INTELLIGENCE · ARTIFICIAL AND COLLECTIVE INTELLIGENCE

Predictive Policing and the Ethics of Intelligence-Driven Knowledge

Predictive policing uses data and algorithms to forecast where crime will happen or who may be involved. This article explains how it works, what the evidence shows about its effectiveness, the ethical problems it raises, and the lessons it holds for any organisation that uses predictive intelligence about people.

ModQu research article · 2,814 words · 26 references (APA 7th edition)

Police forces have always tried to anticipate crime. Officers learn which streets see trouble on Friday nights and which shops are targeted repeatedly. Since the early 2010s, many forces have formalised this knowledge with software that analyses past data to forecast where and when crimes are likely to occur, or which individuals are likely to be involved. This practice, known as **predictive policing**, promised to make policing more efficient and objective. It has also become one of the most contested uses of data-driven intelligence, raising questions about accuracy, fairness, privacy and public trust.

The debate matters beyond policing. Any organisation that uses predictive intelligence to make decisions about people, whether in lending, insurance, hiring, fraud detection or customer management, faces versions of the same problems. Predictive policing is a case study in what happens when intelligence about people becomes the basis for action against them.

What predictive policing is

A RAND Corporation study for the American National Institute of Justice defined predictive policing as the application of analytical techniques, particularly quantitative techniques, to identify likely targets for police intervention and prevent crime or solve past crime through statistical predictions (Perry et al., 2013). The authors identified four broad types of prediction: methods for predicting where and when crimes will occur, for predicting individuals at risk of offending, for predicting the identities of offenders in past crimes, and for predicting likely victims.

The most common tools are **place-based**. They build on decades of research showing that crime is highly concentrated. Sherman, Gartin and Buerger (1989) found that a small share of addresses in Minneapolis accounted for a large share of calls to police, a pattern repeated in many cities. Routine activity theory explains why: crime occurs where motivated offenders meet suitable targets in the absence of capable guardians (Cohen & Felson, 1979). Concentrating patrols at these **hot spots** is one of the best-supported police strategies. A systematic review by Braga, Papachristos and Hureau (2014) found that hot spots policing produced modest but meaningful reductions in crime, without simply displacing it to nearby areas.

Predictive software extends this approach by forecasting which areas are likely to become hot spots in the coming hours or days. **Person-based** tools go further, scoring individuals according to their predicted likelihood of being involved in violence, as offenders or victims, based on factors such as arrest history and social connections.

Predictive policing grew out of a broader movement towards **intelligence-led policing**. Ratcliffe (2016) describes this as a management approach in which data analysis and crime intelligence guide decisions about where to focus resources, with an emphasis on prolific offenders and recurring problems. Predictive tools promised to automate and sharpen that analysis.

What the evidence shows

The evidence on effectiveness is mixed. Mohler et al. (2015) conducted randomised field trials in Los Angeles and in Kent, England, comparing an algorithm based on models of earthquake aftershocks with the predictions of experienced crime analysts. The algorithm predicted more crime than the analysts, and patrols directed by it were associated with modest reductions in crime in the areas tested.

Other evaluations have been less encouraging. Hunt et al. (2014) evaluated a predictive policing experiment in Shreveport, Louisiana, and found no statistically significant reduction in property crime in the districts using predictions compared with control districts. They noted that the way officers used the predictions varied considerably, and that the intervention was as important as the forecast.

Person-based prediction has raised greater concerns. Saunders et al. (2016) evaluated Chicago's Strategic Subject List, which identified individuals considered at highest risk of involvement in gun violence. People on the list were no more or less likely to become victims of homicide or shooting than a comparison group. They were, however, more likely to be arrested for a shooting, which the authors suggested reflected the list being used as a source of suspects rather than as a means of prevention.

Weisburd and Eck (2004), reviewing what police can do to reduce crime, concluded that approaches focused on specific places and problems were more effective than generic increases in patrol. Predictive tools can support such focus. But the evidence suggests that the value lies less in the algorithm than in what police do with its output.

The feedback loop problem

The central ethical and analytic concern is that predictive policing models learn from data produced by policing itself. Recorded crime is not a neutral measure of crime. It reflects which offences are reported by victims, and where police choose to patrol, stop and search. Offences such as drug possession are recorded mainly where police look for them.

