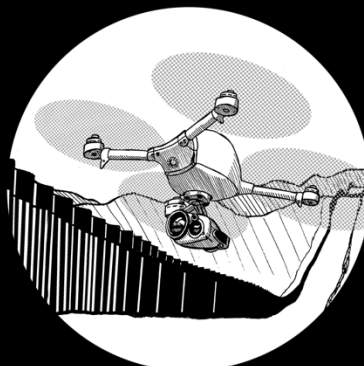
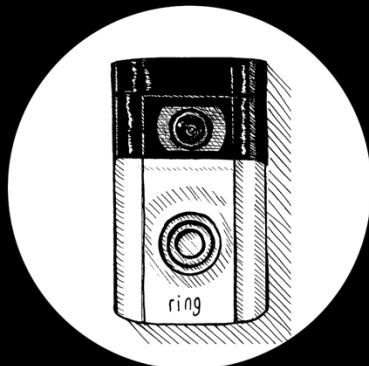
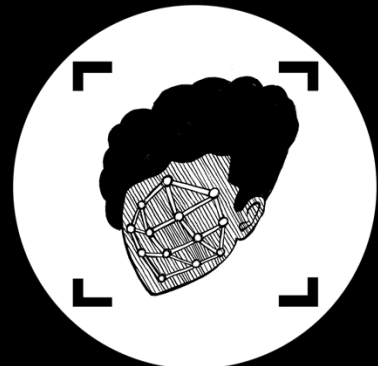
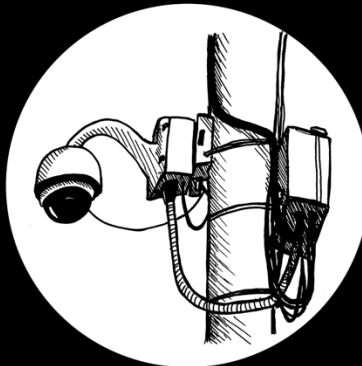
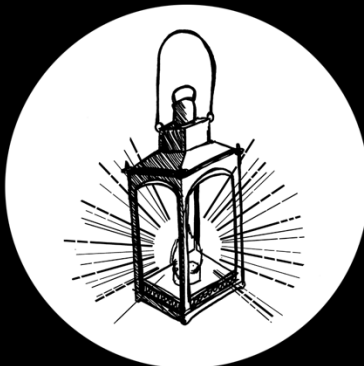
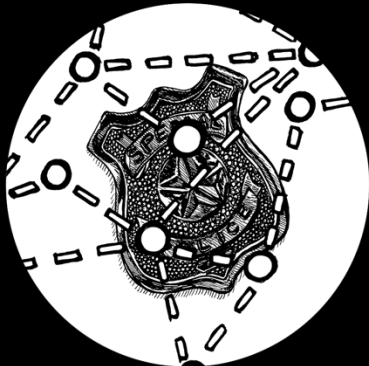


Prediction and Punishment: Critical Report on Carceral AI

Carceral AI shouldn't be used. Read on to learn why.



Written by Dasha Pruss, Hannah Pullen-Blasnik, Nikki Stevens, Shakeer Rahman, Clara Belitz, Logan Stapleton, Mallika G. Dharmaraj, Mizue Aizeki, Petra Molnar, Annika Pinch, Nathan Ryan, Thallita Lima, David Gray Widder, Amiya Tiwari, Ly Xīnzhèn Zhǎngsūn, Jason S. Sexton, and Pablo Nunes (2024).

With contributions by Colin Allen, Chaz Arnett, Laura Bingham, Erin Collins, Frances Corry, Pedro Diogo, Jessica Eaglin, Marissa Gerchick, Tobi Jegede, Gabbrielle Johnson, Donal Khosrowi, Alphoncina Lyamuya, Amreeta Mathai, Darakhshan Mir, Angie Belen Monreal, Santiago Narváez, Ngozi Okidegbe, Sarah Riley, Cierra Robson, Lian Song, Kate Weisburd, and Clifton “Skye” Williamson.

Illustrated by Dasha Pruss.

<https://carceral-ai.com>

Prediction and Punishment is a report collectively envisioned, researched, written, and edited by a group of critical researchers and activists to expose key issues at the intersection of the carceral system and artificial intelligence (AI). The report emerged from a [cross-disciplinary workshop on carceral AI](#) that took place in Pittsburgh, Pennsylvania in February 2024. The report's aim is to provide a resource for researchers, community organizers, and policy-makers to get informed about the impacts of technologies designed to police, incarcerate, surveil, and control human beings. **As a community, we stand against the use of carceral technologies.**

This report begins with an introduction that provides some context and definitions for our analysis and then continues in two main parts:

Part 1: State of Carceral AI discusses our core takeaways, including the ways that the advent of carceral AI is (and is not) novel; the perils of centering the conversation around algorithmic bias; the pernicious role of public-private partnerships; the spread of carceral AI globally and beyond the criminal legal system; the unpredictable human element in how carceral AI is used; and the incompatibility between algorithmic reforms and liberatory futures.

Part 2: Recommendations and Paths Forward discusses our suggested routes to mitigate the use and expansion of carceral AI. These include divesting from carceral technology and reducing the size and scope of the carceral system through low-tech interventions; blocking the rebranding of scrapped carceral AI systems under new names; expanding how we think about 'evidence-based' policy; increasing public access to information about carceral AI systems; building technology that intentionally centers our values; and community building to resist carceral AI.

Table of Contents

Introduction	4
What is carceral AI?	4
A note on the term ‘carceral AI’.	5
Part 1: State of Carceral AI	6
1.1 Carceral AI isn’t new.	6
1.2 The scale, scope, and opacity of carceral AI is unprecedented.	8
1.3 Bias is a red herring.	9
1.4 Carceral AI runs on public-private partnerships.	12
1.5 Carceral AI extends beyond the criminal legal system.	14
1.6 Carceral AI is used in inconsistent, discretionary and unpredictable ways.	16
1.7 Carceral AI will not lead to abolitionist futures.	17
Part 2: Recommendations and paths forward	19
2.1 Invest in communities, not carceral AI.	19
2.2 Anticipate and preemptively block new brands of carceral AI.	20
2.3 Build technology that intentionally centers the values that are important to us.	22
2.4 Update what counts as ‘evidence’ in ‘evidence-based’ policy.	23
2.5 Engage in data activism.	25
2.6 Build community and escape from disciplinary silos.	27
3 Resources	28
3.1 Glossary of terms.	28
3.2 About the workshop	29
3.3 About the authors	31
4 Acknowledgments	33

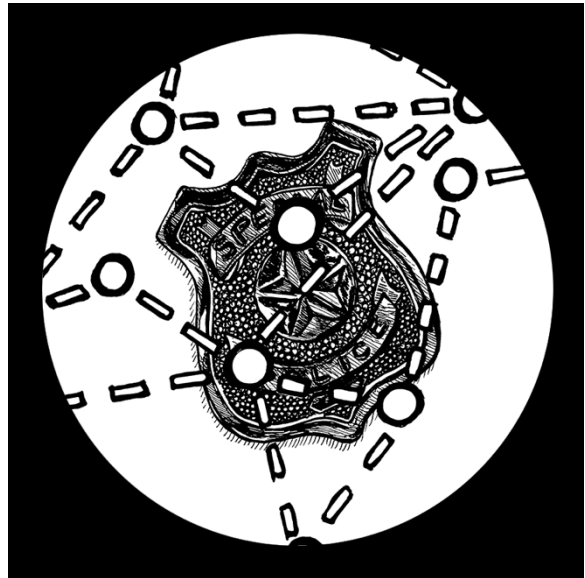
Introduction

What is carceral AI?

This report uses the term **carceral AI** to refer to a growing class of algorithmic and data-driven practices designed to police, incarcerate, surveil, and control people.¹ Carceral AI often reinforces or masks existing structural injustices, expands the reach of carceral systems under the guise of scientific rigor, and interacts in complicated ways with existing legal systems, which are ill-prepared to handle the changes introduced by such technology.

Consider a pair of programs the Los Angeles Police Department (LAPD) developed to mine massive amounts of police data and rate both people and places as sources of future crime. Titled PredPol (literally “predictive policing”) and LASER (“Los Angeles Strategic Extraction and Restoration”), the programs were sold as reforms to increase police accountability and efficiency. Both programs were built in close collaboration with university researchers and software company Palantir, leveraging LAPD’s massive budget, which over the past decade has consumed roughly half of the municipality’s annual discretionary spending.

Despite their high price tag and rhetoric of reform, in practice, carceral AI systems like PredPol and LASER cause the most harm to those our society tends to criminalize – communities of color, immigrants, people with disabilities, and poor people. An analysis by the Stop LAPD Spying Coalition found that one fifth of the people the LASER program labeled “chronic offenders” had zero prior arrests or police contact.² In a city where Black people make up less than a tenth of the population, nearly half of those targeted were Black. Numerous businesses, residences, and community gathering places were marked “crime generators,” and officers were deployed with vague profiles of who to target. LAPD killed 21 people in 2016, the year these programs expanded citywide, including six killings in LAPD’s so-called “LASER zones” in just a short 6-month period. All of the men and boys killed were Black or Latino, four were shot in the back, four were teenagers, and two were under the age of 18.



¹ We build here on the term ‘carceral technology’ used by scholars like Dr. Ruha Benjamin and activist collectives like the Carceral Tech Resistance Network and Against Carceral Tech.

² Stop LAPD Spying Coalition (2021). Automating Banishment: The Surveillance and Policing of Looted Land. <https://automatingbanishment.org/>

Although the use of carceral AI systems like predictive policing is widespread and growing, so is resistance to it. Many activists, academic researchers, community groups, policymakers and artists have stood up to oppose the use of carceral AI systems in their community. One such group of critical voices recently gathered in Pittsburgh, Pennsylvania to exchange their knowledge of carceral AI systems and craft this report.

In *Prediction and Punishment*, we discuss a range of carceral AI systems, including predictive policing tools like PredPol and LASER, facial recognition technology, recidivism risk assessment instruments, automatic license plate readers, border surveillance drones, digital immigration detention technologies, biometric databases, electronic monitoring, and audio gunshot locators (see the glossary for a definition of terms). We advocate against carceral technologies and urge the public, policymakers, and researchers to be wary of these so-called ‘smart’, ‘evidence-based’, or ‘data-driven’ reforms. We conclude by outlining a list of recommended paths forward.

A note on the term ‘carceral AI’.

We understand carceral AI³ as falling under the broader umbrella term of ‘carceral technology’. In this report we build on the work of activist collectives that have resisted and built understanding of carceral technologies, including the Carceral Tech Resistance Network,⁴ Stop LAPD Spying,⁵ O Panóptico, Against Carceral Tech, as well as Dr. Ruha Benjamin’s work on carceral technoscience.⁶ We opted here to use the somewhat more narrow term ‘carceral AI’ to emphasize the outsized role and authority given to **data, algorithms, and prediction** in the technological systems discussed in this report, which are in most cases developed with the intention of augmenting human decision-making.

However, it’s worth emphasizing that carceral AI systems vary widely in their degree of technological sophistication. For instance, facial recognition systems such as Clearview AI rely on deep neural networks – similar to the algorithm used by generative AI systems like Open AI’s Chat GPT – whereas recidivism risk assessment instruments such as the Public Safety Assessment (PSA) can be as simple as a worksheet with numerical weights assigned to each input factor. Both types of systems have been colloquially referred to as ‘AI’ and have generated excessive hype, which this report aims to dispel.

