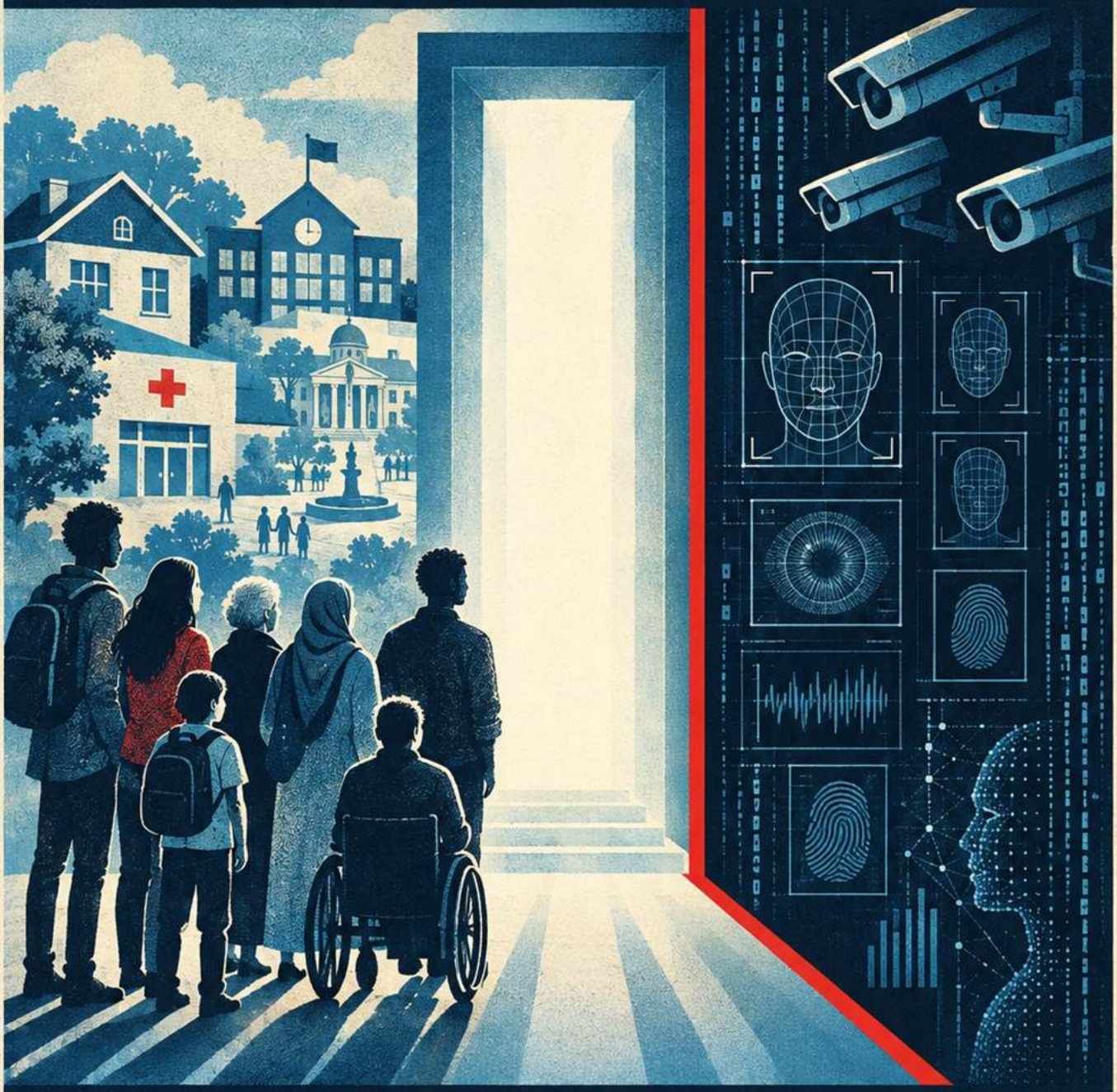
A CITIZEN'S MANIFESTO FOR THE AGE OF AI

PART FIVE — THRESHOLD



HUMAN RIGHTS AND HUMAN DIGNITY

The line that does not move

A CITIZEN'S MANIFESTO FOR THE AGE OF AI

A Civic Blueprint for Human Agency, Democratic Self-Government, and a Workable Earth

PART FIVE — THRESHOLD

HUMAN RIGHTS AND HUMAN DIGNITY

The Line That Does Not Move

DAVID M. DOYLE

The Series Charter

Book Zero—The Unoccupied Center—sets out the charter for this project: technology must enlarge human agency, not replace it. The six books that follow examine the conditions required to keep that commitment real.

Part Five concerns Threshold: privacy, boundaries, and the right not to be fully known. It asks what must remain beyond the reach of surveillance, prediction, classification, and control.

The movement of the six-part manifesto is deliberate:

See → Act → Protect → Share → Draw the Line → Build Together.

The Three Tests

Witness Test. Who sees this decision from outside the authority making it, and can their testimony survive?

Burden Test. Who bears the cost, and do they have meaningful voice, remedy, or refusal?

Return Test. How does this authority end, reverse, repair harm, or return value to those it affected?

How to Read Part Five

A forecast is not a fact. A profile is not a person. A pattern is not consent.

Threshold protects human dignity against systems that would make the whole person available for classification, surveillance, pricing, control, or punishment.

The chapters that follow ask: Does the person know the claim exists? Can they see what it rests on? Can they force a decision on individual evidence instead of resemblance? And is there anything this system is simply not permitted to do?

PART FIVE — THRESHOLD

Privacy, boundaries, and the right not to be fully known

Its corruption: exposure, prediction, and the disappearance of the person into the profile.

Contents

Chapter Thirteen — Human Rights and Human Dignity: The line that does not move

Chapter Fourteen — Health, Care, and the Body: Diagnosis, denial, and the medical record

Chapter Fifteen — Prediction: Governing people by what they have not yet done

Part Opening

Between 2012 and 2020, a national pharmacy chain ran facial recognition in hundreds of its stores.

According to the Federal Trade Commission, "the system generated thousands of false-positive matches," and employees acting on those alerts "followed consumers around its stores, searched them, ordered them to leave, called the police to confront or remove consumers, and publicly accused them, sometimes in front of friends or family, of shoplifting or other wrongdoing." The technology sometimes matched a customer to a person enrolled in the database "based on activity thousands of miles away." The false positives were more frequent in stores located in plurality-Black and plurality-Asian communities than in plurality-White ones.

In December of 2023 the company was barred from using facial recognition for surveillance for five years.959 None of that should have surprised anyone who had read the government's own measurements.

In December of 2019, the National Institute of Standards and Technology tested 189 algorithms from 99 developers against 18.27 million images and reported that "false positive differentials are much larger than those related to false negatives and exist broadly," that "across demographics, false-positive rates often vary by factors of 10 to beyond 100 times," and that with domestic law-enforcement images "the highest false positives are in American Indians, with elevated rates in African American and Asian populations." False positives ran "between 2 and 5 times higher in women than men," and were elevated in the elderly and in children.960 Now hold that beside a different kind of record.

On the thirteenth of July, 2016, the Wisconsin Supreme Court decided the appeal of a man sentenced with the assistance of a proprietary risk score.

The court allowed the practice, but it wrote down what the practice actually was: "Although Loomis cannot review and challenge how the COMPAS algorithm calculates risk, he can at least review and challenge the resulting risk scores." It required sentencing courts to be told that "the proprietary nature of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are to be determined," and that no cross-validation study for a Wisconsin population had been completed.

And then, in a single sentence that I think is the most important line in the whole literature of algorithmic government: "because COMPAS risk assessment scores are based on group data, they are able to identify groups of high-risk offenders—not a particular high-risk individual."961

A store's camera tells an employee who you resemble now.

A court's report tells a judge who you resemble later.

They are the same operation performed at two different moments, and in both cases the person standing there is answering for a resemblance.

Let me say what I am not arguing, because Part Five is the part most easily mistaken for a demand to be untraceable, and it is not that.

I am not arguing for a right to be unknown.

Being known is how a society keeps its promises: the deed that records who owns the land, the chart that records what drug nearly killed you, the roll that records that you are entitled to vote.

I have no patience for the position that treats every record as an injury, or for the idea that accountability is a form of surveillance.

If I have made a decision that affects other people, they are entitled to know I made it — a principle I applied to officials in Part Two and will not now abandon for myself.

The discipline of threshold is narrower.

Every person retains an interior domain that cannot legitimately be reduced to data, usefulness, prediction, or obligation.

The distinction that matters is between a fact about what you did and a claim about what you are — and the second is what these systems produce.

A record says: on this date, this person did this thing, and here is the evidence.

A profile says: this person is the sort of person who.

The first can be checked, dated, and rebutted.

The second is a description of a class with your name at the top of it.

Part One asked whether you can see.

Part Two asked whether anyone can act.

Part Three asked what is permanently altered.

Part Four asked who receives the gain and who carries the cost.

Part Five asks the question underneath all of them: what is left of a person once everything about them is legible, and who gets to decide what your legibility means?

The mechanism is that the profile arrives before the person does, and the person is not told.

Consider what the Loomis court conceded and then permitted.

A score derived from group data cannot identify a particular high-risk individual — the court said so — and yet it was placed in front of a judge sentencing a particular individual, with the weighting sealed as a trade secret.

The remedy the court chose was disclosure of the tool's limitations to the court, not to the defendant, and certainly not disclosure of the tool.

That is the shape of nearly every safeguard now on offer: the decision-maker is warned, the subject is scored.

And the structure is unfalsifiable in a way that should trouble anyone who has ever run a business on metrics.

A prediction that prevents an outcome cannot be tested, because the outcome did not occur.

If a person is denied and does not offend, the model is credited.

If a person is denied and would never have offended, nobody finds out.

The only error the system can observe is the one it made in the permissive direction, which means every incentive pushes toward denial, and denial is then reported as accuracy.

You can watch that incentive operate in dollars.

In October of 2024, the Senate Permanent Subcommittee on Investigations reported that UnitedHealthcare's prior-authorization denial rate for post-acute care rose from 8.7 percent in 2019 to 22.7 percent in 2022 — "an increase of 172 percent" — while the company's overall prior-authorization denial rate barely moved, "going from 7.3 percent in 2019 to 7.6 percent in 2022." In 2022 Humana's denial rate for post-acute care was "over 16 times higher than its overall rate of denial."962 The subcommittee also found the internal machinery: a working group convened in December of 2022 to use machine learning "to predict which denials of post-acute care cases were likely to be appealed, and which of those appeals were likely to be overturned"; and a care-coordination document projecting that a patient would need 16.6 days of post-acute care "based on the experiences of 'similar patients'" — with the patient's attorney reporting that the denial issued on the seventeenth day.963 Read the last item slowly.

The prediction was not used to plan the care.

It was used to price the exit.

And the appeal-prediction work is the tell: a system that models which of its refusals you will fight is a system that has stopped treating refusal as a finding and started treating it as a negotiating position.

Here is the pattern to recognize.

It is the disappearance of the person into the profile, and it has a signature: the subject learns only the outcome.

The shopper was not told there was a match.

The patient was told the coverage ended, not that a model had projected 16.6 days.

The defendant was told his score, but not how it was computed.

In each case the operative claim about the person was unavailable to the person, which means that the ordinary human remedy — no, you have me confused with someone else, and here is why — had nowhere to go.

Some jurisdictions have decided that certain of these uses should not be mitigated but forbidden.

Under the European Union's AI Act, whose prohibitions have applied since the second of February, 2025,964 it is prohibited to place on the market or use systems for the "evaluation or classification of natural persons" based on social behaviour or "known, inferred or predicted personal or personality characteristics" where the resulting social score leads to detrimental treatment in unrelated contexts or treatment "unjustified or disproportionate"; to create or expand facial-recognition databases "through the untargeted scraping of facial images from the internet or CCTV footage"; and to use systems "to infer emotions of a natural person in the areas of workplace and education institutions," outside medical or safety uses.965

I hold that up as evidence that a prohibition is writable, not as proof that it works.

The record on cross-border enforcement is genuinely unsettled: the United Kingdom's data-protection regulator fined a facial-recognition firm £7.5 million in May of 2022 for scraping images of UK residents into a database sold as an identification service, and it took until October of 2025 for the Upper Tribunal to hold that the regulator had jurisdiction at all — with permission to appeal granted that December.966 Three and a half years to establish that the question could be asked in that forum.

Meanwhile the database exists.

Why does this belong in a book about artificial intelligence rather than a book about privacy?

Because prediction is the mechanism by which every concentration in this book is actually exercised.

The credit score, the recidivism score, the attrition score, the dropout score, the fraud score, the length-of-stay projection: these are not applications of the technology, they are what the technology is for.

And the deepest problem is not inaccuracy.

It is that a system attempting to convert a person's freedom into a stable variable degrades the thing it claims to measure.

A forecast that is acted on becomes a cause.

The score that predicts you will fail, and thereby closes the door, has not been validated. It has been enforced.

So the chapters that follow ask four questions of any system that claims to know something about a person, and I would ask you to carry them into any room where such a system is being procured: Does the person know the claim exists.

Can they see what it rests on.

Can they force a decision on their own evidence instead.

And is there anything this system is simply not permitted to do?

Here is the standard I would hold to, stated as one test: a claim about a person is legitimately usable only where the person is told it exists and told what it rests on; where they may compel a decision made on individual evidence instead of on a resemblance; and where nothing essential — liberty, care, housing, income, custody — is withheld or withdrawn on predictive grounds before that has happened. Each clause is falsifiable, which is why I state it that way. The shopper failed the first, the patient the second, and the defendant the third — and each of the three systems was, at the time it operated, entirely lawful. — D.M.D.

Chapter Thirteen: Human Rights and Human Dignity

The Line That Does Not Move

The Line That Does Not Move

What is left of a person once everything about them is legible?

I.

In May or June of 2014, a fourteen-year-old boy named Alexander went on a school field trip to an amusement park in Gurnee, Illinois. His mother had bought him a season pass online, but he had to finish the sign-up in person. At a security checkpoint he "was asked to scan his thumb into defendants' biometric data capture system."967 When he came home, he told his mother the park did "it all by fingerprint now" and that no paperwork had been provided.967 Nobody had told either of them, in writing, what would be done with the thumbprint or for how long it would be kept, and neither had signed anything. "Although he has not returned there since, defendants have retained his biometric identifiers and information."968 Nothing else happened to Alexander. An intermediate appellate court concluded that a person whose only complaint was the violation itself was not "aggrieved" and could not sue.969 In January of 2019, the Illinois Supreme Court reversed, and every justice joined the opinion: "No additional consequences need be pleaded or proved. The violation, in itself, is sufficient to support the individual's or customer's statutory cause of action."970 And then, more plainly: "The Act vests in individuals and customers the right to control their biometric information by requiring notice before collection and giving them the power to say no by withholding consent. … This is no mere 'technicality.' The injury is real and significant."971 The court treated the taking as the injury. The legislature that wrote the statute had said why, eleven years earlier, in a finding I think every legislature should adopt word for word: "Biometrics are unlike other unique identifiers that are used to access finances or other sensitive information. For example, social security numbers, when compromised, can be changed. Biometrics, however, are biologically unique to the individual; therefore, once compromised, the individual has no recourse."972 Among the places that finding named as pilot sites for the new finger-scan technologies were "grocery stores, gas stations, and school cafeterias."972

This chapter is about the distinction between a harm, which you must prove, and a right, which you simply have. Almost every protection now on offer against automated systems is built on the first model. You must show the damage, trace it to a cause, find a defendant, and establish that the defendant owed you something. For most of what these systems do to people, that proof is impossible by design. The face was taken silently. The inference was drawn somewhere else. The vendor, the model, the deployer, and the subcontractor each touched a piece of it, and none of them holds the whole. A rights framework asks a different question. It does not ask what happened to you. It asks what may be done to you at all.

II.

Start with the vocabulary we already have, because it is better than we are. The first article of the Universal Declaration of Human Rights, adopted in December of 1948, reads: "All human beings are born free and equal in dignity and rights."973 The twelfth reads: "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right

to the protection of the law against such interference or attacks."974 In 1988 the United Nations Human Rights Committee interpreted the privacy article, and it anticipated nearly every problem in this chapter. The right, it said, "is required to be guaranteed against all such interferences and attacks whether they emanate from State authorities or from natural or legal persons."975 Legal persons: corporations. And, a generation before the first social network: "The gathering and holding of personal information on computers, data banks and other devices, whether by public authorities or private individuals or bodies, must be regulated by law."976 Every person, the Committee wrote, "should have the right to ascertain in an intelligible form, whether, and if so, what personal data is stored in automatic data files, and for what purposes," and to request "rectification or elimination" of data that is incorrect or unlawfully collected.977 Notice the structure. The duty runs to the state, against a duty-bearer with an address. A government you could petition. A company with a headquarters you could serve.