Lum and Isaac (2016) illustrated the consequence. Using data from Oakland, California, they simulated the application of a widely used predictive algorithm to drug crimes. Public health survey data suggested that drug use was spread fairly evenly across the city. But because past arrests were concentrated in predominantly Black and low-income neighbourhoods, the algorithm directed patrols overwhelmingly to those same areas. Police would then make more arrests there, generating more data that confirmed the prediction.

Ensign et al. (2018) formalised this as a **runaway feedback loop**. When patrols are sent where predicted crime is highest, and crime is recorded only where patrols go, the model's predictions become self-reinforcing. Even a small initial difference between two areas can grow until almost all attention goes to one. Figure 1 shows the loop.

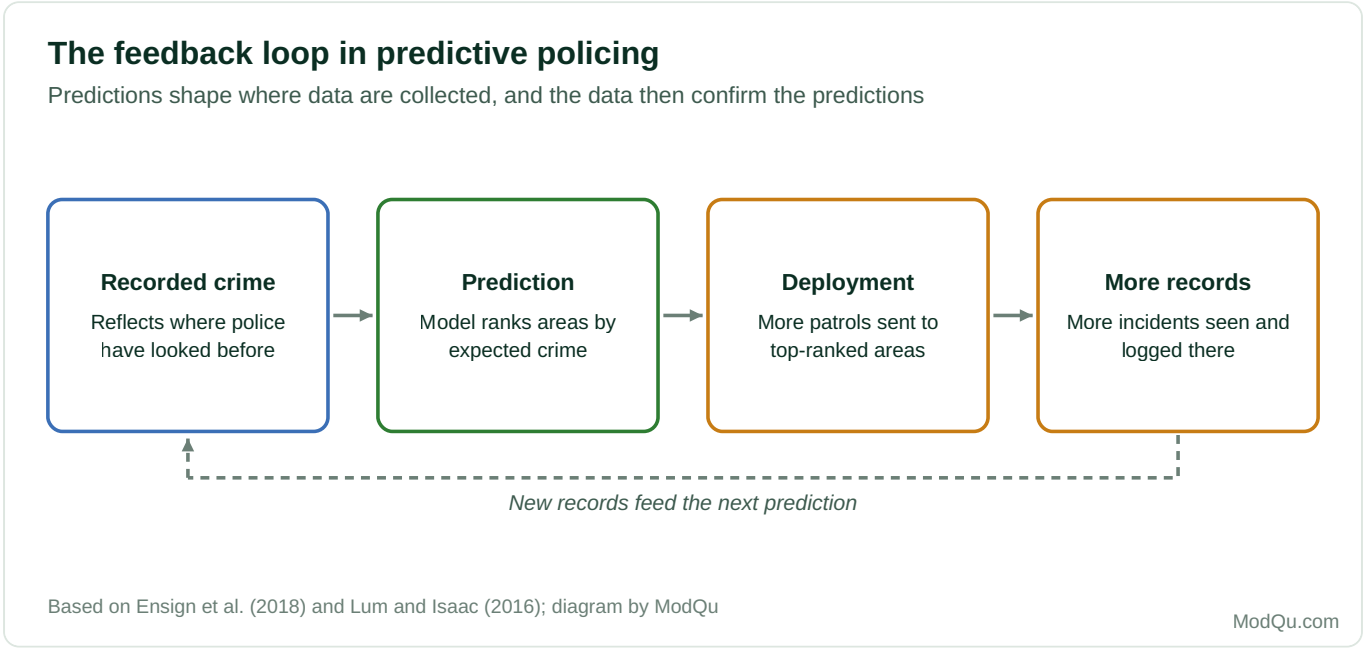


Figure 1. The feedback loop in predictive policing. When recorded crime reflects where police look, predictions based on it send police back to the same places, and the new records appear to confirm the prediction. Based on Ensign et al. (2018) and Lum and Isaac (2016); diagram by ModQu.

O'Neil (2016) highlighted a related risk. When predictive systems include minor offences, which are common everywhere but recorded mainly where police are present, they draw patrols to poorer neighbourhoods and generate records that justify further attention, while serious crimes elsewhere receive less.