³ Artificial intelligence (AI) is a field of computer science that aims to simulate human intelligence through machine learning and deep learning in particular. In common parlance, however, AI is used to refer to a broad range of algorithmic systems, including those that do not rely on machine learning. We use the term ‘AI’ in this latter, colloquial sense.

⁴ <https://www.carceral.tech/> The Carceral Tech Resistance network uses the term ‘carceral tech’ to refer to technologies bound up in the control, coercion, capture, and exile of entire categories of people.

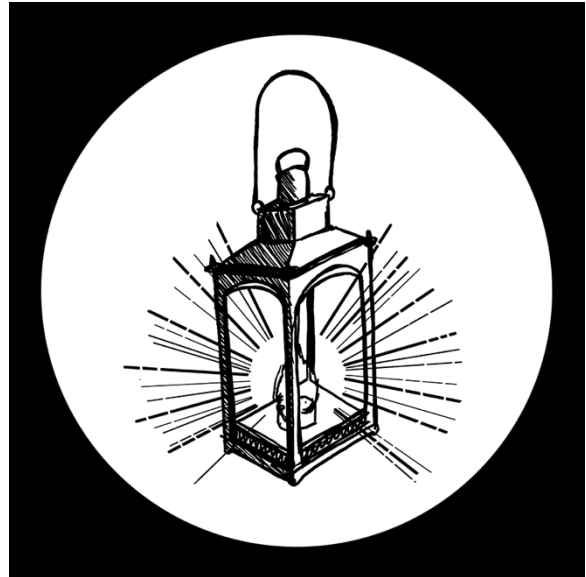
⁵ <https://stoplapdspying.org/>

⁶ Benjamin, R. (2019). *Captivating Technology: Race, Carceral Technoscience, and Liberatory Imagination in Everyday Life*. Duke University Press.

Part 1: State of Carceral AI

1.1 Carceral AI isn't new.

While advances in big data and machine learning are a relatively recent development, the broader ideologies and practices that underpin carceral AI are rooted in long histories of control and colonization of human populations. The emergence of surveillance technologies can be traced to anti-Black surveillance tactics widely used during the transatlantic slave trade, such as 18th century “lantern laws,” which required Black and Indigenous people to carry candle lanterns after dark if not accompanied by a white person.⁷ This practice of counting, surveilling, subdividing, and controlling populations through data related to race, caste, and biometric markers is grounded in **eugenics** – a set of practices that aim to increase the occurrence of ‘desirable’ human traits while eliminating the ‘undesirable’ human traits. This philosophy, which is also deeply rooted in and informed by ableism, continues to underpin many penal policies, which are built on the belief that certain parts of the human population have a proclivity for breaking laws and that these “risky” populations should be identified and isolated from the rest of society.⁸



Caste and race have also been used in colonial contexts such as India, Bengal, Tanzania, Mauritius, and Kenya, to name only a few, as a way to create boundaries and control populations using scientific racism to justify surveillance and domination.⁹ In 1920s Kenya, for instance, the British colonial government required Black African males to wear an identification document known as a *kipande*, which allowed the colonizers to segregate and restrict the movement of Black Africans.¹⁰ Other examples include the use of P.C. Mahalanobis’ anthropometric caste distances in colonial Bengal and the mobility pass for new immigrants in colonial Mauritius.

⁷ Browne, S. (2015). *Dark Matters: On the Surveillance of Blackness*. Duke University Press.

⁸ Simon, J. (2005). Positively Punitive: How the Inventor of Scientific Criminology Who Died at the Beginning of the Twentieth Century Continues to Haunt American Crime Control at the Beginning of the Twenty-First. *Texas Law Review* 84, 2135.

⁹ Bokil, A., Khare, A., Sonavane, N., Bej, S., & Janarthanan, V. (2021). Settled Habits, New Tricks: Casteist policing meets big tech in India.

¹⁰ Raman, B. (2023). Oceanic Mobility and the Empire of the Pass System. *Law and History Review*, 41(3), 565-585.

Colonization has long been a testing ground for technological innovation, and carceral technologies continue to be tested in occupied areas and at borders. Historically, technologies of enumeration like Kenya's *kipande* system have morphed into digital identification and movement monitoring projects, while systematic collection of data for immigration enforcement at the Mexico-border has been part of the United States immigration system since processing newly arrived people at Ellis Island in the 1900s, intensifying after 9/11.^{11,12} However, testing of technologies remains a present-day phenomenon. For example, Israel – one of the world's largest exporters of military equipment – tests new AI-driven surveillance and weapons technologies in occupied Palestinian territories and later exports them globally to police departments and militaries in over 100 countries, including the United States.¹³ In other words, the technologies that comprise carceral AI domestically are built on a legacy of experimentation on oppressed groups abroad.

In the US, police have often been early adopters of technology, using radio systems in the 1920s, measuring their work quantitatively since at least the 1960s,¹⁴ and commonly relying on tools such as Compstat to quantify and manage police activity in larger cities since the 1990s.¹⁵ Nevertheless, the apparent novelty of contemporary data-driven policing programs has played a role in helping police secure more resources and political cover, frequently in response to moments of social upheaval.¹⁶ Police adopt AI technologies under policing “reform” initiatives that can then be further reformed and rebranded (a 2.0 following 1.0) as needed. For example, after community organizing successfully stopped the PredPol and LASER programs described in part 1.1 above, the LAPD repackaged them into a vaguely described new initiative titled “Data-Informed Community-Focused Policing,” and the company PredPol changed its name to Geolitica before its components were ultimately purchased by a larger surveillance technology firm, SoundThinking (which itself is a rebranding of ShotSpotter, creator of an automated gunshot detection system).

Carceral AI is just one part of a complex of technologies that is fueled by longstanding geopolitical agendas of control, conquest and exclusion. Militarism, exclusionary bordering practices, and political agendas built on securitization have given rise to a global industrial complex of carceral technologies, driven by “innovation,” growth, and ultimately, profit.

Suggested readings:

- Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. John Wiley & Sons.

¹¹ <https://www.atlanticcouncil.org/event/digital-identities-and-border-cultures/>

¹² Weitzberg, K. (2020). Biometrics, race making, and white exceptionalism: The controversy over universal fingerprinting in Kenya. *The Journal of African History*, 61(1), 23-43.

¹³ Loewenstein, A. (2023). *The Palestine laboratory: How Israel exports the technology of occupation around the world*. Verso Books.

¹⁴ Fienberg S. E., A. J. Reiss, Soc. Sci. Res. Council, and Workshop Crim. Justice Stat. 1980. “Indicators of Crime and Criminal Justice: Quantitative Studies.” Washington, DC: US Dep. Justice, Bur. Justice Stat.

¹⁵ Brayne, S. (2020). *Predict and surveil: Data, discretion, and the future of policing*. Oxford University Press.

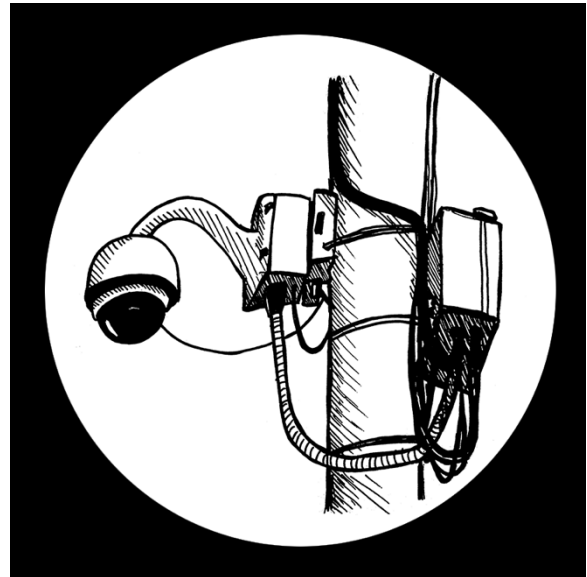
¹⁶ Porter, T. M. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton University Press.

- Browne, S. (2015). *Dark Matters: On the Surveillance of Blackness*. Duke University Press.
- Loewenstein, A. (2023). *The Palestine laboratory: How Israel exports the technology of occupation around the world*. Verso Books.
- Molnar, P. (2024). *The walls have eyes: surviving migration in the age of artificial intelligence*. The New Press.
- *Resisting Borders and Technologies of Violence* (2024). eds. Aizeki, M., Mahmoudi, M., & Schupfer, C.

1.2 The scale, scope, and opacity of carceral AI is unprecedented.

While historical perspective shows that carceral AI repackages long-standing tactics to limit human autonomy, it is important to emphasize which aspects of carceral AI *are* new. Largely, the novelty of carceral AI rests in its unprecedented **scale, scope, and opacity**.

The ability to store, combine, and query **data at scale** and from previously disparate sources gives new uses to old data collection efforts.¹⁷ Private companies also play an increasingly central role, collecting data and selling it to the government either in raw data form or in packaged tools and platforms. There has been a rise in companies specifically marketing products to the criminal legal system, such as for gunshot detection, predictive policing, video analytics and facial recognition, and DNA profiling. These companies often work across global and local scales, collecting data and testing products in carceral and/or global militarized contexts, and then marketing them for the general population. The global border industrial complex is estimated to be between \$70 billion and \$149 billion.^{18,19}



The new computational power to process data has further encouraged the **expansion of data collection**. The systematic and widespread use of surveillance, or “mass surveillance,”²⁰ has augmented the scope of how these technologies get used, placing a heavier emphasis on prediction and “proactive” decisions instead of reaction and explanation. Previous techniques

¹⁷ Brayne, S. (2017). “Big Data Surveillance: The Case of Policing.” *American Sociological Review* 82(2):977–1008.

¹⁸ <https://www.dhs.gov/sites/default/files/publications/DHS%20BIB%202019.pdf>

¹⁹ Higgins, M. (2021). How the \$68 Billion Border Surveillance Industrial Complex Affects Us All. <https://www.vice.com/en/article/how-the-dollar68-billion-border-surveillance-industrial-complex-affects-us-all/>

²⁰ Rule, J. B. (1974). *Private Lives and Public Surveillance; Social Control in the Computer Age*. First Edition. New York: Schocken Books.

are also applied more easily to new arenas where they may be more prone to errors or have unknown risks. DNA software, for example, has expanded DNA analysis into more complex mixtures that contain more people's DNA and to samples with much lower amounts of DNA, a departure from the samples on which the technology was initially developed that renders it difficult or infeasible for a human to verify the results.²¹

Finally, the **opacity of carceral AI tools** makes it more difficult to interpret their outputs, identify potential sources of error and bias, and contest the decision-making processes in their development and use.^{22,23} The quantification of scores and packaging of decision processes into software provides an air of objectivity and abstracts the discretionary choices that go into the process. This predilection for mathematical²⁴ and data-driven decision-making makes it more difficult to challenge decisions, especially where the software is proprietary and therefore hidden from cross-examination.

Suggested readings:

- Brayne, S. (2020). *Predict and Surveil: Data, discretion, and the future of policing*. Oxford University Press.
- Aizeki, M., Bingham, L., & Narváez, S. (2023). The everywhere border: Digital migration control infrastructure in the Americas.
- Molnar, P. (2024). *The walls have eyes: surviving migration in the age of artificial intelligence*. The New Press.
- Zuboff, S. (2018). *The Age of Surveillance Capitalism*. Public Affairs.
- Jefferson, B. (2020). *Digitize and punish: Racial criminalization in the digital age*. University of Minnesota Press.