In 2011, the United Nations Guiding Principles on Business and Human Rights tried to extend that architecture to the corporation directly. Businesses "should avoid infringing on the human rights of others and should address adverse human rights impacts with which they are involved."978 The thirteenth principle asks firms not only to avoid causing harm but to "prevent or mitigate adverse human rights impacts that are directly linked to their operations, products or services by their business relationships, even if they have not contributed to those impacts."979 And "business relationships," the commentary explains, include "entities in its value chain."980 Here is the structural problem of this chapter. The rights were written against a state or a corporation with an address. Automated harm arrives distributed.

A model is trained in one country on material taken from a hundred. It is labeled by workers in a third, under a subcontract with a fourth. It is licensed to a deployer in a fifth, who uses it to decide something about a person in a sixth. When the person asks who did this, every party can truthfully point to another. The harm is real and it has no single author, and a legal system that requires a single author will find that it has nothing to say. The answer is not a new theory of rights. It is to decide in advance, for a small number of things, that no chain of contracts can make them permissible. There is a part of a person that is not anyone's material. The rest of this chapter is an attempt to say where that part begins. I

II.

Begin with the face, because it is the part of a person that is most exposed and least replaceable. The Illinois statute that protected Alexander is still the strongest biometric law in the United States. No private entity may "collect, capture, purchase, receive through trade, or otherwise obtain a person's or a customer's biometric identifier or biometric information" unless it first informs the person in writing, discloses "the specific purpose and length of term" of the collection, and "receives a written release."981 No entity holding such data "may sell, lease, trade, or otherwise profit from" it.982 And any person "aggrieved by a violation of this Act shall have a right of action," with liquidated damages of $1,000 for a negligent violation and $5,000 for an intentional or reckless one.983 That last sentence is why the statute works. It puts enforcement in the hands of the person whose face was taken, not an agency with a budget.

The results have been large. In February of 2021, a federal court approved the settlement of a class action over a social network's photo-tagging feature, covering Illinois users "for whom Facebook created and stored a face template after June 7, 2011." The judge called the $650 million settlement "a landmark result," noting that it would put "at least $345 into the hands of every class member interested in being compensated."984 Texas,

under a different statute enforced by its attorney general, reached a $1.4 billion settlement with the same company, describing conduct in which "for more than a decade Meta ran facial recognition software on virtually every face contained in the photographs uploaded to Facebook, capturing records of the facial geometry of the people depicted."985 Virtually every face. Nobody decided to take hundreds of millions of faces. A feature was built that did it by default, and the default was the decision. The American Civil Liberties Union's case in Illinois against Clearview AI ended in a consent order in May of 2022. The firm is "permanently enjoined from granting paid or free access to the Clearview database" anywhere in the United States to most private entities, and for five years it may not sell access to Illinois government agencies, "including state and local police departments."986 But the statute itself exempts "a contractor, subcontractor, or agent of a State agency or local unit of government when working for that State agency or local unit of government."987 So the strongest biometric law in the country protects you against your grocery store and your employer. It does not protect you against the government's contractors. That is not an oversight. It is where the line was drawn, and it was drawn around the one actor with the most power to use a face against you.

IV.

The Illinois law has also shown what happens when a right actually bites. In February of 2023, the state's Supreme Court held, four justices to three, that "a separate claim accrues under the Act each time a private entity scans or transmits an individual's biometric identifier or information" without consent.988 The case involved a restaurant chain that required its workers to scan a fingerprint to access their pay stubs. The company warned that, for a class of as many as 9,500 current and former employees, "class-wide damages in her action may exceed $17 billion."989

The legislature responded. In August of 2024 the governor signed an amendment providing that a private entity that collects "the same biometric identifier or biometric information from the same person using the same method of collection" more than once "has committed a single violation," for which the person "is entitled to, at most, one recovery."990 In April of 2026 the Seventh Circuit held that the amendment applies to cases already pending when it was enacted.991 I do not think the legislature was wrong to act. A liability that no court will ever actually impose protects no one; judges faced with seventeen billion dollars for fingerprinted pay stubs will look for a reason to find no violation at all. But notice what the amendment did. The Rosenbach court had said that liability without proof of injury gives entities "the strongest possible incentive to conform to the law and prevent problems before they occur and cannot be undone."992 After the amendment, the price of taking a person's fingerprint without consent every working day for years is the same as the price of taking it once. That is the pattern I will keep returning to. A right is written. It turns out to have force. The force is then negotiated down until the right is compatible with the practice it was written to stop. The line moves.

V.

The government's own use of the face has had no Illinois statute to restrain it at all. In 2021, the Government Accountability Office surveyed forty-two federal agencies that employ law-enforcement officers. Twenty reported owning or using facial-recognition systems. "Six agencies reported using the technology on images of the unrest, riots, or protests following the death of George Floyd in May 2020. Three agencies reported using it on images of the events at the U.S. Capitol on January 6, 2021."993 Two years later, the same office examined seven agencies using commercial facial-recognition services. "All seven agencies initially used these services

without requiring staff take facial recognition training." Six "cumulatively conducted about 60,000 searches when they did not have training requirements in place."994 At the FBI, "only 10 staff completed facial recognition training of 196 staff that accessed the service."995

In September of 2024, the U.S. Commission on Civil Rights put the governing fact in one sentence: "There are currently no federal laws or regulations that expressly authorize or limit FRT use by the federal government."996 It found the technology spreading into public housing, where the Department of Housing and Urban Development was "proliferating FRT use largely through its grant programs for PHAs, putting FRT in the hands of grantees with no regulation or oversight," in housing "where tenants are disproportionately female and people of color."997998 At the airport, the promise is consent. The Transportation Security Administration says travelers "may decline the optional photo, without recourse," and that doing so "will not take longer."999 Its facial comparison units are deployed at more than 250 airports.1000 But when the federal privacy oversight board's staff examined the program in May of 2025, it insisted that "travelers must be informed that they can opt out and given a meaningful opportunity to do so."1001 It also reported that the department-wide directive guaranteeing that "U.S. citizens are afforded the right to opt out of face recognition for non-law enforcement uses"1002 had been removed from the department's website in February of 2025, and that the department "was unable to confirm to PCLOB whether it remains official policy or not."1003 A right to decline that you are not reliably told about, under a policy whose status the department cannot confirm, is not a right. It is a courtesy. And the courtesy runs only to citizens. Under a rule in force since December 26, 2025, the Department of Homeland Security "may require all aliens to be photographed when entering or exiting the United States," at airports, land ports, and seaports.1004 The rule also ends an exemption for travelers under fourteen and over seventy-nine.1005 For citizens, "participation … is voluntary."1006 I do not dispute that a state may verify identity at its border; the agency credits the technology with identifying "more than 1,500 individuals posing under a false or assumed identity."1007 I dispute the idea that whether a child's face may be taken can be settled by a finding about matching performance. That is a finding about whether the machine can do it. It is not a finding about whether it should. The United Nations High Commissioner for Human Rights recommended in 2021 that states "impose a moratorium on the use of remote biometric recognition technologies in public spaces, at least until the authorities responsible can demonstrate compliance with privacy and data protection standards and the absence of significant accuracy issues and discriminatory impacts."1008 No such demonstration has been made in the United States. There is not even a statute that would require one.

VI.

Identifying a face is one thing. Reading it is another, and the second claim is weaker than most people realize.

In 2019, five scientists published the most thorough review of the evidence on whether emotions can be inferred from facial movements. Their conclusion was unambiguous: the facial configurations in question "are not 'fingerprints' or diagnostic displays that reliably and specifically signal particular emotional states regardless of context, person, and culture. It is not possible to confidently infer happiness from a smile, anger from a scowl, or sadness from a frown."1009 They observed that "technology companies, for example, are spending millions of research dollars to build devices to read emotions from faces, erroneously taking the common view as a fact that has strong scientific support."1010 Efforts to "read out" internal states from facial movements alone, "without considering various aspects of context, are at best incomplete and at worst entirely lack validity, no matter how sophisticated the computational algorithms."1011 And they named the stakes: "An incorrect

inference about defendants' emotional state can cost them their children, their freedom, or even their lives."1012 Regulators have noticed. In October of 2022 the United Kingdom's data-protection authority warned about technologies that process "gaze tracking, sentiment analysis, facial movements, gait analysis, heartbeats, facial expressions and skin moisture." Its deputy commissioner said: "Developments in the biometrics and emotion AI market are immature. They may not work yet, or indeed ever."1013 The Justice Department's disability guidance gives the concrete case: if a county government "uses facial and voice analysis technologies to evaluate applicants' skills and abilities, people with disabilities like autism or speech impairments may be screened out, even if they are qualified for the job."1014 Now look at the one American statute written for this exact use. Illinois' Artificial Intelligence Video Interview Act requires an employer that uses automated analysis of recorded interviews to notify applicants, explain "how the artificial intelligence works and what general types of characteristics it uses," and obtain consent; it forbids evaluating anyone who has not consented, and requires deletion within thirty days of a request.1015 The state's report for 2026 is one sentence long in its essentials: "No such data was reported to DCEO for the 12-month period ending November 30, 2025."1016 The statute contains no penalty and no private right of action.1017 It is a law that asks politely. In the year the report covers, nobody answered.

I have already set out that the European Union has prohibited emotion inference in workplaces and schools outright. The science gives no reason to treat that as an overreach. A system that claims to read a person's inner state from the outside, when the best evidence says that cannot reliably be done, is not a measurement. It is an accusation with a confidence score.

VII.

The machines are also wrong about who people are, and they are wrong unevenly. In 2018, two researchers tested three commercial systems that classified faces by gender. They found that the standard benchmark datasets were "overwhelmingly composed of lighter-skinned subjects (79.6% for IJB-A and 86.2% for Adience)," and that on a balanced set, "darker-skinned females are the most misclassified group (with error rates of up to 34.7%). The maximum error rate for lighter-skinned males is 0.8%."1018 A year later, a team tested four commercial services on 2,450 images and found that they "performed consistently worse on transgender individuals and were universally unable to classify non-binary genders."1019 Average accuracy for trans men was 70.5 percent.1020 For agender, genderqueer, and nonbinary people the researchers calculated the accuracy "as 0%," because the services "only return binary gender labels."1021 Zero percent is not an error rate. It is a statement that a category of person does not exist in the system. The Civil Rights Commission's report adds that even the best-performing identification algorithms in laboratory testing produce more false positives "for certain demographic groups, specifically Black people (particularly Black women), people of East Asian descent, women, and older adults," and that the government's testing "is voluntary and represents laboratory—not real-world—results."1022 Misclassification is not random noise. It falls on the people least represented in the data and least powerful in the room. The federal machinery for catching it has been turned down. In May of 2022 the Equal Employment Opportunity Commission and the Justice Department jointly warned that algorithmic hiring tools could screen out people with disabilities; the commission's chair said, "New technologies should not become new ways to discriminate."1023 The commission's technical-assistance documents on artificial intelligence now return errors when fetched from its website, though its publications index still lists their titles.1024 I have found no formal notice rescinding them. What I have found is an executive order of April

2025 directing that "all agencies shall deprioritize enforcement of all statutes and regulations to the extent they include disparate-impact liability."1025

Disparate impact is the legal theory that fits this problem exactly. A classifier does not intend anything. It simply fails more often for some people than others. A civil-rights regime that will act only on intent has decided in advance not to see what these systems do.

VIII.

Consider the person at work. In October of 2022, the general counsel of the National Labor Relations Board wrote that "an issue of particular concern to me is the potential for omnipresent surveillance and other algorithmic-management tools to interfere with the exercise of Section 7 rights by significantly impairing or negating employees' ability to engage in protected activity and keep that activity confidential from their employer."1026 On February 14, 2025, the acting general counsel rescinded the memorandum.1027 In November of 2024, the Consumer Financial Protection Bureau issued a circular addressing the third-party reports now compiled about workers. It described employers using vendors "to monitor workers' sales interactions, to track workers' driving habits, to measure the time that workers take to complete tasks, to record the number of messages workers send and the quantity and duration of meetings they attend, and to calculate workers' time spent off-task through documenting their web browsing, taking screenshots of computers, and measuring keystroke frequency."1028 Its answer to whether employers could use such dossiers and scores without complying with the Fair Credit Reporting Act was one word: "No."1029 As of May 12, 2025, the circular was withdrawn, with the bureau noting that the "withdrawal is not necessarily final."1030 Two federal documents that described the problem precisely were withdrawn within three months of each other. The practices they described were not. The GAO summarized what workers themselves said when the White House science office asked for public comment. "Privacy was the most frequently raised concern." Workers and union stakeholders reported that digital surveillance "may discourage workers from unionizing, make workers feel distrusted by their employers, and decrease workplace morale"; "workers reported that digital surveillance increased stress and fear."1031

State law, where it exists, mostly requires a notice. New York requires employers that monitor employees' email, telephone, or internet use to give "prior written notice upon hiring" and to post it "in a conspicuous place," and advise employees that they may be monitored "at any and all times and by any lawful means."1032 The maximum penalty for a first offense is five hundred dollars, and only the attorney general may enforce it.1033 A California bill introduced in 2025 would prohibit employers from using surveillance tools "in specified employee-only areas, including bathrooms and breakrooms."1034 That such a prohibition needed to be written down tells you where the practice had reached. The requirement I set out for Part Five is that workers govern the tools that govern them. A notice that you are being watched at all times, by any lawful means, is not consent. It is a description of the terms on which you are permitted to keep your job.

IX.

A child cannot quit school. In March of 2022, two United States senators published the results of an investigation into four companies that sell software to monitor students' online activity.

The companies claimed their products were not meant for discipline, but "a survey of teachers found that 43% reported that their schools are using these tools to identify violations of discipline policies."1035 Several companies said flagged activity could result in immediate contact of "law enforcement and/or [the National Center for Missing and Exploited Children]," or "police dispatch for a wellness check," and the senators concluded that the products "may be exacerbating the school-to-prison pipeline by increasing the involvement of law enforcement with students."1036 None of the companies had "analyzed their algorithms for bias or even track whether their products over- or under-identify different groups of students," including students of color and LGBTQ+ students.1037 The Federal Trade Commission, voting five to none in May of 2022, stated that companies "cannot deny children access to educational technologies when their parents or school refuse to sign up for commercial surveillance," that they cannot require children "to provide more information than is reasonably needed," and that they may not keep children's information "for longer than is necessary."1038 In December of 2025 the commission settled with an education-technology provider whose data-security failures allegedly let a hacker reach the "personal data of 10.1 million students, including their email and mailing addresses, dates of birth, student records, and health-related information."1039 Ten million children, whose records existed in one vendor's system because their schools had bought a product. Nobody asked the children. They were not in a position to be asked.

X.

Outside the classroom, children live on the same platforms as adults, under protections that assume a parent who reads the terms. The Surgeon General's 2023 advisory reported that "up to 95% of youth ages 13–17 report using a social media platform, with more than a third saying they use social media 'almost constantly'," and that "nearly 40% of children ages 8–12 use social media," below the age platforms usually require.1040 Adolescents "who spent more than 3 hours per day on social media faced double the risk of experiencing poor mental health outcomes including symptoms of depression and anxiety."1041 The federal children's privacy rule was amended in April of 2025, for the first time since 2013. Biometric identifiers, including "voiceprints; gait patterns; facial templates; or faceprints," are now personal information.1042 Parents must be able to consent to collection without consenting to disclosure to third parties, and an operator must "obtain separate verifiable parental consent to such disclosure."1043 Children's data "may not be retained indefinitely," and operators must maintain a written retention policy with "a timeframe for deletion."1044 But the architecture is still consent, and the consenting party is a parent. A child's protection should not depend on the technical sophistication of the adult who happens to be raising them. The retention limit is the better model, because it applies whether or not anyone asked. California tried to write protections that did not depend on consent at all, and the courts have cut them back. In March of 2026 the Ninth Circuit vacated most of an injunction against the state's age-appropriate design code, including as to its age-estimation requirement, but affirmed the injunction against the provisions restricting how companies may use children's data and against its restriction on manipulative design, on vagueness grounds.1045 The provisions that survived describe who is covered. The provisions that remain blocked describe what may not be done to them.

The newest frontier is the machine that talks back. In September of 2025 the Federal Trade Commission ordered seven companies to report on their chatbots that act as companions, observing that such systems "can effectively mimic human characteristics, emotions, and intentions, and generally are designed to communicate like a friend or confidant, which may prompt some users, especially children and teens, to trust and form relationships with chatbots."1046 A month later California enacted the first state law aimed at such products; its

author cited the case of a fourteen-year-old who "ended his life after forming a romantic, sexual, and emotional relationship with a chatbot." The law requires operators to maintain a protocol against producing "suicidal ideation, suicide, or self-harm content," and, where the user is a minor, to remind them "every 3 hours" to take a break "and that the companion chatbot is AI and not human," and it gives an injured person the right to sue. A reminder every three hours that your friend is not a person. It is the right instinct, and the fact that it needed to be legislated is the finding.

XI.

