S. 503, 506, 511 and noting that "[s]tudents *qua* students do not forfeit their First Amendment rights to freedom of speech and expression.").
[171] *Id.* at 389–90 (internal citation omitted).
[172] *Id.* at 395–96.
[173] *Id.* at 392–93.
[174] *Id.* at 392.
[175] *Id.* at 392–93 (citations omitted) ("Greatly affecting this landscape is the recent rise in incidents of violence against school communities. School administrators must be vigilant and

clarifying the school's authority to react to potential threats before violence erupts."[176] The Fifth Circuit stopped short of defining the circumstances under which off-campus speech may be restricted, holding that where, as there, a student directs the speech at the school community, and that speech includes harassment and intimidation of teachers, the school may regulate the speech.[177]

The only circuit that has addressed the issue and declined to hold that schools may, consistent with *Tinker*, regulate off-campus speech is the Third Circuit, and though it has spoken on the matter through a pair of twin opinions, its position remains unsettled. On June 13, 2011, the Third Circuit published two opinions that dealt with regulation of off-campus speech by students. In *J.S. v. Blue Mountain School District*, the Third Circuit found that a school district had violated the First Amendment rights of a student who was disciplined for creating, off campus, a fake MySpace profile that included her middle school principal's official school photograph (though not his name).[178] "The profile contained crude content and vulgar language, ranging from nonsense and juvenile humor to profanity and shameful personal attacks aimed at the principal and his family."[179] In its opinion reversing a grant of summary judgment on behalf of the school with respect to the student's First Amendment claims, the Third Circuit noted that the profile was "so outrageous that no one took its content seriously" and that it could not be viewed at the school in any event, as the school's computers blocked MySpace.[180] Because "[t]here [was] no dispute that [the student's] speech did not cause a substantial disruption in the school,"[181] and because "it was clearly not reasonably foreseeable that [the student's] speech would create a substantial disruption or material interference in school,"[182] the court held that the school's actions in disciplining the student violated her First Amendment rights.[183] In so holding, the Third Circuit noted that "[t]he Supreme Court [has] established a basic framework for assessing student free speech claims in *Tinker*, and we will assume, without deciding, that *Tinker* applies to [the student's off-campus] speech in this case."[184]

---

take seriously any statements by students resembling threats of violence, as well as harassment and intimidation posted online and made away from campus.").

176 *Id.* at 393.

177 *Id.* at 394 ("Therefore, the next question is under what circumstances may off-campus speech be restricted. Our court's precedent is less developed in this regard. For the reasons that follow, and in the light of the summary-judgment record, we need not adopt a specific rule: rather, Bell's admittedly intentionally directing at the school community his rap recording containing threats to, and harassment and intimidation of, two teachers permits *Tinker*'s application in this instance.").

178 J.S. v. Blue Mountain Sch. Dist., 650 F.3d 915, 920 (3d Cir. 2011).

179 *Id.*

180 *Id.* at 920–21.

181 *Id.* at 928.

182 *Id.* at 930.

183 *Id.* at 920.

184 *Id.* at 926.

Five of the *J.S.* judges joined a concurrence arguing that "the First Amendment protects students engaging in off-campus speech to the same extent it protects speech by citizens in the community at large."[185] Those judges objected to the extension of *Tinker* to online speech made by students off of school grounds, arguing that the Supreme Court's subsequent school-speech cases "underscored *Tinker*'s narrow reach."[186] According to the concurring judges, "[a]pplying *Tinker* to off-campus speech would create a precedent with ominous implications. Doing so would empower schools to regulate students' expressive activity no matter where it takes place, when it occurs, or what subject matter it involves—so long as it causes a substantial disruption at school."[187]

Of course, a concurrence is not a holding, even when, as with *J.S.*, it is signed by a plurality of the judges who hear a case. To underscore that the Third Circuit was internally conflicted on this issue, on the same day it issued its opinion in *J.S.*, the Third Circuit also issued an opinion in *Layshock v. Hermitage School District*.[188] That case rather remarkably also dealt with a student who created a fake and embarrassing MySpace profile of his principal.[189] The Third Circuit affirmed the district court's grant of summary judgment to the student regarding his First Amendment claim, holding that the speech at issue did not result in any substantial disruption.[190] The court noted that:

> [B]ecause the School District concedes that [the fake] profile did not cause disruption in the school, we do not think that the First Amendment can tolerate the School District stretching its authority into [the student's] grandmother's home and reaching [the student] while he is sitting at her computer after school in order to punish him for the expressive conduct that he engaged in there.[191]

The court recognized that "*Tinker*'s 'schoolhouse gate' is not constructed solely of the bricks and mortar surrounding the school yard," but nonetheless concluded that the ability of schools to regulate student speech is limited.[192] "It would be an unseemly and dangerous precedent to allow the state, in the guise of school authorities, to reach into a child's home and control his/her actions there to the same extent that it can control that child when he/she participates in school sponsored activities."[193]

Thus, in all jurisdictions that have formally ruled on the question, student speech that occurs off campus, including social-media posts or other online speech, can still be regulated by school officials, and students can be disci-

