

**An Open Letter to the Members of the Massachusetts Legislature Regarding
the Adoption of Actuarial Risk Assessment Tools in the Criminal Justice System
November 9, 2017**

Dear Members of the Massachusetts Legislature:

We write to you in our individual capacities¹ regarding the proposed introduction of actuarial risk assessment (“RA”) tools in the Commonwealth’s criminal justice system. As you are no doubt aware, Senate Bill 2185² – passed by the Massachusetts Senate on October 27, 2017 – mandates implementation of RA tools in the pretrial stage of criminal proceedings. Specifically:

- Section 182 of the bill would amend Massachusetts General Laws chapter 276 to include the following new Section 58E(a):

Subject to appropriation, pretrial services shall create or choose a risk assessment tool that analyzes risk factors to produce a risk assessment classification for a defendant that will aid the judicial officer in determining pretrial release or detention under sections 58 to 58C, inclusive. Any such tool shall be tested and validated in the commonwealth to identify and eliminate unintended economic, race, gender or other bias.³

- Amendment 146 (which was adopted) would add language to chapter 276 requiring that “[a]ggregate data that concerns pretrial services shall be available to the public in a form that does not allow an individual to be identified.”⁴
- Amendment 147 (which was also adopted) would add language providing that “[i]nformation about any risk assessment tool, the risk factors it analyzes, the data on which analysis of risk factors is based, the nature and mechanics of any validation process, and the results of

¹ For purposes of identification, we note that all signatories to this letter are Harvard- and MIT-based faculty and researchers whose work touches on issues relating to algorithms. Most of the undersigned are involved in a research initiative underway at the MIT Media Lab and Harvard University’s Berkman Klein Center for Internet & Society that seeks to examine ethics and governance concerns arising from the use of artificial intelligence, algorithms, and machine learning technologies. See *AI Ethics and Governance*, MIT Media Lab, <https://www.media.mit.edu/projects/ai-ethics-and-governance/overview/> (last visited Oct. 28, 2017); *Ethics and Governance of Artificial Intelligence*, Berkman Klein Ctr. for Internet & Soc’y, <https://cyber.harvard.edu/research/ai> (last visited Oct. 28, 2017).

² S.B. 2185, 190th Gen. Court (Mass. 2017), available at <https://malegislature.gov/Bills/190/S2185.pdf> (last visited Nov. 2, 2017).

³ *Id.* § 182, 1808–12.

⁴ *Id.* Amendment 146, ID: S2185-146-R1, available at <https://malegislature.gov/Bills/GetAmendmentContent/190/S2185/146/Senate/Preview> (last visited Oct. 29, 2017).

any audits or tests to identify and eliminate bias, shall be a public record and subject to discovery.”⁵

As researchers with a strong interest in algorithms and fairness, we recognize that RA tools may have a place in the criminal justice system. In some cases, and by some measures, use of RA tools may promote outcomes better than the status quo. That said, we are concerned that the Senate Bill’s implementation of RA tools is cursory and does not fully address the complex and nuanced issues implicated by actuarial risk assessments.

The success or failure of pretrial risk assessments in the Commonwealth will depend on the details of their design and implementation. Such design and implementation must be: (a) based on research and data; (b) accompanied (and driven) by clear and unambiguous policy goals; and (c) governed by principles of transparency, fairness, and rigorous evaluation.

As the Massachusetts House considers criminal justice reform legislation, and as both houses of the Legislature seek to reconcile their bills, we urge the Commonwealth to engage in significant study and policy development in this area. That study and policy development should ideally take place before the Legislature issues a mandate regarding adoption of risk assessment tools or, at the very least, before any particular tool is developed, procured, and/or implemented. As described herein, we submit that thoughtful deliberation is particularly important in five critical areas.

(1) The Commonwealth should take steps to mitigate the risk of amplifying bias in the justice system.

Research shows the potential for risk assessment tools to perpetuate racial and gender bias.⁶ Researchers have proposed multiple “fairness criteria” to mitigate this bias statistically.⁷ But there remain intrinsic tradeoffs between fairness and accuracy that are mathematically impossible for any RA tool to overcome. Senate Bill 2185 includes a single sentence on eliminating bias; we submit that this issue deserves far more consideration and deliberation.