1.3 Bias is a red herring.

It is now widely established that racial, class, disability, and other biases are baked into carceral AI systems,^{25,26} and much of the existing conversation about these technologies has centered attention on the 'fairness' of algorithmic predictions.^{27,28} For example, facial recognition technology has been found to have significantly higher error rates for Black faces, particularly

²¹ Pullen-Blasnik, H., Eyal, G., & Weissenbach, A. (2024). "Is your accuser me, or is it the software?" Ambiguity and contested expertise in probabilistic DNA profiling." *Social Studies of Science* 54(1): 30-58.

²² Faraj, S., Pachidi, S., & Sayegh, K. (2018). "Working and Organizing in the Age of the Learning Algorithm." *Information and Organization* 28(1):62–70.

²³ O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. 1st edition. New York: Crown.

²⁴ Gilliard, C. and Culik, H. (2018). "The New Pythagoreans." b20, July 30. <https://www.boundary2.org/2018/07/gilliard-culik/>

²⁵ Johnson, G. (2021). Algorithmic bias: on the implicit biases of social technology. *Synthese* 198:9941–9961.

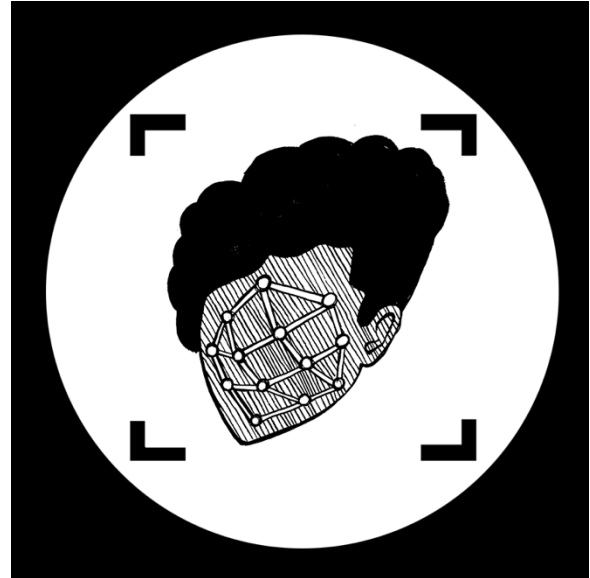
²⁶ Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2022). Machine bias. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

²⁷ Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science advances*, 4(1), eaao5580.

²⁸ Corbett-Davies, S., Gaebler, J. D., Nilforoshan, H., Shroff, R., & Goel, S. (2024). The measure and mismeasure of fairness. *J. Mach. Learn. Res.*, 24(1), 312:14730-312:14846.

women of color,²⁹ which has led police to falsely accuse and arrest Black individuals for crimes they had nothing to do with.³⁰ These disparities have fueled efforts to improve the accuracy of facial recognition and implement fairness benchmarks.³¹

While minimizing algorithmic bias can sometimes be a useful strategy for harm reduction, **too narrow a focus on making carceral AI ‘unbiased’ can work against abolitionist aims and risks obscuring the ways in which carceral technologies are inherently political.** For example, according to a report by O Panóptico, the adoption of facial recognition by police in the Brazilian state of Bahia has had no significant effect on public safety despite the technology’s steep price tag, while basic quality of life issues like sanitation receive comparatively little investment.³² Moreover, with few data protections in place, there are troubling questions about how this mass surveillance data will be used in the future and who will have access to it. In this case, framing the issue around ‘bias’ does not permit us to reject carceral AI on the basis of its societal harms; thus ‘bias’ becomes a red herring that elides more substantive critique.



A narrow focus on algorithmic fairness without enough theoretical grounding can also inadvertently make some false assumptions. First, it sometimes **assumes that non-biased software is possible.** All software contains biases, and the impulse to “eliminate bias” is actually a move towards creating software that is aligned with values of often-powerful creators, and thus perhaps *seems* unbiased. However, “unbiased software” is simply software with biases that we either do not recognize as biased or do not find problematic. A better alternative is to design software to be explicitly *biased towards our values*. We write about this more in section 2.3.

Second, it often **assumes that technological advances, if properly calibrated, will provide an advance over human decision-making.** Algorithms are often introduced as part of an

²⁹ Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR.

³⁰ Hill, K. (2023). Eight Months Pregnant and Arrested After False Facial Recognition Match. *The New York Times*. <https://www.nytimes.com/2023/08/06/business/facial-recognition-false-arrest.html>

³¹ Gerchick, M., & Cagle, M. (2024). When it Comes to Facial Recognition, There is No Such Thing as a Magic Number. *ACLU News and Commentary*. <https://www.aclu.org/news/privacy-technology/when-it-comes-to-facial-recognition-there-is-no-such-thing-as-a-magic-number>

³² Nunes, P., Lima, T. G. L., & Cruz, T. G. (2023). The hinterland will turn into sea: facial recognition expansion in Bahia. Rio de Janeiro: CESeC, O Panóptico.

“objectivity campaign”³³ that paints algorithmic decisions as more impartial, scientific, and reliable than human decision-making. This assumption, however, ignores that people are always involved in deciding how to use and interpret algorithmic outputs. Research shows that police departments use technology to distance themselves from accusations of racial bias,³⁴ judges use risk assessments differently for Black versus white defendants,³⁵ and DNA analysts alternately emphasize or minimize the software’s role in order to maintain authority.³⁶

Third, it **assumes that the variables used by carceral AI are good proxies for complex phenomena like crime**. For example, police reports are often used as a proxy for criminal activity, when they in fact measure where police have gotten involved and filed reports. Higher presence of police patrols in low-income Black neighborhoods makes police more likely to observe violations in those neighborhoods. ShotSpotter, a controversial gunshot audio detection technology, is used in many municipalities to dispatch police to areas with suspected recent gunfire, but its microphones are disproportionately installed in communities of color³⁷ and have low accuracy, leading to heightened fatal interactions with police.³⁸ Additionally, variables such as zip code are highly correlated with race, income, and disability.^{39,40} Using such proxies can lead to a self-fulfilling prophecy, where past bias produces future bias.^{41,42}

Focusing the conversation on algorithmic fairness **accepts that carceral AI will be used, it simply needs to be improved**. As a result, it limits the scope of the questions we ask to how to improve the accuracy of a tool’s output, obscuring questions such as: what are the goals and underlying values of a tool like this? Is it working to create the future we want?

Suggested readings:

- Gerchick, M., & Cagle, M. (2024). When it Comes to Facial Recognition, There is No Such Thing as a Magic Number. ACLU News and Commentary. <https://www.aclu.org/news/privacy-technology/when-it-comes-to-facial-recognition-there-is-no-such-thing-as-a-magic-number>
- Greene, D., Hoffmann, A. L., & Stark, L. (2019). Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning.

³³ Eyal, Gil. 2019. *The Crisis of Expertise*. 1st edition. Cambridge, UK ; Medford, MA: Polity.

³⁴ Ferguson, A. G. 2017. *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York University Press.

³⁵ Stevenson, M. 2018. “Assessing risk assessment in action.” *Minnesota Law Review* 103(1):303-384.

³⁶ Pullen-Blasnik, H., G. Eyal, and A. Weissenbach. 2024. “‘Is your accuser me, or is it the software?’ Ambiguity and contested expertise in probabilistic DNA profiling.” *Social Studies of Science* 54(1): 30-58. <https://doi.org/10.1177/03063127231186646>.

³⁷ Mehrotra, D. & Scott J. (2024). Here Are the Secret Locations of ShotSpotter Gunfire Sensors. *Wired*. <https://www.wired.com/story/shotspotter-secret-sensor-locations-leak/>

³⁸ #StopShotSpotter Campaign. <https://stopshotspotter.com/>

³⁹ Benjamin, R. 2019. *Race after Technology: Abolitionist Tools for the New Jim Code*. Medford, MA: Polity.

⁴⁰ Harcourt, B.E. 2015. “Risk as a Proxy for Race.” *Federal Sentencing Reporter* 27(4):237–43.

⁴¹ Ferguson, A. G. 2019. “Predictive Policing Theory.” Pp. 491–510 in *The Cambridge Handbook of Policing in the United States*, edited by T. R. Lave and E. J. Miller. Cambridge University Press.

⁴² Lum, Kristian, and William Isaac. 2016. “To Predict and Serve?” *Significance* 13(5):14–19.

- Keyes, O., Hutson, J., & Durbin, M. (2019, May). A mulching proposal: Analysing and improving an algorithmic system for turning the elderly into high-nutrient slurry. In Extended abstracts of the 2019 CHI conference on human factors in computing systems (pp. 1-11).
- Nunes, P., Lima, T. G. L., Cruz, T. G. (2023). The hinterland will turn into sea: facial recognition expansion in Bahia. Rio de Janeiro : CESeC, O Panóptico.
https://drive.google.com/file/d/1eP_M11C_P5TFGu-b9wisEQgJVSEiSNha/view
- Pruss, D. (2021). Mechanical jurisprudence and domain distortion: How predictive algorithms warp the law. *Philosophy of Science*, 88(5), 1101-1112.

1.4 Carceral AI runs on public-private partnerships.

Carceral AI is part of a broader trend of the privatization of public sector systems, including the criminal legal system, child welfare, immigration, and education.⁴³ The rise of neoliberal governance – a political-economic approach to market-oriented reform favoring deregulation and privatization – has led to outsourcing portions of work to private companies that then contract for the government.

In some cases, **government agencies hire consultant companies to build in-house programs**, such as New Orleans police hiring Palantir for predictive policing software⁴⁴ and the NYPD hiring Microsoft to build their Domain Awareness System, a mass surveillance system.⁴⁵ In other cases, **government agencies rent or subscribe to services provided by external vendors** that specialize in creating and marketing products specifically to police, courts, prisons, and immigration agencies. Some of the key vendors include Clearview AI, which provides facial recognition programs to police; Sound Thinking, which provides the audio gunshot locator Shotspotter; Geolitica, formerly known as PredPol, which provides predictive policing software; Equivant, formerly known as Northpointe, which created the notorious COMPAS risk assessment;⁴⁶ STRmix and TrueAllele, which provide



⁴³ Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

⁴⁴ Winston, A. (2018). *Palantir has secretly been using New Orleans to test its predictive policing technology*. The Verge.
<https://www.theverge.com/2018/2/27/17054740/palantir-predictive-policing-tool-new-orleans-nopd>

⁴⁵ Kwet, M. (2020). *The Microsoft Police State: Mass Surveillance, Facial Recognition, and the Azure Cloud*. The Intercept.
<https://theintercept.com/2020/07/14/microsoft-police-state-mass-surveillance-facial-recognition/>

⁴⁶ Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine Bias*. ProPublica.
<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

probabilistic DNA profiling software to forensic labs; Vigilant Solutions and Flock Safety, which create and rent automatic license plate readers to cities and police departments; private prison corporations GEO Group and CoreCivic provide electronic monitoring and “community-corrections” services to prison systems; and Palantir, which contracts software to Immigration and Customs Enforcement (ICE) to arrest parents of unaccompanied migrant children.⁴⁷

Companies present their products as more scientific, objective, and data-based than human decision-making, boasting the quantity of data at their disposal and the complexity of their algorithms. Some companies offer data sharing across police departments, allowing inter-city collaborations that were previously infeasible. The involvement of private companies also **introduces a new layer of opacity**: whereas tools built within government agencies are often subject to FOIA requests or public sharing agreements, proprietary tools do not have the same transparency requirements. This can lead to software that exacerbates existing inequalities, either by its design or due to undiscovered errors in its code.