---

185 *Id.* at 936 (Smith, J., concurring).
186 *Id.* at 936–38 (Smith, J., concurring).
187 *Id.* at 939 (Smith, J., concurring).
188 Layshock v. Hermitage Sch. Dist., 650 F.3d 205 (3d Cir. 2011) (en banc).
189 *Id.* at 207–08.
190 *Id.* at 219.
191 *Id.* at 216.
192 *Id.*
193 *Id.*

plined as if the speech was made on campus, provided the speech in some way reaches and disrupts the school community.[194] It is worth noting that where courts have generally given the green light to schools regulating student speech off campus, several state legislatures have taken a different approach. Oregon, for example, explicitly defines "harassment, intimidation, or bullying" in its anti-bullying law as an act that "[t]akes place on or immediately adjacent to school grounds, at any school-sponsored activity, on school-provided transportation or at any official school bus stop . . ."[195] It is also worth noting that while many scholars argue that schools should not regulate off-campus student speech,[196] others disagree, arguing that cyberbullying is particularly harmful and that schools are in the best position to protect their students from it.[197]

2. Prior Restraint Challenge to Online Monitoring

As a general rule, a "prior restraint" is something that "restricts speech in advance on the basis of content and carries a presumption of unconstitutionality."[198] Under this doctrine, the government has less latitude to regulate speech *before* it is made, even if that same speech can be lawfully regulated (and punished) once it is disseminated.[199] "This preference is rooted in a foundational tenet of U.S. law as it departed from English rule: a free society prefers to punish those who abuse rights of speech *after* they break the law, rather than to suppress them and all others beforehand."[200] Prior restraints most often take one of two forms: judicial injunctions (such as a court order forbidding the me-

---

[194] *See* Park, *supra* note 133, at 439–43 ("The latest circuit court precedent . . . reveals that school officials are given significantly broad authority to regulate online student speech that is violent in character and threatens the safety of students and the school. . . . [L]ower courts have given greater deference to the judgment of school officials with regards to speech limitations . . . .").

[195] OR. REV. STAT. § 339.351(2)(b) (2009); *see also* Lee, *supra* note 140, at 849–50 (noting that Alabama, Oregon, and South Carolina all focus on on-campus speech in their cyberbullying laws).

[196] *See, e.g.*, Lee Goldman, *Student Speech and the First Amendment: A Comprehensive Approach*, 63 FLA. L. REV. 395, 430 (2011) ("When student speech occurs outside of school supervision, the speech should receive the same First Amendment protection as a non-student's speech."); Suski, *supra* note 107, at 64 ("To protect students from excessive school surveillance authority and attendant privacy harms, realistic limits need to be imposed on school surveillance authority under the cyberbullying laws . . . .").

[197] Lee, *supra* note 140, at 858, 864 ("[T]here are three reasons that actual cases of cyberbullying require a different analysis: (1) the nature of the harm is unique; (2) other legal remedies are inadequate to protect victims; and (3) schools are in the best position to protect their students.").

[198] Taylor v. Roswell Indep. Sch. Dist., 713 F.3d 25, 42 (10th Cir. 2013) (citations omitted).

[199] *See* Chandran, *supra* note 144, at 291 ("Under the prior restraint doctrine, the government is limited in its ability to restrain protected expression before it is disseminated, even though the same expression could be constitutionally subjected to punishment after the fact through civil and criminal liability.").

[200] *Id.*

dia from publishing certain information) or administrative licensing schemes (such as an ordinance requiring a parade license).[201]

Thus, it is no surprise that when it comes to student speech, the prior restraint doctrine has been invoked most often in regulation of student newspapers or the distribution of certain materials on campus.[202] To date, no court has addressed the application of the prior restraint doctrine to student speech that occurs *off* campus.[203] However, in a fascinating article, scholar Nisha Chandran has examined the issue of whether online monitoring of student speech constitutes an unlawful prior restraint on their First Amendment rights.[204] She concludes that online monitoring of students should be analyzed under a heightened scrutiny standard, as the "underlying policy concerns rendering prior restraints presumptively unconstitutional directly parallel First Amendment concerns with proactive cyberbullying prevention policies . . . ."[205]

Ms. Chandran notes that all of the cases that have examined student speech made off campus to date, including those discussed in Section II.B.1 above, have dealt with that speech *after* it is made, which is a reactive form of regulating speech.[206] Social-media monitoring, by contrast, is proactive; it "diverge[s] from the traditional pattern of punishment following a known student speech violation and transform[s] it into a restraint on expression before dissemination through monitoring surveillance or speech guidelines for private, off-campus speech."[207] Ms. Chandran argues that because proactive online monitoring by schools of their students is not the exclusive remedy to combat the harm of cyberbullying, and that it "strip[s] the speaker of procedural protections characteristic of reactive litigation," such monitoring amounts to forbidden censorship on speech and should be held presumptively unconstitutional.[208] The fact that this monitoring is likely to be fueled by machine learning makes it even more constitutionally suspicious. "Knowing that surveillance technology is often based on computerized algorithms 'triggered' by buzzwords, students may also choose to completely avoid speech on certain topics to avoid discipline even though their speech would have been entirely innocuous."[209] Children and teenagers may be especially vulnerable to this chilling effect,[210] and of course it

---

[201] *Taylor*, 713 F.3d at 42.

[202] Chandran, *supra* note 144, at 294; *see also Taylor*, 713 F.3d at 34 (rejecting argument that denying a student the ability to distribute rubber fetus dolls on campus was an unlawful prior restraint).

