Seeing Like an Algorithmic Error: What are Algorithmic Mistakes, Why Do They Matter, How Might They Be Public Problems?

Mike Ananny[†]

Introduction 1
I. Algorithms and Algorithmic Errors 5
II. Seeing an Algorithmic Error: A Case Study 10
III. Toward a Typology of Algorithmic Errors 17
IV. Algorithmic Errors as Public Problems 20

Introduction

There is a longstanding legal adage that "hard cases make bad law"—that the specific "*this-ness*" of a case limits lawmaking because its facts are so technical, peculiar, or idiosyncratic that its reasoning cannot be generalized.[1] A decision may resolve that particularly challenging issue or conflict, but do little to create broad legal principles or strong public policies for future situations.[2]

Lawmakers, judges, and legal scholars are not alone in struggling with how to move from the particularities of an especially hard case to principles that might resolve similar conflicts. Social scientists, journalists, and activists must similarly decide what to do with a particular harm, offense, or conflict that they discover or are aggrieved by. When is a social transgression something to theorize or use to illustrate a structural force, and when is it an idiosyncratic or insignificant one-off of little general value?[3] The answer often

[†] Associate Professor of Communication and Journalism, University of Southern California.

[1] Frederick Schauer, *Do Cases Make Bad Law?*, 73 U. CHI. L. REV. 883, 884 (2006).

[2] *See* Arthur Corbin, *Hard Cases Make Good Law*, 33 YALE L.J. 78, 78 (1923); *see also* Jeffrey R. Rachlinski, *Bottom-Up Versus Top-Down Lawmaking*, 73 U. CHI. L. REV. 933, 935 (2006) ("The adjudication process necessarily consists of resolving competing arguments—often without compromise, but always with a focus on getting the individual case right. Individual rights in the courts often come at the expense of the public good, at least in the individual case.").

[3] For thoughtful discussions of why to theorize social problems while attending to empirical particulars, see, for example, THEORIZING IN SOCIAL SCIENCE: THE

lies in the people, perspectives, investments, communities, experiences, and assumptions with the power to turn a particular case into a general problem. Activists, scholars, and journalists all have ideas about how the world *should* work—ideas that are often implicit, different, and that change over time, but nonetheless shape what a given era sees as unjust, which injustices are changeable, who is responsible, and how change happens.[4] To be clear, I am not reducing research, activism, or jurisprudence to bald self-interest. But it is both accurate and productive to see that social problems and public attention are always *made*, never found.[5]

Today, this tension between particular cases and general patterns plays out in *algorithmic* errors. How are algorithmic mistakes made, who makes them, and could algorithmic errors be "made" in ways that drive reform? Every time algorithmic systems act—from facial recognition and policing to content moderation and medical diagnosis—they make mistakes. For example, recognition algorithms reduce identities to facial features,[6] policing algorithms reinforce racist surveillance,[7] algorithmic moderation fails to

CONTEXT OF DISCOVERY (Richard Swedberg ed., 2014) and HOWARD S. BECKER, TELLING ABOUT SOCIETY (2007).

[4] For discussions of how the news media decides which social issues to focus on and which injustices attract journalistic attention, see JAMES S. ETTEMA & THEODORE L. GLASSER, CUSTODIANS OF CONSCIENCE (1998) and DANIEL C. HALLIN, WE KEEP AMERICA ON TOP OF THE WORLD 18 (1994).

[5] There is a long-standing and rich body of social theory and empirical research on the forces that construct social and public problems. *See, e.g.*, CELIA LURY, PROBLEM SPACES: HOW AND WHY METHODOLOGY MATTERS (2021); JOSEPH R. GUSFIELD, THE CULTURE OF PUBLIC PROBLEMS: DRINK-DRIVING AND THE SYMBOLIC ORDER (1984); Herbert Blumer, *Social Problems as Collective Behavior*, 18 SOC. PROBLEMS 298 (1971). On the idea that "shocks" to technological systems drive exceptions and frame the scope of problem-solving, see MIKE ANNANY & TARLETON GILLESPIE, PUBLIC PLATFORMS: BEYOND THE CYCLE OF SHOCKS AND EXCEPTIONS (2016).

[6] JESSICA HELFAND, FACE: A VISUAL ODYSSEY (2019).

[7] SARAH BRAYNE, PREDICT AND SURVEIL: DATA, DISCRETION, AND THE FUTURE OF POLICING (2020).

understand speech nuances,[8] and medical diagnoses reflect assumptions about patient populations.[9]

But not all algorithmic mistakes are made in the same way. Some are idiosyncratic one-offs that can be corrected relatively easily while others reveal powerful structural forces that need different kinds of remedies, different theories of change. To know the difference between different types of mistakes, we need to learn to "see like an algorithmic error"—to distinguish among systems, causes, harms, responsibilities, and remedies whenever data-driven, automated systems fail.

My focus is not on the law of algorithmic errors or torts.[10] Instead, I want to use the question of which cases make "good" laws metaphorically to ask which algorithmic errors are "good" mistakes—ones that point to systematic problems that we might think with, design around, regulate against, and use to shape public concerns. To this end, I organize this essay around three questions meant to clarify the meaning and significance of algorithmic mistakes.[11]

First, what, exactly, are algorithmic errors? Using approaches from Science and Technology Studies (STS), I see algorithmic errors as *sociotechnical* constructs—as *relationships* between

[8] TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA (2018).

