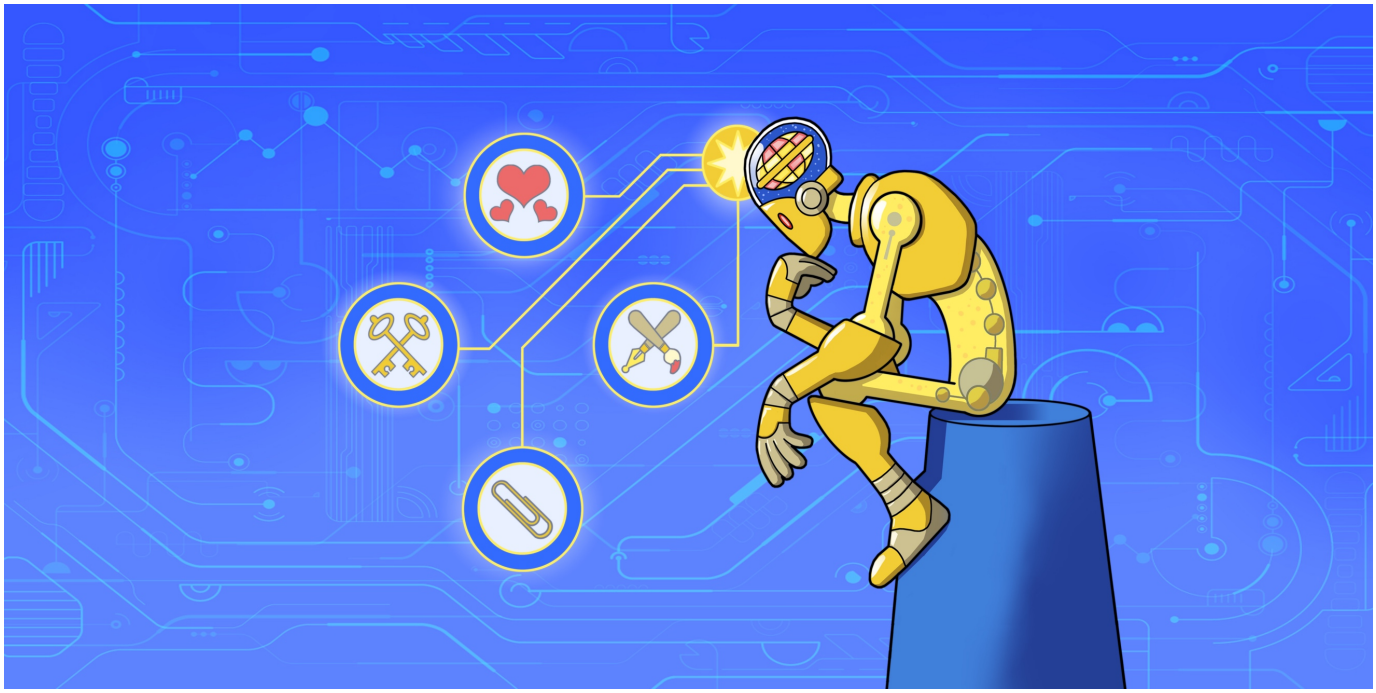
Artificial Intelligence and Algorithmic Tools

A Policy Guide for Judges and Judicial Officers



AI and algorithmic tools are currently being procured and deployed across the country–in criminal cases and beyond–without sufficient vetting, testing, or understanding by those relying on these tools to make decisions. But the choice to procure and implement systems based on AI and algorithmic decision-making is a policy decision that requires careful consideration. To ensure that the introduction of systems based on AI or algorithmic decision-making tools does not reinforce existing biases and inequity, it is critical that court administrators and judges carefully consider the policy implications of decisions to procure and deploy these systems–and that they do so *before* both the procurement and implementation processes actually begin.

POLICY ISSUES TO THINK ABOUT BEFORE PROCUREMENT

1. Analyze the Problem. The first step is to step back and ask: what problems are we aiming to address? What do we know about these problems? The best practice is to gather a baseline dataset on judicial decisions that can be made public to assess how judges are doing–to use the data you already have to better understand the problem you are trying to solve. Otherwise, you will not truly understand the intricacies of the problem, and you will have no threshold against which to determine if a particular tool is actually helping, hurting, or was just a waste of money. This will require some thought as to what metrics you wish to use to measure the success or failure of a particular tool. For example, how will you check for fairness and bias? What comparisons can you make to see if the tool is effective?

