NO PREP

AI in Criminal Justice

Predictive Policing Risks

Includes: Student reading, case study, differentiated questions, teacher key

Name: Period: Date:

AI in Criminal Justice: Predictive Policing Risks

What Is Predictive Policing?

Predictive policing refers to the use of data, statistical algorithms, and AI systems to forecast where crimes are likely to occur, or who is likely to commit them, so that law enforcement resources can be directed accordingly. The concept draws on the idea that crime is not random — it tends to cluster in certain places and times, and certain individuals may have histories that correlate with future illegal activity. If AI can identify those patterns faster and more accurately than human analysts, the argument goes, police can prevent crime rather than just respond to it. Two main approaches exist within predictive policing. Place-based tools analyze historical crime data to identify geographic hotspots where crimes are statistically likely to occur. These tools then direct police patrols to those areas more frequently. Person-based tools are more controversial: they analyze data about individuals — prior arrests, social connections, location history — to generate risk scores that predict which people are most likely to offend in the future. Both types of tools

Name: **Period:** **Date:**

AI in Criminal Justice: Predictive Policing Risks

What Is Predictive Policing?

Predictive policing refers to the use of data, statistical algorithms, and AI systems to forecast where crimes are likely to occur, or who is likely to commit them, so that law enforcement resources can be directed accordingly. The concept draws on the idea that crime is not random: it tends to cluster in certain places and times, and certain individuals may have histories that correlate with future illegal activity. If AI can identify those patterns faster and more accurately than human analysts, the argument goes, police can prevent crime rather than just respond to it. Two main approaches exist within predictive policing. Place-based tools analyze historical crime data to identify geographic hotspots where crimes are statistically likely to occur. These tools then direct police patrols to those areas more frequently. Person-based tools are more controversial: they analyze data about individuals (prior arrests, social connections, location history) to generate risk scores that predict which people are most likely to offend in the future. Both types of tools are currently in use across numerous jurisdictions in the United States, and both have generated serious criticism from civil rights advocates, researchers, and affected communities.

The Feedback Loop at the Heart of Hotspot Policing

At first glance, directing police to areas where crime has historically been more frequent seems straightforwardly reasonable. If a particular neighborhood has high rates of reported robbery, increasing patrols there seems logical. But critics identify a fundamental problem: the crime data those algorithms use reflects where police have been, not necessarily where crime actually occurs. When more officers patrol a neighborhood, they make more arrests there, including arrests for low-level offenses like drug possession that might go undetected in less-patrolled areas.

Those arrests generate new data, which the algorithm interprets as evidence that the neighborhood is high-crime, which directs more officers there, which generates more arrests. This feedback loop means that predictive policing algorithms can become self-fulfilling prophecies. Communities that have historically been over-policed (which in the United States are disproportionately Black and Latino communities) generate more arrest data, which focuses more policing on them, which generates more arrests, which confirms the algorithm's prediction. The algorithm is not measuring crime; it is measuring policing. And because those two things are not identical, the result is a systematic amplification of existing inequities rather than an objective assessment of where crime actually occurs.

COMPAS and the Risk of Predicting People

Perhaps the most controversial predictive tool in criminal justice is not used to direct patrols but to influence sentencing and parole decisions. COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) is a risk assessment algorithm that judges and parole boards use to evaluate the likelihood that a defendant will reoffend. It generates a "risk score" based on answers to a questionnaire that covers factors like employment history, social ties, family criminal history, and neighborhood. The score can influence whether someone receives probation or prison, and how long they remain incarcerated. In 2016, investigative journalists at ProPublica published a landmark analysis of COMPAS scores in Broward County, Florida. They found that the algorithm falsely flagged Black defendants as future criminals at nearly twice the rate it falsely flagged white defendants. Conversely, white defendants were more likely to be incorrectly labeled low-risk, and then go on to reoffend. The company behind COMPAS disputed the methodology, and subsequent academic debate established that multiple definitions of fairness exist and that it is mathematically impossible to satisfy all of them simultaneously with a single algorithm. But the core finding, that a tool influencing real sentences disproportionately mislabeled Black defendants as high-risk, was not effectively refuted.

The Due Process Problem

Predictive policing and risk assessment tools raise profound legal and constitutional questions. In the United States, defendants have the right to confront the evidence against them — a principle rooted in the Sixth Amendment. But when a risk score

influences a sentence and the company that makes the algorithm claims the formula is a proprietary trade secret, defendants cannot examine how the score was generated, what factors drove it, or whether it was calculated correctly. Several courts have faced challenges from defendants who argued that using secret algorithms in sentencing violates their constitutional rights. Courts have generally ruled that risk scores can be used as long as they are not the sole basis for a decision — but the practical reality is that a numerical score carries significant weight with decision-makers. In 2020, a California court ruled in People v. Loomis that using COMPAS scores in sentencing did not violate due process because the score was one factor among many — but critics argued the ruling sidestepped the core problem. If the public, defendants, and even the judges using a score cannot understand how it works, meaningful oversight is impossible. The opacity of the algorithm is itself a form of power: it gives an automated system authority over human lives while insulating its reasoning from scrutiny.