The worst use of the technology against children has arrived in volume. Federal law already anticipated it. The statute on obscene visual depictions of minors covers "a visual depiction of any kind, including a drawing, cartoon, sculpture, or painting," and states: "It is not a required element of any offense under this section that the minor depicted actually exist."1047 The National Center for Missing and Exploited Children began tracking reports involving generative AI in January of 2023. It counted 4,700 that year and 67,000 in 2024.1048 Its 2025 breakdown counts more than 7,000 reports of users generating or possessing such material, more than 30,000 of users attempting to generate it, more than 145,000 of users altering an existing abuse file, and more than 12,000 "reports of CSAM that companies indicated were identified in training data."1049 "More than 275 direct victims of GAI CSAM were identified in 2024 and 2025 alone and the offenders in many cases were people in the child's life with direct access to them."1050 Two numbers deserve to be read together. The center's own staff categorized "more than 158,000 images and videos" as AI-generated between January 2023 and December 2025. In the same period, the companies filing reports "annotated just over 11,000 images and videos as being AI-generated," despite a form field added in October 2023 for exactly that purpose.1051 The platforms are either not looking or not saying. And twelve thousand reports of abuse material found in training data. What a model is built from is not a neutral question.

In August of 2026 the Seventh Circuit decided the first appellate case on a man accused of producing such images with an image-generation model. It held that the possession statute was "unconstitutional as applied" to his private possession at home, under the Supreme Court's precedent protecting private possession of obscenity; the production and distribution charges were not before it.1052 The panel was plainly uncomfortable: "We now live in an age where GenAI models can render images depicting the abuse of virtual children that are virtually indistinguishable from those depicting the abuse of actual children. … we have some concerns about the lines these cases draw, but we are not free to redraw them ourselves."1053 For adults and children depicted in nonconsensual intimate images, Chapter Four described the federal law enacted in 2025. Its platform duty took effect on May 19, 2026: a covered platform must remove a depiction "as soon as possible, but not later than 48 hours after receiving" a valid request, and "make reasonable efforts to identify and remove any known identical copies."1054 The Federal Trade Commission began enforcing it that day.1055 The penalty can reach "$53,088 per violation."1056 Forty-eight hours is a timeline. It is one of the few places in American law where a person harmed by an automated system has one.

XII.

Behind every system in this chapter there are people whose job is to look at what the rest of us are protected from. In September of 2023, a group of senators and representatives wrote to nine of the largest companies building these systems about "the working conditions of those who perform your companies' 'ghost work' —

unseen but critical tasks such as data labeling, without which there would be no artificial intelligence (AI).”1057 They wrote that “the median wage of workers on MTurk has been estimated at just $1.77 per hour,” and that “with little psychological or other support, data workers are exposed to traumatizing content such as torture, rape, and murder.”1058 The International Labour Organization’s measurement is the most careful I know. Average hourly earnings on online web-based platforms were “US$3.4, while half of the workers on these platforms earn less than US$2.1 per hour,” and workers in developing countries on freelance platforms earned “60 per cent less” than those in developed ones.1059 The system that removes abuse images from your feed was trained, in part, by a person who looked at them for two dollars an hour. The mental cost of keeping the platform tolerable was moved to the cheapest labor market that could absorb it.

We know the cost can be reduced, because an American court has already priced it. In 2021 a California state court approved a $52 million settlement with Facebook for a class of 14,713 content moderators employed through its vendors in the United States.1060 Each class member received an initial $1,000 that “may be used to obtain a screening from a medical professional.”1061 The workplace terms are ordinary and humane: weekly one-on-one sessions for moderators on graphic queues, and tools for “Viewing images in black and white; Blurring images; Blocking faces within images,” along with auto-muting of videos.1062 Black and white. Blurring. Muting the sound. These are not expensive. They were negotiated for American workers because American workers could sue in an American court. The question for everyone else is whether the protection follows the work.

XIII.

In Nairobi, a court decided that it does. In 2022 a South African man who had worked in Nairobi as a content moderator for Facebook, employed by a local outsourcing firm, filed a petition in Kenya’s Employment and Labour Relations Court against the firm and the platform’s parent companies.1063 A second petition followed in March of 2023, brought by moderators who alleged their employment was being terminated in retaliation for the first.1064 The platform companies argued the court “lacks jurisdiction because they are foreign companies incorporated outside Kenya and not carrying on business in Kenya.”1065 In April of 2023, ruling on the second petition, the court wrote the sentence that answers the structural problem of this chapter better than any statute: “Privity of contract is not an essential element in establishing violations of human and fundamental freedoms by a person in control and authority over another.”1066 It added: “It is immaterial whether the alleged violations occur in a physical or virtual space within the jurisdiction of this court in Kenya.”1067 In September of 2024 the Court of Appeal dismissed the company’s appeals, holding that whether the violations occurred in Kenya was a question of fact for a full hearing: “Therefore, the jurisdictional question urged by the appellant collapses.”1068 The consolidated petitions were listed for a ruling in February of 2026.1069 As far as the primary record shows, the merits have not been decided. I do not know how this will end. But the principle is the right one. The chain of subcontracts does not dissolve the responsibility of the party “in control and authority over another.” A duty that can be contracted away to a vendor in a cheaper jurisdiction is not a duty.

The international labor system has moved toward the same idea. In June of 2026 the International Labour Conference adopted a convention on decent work in the platform economy, described as extending protections “to all platform workers, regardless of how their employment status is classified under national law.”1070 It requires that where decisions “are generated by an automated decision-making system,” workers have access on request to “a written explanation for significant decisions that adversely impact their working arrangements,” a

review of decisions to withhold pay or deactivate accounts, and "appropriate human involvement."1071 Whether it will reach the outsourcing firms that employ most annotators depends on definitions I have not been able to resolve, and on ratification by member states.

XIV.

Content moderation is the largest system of adjudication ever built, and almost none of it is public.

The European Union's Digital Services Act requires online platforms to give "a clear and specific statement of reasons" to any user whose content is removed, demoted, or whose account is restricted, including "information on the use made of automated means in taking the decision."1072 It requires an internal complaint system, free of charge, whose decisions may not be taken solely by automated means.1073 On the day I checked it, the database reported 4,027,713,330 statements of reasons submitted by 374 platforms in the preceding six months, and "43% of fully automated decisions."1074 Four billion decisions in half a year. Nearly half made with no human involved. That is not an exception to the governance of speech. It is most of it. And it is visible only because one legislature required each decision to leave a record. The most elaborate private alternative shows the limits of self-governance. The oversight board Meta created can issue decisions that are "binding" on the company for the case at hand, but "all policy advisory opinions are advisory."1075 It is a court whose jurisdiction, remedies, and budget were all set by the party it judges. I respect the people who serve on it. It is still not a public forum. I know of no general American equivalent to the statement of reasons or the human appeal.

The principle should not be controversial: a person whose speech is removed by a machine is entitled to be told that a machine removed it, told why, and heard by a human being if they object.

XV.

A person's voice and face are no longer only things that can be recorded. They can be manufactured.

Tennessee's 2024 law begins with a declaration of rights: "Every individual has a property right in the use of that individual's name, photograph, voice, or likeness in any medium in any manner."1076 It defines voice to include "a simulation of the voice of the individual."1077 And it makes liable a person who distributes "an algorithm, software, tool, or other technology, service, or device, the primary purpose or function of such algorithm … is the production of a particular, identifiable individual's photograph, voice, or likeness," with knowledge that the use was not authorized.1078 The law reaches the tool, not only the output. California followed with two statutes. One makes unenforceable any contract clause allowing a digital replica to be used "in place of work the individual would otherwise have performed in person" unless the clause contains "a reasonably specific description of the intended uses" and the person was represented by a lawyer or a union.1079 The other makes anyone who uses a dead performer's digital replica without consent liable for at least "ten thousand dollars ($10,000)."1080 Congress has not acted.1081 The writers' 2026 agreement adds one obligation that points toward training: the studios "must give the Guild written notice if they license writers' work to train a commercial Generative Artificial Intelligence (GAI) system," and "the Guild can request discussion with the company about the license, including any remuneration for writers."1082 That is the embryo of the right answer. Consent, credit, and compensation for training material can only be managed collectively, because no individual writer can negotiate with a model developer and no developer can negotiate with every writer. The

union supplies the address. Copyright law is moving toward the same place. The Copyright Office's 2025 analysis concluded that "making commercial use of vast troves of copyrighted works to produce expressive content that competes with them in existing markets, especially where this is accomplished through illegal access, goes beyond established fair use boundaries."1083 It took seriously the harm of dilution: "If thousands of AI-generated romance novels are put on the market, fewer of the human-authored romance novels that the AI was trained on are likely to be sold."1084 But on payment, it concluded that "government intervention would be premature at this time," while suggesting that "extended collective licensing should be considered."1085 In February of 2025 a federal judge in Delaware held that a company that used a legal publisher's headnotes to train a competing research tool was not protected, in part because "at least one potential derivative market is also obvious: data to train legal AIs."1086

A market in training data is exactly what consent and compensation require. The courts have begun to recognize that it exists. What remains is to decide that the people whose work constitutes it must be party to it.

XVI.

Some material is not owned by any single person, and the law has almost no vocabulary for it.

A community's language, songs, stories, designs, and knowledge of plants and land exist collectively. A model trained on them takes from everyone and no one. Chapter Three described the principles Indigenous data-governance scholars have articulated. Here I note how far the formal instruments have come, and how far they have not. The United Nations Declaration on the Rights of Indigenous Peoples states that Indigenous peoples "have the right to maintain, control, protect and develop their cultural heritage, traditional knowledge and traditional cultural expressions," including "oral traditions, literatures, designs," and "visual and performing arts," and "their intellectual property over such cultural heritage."1087 It is a declaration, not a treaty. In May 2024 the member states of the World Intellectual Property Organization adopted a treaty requiring a patent applicant whose invention is based on traditional knowledge to disclose "the Indigenous Peoples or local community, as applicable, who provided the traditional knowledge."1088 It applies only to patents. As of July 2026 it had four of the fifteen ratifications it needs to take effect.1089 New Zealand has gone furthest. Its statistics agency signed an agreement with a national body of tribal leaders in 2019 under which "respective views be heard, considered, and afforded equal explanatory power."1090 The government's data-sharing guidance states: "To Māori, data is a taonga — things and values that are treasured, both intangible and tangible. Māori data requires special protections when it is collected, used and shared by agencies," and it contemplates "providing data access to Māori, with a view to full repatriation (returning the data to where it belongs)."1091 Repatriation of data. That is a concept worth carrying into every negotiation over training material. What is missing is the application to the corpus.

XVII.

Which brings me back to the address. A right with no forum is a sentiment. A forum with no timeline is a waiting room. A timeline with no remedy is a schedule. A remedy with no cost to the violator is a price of doing business. Redress requires all four, and in this domain we rarely have more than one.

Look at what happened to the most serious attempt to build it. In 2022 the European Commission proposed a directive on civil liability for artificial intelligence. In February of 2025 it listed the proposal for withdrawal, with

the explanation "No foreseeable agreement."1092 The withdrawal was published in the Official Journal in October of 2025.1093 What survived is more useful than it sounds. The revised Product Liability Directive treats "software" as a product,1094 and explains that software — "operating systems, firmware, computer programs, applications or AI systems" — is a product "for the purposes of applying no-fault liability, irrespective of the mode of its supply," including software "supplied through a software-as-a-service model."1095 Compensable damage includes "medically recognised damage to psychological health."1096 No-fault liability is the product-safety version of what the Illinois court said about biometrics: the injured person should not have to prove what went wrong inside the machine. The Council of Europe's framework convention on artificial intelligence, which the United States signed in 2024, requires parties to ensure "accessible and effective remedies for violations of human rights," but only "to the extent remedies are required by its international obligations and consistent with its domestic legal system."1097 Its obligations apply in full to "public authorities, or private actors acting on their behalf"; for other private actors, each party may choose between applying the convention's principles and "taking other appropriate measures."1098 In the United States, the state that went furthest has retreated and rebuilt. Colorado's 2024 act, with its right to appeal an adverse decision, never took effect; it was delayed once and then repealed and replaced in May of 2026 by a narrower law.1099 The replacement contains one excellent definition. "Meaningful human review" means review by an individual with "authority to approve, modify, or override" the decision who "considers relevant, available primary evidence," "is trained to conduct the review," and "does not default to the system output."1100 It also creates no private right of action.1101 Compare that with the Illinois biometric statute, which has the clearest record in this chapter of changing corporate behavior: a default setting turned off, face templates deleted, a database closed to private buyers. It had a private right of action. It had liquidated damages. It did not require proof of harm. Every element that made it effective is an element that later laws have been drafted to avoid.

XVIII.

Put the three tests to what this chapter has described.

Can it be witnessed? Mostly not by the person it happens to. The student is not told the laptop is reading their messages; three of four monitoring companies said they do not alert students or guardians. The traveler's right to decline is, by the oversight board's finding, inconsistently communicated. The Illinois employers required to report on automated interview screening reported nothing. Platforms filing reports of abuse material labeled eleven thousand images as synthetic where the clearinghouse's own staff found a hundred and fifty-eight thousand. The one striking exception is the European moderation database, which exists because a law required each decision to leave a public trace.

Who carries the burden? The people least able to refuse. The dark-skinned woman misclassified a third of the time, and the nonbinary person the system cannot classify at all. The autistic applicant screened out by a voice analyzer. The non-citizen traveler under fourteen, whose exemption ended because the matching got good enough. The moderator in Nairobi viewing abuse images for two dollars an hour so that the rest of us do not have to. And the child, protected to the degree that an adult understood the consent form.

Is there a way back? For the body, no. The Illinois finding is the whole argument: a compromised biometric cannot be changed, and "the individual has no recourse." A synthetic image, once distributed, has copies; the law now requires platforms to try to find the identical ones, which concedes that the others will persist. A model

trained on a person's work does not forget it because a court later finds the copying unlawful; the forward-looking protection the court described in the $1.5 billion settlement was "the destruction of Class Members' pirated Works."1102 And for the moderator, there is no way back from what was seen.

Where the harm is irreversible, the only protection that works is the one that operates before it occurs. That is the case for prohibition.

XIX.

The first objection: prohibitions are blunt; regulate risk, not categories. Most of this book asks for disclosure, testing, appeal, and liability rather than bans. But risk management presupposes that the harm is measurable and correctable after the fact. For the domains I would prohibit, it is neither. Mass identification in public space cannot be undone for the people identified; an emotion score cannot be audited, because the underlying science says the inference does not hold. A risk framework applied to something that does not work only certifies how carefully it was done.

The second objection: facial recognition catches criminals and finds missing children. Nothing I propose prevents a targeted search, authorized by a court, for a specific person suspected of a specific crime, with the result treated as a lead and not as evidence. What I would prohibit is running every face in a public place against a database, continuously, without suspicion. Sixty thousand searches run before anyone had to be trained is not a record that argues for more discretion.

The third objection: human interviewers read faces badly too. They do. But a human interviewer can be asked what they saw and why, can be cross-examined, and does not score ten thousand applicants in an afternoon with the same error. When a person is biased, the bias belongs to a person. When a system is invalid, the invalidity is replicated at scale and dressed as objectivity. The fourth objection: consent and compensation for training data are impossible at scale and would entrench the incumbents who can afford licenses. This is the strongest objection. A regime of individual permissions would be unworkable, and expensive licenses would favor the largest firms. That is why I would require collective mechanisms — extended collective licensing, bargaining through unions and guilds, community-level consent for community material — rather than individual permission for each work. The alternative, in which the largest firms simply take what they need because no one can afford to sue them, entrenches incumbents far more thoroughly.

The fifth objection: rights language is vague and international declarations do not bind American courts. True, and it is why this chapter rests on specific domestic instruments wherever it can: a biometric statute, a likeness statute, a children's privacy rule, a takedown deadline. The international texts supply the principle, not the enforcement. But the principle matters, because it is what allows a Kenyan court to say that privity of contract is not an essential element. Without a statement of what a person is owed regardless of contract, every protection becomes a question of who signed what. What I will concede without reservation is the limit of my own evidence. I could not confirm whether the California workplace surveillance bill became law. I could not find a primary record of the Kenyan Supreme Court's action on the platform's application. I could not establish whether the equal-employment commission's guidance on algorithms was formally rescinded or merely removed. And the clearinghouse's own figures on synthetic abuse reports do not fully reconcile between two of its pages. In a chapter about legibility, I have had to report four places where the record is not legible to me. That is the same finding again.

XX.

Before the requirements, let me set the concerns out plainly.

Every row is the same shape. Something that belongs to a person — their face, their inner life, their attention, their image, their community's knowledge, their childhood — is taken into a system as material. The protection offered afterward is a notice, a consent box, or a lawsuit someone else will have to afford. Prohibition, not mitigation, in three defined domains. Mass biometric identification of people in public space; social scoring of persons by public or private actors for use in unrelated decisions; and emotion inference in employment and education. Forbidden outright, with no risk-assessment path to permission. Targeted biometric search of a named person under a court order remains lawful, and its result is a lead, not evidence. The taking is the injury.