Dirty data

Richardson, Schultz and Crawford (2019) examined American police departments that had used or developed predictive systems while under investigation or court supervision for unlawful practices, such as discriminatory stops or falsified records. They found that data produced by these practices could flow into predictive systems, which would then reproduce the unlawful patterns in their forecasts. They called this **dirty data**: data shaped by flawed, biased or illegal practices that cannot be corrected simply by adjusting an algorithm.

Barocas and Selbst (2016) explain the general mechanism. Data mining can produce discriminatory outcomes without any discriminatory intent, through biased training data, the choice of what to predict, and the use of variables that act as proxies for protected characteristics. Removing race from a model does not remove racial disparities if other variables, such as neighbourhood or prior arrests, carry the same information.

Fairness and its trade-offs

Debates over fairness in predictive policing have been shaped by research on risk scores in criminal justice more broadly. Angwin et al. (2016) reported that a widely used risk assessment tool incorrectly classified Black defendants as high risk at a higher rate than white defendants. The tool's developers responded that scores were equally predictive for both groups. Chouldechova (2017) showed that these two kinds of fairness cannot both be satisfied when the underlying rates of reoffending, as recorded, differ between groups. Choosing between them is a matter of values, not mathematics.

Selbst et al. (2019) argue that a focus on technical definitions of fairness can miss the larger picture. A model is part of a social system that includes how data are produced, how predictions are used, who is affected and who can challenge decisions. Fairness has to be assessed across that whole system, not only in the model's statistics.

Surveillance and discretion

Brayne (2017) conducted fieldwork inside the Los Angeles Police Department and found that big data tools changed not only how police predicted crime but how they conducted surveillance. Databases linked records from many sources, lowering the threshold for inclusion. People who had never been arrested could appear in police systems through contact with others, such as being stopped while in a car with someone of interest. Surveillance spread from targeted individuals to their networks. In a later book, Brayne (2021) argued that data tools did not remove officer discretion but shifted it, from the street to the design and use of systems, often making it less visible.

Ferguson (2017) described the rise of "big data policing" and warned that the combination of predictive tools, extensive databases and surveillance technologies threatens to change the relationship between citizens and police. Ferguson argued that data can also be used to hold police accountable, for example by identifying patterns of stops or uses of force, and that the same technology can serve either purpose depending on governance.

Nissenbaum (2004) offers a useful framework for the privacy issues. Under the theory of **contextual integrity**, information flows are appropriate when they respect the norms of the context in which the information was shared. Information given to a social worker, a health service or a school may be appropriate in that setting but violate expectations when combined into a police risk score.

Legitimacy

Ethical concerns are also practical concerns. Tyler (1990) showed that people obey the law and cooperate with authorities largely because they consider them legitimate, and that legitimacy depends heavily on whether people feel they are treated fairly and with respect. Policing perceived as targeting communities unfairly can reduce cooperation, reporting of crimes and willingness to serve as witnesses. A predictive tool that improves short-term efficiency but damages legitimacy can make crime harder to prevent in the long run.

Accountability and governance

How should predictive policing be governed? Bennett Moses and Chan (2018) examined the assumptions behind predictive policing and found that claims of objectivity and effectiveness were often accepted without adequate evaluation. They argued for transparency about how tools work, independent evaluation of their effects, and clear accountability for decisions informed by them.

In the United Kingdom, Durham Constabulary developed an algorithm called HART to assess the risk that people in custody would reoffend, to help decide who was eligible for a rehabilitation programme. Oswald et al. (2018) analysed the model and proposed a framework for its responsible use. They emphasised that algorithmic predictions should support rather than replace officer judgement, that models should be tested experimentally before wider adoption, and that the proportionality of any intrusion should be assessed. Babuta and Oswald (2019), in a briefing for the Royal United Services Institute, found a lack of clear national guidance on police use of analytics and called for a code of practice covering bias, transparency and accountability.

Kitchin (2017) argues that algorithms should be studied as social and political objects, not only technical ones. Their effects depend on the data they use, the institutions that deploy them and the decisions they inform. Understanding them requires examining the whole context in which they operate. Figure 2 brings these concerns together as a set of tests.