Responses and Reforms

Several cities and states have taken action to limit or end predictive policing programs. Santa Cruz, California became the first U.S. city to ban predictive policing in 2020. Los Angeles ended its Operation LASER program, which targeted individuals with high predictive scores for increased police attention, in 2019 after community protests and a finding that the program had no measurable impact on crime rates. New York City passed a law in 2018 requiring the city to audit algorithms used in government decisions and make the results public — though implementation has been slow and incomplete. The European Union's AI Act classifies AI systems used in law enforcement, criminal justice, and risk assessment as high-risk and requires documentation, testing, human oversight, and transparency before deployment. Advocates in the United States have pushed for similar federal legislation. The debate continues: proponents argue that data-driven decision-making, if done responsibly, is more consistent and less susceptible to individual human bias than the status quo. Critics respond that the data itself is contaminated by historical discrimination and that "more consistent" discrimination is not an improvement. A key question for democratic societies is whether the use of predictive algorithms in law enforcement and sentencing is consistent with foundational principles of justice. The American legal tradition holds that people are presumed innocent until proven

guilty through evidence of their individual actions — not through statistical association with groups or predictions about their future behavior. Risk assessment tools, by design, use group-level patterns to make individual predictions. When those predictions carry legal weight, they challenge this foundational presumption. Resolving that tension — between efficiency and equity, between data-driven consistency and individual justice — is not a technical task at all. It is a political and moral one.

Key Information: Predictive Policing

Detail	Information
Place-based predictive policing	Directs patrols to geographic areas where historical data shows higher crime rates
Person-based predictive policing	Generates individual risk scores based on personal history, social ties, location data
Core feedback loop problem	More patrols generate more arrests, which generate data that directs more patrols to the same areas
COMPAS	Risk assessment algorithm used in sentencing and parole decisions
ProPublica 2016 finding	COMPAS falsely flagged Black defendants as future criminals at nearly twice the rate of white defendants
Due process concern	Defendants cannot examine proprietary algorithm formulas that influence their sentences
Santa Cruz, CA (2020)	First U.S. city to ban predictive policing entirely
EU AI Act classification	Law enforcement and risk assessment AI classified as high-risk, requiring transparency and oversight

Case Study: Jerome's Risk Score

Jerome is 19 years old and was charged with possession of marijuana — a misdemeanor. It is his first offense. Before sentencing, the judge reviews a risk assessment report that includes a COMPAS score. Jerome scored a 7 out of 10 on the "recidivism risk" scale. The report does not explain exactly why. Jerome's public defender requested the algorithm's methodology from the company. The company refused, calling the formula proprietary. Jerome's defender reads the questions on the questionnaire Jerome answered. Several of them ask about family members who have been arrested or incarcerated. Jerome has a brother who served time. Several ask about his neighborhood and whether he has friends who have been arrested. He does. "These questions are about who my family is and where I grew up," Jerome tells his defender. "I can't change any of that. How is this about what I did?" The judge ultimately sentences Jerome to probation rather than prison, saying the score was one factor among several. But Jerome carries the label of "high risk" on his record.

1. Using the reading, identify which specific type of data in COMPAS's questionnaire likely raised Jerome's score, and explain why critics argue this type of data is problematic.

2. Jerome asks, "How is this about what I did?" Based on the reading, explain the constitutional concern his question points to.

Multiple Choice

1. What fundamental problem does the reading identify with using historical arrest data in hotspot policing algorithms?

a) Historical arrest data is usually far too old to be statistically useful for future predictions

b) Most historical arrest databases contain only felony records, which miss minor crimes

c) Historical data cannot be processed quickly enough for real-time crime prediction

d) The data reflects where police have been deployed, not necessarily where crime actually occurs

2. What did ProPublica's 2016 analysis of COMPAS find?

a) The algorithm falsely flagged Black defendants as future criminals at nearly twice the rate it falsely flagged white defendants

b) COMPAS was found to have no statistical relationship with actual reoffending rates

c) White defendants typically received higher risk scores than similarly situated Black defendants in Broward County for that period

d) COMPAS scores were equally accurate for all racial groups studied

3. Why do critics argue that defendants have a due process problem when risk assessment algorithms are used in sentencing?

a) Risk scores are only available in Spanish and English, excluding other language speakers

b) Defendants cannot examine how the proprietary algorithm formula generated their score

c) Risk scores are shared with juries rather than judges in most jurisdictions

d) Defendants are required to pay fees to access their own risk scores

Matching

Write the letter of the correct definition on the line next to each term.

Term	Definition
1. ___ COMPAS	A) A self-reinforcing cycle in which outputs of a system become inputs that amplify the original pattern
2. ___ Feedback loop	B) Privately owned and protected from public disclosure, as when the company that makes an algorithm claims its formula is a trade secret
3. ___ Recidivism	C) Legal principle requiring fair procedures, including the right to examine evidence used against you
4. ___ Due process	D) The tendency of a formerly incarcerated person to reoffend after release
5. ___ Proprietary	E) Risk assessment algorithm used in U.S. sentencing and parole decisions

Fill-in-the-Blank

1. ProPublica's landmark analysis of COMPAS was published in _____ .

2. Santa Cruz, California became the first U.S. city to ban predictive policing in _____ .

3. COMPAS stands for Correctional Offender Management Profiling for Alternative _____ .

Short Answer

Answer each question in 2-4 complete sentences.

1. Explain in your own words why predictive policing based on historical arrest data can become a self-fulfilling prophecy.

2. The reading says it is "mathematically impossible" to satisfy all definitions of fairness simultaneously with a single algorithm. Why does this matter for debates about COMPAS and similar tools?

3. What is the strongest argument in favor of predictive policing, and how do critics respond to it?