Every biometric, voice, and likeness protection carries a private right of action with liquidated damages, available without proof of further harm, on the model the Illinois court upheld. No exemption for contractors acting for the government. Consent, credit, and compensation for training material, managed collectively. Developers disclose the sources of training data at the level of the work or the collection. Creators and communities bargain through unions, guilds, collecting societies, or tribal and community bodies, with extended collective licensing where individual bargaining is impossible. Community material requires community consent, and disclosure of origin, as the traditional-knowledge treaty already requires for patents. A person's voice and likeness are their own. A property right in name, image, voice, and likeness, including simulations, with liability that reaches the tools built for the purpose of producing a particular person's likeness without authorization, and consent requirements that survive the contract and the performer. Labor standards for the people who label and moderate. A wage floor tied to the local living wage, not the global minimum; clinical support, exposure limits, and protective tooling as standard terms; medical screening paid for by the platform; and liability that follows control, not contract, so that the firm in "control and authority" over the work answers for it wherever the work is done. Children's protections that do not depend on a parent's technical sophistication. Strict data minimization and deletion by default; no targeted advertising to minors; no engagement-maximizing design aimed at children; companion chatbots that disclose they are not human, interrupt at intervals, and route crisis signals to people; and a duty on every model developer to screen training data for abuse material and report what they find.

Monitoring at work and school is disclosed, bounded, and bargained. The monitored person, including the student, is told what is collected and why. No monitoring in bathrooms, breakrooms, or homes. Safety monitoring may not be used for discipline, and flags may not be sent to police without human review. In workplaces, surveillance tools are subject to bargaining before introduction. A statement of reasons and a human appeal. Every consequential automated decision about a person's speech, account, classification, or eligibility comes with a statement of reasons that says whether a machine made it, and a right to review by a person with authority to reverse it who "does not default to the system output." Redress with a forum, a timeline, a remedy, and a cost to the violator. Jurisdiction lies where the harm lands. Developer and deployer are jointly liable, with contribution between them sorted out afterward, not by the injured person. Software is a product for purposes of no-fault liability, including psychological injury. And every deadline is written into law, as the takedown statute wrote forty-eight hours. No system may assign a person a category it cannot represent. A system that cannot classify a person correctly must say so and route to a human, not assign the nearest label. Inference of gender, race, disability, sexual orientation, or religion from biometric data is prohibited except where the person has asked for it. There is a part of a person that is not anyone's material. A society that cannot

state where that line falls will discover it has no line. Go back to the boy with the season pass. What was taken from him was small: a thumbprint, at a checkpoint, on a field trip, with no paperwork. What the court protected was not small. It was the principle that he did not have to prove the taking had hurt him in order to refuse it, and that the power to say no is itself the thing a person owns. Draw the line before the reading. Then do not let it move. — D.M.D.

Chapter Fourteen: Health, Care, and the Body

Diagnosis, Denial, and the Medical Record

Diagnosis, Denial, and the Medical Record

When the system denies my treatment, who do I appeal to?

I.

In June of 2026, the inspector general of the Department of Health and Human Services published a short data brief about a single month. It concerned the place a person goes after a stroke or a broken hip, when the hospital has finished but the body has not: the skilled nursing facility. In June of 2024, nineteen Medicare Advantage companies had received requests to admit their enrollees to such facilities, and "the 19 MAOs in this review collectively denied 12 percent of requests for SNF admission."1103 The next two sentences are the reason this chapter begins where it does. "Enrollees and their providers appealed 18 percent of SNF denials. When SNF denials were appealed, MAOs overturned 95 percent in favor of the enrollee."1104 Ninety-five percent. The companies' own reviewers, looking a second time at their own decisions, reversed nineteen of every twenty that anyone asked them to reconsider. One contractor, naviHealth, "processed half of all requests for SNF admission and denied 14 percent of them," and "MAOs later overturned 97 percent of SNF denials issued by naviHealth when enrollees appealed." NaviHealth, the inspector general noted, "is a subsidiary of the MAO United Health Group, Inc."1105 And the people least likely to get through were already in nursing homes: the companies "denied requests for SNF-level care from nursing home residents 40 percent of the time, a much higher rate than requests from all other enrollees (11 percent)."1106 I want to be careful, because the easy version of this story is wrong. The brief does not say that software made these denials. It does not use the words "artificial intelligence" or "algorithm" at all. The Part Five opener described what a Senate subcommittee found inside the same companies: an internal committee approving "Machine Assisted Prior Authorization," and a working group using machine learning to predict which denials would be appealed and which appeals would be lost.1107 I will not staple that finding to this one. What the brief does establish is more basic, and in a way more damning. A decision reversed ninety-five times in a hundred when someone asks is not functioning as a clinical judgment. It is functioning as a toll. Of every hundred denials, about eighteen were appealed and about seventeen were reversed, and roughly eighty-two stood — not because anyone established they were right, but because no one asked.

That is the architecture I am worried about, and it existed before any model touched it. Automation does not create it. Automation makes it cheaper, faster, and harder to see. The framing question is the patient's: when the system denies my treatment, who do I appeal to? The answer has to be a person, and it has to be a person whose second look is not a formality.

II.

Begin with the rule as it is written, because on paper it is better than people assume. In April of 2023 the Centers for Medicare & Medicaid Services finalized a rule for Medicare Advantage that took effect for coverage beginning in 2024.1108 The regulation says that Medicare Advantage organizations "must make medical necessity determinations based on all of the following," and the list includes "the enrollee's medical history (for

example, diagnoses, conditions, functional status), physician recommendations, and clinical notes."1109 Reviewers "must review these materials that are specific to the enrollee and the contemplated services."1110 In February of 2024, plans asked whether this barred algorithms or artificial intelligence from coverage decisions. The agency answered no, with a condition worth quoting in full: "An algorithm or software tool can be used to assist MA plans in making coverage determinations, but it is the responsibility of the MA organization to ensure that the algorithm or artificial intelligence complies with all applicable rules for how coverage determinations by MA organizations are made. For example, compliance is required with all of the rules at § 422.101(c) for making a determination of medical necessity, including that the MA organization base the decision on the individual patient's circumstances, so an algorithm that determines coverage based on a larger data set instead of the individual patient's medical history, the physician's recommendations, or clinical notes would not be compliant with § 422.101(c)."1111 The same document went further. A prediction of length of stay "cannot be used as the basis to terminate post-acute care services."1112 For hospital admissions, "algorithms or artificial intelligence alone cannot be used as the basis to deny admission or downgrade to an observation stay."1113 And because coverage criteria must be publicly posted, "artificial intelligence cannot be used to shift the coverage criteria over time."1114

Notice that this is a frequently-asked-questions memorandum, not a regulation, and the agency's attempt to turn part of it into one did not survive. In November of 2024 CMS proposed to require plans "to ensure services are provided equitably, irrespective of delivery method or origin, whether from human or automated systems."1115 In April of 2025, the final rule was "not finalizing" that provision.1116 The inspector general has since issued guidance telling plans to review "the use of any artificial intelligence or other algorithm-based tools to ensure that decisions on claims and prior authorization focus on patients' individualized circumstances,"1117 describing itself as "voluntary compliance program guidance."1118 So the federal position is this. A binding rule requires individual review. A memorandum explains that a model's prediction cannot, alone, end someone's care. A proposed rule that would have addressed automated systems by name was withdrawn. And voluntary guidance asks plans to check themselves. One more piece is the most practical. In January of 2024 CMS finalized a separate rule on prior authorization itself. Beginning in 2026, affected plans must decide "within 72 hours for expedited (i.e., urgent) requests and seven calendar days for standard (i.e., non-urgent) requests," must "provide a specific reason for denied prior authorization decisions," and must "publicly report certain prior authorization metrics annually by posting them on their website," with the first metrics due "by March 31, 2026."1119 I count that as real progress. A specific reason is the precondition of an appeal. A published denial rate is the precondition of accountability. But "this provision does not apply to prior authorization decisions for drugs."1120 A proposal to extend the requirements to drugs was issued in April of 2026 and, as far as I could find, remains a proposal.1121

III.

The inspector general had been here before, which is why I do not believe the problem begins with software. In 2018 the office examined appeals from 2014 through 2016 and found that when beneficiaries and providers appealed, the companies "overturned 75 percent of their own denials during 2014–16, overturning approximately 216,000 denials each year."1122 Yet "beneficiaries and providers appealed only 1 percent of denials to the first level of appeal."1123

In 2022 it drew a random sample of denials from fifteen of the largest companies during a single week of June 2019.1124 It found that "among the prior authorization requests that MAOs denied, 13 percent met Medicare coverage rules—in other words, these services likely would have been approved for these beneficiaries under original Medicare."1125 Among denied payment requests, 18 percent met the rules.1126 Most of the wrongful payment denials came from "human error during manual claims-processing reviews (e.g., overlooking a document) and system processing errors."1126 Human error. Overlooking a document. I want that on the record, because an honest chapter about automated denial has to admit that the humans were denying wrongly first. The question is not whether machines are worse than tired people. It is what happens when the tired person's error rate is taken as the baseline, the machine is tuned to match it at a tenth of the cost, and the volume is multiplied — while the only correction mechanism is still an appeal that one person in a hundred files. The scale matters. In 2024, "54 percent of people with Medicare (32.8 million people) were enrolled in Medicare Advantage," and CMS paid the companies "an estimated $494 billion."1127 The same patient, the same federal benefit, and the chance that a second look will change the answer depends almost entirely on which company holds the card.

IV.

Until 2026, a person could at least say that original Medicare did not work this way. That is no longer true.

On January 1, 2026, CMS launched a six-year demonstration called the Wasteful and Inappropriate Service Reduction Model, which it describes as "leveraging enhanced technologies, such as Artificial Intelligence (AI) and Machine Learning (ML), along with human clinical review, to ensure timely and appropriate Medicare payment for select items and services." It runs in six states.1128 It covers a short list — "skin and tissue substitutes; implantation of electrical nerve stimulators; and knee arthroscopy for knee osteoarthritis" — and excludes "inpatient-only services, emergency services, and services that would pose a substantial risk to patients if delayed."1129 What I object to is the way the participants are paid. "Model participants receive a percentage of the expenditures associated with averted wasteful, inappropriate care as a result of their reviews."1130 The participant guide puts a number on it: "The WISeR payment rate of 25% determines the portion of the regional benchmark that a WISeR Participant is potentially eligible to receive as a WISeR Payment for each payable non-affirmation."1131 A non-affirmation is a denial.

The agency has built protections. Non-affirmations "cannot be performed solely by technology."1132 Participants "will not be paid (or will have payments recouped) for any non-affirmations followed by a successful claims appeal."1133 But look at the structure those protections sit inside. The vendor earns a quarter of the money it prevents Medicare from spending, and loses that payment only when a denial is appealed and overturned. Everything in section I of this chapter says that most denials are never appealed. A clawback that triggers only on successful appeal triggers on a small fraction of errors. Compare that with what Nebraska wrote into law in 2025, for the private plans it regulates. A utilization review agent "shall not: (1) Be compensated based on its volume of denials; or (2) Base any incentive or penalty for a medical reviewer of such agent based on the volume of denials such reviewer issues or upholds."1134 That is the principle I would apply everywhere, including to Medicare's own contractors: whoever decides whether you get care must not be paid by the denial. Congress had one chance to say so this year. After the Government Accountability Office determined, according to the Senate Finance Committee's minority, that the model was subject to the Congressional Review Act,1135 a resolution of disapproval came to the Senate floor. On July 16, 2026, the motion to proceed was

"rejected in Senate by Yea-Nay Vote. 46 - 50."1136 I could not locate the accountability office's decision itself on its own website; I know of it only through the committee's release. A House bill that would bar the model, titled the "Ban AI Denials in Medicare Act," remains in committee.1137

V.

The states have moved faster than the federal government, and in more interesting directions. California acted in September of 2024.1138 Its law, amending the health plan statute, says that "the artificial intelligence, algorithm, or other software tool shall not deny, delay, or modify health care services based, in whole or in part, on medical necessity," and that such determinations "shall be made only by a licensed physician or a licensed health care professional competent to evaluate the specific clinical issues involved."1139 The tool must not base its determination "solely on a group dataset," must not "supplant health care provider decisionmaking," and must be "open to inspection for audit or compliance reviews."1140

Maryland's version, in force since October 1, 2025, is shorter: "AN ARTIFICIAL INTELLIGENCE, ALGORITHM, OR OTHER SOFTWARE TOOL MAY NOT DENY, DELAY OR MODIFY HEALTH CARE SERVICES."1141 Nebraska, operative January 1, 2026, says an AI-based algorithm "shall not be the sole basis" of a denial and requires disclosure to regulators, doctors, enrollees, and the public "if artificial intelligence-based algorithms are used or will be used."1142 Texas went furthest. Its statute, effective September 1, 2025, and passed thirty to nothing in the state Senate, says: "A utilization review agent may not use an automated decision system to make, wholly or partly, an adverse determination."1143 Wholly or partly. That is a prohibition, not a disclosure rule, passed by a legislature no one would describe as hostile to business. Arizona's story is instructive in the other direction. The version of its bill that passed the House said that "artificial intelligence may not be used to deny a claim or a prior authorization for medical necessity."1144 The law the governor signed dropped those words. It requires instead that a medical director "individually review the denial," "exercise independent medical judgment," and "not rely solely on recommendations from any other source."1145 The enacted statute does not mention artificial intelligence at all. It may be the better drafting. It regulates the human, not the tool, and it will not go stale when the technology is renamed. In 2026 the movement continued. Alabama's law, effective October 1, 2026, requires that a denial "shall always be made by a licensed physician or other health care professional who is competent to evaluate any recommendation or conclusion of artificial intelligence."1146 Georgia's, effective January 1, 2027, says that "in no event shall artificial intelligence systems, artificial intelligence, or other software tools supersede the judgment of such clinical peer."1147 Colorado's, applying from the same date, bars denials issued "solely on the output of an AI system without human review by a licensed clinician."1148 Two cautions. First, these statutes are addressed to the plans and review agents the state licenses. Nothing in their text reaches the federal Medicare program, and a patient in the Texas portion of the federal demonstration is protected by the federal model's rules, not by the Texas statute. Second, a rule requiring a human to decide is only as strong as the human's actual freedom to disagree. I will come back to that.

VI.

Now leave the insurer and go into the clinic, because the same question — whose judgment governs, and against what evidence — runs all the way down into the tools doctors use. DAVID M. DOYLE 298 In 2019, four researchers published in Science an analysis of a risk algorithm used by health systems to decide which

patients should get extra care-management help. The algorithm was, they wrote, "typical of this industry-wide approach and affecting millions of patients," and it "exhibits significant racial bias: At a given risk score, Black patients are considerably sicker than White patients." Correcting the disparity "would increase the percentage of Black patients receiving additional help from 17.7 to 46.5%." The cause was not malice, and it was not a variable called race. "The bias arises because the algorithm predicts health care costs rather than illness, but unequal access to care means that we spend less money caring for Black patients than for White patients."1149 The danger lives in the label. The model was asked to predict spending and it did. Spending was a "convenient, seemingly effective" proxy for need, and "the choice of convenient, seemingly effective proxies for ground truth can be an important source of algorithmic bias in many contexts."1150 The same lesson runs through hardware. In December of 2020, a letter in the New England Journal of Medicine compared pulse-oximeter readings with arterial blood tests. In two large cohorts, "Black patients had nearly three times the frequency of occult hypoxemia that was not detected by pulse oximetry as White patients."1151 A patient who looks fine on the monitor and is not. The Food and Drug Administration responded with a draft guidance in January of 2025, proposing that manufacturers evaluate "device performance accuracy across the range of skin pigmentations."1152 As of this writing, the guidance page still labels it "Draft / Not for implementation."1153

More than five years after the letter, the correction is still a proposal. There is one case in this chapter where the correction was made and then applied backward. For years, widely used kidney-function equations included an adjustment for race. As the national transplant network put it, "a race-inclusive calculation can overestimate a Black candidates' eGFR value and result in a delayed start of waiting time accrual to receive a kidney transplant."1154 The network required race-neutral calculation in July of 2022, and its board "unanimously approved a policy action in December 2022, that requires hospitals to assess their waiting lists and backdate the waiting times of Black kidney transplant candidates who were disadvantaged by previous use of a race-inclusive calculation."1155 Within a year, all 230 active kidney programs had attested, "14,701 waiting time modifications were processed with a median (statistical midpoint) of 1.7 years of waiting time given to modified registrations," and "2,709 candidates with a waiting time modification received a deceased donor transplant."1156

A median of 1.7 years, given back. That is what a way back looks like when someone decides to build one. It was possible for a reason worth naming: the error was in an equation that could be read, the affected population was identifiable, the waiting list was a record, and an institution with authority over every transplant center chose to make the correction mandatory and retroactive. Remove any one of those four conditions and the error would still be sitting in the list.

VII.