[203] Chandran, *supra* note 144, at 296 ("[N]o case to date has addressed the application of prior restraint law to off-campus restrictions on student speech.") (emphasis omitted).

[204] *Id.* at 279.

[205] *Id.* at 301.

[206] *Id.*

[207] *Id.* at 301–02 (emphasis omitted).

[208] *Id.* at 302.

[209] *Id.* at 304.

[210] *See A Machine of Paranoia: How Concerns for Student Safety May Chill Speech*, NAT'L COALITION AGAINST CENSORSHIP (Sept. 18, 2014), http://ncac.org/blog/a-machine-of-

will be compounded by the fact that the algorithms may flag seemingly innocuous terms and report that they are statistically correlated with a higher likelihood of violence.

How likely might a court be to rule that a school's social-media monitoring program is an unlawful prior restraint on student speech? If the willingness of federal appellate courts to extend *Tinker* to online speech made by students off campus is any indication, not especially likely. As discussed above, each circuit that has ruled on the issue has held that schools can regulate student online speech when it disrupts the school environment (or even when it reasonably might), and courts have also indicated a willingness to embrace more regulation of student speech in light of school shootings.[211] For those courts, a school district that can argue its online monitoring is an attempt to prevent school shootings is likely to be given more latitude, even if they adopt a heightened scrutiny standard. Further, a court that ruled that online monitoring is unlawful would perhaps have to invalidate as unconstitutional any state cyberbullying law that required schools to monitor their students' online activity in any way. Therefore, despite Ms. Chandran's intriguing arguments, it is unlikely that the Supreme Court will ultimately rule that online monitoring of students operates as an unlawful prior restraint on their speech.

C. Fourth Amendment Challenges

Just as students do not leave their First Amendment rights at the schoolhouse gates, even as schools have power to regulate certain kinds of student speech, so too do students retain Fourth Amendment rights, even as schools have the power to conduct certain kinds of searches. Further, "Fourth and First Amendment[] [violations] may be interconnected: knowledge of unreasonable searches regarding personal communication often chills speech by causing speakers to self-censor."[212]

The Fourth Amendment provides in relevant part that "[t]he right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated . . . ."[213] The seminal case outlining the power of schools to conduct searches of their students is *New Jersey v. T.L.O.*[214] In that case, a high school student alleged that her vice principal's in-school search of her purse, and his subsequent handing over of certain drug paraphernalia found therein to the police, was an unlawful search and sei-

paranoia-how-concerns-for-student-safety-may-chill-speech [https://perma.cc/AB7Q-NYHR] ("[S]urveillance can facilitate an anxious culture of self-censorship. Youth—especially in places of learning—will feel as though their school is watching them constantly, on and off campus, whenever they post.").

[211] *See supra* Section II.B.1.

[212] Chandran, *supra* note 144, at 285.

[213] U.S. CONST. amend. IV.

[214] New Jersey v. T. L. O., 469 U.S. 325 (1985).

zure that violated her Fourth Amendment rights.[215] The Supreme Court rejected the argument that the search was unlawful, holding instead that it was reasonable under the circumstances.[216]

In rejecting the Fourth Amendment argument, the Court made several crucial holdings about the extent to which the Fourth Amendment applies to school students. First, the Court held that the Fourth Amendment's prohibition on unreasonable searches and seizures *does* apply to searches that are conducted by public school officials, an idea that the state of New Jersey had challenged and which courts at the time were divided on.[217] The Court cited to the holdings in *Tinker* and in *Goss v. Lopez*, reasoning that "[i]f school authorities are state actors for purposes of the constitutional guarantees of freedom of expression and due process, it is difficult to understand why they should be deemed to be exercising parental rather than public authority when conducting searches of their students."[218] The Court acknowledged that maintaining discipline can be a difficult task in a school but concluded that the situation "is not so dire" that students should be treated like prisoners and afforded no legitimate expectation of privacy while at school.[219]

Second, the Court held that although students maintain some Fourth Amendment rights while in school, they do so with important limitations. In language that sounds hopelessly quaint in the era of smart phones, the Court noted that school "students may carry on their persons or in purses or wallets such non-disruptive yet highly personal items as photographs, letters, and diaries."[220] Nonetheless, the Court recognized that students are to be afforded somewhat less Fourth Amendment protection than adults, noting that "the school setting requires some easing of the restrictions to which searches by public authorities are ordinarily subject."[221] The Court held that school officials need not seek a warrant before searching students, and that they also need not have "probable cause" for the search.[222] Rather, the Court held that "the legality of a search of a student should depend simply on the reasonableness, under all the circumstances, of the search."[223]

---

[215] *Id.* at 328–29.
[216] *Id.* at 347–48.
[217] *Id.* at 333–34.
[218] *Id.* at 336.
[219] *Id.* at 338–39 ("We are not yet ready to hold that the schools and the prisons need be equated for purposes of the Fourth Amendment.").
[220] *Id.* at 339.
[221] *Id.* at 340.
[222] *Id.* at 341 ("We join the majority of courts that have examined this issue in concluding that the accommodation of the privacy interests of schoolchildren with the substantial need of teachers and administrators for freedom to maintain order in the schools does not require strict adherence to the requirement that searches be based on probable cause to believe that the subject of the search has violated or is violating the law.").
[223] *Id.* at 341.