[9] David Armstrong, *Clinical Prediction and the Idea of a Population*, 47 SOC. STUDS. SCI. 288, 290-91, 298 (2017).

[10] For discussions of legal remedies for algorithmic and robotic harms, see Mark A. Lemley & Bryan Casey, *Remedies for Robots*, 86 U. CHI. L. REV. 1311 (2019); Karni Chagal-Feferkorn, *The Reasonable Algorithm*, 2018 U. ILL. J.L. TECH. & POL'Y 111; Ryan Calo, *Robots as Legal Metaphors*, 30 HARV. J.L. & TECH. 209 (2016).

[11] At the outset, I use "error," "mistake," "breakdown," and "failure" interchangeably in this paper. I do not see these words as synonymous, but their precise differences live in the particularities of algorithmic contexts (different words are better for different events) and in theoretical distinctions between intent, expectation, and anticipation that are beyond the scope of this paper. For an excellent discussion of unintended versus unanticipated consequences of sociotechnical systems, see Nassim Parvin & Anne Pollock, *Unintended by Design: On the Political Uses of "Unintended Consequences"*, 6 ENGAGING SCI., TECH., & SOC'Y 320 (2020)

people and machines that have somehow failed, broken down, behaved in unexpected ways.[12]

Second, how can a seemingly straightforward algorithmic mistake generate new ways to see how algorithmic breakdowns usually have different causes, significances, and remedies, depending on how you understand algorithmic systems? Using the story of a recent algorithmic error in remote proctoring software, I show how algorithmic mistakes are not found but *made*. They are made by people deciding how broadly to see a sociotechnical system, the forces that create it, and the factors that could be behind its breakdown. These choices about how to define an algorithm and its error define what I call "seeing like an algorithmic error".

Finally, how can different ways of "seeing like an algorithmic error" suggest different ways to cast algorithmic breakdowns as public problems[13]? To see an algorithmic mistake as something that creates shared consequences and needs public regulation—versus as an idiosyncratic quirk requiring private troubleshooting—is to see algorithmic errors in ways that resist the individualization and privatization of failures. It is to understand them as systematic,

[12] My use of the term "sociotechnical" grows out of Science and Technology Studies (STS) work that insists upon seeing phenomena as inextricable relationships between social and technological forces. I.e., technologies are not neutral tools used by people with good or bad intentions—rather, whenever people and computational systems meet (in everything from the engineering cultures that make Facebook's Newsfeed to the people who use Siri's voice recognition) there are collisions between what people think they are, do, and could be, and what systems are, do, and are thought to be. For an introduction to this field, see SERGIO SISMONDO, AN INTRODUCTION TO SCIENCE AND TECHNOLOGY STUDIES (2d ed. 2009) and Harry Collins, Robert Evans & Martin Weinel, *STS as Science or Politics?*, 47 SOC. STUDS. SCI. 580 (2017).

[13] I borrow this phrase from a host of studies showing the analytical and empirical power of adopting a critical perspective on a variety of social constructs like states, markets, algorithms, surveys, and infrastructures. *See, e.g.*, Nick Seaver, *Seeing Like an Infrastructure: Avidity and Difference in Algorithmic Recommendation*, 35 CULTURAL STUDS. 771 (2021); Rebecca Uliasz, *Seeing Like an Algorithm: Operative Images and Emergent Subjects*, 36 AI & SOC'Y 1233 (2020); Marion Fourcade & Kieran Healy, SEEING LIKE A MARKET, 15 SOCIO-ECON. REV. 9 (2016); John Law, *Seeing Like a Survey*, 3 CULTURAL SOCIO. 239 (2009); JAMES C. SCOTT, SEEING LIKE A STATE (1999).

structural breakdowns that reveal normative investments and demand interventions on behalf of collectives.

Just as some legal cases may make "good" laws—because they advance legal principles and enrich jurisprudence—some algorithmic errors may make "good" public problems. If the "this-ness" of an algorithmic error is unjust conditions that people cannot avoid, then the algorithmic error can be made into a *public* problem. When algorithmic errors are public problems, they are not idiosyncratic quirks for software companies to debug privately and on their own timeline. They are instead powerful provocations showing—exactly—how a system has failed, why it has failed, what its successful operation would look like, who benefits from its failures, and how reformers can fix the mistake, remedy the harms, and prevent future errors. As I try to show in this essay, "seeing like an algorithmic error" means turning seemingly simple quirks and individually felt glitches into shared social consequences with the power to shape social life—that is, into public problems.[14]

I. Algorithms and Algorithmic Errors

By "algorithms" I mean more than computational instructions that transform data from one state to another. Though this is a common and largely uncontroversial way to see algorithms that has long dominated computer science and engineering practices, more recent academic research, popular press, and regulatory efforts rightly take a more expansive view of "algorithms." They are indeed computational instructions that live in machine code, but they are also drivers of surveillance cultures that feed on and create vast