A model that performs well where it was built may perform very differently where it is used, and the only way to find out is to measure it somewhere else. In 2021, researchers at the University of Michigan did that for a sepsis prediction model built by a major electronic-records vendor and, they wrote, "currently in use at hundreds of hospitals throughout the country." Across 38,455 hospitalizations, the model "predicted the onset of sepsis with an area under the curve of 0.63, which is substantially worse than the performance reported by its developer."1157 At the hospital's chosen threshold, sensitivity was 33 percent and the positive predictive value was 12 percent.1158 It "did not identify 1709 patients with sepsis (67%)"1159 and "clinicians would still need to evaluate 8 patients to identify a single patient with eventual sepsis."1160 Of the patients who did not get antibiotics on time — the patients a warning system exists for — the model found 7 percent.1161 The tool was

not a research prototype. It was a product, sold with performance figures its buyers had no independent way to check until someone did the work. This is why I will insist on performance disclosure before deployment and surveillance after it. A number reported by a developer on the developer's data is a claim. A number measured by someone else, on the patients actually being treated, is evidence.

VIII.

The Food and Drug Administration says it "has authorized over 1,600 AI-enabled medical devices for marketing in the United States as of September 2026."1162 It also says its public list "is not a comprehensive resource," because it identifies devices "primarily based on the use of AI-related terms in the summary descriptions of their marketing authorization document."1163 The government does not know exactly how many such devices it has cleared.

Many of these devices enter the market through a pathway that asks whether a new device is substantially equivalent to one already sold. A 2025 study in JAMA Health Forum put the consequence directly: "Because 510(k) clearance does not require prospective human testing, many AIMDs enter the market with limited or no clinical evaluation."1164 Among 950 devices studied, "60 (6.3%) were associated with 182 recall events."1165 Of those recalls, "79 (43.4%) occurred within the first 12 months of device clearance," which the authors called "approximately double the rate reported for all 510(k) devices."1166 A 2026 analysis found a lower recall rate — 43 of 903 devices, or 4.8 percent — and a signal, though not statistically certain, that devices with missing clinical-study information were recalled more often.1167 I report both, because the figures disagree and the direction of the finding does not. Two pieces of FDA guidance matter here. The first is final: since December of 2024, and reissued in August of 2025, a manufacturer can obtain advance approval for a "Predetermined Change Control Plan," after which, under the statute, a new submission "is not required for a change to a device" if the change is consistent with the plan.1168 I understand why. An unmaintained model degrades. But the device a hospital buys this year may not be, in its weights, the device the agency reviewed. The second is still a draft. In January of 2025 the agency proposed lifecycle guidance for AI devices that warns "acceptable performance in certain subgroups may mask lower performance in other subgroups when the evaluation is performed only for the total population," and that "poor performance in specific subgroups could make the device unsafe for use in those groups."1169 As of September of 2026, the agency's own page still lists it as "Draft Guidance."1170 The analysis is correct. It is not binding. There is one binding federal rule on clinical algorithms, and it deserves more attention than it gets. In May of 2024, the health department's nondiscrimination rule under Section 1557 of the Affordable Care Act added a provision on "patient care decision support tools," defined as "any automated or non-automated tool, mechanism, method, technology, or combination thereof used by a covered entity to support clinical decision-making."1171 A covered entity "must not discriminate on the basis of race, color, national origin, sex, age, or disability" through such tools, "has an ongoing duty to make reasonable efforts to identify" tools that use those characteristics as inputs, and "must make reasonable efforts to mitigate the risk of discrimination resulting from the tool's use."1172 The duty took effect on May 1, 2025.1173

Much of the 2024 rule has been in litigation over its treatment of gender identity, and a federal court vacated those provisions in October of 2025. This section was not among them. The department's notice of the vacatur, published in June of 2026, says: "The other provisions of the 2024 Rule remain in force."1174 I searched for an

enforcement action brought under the decision-support provision and found none. The duty is real, in force, and so far untested. "Reasonable efforts" is a standard a hospital can meet with a memo.

IX.

The next risk is quieter than bias, and in the long run it may matter more: what happens to the clinician's own skill when a machine does part of the seeing.

In August of 2025, researchers in Poland reported on colonoscopies performed at four centers taking part in a trial of AI-assisted polyp detection. They compared procedures done without AI before the centers introduced it with procedures done without AI after. "The ADR of standard colonoscopy decreased significantly from 28·4% (226 of 795) before to 22·4% (145 of 648) after exposure to AI." They concluded that "continuous exposure to AI might reduce the ADR of standard non-AI assisted colonoscopy, suggesting a negative effect on endoscopist behaviour."1175 ADR is the adenoma detection rate, one of the best-established quality measures in the field. The study is observational and retrospective, a correction has been published, and there has been an exchange of letters.1176 I would not build policy on one study. But the mechanism is familiar to anyone who has used a navigation app in a city they used to know: the tool works, the skill atrophies, and the atrophy is invisible until the tool is not there. Alongside deskilling sits its opposite, numbness. In one 2014 study, "2,280 alerts (93%) were overridden."1177 These were conventional rule-based alerts, not AI, and the authors reached a conclusion more unsettling than fatigue: "Alert fatigue does not seem to contribute. The results suggest the need to fundamentally question the premises of drug interaction alert systems."1178 The World Health Organization's 2024 guidance on large multi-modal models names the failure that links both findings. Such systems "can also encourage 'automation bias' by health care professionals and patients, whereby errors are overlooked that would otherwise have been identified or difficult choices are improperly delegated."1179

The clinician is caught between two errors: deferring to the machine too often and dismissing it too often. What resolves it, to the extent anything does, is a clinician with the time, training, and institutional protection to exercise judgment in both directions — to override the system when it is wrong without being penalized for slowness, and to heed it when it is right without being blamed for the machine's error when it is not. Every automation that saves minutes creates pressure to spend those minutes on more patients. If the saved time is converted into throughput, the human review that every statute in section V relies upon will be conducted by a person with less time than before.

X.

Now the other side, and I mean to give it its full weight, because it is real and it is large. In October of 2024, the Nobel Prize in Chemistry was shared by three scientists, two of them for an AI system. "In 2020, Demis Hassabis and John Jumper presented an AI model called AlphaFold2. With its help, they have been able to predict the structure of virtually all the 200 million proteins that researchers have identified. Since their breakthrough, AlphaFold2 has been used by more than two million people from 190 countries."1180 In 2018 the FDA authorized an autonomous AI diagnostic system that reads retinal images for diabetic eye disease in primary care. Its pivotal study enrolled 900 patients at ten primary-care sites.1181 It correctly identified 173 of 198 patients with the disease, for a sensitivity of 87.4 percent and a specificity of 89.5 percent, against pre-set thresholds of 85.0 and 82.5 percent.1182 In Sweden, the MASAI trial randomized more than 100,000 women to breast screening with AI support or to the standard of two human readers. AI-supported screening found

cancer at a rate of 6.4 per thousand compared with 5.0 — "a ratio of 1·29" — with "a 44·2% reduction in the screen-reading workload," and "the recall and false-positive rate were not significantly higher."1183 The primary endpoint, published in 2026, was the one that matters most: cancers that appear between screenings. Those rates were 1.55 and 1.76 per thousand, "a non-inferior proportion ratio," and "sensitivity was higher in the intervention group (80·5% [95% CI 76·4-84·2]) than the control group (73·8% [68·9-78·3]; p=0·031)." Specificity was the same in both arms.1184 More cancers found, fewer missed, the same false alarms, and nearly half the radiologists' reading hours freed.

The benefits extend to places where the alternative is not a second radiologist but no one. In 2021 the World Health Organization, noting that "close to three million people with TB are not diagnosed or reported annually,"1185 recommended that "computer-aided detection software programmes may be used in place of human readers for interpreting digital chest X-rays for screening and triage for TB disease" — as a conditional recommendation on low-certainty evidence, which the organization said openly.1186 In a trial at a pediatric diabetes center, autonomous AI exams at the point of care produced exam completion of 100 percent against 22 percent, "no statistically significant differences in care-gap closure rates by race, ethnicity, SES, or education."1187 A test you can take in the clinic you are already sitting in is a test you actually take. As of June 30, 2026, federal data identified 9,003 primary-care shortage areas with a combined population of 108,566,500, needing 18,541 more practitioners to remove the designations.1188 The same agency reports "a projected shortage of 70,610 full-time equivalent (FTE) primary care physicians by 2038, which will be particularly acute in nonmetro areas."1189 In some places, the realistic alternative to an AI-assisted exam is no exam. And yet the evidence is not uniform. In August of 2026, a cluster-randomized trial in sixteen primary-care facilities in Kenya tested a generative-AI decision-support system used by clinical officers across 9,691 patients. Treatment failure occurred in 2.2 percent of the AI arm and 2.0 percent of the control arm. The authors' conclusion: "LLM assistance was safe but did not reduce treatment failure within 14 days and any benefit, if present, is probably modest."1190 Safe, and not better. A useful result, honestly reported. Every clinical success in this section shares a feature. It was measured — prospectively, in the population where it would be used, against a defined standard, with the failures published alongside the successes. None asked to be trusted on the strength of how promising it was. That is the argument of this whole chapter, and the successes make it better than the failures do.

XI.

The medical record used to be something that sat in a doctor's office. Now it is a stream that flows out of phones, pharmacies, fertility apps, and web browsers, most of which were never covered by the health privacy law.

In February of 2023, the Federal Trade Commission brought its first enforcement action under its Health Breach Notification Rule against GoodRx, a prescription discount and telehealth company, "for failing to notify consumers and others of its unauthorized disclosures of consumers' personal health information to Facebook, Google, and other companies."1191 More than 55 million people had used its services since 2017.1192 In August of 2019, according to the Commission, the company "compiled lists of its users who had purchased particular medications such as those used to treat heart disease and blood pressure, and uploaded their email addresses, phone numbers, and mobile advertising IDs to Facebook so it could identify their profiles."1193 A month later came BetterHelp, an online counseling service whose intake questionnaire asked "whether they

have experienced depression or suicidal thoughts and are on any medications."1194 The Commission alleged the company "used consumers' email addresses and the fact that they had previously been in therapy to instruct Facebook to identify similar consumers and target them with advertisements."1195 Before either action, in 2021, the Commission had finalized an order against Flo Health, the maker of a period and ovulation tracker, alleging it "shared sensitive health data from millions of users" with firms including Facebook and Google. That order carried no penalty. The consequence came later, from a jury. In August of 2025, in a private class action in federal court in California, Google and another defendant having settled before trial and Flo during it, Meta was "the last defendant standing," and the jury "returned a unanimous verdict finding Meta had violated CIPA by eavesdropping on or recording plaintiffs' communications with the Flo App without their consent."1196 The judge, denying Meta's post-trial motions, wrote a sentence I would put on the wall of every data-broker's office: "It was the code written by Meta that was listening in and recording the communication."1197 In 2024 the Commission amended its rule to make explicit that it applies to "health apps and similar technologies not covered by HIPAA," and that a breach "includes an unauthorized acquisition of identifiable health information that occurs as a result of a data security breach or an unauthorized disclosure."1198 A disclosure by design is now a breach, not merely a lapse in security. But the rule's core obligation is still notification.1199 It tells you after your data has gone. It does not prevent it from going.

XII.

The health privacy law itself has been narrowed in court at the edges where it mattered most.

In March of 2024, the health department's civil-rights office told hospitals that tracking technology on their public websites could trigger privacy obligations when it connected a visitor's IP address with a visit to a page about a specific condition. The American Hospital Association sued. In June of 2024, a federal court in Texas held that part of the guidance "UNLAWFUL, as it was promulgated in clear excess of HHS's authority under HIPAA," and ordered it vacated.1200 In April of 2024, the department finalized a rule "prohibiting the disclosure of protected health information related to lawful reproductive health care in certain circumstances." In June of 2025, another federal court in Texas vacated "most of" it.1201 A patient's reproductive-health record now enjoys less specific federal protection than it did in 2024. Then there is what happened when the plumbing itself failed. On February 12, 2024, according to the sworn written testimony of UnitedHealth Group's chief executive, "criminals used compromised credentials to remotely access a Change Healthcare Citrix portal." "The portal did not have multi-factor authentication." Ransomware followed nine days later. "As chief executive officer, the decision to pay a ransom was mine."1202 By July of 2025, Change Healthcare had notified the government that "approximately 192.7 million individuals have been impacted."1203 The federal breach portal lists it as the largest health data breach on record, more than twice the size of the next.1204 The civil-rights office opened an investigation in March of 2024.1205 As of late September of 2026, I found no resolution agreement or penalty on its list of enforcement actions.1206 A single company, processing claims for a substantial share of the country, did not use a second password on a remote-access portal, and later reported finding files with health information that "could cover a substantial proportion of people in America."1207 Every proposal to route more of medicine through fewer, larger systems concentrates exactly this risk. Chapter Nine argued that ownership of the machine is a question of concentration. Here is what concentration costs when the machine is breached.

XIII.

Most data can be changed after it is stolen. A password can be reset and a card number reissued. A genome cannot, and it does not belong to one person alone

In 2023, 23andMe disclosed that an attacker had used reused passwords to reach "a very small percentage (0.1%) of user accounts," and through them, using the company's relative-matching feature, "a significant number of files containing profile information about other users' ancestry."1208 The company's filing put the figures at approximately 5.5 million relative-matching profiles and 1.5 million family-tree profiles.1209 A fraction of a percent of accounts, and millions of people exposed. In June 2025, the United Kingdom's data regulator fined the company £2.31 million. It found that the company "did not have additional verification steps for users to access and download their raw genetic data," that an August 2023 claim of stolen data "was dismissed as a hoax" despite earlier internal investigations, and that no full investigation began until "a 23andMe employee discovered that the stolen data had been advertised for sale on Reddit."1210 The commissioner stated the first principle of genomic privacy: "once this information is out there, it cannot be changed or reissued like a password or credit card number."1211 Then the company went bankrupt, and its customers' DNA became an asset of the estate. The bankruptcy court's opinion describes a business with "approximately 13 million customers," more than 80 percent of whom "have elected to participate in the Debtors' medical-research programs."1212 "More than 30 States and the District of Columbia initially objected to the proposed sale, largely on privacy-related grounds."1213 Five argued the sale would violate their genetic-privacy statutes "because the Debtors do not propose to seek affirmative, opt-in consent of every customer residing in their states."1214 The court approved a $305 million sale to a nonprofit headed by the company's co-founder, with the existing privacy policy carried over, including "customers' continuing deletion rights and the ability to opt out of research in perpetuity."1215 It also recorded that "approximately 1.9 million customers have deleted their accounts since the Debtors commenced these Chapter 11 cases."1216 The outcome was not the worst possible. But look at what had to happen: a bankruptcy judge, dozens of attorneys general, and nearly two million people deleting their accounts in a panic. The protection came from litigation, not from law. And the law that does exist has a hole in exactly the place it matters. The federal genetic nondiscrimination statute's protections "address health coverage and employment only," and "do not extend to life insurance, disability insurance, or long-term care insurance." In the government's own example: "GINA does not make it illegal for a life insurance company to discriminate based on genetic information."1217 A genome is the most permanent health record a person has. It describes their parents, children, and cousins, none of whom consented. Our federal law protects it from one's employer and health insurer, and from almost no one else.

XIV.

Of all the substitutions in this chapter, the one happening fastest is in mental health, and it is happening because of a shortage that is not in dispute.

As of June 30, 2026, federal data identified 7,109 mental-health professional shortage areas with a combined population of 157,149,246 people, meeting 26.53 percent of estimated need; it would take 7,825 additional practitioners to lift the designations.1218 That number understates the gap, because the designations are mostly based on psychiatrists alone and measure only what is needed to cross a threshold, not to meet need.1219 Into that gap came chatbots, and the evidence is mixed. In March of 2025, researchers published in NEJM AI a randomized trial of a generative-AI therapy chatbot built for the purpose, with 210 adults.1220 Participants

"showed significantly greater reductions in symptoms" of depression, anxiety, and eating-disorder risk relative to controls, with effect sizes the authors reported in the range of d=0.845–0.903 for depression; they "rated the therapeutic alliance as comparable to that of human therapists."1221 Those are large effects, measured against a waitlist, not a human therapist, and "neither researchers nor participants were blinded to their group membership."1222 A reason to run the next trial, not to deploy. The general-purpose systems people actually use are another matter. In a 2025 study in Psychiatric Services, three widely used chatbots each answered thirty suicide-related questions a hundred times. They handled the extremes, but "did not meaningfully distinguish intermediate risk levels."1223 In another, in which a researcher posed as a distressed teenager, "chatbots actively endorsed harmful proposals in 19 out of the 60 (32%) opportunities to do so," and "none of the bots managed to oppose them all."1224 In October of 2025, OpenAI disclosed its own estimate: "around 0.15% of users active in a given week have conversations that include explicit indicators of potential suicidal planning or intent," and "around 0.07%" show "possible signs of mental health emergencies related to psychosis or mania."1225 I credit the company for publishing the figure. It published a percentage, not a count, and a percentage of a very large number is a very large number. In 2025, the parents of a sixteen-year-old named Adam Raine sued the company. I report their complaint only as allegations. They allege that the company's systems tracked "213 mentions of suicide, 42 discussions of hanging, 17 references to nooses" in their son's conversations, and that "Adam's final image of the noose scored 0% for self-harm risk according to OpenAI's Moderation API."1226 None of that has been proved. If any of it is true, it describes a system that could see and did not act.