Third, the Court provided guidance on what a "reasonable" search would look like within the confines of a school. The Court outlined a two-prong inquiry, citing to the seminal Fourth Amendment case *Terry v. Ohio*.[224] First, reasonableness should be assessed by considering "whether the . . . action was justified at its inception,"[225] a standard which is satisfied "when there are reasonable grounds for suspecting that the search [would] turn up evidence that the student has violated or is violating either the law or the rules of the school."[226] The second prong of the analysis asks "whether the search as actually conducted was reasonably related in scope to the circumstances which justified the interference in the first place,"[227] a standard that is met "when the measures adopted are reasonably related to the objectives of the search and not excessively intrusive in light of the age and sex of the student and the nature of the infraction."[228]

The Court returned to students' Fourth Amendment rights in *Vernonia School District v. Acton*,[229] where a middle school football player challenged his school's requirement that athletes consent to undergo random drug tests as a Fourth Amendment violation.[230] In rejecting that argument, the Court held that suspicionless searches (such as a requirement that any athlete undergo drug testing even if there is no particular suspicion about the drug use of that particular athlete) in the context of school students do not necessarily violate the Fourth Amendment.[231] The Court reiterated *T.L.O.*'s holding that the "ultimate measure of the constitutionality of a governmental search is 'reasonableness,'"[232] and noted that such an inquiry is not made in a vacuum and "cannot disregard the schools' custodial and tutelary responsibility for children."[233]

The Supreme Court has not weighed in on whether online student surveillance violates the Fourth Amendment. Indeed, the Court has not weighed in at all on whether *any* search conducted by school officials that occurs outside of school is protected by the Fourth Amendment.[234] But one district court did address the issue in ruling on a motion to dismiss.[235] In *R.S. v. Minnewaska Area*

---

[224] *Id.* (citing Terry v. Ohio, 392 U.S. 1, 20 (1968)).
[225] *Id.*
[226] *Id.* at 342.
[227] *Id.* at 341 (internal quotation marks omitted).
[228] *Id.* at 342.
[229] Vernonia Sch. Dist. 47J v. Acton, 515 U.S. 646 (1995).
[230] *Id.* at 651.
[231] *Id.* at 654.
[232] *Id.* at 652.
[233] *Id.* at 656.
[234] Suski, *supra* note 134, at 95 ("The Supreme Court did not [in *Acton*], and has not subsequently, addressed whether schools have any authority to search students outside the time and space of the physical school setting or any limits thereof.").
[235] R.S. v. Minnewaska Area Sch. Dist. No. 2149, 894 F. Supp. 2d 1128, 1132 (D. Minn. 2012).

School District,[236] a twelve–year–old girl argued that her school violated the Fourth Amendment when it disciplined her for out-of-school Facebook posts she made, including one expressing her dislike of a school employee (her Facebook account was private, not public).[237] The school officials even made the girl "involuntarily surrender" her email and Facebook passwords to them when they learned that she and another student "had an out-of-school sex-related conversation" so that they could review more of her posts.[238] Because the district court was only ruling on the school's motion to dismiss, it accepted as true the allegations the girl made for purposes of that review, and concluded that they "amount[ed] to violations of [the girl's] constitutional rights and that those rights were clearly established at the time of the alleged conduct."[239]

In so holding, the district court first examined whether the student had a reasonable expectation of privacy in her Facebook posts and direct messages.[240] The court placed weight on the fact that although the student's Facebook posts were semi-public, in that they could be viewed by any of her Facebook friends, her direct messages were private and accessible only to her, thus making them more akin to email.[241] Therefore, the court determined that she did have a reasonable expectation of privacy in at least some of the online material that the school had accessed.[242] The court cited to *T.L.O.*, concluding that the school officials had no reasonable grounds to believe that the search they conducted would yield evidence that the student had violated school rules.[243]

Even in the absence of guidance from the Supreme Court or any appellate courts, the holdings of *T.L.O.* and *Acton* offer some insight into how successful such a challenge to online monitoring of students would be. Professor Emily Suski has argued that schools' surveillance of their students' online activity "actively fails" the two-prong test set out in *T.L.O.* for whether a search is reasonable.[244] First, she argues that it fails the prong that the search be "justified at

---

[236] *Id.* at 1128.

[237] *Id.* at 1132–33.

[238] *Id.* at 1133.

[239] *Id.*

[240] *Id.* at 1142.

[241] *Id.* ("[A]t least some of the information and messages accessed by the school officials were in R.S.'s exclusive possession, protected by her Facebook password. R.S. controlled those items until she involuntarily relinquished her password. As with a private letter, the content of R.S.'s electronic correspondence was available only to her and her correspondent.").

[242] *Id.*

[243] *Id.* at 1143 ("It is difficult for the Court to discern what, if any, legitimate interest the school officials had for perusing R.S.'s private communications . . . . Moreover, the school officials had no reason to believe that the search would return evidence of illegal behavior or violations of school policy.").