[14] There is a rich emerging literature on "glitches" and "errors" in media technologies and data-driven systems, including how data messiness, mistakes, and misinterpretations show how particular people think systems work and should work. *See, e.g.*, Nanna Bonde Thylstrup, *Error*, *in* UNCERTAIN ARCHIVES: CRITICAL KEYWORDS FOR BIG DATA 191, 193-94 (Nanna Bonde Thylstrup, Daniela Agostinho, Annie Ring, Catherine D'Ignazio & Kristin Veel, eds., 2021); Rebecca Schneider, *Glitch*, *in* UNCERTAIN ARCHIVES: CRITICAL KEYWORDS FOR BIG DATA 259, 266-67 (Nanna Bonde Thylstrup, Daniela Agostinho, Annie Ring, Catherine D'Ignazio & Kristin Veel, eds., 2021); Lisa Gitleman, *Misreading*, *in* UNCERTAIN ARCHIVES: CRITICAL KEYWORDS FOR BIG DATA 346, 347-52 (Nanna Bonde Thylstrup, Daniela Agostinho, Annie Ring, Catherine D'Ignazio & Kristin Veel, eds., 2021).

amounts of data.[15] Especially in the context of machine learning and artificial intelligence, algorithms' stability and reliability grow out of seemingly objective statistical models and tests that rest upon histories of training data, politics of categorization, and planetary energy resources.[16] They are produced in response to commercial demands for faster, more fine-grained, and more powerful ways to classify consumer preferences and behaviors.[17] And they drive systems that both analyze and process language, creating descriptions of the world that people use to reflect upon their identities, communicate with others, and create public life.[18] Algorithms are both "traps" that sequester people in particular cultural worldviews,[19] and "societies" that transform how "people interact, associate, and think."[20] They simultaneously give people options for what to do, and signal what people are expected to do and what most people do.

But algorithms fail, in different and intertwined ways. They rely on incomplete datasets, partial categorizations, inaccurate and unjust assumptions, extractive business models, reductionist understandings of identity and culture, and generally odious aesthetics about the human value of automation. Because "algorithms" are almost everywhere and have such complex dynamics, we need to be "precise in our outrage"[21] at their failures. Once you start noticing them, algorithmic errors are almost everywhere and increasingly frequent, but they are usually hard to neatly categorize into discrete causes and harms. Attempts to do so

[15] *See generally* SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM (2019).

[16] KATE CRAWFORD, ATLAS OF AI 14-17 (2021).

[17] FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015).

[18] Tarleton Gillespie, *The Relevance of Algorithms*, *in* MEDIA TECHNOLOGIES: ESSAYS ON COMMUNICATION, MATERIALITY, AND SOCIETY 167, 167 (Tarleton Gillespie, Pablo J. Boczkowski & Kirsten A. Foot eds., 2014).

[19] Nick Seaver, *Captivating Algorithms: Recommender Systems as Traps*, 24 J. MATERIAL CULTURE 421, 425-27 (2019).

[20] Jenna Burrell & Marion Fourcade, *The Society of Algorithms*, 47 ANN. REV. SOCIO. 213, 213 (2021).

[21] Karen Levy, *The Case for Precise Outrage*, DATA & SOC'Y (Feb. 2, 2016), https://points.datasociety.net/the-case-for-precise-outrage-407884d2d3b5#.mkqjm2xc8.

usually show not only the expansiveness of the algorithm system, but also a critic's particular political investments.

To take a few recent examples: In January 2020, in front of his wife and young daughters, the Detroit Police Department handcuffed Robert Julian-Borchak Williams, detained him for 30 hours, and required him to post a personal bond before arraignment, all because a racially discriminatory facial recognition system mistook him as a local robber.[22] When researchers reviewed over 600 machine learning models and tools developed to help medical professionals diagnose Covid-19 patients and predict illness severity, they found that not one of the systems was clinically useful and actually, through a series of training and testing errors, many systems may have harmed patients.[23] In April 2021, the UK Air Accidents Investigation Branch discovered that the airline TUI had been systematically miscalculating flight loads because its software automatically classified passengers registered as "Miss" as children (weighing approximately 77 lbs) and not adults (weighing approximately 152 lbs), attributing the error "to cultural differences in how the term Miss is understood."[24] And in 2021, Twitter apologized for errors in its image cropping system's "saliency algorithm" after an internal audit found that the algorithm relied on datasets of human eye movements to see images of white people as more salient than Black people.[25] The list could go on: the Partnership on AI even maintains an "Artificial Intelligence Incident

[22] Kashmir Hill, *Wrongfully Accused by an Algorithm*, N.Y. TIMES (June 24, 2020), https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html.

[23] Will Douglas Heaven, *Hundreds of AI Tools Have Been Built to Catch COVID. None of Them Helped*, MIT TECH. REV. (July 30, 2021), https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/.

[24] Thomas Claburn, *Airline Software Super-Bug: Flight Loads Miscalculated Because Women Using 'Miss' Were Treated as Children*, THE REGISTER (Apr. 8, 2021), https://www.theregister.com/2021/04/08/tui_software_mistake/.

[25] Rumman Chowdhury, *Sharing Learnings About Our Image Cropping Algorithm*, TWITTER BLOG (May 19, 2021), https://blog.twitter.com/engineering/en_us/topics/insights/2021/sharing-learnings-about-our-image-cropping-algorithm.

Database" of over 1700 "unforeseen and often dangerous failures" of machine learning systems.[26]

All these examples—from sentencing, medical diagnosis, transportation logistics, and social media—show how common algorithmic errors are in so many aspects of life. They also point to what sociotechnical scholars of algorithmic systems have argued for years: that algorithmic systems are not just computational code, but intertwined and often invisible assemblages of people, classifications, calculations, institutions, risks, and values.[27] To say that an algorithm failed or made a mistake is to take a particular view of what *exactly* has broken down—to reveal what you think an algorithm is, how you think it works, how you think it should work, and how you think it has failed.

