ALGORITHMIC SYSTEMS IN EDUCATION: INCORPORATING EQUITY AND FAIRNESS WHEN USING STUDENT DATA

AUGUST 2019

A SERIES OF PAPERS ON STUDENT PRIVACY

CENTER FOR DEMOCRACY & TECHNOLOGY

ABOUT CENTER FOR DEMOCRACY & TECHNOLOGY

The Center for Democracy & Technology is a 501(c)(3) working to promote democratic values by shaping technology policy and architecture, with a focus on the rights of the individual. CDT supports laws, corporate policies, and technological tools that protect privacy and security and enable free speech online. Based in Washington, D.C., and with a presence in Brussels, CDT works inclusively across sectors to find tangible solutions to today's most pressing technology policy challenges. Our team of experts includes lawyers, technologists, academics, and analysts, bringing diverse perspectives to all of our efforts.

Learn more about our experts and the issues we cover: https://cdt.org/

ABOUT STUDENT PRIVACY

CDT's vision for the *Student Privacy Project* is to create an educated citizenry that is essential to a thriving democracy by protecting student data while supporting its responsible use to improve educational outcomes. To achieve this vision, CDT advocates for and provides solutions-oriented resources for education practitioners and the technology providers who work with them, that center the student and balance the promises and pitfalls of education data and technology with protecting the privacy rights of students and their families.

AUTHORED BY

Hannah Quay-de la Vallee, Senior Technologist
Natasha Duarte, Policy Analyst

Cover and layout designed by Tim Hoagland

Table of Contents

Cover and layout designed by: Timothy Hoagland

CENTER FOR DEMOCRACY & TECHNOLOGY

Algorithmic Systems in Education: Incorporating Equity and Fairness When Using Student Data

Executive Summary

Some K-12 school districts are beginning to use algorithmic systems to assist in making critical decisions affecting students' lives and education. Some districts have already integrated algorithms into decision-making processes for assigning students to schools, keeping schools and students safe, and intervening to prevent students from dropping out. There is a growing industry of artificial intelligence startups marketing their products to educational agencies and institutions. These systems stand to significantly impact students' learning environments, well-being, and opportunities. However, without appropriate safeguards, some algorithmic systems could pose risks to students' privacy, free expression, and civil rights.

This issue brief is designed to help all stakeholders make informed and rights-respecting choices and provides key information and guidance about algorithms in the K-12 context for education practitioners, school districts, policymakers, developers, and families. It also discusses important considerations around the use of algorithmic systems including accuracy and limitations; transparency and explanation; and fairness and equity.

To address these considerations, education leaders and the companies that work with them should take the following actions when designing or procuring an algorithmic system:

- **Assess the impact of the system and document its intended use:** Consider and document the intended outcomes of the system and the risk of harm to students' well-being and rights.
- **Engage stakeholders early and throughout implementation:** Algorithmic systems that affect students and parents should be designed with input from those communities and other relevant experts.
- **Examine input data for bias:** Bias in input data will lead to bias in outcomes, so it is critical to understand and eliminate or mitigate those biases before the system is deployed.
- **Document best practices and guidelines for future use:** Future users need to know the appropriate contexts and uses for the system and its limitations.

Once an algorithmic system is created and implemented, the following actions are critical to ensuring these systems are meeting their intended outcomes and not causing harm to students:

- **Keep humans in the loop:** Algorithmic decision-making systems still require that humans are involved to maintain nuance and context during decision-making processes.
- **Implement data governance:** Because algorithmic systems consume and produce a lot of data, a governance plan is needed to address issues like retention limits, deletion policies, and access controls.

- **Conduct regular audits:** Audits of the algorithmic system can help ensure that the system is working as expected and not causing discriminatory outcomes or other unexpected harm.
- **Ensure ongoing communication with stakeholders:** Regular communication with stakeholders can help the community learn about, provide feedback on, and raise concerns about the systems that affect their schools.
- **Govern appropriate uses of algorithmic systems:** Using an algorithm outside of the purposes and contexts for which it was designed and tested can yield unexpected, inaccurate, and potentially harmful results.
- **Create strategies for accountability and redress:** Algorithmic systems will make errors, so the educational institutions employing them will benefit from having plans and policies to catch and correct errors, receive and review reports of incorrect decisions, and provide appropriate redress to students or others harmed by incorrect or unfair decisions.
- **Ensure legal compliance:** While legal compliance is not enough to ensure that algorithmic systems are fair and appropriate, algorithmic systems must be held to the same legal standards and processes as other types of decision-making, such as FERPA and civil rights protections.

Introduction

K-12 educational agencies and institutions are navigating a growing market of algorithmic decision-making systems designed to transform district and school functions such as assigning students to schools,[1] preventing dropout,[2] and keeping students safe.[3] These decisions can significantly affect students' experiences, relationships, and future opportunities, whether by determining which school a student attends and thus what teachers and extracurriculars are available to her, or by deciding whether or not that student is a threat to school safety, thus impacting how she is perceived throughout her educational career and beyond. Therefore, it is important that education practitioners and policymakers understand how these systems work, their limitations, and their implications for students and families, so they can take steps to ensure that algorithmic systems support all students and do not introduce or perpetuate discriminatory decision-making, privacy risks, or other harms.

Education practitioners and policymakers who do not have a background in computer or data science might assume that they are unqualified to evaluate decision-making processes that involve the use of algorithms in their schools and districts. However, many aspects of these systems can be scrutinized without specific technical expertise.[4] Designing and implementing these systems involves a series of policy choices about how to solve education-related problems and how to best support and interact with students and their families. These are questions that educators and school officials are uniquely

[1] *See infra* text accompanying notes 17–19.
[2] *See infra* text accompanying note 20-21.
[3] *See infra* text accompanying notes 22–38.
[4] *See* Aaron Rieke, Miranda Bogen & David G. Robinson, Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods 5, February 27, 2018, https://www.omidyar.com/sites/default/files/file_archive/Public%20Scrutiny%20of%20Automated%20Decisions.pdf.

qualified to answer. Education practitioners can and should use their expertise to help ensure that algorithmic systems are designed and governed in ways that support the achievement and well-being of all students. Policymakers also have an important role to play in setting the agenda for which types of algorithmic systems are used in schools and ensuring that they preserve civil rights, equal opportunity, privacy, and healthy school environments.

Overview of Algorithmic Systems

As in other sectors, K-12 educational institutions have been experimenting with computer algorithms to support human decision-making in areas that directly impact students' lives and educations. Although some of these algorithms may be developed locally by school districts, many are developed by companies and licensed to schools across the country.[5] It may be tempting for local policymakers and education practitioners to defer judgment of these systems to outside experts, but the people who live and work in a school district are often in the best position to decide whether a particular technology is right for their schools and to observe its impacts on students. Thus, it is important for all stakeholders to understand some basic concepts about algorithmic systems.

Algorithms are Designed by Humans and Reflect Human Value Judgments and Biases

An **algorithm** is a process performed by a computer to answer a question or carry out a task, such as sorting students into schools or classifying social media posts.[6] Although they involve math, algorithms are not neutral decision-makers. Subjective human judgments dictate the purpose, design, and function of an algorithm and influence its outcomes.

First, humans decide what problems or objectives to address with algorithms. For example, when designing an algorithm that sorts students into different schools, the school district officials and the algorithm's developers must balance a range of value-laden considerations:[7] How far should students have to travel to get to school? Should districts optimize for school diversity, and what should diversity look like? How should priority be assigned to families' choices? Should a student's financial situation or

---

[5] *See, e.g.*, Jodi Hillman, *AI in Education: 10 Companies to Watch in 2018*, Disruptor Daily (Oct. 23, 2017), https://www.disruptordaily.com/ai-education-10-companies-using-ai-2017/; Faiza Patel et al., Brennan Center for Justice, School Surveillance Zone (Apr. 30, 2019), https://www.brennancenter.org/analysis/school-surveillance-zone; Mass. Inst. of Tech. News, *What 126 studies say about education technology* (Feb. 26, 2019), http://news.mit.edu/2019/mit-jpal-what-126-studies-tell-us-about-education-technology-impact-0226.