[244] Suski, *supra* note 134, at 94. Professor Suski acknowledges that the online monitoring may not be considered a "search" for Fourth Amendment purposes. *Id.* at 96 ("Whether students have a reasonable expectation of privacy in many online and electronic communications is at best questionable. Frequently used online tools and services like Google make

its inception" because "broad surveillance authority provided by the cyberbullying laws has no justification other than an undifferentiated understanding that cyberbullying does happen sometimes among some students."[245] Second, she argues that online monitoring is not reasonably related in scope to the justification of preventing cyberbullying because the scope of the monitoring is broad, occurring twenty-four hours a day.[246] To justify such a broad scope, "schools would have to suspect that all students are engaged in cyberbullying at all times."[247] However, Professor Suski also acknowledges that this argument is not airtight and that schools would have a "decent argument" in support of the surveillance under *Acton*.[248] "While not as limited a search as the drug testing in *Acton*, the argument exists that school surveillance does respond to a strong need for school intervention and discipline in order to combat cyberbullying, much like the searches in *Acton*."[249]

Further support for the argument that online surveillance of students violates their protected expectation of privacy might be found in a seemingly unrelated Supreme Court opinion dealing with GPS tracking of cars. In *United States v. Jones*, the police had placed a GPS tracker on the defendant's car.[250] The district court suppressed the data that had been gathered when the car was parked at the defendant's private residence but refused to exclude the other information gathered by that tracker, concluding that a person who is travelling on public roads has no expectation of privacy in his or her movements.[251] The D.C. Circuit disagreed, and the Supreme Court likewise sided with the defendant.[252] Writing for the majority, Justice Scalia rejected the government's argument that there was no Fourth Amendment violation here, as the data at issue merely provided the car's location on public roads, roads that "were visible to all."[253] In rejecting that argument, Justice Scalia made clear that even though the police could have lawfully surveilled the car without a warrant by traditional means such as having officers follow the car, the surveillance by GPS violat-

clear that users' expectation of privacy in their searches and posts is limited."). *But see* United States v. Meregildo, 883 F. Supp. 2d 523, 525 (S.D.N.Y. 2012) ("Whether the Fourth Amendment precludes the Government from viewing a Facebook user's profile absent a showing of probable cause depends, *inter alia*, on the user's privacy settings. When a social media user disseminates his postings and information to the public, they are not protected by the Fourth Amendment.") (citations omitted); *R.S.*, 894 F. Supp. 2d at 1142 (holding that even if a user's public Facebook wall is not private, their direct messages within the Facebook app are).

[245] Suski, *supra* note 134, at 94.

[246] *Id.* at 95.

[247] *Id.*

[248] *Id.* at 117.

[249] *Id.*

[250] United States v. Jones, 565 U.S. 400, 403 (2012).

[251] *Id.*

[252] *Id.* at 404, 413.

[253] *Id.* at 406.

ed the Fourth Amendment.[254] He further noted that the Court was unaware of any cases that would support an argument that "what would otherwise be an unconstitutional search is not such where it produces only public information."[255]

Using the logic of the *Jones* holding, an advocate might be able to persuade the Court that online surveillance of students by geofencing is akin to having a warrantless GPS placed on their cars. Even though the students are posting on a public website, they are like Jones when he drove on public roads. School officials can lawfully view their students' posts through traditional surveillance, just as the police in *Jones* could have tailed his car, but the added element of using technology to conduct this surveillance turns it into an unlawful search. Of course, the *Jones* majority relied on the fact that the police had touched the undercarriage of the car while placing the GPS monitor, a physical intrusion that is not present with purely online surveillance.[256] Nonetheless, Justice Scalia did not foreclose the idea that electronic surveillance without a physical act could still violate the Fourth Amendment.[257] "It may be that achieving the same result through electronic means, without an accompanying trespass, is an unconstitutional invasion of privacy, but the present case does not require us to answer that question."[258]

One further point must be made in discussing potential Fourth Amendment violations. We live in the era of "surveillance capitalism," described as "a new economic order that claims human experience as a free source of raw material."[259] The online activity of any person is an extremely valuable commodity in this new world order, and third-party marketers and others are willing to pay a premium to access it.[260] Indeed, the online activities of schoolchildren are an especially valuable commodity and children are a highly prized audience, given that brand loyalty impressions created during youth can last for a lifetime.[261] Thus, students' interest in their privacy with respect to their online ac-

---

[254] *Id.* at 412.

[255] *Id.* at 409.

[256] *Id.* at 410.

[257] *Id.* at 412.

[258] *Id.*

[259] Jacob Silverman, *How Tech Companies Manipulate Our Personal Data*, N.Y. TIMES (Jan. 18, 2019), https://www.nytimes.com/2019/01/18/books/review/shoshana-zuboff-age-of-surveillance-capitalism.html [https://perma.cc/HXH7-WG7U]; *see also* Shoshana Zuboff, *Big Other: Surveillance Capitalism and the Prospects of an Information Civilization*, 30 J. INFO. TECH. 75, 75 (2015).

[260] *See* Max Eddy, *How Companies Turn Your Data Into Money*, PCMAG (Oct. 10, 2018), https://www.pcmag.com/article/364152/how-companies-turn-your-data-into-money [https://perma.cc/B2K9-7BGK].