Ethical tests for intelligence-driven policing

Questions to answer before, during and after a predictive tool is used

Test	Question to answer
Accuracy	Does it predict better than existing methods, on independent evaluation?
Fairness	Are errors and burdens spread fairly across communities?
Data quality	Do the records reflect crime, or past patterns of enforcement?
Transparency	Can officers, courts and the public understand how it works?
Accountability	Who answers for decisions, and how can they be challenged?
Proportionality	Is the intrusion justified by the benefit, and limited to need?
Legitimacy	Does it support or damage public trust in policing?

Compiled from Oswald et al. (2018), Babuta and Oswald (2019) and Bennett Moses and Chan (2018); table by ModQu

ModQu.com

Figure 2. Ethical tests for intelligence-driven policing. Each test should be answered before a tool is adopted and reviewed regularly while it is in use. Compiled from Oswald et al. (2018), Babuta and Oswald (2019) and Bennett Moses and Chan (2018); table by ModQu.

Prediction is not the only option

Much of the debate assumes that the choice is between prediction and no data use at all. In practice, there is a wider range of options. Data can be used to understand problems rather than to target people: analysing why certain places attract repeated crime, such as poor lighting, unmanaged licensed premises or abandoned buildings, can lead to changes that remove the opportunity for crime without increasing enforcement. Weisburd and Eck (2004) found that problem-oriented approaches of this kind were among the most effective police strategies. Predictions can also direct services rather than enforcement, for example by sending outreach workers or victim support to areas or people at high risk.

The same predictive output can therefore lead to very different interventions, with very different ethical implications. A forecast that a street is likely to see burglaries could lead to more stops of young people in the area, or to advice for residents on securing their homes. Perry et al. (2013) emphasised that prediction is only one step in a larger process, and that the response chosen matters as much as the forecast. Ethical assessment should focus on the whole chain from data to action.

Questions to ask before adopting a predictive tool

Any organisation, public or private, considering a predictive tool that affects people can work through a short set of questions:

- **What exactly will be predicted, and why that?** Predicting arrests is not the same as predicting crime; predicting past investigations is not the same as predicting fraud.
- **How were the training data produced?** Which past decisions shaped them, and could those decisions have been biased?
- **What will be done with the predictions?** Will they lead to help, scrutiny or sanctions, and for whom?
- **How will effectiveness be tested?** Is there a plan for a controlled comparison, rather than relying on anecdotes or vendor claims?
- **How will unequal effects be detected?** Which groups will be compared, on which measures, and how often?
- **Who can see how the system works, and who can challenge it?** Can affected people find out that a prediction was used and contest it?
- **When will it be stopped?** What evidence would lead the organisation to withdraw the tool?

Common mistakes

- **Treating recorded data as ground truth.** Records describe what an organisation observed, not everything that happened.
- **Accepting vendor claims of accuracy without independent evaluation.** Accuracy must be tested on local data and in local conditions.
- **Assuming that removing sensitive variables ensures fairness.** Proxies carry the same information.
- **Judging success by the model's own outputs.** A rise in detections in targeted areas may only show that more attention produces more records.
- **Ignoring the people affected.** Those subject to predictions often know where systems go wrong and can identify problems early.

Lessons for business intelligence

Most businesses do not police anyone, but many use predictive intelligence to make decisions about people: which customers to offer credit, which claims to investigate for fraud, which job applicants to interview, which employees to flag as at risk of leaving. The lessons of predictive policing apply directly.

1. **Know where your data come from.** If past decisions shaped the data, such as which claims were investigated or which customers were approved, the data reflect those decisions, not only the underlying reality.
2. **Watch for feedback loops.** If a model decides where to look, and new data come only from where you looked, the model will confirm itself. Reserve some effort for random or independent checks.
3. **Test for unequal effects.** Measure error rates and outcomes across groups, not only overall accuracy.
4. **Evaluate the intervention, not just the prediction.** A good forecast is useless or harmful if the action taken in response does not help.

5. **Keep people accountable.** Decisions informed by models should have a named owner and a route for challenge.
6. **Consider legitimacy.** Customers, employees and regulators who believe a system is unfair will respond accordingly, whatever its statistical properties.