[6] This is a basic summary derived from previous definitional work. For other useful definitions and descriptions of algorithms, *see also, e.g.*, Ctr for Democracy & Tech., Digital Decisions, https://cdt.org/issue/privacy-data/digital-decisions/; AI Now Institute, Algorithmic Accountability Policy Toolkit 2 (Oct. 1, 2019), https://ainowinstitute.org/aap-toolkit.pdf; Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 Calif. L. Rev. 671, 677–84 (2016), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899.

[7] *See, e.g.*, Matt Kasman & Jon Valant, Brookings Institution, The Opportunities and Risks of K-12 Student Placement Algorithms (Feb. 28, 2019), https://www.brookings.edu/research/the-opportunities-and-risks-of-k-12-student-placement-algorithms/.

other risk factors ever be considered in school assignment, and if so, how? These are questions of policy, not math, and their answers might depend in part on the specific context and characteristics of the district. Education practitioners, district-level officials, and students and parents themselves have critical localized insight into these questions and their potential effects on students.

Human judgments are also deeply embedded in the design of *how* an algorithm works. Algorithms rely on certain inputs, called "features" — such as the words in a social media post[8] or the order in which students ranked their school choices[9] — to determine the outputs. Often, the algorithm's developers decide exactly which features to input and how important each feature will be in determining the output (the feature's "weight"). For example, an algorithm might be designed to flag social media posts as "bullying" if they contain any of a predetermined set of words.[10] In this case, the developers define the words that will cause the algorithm to flag a post — perhaps based on their own personal observations, academic literature, popular culture, or some combination thereof.[11]

Some algorithmic systems are developed using **machine learning** techniques.[12] In machine learning, a computer "learns" from examples (called "training data"), which are usually selected by the developers.[13] For example, instead of designing an algorithm to classify a social media post as "bullying" if it contains one of a predetermined set of words, developers might train a computer to recognize bullying by feeding it examples of social media posts that the developers (or others) have labeled as bullying or non-bullying. Based on the examples, the computer generates a model for classifying new social media posts as bullying or not. In this case, the developers may not be selecting features or weights, but they are selecting the training examples that the computer learns from, the underlying model, and the parameters (such as how much error is acceptable). If the developers' sample under- or over-represents certain groups of students, or is skewed to include only a certain type of bullying (e.g., verbal abuse vs. physical abuse), those biases will be reflected in the learned model.

---

[8] *See* Natasha Duarte, Emma Llansó & Anna Loup, Ctr for Democracy & Tech., Mixed Messages: The Limits of Automated Social Media Content Analysis 10–11 (Nov. 2017), https://cdt.org/insight/mixed-messages-the-limits-of-automated-social-media-content-analysis/.

[9] *See infra* text accompanying notes 17–19.

[10] *See* Tom Simonite, *Schools Are Mining Students' Social Media Posts for Signs of Trouble*, Wired (Aug. 20, 2018), https://www.wired.com/story/algorithms-monitor-student-social-media-posts/.

[11] *See* Duarte, Llansó & Loup, *supra* note 8, at 10–11, 16–17; Haley Zapal, Bark, *Teen Slang Through the Ages*, https://www.bark.us/blog/teen-slang-through-the-ages/.

[12] The term artificial intelligence (AI) is often used to refer to algorithmic systems. AI is an umbrella term whose meaning often depends on who is using the term. AI is often used to describe more complex processes such as machine learning, but it has also been used to refer to simpler statistical functions or any system in which a computer processes data to produce outputs.

[13] For a more detailed but still accessible description of machine learning, *see* CDT, Digital Decisions, *supra* note 6; Barocas & Selbst, *supra* note 6, at 677–684.

Algorithms Do Not Exist in a Vacuum

Algorithms are part of and interact with larger systems involving laws, policies, human decision-makers, and social factors.[14] For example, school assignment algorithms are just one component of a larger system that determines where students attend school. A student's geographic area and that district's policies determine the universe of school options that families have in the first place. The district or state may have policy goals, such as good student-teacher ratios, efficient busing, or school diversity, that the algorithm is designed to bring about. In most cases, human decision-makers are reviewing the algorithm's outputs and making final decisions. All of these factors — the school district's goals, the relevant laws, the human decision-makers, and the pre-existing social factors present in the district's schools — will influence student and school outcomes. Thus, evaluating the function and effects of an algorithm requires us to look at the system as a whole.[15] This issue brief uses the term "**algorithmic system**" to refer holistically to a decision system that involves algorithms, human decision-makers, legal and social structures, and other forces.[16]

Examples of Algorithmic Systems in Education

The education sector is beginning to use algorithmic decision-making systems in a number of ways. This brief will focus on three of the more common uses: school assignment systems, dropout early warning systems, and school safety systems.

School Assignment Systems

Some school systems offer school choice to families: families can request that their children attend a certain school, rather than being assigned to a school based on where the family lives. These school systems may choose to create a centralized enrollment process that uses an algorithmic system to decide which children attend which schools. Because schools in the system likely vary in desirability, it is unlikely that every child will get to attend their first-choice school. As an example, consider DC's school lottery system, called My School DC.[17] The My School DC algorithm distributes students to schools in such a way that it reports that most students are placed in a school that is relatively high on their

---

[14] *See, e.g.*, Andrew Selbst et al., *Fairness and Abstraction in Sociotechnical Systems*, ACM Conference on Fairness, Accountability, and Transparency (2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3265913; Alexandra Chouldechova, *A case study in algorithm-assisted decision making in child maltreatment hotline screening decisions*, ACM Conference on Fairness, Accountability, and Transparency 13–14 (2018), http://proceedings.mlr.press/v81/chouldechova18a/chouldechova18a.pdf; Ctr for Democracy & Tech., Comments on the European Commission's High Level Expert Group on Artificial Intelligence's Draft Ethics Guidelines for Trustworthy AI 1, 3–4 (Feb. 4, 2019), https://cdt.org/files/2019/02/comment_-EU-Commission-HLEG-AI-guidelines-1.pdf.

[15] *See* sources cited *supra* note 14.

[16] These systems are sometimes referred to as automated decision systems. *See, e.g.*, AI Now, Algorithmic Accountability Policy Toolkit, *supra* note 6, at 2.

[17] My School DC, My School DC: The Public School Lottery, https://www.myschooldc.org/.

preference list without overcrowding schools.[18] To enter into the lottery, parents or guardians submit an application for their child listing the schools they want to apply to in order of descending preference. In addition to the order of preference stated by the parent, the lottery also considers other factors like whether the student has a sibling at a given school or whether the school is in close proximity to the student's home.[19] For school assignment algorithms, the weights the designers assign to the various factors influence which students are assigned to which schools. These weights are essentially value judgments about which factors in the algorithm should matter — and how much — when determining which students attend which schools.

Dropout Early Warning Systems

Dropout early warning systems are intended to flag students who are at risk of dropping out of school before graduating. These systems can be quite simple (like attendance thresholds or GPA), but there are far more complex versions. The dropout early warning system used by the Kentucky Department of Education uses machine learning over a broad range of factors like attendance, behavioral information, home and family stability, demographics, and how the student is faring related to other students in a similar situation to assign students a Graduation Related Analytic Data (GRAD) score, which is an indicator of how likely the student is to progress to the next grade level or to graduation.[20] In addition to assigning the student an overall GRAD score, the system also calculates an *impact score* for the categories that factor into the GRAD score, which indicates how much a given category is impacting the overall score. This helps educators understand where to focus their interventions to have the greatest effect.