For some people, the algorithmic system may not have failed at all and is behaving as intended and properly enabling a particular worldview. As Louise Amoore argues, algorithmic outcomes "that might appear as errors or aberrations are in fact integral to the algorithm's form of being and intrinsic to its experimental and generative capacities."[28] In another view, underpinning the Partnership on AI's "AI Incident Database", algorithmic errors are "unforeseen and often dangerous failures" that can do harm when "deployed to the real world."[29] Scholars, practitioners, and regulators alike seem unclear on what algorithmic errors *are*. Are they unforeseen malfunctions, unavoidable side-effects of "permanently beta" software cultures,[30] statistical calculations of

[26] *About*, AI INCIDENT DATABASE, https://incidentdatabase.ai/about (last visited Aug. 29, 2022).

[27] *See, e.g.*, Burrell & Fourcade, *supra* note 20, at 221-226; Nick Seaver, *Algorithms as Culture*, 4 BIG DATA & SOC'Y 1, 4-5 (2017); Mike Ananny, *Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness*, 41 SCI., TECH. & HUM. VALUES 93, 98-99 (2016); Gillespie, *supra* note 18, at 179-82.

[28] LOUISE AMOORE, CLOUD ETHICS: ALGORITHMS AND THE ATTRIBUTES OF OURSELVES AND OTHERS 23 (2020).

[29] *Supra* note 26.

[30] Gina Neff & David C. Stark, *Permanently Beta: ResponsiveOorganization in the Internet Era*, in SOCIETY ONLINE: THE INTERNET IN CONTEXT (P. Howard & S. Jones eds., 2004).

probability and acceptable risk,[31] distributions of responsibility between people and machines,[32] or political choices about which technological consequences to anticipate and preempt and which to label unknowable and thus "unintended"?[33]

If algorithms are computationally calculated, institutionally produced, culturally meaningful sociotechnical constructions, then so are their errors. But while it is now commonplace to call out "bias" in algorithmic systems—highlighting incorrect and unjust results[34] and suggesting technical interventions[35]—makers, scholars, regulators, and targets of algorithms would benefit from more precisely defining, classifying, and triaging algorithmic errors. Instead of trying to say exactly what an algorithmic error is or is not, a more pragmatic approach asks what is at stake in seeing an error in a particular way? If algorithms and their errors can be described in so many different ways, we can more generatively ask how and why certain people see an algorithmic event as an error—or mistake, failure, breakdown, glitch, bug, unanticipated consequence, unforeseen outcome, necessary step for innovation—while others see no error at all, just a system working as intended.

To "see like an algorithmic error" means seeing a sociotechnical scene expansively, creatively, and with a degree of

[31] Mike Ananny, *Probably Speech, Maybe Free: Toward a Probabilistic Understanding of Online Expression and Platform Governance*, KNIGHT FIRST AMENDMENT INST. (August 21, 2019), https://knightcolumbia.org/content/probably-speech-maybe-free-toward-a-probabilistic-understanding-of-online-expression-and-platform-governance.

[32] Madeleine Clare Elish, *Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction*, 5 ENGAGING SCI., TECH., & SOC'Y 40, 41 (2019).

[33] Parvin & Pollock, *supra* note 11, at 323-24.

[34] *See, e.g.*, SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM (2018); Lucas D. Introna & Helen Nissenbaum, *Shaping the Web: Why the Politics of Search Engines Matters*, 3 INFO. SOC'Y 169 (2000); Susan Leigh Star & Martha Lampland, *Reckoning with Standards*, *in* STANDARDS AND THEIR STORIES 3, 6-7 (Martha Lampland and Susan Leigh Star eds., 2008).

[35] *See, e.g.*, Christian Sandvig et al., *An Algorithm Audit*, *in* DATA AND DISCRIMINATION: COLLECTED ESSAYS 6, 8-9 (Seeta Pena Gangadharan ed., 2014); Timnit Gebru et al., *Datasheets for Datasets*, 64 COMMC'NS ASS'N COMPUTING MACH. 86, 88-91 (2021); Margaret Mitchell et al., *Diversity and Inclusion Metrics in Subset Selection*, 20 PROC. 2020 CONF. ON A.I., ETHICS, AND SOC'Y AAAI/ACM 117, 121-22 (2020).

detachment and analytical humility that acknowledges errors as coming from many different forces, value systems, and calls for remedies. While algorithmic errors may understandably fuel quick outrage and political entrenchments, they might also be opportunities to think beyond a single instance and idiosyncratic harm, to recast algorithmic errors as public problems with complex structural dynamics that go beyond any single perspective or normative investment.

II. Seeing an Algorithmic Error: A Case Study

To illustrate how complex and fraught it can be to "see like an algorithmic error", I want to tell a personal story about my own experiences as part of a task force my institution created to provide guidance on the use of information technologies for online student assessment. Throughout this story I use "error" in an expansive way, illustrating how the failures of an electronic proctoring system could be understood as, among other things, technical mistakes in a facial detection system, institutional failures to ensure that a system represents pedagogical values, and economic forces that budget a certain amount of error in exchange for pedagogical scale. The purpose here is to show how—depending on how you understand an algorithmic system—different types of errors within it will be more or less acceptable or alarming.

Like many universities, when the Covid-19 pandemic moved our school to online instruction in March 2020, we were faced with an urgent need to address a host of challenges. Some of these had been percolating for years and were well understood by many, while others were appearing for the first time, or at least taking on a newfound urgency.