A business illustration

An insurance company introduces a model to flag claims for fraud investigation. It is trained on past investigations, which were concentrated on claims from certain postcodes because investigators had historically focused there. The model learns that these postcodes predict fraud and flags their claims at a high rate. Investigators find fraud among the flagged claims, as they would in any large sample, and these findings are added to the training data. Within two years, honest customers in those postcodes face far more delays and scrutiny than customers elsewhere, and complaints to the regulator rise.

An internal review identifies the loop. The company changes its approach: it investigates a random sample of claims from all areas each month to obtain unbiased data on where fraud actually occurs, retrains the model on that sample, monitors flag rates and outcomes by area, and requires investigators to record the reason for each investigation beyond the model's score. Fraud detection improves, and the disparity between areas falls.

Conclusion

Predictive policing shows both the promise and the risks of intelligence-driven decision-making. Forecasts based on data can help direct limited resources where they are most needed, and focused strategies such as hot spots policing have genuine support. But predictions are only as good as the data behind them, and data produced by past enforcement can embed and amplify past patterns. Feedback loops, dirty data, conflicting definitions of fairness, expanding surveillance and threats to legitimacy all follow when predictive tools are adopted without careful governance. The ethical use of predictive intelligence, in policing or in business, requires understanding where data come from, testing effects on different groups, evaluating outcomes independently, keeping people accountable and remembering that public trust is itself a resource that intelligence should protect.

References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine bias*. ProPublica.

Babuta, A., & Oswald, M. (2019). *Data analytics and algorithmic bias in policing* (Briefing Paper). Royal United Services Institute.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*(3), 671–732.

Bennett Moses, L., & Chan, J. (2018). Algorithmic prediction in policing: Assumptions, evaluation, and accountability. *Policing and Society, 28*(7), 806–822.

Braga, A. A., Papachristos, A. V., & Hureau, D. M. (2014). The effects of hot spots policing on crime: An updated systematic review and meta-analysis. *Justice Quarterly, 31*(4), 633–663. https://doi.org/10.1080/07418825.2012.673632

Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review, 82*(5), 977–1008. https://doi.org/10.1177/0003122417725865

Brayne, S. (2021). *Predict and surveil: Data, discretion, and the future of policing*. Oxford University Press.

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data, 5*(2), 153–163. https://doi.org/10.1089/big.2016.0047

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review, 44*(4), 588–608. https://doi.org/10.2307/2094589

Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., & Venkatasubramanian, S. (2018). Runaway feedback loops in predictive policing. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (pp. 160–171).

Ferguson, A. G. (2017). *The rise of big data policing: Surveillance, race, and the future of law enforcement*. New York University Press.

Hunt, P., Saunders, J., & Hollywood, J. S. (2014). *Evaluation of the Shreveport predictive policing experiment*. RAND Corporation.

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society, 20*(1), 14–29.

Lum, K., & Isaac, W. (2016). To predict and serve? *Significance, 13*(5), 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x

Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2015). Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association, 110*(512), 1399–1411. https://doi.org/10.1080/01621459.2015.1077710

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review, 79*(1), 119–157.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Oswald, M., Grace, J., Urwin, S., & Barnes, G. C. (2018). Algorithmic risk assessment policing models: Lessons from the Durham HART model and 'experimental' proportionality. *Information & Communications Technology Law, 27*(2), 223–250.

Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., & Hollywood, J. S. (2013). *Predictive policing: The role of crime forecasting in law enforcement operations*. RAND Corporation.

Ratcliffe, J. H. (2016). *Intelligence-led policing* (2nd ed.). Routledge.

Richardson, R., Schultz, J. M., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review Online, 94*, 15–55.

Saunders, J., Hunt, P., & Hollywood, J. S. (2016). Predictions put into practice: A quasi-experimental evaluation of Chicago's predictive policing pilot. *Journal of Experimental Criminology, 12*(3), 347–371.

Selbst, A. D., boyd, d., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 59–68). ACM. https://doi.org/10.1145/3287560.3287598

Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology, 27*(1), 27–56. https://doi.org/10.1111/j.1745-9125.1989.tb00862.x

Tyler, T. R. (1990). *Why people obey the law*. Yale University Press.

Weisburd, D., & Eck, J. E. (2004). What can police do to reduce crime, disorder, and fear? *The Annals of the American Academy of Political and Social Science, 593*, 42–65. https://doi.org/10.1177/0002716203262548