In response to critique, **companies routinely rebrand** (Predpol becomes Geolitica, Northpointe becomes Equivant, ShotSpotter becomes SoundThinking). They also consolidate power through parent companies and **expand into other surveillance technology markets**, such as Flock Safety launching gunshot detection that pairs with their automatic license plate readers.⁴⁸

Public-private surveillance partnerships extend beyond companies that market directly to criminal legal system agencies. Flock Safety partners with homeowner associations that then share data with law enforcement as well. Amazon Ring, the video doorbell, sells homeowners’ camera footage to police departments in partnership agreements. Social media companies also share information, including in private messages, with police departments. This process of “infrastructural obfuscation” **hides surveillance technology as part of the community infrastructure**, blurring the boundaries between police and civilian surveillance.⁴⁹

Suggested readings:

- Balasubramaniam, G. S., Belitz, C., & Chan, A. S. (2024). Bridging Informational Divides: A Community-Centered Analysis of “Public Safety” Surveillance Technology. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1-13).
- Bridges, L. (2021). Infrastructural obfuscation: unpacking the carceral logics of the Ring surveillant assemblage. *Information, Communication & Society*, 24(6), 830-849.

⁴⁷ Mijente (2019). The war against immigrants: Trump’s tech tools powered by Palantir. https://mijente.net/wp-content/uploads/2019/08/Mijente-The-War-Against-Immigrants_-Trumps-Tech-Tools-Powered-by-Palantir_.pdf

⁴⁸ Miller, B. (2021). *Exclusive: Flock Safety Reveals Gunshot Sensors Tied to Cameras*. GovTech. <https://www.govtech.com/biz/exclusive-flock-safety-reveals-gunshot-sensors-tied-to-cameras>

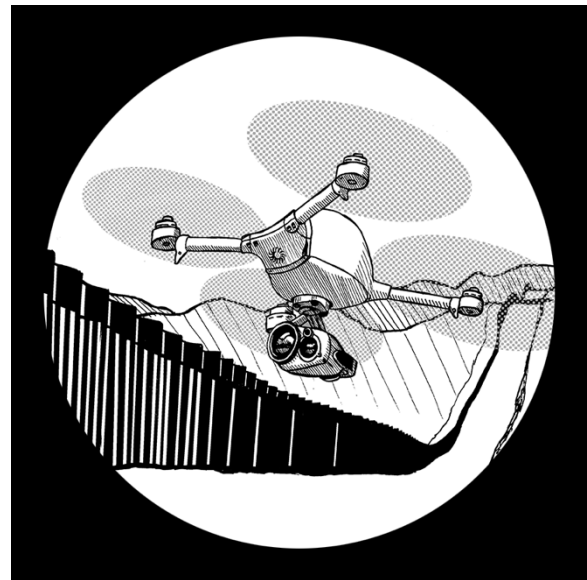
⁴⁹ Bridges, L. 2021. “Infrastructural Obfuscation: Unpacking the Carceral Logics of the Ring Surveillant Assemblage.” *Information, Communication & Society* 24(6):830–49. doi:[10.1080/1369118X.2021.1909097](https://doi.org/10.1080/1369118X.2021.1909097).

- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Mijente (2019). Who's Behind ICE? The Tech Companies Fueling Deportations. https://mijente.net/wp-content/uploads/2023/02/Who-is-Behind-ICE-The-Tech-and-Data-Companies-Fueling-Deportations_v4.pdf
- Wang, J. (2018). *Carceral capitalism* (Vol. 21). MIT Press.

1.5 Carceral AI extends beyond the criminal legal system.

Policing, surveillance, incarceration, and captivity extend beyond formal carceral institutions like prisons, police departments, or the criminal legal system. AI and digital technologies allow for new ways for policing and incarceration to be brought out of prisons into homes, schools, hospitals, refugee camps, borders, streets, and neighborhoods. We highlight three ways that carceral technologies extend incarceration to illustrate our argument.

First, technologies are used to **create de facto prisons**, often to automate incarceration outside of formal prisons. For example, e-carceration technologies like ankle monitors and cameras are used to surveil and control people serving criminal sentences in their own homes, instead of incarcerating them in prisons⁵⁰ and immigration detention facilities. Outside the criminal justice context, border control technologies like aerial drones – which are part of a patchwork of technologies that make up the so-called “smart border” – allow governments and NGOs to surveil and control the movement of individuals at the border and in refugee camps, turning those spaces into de facto prisons.⁵¹ Moreover, the deployment of carceral technologies at the US-Mexico border has increased the number of deaths in borderlands as migrants are forced to take more dangerous routes to evade detection.⁵²



Second, **technologies reify policing and surveillance** outside the formal policing or criminal legal systems, such as in care work. For example, algorithms in the child welfare or family policing system, such as the Allegheny Family Screening Tool in Pennsylvania, justify the

⁵⁰ Weisburd, K. 2023. *The Carceral Home*. GWU Law. <https://ssrn.com/abstract=4687723>.

⁵¹ Molnar, P. 2020. “Technological Testing Grounds: Migration Management Experiments and Reflections from the Ground Up.” *European Digital Rights (EDRi) and Refugee Law Lab*. <https://edri.org/our-work/technological-testing-grounds-border-tech-is-experimenting-with-peoples-lives/>.

⁵² Chambers, S. N., Boyce, G. A., Launius, S., & Dinsmore, A. (2021). Mortality, Surveillance and the Tertiary “Funnel Effect” on the U.S.-Mexico Border: A Geospatial Modeling of the Geography of Deterrence. *Journal of Borderlands Studies*, 36(3), 443–468.

investigation and separation of families, with Black and Brown parents, queer and trans parents, and disabled parents disproportionately impacted by family policing and separation.^{53,54,55} Furthermore, electronic visit verification or remote disability service provision technologies place both caregivers and the disabled people they support under digital surveillance, restricting their autonomy, privacy, and freedom of movement.^{56,57}

Third, technologies **facilitate new forms of cooperation** between police and entities outside the formal carceral system. For example, hospitals, hotlines, and social media sites like Facebook and Instagram use technologies to predict if people are experiencing suicidality, which have been used to justify police involvement through so-called "wellness checks" or involuntary detention in psychiatric facilities.^{58,59,60} Police also increasingly use social media data to secure criminal charges and convictions, such as in gang prosecutions.⁶¹ Additionally, civilians actively engage in state surveillance by sharing mugshots, doorbell camera footage, and court records online in the pursuit of personal and communal safety.⁶² This widespread sharing and accessibility of information across private and public databases perpetuates the cycle of incarceration, impeding the reintegration of formerly incarcerated individuals into society.⁶³

Suggested readings:

- Aizeki, M., Bingham, L., & Narváez, S. (2023). The everywhere border: Digital migration control infrastructure in the Americas.

⁵³ Roberts, Dorothy. Torn apart: How the child welfare system destroys Black families--and how abolition can build a safer world (2022). Basic Books.

⁵⁴ Stapleton, Logan, et al. Imagining new futures beyond predictive systems in child welfare: A qualitative study with impacted stakeholders (2022).

⁵⁵ Abdurahman, J. Khadijah. Calculating the souls of Black folk: Predictive analytics in the New York city administration for children's services (2021). In Colum. J. Race & L. Forum.

⁵⁶ Eubanks, Virginia and Mateescu, Alexandra. 'We Don't Deserve This': New App Places US Caregivers Under Digital Surveillance (July 28, 2021). Economic Hardship Reporting Project.

<https://economichardship.org/2021/07/we-dont-deserve-this-new-app-places-us-caregivers-under-digital-surveillance/>.

⁵⁷ Ohio Department of Developmental Disabilities. Technology First (Last accessed: April 7, 2024).

<https://dodd.ohio.gov/about-us/resources/tech-first/Technology-First/>.

⁵⁸ Stapleton, Logan, et al. "If This Person is Suicidal, What Do I Do?": Designing Computational Approaches to Help Online Volunteers Respond to Suicidality (2024). ACM CHI'24.

<https://doi.org/10.1145/3613904.3641922>.

⁵⁹ Marks, Mason. Artificial Intelligence Based Suicide Prediction (2019). Yale Journal of Law and Technology. <https://ssrn.com/abstract=3324874>.

⁶⁰ Martin Kaste, Facebook Increasingly Reliant on A.I. To Predict Suicide Risk (Nov. 17, 2018). NPR, All Things Considered.

<https://www.npr.org/2018/11/17/668408122/facebookincreasingly-reliant-on-a-i-to-predict-suicide-risk>.

⁶¹ Lane, J., Ramirez, F. A., & Pearce, K. E. (2018). Guilty by visible association: Socially mediated visibility in gang prosecutions. *Journal of Computer-Mediated Communication*, 23(6), 354-369.

⁶² Brayne, S., Lageson, S., & Levy, K. (2023). Surveillance deputies: When ordinary people surveil for the state. *Law & Society Review*, 57(4), 462–488. doi:10.1111/lasr.12681

⁶³ Lageson, S. E. (2020). *Digital punishment: Privacy, stigma, and the harms of data-driven criminal justice*. Oxford University Press.

- Brown, L., Shetty, R., Scherer, M., & Crawford, A. (2022). [*Ableism And Disability Discrimination In New Surveillance Technologies: How new surveillance technologies in education, policing, health care, and the workplace disproportionately harm disabled people*](#). Center for Democracy & Technology.
- Bossewitch, J., Brown, L., Gooding, P., Harris, L., Horton, J., Katteri, S., Myrick, K., Ubozoh, K., & Vasquez, A. (2021). [*Digital Futures in Mind: Reflecting on Technological Experiments in Mental Health & Crisis Support*](#). University of Melbourne.
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Molnar, P. (2024). *The walls have eyes: surviving migration in the age of artificial intelligence*. The New Press.
- Roberts, D. (2022). *Torn apart: How the child welfare system destroys Black families – and how abolition can build a safer world*. Basic Books.
- Schenwar, M. & Law, V. (2021). *Prison by any other name: The harmful consequences of popular reforms*. The New Press.

1.6 Carceral AI is used in inconsistent, discretionary and unpredictable ways.

Carceral AI is often promoted as an “evidence-based” reform to remove subjectivity from human decision-making, but the use of such technology does not remove human input so much as obfuscate it. In practice, human decision-makers like judges, prosecutors, police officers, and probation officers **differ widely** in their adherence to algorithmic recommendations and **follow them inconsistently** in different contexts, leading to new sources of conflict and less clear accountability.