Our task force was specifically charged with considering privacy issues associated with using online tools like Zoom, Blackboard, and Respondus to create online environments and assess student learning. Some academic programs had been using these and similar technologies for years while others were encountering them for the first time, with many students and faculty alike adjusting their expectations of teaching and learning almost overnight. While our task force's initial discussions focused on privacy questions associated with contact tracing apps and reporting medical symptoms, in response to widespread news and social media reports, we were asked to investigate the possibility that our system for proctoring students' online exams was systematically treating students of color differently than other students.

While a complete review of the functions, deployments, failures, and resistances against remote proctoring tools (RPTs) is beyond the scope of this paper,[36] in general, an RPT is software that universities require students to install on their home computers and use during timed examinations, in place of the human proctoring that would normally be used to supervise in-person exams. Though there are variations among their products, several companies (e.g., ExamSoft, Respondus, Proctorio, ProctorU, and Honorlock) have developed tools that: lock a student's computer to make available only a particular web browser or screen; monitor students' keystrokes and mouse movements for "suspicious" behavior that might signal cheating; use the cameras and microphones of students' computers to listen for ambient sounds like whispered answers, watch backgrounds for any suspicious movements, and monitor students' faces, head movements, and eye gazes for any expressions, motions, or fixations that the software defines as indicative of cheating. The software identifies these supposedly suspicious actions, sounds, and motions only because it has been "taught" to see them as indications of cheating through machine learning techniques. These techniques classify what they observe according to patterns represented in datasets, data training, and computational models of unacceptable behavior. As the makers of such tools are

[36] For critical scholarship and explanatory reporting on the use of remote proctoring systems, see Britt Paris, Rebecca Reynolds & Catherine McGowan, *Sins of Omission: Critical Informatics Perspectives on Privacy in e-Learning Systems in Higher Education*, 73 J. ASS'N INFO. SCI. & TECH. 708, 719-20 (2021); Nora Caplan-Bricker, *Is Online Test-Monitoring Here to Stay?*, NEW YORKER (May 27, 2021), https://www.newyorker.com/tech/annals-of-technology/is-online-test-monitoring-here-to-stay; Todd Feathers, *Schools Are Abandoning Invasive Proctoring Software After Student Backlash*, VICE (Feb. 26, 2021), https://www.vice.com/en/article/7k9ag4/schools-are-abandoning-invasive-proctoring-software-after-student-backlash; Drew Harwell, *Cheating-detection Companies Made Millions During the Pandemic. Now Students Are Fighting Back.*, WASH. POST (Nov. 12, 2020), https://www.washingtonpost.com/technology/2020/11/12/test-monitoring-student-revolt/; Shea Swauger, *Software that Monitors Students During Tests Perpetuates Inequality and Violates Their Privacy*, MIT TECH. REV. (Aug. 7, 2020), https://www.technologyreview.com/2020/08/07/1006132/software-algorithms-proctoring-online-tests-ai-ethics/.

quick to stress, these tools do not *decide* that a student has cheated, they simply *identify* patterns that they argue are statistically correlated with cheating, leaving schools to investigate and decide whether an event the software flags is indeed cheating.

These systems are problematic for many reasons, as popular press accounts and social media complaints document. Students, especially those in shared living situations, often cannot create the kind of silent and visually static environments that such proctoring systems expect; family members and roommates may enter the exam scene for reasons unrelated to cheating. Students who think by habitually looking up or away for any length of time may be flagged as potential cheaters more than students who train themselves to stare at the camera. Because proctoring systems often do not allow the use of virtual backgrounds, students are forced to reveal their home environments to the camera and, potentially, a professor investigating a possible instance of cheating. Though universities and companies stress that such data is anonymized and only used in the aggregate to improve an algorithm's accuracy, students using these systems are effectively forced to submit their keystroke, mouse, audio, and video data to machine learning datasets. There is usually no way to opt out of remote proctoring and still take a test.

The question that our task force was asked to consider was whether the facial detection system that our university's remote proctoring system used to track students' head movements and eye gazes systematically treated students of color differently from other students. Though it was not the software we used, ExamSoft was publicly criticized for telling students of color that they should take extra steps to make sure that they were properly illuminated. They told students to front-light themselves and be sure to hold their heads especially still, to avoid having their exams flagged for review. We knew that many remote proctoring systems used similar facial recognition systems (competitors often use the same off-the-shelf datasets, computational models, and pattern-matching algorithms) and, indeed, we confirmed that our vendor's remote proctoring system had a higher error rate for dark-skinned versus light-skinned students. They similarly suggested that students of color should front-light themselves and be especially careful to minimize head movements. Our algorithm was systematically treating our students of color differently than our other students. Since our university aimed to treat all students equally, we were arguably failing.[37]

Our task force's first step was to describe the error. In the narrowest sense, the error was *technically* in the part of the remote proctoring system that detected the *presence* of a face (the system did not recognize *particular* faces). Our vendor assured us—as many machine learning designers often do—that, with more data and better computational models, the system would, over time, detect potential instances of cheating equally, regardless of skin color. They promised a software update that they said would improve the system's accuracy. However, they also said that we would not be able to have this improvement independently verified. It was unclear whether we would be told the new error rate, what thresholds would be used to train the new model, or what fraction of potentially cheating students were students of color.