[261] *See* Jennifer Comiteau, *When Does Brand Loyalty Start?*, ADWEEK (Mar. 24, 2003), https://www.adweek.com/brand-marketing/when-does-brand-loyalty-start-62841 [https://perma.cc/QAA7-Y59K] (noting that "American children become 'brand-conscious' at about 24 months, and by 36–42 months they make the connection that a brand can say something about their personalities—they are strong or cool or smart.").

tivity is in some ways uniquely heightened, "as schools collecting sensitive information about their students may subsequently put this private data in the hands of for-profit companies."[262] Khaliah Barnes, a lawyer at the Electronic Privacy Information Center in Washington, says that "[s]tudents are currently subject to more forms of tracking and monitoring than ever before," and that "there are too few safeguards for the amount of data collected and transmitted from schools to private companies."[263] Updates made in 2013 to the Federal Education Rights Privacy Act (FERPA) actually compounded this problem, rather than improving it, as the changes "permit schools to share student data, without notifying parents, with companies to which they have outsourced core functions like scheduling or data management."[264] Some state lawmakers are so concerned about the possibility of schools selling student data that they have passed laws prohibiting schools from selling or otherwise sharing certain data about their students.[265]

Accordingly, students may have a successful Fourth Amendment claim arguing against online monitoring of them, especially if the schools are not being careful about who can access the data that they are collecting and what they do with it. But, as with a First Amendment argument, the Supreme Court is likely to be sympathetic to a school's proffered explanation that it is doing this monitoring to prevent school shootings or other forms of violence, and *Acton* opens up leeway for schools to conduct such "suspicionless" searches.

D. Equal Protection

Students who are disciplined for their off-campus online activity may have claims under the First and Fourth Amendment, as discussed above, and there may even exist arguments for students who are not disciplined but are simply subject to such online surveillance. For the narrower class of students of color who are the victims of disproportioned punishment for their online expression, they may also be able to challenge any discipline they receive under the Equal

---

[262] Chandran, *supra* note 144, at 314.

[263] Natasha Singer, *Deciding Who Sees Students' Data*, N.Y. TIMES (Oct. 5, 2013), https://www.nytimes.com/2013/10/06/business/deciding-who-sees-students-data.html [https://perma.cc/6CRJ-P5ZK].

[264] *Id.*; *see also* Kevin Miller, *Total Surveillance, Big Data, and Predictive Crime Technology: Privacy's Perfect Storm*, 19 J. TECH. L. & POL'Y 105, 112 (2014) ("Not only are the companies providing learning data systems often not clear about with whom they share data, parents are concerned about what will eventually come of behavioral data and other assessments—and whether that information will permanently limit their child's future.").

[265] Chandran, *supra* note 144, at 314; *see also* Natasha Singer, *With Tech Taking Over in Schools, Worries Rise*, N.Y. TIMES (Sept. 14, 2014), https://www.nytimes.com/2014/09/15/technology/with-tech-taking-over-in-schools-worries-rise.html [https://perma.cc/UKS5-ZPXN] (noting that California passed a law "prohibiting educational sites, apps and cloud services used by schools from selling or disclosing personal information about students from kindergarten through high school; from using the children's data to market to them; and from compiling dossiers on them.").

Protection clause of the Fourteenth Amendment. For the reasons that follow, though, such claims are unlikely to succeed.

The Fourteenth Amendment's Equal Protection Clause prohibits state actors from denying any person "the equal protection of the law."[266] If the challenged policy is "facially neutral but its application results in racially disproportionate outcomes,"[267] courts will apply strict scrutiny when examining it.[268] As is discussed above, there is already some evidence that there are racial disparities in the way that schools punish student online speech,[269] which is consistent with other evidence surrounding national school discipline trends.[270] "However, a challenge under the Equal Protection Clause to [discipline for online activity] will nonetheless be difficult to maintain as the Supreme Court has consistently held that statistical evidence alone, absent discriminatory intent or purpose, is not enough."[271]

In *Washington v. Davis*, the Supreme Court denied an equal protection claim brought by black applicants to the police force in Washington, D.C.[272] Because more blacks than whites failed the written test, and because the written test was not shown to correlate with success as a police officer, the plaintiffs challenged the test under the Equal Protection Clause.[273] The Court rejected the argument, noting that "our cases have not embraced the proposition that a law or other official act, without regard to whether it reflects a racially discriminatory purpose, is unconstitutional *solely* because it has a racially disproportionate impact."[274] Although the Court conceded that a racially disproportionate impact is "not irrelevant," it nonetheless held that "[s]tanding alone, [racially disproportionate impact] does not trigger the rule, that racial classifications are to be subjected to the strictest scrutiny and are justifiable only by the weightiest of considerations."[275]

Courts that have applied *Washington* to school discipline cases where the plaintiffs pointed out racially disproportionate impacts "have demanded more than statistical evidence and have looked for evidence of racially discriminatory

---

266 U.S. CONST. amend. XIV § 1.

267 Cyphert, *supra* note 98, at 916 (citing Yick Wo v. Hopkins, 118 U.S. 356 (1886)).

268 Hunt v. Cromartie, 526 U.S. 541, 546 (1999) (citations omitted).

269 *See, e.g.*, Jambulapati, *supra* note 91(explaining that in one Alabama school district, twelve of the fourteen students who were expelled in a school year for the content of their social media were African-American, despite the fact that African-American students made up only 40 percent of the district).

270 *See, e.g.*, Balingit, *supra* note 97 (in the 2015–2016 school year, according to federal data, "[b]lack students faced greater rates of suspension, expulsion and arrest than their white classmates . . . disparities that have widened despite efforts to fix them.").

271 Cyphert, *supra* note 98, at 916.

272 Washington v. Davis, 426 U.S. 229, 232 (1976).

273 *Id.* at 235.