It is important to note that the dropout early warning system is intended to help educators and staff understand which students need intervention and in which areas; it is not an intervention itself. Kentucky's algorithmic system is complemented by a variety of intervention programs to help educators act on the recommendations from the system in a meaningful way.[21]

---

[18] The My School DC algorithm is based on a Nobel Prize winning algorithm that serves as the basis for a number of school systems such as New York City public schools. Thomas Toch, *The Lottery That's Revolutionizing DC Schools*, The Washington Post Magazine (Mar. 20, 2019), https://www.washingtonpost.com/news/magazine/wp/2019/03/20/feature/the-lottery-thats-revolutionizing-d-c-schools/.

[19] My School DC, Lottery Preferences, https://www.myschooldc.org/faq/key-terms#preference.

[20] Infinite Campus, Early Warning, https://content.infinitecampus.com/sis/latest/documentation/early-warning/; Kentucky Department of Education, Early Warning and Persistence to Graduation Data Tools (Apr. 2, 2019), https://education.ky.gov/educational/int/Pages/EarlyWarningAndPersistenceToGraduation.aspx.

[21] Joannah Hornig Fox, Erin S. Ingram & Jennifer L. Depaoli, For All Kids: How Kentucky is Closing the High School Graduation Gap for Low-Income Students, Civic Enterprises and the Everyone Graduates Center at Johns Hopkins University (July 26, 2016), https://www.americaspromise.org/sites/default/files/d8/18571_Civic_KY_CaseStudy_v15.pdf.

School Safety Systems

Tasked with ensuring the safety of their students, some local and state education agencies are turning to technological solutions, including algorithmic decision-making systems. One goal of these systems might be to predict which students may be at risk of committing a violent act, often by monitoring student social media posts, an approach that currently lacks empirical grounding (see box on *Mixed Messages and School Safety: The Limitations and Risks of Social Media Monitoring*).[22] These systems access students' social media information, typically by using the social media platforms' Application Programming Interfaces (APIs).[23] Using this data, the systems aim to flag posts or accounts of concern. The methods used to make this determination vary from simple keyword flagging systems to machine learning algorithms that factor in a broader range of data. The keyword flagging systems simply scan posts for words on a preset list, like "bomb," "gun," or "shooting" (a method that fails to capture many linguistic nuances such as hyperbole, sarcasm, and words with contextual or slang meanings). The more complex systems may incorporate additional information such as the social graph (the web of connections between people on a social network), sentiment analysis of posts (e.g., whether the post seems to be generally positive or negative in tone), word embeddings (a machine learning technique that aims to identify words with similar meaning based on how they are used in context), or the speaker's demographics (such as their stated age or gender).

More complex systems may be "fixed," or they may "learn" through continued use. In fixed algorithms, the designers set specific weights for each of the factors considered, and a post or account is flagged if it reaches a certain score. For algorithms that learn, the designer sets initial weights, but in an effort to make the system more accurate over time, those weights change depending on feedback from school administrators or other sources.

Social media information may be part of a larger threat assessment database, as in the case in Florida, where the Marjory Stoneman Douglas High School Public Safety Act mandates that social media information from students be combined with data from other sources such as law enforcement and social services.[24] (The database went live in August 2019, despite bureaucratic issues and unresolved privacy and governance concerns that delayed its implementation by eight months[25]). This larger database may also feed into an algorithmic decision system similar to those that operate using only social media information.

---

[22] Aaron Leibowitz & Sarah Karp, *Chicago Public Schools Monitored Social Media for Signs of Violence, Gang Membership*, ProPublica Illinois (Feb. 11, 2019), https://www.propublica.org/article/chicago-public-schools-social-media-monitoring-violence-gangs#; Lynn Jolicouer, *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.

[23] APIs are a portal that allows two tools to communicate with one another & pass info between them. In this case, it allows the social media monitoring tool to access posts & other data that are held by the social media platform.

[24] Benjamin Herold, *To Stop School Shootings, Fla. Will Merge Government Data, Social Media Posts*, EdWeek (July 26, 2018), https://www.edweek.org/ew/articles/2018/07/26/to-stop-school-shootings-fla-will-merge.html.

[25] Fla. Laws 2018-3 (S.B. 7026), http://laws.flrules.org/2018/3; Katya Schwenk, *Florida Launches School Safety Database Despite Privacy Concerns*, EdScoop (Aug. 6, 2019), https://edscoop.com/florida-launches-school-safety-database-despite-privacy-concerns/.

Mixed Messages and School Safety: The Limitations and Risks of Social Media Monitoring

Social media monitoring algorithms present special concerns, especially when used to predict safety issues such as mass shootings.

Social media monitoring algorithms have technical limitations that make them unreliable for predicting acts of violence.

- Algorithms can sort large amounts of social media posts according to their words, phrases, and other features.[26] However, algorithms do not possess human-like abilities to interpret the meaning or intent of the speaker.[27] **Social media monitoring algorithms tend to err in recognizing humor, sarcasm, slang, and novel uses of language.**[28]
- While human review is essential, **even humans often err in understanding the meaning of social media posts.** Age, gender, racial, and cultural differences can inhibit reviewers' ability to make sense of social media posts and magnify the risk of mistaking a joke for a serious threat.[29]
- **Algorithms cannot reliably predict acts of mass violence or terrorism.** While even one mass shooting is too many, these events are statistically rare. Because of their small sample size and the complexity of factors surrounding them, they cannot be reliably predicted with algorithms.[30]
- **Because of these limitations, social media monitoring algorithms tend to generate large numbers of false positives.**[31] A high number of false positives can overwhelm schools with unhelpful information and subject students to unnecessary scrutiny.

---

[26] *See* Duarte, Llansó & Loup, *supra* note 8, at 9–12.
[27] *Id.* at 6, 19–20.
[28] *Id.* at 19 (citing Ahmed Abbasi, Ammar Hassan & Milan Dhar, *Benchmarking Twitter Sentiment Analysis Tools*, Proceedings of the 9th Language Resources and Evaluation Conference (2014), https://pdfs.semanticscholar.org/d0a5/21c8cc0508f1003f3e1d1fbf49780d9062f7.pdf); *Id.* at 20 (citing Nikhil Sonnad, *Alt-right Trolls are Using These Code Words for Racial Slurs Online*, Quartz (Oct. 1, 2016), https://qz.com/798305/alt-right-trolls-are-using-googles-yahoos-skittles-and-skypes-as-code-words-for-racial-slurs-on-twitter/).
[29] *See, e.g.,* danah boyd, *For Privacy, Teens Use Encoded Messages Online*, Science Friday (Feb. 27, 2014), https://www.sciencefriday.com/articles/for-privacy-teens-use-encoded-messages-online/; Alex Hern, *Facebook Translates 'Good Morning' into "Attack Them", Leading to Arrest*, The Guardian (Oct. 24, 2017), https://www.theguardian.com/technology/2017/oct/24/facebook-palestine-israel-translates-good-morning-attack-them-arrest; Desmond Patton, *Annotating Twitter Data from Vulnerable Populations: Evaluation Disagreement Between Domain Experts Graduate Student Annotators*, 52nd Hawaii International Conference on System Sciences (Jan. 9, 2019), https://scholarspace.manoa.hawaii.edu/bitstream/10125/59653/0213.pdf; Zapal, *supra* note 11.
[30] Letter from computer science experts to Elaine C. Duke, Acting Secretary of Homeland Sec., Dept. of Homeland Sec. (Nov. 16, 2017), https://www.brennancenter.org/sites/default/files/Technology%20Experts%20Letter%20to%20DHS%20Opposing%20the%20Extreme%20Vetting%20Initiative%20-%2011.15.17.pdf.
[31] *See, e.g.*, Aaron Leibowitz, *Could Monitoring Students on Social Media Stop the Next School Shooting?*, NY Times (Sept. 6, 2018), https://www.nytimes.com/2018/09/06/us/social-media-monitoring-school-shootings.html; Duarte, Llansó & Loup, *supra* note 8, at 17–19 (citing Leibowitz, *supra* note 31).