We also located the error in the system's *expectation* that students were taking exams in environments that were free of audio and visual distractions—the type of environments we had previously created for them in on-campus rooms but that we did not create for them in remotely proctored exams. Yes, the system's facial detection system treated students of color differently from other students; but it *also* treated students differently depending on whether they were able to create quiet and visually static test-taking

[37] There is an increasingly large body of scholarship on systematic biases of facial detection and recognition systems, and the misuse of artificial intelligence technologies to further oppression of historically marginalized and disempowered groups. *See, e.g.*, Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACH. LEARNING RSCH. 1, 10-12 (2018); CRAWFORD *supra* note 16, at 109-11; RUHA BENJAMIN, RACE AFTER TECHNOLOGY (2019); Sasha Costanza-Chock, *Design Justice, A.I., and Escape from the Matrix of Domination*, J. DESIGN & SCI. (July 18, 2018), https://jods.mitpress.mit.edu/pub/costanza-chock/release/4. Additionally, as this paper was going to press, a federal judge ruled that Cleveland State University violated the 4th Amendment when it used electronic proctoring software "to virtually scan the bedroom of a chemistry student before he took a remote test," *see* Amanda Holpuch & April Rubin, *Remote Scan of Student's Room Before Test Violated His Privacy, Judge Rules*, NY TIMES (Aug. 25, 2022), https://www.nytimes.com/2022/08/25/us/remote-testing-student-home-scan-privacy.html.

environments, an ability that we suspected correlated with a student’s socioeconomic status.

We also questioned whether the system was failing in its approach to identifying and investigating potential instances of cheating. If our remote proctoring system was flagging students of color and (potentially) lower-income students for potential academic violations at higher rates than their counterparts, were our faculty sufficiently aware of the structural forces driving such flagging, and their own implicit biases, to consider each potential case of cheating justly? What would faculty think if a supposedly neutral algorithm repeatedly found that Black students “cheated” more than others?

Even more broadly, though it was beyond the scope of our task force, the error was also in the economic models that drove the university to run large classes that tended to use such standardized forms of assessment and student surveillance. Indeed, the problem was not confined to remote proctoring of timed exams; our university also used plagiarism detection software designed to quickly make statistical judgments about the likelihood that a student’s written work was not their own. We had no official, campus-wide student honor code for exams and instead relied on forms of assessment and surveillance that some faculty had long abandoned as pedagogically ineffective, but that others relied upon and saw as integral to ensuring academic integrity. In many ways, the failures of the remote proctoring algorithms simply highlighted larger institutional challenges: the university’s business model needed large classes that used smallest amount of labor possible for standardized forms of assessment that could be scaled, replicated, and audited relatively easily. The economics and pedagogical rationales of core parts of the university already fit perfectly with the promises of remote proctoring software.

So where, exactly, was the error? It was most certainly in the datasets, models, and machine learning systems that collectively treated students of color differently from others. It was partly in our failure to ask the question when we first licensed the software, the vendor’s failure to discover or disclose the error at the outset, and the advice that dark-skinned students could “fix” the system for themselves by shining bright lights onto their faces while taking exams.[38] The error was also in the industry-wide infrastructures, machine learning cultures, and business models that propagated so easily amongst so many remote proctoring companies. There was something wrong with a technology industry that seemed to so easily and uncritically share datasets, machine learning designs, and discriminatory troubleshooting recommendations.

The error was in our university and many universities like ours, but it was an error that resisted easy solutions. Though many faculty had long abandoned standardized, timed, large-scale testing as the gold standard of student assessment, other faculty still subscribed to it, argued for its value, and drew large classes and tuition revenue with relatively small marginal costs. We had intertwined economic models, pedagogical theories, and models of academic integrity in ways that made us rely upon systems that promised—at once—organizational efficiency, standardized assessment, academic integrity, and brand protection. How could we and other universities *not* use remote proctoring systems? Though we lacked evidence on this point, we also questioned whether all of our faculty and teaching assistants could be educated quickly enough on how systematic racism and implicit bias can appear in seemingly neutral algorithmic systems, so investigators could see discriminatory patterns among students flagged as potential cheaters. Broader cultural fixes were needed, and those would take time.

The Provost publicly accepted our task force's recommendation to "discontinue the use of Respondus Monitor, an online exam proctoring program that uses artificial intelligence . . . [due to] a number of concerns about fairness and privacy."[39] Additionally, the Provost directed our university's "Center for Excellence in Teaching" to support faculty impacted by this discontinuance.

[38] This advice to Black students that they should illuminate themselves is especially tragic in the context of "lantern laws in eighteenth-century New York City that mandated enslaved people carry lit candles as they moved about the city after dark." SIMONE BROWNE, DARK MATTERS: ON THE SURVEILLANCE OF BLACKNESS 11 (2015). Black people were again being told to take personal responsibility for making themselves visible to systems of surveillance and control.

[39] Letter from Charles F. Zukoski, Provost, Univ. S. California, to Univ. S. California Faculty (Jan. 26, 2021), https://www.provost.usc.edu/spring-2021-update.

The policy decision meant that none of our students would be subjected to the facial detection system that had initially prompted our task force's review. And, to be clear, our task force found no evidence of harm or discrimination; ours was a relatively responsible preemptive attempt to improve our university's electronic proctoring. The rationale that dominated our recommendation narrowly focused on the system's *unreliability*. It could not *reliably* classify students' actions, treat students equitably, lead to determinations of cheating. While this rationale was true and helped spur a policy that removed the particularly inequitable condition, the focus on reliability was limiting. Our university still uses plagiarism detection software, we still rely on large classes as sources of revenue, faculty autonomy still allows for surveilled student assessment, and we took no position on discriminatory machine learning.