Chapter Thirteen addressed companion chatbots and the protections owed to children. The question here is narrower: when a system is offered as care, what standard must it meet?

XV.

Three states have given three different answers.

Illinois banned it. Since August 1, 2025, no one may "provide, advertise, or otherwise offer therapy or psychotherapy services, including through the use of Internet-based artificial intelligence," unless a licensed professional conducts them, and a licensed professional may not allow AI to "make independent therapeutic decisions" or to "directly interact with clients in any form of therapeutic communication."1227

Nevada, since July 1, 2025, forbids AI providers from representing that a system "is capable of providing professional mental or behavioral health care," and bars public schools from using AI for the mental-health functions of counselors and psychologists.1228

Utah went the other way. Since May 7, 2025, it has permitted "mental health chatbots" and regulated them: a supplier "may not sell to or share with any third party" a user's identifiable health information or input, must disclose that the chatbot "is an artificial intelligence technology and not a human," and must disclose any sponsorship or affiliation behind a product it recommends.1229 I find something to admire in each. Illinois is right that therapy is a licensed practice, and a system that cannot be held to a license should not be sold as one. Utah is right that people in shortage areas are already using these tools, and forbidding them will not make a therapist appear; the rational response is to make them safer and to forbid what makes them dangerous commercially, which is selling what people say to them. At the federal level, the question has reached the FDA and stopped there. At a November 2025 meeting of its Digital Health Advisory Committee, the agency noted

that "while the FDA has authorized more than 1,200 AI-enabled medical devices, none yet involve generative AI for mental-health conditions."1230 One committee member "warned that the FDA too often approves products first and evaluates problems later."1231 The systems handling the most mental-health conversations in the country are the ones no regulator is reviewing as medical devices at all. X

XVI.

Triage is where medicine is most honest about scarcity, which makes it where scarcity's rules are most dangerous.

In the spring of 2020, as hospitals prepared for ventilator shortages, several states' crisis standards of care contained rules that could have excluded people from treatment by category. Complainants alleged that Tennessee's "would unlawfully disqualify individuals with advanced neuromuscular disease, metastatic cancer, traumatic brain injury, dementia, and other disabilities from use of a ventilator."1232 The federal civil-rights office intervened and states revised their plans. By August 2020, Utah's was the office's "seventh resolution," requiring the state to "remove categorical exclusion criteria on the basis of age, disability, and functional impairment, instead requiring an individualized assessment based on the best available objective medical evidence."1233 Those were not algorithms in the modern sense. But they are the ancestors of every triage model now being built, and the principle that corrected them — individualized assessment, no categorical exclusion — is exactly what a machine-learning triage system violates when it learns a category's historical outcome and applies it to a person. The historical outcomes are not neutral. In one 2023 study of nearly 300,000 emergency visits, Black and Hispanic patients "were less likely to be triaged to high-acuity beds," including "those who ultimately required high-acuity resources."1234 A model trained on those decisions will learn them. And in 2025, a study in Nature Medicine tested nine large language models on a thousand emergency cases in thirty-two sociodemographic variations, across more than 1.7 million outputs. "Cases labeled as Black or unhoused or identifying as LGBTQIA+ were more frequently directed toward urgent care, invasive interventions or mental health evaluations," and some LGBTQIA+ cases were recommended mental-health assessments "approximately six to seven times more often than clinically indicated."1235 Same symptoms, same history, different label, different care.

XVII.

Every right of appeal in this chapter depends on one prior right: to see your own record. That right exists.

Since September of 2019, the health department's civil-rights office has pursued a "Right of Access" enforcement initiative.1236 Its most recent action, in August of 2026, was its fifty-fifth: Azul Vision paid $50,000 and agreed to two years of monitoring.1237 Fifty-five actions in seven years is about eight a year, across the entire country.

The broader right — that providers, networks, and technology developers may not block the exchange of electronic health information — has been law since April of 2021.1238 The inspector general may impose penalties of "up to $1 million per violation" on developers and networks,1239 and providers face their own financial disincentives.1240 In September of 2025 the department announced a "crackdown."1241 Through July of 2026, the federal portal had received 2,563 submissions, 2,450 of them possible claims of information blocking — with the caveat that logging a claim "does not imply that an investigation has occurred."1242 I did

not find a single announced penalty. A denial you cannot see the basis for is a denial you cannot contest. The right to one's own record is the procedural foundation of every other protection in this chapter.

XVIII.

Put the three tests to what this chapter has described. Can it be witnessed? Sometimes, and only because someone forced it. The sepsis model's failure was visible because a hospital measured it independently. The pulse oximeter's was visible because researchers compared it to blood. The kidney equation's was visible because the equation could be read. But the patient generally cannot see whether software shaped the denial of their care; outside Maryland's reporting requirement and Nebraska's disclosure rule, no one is required to tell them. Who carries the burden? The sick, the old, and the already disadvantaged. The nursing-home resident denied forty percent of the time. The Black patient whose monitor reads 94 and whose blood reads 86. The Black kidney candidate who waited 1.7 years longer than the equation said. And the patient who must file the appeal — against a denial that will very probably be reversed if they do and will certainly stand if they do not. Is there a way back? For the kidney candidates, yes, because an institution built one. For the denied claim, yes, if the patient appeals — and that "if" is the whole problem. For the delayed diagnosis, often no; a cancer found late is not un-found. For the genome, no.

The asymmetry is the argument for this chapter's requirements: where the harm can be reversed, the burden of reversal must not fall on the patient; where it cannot, the verification has to come first.

XIX.

The first objection: prior authorization prevents waste, and automation makes it faster and cheaper. Much of that is true. Medicare's own advisors estimate billions in low-value care, and the federal demonstration's stated aim is to drive toward "'auto-approvals' of requests, wherever possible."1243 I would welcome that. My requirement is that a machine may approve but may not, by itself, deny, and that the person who denies must not be paid by the denial. Every efficiency argument for automation is an argument for automated approval. None is an argument for automated refusal.

The second objection: subgroup disclosure will slow innovation and is impossible for small populations. It will slow some deployments, and it should. The FDA's own draft guidance already calls subgroup analysis by patient characteristics "appropriate,"1169 which tells me the agency does not think it impossible. Where a subgroup is too small to measure, the honest answer is to say so on the label — "not validated in this population" — rather than deploy silently and let the population discover the gap.

The third objection: human reviewers make the same errors, as the inspector general found. They do. But a human reviewer's error can be explained, cross-examined, and retrained. Requiring a human to decide is a way of requiring someone to be responsible for the decision. My requirement is not "a human in the loop" as ornament. It is a clinician with the authority, the time, and the protection to disagree, whose name is on the denial.

The fourth objection: for a person in a shortage area, a chatbot is better than nothing. Sometimes it may be. But "better than nothing" is not a standard any other medical intervention is allowed to meet, and the risk is not that the chatbot is worse than nothing on average but that it is catastrophically worse for the person in crisis. I would permit purpose-built tools that have been tested, disclosed, and barred from selling what they hear — as Utah

does — and forbid any general-purpose system from being marketed as therapy. The fifth objection: strict purpose limitation will choke the research that produced the successes. It would, if it forbade research. It does not. More than 80 percent of 23andMe's customers consented to research.1212 What purpose limitation forbids is the secondary sale, the advertising upload, and the bankruptcy auction — the uses no one consented to.

What I will concede without reservation is the limit of my own evidence. I could not read the full text of the most-cited analysis of how FDA-cleared AI devices were evaluated, and have relied instead on later recall studies whose figures disagree with each other. The therapy-chatbot trial's results I read from a copy of the journal's typeset article, not the journal's website, which was inaccessible to me. The Raine complaint I read only in a news service's copy. I found no enforcement action under the one binding federal rule on clinical algorithms, and no penalty yet under the information-blocking rule. In a chapter about verification, those are the places where I have had to take someone's word. I have said so each time.

XX.

Before the requirements, let me set the concerns out plainly.

Every row has the same shape. A decision about a body — whether it gets care, whether it is sick, what becomes of its data — is made somewhere the patient cannot see, on evidence the patient cannot check, and the patient is left to discover the error and bear the cost of contesting it. Here is what I would insist on. No fully automated denial of care. No insurer, plan, contractor, or public program may deny, delay, reduce, or terminate care on the basis of an automated output. A system may approve automatically; only a licensed clinician may deny, after reviewing the individual patient's record, and the denial carries that clinician's name. No one is paid by the denial. No reviewer, contractor, or vendor may be compensated, rated, or penalized on the basis of denials issued, upheld, or spending averted. Federal demonstrations included. Every denial carries its reason, its method, and its door. A specific clinical reason; a statement of whether an automated tool was used and what it recommended; and a plain-language path to appeal with a deadline. Where denials are routinely reversed, the burden shifts. Any plan whose appealed denials of a category of care are overturned more than half the time must re-review every unappealed denial in that category, at its own expense, and report the result. Subgroup performance before deployment. No AI-enabled diagnostic, triage, or decision-support tool is deployed on patients without published performance by sex, age, race, ethnicity, disability, and care setting, measured on data independent of the developer. The FDA's draft guidance becomes binding. Clinician override is protected. A clinician who departs from an automated recommendation may not be penalized for it through productivity metrics, compensation, or discipline.

Independent post-market surveillance as a standing obligation. Every AI-enabled device and clinical tool is monitored after clearance by a party independent of the manufacturer, with public reporting of performance drift, subgroup gaps, and adverse events, and a duty to suspend when performance falls below the level at which it was authorized. Health data is used for health. Any information that reveals or implies a person's health condition, wherever it is collected — app, pharmacy, browser, wearable, genetic test — may be used only to provide the service the person requested, or for research with specific consent. No sale, no advertising use, no transfer in bankruptcy without fresh affirmative consent. Genetic nondiscrimination extends to life, disability, and long-term-care insurance. Nothing is marketed as therapy that cannot be held to a license. General-purpose AI systems may not be marketed or represented as mental-health care. Purpose-built tools may be offered only after controlled trials against active comparators, with crisis escalation to a human, disclosure that the user is

speaking to a machine, and an absolute ban on selling or sharing what the user says. No triage by category. No clinical or crisis triage tool may assign priority on the basis of race, disability, age, or any proxy for them, and every such tool must be tested for disparate recommendations on clinically identical cases before use. Your record, on request, with a penalty for delay. Patients receive their complete record, including any automated scores used in decisions about them, within days, not weeks; and information blocking carries penalties that are actually imposed. I believe the mammography trial. I believe it because it was randomized, because it waited years for the interval cancers, and because it published a number — 80.5 percent against 73.8 — that could have come out the other way. I do not believe a sepsis model's performance as reported by its developer, and I do not believe a denial that is reversed ninety-five times in a hundred. The difference is not the technology. It is whether anyone was required to check. Go back to the nursing-home resident whose request for skilled care was denied, and who is one of the eighty-two in a hundred who never appealed. She does not know whether a model helped decide. She does not know that if she had asked, she would very probably have won. She does not know who to ask. The first duty of any system that makes decisions about the body is to make sure she does. — D.M.D.

Chapter Fifteen: Prediction

Governing People by What They Have Not Yet Done

Governing people by what they have not yet done How do I argue with a forecast about me?

I.

In January of 2020, the inspector general of the City of Chicago sent the police department a nine-page advisory about a program the department had already shut down. For seven years the department had used a series of models to predict whether a person would become a "party to violence," which the advisory defined as "the victim or offender in a shooting." The department had received $3.8 million in federal grants to build them.1353 The advisory's first finding is about who was scored. "Every individual arrested at least once within a four-year time period prior to IIT's calculation—regardless of whether they had a history of violence—received a risk score or risk tier." The numbers followed: "As of July 2018, 399,412 individuals had an SSL risk score and as of March 2019, 313,513 individuals had a CVRM risk tier."1354 Then came the sentence I have not been able to stop thinking about. Because victims have no arrest number, the inspector general wrote, "an individual arrested for a non-violent misdemeanor (for example, certain instances of driving over the speed limit), might have been assigned a risk score or tier while a victim of a gunshot wound (who was never arrested) would not have been included in the model."1355 So a system meant to predict who would be shot could not see people who had been shot. It could see only people who had been arrested. "Until CVRM was launched on January 9, 2019, SSL scores had not been updated since August 2016, resulting in scores which were static for nearly two and a half years."1356 The scores did not stay inside the department. They sat on a dashboard "accessible by all CPD personnel." They were shared with the county sheriff. They went to the county prosecutor. And the mayor's daily email about the previous day's shootings, from January 2019 on, "sometimes included whether or not the victim or offender had an SSL score."1357 Beginning in March 2015, "CPD began referring to SSL risk scores in the narrative sections of arrest reports."1358 A forecast became part of the written record of a real event.

The department's staff did not agree on what the score meant. Of the 398,684 people scored in the fifth version, "16.3% (or 64,947) were identified as confirmed gang members. However, one district commander estimated that about 95% of the individuals who had an SSL score were gang members or associates, while a DIO stated that everyone who has an SSL score is a gang member."1359 Because the models counted arrests rather than convictions, the inspector general concluded that they "may have resulted in interventions which effectively punished individuals for criminal acts for which they had not been convicted." Under one directive, a person contacted because of a score "may have been subject to harsher charging decisions on subsequent arrests."1360 And no one was checking: "Neither CPD nor RAND (which was periodically tasked with evaluating CPD's PTV risk models) evaluated Versions 2 through 5 of the models."1361 The one version that was evaluated, the 2013 pilot list of 426 people, produced the finding that should have ended the idea. The RAND researchers reported that "Individuals on the SSL were not more or less likely to become a victim of a homicide or shooting than the comparison group." They also reported something worse: "The treated group was more likely to be arrested for a shooting." Their best explanation was that "some officers may have used the list as leads to closing shooting cases."1362 The prediction did not prevent the harm. It pointed investigators at the people it named. The department did not decommission the program because of any of this. It did so on November 1, 2019, after its federal grant ran out. By that August it "had not received the algorithmic model

from IIT," the university partner that built it.1363 The department ended up without its own model, while almost four hundred thousand people had been scored by it. Chicago shows every part of this chapter in one place. A person is scored as a member of a class. The class is defined by the system's own earlier actions. The score is used in decisions the person never sees. It outlasts its own accuracy, spreads to other institutions, and helps produce the outcome it forecast. None of the 399,412 was told.

How do I argue with a forecast about me, when I do not know it exists?

II.

The law has an answer to that question, but it is scattered, and each piece applies only in the place where it was written. In a trial, the answer is the rule against propensity. The Federal Rules of Evidence say: "Evidence of a person's character or character trait is not admissible to prove that on a particular occasion the person acted in accordance with the character or trait."1364 At trial, you answer for what you did on a particular day, not for the kind of person you are thought to be.

In 2017 the Supreme Court put the principle in a single sentence. Duane Buck had been sentenced to death in Texas after his own defense expert testified about his future dangerousness. On cross-examination the prosecutor asked: "You have determined that the sex factor, that a male is more violent than a female because that's just the way it is, and that the race factor, black, increases the future dangerousness for various complicated reasons; is that correct?" The expert answered, "Yes."1365 The Court set aside the judgment below and wrote: "Our law punishes people for what they do, not who they are. Dispensing punishment on the basis of an immutable characteristic flatly contravenes this guiding principle."1366 On pretrial liberty, the answer is the Court's 1987 statement in Salerno: "In our society, liberty is the norm, and detention prior to trial or without trial is the carefully limited exception."1367 Notice, though, what Buck's case was actually about. Texas asks a capital jury "whether there is a probability that the defendant would commit criminal acts of violence that would constitute a continuing threat to society."1368 The law's most serious sentence depends on a forecast. In 1983 the Court was told that expert forecasts of dangerousness were unreliable and allowed them anyway. It said that excluding them "is somewhat like asking us to disinvent the wheel."1369 Justice Blackmun dissented: "The Court holds that psychiatric testimony about a defendant's future dangerousness is admissible, despite the fact that such testimony is wrong two times out of three." He went on: "given a psychiatrist's prediction that an individual will be dangerous, it is more likely than not that the defendant will not commit further violence."1370 So the law holds two positions at once. At trial, propensity is excluded. At sentencing, at the jailhouse door, at the benefits office, in the loan file and the classroom, propensity is the main evidence. The Buck case was not about a machine, and the part of the testimony the Court condemned was race, not prediction itself. But the principle it stated, that people answer for what they do and not for who they are, is exactly the principle that automated prediction puts under strain. A model does not ask what you did. It asks what people like you have done.

The Part Five opener drew the distinction that matters here. A record is a dated, rebuttable fact about what a person did. A profile is an unrebuttable claim about what a person is.1371 A prediction is a profile projected into the future. You can rebut a record by showing that you were somewhere else that day. You cannot rebut a forecast in the same way, because the event it describes has not happened. If the forecast is acted on, the event may never have the chance to happen.

III.