CENTER FOR DEMOCRACY & TECHNOLOGY

Social media monitoring algorithms create high risks to students' rights and wellbeing, especially when used to identify individuals as at risk of committing a violent act.

- **Social media monitoring invades students' privacy.** Systematic monitoring can reveal sensitive information about a student's personal life, such as their sexual orientation.
- **Flagging students as safety risks can impact their educational experiences and opportunities well into the future.** Even if a flagged student is ultimately cleared, she could still face negative repercussions if her records indicate that she was flagged as a safety concern.
- **Social media monitoring can chill expressive activities that are critical for young people's development.** Surveilling students' speech risks dissuading them from expressing their views, engaging in political organizing, or discussing sensitive issues such as mental health.[32]
- **The risk of harm from surveillance is likely to be concentrated in minority or marginalized communities**, including students of color, immigrants, and Muslim students or other religious minorities.[33] These groups may face a higher risk of punishment or law enforcement contact based on a flagged post[34] and may be particularly chilled for fear of punishment.[35] Algorithms can amplify societal bias and tend to misinterpret the posts of minority speakers more often.[36]
- Rather than relying on tools for predicting violence, educational agencies and institutions should take a more holistic and inclusive approach to school safety. For more information about data-driven school safety initiatives and social media monitoring, please see the following resources:
  - *Technological School Safety Initiatives: Considerations to Protect All Students*[37]
  - *Mixed Messages: The Limitations of Social Media Content Analysis*[38]

[32] Emily Witt, *From Parkland to Sunrise: A Year of Extraordinary Youth Activism*, New Yorker (Feb. 13, 2019), https://www.newyorker.com/news/news-desk/from-parkland-to-sunrise-a-year-of-extraordinary-youth-activism.
[33] Duarte, Llansó & Loup, *supra* note 8, at 13–16; Patton, *supra* note 28; Advancement Project, We Came to Learn: A Call to Action for Police-Free Schools (Sept. 13, 2018), https://advancementproject.org/wecametolearn/.
[34] Sarah Sparks & Alyson Klein, *Discipline Disparities Grow for Students of Color, New Federal Data Show*, Education Week (Apr. 24, 2018), https://www.edweek.org/ew/articles/2018/04/24/discipline-disparities-grow-for-students-of-color.html.
[35] *See* Sarah Brayne, *Surveillance and System Avoidance: Criminal Justice Contact and Institutional Attachment*, American Sociological Review (Apr. 4, 2014), https://journals.sagepub.com/doi/10.1177/0003122414530398.
[36] Duarte, Llansó & Loup, *supra* note 8, at 13–16 (citing Tolga Bolukbasi et al., *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*, Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS) (2016), https://papers.nips.cc/paper/6228-man-is-to-computer-programmer-as-woman-is-to-homemaker-debiasing-word-embeddings.pdf; Jieyo Zhao et al., *Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints*, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) (2017), https://arxiv.org/pdf/1707.09457; Jeff Larson, Julia Angwin & Terry Parris Jr., *Breaking the Black Box, How Machines Learn to Be Racist*, Episode 4, Artificial Intelligence, ProPublica (Oct. 19, 2016), https://www.propublica.org/article/breaking-the-black-box-how-machines-learn-to-be-racist?word=Trump; Su Lin Blodgett & Brendan O'Connor, *Racial Disparity in Natural Language Processing: A Case Study of Social Media African-American English* 1-2, Proceedings of the Fairness, Accountability, and Transparency in Machine Learning Conference (2017), https://arxiv.org/pdf/1707.00061.pdf.); Leila Ettachfini, *Court Reporters May Be Writing Down Black People's Testimonies Wrong*, Vice (May 23, 2019), https://www.vice.com/en_us/article/ywynzj/court-reporters-write-down-black-testimonies-wrong-study; John Eligon, *Speaking Black Dialect in Courtrooms Can Have Striking Consequences*, NY Times (Jan. 25, 2019), https://www.nytimes.com/2019/01/25/us/black-dialect-courtrooms.html.
[37] Ctr for Democracy & Tech. & Brennan Ctr for Justice, Technological School Safety Initiatives: Considerations to Protect All Students (May 24, 2019), https://cdt.org/files/2019/06/2019-05-24-School-safety-two-pager-Final.pdf.
[38] Duarte, Llansó & Loup, *supra* note 8.

Concerns About Algorithmic Decisions

Although algorithmic systems can provide value in the educational context, these systems can also cause harm to the very students they are intended to help. These harms often fall into three overarching categories.

The System Does Not Fulfill Expectations or is Not the Correct Tool for the Problem

A major concern with all algorithmic decision-making systems is that the system may not be effective for the purpose for which it is being used. For instance, a social media monitoring system that relies too heavily on keywords to flag concerning posts may end up flagging a student who was simply informing her friends that her recent track meet was "the bomb." Although this may not seem like a major problem, if these systems are feeding into a broader monitoring program, these seemingly small errors can lead to unnecessary interactions between law enforcement and students, which can have serious consequences for the student. This can lead to students being over-policed — if law enforcement then views them as a concern — and erode trust between the school and students and their families.

In addition to false positives (overreporting), false negatives (underreporting) can also be a problem. That same keyword-reliant algorithm that flagged the post with the phrase "the bomb" could be searching for keywords or phrases like "suicide" or "kill myself" to try to prevent self-harm incidents. However, it may miss more nuanced signals like a student sending goodbye messages to their friends. If the school relies too heavily on this algorithmic system, it may miss real-world indications that the student is in need of assistance.

The System May Exhibit Bias

Algorithmic systems are often perceived as rational arbiters that do not fall prey to human frailties like bias. However, algorithms are just as capable of bias as humans — or rather, human designers can embed their biases into the algorithm. Imagine, for example, a dropout early warning system designer who does not believe that attendance can play an important role in academic performance. The designer may then choose to not incorporate attendance data or to give it less weight than a different designer with different beliefs, resulting in a system that does not flag certain types of at-risk students (namely, those who are frequently truant).

Algorithmic bias may also arise from the data fed into the algorithm or fundamental facts about the world in which the algorithm operates, rather than the algorithm itself; that is, to use a colloquial phrase: "garbage in, garbage out." For instance, consider a dropout early warning system that factors in law enforcement interactions. Students of color are more likely to have a law enforcement interaction than their white peers.[39] Consequently, students of color may disproportionately trigger the dropout

[39] Jordan Green, *More cops in schools affect children of color the most*, Justice Policy Institute (Mar. 15, 2018), http://www.justicepolicy.org/news/12013; Evie Blad & Alex Harwin, *Black Students More Likely to Be Arrested at*

early warning system as compared with their white peers. Depending on the interventions used by the school, this could be damaging to white students if they are less likely to receive needed interventions than their peers of color. Alternatively, students of color could be disproportionately pushed onto lowered expectations tracks (the school only aims to graduate the student, not to push them to be college-ready, for instance).