STEPHANIE LACAMBRA, CRIMINAL DEFENSE STAFF ATTORNEY 815 EDDY STREET, SAN FRANCISCO, CA 94109 USA
phone +1.415.436.9333 x130 fax +1.415.436.9993 email stephanie@eff.org website eff.org

2. Conduct Due Diligence. Next, you must responsibly investigate not only whether the tool you are considering for procurement could reliably address the problem you are trying to solve, but that it will not result in serious unintended consequences or reinforce existing biases and inequity within the system. We review some of the important questions to ask in our blog: eff.org/questionsai

    a. Consult Independent Data Scientists. You should not rely on the marketing materials from vendors, but instead seek the opinion of independent data scientists like the Human Rights Data Analysis Group (hrdag.org) to conduct rigorous testing of the AI tool you are considering, including the source code. Sometimes just publishing (or analyzing) the tool isn't sufficient. If it's AI based on machine learning, it may be necessary for the training and testing datasets to be published and reviewed as well. Datasets must be analyzed to assess the variables and proxies upon which they rely, and to identify and measure any statistical biases–including omitted variable biases.

    b. Analyze the Design Process. You must consider not only the model's operation (e.g., why the model made the decision it did), but also the design process (e.g., why the model was designed that way). This means looking at the developmental and methodological materials documenting the decision-making processes in a system's creation, design, and use – decision-making processes that are in effect policy decisions that will directly impact the model's outcomes. Such documentation could show, for instance, that a design team tested a model with and without certain data and found that using the data reduced the disproportionate impact of the model; or that a team considered adding additional features to create a more accurate and fair model but, after discovering that such features were exceedingly difficult or costly to implement, the company decided to use a less costly proxies that reduced the model's accuracy and fairness.

3. Provide an Opportunity for Community Notice and Feedback. There should be a period and process for:

    a. Notice to the potential target community, and

    b. An opportunity for community members and stakeholders to examine the potential ramifications of the tool's deployment in their community, and to weigh in on which algorithmic fairness metric will be used.

4. Be Transparent. Any system considered for procurement should have its source code, testing data, and developmental and methodological materials documenting decision-making processes in a system's creation, design, and use published and readily available for public review and testing. Proper transparency will ensure that a person impacted by the system's decision-making will have sufficient access to determine and meaningfully challenge how an adverse decision was made.

STEPHANIE LACAMBRA, CRIMINAL DEFENSE STAFF ATTORNEY 815 EDDY STREET, SAN FRANCISCO, CA 94109 USA
phone +1.415.436.9333 x130 fax +1.415.436.9993 email stephanie@eff.org website eff.org

ELECTRONIC FRONTIER FOUNDATION

POLICY ISSUES TO THINK ABOUT BEFORE IMPLEMENTATION:

1. Properly Train Users (i.e., judges). Judges need to understand what inputs are being used and their weights, as well as how to properly use the tool and check for errors. This requires some basic training in both statistics and on the potential limits and shortcomings of the specific AI or algorithmic tools they will be using.

2. Control for Automation Bias. The tendency to view machines as objective and inherently trustworthy – even when they are not – is referred to as "automation bias." To safeguard against such bias, care must be taken to see that outputs are thoroughly explained in a narrative report rather than a numerical score.

3. Build in an Opportunity for Redress. Target individuals should have the right and opportunity to meaningfully challenge the use or outcome of the algorithmic/AI system in an adversarial setting, with due process protections. This process must be set up before the system is implemented to protect rights, ensure that issues are promptly identified and corrected, and avoid unintended negative consequences.

4. Identify a Fairness Metric and Calibrate Accordingly. The fairness of every algorithmic or AI system must be analyzed prior to implementation and throughout the system's use. For example, does the algorithm treat like groups similarly, or disparately? Is the system optimizing for fairness, for public safety, for equal treatment, or for the most efficient allocation of resources? The chosen fairness metric and the results of any fairness analysis should be made public, so stakeholders may review and inform which fairness metric is appropriate.