While practitioners hold up the objective backing that the algorithm provides, researchers have found that algorithmic recommendations often get applied selectively and in ways that can **increase discrimination in practice**. For example, judges using a sentencing risk assessment tool in Virginia provided harsher sentences for Black defendants than for white defendants who received the same risk score.⁶⁴ In other cases, practitioners exhibit “algorithm aversion,” or resist engaging with carceral AI at all. Police officers have been found to ignore predictive policing hotspots in favor of their own judgment in an effort to reaffirm

⁶⁴ Stevenson, Megan. 2018. “Assessing risk assessment in action.” *Minnesota Law Review* 103(1):303-384.

their decision-making authority,⁶⁵ and judges and pretrial officers have been observed ignoring or overriding risk assessment recommendations.^{66,67,68} These powerful decision-makers express resentment to how the technology threatens their authority, and their resistance to its implementation can at times render it obsolete. While technology with minimal impact is better than technology with negative impact, a better alternative would be low-tech decarceration efforts that do not rely on individual decision-makers' alignment with policy goals.

Carceral AI may also be **used beyond its initial purpose** in a process deemed “function creep,” such as being repurposed to surveil workers.⁶⁹ Police agencies collect massive amounts of dragnet surveillance data without a clear plan for how it will be used or shared later on.⁷⁰ Access to technology may also further **reproduce existing power imbalances**. Prosecutors have greater access to technology providers, data, and the experts doing analysis than public defenders, exacerbating an already prevalent imbalance in the use of surveillance technology,⁷¹ risk assessment instruments,⁷² and DNA profiling.⁷³

Finally, while the exact responses to technological intervention vary, its use consistently adds a level of **opacity** into the decision-making process and **obscures accountability** for the ultimate judgment. The “black box” nature of algorithms, particularly proprietary software whose source code is hidden from public scrutiny, makes it more difficult for defendants to contest life-altering decisions or pinpoint potential sources of error.

Suggested readings:

- Albright, A. (2023). The hidden effects of algorithmic recommendations. https://apalbright.github.io/pdfs/Algo_Recs_July_2023.pdf
- Brayne, S., & Christin, A. (2021). Technologies of crime prediction: The reception of algorithms in policing and criminal courts. *Social problems*, 68(3), 608-624.

⁶⁵ Brayne, S. (2020). *Predict and surveil: Data, discretion, and the future of policing*. Oxford University Press.

⁶⁶ Christin, Angèle. 2017. “Algorithms in Practice: Comparing Web Journalism and Criminal Justice.” *Big Data & Society* July-December 2017:14.

⁶⁷ Pruss, D. (2023). Ghosting the machine: Judicial resistance to a recidivism risk assessment instrument. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 312-323).

⁶⁸ Riley, S. (2024, June). Overriding (in) justice: pretrial risk assessment administration on the frontlines. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 480-488).

⁶⁹ Brayne, Sarah and Angèle Christin. 2021. “Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts.” *Social Problems* 68(3):608–24.

⁷⁰ Brayne, S. (2020). *Predict and surveil: Data, discretion, and the future of policing*. Oxford University Press.

⁷¹ Warren, Rachel B., and Niloufar Salehi. 2022. “Trial by File Formats: Exploring Public Defenders’ Challenges Working with Novel Surveillance Data.” *Proceedings of the ACM on Human-Computer Interaction* 6(CSCW1):1–26.

⁷² Riley, Sarah A. 2024. “Overriding (In)justice: Pretrial Risk Assessment Administration on the Frontlines.” *ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*.

⁷³ Aronson, Jay D. 2007. *Genetic Witness: Science, Law, and Controversy in the Making of DNA Profiling*. None edition. New Brunswick: Rutgers University Press.

- Pruss, D. (2023). Ghosting the machine: Judicial resistance to a recidivism risk assessment instrument. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 312-323).
- Pullen-Blasnik, H., Eyal, G., & Weissenbach, A. (2024). 'Is your accuser me, or is it the software?' Ambiguity and contested expertise in probabilistic DNA profiling. *Social Studies of Science*, 54(1), 30-58.
- Riley, S. (2024). Overriding (in) justice: pretrial risk assessment administration on the frontlines. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 480-488).
- Scherer, M. (2021). [Warning: Bossware May Be Hazardous to Your Health](#). Center for Democracy & Technology.
- Stevenson, M. (2018). Assessing risk assessment in action. *Minn. L. Rev.*, 103, 303.

1.7 Carceral AI will not lead to abolitionist futures.

Carceral AI systems are often sold as reforms to reduce bias and increase efficiency in the criminal legal system. Predictive policing tools are intended to increase police accountability and efficiency; risk assessment tools are intended to make sentencing more consistent and objective; electronic monitoring is supposed to reduce overcrowding in prisons by detaining people in their homes. These reforms **limit their objectives to the maintenance and practicality of the current system**, rather than questioning the hierarchical and exploitative principles underlying the system's function. In doing so, carceral AI locks us into false dichotomies (e.g., incarcerated in a jail versus incarcerated at home), without the possibility of liberatory alternatives (e.g., not incarcerated at all).



Abolitionist philosophy calls for an end to the reliance on imprisonment and policing and proposes alternatives that engage holistically with the structural conditions underlying violence and suffering. In this vein, abolitionists often advocate for 'non-reformist reform', change that is not cordoned by what is possible within the current system but rather **change that is premised on what should be made possible long-term given human needs**.⁷⁴ Non-reformist reform prioritizes decarceration and reinvestment of resources back into communities.

By contrast, reforms grounded in carceral AI intensify eugenic logic (1.1) and tend to **pull resources away from addressing root causes of human suffering**. Carceral AI systems are

⁷⁴ Gorz, A. (1967). *Strategy for Labor*. Beacon Press.

expensive to develop, justifying increases to police and prison budgets rather than reinvesting resources into communities, such as funding reentry support programs and education. For example, the Sentence Risk Assessment Tool in Pennsylvania ultimately had no impact on sentencing, despite taking 10 years to develop and being funded by taxpayer dollars.⁷⁵

Finally, **carceral AI systems are rubber stamp reforms**, allowing legislatures to claim some measure is being taken to address the crisis of mass incarceration while leaving the core problems unaddressed. For instance, the number of people on electronic monitoring has increased fivefold from 2005 to 2021, driven in part by the COVID-19 pandemic. Despite the increase in this so-called “alternative” to incarceration, research has shown that it places intense financial and psychological burdens on monitored individuals and their families, making it more difficult for them to reintegrate into society.^{76,77} By contrast, community organizations have long advocated for low-tech non-reformist reforms to reduce prison populations, such as eliminating cash bail and releasing elderly populations from prison. See our recommendations for non-reformist reforms in Section 2.5.

Suggested readings:

- Arnett, C. (2019). From Decarceration to E-carceration. *Cardozo L. Rev.*, 41, 641.
- Barabas, C. (2022). Refusal in data ethics: Re-imagining the code beneath the code of computation in the carceral state.
- Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. John Wiley & Sons.
- Butler, P. (2015). The system is working the way it is supposed to: the limits of criminal justice reform. *Geo. LJ*, 104, 1419.
- Critical Resistance (2020). Reformist reforms vs. abolitionist steps in policing. <https://criticalresistance.org/resources/reformist-reforms-vs-abolitionist-steps-in-policing/>
- Kilgore, J. (2022). *Understanding E-Carceration: Electronic Monitoring, the Surveillance State, and the Future of Mass Incarceration*. The New Press.
- Weisburd, K. (2019). Sentenced to surveillance: Fourth Amendment limits on electronic monitoring. *NCL Rev.*, 98, 717.

⁷⁵ Pruss, D. (2023). Ghosting the machine: Judicial resistance to a recidivism risk assessment instrument. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 312-323).

⁷⁶ Kilgore, J. & Dolinar, B. (2023). Cages Without Bars Are Widening the Net: The Explosion of Electronic Monitoring

⁷⁷ Kofman, A. (2019). *Digital Jail: How Electronic Monitoring Drives Defendants Into Debt*. ProPublica. <https://www.propublica.org/article/digital-jail-how-electronic-monitoring-drives-defendants-into-debt>

Part 2: Recommendations and paths forward

2.1 Invest in communities, not carceral AI.

In part 1, we described how AI systems have expanded the reach of the carceral system beyond the prison to our neighborhoods, our borders, and our homes. Mass surveillance and sorting of people for punishment has threatened our civil liberties and has had clear harms that are disproportionately felt by marginalized communities. Consequently, our first recommendation about carceral AI is quite basic: **it should not be developed, funded, or used.**

Technology is not the solution to our crises, themselves the result of deliberate discriminatory policy choices like the legacy of the US war on crime and drugs. Instead, we advocate for **reducing the size and scope of the carceral system through low-tech community-oriented interventions.** We do not need additional surveillance or data to know that we need policy interventions that center decarceration and care-based approaches. These approaches should always be prioritized over technological innovation that replaces, expands, or reaffirms parts of the current system.



To this end, we urge researchers to be cautious when pursuing technology-based harm reduction or “AI for social good” approaches, such as recidivism and crime prediction, as these strategies may implicitly reaffirm the normalcy of the current dehumanizing system and engage in extractivist and colonial research practices. Instead, we advocate for researchers to focus their attention on technologies that center participatory and liberatory values (2.3) and engage in data activism (2.5).

We urge policymakers and the public to **divest from carceral AI systems.** Resources should instead be directed to other models of decarceration, community care, housing programs, universal healthcare and education. Formal models for violence and crime reduction include *Community Violence Intervention*, *Public Health-Based Approaches*, and *Mentorship and Workplace Programs*. For example, reentry support can assist formerly incarcerated individuals through peer mentorship, workforce development training, transitional housing, and discussion groups. By reducing the systemic barriers that people caught up in the criminal legal system face, like low employment, lack of healthcare, limited educational opportunities, and insecure housing, money can be put towards addressing root causes of incarceration and criminalization, rather than further entrenching carceral systems in our communities. Locally based initiatives

require material resources to operate and are a proven alternative to investing in expensive, unproven, and potentially harmful technologies.

Suggested reading:

- Green, B. (2019). “Good” isn’t good enough. In *Proceedings of the AI for Social Good workshop at NeurIPS* (Vol. 17).
- Peirce, J., Bailey, M., Kajeepeta, S., Crutchfield, C. (2021) A Toolkit for Jail Decarceration in Your Community. Vera Institute.
<https://www.vera.org/a-toolkit-for-jail-decarceration-in-your-community>
- Mijente (2019). Take Back Tech: How to expose and fight surveillance tech in your city.
https://mijente.net/wp-content/uploads/2019/07/Tech-Policy-Report_v4LNX.pdf
- Teng, S. and Nuñez S. (2019). Measuring Love in the Journey for Justice: A Brown Paper. <https://latinocf.org/wp-content/uploads/2019/07/Shiree-Teng-Measuring-Love.pdf>

2.2 Anticipate and preemptively block new brands of carceral AI.

As we saw in sections 1.1 and 1.4, carceral AI systems and the companies that produce them often rebrand themselves in response to public criticism, reemerging under new names and appropriating the language of public calls for accountability. Recall the rebranding of PredPol and Project LASER under the guise of “community policing” and the rebranding of the discredited gunshot audio detection system ShotSpotter as SoundThinking. It is essential to see past the marketing of carceral AI systems as innocuously improving our social problems and see them for what they are: technologies that reaffirm and amplify harmful carceral logics.