The System May Infringe on Students' Rights

Algorithmic decision-making systems are heavily dependent on data, and it is important for designers and users to understand how and from where that data is sourced. An algorithmic system might involve the collection, processing, and monitoring of students' personal information in a way that is privacy-invasive, could chill students' First Amendment-protected activities, or inhibit their ability to learn. Social media monitoring, for instance, can be implemented in such a way that students' activities are monitored well beyond the school walls. This sort of overbroad surveillance can be a privacy violation to students and, over time, may cause a chilling effect on their speech and expressions, associations,[40] and movements. Certain threat assessment systems, like the Florida system discussed previously, require that very personal information about students, such as disciplinary information, be fed into a broader threat assessment database that may be accessible to stakeholders outside the education system like law enforcement.[41] This type of information-sharing may lead students and their families to avoid seeking out essential services, like mental health care, out of fear that their child will be perceived as a danger to others. In addition to the fundamental harm this can cause to students, it also limits the efficacy of these systems as students begin to "hide" data from the systems through means like hidden social media accounts, meaning the system is operating with low-quality data — a practice doomed to produce low-quality results.

Considerations for Algorithmic Decisions

Unfortunately, there is no checklist or defined set of procedures that can ensure that algorithmic systems will be free from harmful bias or other negative consequences. Instead, each system must be evaluated based on its specific function and design and the context in which it will be used, and must be evaluated on an ongoing basis to spot potential problems. Everyone plays a role in this process. Educators and school officials, who work directly with students and families, have valuable experience and localized knowledge that should inform algorithmic development and evaluation. Because they interact with students everyday, they may be more likely to recognize problems with how algorithmic systems are affecting students on the ground. While school- and teacher-level feedback should be part

School, EdWeek (Jan. 24, 2017), https://www.edweek.org/ew/articles/2017/01/25/black-students-more-likely-to-be-arrested.html.

[40] Blad & Harwin, *supra* note 39.

[41] Sidney Fussell, *Parkland is Embracing Student Surveillance*, The Atlantic (Jan. 29, 2019), https://www.theatlantic.com/technology/archive/2019/01/parklands-high-school-adds-new-surveillance-technology/581368/; Florida Department of Education, Department of Education Announces the Florida Schools Safety Portal (Aug. 1, 2019), https://nces.grads360.org/#communities/data-governance/publications/15066.

of the algorithmic development and management process, educators should not be overburdened with managing algorithmic systems at the expense of their focus on educating students. Developers, vendors, and district- and state-level officials cannot abdicate the responsibility of ensuring that algorithmic systems do not harm students.

Over the past decade, a community of research and practice has developed around the need to mitigate harm in algorithmic systems.[42] This has led to the development of many statements of principles, declarations, and guidance documents generally aimed at ensuring that algorithmic systems are human-centric, do not perpetuate harmful bias, and follow various legal and ethical standards.[43] Drawing on this work, this section discusses considerations to guide the design, use, and governance of algorithmic systems in education so that those systems will contribute to student success and not detract from it.

[Accuracy and Limitations]

When an algorithm stands to affect student outcomes, it should be accurate, work as intended, and be the right fit for the problem it is trying to solve. While this may seem fundamental, it is common to see overblown or misleading claims[44] about a system's capabilities, and to see experimental, unproven models released for general consumption.[45]

Does the Proposed Solution Fit the Problem?

As a threshold matter, the design of an algorithmic system should be the right fit for the problem it is trying to solve. Although some companies may offer "off-the-shelf" algorithmic systems for any school district to use, these tools may not take into account the specific needs and characteristics of a

[42] *See, e.g.*, FAT/ML, Fairness, Accountability, and Transparency in Machine Learning, https://www.fatml.org/; Partnership on AI, https://www.partnershiponai.org/about/#our-work; AI Now Institute, https://ainowinstitute.org/.

[43] *See, e.g.*, FAT/ML, Principles for Accountable Algorithms and a Social Impact Statement for Algorithms, https://www.fatml.org/resources/principles-for-accountable-algorithms; Association for Computing Machinery, Statement on Algorithmic Transparency and Accountability (Jan. 12, 2017), https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf; Leadership Conference on Civil & Human Rights, Civil Rights Principles for the Era of Big Data (Feb. 27, 2014), https://civilrights.org/2014/02/27/civil-rights-principles-era-big-data/; Access Now, The Toronto Declaration: Protecting the Rights to Equality and Nondiscrimination in Machine-Learning Systems (May 16, 2018), https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/; European Commission High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy Artificial Intelligence (Apr. 8, 2019), https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

[44] *See* Testimony of Rashida Richardson before the Senate Subcommittee on Communications, Technology, Innovation & the Internet of the Senate Committee on Commerce, Science & Transportation, 116th Cong. 7 (June 25, 2019), https://ainowinstitute.org/062519-richardson-senate-testimony.pdf.

[45] *See* Shan Jiang, John Martin & Christo Wilson, *Who's the Guinea Pig?: Investigating Online A/B/n Tests in-the-Wild*, Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency 201-210 (Jan. 2019), https://dl.acm.org/citation.cfm?id=3287565.

particular school district.[46] Ideally, algorithmic systems should be designed with insight about and input from the area and population in which they will be used. For example, a school assignment algorithm designed to work for a rural district where schools are far apart and most students take the school bus may be unreliable for assigning students to schools in a densely populated urban district.

What are the Model's Capabilities and Limitations?

It is also important to understand the limits of any algorithmic system for capturing and addressing a complex issue. For example, college ranking algorithms might measure a range of variables that have a reasonable connection to overall school quality, like class sizes, professor qualifications and salaries, number of classes and extracurriculars offered, and average SAT scores of admitted students, but there may be other relevant factors that the algorithms do not measure, such as students' interactions with one another, or the availability of low-cost housing so that students do not have to work to afford school.[47] Some problems are not appropriate to address with algorithms at all. Statistical models are generally not reliable for measuring highly subjective concepts like a person's state of mind or mental health.[48] Rare events like acts of terrorism or mass shootings cannot be statistically predicted because of their small sample size and the complexity of potential factors surrounding them.[49]

Is it Accurate — and What Do We Mean by Accuracy?

Testing an algorithmic system's accuracy is not always straightforward or scientific. An algorithm that detects whether there is a face in a picture may be fairly simple to validate. Most people can determine whether there is a face in a photo, so the developer can test the algorithm against a set of photos labelled by humans and calculate how often the algorithm got it right. They can also look at what the algorithm got wrong (e.g., did it think a painting in the background was a live human face?) and determine what kind of additional data or tuning would make it better. However, for more subjective determinations, like whether a social media post is bullying or whether a student is at risk of dropping out of school, there may not be any ground truth against which to measure an algorithm's performance. Developers can still attempt to design validation studies, such as testing a bullying algorithm on human-labelled social media posts, but there will always be some subjectivity and thus some uncertainty about the algorithm's accuracy.[50]

---

[46] *See, e.g.*, Duarte, Llansó & Loup, *supra* note 8, at 12–13.
[47] Mary Cunningham & Graham MacDonald, Housing as a platform for improving education outcomes among low-income children, Urban Institute (May 2012), https://www.researchgate.net/profile/Heather_Schwartz/publication/267687704_Housing_as_a_Platform_for_Improving_Education_Outcomes_among_Low-Income_Children/links/546621100cf25b85d17f58d7/Housing-as-a-Platform-for-Improving-Education-Outcomes-among-Low-Income-Children.pdf.
[48] Duarte, Llansó & Loup, *supra* note 8, at 16–17.
[49] *See, e.g.*, *supra* note 30 and accompanying text.
[50] *See, e.g.*, Duarte, Llansó & Loup, *supra* note 8, at 17.