⁵ *Id.* Amendment 147, ID: S2185-147 (2017), available at <https://malegislature.gov/Bills/GetAmendmentContent/190/S2185/147/Senate/Preview> (last visited Oct. 29, 2017).

⁶ See, e.g., Alexandra Chouldechova, *Fair prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments*, arXiv:1703.00056 (submitted on Feb. 28, 2017), available at <https://arxiv.org/abs/1703.00056> (last visited Oct. 28, 2017); Devlin Barrett, *Holder Cautions on Risk of Bias in Big Data Use in Criminal Justice*, Wall St. J., Aug. 1, 2014, <https://www.wsj.com/articles/u-s-attorney-general-cautions-on-risk-of-bias-in-big-data-use-in-criminal-justice-1406916606> (last visited Oct. 28, 2017); Michael Tonry, *Legal and Ethical Issues in the Prediction of Recidivism*, 26 Fed. Sentencing Reporter 167, 173 (2014).

⁷ Richard Berk et al., *Fairness in Criminal Justice Risk Assessments: The State of the Art*, arXiv preprint arXiv:1703.09207 (submitted on Mar. 27, 2017, last rev. 28 May 2017), available at <https://arxiv.org/abs/1703.09207> (last visited Oct. 28, 2017).

Before implementing any RA tool, the Commonwealth should consider developing specific criteria along the following lines:

- (a) The Commonwealth should develop fairness criteria that mitigate the risk of an RA tool exacerbating bias on the basis of race, gender, and other protected classes.
- (b) The Commonwealth should craft rules and guidelines for identifying and ethically handling “proxy variables” (which correlate with race, gender, and other protected characteristics in any RA tool) and addressing other means by which such characteristics may be inferred from ostensibly neutral data. Notably in this regard, the state of California – which moved toward use of pretrial risk assessment tools relatively early – is now actively considering legislation to eliminate housing status and employment status from risk assessments, because these variables are strong proxies for race and class.⁸ If passed, such legislation would require counties to alter and adapt the patchwork of individual pretrial risk assessment tools in use across that state.⁹ We submit that the Commonwealth might learn from this example by putting in work upfront to fully understand bias and address proxies, rather than moving forward with implementation and specifying change at a later date.
- (c) The Commonwealth should create guidelines that govern data used in the development and validation of RA tools, to ensure tools deployed in Massachusetts are appropriately well-tailored to local populations and demographic structures.

(2) The Commonwealth should clarify procedures for validation and evaluation of risk assessment tools.

Research has shown that RA tools must be evaluated regularly and repeatedly to ensure their validity over time.¹⁰ In providing for adoption and use of risk assessments, the Commonwealth should take the opportunity to establish baselines concerning such review and evaluation. In particular, we urge the development of the following kinds of specifications:

- (a) The Commonwealth should require mandatory, jurisdiction-by-jurisdiction validation checks, including rigorous comparison of a given tool’s predictions to observed results (such as re-conviction and failure to appear in court).

⁸ See Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 Stan. L. Rev. 803 (2014), available at <https://www.stanfordlawreview.org/print/article/evidence-based-sentencing-and-the-scientific-rationalization-of-discrimination/> (last visited Nov. 2, 2017).

⁹ See S.B. 10, 2017-2018 Reg. Sess. (Cal. 2017), available at http://leginfo.ca.gov/faces/billNavClient.xhtml?bill_id=201720180SB10 (last visited Nov. 1, 2017).

¹⁰ See *Risk and Needs Assessment and Race in the Criminal Justice System*, Justice Ctr., Council State Gov’ts (May 31, 2016), <https://csgjusticecenter.org/reentry/posts/risk-and-needs-assessment-and-race-in-the-criminal-justice-system/> (last visited Oct. 28, 2017).