To better understand how to identify and resist new branches of carceral AI, we can learn from past successful organizing efforts. From 2014 to 2019 in Ramsey County, Minnesota (home to the capital city St. Paul), the county attorney’s office sought to implement an early-warning predictive algorithm to identify “at-risk students before they turn to crime.”⁷⁸ The county and the school district signed a data-sharing agreement that would allow the risk assessment algorithm to be trained on school data and juvenile justice records. When parents and other community members learned about these plans in 2018, they organized to oppose the data-sharing agreement, worried that the algorithm would exacerbate the well-documented school-to-prison pipeline and justify further surveillance and criminalization of Black, Brown, and Indigenous youth in the district. The county attorney’s office acquiesced and ended their plan to build the algorithm in 2019.

We highlight this example first to emphasize the community’s strategy to target not the algorithm itself, but the data sharing agreement between the school district and the county attorney’s office. The most effective strategy to contest new forms of carceral AI is to **identify and block the conditions that allow for it to exist**, before these harmful technologies have the chance to

⁷⁸ Horner, S. (2019). Ramsey County, St. Paul abandoned grant offer to address violence after mayor raised questions. *Twin Cities*.
<https://www.twincities.com/2019/11/03/ramsey-county-st-paul-crime-grant-money-violence-intervention/>

be developed. We encourage advocates and organizers to preempt overzealous data sharing practices with carceral institutions to prevent the development of new carceral technologies from collated data.

Furthermore, the Ramsey County example illustrates rhetorical strategies used to market and repurpose carceral AI to new contexts. Although the design of the algorithm proposed in Minnesota was nearly identical to the recidivism risk assessment algorithms used to judge people throughout the adult criminal legal system, the county attorney's office claimed that the new algorithm would identify which children and families "get support and services" to prevent children from turning to crime. Thus, the government attempted to **rebrand carceral AI as benevolently improving preventive and supportive measures for "at risk" groups**. We see this pattern recurring throughout domains like housing, public benefits, child welfare, and healthcare. Los Angeles County, for instance, is in the process of creating an algorithm to help child welfare workers decide which families to investigate. This algorithm is similar to other child screening algorithms like the better-known Allegheny Family Screening Tool; however, Los Angeles County claims that their algorithm will be used to identify families who "might benefit from additional engagement and support" – thereby framing carceral AI as supporting the people who are already scrutinized and surveilled by the child welfare system.⁷⁹

Suggested reading:

- Carceral Tech Resistance Network and Inter-Faith Peace and Action Collaborative (2023). on the testing + procurement of a gunfire detection surveillance system. <https://static1.squarespace.com/static/5d7edafcd15c7f734412daf2/t/641c9727c03d2c595c2729a5/1679595324529/2023-01-29+CTRN+x+IPAC+x+StopShotSpotter.pdf>
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Stop the Cradle to Prison Algorithm Coalition. Improving Outcomes for Kids & Families: Beyond Predictive Analytics & Data Sharing (Last accessed April 7, 2024). <https://www.tciamn.org/cpa-journey>.
- Stop LAPD Spying Coalition. Predictive Policing Archives (Last accessed April 7, 2024). <https://stoplapdspying.org/action/our-fights/data-driven-policing/predictive-policing/>.

2.3 Build technology that intentionally centers the values that are important to us.

As we discussed in section 1.3, **all software is biased**, at all levels.⁸⁰ This means that data is biased, methods of software creation are biased, and, of course, the results are biased. The path forward is not to seek to eliminate bias, but to be explicit about what our biases are. We can use the notion of bias to think about what we might want our software to be oriented towards – can we build software that is biased towards justice? Towards equity? Towards eliminating systemic harms? How do we start out by asking better questions that might lead to technology oriented to different problems?

⁷⁹ The Los Angeles County Risk Stratification Pilot: An Overview and One Year Update. (2022). https://dcfs.lacounty.gov/wp-content/uploads/2022/08/Risk-Stratification-One-Year-Update_8.24.22.pdf

⁸⁰ Winner, L. (2017). Do artifacts have politics?. In *Computer ethics* (pp. 177-192). Routledge.

Of course, building software based on our individual values is not a straightforward task, and there are real barriers to doing so. For example, many of us have to work jobs that do not align with our values. We might lack the technical skills to execute what we envision. We also might look around us and not see examples of others doing work that we want to emulate. However, we believe that it is possible to build technology that does not reproduce existing harms, or perhaps even sits outside of the harm-based framework.

A common question at this point is “can technology be part of a non-reformist reform?” Those who say no might insist that computing technology is grounded in capitalism and white supremacy,⁸¹ and that rather than recover the tools we have, we need to envision new tools. However, if we are frozen by the history or foundations of a tool, we will overlook many ways forward in search of something that is “purely good.” But we resist purity⁸² and instead we think with Ruth Wilson Gilmore’s invocation of Audre Lorde’s famous quote: the issue is not the “master’s tools,” but the apostrophe s in master’s. That is, the issue is the ownership and control of the tool, and not the tool itself.⁸³ Gilmore and others argue that we must be attendant to issues of ownership and control.

What might this look like in practice? It could look like beginning with a theory of change or other framework that resonates with one’s community, values, or goals. For example, one might start with activist and community organizer adrienne maree brown’s emergent strategy framework, and use that to guide each step of the development process.⁸⁴ Another approach would be to use a more “traditional” academic framework about technology and data ethics. For example, one might follow the seven principles of data feminism and apply them as recommended in that work.

Regardless of which framework one begins with, it is important to decenter the technology itself. That is, we do not recommend an approach like “we want to build an AI system, we just need to find a problem to solve.” Instead, we believe that any community situation that is improved by technology will be identified *by* that community and will be built in collaboration *with* that community. Participatory approaches to building technology have been in use for decades and while they are imperfect at best (because power imbalances are especially acute between software makers and community software users), they are a good starting point for thinking about how to engage community activists in technology development. We outline some ways to rethink what gets considered ‘evidence’ in this space in section 2.4 and ways to engage in data activism in section 2.5 – sousveillance, script flipping, and counterdata generation.

⁸¹ Dinerstein, J. (2006). Technology and Its Discontents: On the Verge of the Posthuman. *American Quarterly*, 58(3), 569–595.

⁸² Shotwell, A. (2016). *Against Purity: Living Ethically in Compromised Times* (1st edition). Univ Of Minnesota Press.

⁸³ Gilmore. (1993). Public Enemies and Private Intellectuals: Apartheid USA. *_Race & Class._*, 35_(1), 69–78.

⁸⁴ brown, adrienne maree. (2017). *Emergent Strategy: Shaping Change, Changing Worlds*. AK Press.

Suggested reading:

- Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. The MIT Press.
- D'Ignazio, C., & Klein, L. F. (2023). *Data feminism*. MIT press.
- Lewis et al. (2020). Indigenous Protocol and Artificial Intelligence Position Paper. <https://www.indigenous-ai.net/position-paper/>
- On participatory design / participatory technology creation processes:
 - Delgado, F., Barocas, S., & Levy, K. (2022). An Uncommon Task: Participatory Design in Legal AI. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1), 51:1-51:23. <https://doi.org/10.1145/3512898>
 - Delgado, F., Yang, S., Madaio, M., & Yang, Q. (2023). The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice. *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–23. <https://doi.org/10.1145/3617694.3623261>

2.4 Update what counts as ‘evidence’ in ‘evidence-based’ policy.

As we discussed in the first part of the report, carceral AI systems are often sold as ‘smart’ or ‘evidence-based’ reforms, referring to their reliance on large datasets or new technologies. However, the ‘evidence-based’ label is often a misnomer, helping states capitalize on the scientific authority that is associated with the objectivity of data and algorithmic systems while ignoring empirical and community-sourced evidence of the technologies’ on-the-ground impacts.

Sources of evidence showing that carceral AI systems are ineffective or harmful, misguided, and unwanted by marginalized communities get sidelined in evidence-based policy. In the context of ‘evidence-based’ sentencing, for instance, empirical evidence about the impacts of risk assessment instruments shows that the tools have minimal impacts on sentencing, and that those effects are distributed in arbitrary ways and tend to wash out over time.⁸⁵ Critical voices from system-impacted communities most clearly call out the harms of carceral AI systems but their concerns rarely get taken seriously as evidence.

We maintain that in order for policy around carceral AI systems to be ‘evidence-based’, the following updates should be made to how ‘evidence’ gets generated and what counts as evidence in the first place.

Embracing “small” and local data. Evidence-based policy must think beyond “big data” and speculative lab-based studies toward data on the impacts carceral AI has on the ground in real situations, organizations, and cities, with attention to interactions between people and technology. Qualitative context is important, and first-hand testimony of data gathered by local community members must play a central role.

Decentering the expertise of academic researchers. Academic researchers can play a crucial role in generating data on carceral AI systems, but their voices can easily drown out the

⁸⁵ Stevenson, M. (2018). Assessing risk assessment in action. *Minn. L. Rev.*, 103, 303.

expertise of system-impacted communities or reaffirm the systems that community organizations fight to dismantle.⁸⁶ Because most academic researchers in the global north come from privileged race, class, and able-bodied backgrounds and often have no personal relationship to incarceration, they may be blind to default dominant assumptions that pervade academic research in this space, leading too often to extractivist relationships with communities.

All experts are chosen, and we must decenter academics as experts toward academics as learners and vectors. Academic researchers without lived experience must embrace humility, accept that they are not the definitive experts, and adopt the role of learners and amplifiers, using their power strategically to make room for impacted communities to set the terms of research and conduct it themselves wherever possible.

Impacted communities must be at the helm of research on carceral AI. Participatory action research is a methodology that allows participants to directly conduct research themselves to challenge inequalities and answer questions important to them, producing knowledge in collaboration with researchers. Participants here are treated as experts in their own stories and experiences and willingly share this data with academic researchers, rather than being treated as test subjects. Citizen science is another kind of participatory research in which members of the public collaborate to gather evidence about something happening in their community.

Suggested reading:

- Dillahunt, T. R., Lu, A. J., & Velazquez, J. (2023, July). Eliciting alternative economic futures with working-class Detroiters: Centering afrofuturism in speculative design. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (pp. 957-977).
- Farrell, L., Young, B., Willison, J. B., and Fine, M. (2021). Participatory Research in Prisons. https://www.urban.org/sites/default/files/publication/104153/participatory-research-in-prisons_0.pdf
- Molnar, P. (2024). “Nothing about us without us”: People on the move interrogate border tech with the Migration and Technology Monitor. <https://www.openglobalrights.org/people-on-move-interrogate-border-tech-migration-tech-nology-monitor/>
- Prison Journalism Project. <https://prisonjournalismproject.org/>

2.5 Engage in data activism.

Having emphasized caution about the limits of academic researchers’ expertise in the context of carceral AI, we note several recommendations of concrete ways that academic researchers can contribute their particular skillsets.

Resisting disparities in access to knowledge. It is common for people in different positions of power to have unequal access to information. Consider the informational imbalances faced by

⁸⁶Green, B. (2019). “Good” isn’t good enough. In *Proceedings of the AI for Social Good workshop at NeurIPS* (Vol. 17).

public defenders and prosecutors, city residents and police officers, and undocumented people and border enforcement agents, respectively. Academic researchers have a responsibility to resist these imbalances by ensuring that their work is open access (that is, not behind a paywall), such as by using creative commons licenses, and engaging in translational work to ensure that their work is publicly legible. The latter may require academics to work in interdisciplinary partnerships – no single academic needs to have all these skills individually.