5. Conduct Pilot Tests. The algorithmic/AI system should be rolled out in a small pilot program to test its efficacy in achieving the stated goal before being deployed across the entire jurisdiction. This allows for comparison with a baseline control group that should be made public to allow stakeholders to comment on the tool's efficacy.

POLICY ISSUES TO THINK ABOUT DURING/AFTER IMPLEMENTATION:

1. Track How the Tool is Used. Ensure that client information is anonymized, but make sure to collect data and analyze whether the system is having a disproportionate impact on protected classes of people (by race, ethnicity, gender, etc.).

2. Analyze Continuously. There should be a continuous review process to evaluate the outcomes of any systems implemented, compare the results for progress toward the stated goal, and correct any disparate impacts.

3. Make Data Publicly Accessible. Ensure the anonymized data collected belongs to the courts and will be made available to the public as a check on whether the tool is functioning properly and as intended and does not have a disproportionate impact on protected classes of individuals.

STEPHANIE LACAMBRA, CRIMINAL DEFENSE STAFF ATTORNEY 815 EDDY STREET, SAN FRANCISCO, CA 94109 USA
phone +1.415.436.9333 x130 fax +1.415.436.9993 email stephanie@eff.org website eff.org

GLOSSARY

algorithm – an algorithm is a series of instructions for arriving at a conclusion. Algorithms are designed by humans, sometimes using machine learning systems.

algorithmic tools - an algorithm to make a decision or to provide information to a human decision-maker. An algorithmic tool can be as simple as "Fill out a questionnaire. Count the number of 'Yes' answers. If greater than some threshold, do X. Otherwise, do Y." In other words, for a tool to be algorithmic, you don't necessarily need a computer. The Arnold Foundation's Public Safety Assessment is an example of an algorithmic tool.

Artificial Intelligence ("AI") - a field of computer science that focuses on getting computers to do tasks and/or behave in ways that traditionally have only been done by humans. Often, we call a thing "AI" until it has become commonplace to have the function carried out by a computer, at which point it gets another name. An example of this is a web search (like Google or Bing). Twenty years ago, we might have called a simple web search "AI", but it has become so commonplace that we've developed shorthand terms for it like "Googling" or simply "web browsing". The same thing is true of most computer gaming systems. In the 60's, they probably would've called a system that could play chess or checkers "AI". Now it's just referred to as a computer or video game.

data analytics - a fancy term for statistics

dataset – a collection of related points of information that can be organized and analyzed as a unit by a computer; for example, a data set might contain a collection of business data like names, salaries, contact information, and sales figures.

fairness metric – a measurement used in statistical analysis to determine whether algorithms are operating fairly.

input – the information that is put in, taken in, or operated on by the algorithm.

machine-learning – a particular sub-area of AI that also incorporates techniques from statistics, in which a computer program "learns" from existing data in order to produce some desired output – either an action to take in response to some input data (if we're talking about robots or drones), or a prediction of some outcome. Not all AI uses machine-learning, but machine-learning is probably the most popular sub-field of AI these days.

omitted variable bias – a bias that occurs when an algorithmic system does not have enough information to make a truly informed prediction and learns to rely on an available, but inadequate proxy variable. For example, if a system was predicting for a person's educational success, but lacked information about their intelligence, studiousness, or access to supportive resources, it might use their postal code or socio-economic status as a proxy variable for these things.

STEPHANIE LACAMBRA, CRIMINAL DEFENSE STAFF ATTORNEY 815 EDDY STREET, SAN FRANCISCO, CA 94109 USA
phone +1.415.436.9333 x130 fax +1.415.436.9993 email stephanie@eff.org website eff.org

output – the information that is produced by the algorithm and bears some relationship to the input.

proxy - a variable that is not in itself directly relevant, but that serves in place of an unobservable or immeasurable variable.

statistical bias – a feature of a statistical technique or of its results whereby the 'expected value' of the results differs from the underlying truth.

validation data – data used to see how well your model performs.

STEPHANIE LACAMBRA, CRIMINAL DEFENSE STAFF ATTORNEY 815 EDDY STREET, SAN FRANCISCO, CA 94109 USA
phone +1.415.436.9333 x130 fax +1.415.436.9993 email stephanie@eff.org website eff.org