274 *Id.* at 239.

275 *Id.* at 242 (citation omitted).

intent or animus."[276] A district court addressed the question directly in *Fuller v. Decatur Public School*, where students who were expelled for fighting at a football game challenged their expulsions under the Equal Protection Clause.[277] Despite the fact that the school district acknowledged that African American students made up 82% of the students who were expelled despite comprising only 46–48% of the student body, the court rejected the students' equal protection claim.[278] Although the court recognized that such statistics "could lead a reasonable person to speculate that the School Board's expulsion action was based upon the race of the students," the court nonetheless held that it could not "make its decision solely upon statistical speculation."[279] Rather, without any evidence of actual racial animus on the part of the school officials, the claim failed, as "the law is clear that a claim of racial discrimination and violation of equal protection cannot be based upon mere statistics standing alone."[280]

Other courts have joined *Fuller* and held that *Washington* requires more than statistical evidence to sustain an equal protection claim on behalf of students of color who are disciplined by their schools.[281] Accordingly, in the absence of any information suggesting that the policy or school officials specifically targeted students of color because of their race, it is unlikely that they would have a successful equal protection claim if they are disciplined by their schools for online expression.[282]

III. POLICY RECOMMENDATIONS

Scholars have called upon the Supreme Court to grant certiorari to cases involving school monitoring and regulation of student online speech in order to provide clarity and consistency for the lower courts to follow.[283] The Court

---

[276] Cyphert, *supra* note 98, at 917.

[277] Fuller v. Decatur Pub. Sch. Bd. of Educ. Sch. Dist. 61, 78 F. Supp. 2d 812, 814, 823 (C.D. Ill. 2000), *aff'd sub nom*., 251 F.3d 662 (7th Cir. 2001).

[278] *Fuller*, 78 F. Supp. at 824–25.

[279] *Id.* at 824.

[280] *Id.* at 825.

[281] *See* Cyphert, *supra* note 98, at 917–18 (noting that "[o]ther courts have used this same analysis in rejecting [e]qual [p]rotection claims brought by suspended or expelled students of color, holding that 'statistical proof that black students are disciplined more frequently and more severely than white and Mexican-American students has limited probative value,' " and concluding that "data alone would not be enough to establish an [e]qual [p]rotection claim on behalf of students of color who were suspended or expelled from" their schools) (citation omitted).

[282] The *Loomis* decision, discussed above in Part I, is not probative of how a court might handle an equal protection claim brought by a student challenging online monitoring because in that case, the defendant made a due process challenge, not an equal protection one. State v. Loomis, 881 N.W.2d 749, 766 (Wisc. 2016), *cert. denied*, 137 S. Ct. 2290, (2017) ("Notably, however, Loomis does not bring an equal protection challenge in this case.").

[283] *See, e.g.*, Mendola, *supra* note 79, at 182–87 (urging the Supreme Court to rule on these cases and proposing the adoption of an adapted version of the *Tinker* substantial disruption test).

may decline to do so, or it may take the cases and then reject any legal arguments against online surveillance. This Article has attempted to predict what the Court might ultimately do, but it seems increasingly likely that this kind of surveillance will come to pass, and thus it is important to examine what best practices might look like.

Of course, there are potential benefits to surveillance, and they are worth noting. "By ensuring that students' Internet usage does not substantially interfere with their peers' learning, schools encourage the development of their students, who may worry less about Internet threats and more about their education."[284]

A. Invest in High Quality High School Counselors

Contracting with a third party to conduct online surveillance on your students is not just legally murky and ethically complicated—it's also expensive.[285] Many contracts cost tens-of-thousands of dollars per year.[286] Investing that money instead into counselors is more likely to address the root causes of violence in schools, whether that violence is self-harm or not.[287] To the extent schools are monitoring their students' online accounts to flag students at risk of death by suicide, a school counselor may be the only mental health professional that students have access to. Counselors can also be instrumental in preventing students from harming fellow students in episodes of school violence, as some experts conclude that school shootings are the result of gaps in the provision of mental health services.[288] Further, unlike online surveillance, which has not been proven effective, "[c]ounselors are well-tested: they have been standard in most public schools since the late twentieth century and their presence has proven to be effective in supporting and guiding students."[289] Counselors could help educate students on proper cyber usage and address some of the underlying issues that impact cyberbullying. According to researchers, the relationships counselors form with their students are critical in preventing school vio-

---

[284] *Id.* at 174; *see also* Chastel, *supra* note 88 (detailing how algorithms performing predictive analysis of at-risk students allow early intervention and promote student success).

[285] *See, e.g.*, Mendola, *supra* note 79, at 189 (noting that third-party surveillance can cost $40,000 per year).

[286] *Id.*

[287] *See* Chatterjee, *supra* note 84 (noting that while "[m]ental health issues don't cause school shootings . . . [as] only a tiny, tiny percentage of kids with psychological issues go on to become school shooters . . . mental health problems are a risk factor.").

[288] *See* Christopher J. Ferguson et al., *Psychological Profiles of School Shooters: Positive Directions and One Big Wrong Turn*, 11 J. POLICE CRISIS NEGOT., 141, 153 (2011) ("In many ways it is apparent to us that the issue of school shooters, like other mass homicide perpetrators, is very much a failure of the mental health system or, in fairness, a failure of society more broadly to provide adequate mental health services.").