Our task force discussions touched upon technical architectures, policy frameworks, procurement commitments, pedagogical theories, and economic models—but the "fix" that the administration announced was motivated by a technical rationale (a lack of system reliability) and appeared as a limited decision (discontinuing the use of one system). By seeing and fixing the error this way but not some other way, our institution signaled what parts of an algorithmic error it thinks are salient and controllable, and how big a fix is it could imagine and implement.

The task force's version of "seeing like an algorithmic error" was limited to a narrow technical and policy sense, and not illustrative of the larger concept described here. We could have seen the error and fixed it differently. We could have seen the error as an indication that our educational mission was incompatible with mass student surveillance and discontinued the use of all proctoring systems and plagiarism detectors. We could have seen the error as an indication that our classes were simply too big to be good learning and assessment environments, reducing class sizes until they allowed for evaluations that did not need automated surveillance. We could have seen the error as a challenge to the very idea of supervised testing, implementing an honor code that would remove any requirement to observe student test-taking. We did none of these things and saw the error as a technological bug that could be fixed by discontinuing the use of one feature of one surveillance system.

III. Toward a Typology of Algorithmic Errors

This case and the examples I highlighted at the outset prompted different corrections and remedies. The Wayne County prosecutor's office apologized to Williams for his arrest and detention and claimed that facial recognition evidence alone should never prompt police to act.[40] In response to critical scholarship and investigative journalism, the World Health Organization (WHO) increased advocacy for "emergency data-sharing contracts" that would force the makers of machine learning systems to share data and models during international crises. The TUI airline updated their software and now says that its flight load calculations will rely on passenger age and not assumptions about the marital status of female passengers. And, in response to the public outcry over the cropping algorithm's race-based differences, Twitter announced that they would discontinue the algorithm and instead offer users a way to manually select image previews. My own institution announced that it would discontinue use of the remote proctoring system's real-time monitoring and offer faculty support for alternative forms of assessment.

These corrections and remedies show just how differently algorithmic errors can be seen, how embedded they are in systems of prediction and control, and how much seemingly technical fixes are always intertwined with larger questions of ethics, institutional mission, and normative ideals. Knowing that both facial recognition and incarceration systems disproportionately mistreat Black American men, why is the algorithm still being used at all? Why do data scientists need the WHO to force action instead of rejecting the secret and proprietary practices that lead to unaudited and unaccountable machine learning systems? What other gender-based assumptions underpin seemingly innocuous enterprise software categories, and how might TUI have used its algorithmic error to drive broader change about the politics of data labeling? Why does it take user outcry and journalistic pressure to force an internal audit of a social media platform, and why did independent researchers not have access to the Twitter infrastructures and cultures that produced

[40] Press Release, Wayne County Prosecutor's Office, WCPO Statement in Response to New York Times Article *Wrongfully Accused by an Algorithm* (June 24, 2020), https://int.nyt.com/data/documenthelper/7046-facial-recognition-arrest/5a6d6d0047295fad363b/optimized/full.pdf.

the cropping algorithm to begin with? And, if remote proctoring systems can “fix” their facial detection algorithms to surveil all skin colors equally well, will universities use that reliability as a reason to continue with classroom sizes, economic models, and surveillance-based assessment, without ever questioning why the university needed the system in the first place?

By examining the forces that create algorithmic errors and why some people find some fixes acceptable, we can start to build a typology of algorithmic error that shows not only what errors are, but why they matter and what their fixes and ameliorations reveal.

For example, if an algorithmic system’s breakdown is seen as the product of “biased datasets,”[41] and responded to with larger or more diverse datasets, then the “mistake” was seen as a failure to include as many people as possible in a dataset. It leaves little room to see inclusions as also potentially harmful, or to question whether a dataset can ever be a “complete” image of human identity or behavior. And it leaves little room to ask whether the system should exist at all. A criminal sentencing algorithm built on a dataset that includes all people ever incarcerated may be considered “complete,” but if it reinforces racist incarceration patterns it leaves little room to see the error as part of histories of discriminatory policing, underinvestment in racialized communities, media depictions of criminality, or to question whether prisons should exist at all. The algorithm is “successful” because the scope of its error is contained to the completeness or “biases” of the dataset that trained it, and the assumption that the institution it serves is acceptable.

If an algorithmic system’s breakdown is seen as a third-party developer misusing an algorithmic infrastructure to create a problematic derivative—as Amazon claimed when the ACLU used the company’s Rekognition facial recognition system to argue that

[41] For a critical discussion of the limits of using “bias” as a way to frame data-based injustices, see, for example, Anna Lauren Hoffmann, *Data Violence and How Bad Engineering Choices Can Damage Society*, MEDIUM (Apr. 30, 2018), https://medium.com/s/story/data-violence-and-how-bad-engineering-choices-can-damage-society-39e44150e1d4 and Anna Lauren Hoffman, *Terms of Inclusion: Data, Discourse, Violence*, 23 NEW MEDIA & SOC’Y 3539, 3546-47 (2020).

the system was racist[42]—then the breakdown is seen as a failure to properly understand a toolkit, use recommended error thresholds, and appreciate a machine learning's statistical properties. Blaming a third-party developer for misunderstanding an algorithmic architecture leaves little room to question which default error thresholds are acceptable, who has the power to set such scales, and whether a system should be deployed with systematically uneven error profiles.