[Transparency, Explanation, and Redress]

Educational institutions and their vendors should be open and transparent about how and when algorithms are used to make decisions affecting students and their families. Transparency serves several purposes. For the general public, transparency can be a means of holding systems accountable in the same manner that public records laws allow the public to observe whether the government is properly carrying out its functions.[51] For the individuals and communities who will be directly affected by an algorithmic system, informing them about the system may be necessary to protect their rights and to seek community input on whether and how the system should be used. Schools and districts using third-party algorithms need to know how the system works in order to use it safely and appropriately.[52] While there are many different types of transparency, the literature and policy around algorithms focuses on the concept of "explanation"—that is, the provision of meaningful information about a system.[53] Meaningful information about an algorithmic system might include, for example, the mere existence of the system, how it is designed to work and to be used, the reasoning behind individual decisions made using the algorithmic system, and the impact of those decisions on individuals. The appropriate content and format of explanations will likely vary depending on whom the explanation is for.

Students and parents should be made aware of algorithmic decisions that stand to impact their lives and opportunities. They should also have access to any explanations or information necessary to effectively assert their legal rights (such as the right to access education records under the Family Educational Rights and Privacy Act (FERPA)),[54] correct any inaccurate records, and challenge decisions that could interfere with their ability to pursue an equitable education in a safe environment. This information should be provided in an accessible format that does not require technical expertise to understand.

In most cases, the use of algorithms by public institutions should be disclosed to the public. Regulators and policymakers may need information from schools and districts in order to perform oversight.[55] Explanations are also critical for informing other users of an algorithmic system about how to properly

---

[51] *See, e.g.*, Robert Brauneis & Ellen Goodman, *Algorithmic Transparency for the Smart City*, 20 Yale J. L. & Tech. 103 (2018), https://www.yjolt.org/sites/default/files/20_yale_j._l._tech._103.pdf.

[52] *See, e.g.*, Timnit Gebru, *Datasheets for Datasets*, https://arxiv.org/abs/1803.09010; Clare Garvie, Garbage In, Garbage Out: Face Recognition on Flawed Data, Georgetown Law Center on Privacy & Technology (May 16, 2019), https://www.flawedfacedata.com/.

[53] The European General Data Protection Regulation provides a right to access "meaningful information about the logic involved" in some decisions involving automated processing. Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC, 2016 O.J. (L 119) 1 (EU), art. 15, https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[54] 34 C.F.R. § 99.10.

[55] *See, e.g.*, Angel Diaz, Brennan Center for Justice, Testimony before the New York City Automated Decision Systems Task Force (May 30, 2019), https://www.brennancenter.org/analysis/testimony-new-york-city-automated-decision-systems-task-force.

use it.[56] Often the person using a system is not the same person who designed it. A system's users need to know its capabilities, limitations, and appropriate uses.

[Fairness and Equity]

Educational institutions have a legal and moral imperative to treat students fairly and protect their rights to an equal education.[57] This includes ensuring that algorithmic systems do not cause discriminatory, marginalizing, or otherwise detrimental impacts on students and their families.

What Do We Mean by Fairness?

There is no singular definition of fairness for the design of algorithmic systems.[58] The definition depends on the system's intended use, its potential consequences, and who will be impacted by it. As a baseline, algorithmic systems should not contribute to or facilitate the marginalization of any group, especially protected or vulnerable classes. Education environments present a number of considerations that should guide the definition and development of fair algorithmic systems.

Because of their age, their dependence on adults, and their inability to opt out of a school's policies and practices, K-12 students as a group should be treated with special care. Algorithmic decision systems, particularly those that involve monitoring students, should not inhibit their personal development (including their ability to express themselves and form relationships) or subject them to public ridicule or unnecessary disciplinary interactions.[59] Data created about students can follow them around and impact their future opportunities,[60] such as their ability to get into college or how they are treated when they transfer schools.[61] Algorithmic systems that result in the creation and retention of negative inferences or derogatory records about students could have lasting harm.

Equal education opportunities are protected by federal civil rights law.[62] Algorithmic systems must be held to the same or higher nondiscriminatory standards as any other education policy or practice, as these systems can sometimes amplify and propagate systematic discrimination.

---

[56] *See* Margaret Mitchell, *Model Cards for Model Reporting*, Conference on Fairness, Accountability, and Transparency (Oct. 5, 2018), https://arxiv.org/abs/1810.03993.
[57] 20 U.S.C. § 1701 et seq.
[58] Arvind Narayanan, Tutorial: 21 fairness definitions and their politics, https://www.youtube.com/watch?v=jIXIuYdnyyk.
[59] *See* CDT & Brennan Ctr., *supra* note 37; Natasha Duarte, *Six Considerations Missing from the School Safety and Data Conversation*, Ctr. for Democracy & Tech. (Mar. 13, 2019), https://cdt.org/blog/six-considerations-missing-from-the-school-safety-and-data-conversation/.
[60] Hannah Quay-de la Vallee, *Deletion and Student Privacy: I Forgot to Remember to Forget*, Ctr for Democracy & Tech. (Dec. 12, 2018), https://cdt.org/blog/deletion-and-student-privacy-i-forgot-to-remember-to-forget/.
[61] Elizabeth Laird & Hannah Quay-de la Vallee, Ctr for Democracy & Tech., Protecting Privacy While Supporting Students Who Change Schools (June 2019), https://cdt.org/insight/protecting-privacy-while-supporting-students-who-change-schools/.
[62] 20 U.S.C. § 1701 et seq.

How Can Algorithmic Systems Exacerbate Inequity?

At least three characteristics of algorithmic systems make them potential vehicles for discrimination. First, as discussed earlier in *Concerns about algorithmic decisions*, they are the product of human judgments. Algorithmic systems embed their developers' biases about what outcomes to optimize for and what features to focus on. For example, several states are creating risk assessment teams to evaluate students for potential safety concerns.[63] The teams are tasked with, among other things, identifying behaviors that raise safety concerns.[64] Because "concerning behavior" is highly subjective, the life experiences and implicit biases of people creating the risk assessments will inevitably influence the targeted behaviors. The targeted behaviors could end up reflecting cultural or social norms that are unfamiliar or unrelatable to the task force members. Because algorithmic systems typically work at a larger scale than individual human actors, and because they lack understanding of nuance and context that might temper human judgments, algorithmic systems may amplify any biases that the human designers introduce into the system.

Second, algorithms rely on observations about datasets, and those datasets can reflect patterns of discrimination or the biases of the people who created or collected the data. Algorithmic systems can amplify the bias reflected in their training data. For example, a model trained to automatically generate labels for images of people doing activities tended to mislabel men as women when the men were pictured in a kitchen.[65] In the training data, women were about 33% more likely than males to be cooking, but the trained algorithm was 68% more likely to label people cooking as females.[66] For under-represented groups, training sets may not include enough data about them to make accurate decisions. An algorithmic system may have a good overall accuracy rate but still perform poorly for some groups. This has occurred in facial recognition systems, which tend to have lower accuracy rates for identifying people of color, young people, and women.[67]

Many aspects of U.S. education systems — especially decisions about where students attend school and how they are disciplined — have racist legacies.[68] Segregationist policies still have lasting impacts on access to quality instruction and resources; a student's neighborhood (and by extension their race and/or economic status) can determine the quality of their education.[69] Students of color tend to be

---

[63] *See, e.g.*, Fla. Laws 2019–22, http://laws.flrules.org/2019/22.
[64] *See id.*
[65] Jieyo Zhao et al., *Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints*, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) (2017), https://arxiv.org/pdf/1707.09457.
[66] *Id.*
[67] *See* Gender Shades: How well do IBM, Microsoft, and Face++ AI services guess the features of face?, http://gendershades.org/; Julia Angwin & Jeff Larson, *Bias in Criminal Risk Scores Is Mathematically Inevitable, Researchers Say*, ProPublica (Dec. 30, 2016), https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say.
[68] *See generally, e.g.*, Advancement Project, *supra* note 33.
[69] Keith Meatto, *Still Separate, Still Unequal: Teaching about School Separation and Educational Inequality*, NY Times (May 2, 2019),

disciplined more often and more severely than white students,[70] are more likely to have law enforcement officers in their schools,[71] and are more likely to have law enforcement interactions both inside and outside of school.[72] Algorithmic systems must be scrutinized to ensure that they do not repeat these patterns and, when possible, should seek to remediate them.