[289] Mendola, *supra* note 79, at 189; *see also* Chatterjee, *supra* note 84 ("Time and time again, psychologists and educators have found that surrounding a young person with the right kind of support and supervision early on can turn most away from violence.").

lence: "[c]onnecting with these students, listening to them and supporting them, getting them the help they need . . . can help prevent future attacks and make schools a safer place for all children."[290]

B. Provide Students and Families with Transparency and Privacy Protections

Students and their families should not only be made aware that their school is engaging in online surveillance, they should be given an opportunity to meaningfully engage in conversations about the practice. Schools have had to respond to concerns as they learn that parents are wary of third parties receiving data about their children without their consent.[291] For example, the software company inBloom, which had funding from the Bill & Melinda Gates Foundation and the Carnegie Corporation of New York, once had contracts with nine different states to store the data of more than eleven million students in a cloud-based system.[292] But it shut down abruptly in 2014 after parent activists decried the lack of privacy the company was providing to student data.[293] Whether or not inBloom was actually careless with student data, the fact that parents were not given an opportunity to learn more about the services it provided was enough to doom the company.[294] State legislators are listening. In 2015, Delaware passed the Student Data Privacy Protection Act, which forbids third-party software companies from selling student data or engaging in targeted advertising based on student data, and which defines student data to include, among other things, a students' geolocation data, instant messages, photos, and search activity.[295] Any school that is considering hiring a third party to monitor its students' online activities should be transparent about who will be doing that monitoring, who will have access to the data that is collected, and should also host information sessions to address questions or concerns from parents and students.

---

[290] Chatterjee, *supra* note 84.

[291] *See, e.g.*, Ainsley Harris, *Privacy Concerns Force InBloom, A Data Repository for Schools, To Shut Down*, FAST COMPANY (Apr. 21, 2014), https://www.fastcompany.com/3029451/privacy-concerns-force-inbloom-a-data-repository-for-schools-to-shut-down [https://perma.cc/JTX6-XKM6].

[292] *Id.*

[293] *See id.* (noting a movement started by an education advocacy group helped promulgate legislation that blocked inBloom from the market, all but ensuring its inability to continue operation).

[294] Chastel, *supra* note 88 ("[M]any parents are extremely uncomfortable with the idea of not being informed about precisely what data is being collected, and more importantly, how it's being used.").

[295] DEL. CODE ANN. tit. 14, §§ 8101A, 8102A, 8105A (2016).

C. Take a More Intentional and Multidisciplinary Approach to the Use of Machine Learning

Predictive algorithms are here to stay. But that does not mean they cannot be improved, and that we cannot learn early lessons from their deployment in the criminal justice system and enact some best practices to help guide their use in our school system. Indeed, scholars have already begun to outline important technical ways that the stages of machine learning could be improved in terms of guarding against bias, such as not using certain approaches where outputs cannot be explained (such as convolutional neural networks).[296]

One non-technical best practice is for the teams who develop the kinds of predictive algorithms addressed in this Article to be more multidisciplinary. Rashida Richardson, director of policy research at New York University's AI Now Institute, which "studies the social implications of artificial intelligence," warns that:

> [P]eople making these algorithms don't necessarily understand all the social, and even political, aspects of the data they're using . . . . Researchers may not understand a lot of the nuances of what people in the education and legal policy world call school climate. That includes safety and behavioral issues . . . . The kind of school you're in will often dictate how behavior is dealt with and how discipline is handled.[297]

Others echo Richardson's concern about a lack of a multidisciplinary approach. "As machine learning has expanded beyond its roots in the worlds of computer science and statistics into nearly every conceivable field, the data scientists and programmers building those models are increasingly detached from an understanding of how and why the models they are creating work."[298] Schools should ask questions about the team that developed the algorithm before they enter into a contract with a company. Was it multi-disciplinary? Did they consult with education experts? Legal experts? These are important questions with serious consequences.

CONCLUSION

> In the end, as technology places ever more powerful tools in the hands of those without an understanding of how they work, we are creating great business and societal risk if we don't find ways of building interfaces to these models such

---

[296] *See* Lehr & Ohm, *supra* note 15, at 715 ("By focusing on the many neglected middle stages of machine learning, legal scholars and policymakers will find creative new methods for detecting and ameliorating harm . . . if decision-making could lead to imprisonment or the loss of life, perhaps particularly unexplainable approaches such as convolutional neural networks should not be used.").

[297] Rieland, *supra* note 80 (internal quotation marks omitted).

[298] Kalev Leetaru, *A Reminder That Machine Learning Is About Correlations Not Causation*, FORBES (Jan. 15, 2019), https://www.forbes.com/sites/kalevleetaru/2019/01/15/a-reminder-that-machine-learning-is-about-correlations-not-causation [https://perma.cc/47E8-RF28].

> that they are able to communicate these distinctions and issues like data bias to their growing user community that lacks an awareness of those concerns.[299]

School students are an especially vulnerable population. As schools work toward the essential goal of protecting students, they must be careful to respect their legal rights and their privacy. Machine learning technology is still so new and is rapidly evolving. But lessons can and should be learned from its use in the criminal justice system. If we truly want to keep our children safe, both from violence and from bias, we must tread carefully and deliberately.

---

[299] *Id.*

[THIS PAGE INTENTIONALLY LEFT BLANK]