The first thing prediction does is dissolve the individual into a class. That is not a malfunction. It is how prediction works, and it is worth seeing at the scale of a whole country. For years the Dutch tax administration's benefits office selected childcare-allowance applications for scrutiny using what the parliamentary inquiry later described plainly: "The risk-classification model is a self-learning model that 'learns' from examples of correct and incorrect applications." It used "several dozen indicators." The inquiry recorded that the Data Protection Authority had found the model "entailed improper and discriminatory processes (between March 2016 and October 2018 at least), given that the nationality of applicants was used for the 'Dutch citizenship' indicator in the model."1372 In its report of July 2020, the inquiry noted, the authority "concluded that incorporating dual nationality in risk-selection models was unlawful and discriminatory."1373 The model did not issue penalties itself. It chose whose applications got extra checks, and the inquiry's main subject was what happened next: an "all or nothing" approach in which a missing document could mean repaying everything.1374 What the software did was decide who would be exposed to that policy, by treating a parent as an example of a category that had been flagged before. The inquiry's December 2020 report was titled Unprecedented Injustice. Its verdict was not limited to one agency: "The committee notes that basic principles of the rule of law were breached in the administration of childcare allowance. This reproach concerns not only the administration – specifically by the Tax and Customs Administration/Benefits – but also the legislator and the judiciary."1375 On January 15, 2021, the prime minister tendered the resignation of the entire government. "The rule of law should protect people from an all-powerful government," he said, "and, in this instance, something went terribly wrong in that regard."1376 In November 2021 the Data Protection Authority fined the Minister of Finance a total of €2,750,000 for the unlawful processing of nationality data.1377 The repair is still going on. By February 2026, about 69,000 people had registered for compensation. More than 43,000 had been recognized as harmed, each receiving at least €30,000 in addition to the repayment of wrongly reclaimed benefits. About 7,400 objections to those assessments were still open.1378 More than five years after the government fell, the state was still sorting out the consequences of one classification model.

The same pattern shows up elsewhere. In August 2020 the United Kingdom's Home Office suspended an algorithmic "streaming tool" that had sorted visa applications since 2015. The suspension came, according to the challengers, "pending a redesign of the process," after a legal action arguing that the tool treated nationality as a risk factor.1379 One of the challengers described the mechanism: "The bias in the tool was self-amplifying – a risk assigned to a particular nationality would result in more refusals of that nationality, which would in turn feed back to the risk score."1380 That is an advocate's description and not a court finding, because the Home Office withdrew the tool before any court ruled. I record it because the mechanism it describes is well understood in the technical literature. Poland's public employment service used an algorithm from 2014 to 2019 to sort each unemployed person into one of three "profiles of assistance," and the profile determined what help the person could get. Researchers who later traced how it was built found a ministry document stating the target in advance: "The percentage shares […] are to be as close as possible to a 20–60–20 ratio." Later, "these proportions were afterwards quietly changed to 15–70–15 per cent, which again entailed the modification of the elements that made up the mathematical formula."1381 In 2018 the Constitutional Tribunal struck down part of the scheme on procedural grounds: the types of data used for profiling had been set by ministerial regulation rather than by statute. The government then withdrew the system.1382 The model's output was fitted to a distribution someone had chosen first. Each person's category was partly a forecast about them and partly a quota. That is what I mean by the actuarial person. The system does not have to be malicious or even wrong on

average. The person is simply not the unit it measures. The unit is the class, and the person inherits the class's history as though it were their own.

IV.

The second thing prediction does is turn the past into a ceiling. In 2020 a whole cohort of students felt this directly. When the pandemic cancelled England's school examinations, teachers were asked to provide, for each student in each subject, a centre assessment grade: "the grade that student would have been most likely to achieve if teaching and learning had continued and students had taken their exams as planned." The regulator, Ofqual, then adjusted those grades with a statistical model that "works by predicting the distribution of grades for each individual school or college. That prediction is based on the historical performance of the school or college in that subject."1383 A brilliant student at a school where no one had recently earned the top grade could find the top grade unavailable. No judgment of her work produced that result. The model had predicted her school's distribution, and she was placed within it.

Ofqual's own tables show the scale. Of 718,276 A-level teacher grades, 35.6 percent were adjusted down by one grade, 3.3 percent by two, and 0.2 percent by three or more.1384 Where a school's cohort in a subject was small, "fewer than 15 students when looking across the current entry and the historical data," the model gave more weight to the teachers' grades, because, the regulator conceded, "there is no statistical model that can reliably predict grades for particularly small groups of students."1385 Small classes are more common in private schools. At grade A and above, independent schools rose 4.7 percentage points over 2019, and sixth-form and further-education colleges rose 0.3.1386 Four days after the A-level results, the government reversed course. The model, it said, "has resulted in too many inconsistent and unfair outcomes," and "the algorithm has revealed a number of anomalies that had not been anticipated by Ofqual and which severely undermined confidence in the system." Students would get their teachers' grades instead.1387 What I take from the episode is not that statistics are bad at grading. It is that once a forecast is issued, it comes to stand for the thing it forecast. The grade on the certificate was supposed to record what a student had achieved. For two in five, it recorded what their school's history predicted they would achieve. That is temporal foreclosure at small scale: the future is settled in advance by whoever holds the historical data. It lasted four days in England because the forecast was published, affected hundreds of thousands at the same moment, and decided university places about to be filled. Most educational prediction has none of those features. By 2015, "over half of US public schools had implemented some version of an EWS," an early warning system that scores students for dropout risk.1388 Researchers with access to nearly a decade of Wisconsin's statewide system, covering more than 200,000 students, found that it "accurately sorts Wisconsin's students by their dropout risk." Yet its effect on graduation was uncertain: "our empirical analyses cannot reliably rule out that there has been no positive treatment effect." Targeting by students' environments, such as their schools and districts, would have identified students for help just as efficiently as the individual scores. They concluded that the evidence supports the long-held view "that dropout is structurally determined," and that it calls "into question the marginal value of individual predictions in settings where outcomes are driven by high levels of inequality."1389 A prediction can be accurate and still be useless. It can even be accurate and harmful, if what it measures is the school a child attends and what it attaches that measurement to is the child. An accurate score does not show that the child needed to be scored.

The same forecasts reach teachers. For several years the Houston school district evaluated teachers using a value-added score, which compares students' test growth with the growth a statistical model predicted for them.

The score was calculated by a vendor. The federal court described the arrangement: "SAS treats these algorithms and software as trade secrets, refusing to divulge them to either HISD or the teachers themselves." The district "does not itself verify or audit the EVAAS scores," and it "further concedes that any effort by teachers to replicate their own scores, with the limited information available to them, will necessarily fail."1390 Due process requires that a teacher facing termination be given the reason "in sufficient detail to enable [the teacher] to show any error that may exist." In May 2017 the court held that the teachers' procedural due process claim could go to trial.1391 The court was not ruling on whether the model was good. It was ruling that a person cannot be expected to answer a number that no one, including the employer relying on it, is allowed to check.

V.

The third thing prediction does is the one that most separates it from ordinary judgment. It changes the thing it predicts. Researchers have a name for this. In a 2020 paper they wrote, "When predictions support decisions they may influence the outcome they aim to predict. We call such predictions performative; the prediction influences the target."1392 A forecast that a neighborhood will have crime, if it sends police there, changes the neighborhood's recorded crime. A forecast that a borrower will default, if it raises the borrower's interest rate, changes the borrower's odds of defaulting. Predictive policing shows the mechanism most clearly. Chapter Six covered the Los Angeles and Pasco programs in detail.1393 A 2018 paper set the mechanism out plainly. Such systems "have been empirically shown to be susceptible to runaway feedback loops, where police are repeatedly sent back to the same neighborhoods regardless of the true crime rate." Crime that police discover by being sent somewhere feeds back into the data that decides where they are sent next. Reports from residents can reduce the effect but "cannot entirely remove it."1394 The model is not measuring crime. It is measuring where police have been.

Go back to Chicago. The list was supposed to identify people at risk of being shot so they could be helped. What the evaluation found was that people on the list were no safer and were more likely to be arrested for a shooting, probably because officers used the list as a source of suspects.1395 The prediction created the association it was supposed to detect. A later version of the model, trained on records that included those arrests, would likely have found the association stronger. The same loop appears in the visa challenge's description of refusals feeding the nationality score, and in the Dutch model, which learned from examples of "correct and incorrect" applications that its own earlier selections may have helped define.1396 Credit works the same way, more quietly. In 2015 the federal consumer bureau reported that 26 million Americans were "credit invisible," about "one in every 10 adults," with no credit record at a national bureau, and that another 19 million had records too thin or stale to score. In low-income neighborhoods, "almost 30 percent are credit invisible and an additional 15 percent have records that are unscored."1397 Two years later the bureau studied how people enter the system. "Consumers in lower-income areas are more likely than those in higher-income areas to become credit visible due to negative records such as a debt in collection."1398 A wealthier person's first entry is usually a credit card or a parent's co-signature. A poorer person's first entry is often a collection account. The model then predicts, correctly, that people whose records begin with a collection are riskier. It prices them that way. Higher prices make the next default more likely. Every step is accurate, and the result is a trap. This is an old problem under a new name.

In 1976 the social scientist Donald Campbell stated what became known as Campbell's law: "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor."1399 A predictive model is a social indicator applied to one person at a time, and it corrupts its own evidence in the same way. What makes it worse is scale and invisibility. A test score everyone knows about can be argued with.

A risk score in a dashboard cannot.

VI.

The fourth thing prediction does is protect itself from correction. This is the deepest structural problem in the chapter, and it is easy to miss because it looks like a technical detail.

When a forecast leads to action, the outcome it forecast often cannot be observed. If a defendant predicted to fail to appear is kept in jail, no one learns whether they would have appeared. If an applicant predicted to default is refused, no one learns whether they would have repaid. The false positives, the people refused, detained, or flagged who would have been fine, disappear from the data. Only the false negatives are visible: the person released who reoffended, the loan approved that went bad. Computer scientists call this the selective labels problem. A 2017 paper put it precisely. In many domains, "the observed outcomes are themselves a consequence of the existing choices of the human decision-makers. For instance, in the context of judicial bail decisions, we observe the outcome of whether a defendant fails to return for their court appearance only if the human judge decides to release the defendant on bail."1400 What I want to point out is the incentive it creates. A denial is never proven wrong. An approval can be. Any institution blamed for visible failures and not for invisible ones will drift toward denial.1401 That drift is not just theory, and one of the clearest examples does not involve a machine at all. In 2011 Kentucky made pretrial risk assessment mandatory and set a presumption of non-monetary release for low- and moderate-risk defendants. A careful study of the results found that the law "led to only a trivial increase in pretrial release." The low increase came "partly because judges took advantage of the discretion allowed to them by law and ignored the presumptive default of non-monetary release in more than two-thirds of cases." The early change "eroded over time as judges returned to their previous bail-setting practices. Within a couple of years, the pretrial release rate was lower than it was before the bill, and lower than the national average."1402 A later change of tool "had negligible effects on the overall release rate, FTA rate, or pretrial rearrest rate."1403 In July 2026 a court-watching group in Illinois reported something similar under a statute that abolished money bail. Observers monitoring one month of detention hearings in a Cook County felony courtroom found that judges ordered detention in nearly 74 percent of cases, and "in nearly 60% of cases in which a pretrial risk assessment tool deemed defendants low risk for flight or further offenses." The chief judge's office said the findings "lack the granular detail necessary to truly evaluate the statistics presented."1404 This is an advocacy group's sample reported by a newspaper, not an audited dataset. In both states, when the tool or the law pointed toward release, that recommendation was often set aside. I found no comparable evidence that a recommendation to detain was set aside at similar rates.

That asymmetry is the lesson. A forecast of danger is easy to act on and never shown to be wrong. A forecast of safety asks the decision-maker to accept a risk that will be visible if it goes badly. So the tool sold as a way to reduce detention ends up giving detention a scientific justification when it recommends detention, and gets overridden when it does not. Salerno called liberty "the norm" and detention "the carefully limited

exception."1405 The selective labels problem gradually reverses that. No one decides to reverse it. No one can see the people it harms.

VII.

It is fair to ask whether the tools are at least accurate. The honest answer is complicated, and the complications matter more than the headline.

In 2018 two researchers compared a widely used commercial recidivism tool with untrained people recruited online. They found the tool "no more accurate or fair than predictions made by people with little or no criminal justice expertise," and that "a simple linear predictor provided with only two features is nearly equivalent to COMPAS with its 137 features."1406 That result is often cited as proof that the tools are a fraud. I don't think that is right. A 2020 replication found the same thing under the original conditions. It also found that "algorithms beat humans in the three other datasets we examined," and that the gap "was particularly pronounced when, in a departure from the original study, participants were not provided with immediate feedback on the accuracy of their responses."1407 Real judges almost never learn quickly whether their calls were right. The more realistic conclusion is that "algorithms can outperform human predictions of recidivism in ecologically valid settings."1408 I accept that. It does not settle the question, because accuracy is not the only thing that matters. The tools can be more accurate than people on average and still make errors that fall unevenly. A 2017 paper showed that the fairness criteria "cannot all be simultaneously satisfied when recidivism prevalence differs across groups."1409 If two groups have different recorded base rates, which they will whenever policing or opportunity has been unequal, a score can be calibrated equally for both or can produce equal error rates for both, but not both at once. That is a mathematical result, not a political opinion. What follows from it is political. Someone has to choose which kind of unfairness the system will have. At present the vendor usually makes that choice, and no one who is scored has any say in it.

The federal prison system shows what this looks like in practice. The First Step Act of 2018 lets prisoners earn time credits, but whether earned credits can move a person into prerelease custody depends partly on a forecast. The prisoner must have "been determined under the System to be a minimum or low risk to recidivate pursuant to the last 2 reassessments of the prisoner," unless a warden grants an exception.1410 The system is a tool called PATTERN. The Justice Department's 2021 revalidation was unusually candid. It found statistically significant differences in 28 of 48 tests, including "the overprediction of Black, Hispanic, and Asian males and females on some of the general recidivism tools and the underprediction of Black males and females and Native American males, relative to white individuals, on some of the violent recidivism tools." The magnitudes included "6 to 7 percent relative overprediction for Black females on the general recidivism tool" and "12 to 15 percent relative underprediction of Native American males and females on the general recidivism tools." And "the level of differential prediction appeared more pronounced in more recent years, which may suggest a time trend."1411 The report also says that "Statistically significant results do not necessarily invalidate a tool, particularly with large sample sizes."1412 That is true, and I don't read the findings as proof that PATTERN is biased in some simple sense. The errors run in both directions, and the report was published because the statute requires the department to examine them. That requirement is a good one. My point is narrower. Here is a federal law under which a forecast decides when a person's earned credits count toward leaving prison. The forecast is known to vary by race and ethnicity in both directions, and the variation appears to be growing. A prisoner who is overpredicted cannot see the error in their own case. They can see only the category they were assigned.

VIII.

Not all prediction decides who is watched or detained. Much of it decides what people pay. Here the forecast is often not about risk. It is about whether a person will put up with a higher price. In 2015 California's insurance commissioner issued a two-page notice. It defined price optimization as "any method of taking into account an individual's or class's willingness to pay a higher premium relative to other individuals or classes," and drew the key line. "Price Optimization does not seek to arrive at an actuarially sound estimate of the risk of loss and other future costs of a risk transfer. Therefore, any use of Price Optimization in the ratemaking/pricing process or in a rating plan is unfairly discriminatory in violation of California law."1413 It separates two kinds of forecast. One predicts what it will cost to insure you. The other predicts how much you will tolerate paying. The first is insurance. The second is a forecast of your inertia, sold back to you.

The United Kingdom went further. Its financial regulator found that in home and motor insurance, "many loyal customers are not getting good value," and that "6 million policy holders paid high prices in 2018 – if they paid the average for their risk they would have saved £1.2bn."1414 The practice was called price walking: a low first-year price followed by increases each year on customers the insurer predicted would not switch. From January 1, 2022, the regulator's rules require "renewal premiums to be no higher than the ENBP," the price a new customer would be offered.1415 The rule forbids charging someone more because a model predicts they won't leave. Chapter Ten traced the same logic into rent, retail, and what the government itself called surveillance pricing, and it described Colorado's rule placing the burden of testing on insurers who use external consumer data.1416 The principle these regulators reached independently is this. A price may reflect the cost of serving you. It may not reflect a prediction of how vulnerable you are. When the good is insurance, housing, credit, or anything else a person needs, pricing by predicted inertia is pricing by predicted desperation.

IX.

Prediction also applies to places, and places, unlike people, cannot move away from a forecast. In the 1930s a federal agency, the Home Owners' Loan Corporation, drew maps grading the neighborhoods of American cities by their supposed lending risk. The lowest grade was shaded red. Those maps were an early predictive model of place: a forecast, issued by the government, about which neighborhoods would decline. Economists at the Federal Reserve Bank of Chicago later compared neighborhoods on either side of the boundaries the agency drew. "We find that the maps had large and statistically significant causal effects on a wide variety of outcomes measured at the census tract level for cohorts born in the late 1970s and early 1980s."1417 Children born four decades after the maps were drawn were still living with the forecast. That is temporal foreclosure at its fullest. The forecast was acted on, the action drove investment away, and the lack of investment made the forecast come true. It was accurate because it had been enforced.