Third, algorithmic systems are generally designed to make decisions at scale. They apply the same general decision-making logic to everyone. While this kind of consistency could combat discrimination under the right constraints, it also means that bias in the algorithm can become quickly systematized.

What Should Schools and Districts Do?

While it is important to understand the shortcomings of algorithmic decision-making systems and be able to identify contexts in which they are not the right approach, there are many situations in which these systems can be a valuable tool. In these cases, it is still important to avoid or mitigate the concerns discussed in this brief. In this section, we discuss some of the ways a school or district can incorporate these systems responsibly when they are designing and procuring algorithmic systems, as well as when they are providing ongoing support and maintenance of these systems. In addition, the appendix provides questions these organizations can ask vendors when procuring such a system to help ensure it manages these concerns. These guidelines and failsafes should be codified in technology vendor or licensing agreements.

[Recommendations for Designing and Procuring Algorithms]

There are a number of steps that designers of algorithmic decision-making systems can take to make the systems more effective and limit the likelihood that those systems will inadvertently cause harm to students.

Assess the System and Document its Intended Use

Each system and intended use will have its own set of potential benefits and risks that must be considered before implementation. Several other organizations, policymakers, and scholars have proposed risk or impact assessment frameworks that could help guide the education sector.[73] Regardless

https://www.nytimes.com/2019/05/02/learning/lesson-plans/still-separate-still-unequal-teaching-about-school-segregation-and-educational-inequality.html.

[70] Moriah Balingit, *Racial disparities in school are growing, federal data show*, Washington 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?utm_term=.0f0a32e9fcae.

[71] *See* Advancement Project, *supra* note 33.

[72] *See Id.*

[73] *See* Dillon Reisman et al., Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability, AI Now Institute (Apr. 2018), https://ainowinstitute.org/aiareport2018.pdf; Andrew Selbst, *Disparate Impact in Big Data Policing*, 52 Ga. L. Rev. 109, 169–78 (2017), https://par.nsf.gov/servlets/purl/10074337; Cory Booker, Booker, Wyden, Clarke Introduce Bill Requiring

of the precise framework used, educational institutions should assess and document the intended outcomes of the system and the risk of harm to students' well being and rights—both if the system performs correctly and if it errs.

Engage Stakeholders Early and Throughout Implementation

Algorithmic decision-making systems often have a wide range of stakeholders. For instance, the set of stakeholders for a threat assessment system may include students, their families, faculty and staff at the school, the school's School Resource Officer, any organizations that feed data into the system such as law enforcement, and mental health providers at the school. Each of these stakeholders will have a different perspective about how the system could work, and, importantly, concerns about the system and whether the system actually addresses the problem to be solved. Soliciting opinions and feedback during the design process can help the designers of the system account for and mitigate these concerns early.

Examine Input Data for Bias

Algorithmic systems rely on data. Unfortunately, this means that any bias exhibited by these data sets will be reflected in the algorithmic system's outputs. This can perpetuate, and even amplify, the existing biases of the data set. System designers should interrogate any input data for bias, such as juvenile justice data where students of color may be overrepresented. If bias is detected, the system must incorporate mitigations, such as giving less-biased data sources more weight, using techniques like oversampling to minimize the effects of bias, or, if possible, avoid the biased data entirely.

Document Best Practices and Guidelines for Future Use

Because of the complexity of many algorithmic systems, it is easy for a new user unfamiliar with the design and capabilities of the system to misuse it. Thus, it is important to carefully document the design and implementation of the system, any best practice guidelines for using it, and, crucially, the limitations of the system and cases where it should *not* be relied on. This documentation should include information about the logic behind the design choices, such as how certain stakeholder engagement influenced those choices. This will help future users of the system ensure that those considerations are carried forward throughout the lifetime of the system, particularly if the system is expanded or altered in the future.

Companies to Target Bias in Corporate Algorithms (Apr. 10, 2019), https://www.booker.senate.gov/?p=press_release&id=903.

Ongoing Management of an Algorithmic System

Once an algorithmic system is deployed, the work has only just begun. In addition to careful system design, algorithmic systems should be carefully managed on an ongoing basis. This management can help make the system effective and safe over time.

Keep Humans in the Loop

Ensuring that humans are still involved in algorithm-driven processes can maintain nuance and context during decision-making processes. An obvious way to keep humans involved is to give a human final say in any decision, such as whether a particular social media post is actually a threat of self harm. However, there are certainly cases where that would be excessive and limit the utility of the algorithmic system. For example, it is probably unnecessary to review every single placement from a school assignment algorithm. It is important to note that there are also cases where human intervention may be insufficient, such as when a threat assessment protocol labels a student as a threat. School officials may then override that determination, and state the student is not, in fact, a threat, but that student may already be on law enforcement's radar, exposing him to unnecessary scrutiny and surveillance. To avoid situations like this, it is important to design the system so that humans can intercede before the student is harmed. In this example, that might mean that the school officials are responsible for reporting the outputs of the threat assessment system to law enforcement, thus giving them a chance to override the decision or to include additional context they feel is relevant, or it could mean that if school officials override the threat assessment determination, law enforcement must also adjust their records to reflect that change.

Implement Data Governance

Algorithmic systems almost always involve the collection, use, and creation of a lot of data. There is the data used to train the algorithm, the inputs analyzed by the trained algorithm, the algorithm's outputs, and any other data created from the outputs or from feedback. All of these datasets must be managed to protect privacy and security, quality decision-making, and data integrity. Privacy and security considerations include how data will be retained, when it will be deleted, who will be able to access it, and how the outputs will be used in the creation of new student records. For example, if a student is identified as at-risk of dropping out of school, how will that information be connected with the student and who can it be shared with? It is important to develop and follow data governance procedures (i.e. the overall management of data, including its availability, usability, integrity, quality, and security[74]) for this information to protect students from the harm of data exposure or misuse. CDT has written

[74] National Center for Education Statistics Institute of Education Sciences, SLDS Issue Brief: Communicating the Value of Data Governance (2017), https://nces.grads360.org/#communities/data-governance/publications/15066.

guidance on deletion and retention,[75] as well as data portability,[76] that may be of use in establishing governance procedures.

Conduct Regular Audits

Another strategy to mitigate potential harms of an algorithmic system is to regularly audit the system to ensure it is performing as expected and not exhibiting significant bias.[77] Auditing can be done internally (by the educational institution or the vendor), externally (by an outside expert), or both. Audits could take a number of forms such as looking at the outputs of the system as a whole to look for trends in the decisions (this approach would be suitable for something like a school assignment system where the district is interested in the movement and distribution of the student population as a whole), or spot-checking individual instances for accuracy (such as looking at individual post flagged by a social media monitoring system to see if it is over-reporting, or looking at students who dropped out to determine if the dropout early warning system is regularly failing to flag at-risk students). Auditing methods should be designed to fit the context of the system. Data sources should also be audited regularly to ensure that they are not feeding bias or inaccuracy into the overall system.

Ensure Ongoing Communication with Stakeholders

In addition to continuing to audit the system, community stakeholders should also be kept in the loop about any algorithmic decision systems that affect them. This means ensuring that newcomers to the school are aware of the systems, their uses, and their data inputs, and also that existing users and subjects of the systems are kept abreast of any changes or expansions to the systems.