- (b) The Commonwealth should insist that RA tools are tested on a regular basis to measure the disparate impact of tool error rates by race, gender, and other protected classes and should ensure that researchers have access to data and algorithms necessary to support robust testing.
- (c) The Commonwealth should develop processes to promote regular (e.g., bi-annual) external oversight of validation checks of RA tools by an independent group – possibly a standing commission – which includes perspectives of statisticians, criminologists, and pretrial and probation service workers specific to the relevant jurisdiction.

(3) The Commonwealth should promulgate procedures for effective deployment of risk assessment tools.

Risk assessment tools employ statistical methods to produce risk scores. Representatives of the court system (usually, judges) use those numerical scores as one input in their pretrial decision-making processes, in the context of applicable legal standards. Use of an RA tool in a given case may involve a combination of statistical methods, fact determinations, and policy considerations. It is vital that all stakeholders in the pretrial pipeline be trained to accurately interpret and understand RA tools and the meaning (and limitations) of the risk assessment scores they produce.

By way of example, the classification of a risk category applicable to a particular criminal defendant with respect to a given risk score (e.g., high risk, medium risk, or low risk) is a matter of policy, not math. Tying the definition of terms like “high risk” to scores that are the products of RA tools can influence both: (a) decision-making by prosecutors, defendants, and judges in a pretrial setting (who may place undue emphasis on numerical scores generated by computers); and (b) public perception of the specific outcomes of RA tools. It is essential that the Commonwealth make clear how those risk scores are generated and what they purport to predict.

In this regard, we suggest the following:

- (a) The Commonwealth should mandate continual training processes for all system actors to ensure consistency and reliability of risk score characterizations, irrespective of race, gender and other immutable characteristics.
- (b) The Commonwealth should require timely and transparent record-keeping practices that enable the auditing and adjustment of RA classifications over time.
- (c) The Commonwealth should dictate a consistent decision-making framework to support appropriate interpretation of risk assessment predictions by all actors in the pretrial system. This framework should be regularly updated to reflect ongoing research about what specific conditions (i.e. electronic monitoring, weekly supervision meetings, etc.) have been empirically tested and proven to lower specific types of risk.

- (d) The Commonwealth should provide adequate funding and resources for the formation and operation of an independent pretrial service agency that stands separate from other entities in the criminal justice system (such as probation offices and correctional departments). This agency will deal with the increased supervision caseload of individuals who are released prior to their trial date.
- (e) The Commonwealth must ensure that updates to RA tools are accompanied by a detailed articulation of new intended risk characterizations.

(4) The Commonwealth should ensure that RA tools adequately distinguish among the types of risks being assessed.

A variety of risks may be relevant to a pre-trial determination such as bail. These risks may include (for example) the risk that a defendant will fail to appear for a hearing; the risk that a defendant will flee the jurisdiction; and the risk that defendant will engage in new criminal activity. Each of these risks may require different assessments, based on different factors, and each may need to be separately considered and weighed in accordance with applicable legal standards in the context of a given pretrial decision.

Despite this complexity, most pretrial RA tools do not adequately differentiate among types of risks they purport to predict. An individual may be assigned a score indicating high risk in one category but not another, and the output report may not delineate this distinction. This can have significant implications for pretrial release decisions. A high risk of failure to appear in court due to mental health issues is not the same as a high risk that a defendant will commit a violent crime while awaiting trial. We urge the Legislature to ensure that RA tools adopted in the Commonwealth adequately differentiate among types of risks being assessed, so that courts can effectively identify appropriate conditions to place on defendants for release.

(5) The Commonwealth should give careful consideration to the process of developing or procuring RA tools, fully exploring the possibility of developing tools in-house, and establishing basic requirements for any tools developed by private vendors.

When a government entity seeks to adopt and implement any technological tool, it can do so in one of two ways. First, it can develop the tool on its own (relying on government personnel and/or outside developers). Second, it can purchase or license existing technology from a private outside vendor. In this regard, we submit that all of the factors identified in this letter should be considered by the Commonwealth with an eye toward informing two key decisions:

- (a) a decision about whether Massachusetts should develop new risk assessment tools or procure existing ones; and
- (b) establishing and enforcing concrete procurement criteria in the event the Commonwealth chooses to buy or license existing technology.