Public pedagogy. One of the ways that academics can have a positive impact is through teaching and disseminating knowledge. Academic researchers can use their privileged positions to promote work by scholars from marginalized backgrounds and engage in decentralized sites of knowledge sharing outside of standard academic publications, such as community workshops, public teach-ins, zines, websites, white papers, and policy briefs. Academics should promote decentralized networks of knowledge production and sharing and normalize the value of community engagement and relationships.

Alternate data sources and analyses. Counterdata are an alternative to data officially reported by the state and can provide data that are otherwise missing from the public record.⁸⁷ These alternate data sources can be used to challenge official definitions, measurements, and analyses, calling out state inaction, mobilizing public attention, and promoting policies to help repair communities. For example, in her book *Counting Feminicide: Data Feminism in Action*, Catherine D'Ignazio details how activists gather counterdata on feminicide to challenge piecemeal and biased state narratives throughout Latin America.

We have included below a list of recommended approaches and alternate data sources that academic researchers can engage in to actively push back on data and narratives that uphold the status quo.

- a. **Sousveillance**, a term coined by engineer Steve Mann, comes from the French word 'sous', meaning 'below', as opposed to 'sur', meaning 'above'. Unlike surveillance, in which authorities or institutions observe individuals, sousveillance inverts this power dynamic – members of the public observe authorities, such as by recording a police interaction on their phone.⁸⁸ Sousveillance is thus a powerful strategy for creating counterdata. An example of sousveillance is copwatching, in which community members collectively observe and document police abuses, with roots in the Black Panthers' armed watches of police. The [Watch the Watchers](#) project by Stop LAPD Spying is an excellent example of the power of leveraging publicly accessible data for copwatching.
- b. **Script flipping.** Instead of studying marginalized populations in low positions of power, researchers may engage in what anthropologist Laura Nader (1972) calls "studying up" – gathering data on "the most powerful strata" of society in order to "understand those who shape attitudes and actually control institutional structures." For instance, instead of

⁸⁷ Catherine D'Ignazio, *Counting Feminicide: Data Feminism in Action*. Chapter 8: A Toolkit for Counterdata Science. <https://mitpressonpubpub.mitpress.mit.edu/pub/cf-chap8/release/2>

⁸⁸ Sousveillance is an example of what anthropologist Laura Nader (1972) calls "studying up" – gathering data on "the most powerful strata" of society in order to "understand those who shape attitudes and actually control institutional structures."

making predictions about criminal defendants, such as predictions of recidivism or failure to appear in court, academics may turn the gaze upward to build algorithms that predict the risk the carceral system poses to the individual,⁸⁹ a judge's risk of failing to adhere to the law,⁹⁰ or the city blocks where financial crimes are likely to occur.⁹¹

- c. **Freedom of information laws.** Many cities, states and countries have laws requiring government transparency through the release of meeting minutes, collected datasets, voting records, and so on. These data can in principle be requested by the general public but may be inaccessible in practice. Academics are in positions of power relative to most members of the public and use their resources to request and use these data. For example, FOIA requests allowed researchers at the University of Illinois Champaign to map the usage of Automatic License Plate Readers around the county.⁹²
- d. **“New” and social media.** While social media platforms like Instagram, TikTok, and X (formerly Twitter) can be a site of surveillance and mis- and disinformation, they can also allow for the spread of alternative narratives, counterdata, and peer-to-peer storytelling. Academics can participate in these data communication narratives themselves or be supportive of those who do so. Prison TikTok is one example in which incarcerated and formerly incarcerated people post content online – answering questions, dispelling myths, and participating in social media trends and challenges – breaking down boundaries between the general public and prison life, enabling viewers to see into a realm of society that is normally closed off.⁹³ MigrantTok is another, with people on the move sharing their experiences and journey while crossing borders and navigating immigration systems.⁹⁴

Suggested reading:

- Entries on Counterdata and Missing data in *Keywords of the Datafied State*. <https://datasociety.net/library/keywords-of-the-datafied-state/>
- D'Ignazio, C., & Klein, L. F. (2023). *Data feminism*. MIT press.
- Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. The MIT Press.

⁸⁹ Meyer, M., Horowitz, A., Marshall, E., & Lum, K. (2022). Flipping the Script on Criminal Justice Risk Assessment: An actuarial model for assessing the risk the federal sentencing system poses to defendants. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 366-378).

⁹⁰ Barabas, C., Doyle, C., Rubinovitz, J. B., & Dinakar, K. (2020, January). Studying up: reorienting the study of algorithmic fairness around issues of power. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 167-176).

⁹¹ Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. John Wiley & Sons.

⁹² Balasubramaniam, G. S., Belitz, C., & Chan, A. S. (2024). Bridging Informational Divides: A Community-Centered Analysis of “Public Safety” Surveillance Technology. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

⁹³ Borello, S., Fetherston, D., & Tutrone, K. (2020). Prisoners Are Going Viral on TikTok. *VICE*. <https://www.vice.com/en/article/prison-inmates-are-going-viral-on-tiktok/>

⁹⁴ Gerber, M. (2023). ‘Their only lifeline’ for migrants at the U.S. border: smartphones and TikTok. *LA Times*. <https://www.latimes.com/business/story/2023-05-17/how-tiktok-and-other-social-media-changed-the-way-people-migrate-to-the-u-s-in-the-title-42-era>

- Meng, A., DiSalvo, C., & Zegura, E. (2019). Collaborative data work towards a caring democracy. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1-23.

2.6 Build community and escape from disciplinary silos.

Finally, studying and contesting carceral AI requires interdisciplinary and cross-community collaborations. Any particular method is by its nature a partial view of the carceral AI landscape, and people have access to different information streams depending on their positionality and approach. It is impossible to collectively reimagine alternatives to existing carceral AI systems when we are all isolated in our respective bubbles. The interdisciplinarity of the Pittsburgh carceral AI workshop was thus its major strength, revealing how researchers and activists approach and communicate about carceral AI systems in a variety of ways while converging on a common set of findings and understandings.

We want to emphasize the importance of building and strengthening bridges between affected communities, academics, activists, policymakers, artists, and coders/makers. This can take the form of future events following the format of the carceral AI workshop in Pittsburgh, community events, art exhibits, teach-ins, and virtual spaces like listservs and social media sites. We encourage anyone interested in organizing or participating in a future event like this to reach out to us at carceral.ai@gmail.com.

3 Resources

3.1 Glossary of terms

- ❖ **Predictive policing:** software used by police departments to target specific areas for patrols. These areas are determined by past data such as arrests, crime reports, and calls for service and are then used to predict the locations (or individuals) more likely to have future crime. Famously, these systems often lead to feedback loops where predictions become self-fulfilling prophecies.⁹⁵
- ❖ **Risk assessment instruments:** tools used to estimate the likelihood, or “risk,” of an (often negative) outcome based on known information. These tools are used throughout the criminal legal system to aid in decision-making, commonly in the courts during pretrial to assess bail and likelihood to appear, at sentencing for duration of sentence, at prison intake to assign to programs, at parole to decide recidivism risk, as well as in other areas of law enforcement such as determining child welfare. These tools can be predictive models that take in information about past behavior, schooling, family relations, and other available information. They may also be as technologically simple as a checklist.
- ❖ **Automatic license plate readers (ALPRs):** cameras that automatically capture license plate information of cars driving past. These cameras can be mounted to poles, streetlights, overpasses, or police patrol cars. They upload the license plate photos to a central database along with the location and time. This information can then be used to identify all cars at a location of interest or to track cars of interest across locations.⁹⁶
- ❖ **Border surveillance systems:** broad class of technologies of surveillance and migration control such as aerial drones, AI-powered towers, robodogs, biometric data collection, and other types of tools.⁹⁷
- ❖ **Electronic monitoring systems:** devices, usually attached to the ankle or wrist, that monitor a person’s location and occasionally their blood alcohol and/or breath. They are most often used as a form of digital incarceration, where the individual is prevented from leaving their house or a small radius without advance permission or they risk incarceration. Their use has become increasingly common and can be before trial, after conviction, or after release from prison. They are also used by ICE and in drug rehabilitation programs.⁹⁸
- ❖ **Audio gunshot locators:** audio sensors that are triggered by loud noises such as gunshots. These sensors are typically mounted on streetlights or buildings and, when triggered, triangulate the location of the noise and send an alert of the location and time.⁹⁹ Examples: ShotSpotter.
- ❖ **Facial recognition:** a method of identifying individuals based on facial features in a photo, video, or in person. These systems rely on using large databases of images of

⁹⁵ See a more [in-depth report on predictive policing](#) from the Electronic Frontier Foundation.

⁹⁶ See a more [in-depth report on ALPRs](#) from the Electronic Frontier Foundation.

⁹⁷ <https://www.vox.com/recode/2019/5/16/18511583/smart-border-wall-drones-sensors-ai>

⁹⁸ See a more [in-depth report on electronic monitoring](#) from the Electronic Frontier Foundation.

⁹⁹ See a more [in-depth report on gunshot detection](#) from the Electronic Frontier Foundation.

people's faces to train the algorithm. They can then be used for face verification, or confirming that a known individual matches to their known profile, or face identification, which tries to estimate the likelihood that an unknown face matches to a known profile. Facial recognition is used by police and private companies and can be used to identify someone in surveillance footage, verify someone's identity compared to their ID, or find someone in a crowd.¹⁰⁰

- ❖ **Probabilistic genotyping/Probabilistic DNA profiling:** software used to examine found DNA mixtures to determine the likelihood that a known individual contributed DNA to the sample. While DNA is unique to each individual, in practice it is often found in very small amounts, in combination with DNA from other people, and only in partial profiles (not with all the information). The software assesses the probability that a known DNA profile contributed DNA to the found mixture. DNA profiles are recorded and stored in databases ranging from local to federal levels that can be accessed for comparison in different scenarios.

3.2 About the workshop

Much of the information in this report was sourced directly from presentations by participants in [Prediction and Punishment: Cross-Disciplinary Workshop on Carceral AI](#), which took place at the Center for Philosophy of Science at the University of Pittsburgh in February 2024. The event was organized by Dasha Pruss and Colin Allen and was generously supported by the Center for Philosophy of Science at the University of Pittsburgh, the Center for Ethics and Policy at Carnegie Mellon University, Pitt Cyber, the Embedded EthiCS program at Harvard University, Professor Gayle Rogers, and Professor Ronald Brand.

Below, you can find the full list of workshop presentations.

Keynote Speakers

Rahman, Shakeer. Stop LAPD Spying Coalition. "Reform and Revolt."

Williamson, Clifton "Skye." Transforming Lives, REFORM Alliance. "The ESR Strategy for Preserving Freedom in the Digital Age: The Transformative Resonance of Collective Wisdom in Action."

Nunes, Pablo. Centro de Estudos de Segurança e Cidadania (CESeC). "Surveillance and Police Brutality in Brazil: Notes on the Use of Facial Recognition."

Johnson, Gabbrielle. Claremont McKenna College, Department of Philosophy. "Precarious Predictions in Automated Immigration Decision-Making."

Stevenson, Megan. University of Virginia Law School. "Origins of Predictive Algorithms in Criminal Justice."

Presentations

¹⁰⁰ See a more [in-depth report on facial recognition](#) from the Electronic Frontier Foundation.

Belitz, Clara. "Bridging Informational Divides: A Community-Centered Analysis of 'Public Safety' Surveillance Technology." Department of Information Science, University of Illinois Urbana-Champaign.