Govern Appropriate Uses of Algorithmic Systems

Good governance of algorithmic systems requires limiting them to very specific appropriate uses. Machine learning models are not one-size-fits-all tools. When a model is trained and validated to do a task in a particular domain, such as recognizing cyberbullying among middle schoolers on Snapchat, we cannot assume that it will have similar accuracy rates when used in a different domain, such as high schoolers' Tweets, without adding training data from that domain. In order to maintain fair and high-quality decision-making, algorithmic systems need clear policies about how they can and cannot be used and who can use them.

---

[75] Elizabeth Laird & Hannah Quay-de la Vallee, Center for Democracy & Technology, Balancing the Scale of Student Data Deletion and Retention in Education (Mar. 14, 2019), https://cdt.org/insight/report-balancing-the-scale-of-student-data-deletion-and-retention-in-education/.
[76] Laird & Quay-de la Vallee, *supra* note 61.
[77] Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. Pa. L. Rev. 633 (2017), https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/.

Create Strategies and Protocols for Accountability and Redress

Because there is no way to guarantee error- and bias-free algorithms, they will make mistakes. At the most basic level, accountability means that there is someone a student or parent can contact if they think they have been subjected to an incorrect or unfair decision. When algorithmic systems make mistakes, who is accountable? How can individuals seek corrections or redress? These questions should have concrete, operational answers (i.e., there should be specific actions in place that parents and the school can take in the event that the system errs) and those answers should be documented and communicated both internally and externally as part of a governance plan. There may need to be several layers of accountability; for example, the school or district may be directly responsible for assisting a student or parent, but the system's vendor may also be accountable for investigating whether the student-level issue is indicative of a deeper issue with the algorithmic system, and, if it is, fixing those problems.

Students and families should have access to redress for decisions that interfere with their rights, ability to learn, or educational opportunities. Appropriate redress can vary depending on the context of the decision and the degree and type of harm. It is particularly important to have frameworks in place to provide redress for instances of significant harm to isolated individuals. For example, the results of a behavioral threat assessment can be included as part of students' records, so if a student is labeled as a threat as a result of inaccurate information, it is critical that parents have a right to access and correct that information and have the student's record corrected. Alternatively, if a district is using school assignment algorithms, it is reasonable to expect the district to allow families to correct inaccurate records about them in the system, but it is not necessarily reasonable to expect the district to reverse school assignment decisions mid-year as a form of redress (unless doing so is necessary to preserve a legal right).

Ensure Legal Compliance

An algorithmic system will of course need to comply with any applicable laws. These could include FERPA, civil rights laws, and state and local student privacy and oversight laws, just to name a few. If an algorithm involves monitoring students, it could implicate the First and Fourth Amendments. Several local governments have also passed laws creating task forces or other structures to study and oversee public agencies' use of algorithmic systems. Enforcement of these laws may require schools or districts to meet certain reporting requirements, limit or eliminate certain uses, and/or to cooperate with task force studies. Compliance is likely to be easier with intelligible models than with black box models, since intelligible models allow their users to see exactly what is happening and where there might be compliance issues.

Conclusion

Schools are increasingly turning to algorithmic decision-making systems to assist with complex decisions or simply to expedite or formalize decisions previously made by humans. When implemented and used

correctly, these tools can provide insights not noticeable by humans or improve efficiency. However, they can also introduce or amplify bias, infringe on students' rights, make students feel surveilled, or drain resources. It is important that schools incorporating algorithmic systems pick the right systems and use them in ways that do not infringe on students' rights or expose them to bias.

APPENDIX

A

CENTER FOR DEMOCRACY & TECHNOLOGY

CENTER FOR DEMOCRACY & TECHNOLOGY

Appendix A:

What to Ask Your Vendor: Procuring an Algorithmic Decision-making System

Procuring an algorithmic decision-making system can be daunting as it requires the procurer to evaluate a system that may not fall in their area of expertise. This appendix offers expectations of what vendors may provide as well as presents sample questions that a procuring school or district can ask the vendor to help understand whether to procure an algorithmic decision system and how they can ensure that their system does not inadvertently harm or negatively impact students.

- **Validation and impact assessment**: A vendor should be able to show how well their system works and whether it has disparate accuracy rates that could result in discriminatory outcomes. Vendors should provide documentation showing how they tested their system for accuracy, and the results of the tests (accuracy and error rates), disaggregated across groups that may be at risk of discriminatory outcomes, such as women, people of color, non-native English speakers, LGBTQ+ people, and people from low-income communities. Vendors should demonstrate that they have conducted both quantitative and qualitative analyses of their systems to assess the potential risk to students, including privacy and discriminatory harms, and attest that they have taken steps to mitigate those risks.
    - What are the system's accuracy, false positive, and false negative rates?
    - Are false positives and false negatives evenly distributed across different student groups, such as race and gender?
- **Governance and documentation of appropriate uses**: A vendor should be able to articulate the appropriate and inappropriate uses of an algorithmic system, as well as its limitations, and provide instructions for how to properly use the system. If the vendor makes claims in its marketing materials or sales pitch about what the system can do, those claims should be demonstrated in the vendor's validation studies or other documentation.
    - Under what circumstances can this algorithm be trusted to produce accurate outputs that can be reliably used in school decision-making processes?
    - Under what circumstances does the system's accuracy drop?
    - Are there any foreseeable uses of the system for which the vendor cannot confidently test and report its accuracy?
- **Humans in the loop**: Most algorithmic systems used in education will involve a human decision-maker interpreting, evaluating, and making decisions based in part on an algorithm's outputs. Vendors should provide training and instructions to human decision-makers, whether they work for the vendor or for the educational agency or institution. These materials should be reviewable before procurements are finalized. Training or instruction materials should, in part, help human decision-makers understand the logic behind the algorithm's outputs and how much confidence to place in them.
    - What is the training or educational background of people employed or contracted by the vendor to review the algorithm's outputs?
        - Do the vendor's reviewers demographically represent the relevant student population?
    - What steps does the vendor take to mitigate the potential influence of human decision-makers' racial, gender, or other biases in interpreting algorithmic outputs?

- **Data governance**: There are at least two types of data for which the vendor should have documented policies: the training data used to develop and test the algorithm(s) and any data the vendor receives or retains as a result of the educational institution's use of the algorithmic system (e.g., outputs or feedback data). For training data, the vendor should provide documentation of the data sources, data types, and who is represented in the data, including any groups that are over- or under-represented. If the vendor collects or retains any data from the educational institution's use of the system, the vendor should be able to provide a detailed data governance plan, including retention, deletion, de-identification, access control, and purpose limitation policies and procedures.
    - Is the data used to train the system representative of the student population that will be subject to the system?
    - Will the vendor collect or retain any data that results from the educational institution's use of the system? If so, how will the data be protected?
- **Audits**: The vendor should have a plan and design in place to conduct regular audits of the system to ensure that it is working as expected and producing fair, nondiscriminatory outcomes. Ideally, the vendor should contract with an independent auditor.
    - How often will the vendor audit the system to ensure that it is still working as expected and not producing discriminatory outputs?
- **Communication with stakeholders**: The vendor should provide documentation and explanations of its models that the educational institution can share with key stakeholders, including parents and students, oversight bodies, and potentially the public (particularly if the school is a public institution). Documentation should include explanations of how the system works, its intended and appropriate uses, validation studies, impact assessments, and audit reports. Vendors should also consider attending any school board or other governmental hearings designed to inform stakeholders about the system and answer their questions.
    - Are there any terms in the vendor's contract restricting the educational institution's ability to access or share documentation or information about the system?
- **Accountability and redress**: In addition to ongoing audit provisions, the vendor should have infrastructure and protocols in place to address any issues exposed by the audits, as well as issues brought to their attention by students, families, or other stakeholders.
    - What are the vendor's protocols for responding to audit findings?
    - How does the vendor handle complaints and other external feedback?
    - What channels does the vendor have in place to receive complaints?
    - Who is responsible for handling complaints? Is there a single point of responsibility?