An error may also be framed as a necessary step for improving an algorithmic system—as Tesla claimed when it defended the "Insane" and "Ludicrous" versions of its self-driving car software that allowed for more aggressive acceleration and lane changing.[43] The increased risks and potentials for failure, the company claimed, were part of its attempt to improve its autopilot systems. In this case, an error is not seen as an error at all. It is a responsible engineering strategy to improve software by deploying mistake-prone systems and using errors to illustrate system limitations. The public forced to contend with such "insane" and "ludicrous" cars is enrolled as unwilling participants in an experiment that is designed to have errors. In this case, the company offers no fix because the error is a desired outcome and key to an algorithm's improvement.

If you look at these examples as moments when people "see like an algorithmic error," an analytical approach begins to emerge. Depending on how an algorithmic error is framed and responded to, different *parts* of a sociotechnical system seem more or less alterable, worthy of reform, or able to be altered. And different people seem more or less acceptable as victims of algorithmic mistakes. It is easier to make a seemingly biased dataset larger and more inclusive than it is to question whether algorithmic surveillance capitalism is okay. It is easier to blame a third-party developer for misusing a toolkit than it is to trace which default error thresholds are ethical. It is easier to think that algorithmic systems

[42] Davey Alba, *Amazon Rekognition Falsely Matched 28 Members of Congress with Arrest Mugshots*, BUZZFEED NEWS (July 26, 2018), https://www.buzzfeednews.com/article/daveyalba/amazon-rekognition-facial-recognition-congress-false.

[43] Faiz Siddiqui, *Tesla Tempted Drivers with 'Insane' Mode and Now Is Tracking Them to Judge Safety. Experts Say It's Ludicrous*, WASH. POST (Oct. 10, 2021), https://www.washingtonpost.com/technology/2021/10/10/tesla-full-self-driving/.

can only improve through experimental encounters with the world than it is to question which people should be subjected to algorithmic experiments, and what informed consent for such subjugation requires. It is easier to ask students of color to illuminate themselves, to ask software companies to make a technical fix, or to stop using a software feature than it is to ask how a learning model requiring large-scale, standardized student surveillance balances ethical pedagogy and faculty autonomy.

To "see like an algorithmic error" means engaging with the myriad social, technological, economic, cultural, and political forces that make algorithms, asking which collision of forces have failed, and being honest about just how far and how fast remedies can go.

IV. Algorithmic Errors as Public Problems

If algorithmic failures are the product of "broken world" relationships,[44] we might better see how algorithmic errors can be both diagnostic and generative. They can be diagnostic because, by understanding how people see algorithmic errors as *mis*arrangements of people and computation, we can get a better sense of how they define *proper* arrangements—which errors they can recognize, anticipate, prevent, tolerate, distribute, explain, ameliorate, avoid, and resist. Being precise in diagnosing an error can be generative because it can reveal *communities of algorithmic error*—people who see and diagnose errors similarly, who strive for fixes together. But such communities of interpretation might achieve a kind of precision as they collide with people who think about, experience, relate to, and call attention to errors differently. Some errors may be highly visible and salient to some affected communities while others may never be seen or felt at all. If we can see how algorithms fail differently for different people, we can see fault lines of inequality—how errors and their harms are unevenly distributed and reinforce power imbalances. This lets us more accurately call attention to the causes and scopes of computational injustices.

[44] Sarah Sharma, *A Manifesto for the Broken Machine*, 35 CAMERA OBSCURA 171 (2020); Steven J. Jackson, *Rethinking Repair*, *in* MEDIA TECHNOLOGIES: ESSAYS ON COMMUNICATION, MATERIALITY, AND SOCIETY 221, 221-22 (Tarleton Gillespie, Pablo J. Boczkowski & Kirsten A. Foot eds., 2014).

Seeing like an algorithmic error might help create *public problems*, instead of idiosyncratic failures or technical missteps. Following scholars like Dewey,[45] Marres,[46] Gusfield,[47] and Napoli,[48] public problems are never found; they are always *made*, through communities of interpretation and technological conditions of the day. People and machines make problems together. They use inquiry, language, experiences, and materials to show how problems are *differently* significant, relevant, and inextricably shared consequences that require collective governance.

If algorithmic errors are seen as things that people cannot opt out of, that require collective action, and that create new shared consequences, then algorithmic errors become *public problems*.

Turning algorithmic errors into public problems takes work. It means seeing seemingly private, individual errors in system design, datasets, models, thresholds, testing, and deployments—as well as the funding and imagination that birth such systems—as collective concerns. If technologists, policymakers, scholars, and activists can "see like an algorithmic error"—and surface their different ways of doing so—perhaps algorithmic mistakes can be better planned for, ameliorated, or avoided altogether. Algorithmic breakdowns will continue, but we may know better how to deal with them if we learn to interrogate the sociotechnical forces that make algorithmic errors, seeing some errors as "good" and generative illustrations of public problems that need governance, regulation, and reform.

[45] JOHN DEWEY, THE PUBLIC AND ITS PROBLEMS (1954).

[46] NOORTJE MARRES, MATERIAL PARTICIPATION (2012).

[47] GUSFIELD, *supra* note 5.

[48] PHILIP M. NAPOLI, SOCIAL MEDIA AND THE PUBLIC INTEREST (2019).