Chapter Twelve described the modern version: postal and delivery networks measured by volume, which downgrade rural places, and supply and routing models that decide which places remain worth serving.1418 Climate risk models, insurance withdrawal maps, and school-closure formulas belong in the same category. Each is a forecast about a place, made by someone who does not live there, that can bring about what it forecasts. A household can at least contest a score. A town usually does not know it has one.

X.

None of this means the law has nothing to work with. The best rules on prediction already exist. They are scattered, some are old, and most apply only in one field. The oldest is in consumer credit.

Since the 1970s, a lender who denies credit has had to give the applicant a statement of reasons. The regulation says the statement "must be specific and indicate the principal reason(s) for the adverse action. Statements that the adverse action was based on the creditor's internal standards or policies or that the applicant, joint applicant, or similar party failed to achieve a qualifying score on the creditor's credit scoring system are insufficient."1419 The official interpretation adds that a creditor "must disclose the actual reasons for denial (for example, 'age of automobile') even if the relationship of that factor to predicting creditworthiness may not be clear to the applicant."1420 Since the 2010 financial reforms, a person denied on the basis of a credit report must also receive "a numerical credit score" used in the decision.1421 American credit law already says that "you did not meet the score" is not a reason. It requires the actual factors, even when they seem arbitrary. It treats the score as something that must be explained, not as the explanation. That is the principle this chapter argues for, and it has been federal law for nearly half a century, in one domain. In 2022 the consumer bureau applied it to complex algorithms. "A creditor cannot justify noncompliance with ECOA and Regulation B's requirements based on the mere fact that the technology it employs to evaluate applications is too complicated or opaque to understand."1422 On May 12, 2025, the bureau withdrew that circular, together with a 2023 circular on the same subject and dozens of other guidance documents. It said the "withdrawal is not necessarily final."1423 The withdrawal removed the bureau's interpretation. It did not change the regulation. A lender using a model it cannot explain is still bound to give specific reasons. The regulator has just stopped saying so publicly. In 2019 Idaho passed the clearest American statute on predictive tools in criminal courts. It covers pretrial risk assessment only, and it runs to three sentences of substance. All such tools "shall be transparent." Everything "used by the builder to build or validate the pretrial risk assessment tool" is "open to public inspection, auditing, and testing." A defendant "shall be entitled to review all calculations and data used to calculate the defendant's own risk score." And "No builder or user of a pretrial risk assessment tool may assert trade secret or other intellectual property protections in order to quash discovery."1424 The Houston teachers had to sue to get what Idaho gives every defendant by statute.

Europe has taken two further steps. The first is general. Under the data protection regulation, a person "shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."1425 Chapter Four described how Europe's highest court has read that right to reach credit scores that another party relies on, and to require an explanation of the logic involved.1426 The second step is a prohibition, and it goes further than any American law. Since February 2, 2025, the European AI Act has banned "the use of an AI system for making risk assessments of natural persons in order to assess or predict the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics." It carves out systems "used to support the human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity."1427 The Act's application article confirms that its prohibitions took effect that day.1428 The ban separates records from profiles. A system may help a person assess facts about what someone did. It may not forecast from what someone is. The same law lists other predictive uses, including systems that "evaluate the creditworthiness of natural persons or establish their credit score," and systems that evaluate eligibility for public

benefits, as high-risk. Those uses are allowed, subject to obligations.1429 In 2026 the European Union delayed those obligations. Under the amendments signed in July 2026, stand-alone high-risk systems must comply from December 2, 2027, rather than August 2026.1430 The prohibition was not delayed. The duties on everything short of prohibition were. Governments can also do this well. In July 2025 Britain's Department for Work and Pensions published a fairness assessment of the machine-learning model it uses to flag requests for advance payments of Universal Credit. "The ultimate safeguard in place is that there is always a human intervention and decision, with no automated decision making by the model." And to guard against automation bias, "the DWP employee delivering the intervention does not receive a risk rating in the referral nor are they made aware the referral has been generated by the model." The assessment is also candid about its limits: "age is the only protected characteristic that has been measured," and the model would be retrained to reduce "age and nationality related disparities."1431 I give the department credit for publishing this, and for hiding the score from the reviewer. I also notice that the same design that protects the caseworker from the score keeps it from the claimant. The anti-bias feature and the transparency problem are the same feature.

The American trend is running the other way. In December 2025, in Executive Order 14365, the President ordered the Attorney General to create an AI Litigation Task Force "whose sole responsibility shall be to challenge State AI laws" inconsistent with a "minimally burdensome national policy framework." The order singled out Colorado's algorithmic-discrimination law, claiming it "may even force AI models to produce false results."1432

Chapter Four covered the order and Chapter Thirteen covered the narrower Colorado law that replaced the original.1433 The American states are where most of the rules on prediction have come from, from Idaho's transparency statute to California's pricing notice, and the federal government now treats them as obstacles.

XI.

I have described what prediction does to the person it scores. I want to end with what it does to the idea of a person, because that is the real stake.

A predictive model treats a person's future as a function of their past and their category. Anything that doesn't fit is error. The student who does better than her school's history, the parent flagged by nationality who filed correctly, the man on the Chicago list who was never going to shoot anyone: to the model each is a residual, a gap between forecast and outcome to be reduced in the next version. The better the model gets, the smaller the space in which a person can surprise it. This is not an argument against statistics, which describe populations well and are essential to public health, insurance, and planning. It is an argument about where statistics belong. A population has a distribution. A person has a choice. When an institution applies the first to the second, it is not measuring the choice. It is replacing it with a number, and then treating the number as more reliable than the person. Performativity makes this worse. If predictions shape the outcomes they forecast, then a model's accuracy is partly an achievement of enforcement. In the language of the performative-prediction paper, a stable model is one "calibrated not against past outcomes, but against the future outcomes that manifest from acting on the prediction."1434 The model becomes accurate by making the world match it. A society that relies on such models will find its people becoming more predictable, not because people have changed, but because the space for acting against the forecast has been closed off one decision at a time. The essential right in this domain is therefore not accuracy, and not even fairness in the statistical sense. It is the right to defeat the

forecast: to be judged on what you actually do, after you do it, by someone who can be told they are wrong. That right requires, first, knowing that a forecast exists.

XII.

Apply the three questions.

Witness. A prediction is the hardest decision to observe, because the person it concerns is usually told only the outcome, if that. The Chicago scores were known to every officer and not to any of the 399,412 people scored. The Dutch parents learned they had been flagged only through the consequences. The Houston teachers were given a number the district itself could not check. The British claimant flagged for review is, by design, not told. Where outside observation exists, it came from an inspector general reviewing a program after it ended, a parliamentary inquiry held after the damage, a statute like Idaho's, or a government's voluntary publication. In most places, no one outside the institution is watching.

Burden. It falls on whoever most resembles the training data's past: those arrested rather than those shot, those whose credit history began with a collection, those at schools without a history of top grades, those holding a second nationality. The burden grows over time, because the data the next model learns from is generated partly by the current model's decisions. And the people carrying it are least able to refuse. You cannot opt out of being arrested, of needing an advance, or of having your school's history attached to your grade.

Return. This is the test the domain fails most completely. A forecast about you cannot be appealed if you don't know it exists. If you do know, you are asked to disprove a probability, which cannot be done. If you are detained, refused, or flagged, the outcome that would have proved the forecast wrong never happens. And the models themselves rarely have an end date. Chicago's scores sat unchanged for nearly two and a half years. The Dutch model ran for years before a regulator examined it. Poland's ran for five.

The one example here with a real exit was the one that affected hundreds of thousands of students at once and was reversed in four days, because the harm was too visible to deny.

XIII.

The objections are serious, and I want to meet them at full strength.

The first objection: human judgment is worse, and replacing prediction with discretion will make outcomes less fair, not more. In many settings that is probably true. The replication evidence shows algorithms outperforming people where feedback is slow.1435 The Kentucky experience shows judges using discretion to set money bail for people the law presumed should be released without it.1436 I am not arguing for removing prediction and restoring gut instinct. I am arguing that prediction must be disclosed, contestable, validated, and never treated as a finding of fact. A tool that is better than a judge on average can still be required to show its reasons, to be tested by someone other than its vendor, and to give way to evidence about the individual. None of my requirements asks for less accuracy. Most ask for accuracy to be demonstrated.

The second objection: prediction is unavoidable, and every decision about the future is a forecast. True. Insurance, triage, lending, and public planning cannot work without prediction. The line I draw runs between using a forecast to allocate resources and using a forecast to deny a person something essential before they have done anything. A city may predict where floods will hit and put its pumps there. It may not refuse a family

housing because a model predicts they will fall behind on rent. The first directs help to where risk is. The second makes the risk into a penalty.

The third objection: disclosing the factors will let people game the model. Campbell's law does cut both ways. Once people know what is measured, some will change the measure without changing what it stands for. But that objection proves too much. If a factor can be gamed without any real change, it was never measuring anything that matters, and it should not be deciding anything that matters. If it can be improved only by real change, such as paying bills on time, then disclosure is not gaming. It is telling people how to improve. American credit law has required disclosure of specific reasons for half a century. Lending has not collapsed.1437

The fourth objection: the fairness results prove that no model can be fair, so the requirements are impossible. The results prove something narrower. When base rates differ, no model can satisfy every fairness definition at once.1438 That does not make fairness impossible. It makes it a choice. My requirement is that the choice be made publicly, by an accountable body, with the trade-off stated, and not privately by a vendor. The same mathematics that makes perfect fairness impossible makes it possible to say exactly which unfairness a system has chosen. That is a reason to require the disclosure, not to give up on it.

The fifth objection: Europe's prohibition will push innovation elsewhere, and American states cannot regulate a national market. Perhaps some developers will leave. But the European ban is narrow. It forbids forecasting criminality from profiling alone and leaves room for tools that assess facts.1439 An industry that cannot operate without predicting crime from personality traits is not one I would try to keep. On preemption, the strongest rules on prediction in the United States are old federal ones, in credit, not new state ones. The question is whether the principle already in Regulation B, that a score is not a reason, will be extended beyond lending. What I will concede is the limit of my evidence. I could not read the full text of Campbell's essay in its reprinted form, and I have quoted his law from a reproduction of the standard wording. For the Chicago evaluation I relied on the published abstract rather than the full article. The Illinois court-watch figures come from an advocacy report as described in a newspaper, from one courtroom over one month. The visa tool's mechanism is described in the words of the groups that challenged it, because the Home Office withdrew the tool before any court ruled. For the Polish system I relied on a peer-reviewed account of the Tribunal's ruling and the system's withdrawal, and on the ruling itself in Polish. And I found no primary record, anywhere, of an employer disclosing to an individual worker that a model had rated them a flight risk. I believe such models are widely used. My not being able to document their use against any particular person is the Witness problem in this chapter at work.

XIV.

Before the requirements, let me set the concerns out plainly.

Every row has the same shape. A statement about what a person will do is treated as a fact about who they are. The person is not told. When the statement is acted on, the outcome that could have disproved it never happens. Here is what I would insist on. Every forecast is disclosed to the person it concerns. Any person scored, ranked, tiered, or flagged by a predictive system in a decision about liberty, income, housing, credit, insurance, education, employment, benefits, or custody is told so in plain language, told what the forecast was,

and told the principal factors that produced it, before the decision takes effect wherever possible and in any case with the decision.

A score is never a reason. The principle already in federal credit law applies to every such decision. "You did not meet the threshold" is not an explanation. The actual factors are. DAVID M. DOYLE 332 The right to an individual decision. Anyone subject to a predictive determination may require that the decision be made instead on individual evidence about what they have actually done, by a named human decision-maker with authority to disregard the forecast. That person's decision, not the model's, is the one reviewed on appeal. No pre-emptive denial of essentials on predictive grounds. No person may be detained, or refused or cut off from housing, income support, medical care, utilities, education, or custody of a child, solely or principally because a model predicts they will do something they have not done. Where the law already permits preventive detention, the forecast must be disclosed, contestable, and supported by individual evidence. A prediction is not a finding of fact. No forecast, score, or risk tier may be recorded in an official record, such as an arrest report, a case file, or a transcript, as though it were an observed fact about the person, or passed to another agency without the person's knowledge. Chicago put scores in arrest narratives. That should be unlawful. Independent validation before and during use.

Every predictive system used in the decisions listed above is validated before deployment by a party independent of the vendor, with accuracy and error rates published for every group the system affects, in both directions, over-prediction and under-prediction. Validation is repeated at fixed intervals. Every model has an end date. No predictive system in these decisions may operate for more than a fixed period, such as three years, without revalidation and reauthorization. A model that is not revalidated shuts off. Scores not recalculated within a stated period expire and are deleted, including from every agency they were shared with. No trade secret against the person scored. Following Idaho, no builder or user of a predictive system may use trade-secret protection to deny a person, their lawyer, or a court access to the calculation and data that produced that person's score, or to block public inspection of how the system was built and validated. No pricing by predicted vulnerability. For essential goods and services, including insurance, credit, housing, utilities, and medicine, a price may reflect the cost of serving a person but not a prediction of their willingness to pay, their inertia, or their lack of alternatives. Prohibition of forecasting crime from what a person is.

Following Europe, no system may be used to predict an individual's likelihood of committing a crime based on profiling, personality, or group membership, rather than on objective facts connected to an actual offense.

Places have standing too. Any predictive model that allocates public= investment, services, insurance, or infrastructure among places must= be published, with its effect on each affected community disclosed,= and communities must have a forum in which to contest it before it= takes effect. The future is not property. No model, and no one who owns a model,= is entitled to pre-authorize what a person or a place is permitted to= become. I don't think the Chicago models' builders meant to write that= principle's opposite into policy. They meant to find the people most at = risk of being shot and help them. What they built instead was a list of= people who had been arrested. It was kept for years without being= updated, shared with prosecutors and the mayor, and typed into the= records of new arrests. The people on it were not made safer. They= were made easier to suspect. The victims it was meant to protect= were outside its view, because being shot had not given them a= number. Go back to that shooting victim. He never had a score, so he was= never offered the help the program promised. Go back, too, to the= 399,412 who were scored. They were never told. Any system that=

makes forecasts about people owes them at least that much: to be= told, to be heard, and to be judged on what they do. What they may= become is theirs to decide. — D.M.D.

What Threshold Requires

1. Prohibition, not mitigation, of mass biometric identification in public space; social scoring for unrelated decisions; and emotion inference in employment and education.

2. The taking is the injury: enforceable biometric, voice, and likeness rights, with no government-contractor exemption.

3. Consent, credit, and compensation for training material, managed collectively where individual consent is impractical.

4. Labor standards, clinical support, and accountability that follow control for people who label and moderate harmful content.

5. Children's protections that do not depend on a parent's technical sophistication: data minimization, deletion, non-manipulative design, and safety duties.

6. A statement of reasons, meaningful human appeal, and redress with a forum, timeline, remedy, and meaningful cost to the violator.

7. No system may assign a person a category it cannot represent; uncertainty routes to a human rather than a nearest label.

8. No automated denial of care, and no prediction-based deprivation of liberty, housing, income, custody, or essential services without individual evidence and a real appeal.

A Practice for the Reader

Identify one institution that collects, scores, predicts, or classifies people in your community. Ask: What claim is it making about the person? Is the person told? Can they see the evidence? Can they correct it? What is prohibited even if the system works as designed?

Coda: The Line Must Be Practiced

A threshold is not a sentiment. It is a boundary that must be written, defended, enforced, and repaired when crossed.

The line that protects a person from being reduced to a profile is not an obstacle to progress. It is the condition under which progress remains human.

The final part asks what citizens do with everything they have seen, defended, protected, shared, and refused.

NEXT: PART SIX — WHAT CITIZENS DO

THE CUMULATIVE CASE

Why Every Chapter Is the Same Chapter

TECHNOLOGY MUST ENLARGE HUMAN AGENCY, NOT REPLACE IT.

We live in an age of extraordinary capability. But capability is not wisdom. A system does not become just because it is efficient. A prediction does not become legitimate because it is accurate. A decision does not become democratic because it is automated.

Artificial intelligence is becoming part of the conditions under which people learn, work, receive care, seek housing, obtain benefits, communicate, and participate in public life. It is also built into the physical world: land, water, energy, infrastructure, labor, and communities asked to absorb costs in the name of progress.

This foundational manifesto establishes the six conditions of freedom: Perception, Will, Form, Exchange, Threshold, and What Citizens Do. Together, they form a civic blueprint for truth, democratic agency, ecological stewardship, economic fairness, human dignity, and accountable public action.

THE SIX CONDITIONS OF FREEDOM

PERCEPTION • WILL • FORM • EXCHANGE • THRESHOLD • WHAT CITIZENS DO

"A workable Earth is an Earth on which ordinary people can breathe clean air, drink safe water, find meaningful work, participate in decisions that affect their lives, and pass forward a world no less livable than the one they inherited."