Bingham, Laura, Mizue Aizeki and Santiago Narváez. "The Everywhere Border." Temple University School of Law; Surveillance Resistance Lab; Red en Defensa de los Derechos Digitales (R3D).

Collins, Erin and Ngozi Okidegbe. "The Algorithmic Accuracy Obsession." Richmond School of Law; Boston University School of Law and Department of Computing & Data Sciences.

Corry, Frances. "Combine data, control crime? Early imaginaries of predictive policing under the Johnson Administration, 1963-1969." Department of Information Culture & Data Stewardship, University of Pittsburgh.

Dharmaraj, Mallika. "Surveillance in Transit: A Chronology of Power, Complicity, and Policing in the Indian Ocean World, c. 1850-Present." Leverhulme Centre for the Future of Intelligence, Cambridge University.

Diogo, Pedro. "Bahia as an experiment at digitizing racial state terror in Brazil: Investigating the law enforcement application of facial recognition in Brazilian state of Bahia through racial lens." Laboratory of Public Policy and Internet (LAPIN).

Eaglin, Jessica. "Fixing Race Through Informational Criminal Legal Practices." Cornell Law School.

Jegade, Tobi, Marissa Gerchick and Amreeta Mathai. "Lifting the Veil on Algorithm Design and Deployment in the Criminal Legal System." ACLU; NYCLU.

Khosrowi, Donal. "High-risk, because they say so – How risk assessment tools may produce rather than predict criminal behavior." Institute of Philosophy, Leibniz University Hannover.

Lima, Thallita G. L. "Between 'efficiency' and failures: a critical analysis of the use of facial recognition in Brazil." Department of International Relations, Pontifícia Universidade Católica do Rio de Janeiro; Centro de Estudo de Segurança e Cidadania (CESeC).

Lyamuya, Alphoncina. "Humanitarian Experimentation with Predictive Technologies." Annenberg School for Communication and Journalism, University of Southern California.

Molnar, Petra. "Refugee Camps as Prisons: Carceral Technologies of Immigration Detention." Refugee Law Lab, York University; Migration and Technology Monitor Project.

Monreal, Angie Belen. "'They Caught Me and Took My DNA': The Advancement of Technology at the U.S. Southern Border and its Implications." Department of Sociology, UC Irvine.

Pinch, Annika. "Navigating the Ecosystem of Big Data and Algorithms after Incarceration." School of Communication, Northwestern University.

Pullen-Blasnik, Hannah. "Carceral Algorithms and Contested Expertise: Comparisons in Probabilistic DNA Profiling and Facial Recognition Technology." Department of Sociology, Columbia University.

Riley, Sarah. "Infrastructural spillover effects: how algorithms beget themselves in the criminal legal system." Department of Communication, Stanford University.

Robson, Cierra. "Risk Roulette: How Lawyers Make Pretrial Risk Assessment Tools Matter in Criminal Court." Department of Sociology, Harvard University.

Ryan, Nathan and Darakhshan Mir. "Flipping the script on the Pennsylvania Department of Corrections: counterdata and counteranalysis." Department of Mathematics, Bucknell University; Department of Computer Science, Bucknell University.

Sexton, Jason. "Assessing Eugenic Logics Behind Carceral AI: Perspectives from Theological and Religious Studies." Department of Sociology and Institute of the Environment and Sustainability, UCLA.

Song, Lian. "Organizing for Epistemic Justice: Emergent Coalitions and Anti-Surveillance Resistance in San Diego." Department of Communication and Science Studies, UC San Diego.

Stapleton, Logan. "Carceral Care Tech: How Predictive Technologies Strengthen Policing in Care Domains." Department of Computer Science, Vassar College; Department of Computer Science, University of Minnesota.

Stevens, Nikki. "Modeling Abolition: Building a New World from the Data of the Old." Department of Urban Studies and Planning, MIT.

Weisburd, Kate. "The Carceral Home." George Washington University School of Law.

3.3 About the authors

On February 9-10, 2024, 36 academic researchers, activists, and community organizers from around the world gathered at the University of Pittsburgh to share their experiences researching and resisting carceral AI systems. Following the event, 17 of us collaboratively drafted this report. You can learn about each of our backgrounds below.

Dasha Pruss is a researcher, activist, and abolitionist. She researches the impacts of algorithmic decision-making systems in the US criminal legal system. Dasha is an Assistant Professor of Philosophy and Computer Science at George Mason University and a Faculty Associate at the Berkman Klein Center for Internet & Society.

Hannah Pullen-Blasnik is a researcher whose work examines the criminal legal system and urban development to understand how political and economic power gets regulated and contested, and how algorithms and surveillance technologies alter these dynamics. She is currently a Sociology PhD Candidate at Columbia University running Columbia Incite's Criminal Legal Algorithms, Technology, and Expertise project and affiliated with the Movements Against Mass Incarceration Lab, the Data and Racial Inequality Project, and the Columbia Justice Lab.

Nikki Stevens is a critical technology researcher, software engineer, and open-source contributor. Stevens researches the ways that systems of oppression, like white supremacy and transphobia, are both foundational to and recreated within data infrastructures. They are currently a postdoc at MIT where they are writing a book titled *Abolitionist Engineering*.

Shakeer Rahman is an attorney with the Stop LAPD Spying Coalition, a community group based in the Skid Row neighborhood of downtown Los Angeles.

Clara Belitz is an academic researcher and community organizer, currently pursuing her PhD in Information Sciences at the University of Illinois Urbana-Champaign. Her work focuses on how social identities are represented and shaped by predictive technologies, particularly in the K-12 classroom.

Logan Stapleton is a professor at Vassar College, a PhD student in the GroupLens lab at University of Minnesota, and a visiting student at Carnegie Mellon University. Logan's research has focused on algorithmic technologies used for suicide prevention, mental healthcare, and family policing (child welfare).

Mallika G. Dharmaraj is an academic researcher, the Growth and Development lead at Logic(s) Magazine, and a Technical Advisor of the Trans Rights Now Collective. As a writer and organizer, her work focuses on the politics of artificial intelligence and digital technologies, particularly in anti-caste and transfeminist contexts, while foregrounding complicity as a primary ethical obligation (writing as a caste-/class-privileged person). She holds an MPhil (With Distinction) in the Ethics of AI, Data, & Algorithms at the

University of Cambridge and an A.B. (High Honors) from Harvard University in Computer Science and South Asian Studies.

Mizue Aizeki is Executive Director and founder of the Surveillance Resistance Lab. She is a co-editor of *Resisting Borders and Technologies of Violence* (2023), and her photographic work appears in *Dying to Live, A Story of U.S. Immigration in an Age of Global Apartheid* (City Lights Books, 2008) and *Policing the Planet: Why the Policing Crisis Led to Black Lives Matter* (Verso, 2016).

Petra Molnar is a lawyer and anthropologist specializing in migration and human rights. She co-runs the [Refugee Law Lab](#) at York University and is a faculty associate at Harvard's Berkman Klein Center for Internet and Society. She has crossed many borders and is the author of [The Walls Have Eyes: Surviving Migration in the Age of Artificial Intelligence](#).

Annika Pinch is an academic researcher whose work examines the intersection of identity, stigma, and digital media, with a focus on how marginalized individuals navigate and engage with social media platforms to advocate for themselves and their communities. Annika is currently a PhD Candidate at Northwestern University in the Media, Technology, & Society program.

Nathan Ryan is a Professor of Mathematics at Bucknell University. For the past few years he's been involved in transdisciplinary research on the algorithms used by the Pennsylvania Department of Corrections; this work has been done in collaboration with computer scientists, statisticians, geographers, individuals who are formerly incarcerated and undergraduate students. This work has been supported by Google, Mozilla and the National Endowment for the Humanities.

Thallita Lima holds a PhD in International Relations from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio), where she also completed her master's degree. She earned her undergraduate degree in International Relations from the Federal Rural University of Rio de Janeiro (UFRRJ). Currently, she is the research coordinator of the Panopticon Project at the Center for Studies on Security and Citizenship (CESeC).

David Gray Widder is a member of Against Carceral Tech Pittsburgh, and a Postdoctoral Fellow in the Digital Life Initiative at Cornell Tech in New York City. He studies how people creating "Artificial Intelligence" systems think about the downstream harms their systems make possible, and the wider cultural, political, and economic logics which shape these thoughts.

Amiya Tiwari is a Junior at Harvard pursuing a degree in Social Studies. Her academic focus is in understanding punitive public opinion, advancing criminal-legal reform, and exploring alternatives to incarceration. Driven by a commitment to reimagine justice, she

aspires to become a lawyer to build criminal-legal pathways that not only protect but respect and uphold the dignity of all people.

Ly Xīnzhèn Zhǎngsūn, J.D. (Lydia X. Z. Brown) is an internationally recognized advocate, organizer, attorney, strategist, scholar, and writer whose work addresses interpersonal and state violence targeting disabled people at intersections of race, class, gender, sexuality, faith, language, and nation. Their other interests include carcerality and institutional violence, asexuality as queerness, algorithmic harm as an accelerating force of systemic injustice, and the ableism-racism nexus of transracial and transnational adoption. Ly Xīnzhèn is Assistant Teaching Professor of Disability Studies at Georgetown University, as well as Law and Public Policy Discipline Coordinator for the Leadership Education in Neurodevelopmental Disabilities Program. They have taught courses on disability, race, and gender to students at all levels from middle grades to graduate students. Ly Xīnzhèn founded the Autistic People of Color Fund, a project of collective care, redistributive justice, and mutual aid, and they have been creating Disability Justice Wisdom Tarot since 2020. Often, their most important work has no title, job description, or funding, and probably never will.

Jason S. Sexton is a social theorist, cultural historian, and religion scholar based at UCLA's Sociology Department and the Institute of the Environment and Sustainability.

Pablo Nunes is a Brazilian political scientist, with a PhD from the Institute of Social and Political Studies (Iesp) and the coordinator of the Center for Studies on Public Security and Citizenship (CESeC). He coordinates Panóptico, which monitors the use of new technologies by the police, and the Network of Public Security Observatories, which operates in nine Brazilian states. Currently, he is a faculty member in the graduate program in Social Urbanism at Insper. He conducts research and activism focused on public security policies, the use of new technologies, combating racism, and reducing police lethality.

4 Acknowledgments

We are grateful to all the participants of the carceral AI workshop for collectively planting the seeds for this report. The carceral AI workshop was organized by Dasha Pruss and Colin Allen and was generously supported by the Center for Philosophy of Science at the University of Pittsburgh, the Center for Ethics and Policy at Carnegie Mellon University, Pitt Cyber, the Embedded EthiCS program at Harvard University, Professor Gayle Rogers, and Professor Ronald Brand.

Our report also builds on the work of activist collectives that have resisted and built understanding of carceral technologies, including the Carceral Tech Resistance Network, Stop LAPD Spying, O Panóptico, Against Carceral Tech, as well as Dr. Ruha Benjamin's work on carceral technoscience.

Illustrations in this report are by Dasha Pruss and were inspired by artwork by Douglas Lopes for O Panóptico; David Gray Widder; and Simon Montag for the Atlantic. The carceral AI website (<https://carceral-ai.com>) was designed by Amiya Tiwari and funded by the Berkman Klein Center for Internet and Society at Harvard University.

We are grateful to the Embedded EthiCS research group and the Berkman Klein Center for Internet and Society at Harvard University for their feedback on early drafts of this report.