Recommended Best Practices for policy makers, data scientists, and the lay public when considering the use of algorithmic decision-making systems that impact individuals or groups in a significant and material way. *

For policy makers:

1. Expressly delineate the policy goals and objectives to be achieved through the algorithmic decision-making system.

2. Understand the inputs used in the algorithm and how each input is weighted.

3. Involve computer scientists, lawyers and social scientists in the design, evaluation, implementation and validation of the algorithm to better insure that the algorithm satisfies policy goals and objectives. Consider establishing an independent panel of experts to evaluate and test the algorithm.

4. Promote transparency of the algorithm by requiring disclosure (either to the public, the relevant agency or the expert panel), of one or more of the following:

a. the source code;
b. the data sets or other "inputs", including the relative weight assigned to each input, used by the algorithm
c. the key decision-making procedural rules or logic used by the algorithm
d. the identification, logging and tracking of errors in data sources and algorithm operation in a form that is capable of being audited.

5. Promote accountability and fairness of the algorithm by requiring, at a minimum, the following

a. a defined avenue of redress for or a review of adverse consequences of an algorithmic decision on an individual;
b. an opportunity to challenge the data inputs used by the algorithm in terms of accuracy, completeness or validity;
c. an explanation in lay terms of how the algorithm works and the basis for any specific decision made;
d. a publicly disclosed procedure for validation and evaluation of the algorithm to ensure the algorithm does not create discriminatory or unjust impacts on protected groups or produce unintended consequences; and
e. regular auditing of the algorithm design and operation and publication of the audit results.

6. Require a predetermined sunset provision for use of the algorithm to insure regular review, updating or reconsideration of the underlying software in light of any revised policy goals.

7. Before using algorithmic systems in criminal sentencing, the relevant government agency should, in addition to the recommendations above, determine whether its primary objectives in sentencing are retribution, deterrence, utilitarian, rehabilitative or a combination thereof and assess the design and operation of the algorithmic system in light of the desired objectives. In any event, any algorithmic

system used in sentencing

a. should not be outcome determinative, but only one of many factors to consider;
b. should be advisory to the sentencing judge;
c. should, wherever practicable, be tailored to the specific jurisdiction and relevant population and take into account any state statutes that identify permissible or impermissible defendant status variables such as race, national origin, religion or gender;
d. should be used to identify appropriate risk factors and offender needs in imposing terms and conditions of probation and supervision, but not to determine the severity of a sentence or whether an offender should be incarcerated;
e. should contain certain cautions, as appropriate, about their limitations and accuracy, including, among other things, whether the algorithm assigns scores based on group data and whether a statistically sound validation study of the system has been conducted based on the relevant sentencing population See State of Wisconsin v. Loomis, Wis. Sup. Ct. (July 13, 2016);
f. should be the subject of a periodic validation study conducted by an independent third-party to assess whether the algorithm is working as intended and does not produce any inappropriate discriminatory, disparate or unjust impacts.
f. Should include appropriate due process mechanisms for a defendant to challenge the data inputs used by the algorithm in terms of accuracy, completeness, validity and the relative weight given to the inputs.

For students of data science:

1. Realize that models are only an imperfect snapshot of reality, and are not the same as truth.

2. Be skeptical about the “unbiasedness” of data. Interrogate sources and think about ways in which data collection might introduce or amplify biases.

3. Realize that the training process searches for *some* model that fits the data, not *the* model.

4. Question the predictions of the model on unseen data. Make efforts to audit and explain the model to understand its decision-making process.

5. Understand that the predictions of a model might be used in decisions that affect real people. Be circumspect about claims of accuracy, and be open and clear about possible sources of uncertainty.

6. Subscribe to a “Hippocratic Oath for Data Scientists” such as that formulated by IBM technologist Marie Wallance https://wiki.p2pfoundation.net/Hippocratic_Oath_for_the_Data_Scientist or British government chief scientific advisor scientist David King http://blogs.nature.com/news/2007/09/hippocratic_oath_for_scientist.html or Harvard researcher Alison Hill http://www.pbs.org/wgbh/nova/next/body/scientific-oath/

7. Insist that computer science curriculum include a component that discusses the ethical issues and challenges inherent in big data, data analytics and automated decision-making systems.

For the lay public

1. Understand that automated decision-making is a result of a training process, and is only as good as the process by which the decision-maker is trained. Ask how the model was trained before believing its conclusions.

2. Machine learning is good at finding patterns, like humans. But like the patterns humans find, ML patterns are not necessarily real.

3. Accuracy is not the same as being fair. An algorithm that makes a few mistakes (as recorded statistically) might make all its mistakes on one well defined subgroup. For this group, the mistakes are many, not few.

4. Automated decision-making works well “on average”. But it is poor at making decisions on people who look like fringe cases. And we don’t know what the algorithm thinks a fringe case is.

3. Above all, remember that automated decision-making exploits apparent correlations. The automated systems don’t usually learn actual causal relationships. Correlation and causation are not necessarily the same, legally or scientifically.

*These recommendations are informed by the ACM “Statement on Algorithmic Transparency and Accountability” (January 12, 2017), The White House “Report on Algorithmic Systems, Opportunity and Civil Rights” (May 2016), the “Principles for Accountable Algorithms” developed by the German based non-profit organization Schloss Dagstuhl-Leibniz Center for Informatics (July 2016), the Pew Research Center Report “Code-Dependent: Pros and Cons of the Algorithmic Age” (February 2017) and the writings of various academics and commentators.