To the first point (re: whether to develop new tools or procure existing ones) – it is worth being mindful of cautionary tales such as the experience of local jurisdictions that sought to upgrade their voting infrastructures and implement electronic voting in the wake of the disputed 2000 United States presidential election.¹¹ Nearly twenty years later, many municipalities find themselves bound by undesirable contracts with a handful of outside vendors that offer unreliable voting machines and tallying services. Some of these vendors assert intellectual property protections in ways that complicate effective audits of the machines' accuracy and integrity.¹² Dissatisfaction with vendors is rarely sufficient to occasion a change in course, because of sunk costs and the burdens of reworking locked-in procedures. The Commonwealth must strive to avoid a structural repeat of governments' regrets around proprietary private voting infrastructure. There are strong arguments that the development of risk assessment tools for the justice system should be undertaken publicly rather than privately, that results should be shareable across jurisdictions, and that outcomes should be available for interrogation by the public at large.

To the second point (re: criteria for procurement) – we are hopeful that this document can serve as the basis for a roadmap toward development of comprehensive procurement guidelines in the event that the Commonwealth decides to buy or license existing tools developed by private vendors rather than developing its own tools. Stated simply, procurement decisions cannot be based solely on considerations of cost or efficiency and must be driven by principles of transparency, accountability, and fairness. Those principles must be codified to ensure that the Commonwealth and its citizens leverage their purchasing power with vendors to understand what tools are being procured and ensure those tools operate fairly. Private vendors may raise concerns about scrutiny of their technologies and the algorithms they employ given proprietary business considerations. But, the Commonwealth must balance those private pecuniary interests against the overwhelming public interest in ensuring our criminal justice system satisfies fundamental notions of due process. The transparency measures described in Amendment 147 are welcome additions to the Senate Bill, and we urge consideration of additional measures that support fully-informed decision-making on this important issue.¹³

¹¹ See, e.g., Andrew W. Appel et al., *The New Jersey Voting-Machine Lawsuit and the AVC Advantage DRE Voting Machine*, in EVT/WOTE'09: Electronic Voting Technology Workshop / Workshop on Trustworthy Elections (2009), available at <https://www.cs.princeton.edu/~appel/papers/appel-evt09.pdf> (last visited Nov. 2, 2017).

¹² See, e.g., Alex Halderman, *How to Hack an Election in 7 Minutes*, Politico (Aug. 6, 2016), <https://www.politico.com/magazine/story/2016/08/2016-elections-russia-hack-how-to-hack-an-election-in-seven-minutes-214144> (last visited Oct. 28, 2017); David S. Levine, *Can We Trust Voting Machines?*, Slate (Oct. 24, 2012), www.slate.com/articles/technology/future_tense/2012/10/trade_secret_law_makes_it_impossible_to_independently_verify_that_voting.html (last visited Oct. 28, 2017).

¹³ By way of example, a recently proposed New York City Council Local Law would amend the administrative code of the City of New York to require agencies that use algorithms in certain contexts to both: (a) publish the source code used for such processing; and (b) accept user-submitted data sets that can be processed by the agencies' algorithms and provide the outputs to the user. See Introduction No. 1696-2017, N.Y.C. Council (2017), available at

In conclusion, decisions around confinement and punishment are among the most consequential and serious that a government can make. They are non-delegable, and any technological aids that are not transparent, auditable, and improvable by the state cannot be deployed in the Commonwealth. Massachusetts has wisely avoided jumping rapidly into the use of RA tools. It is now in a position to consider them with the benefit of lessons from jurisdictions that have gone first. We submit that – given that the potential benefits and dangers of pretrial RA tools rest on the details of tool development, oversight, and training, informed by clear policy goals – it is imperative that laws and regulations governing the introduction of pretrial RA tools be clear, concrete, specific, and data-driven. We are happy to assist in this effort.

Respectfully submitted,

Chelsea Barabas
Research Scientist,
MIT Media Laboratory

Christopher Bavitz
WilmerHale Clinical Professor of Law,
Harvard Law School

Ryan Budish
Assistant Research Director
Berkman Klein Center for Internet & Society

Karthik Dinakar
Research Scientist,
MIT Media Laboratory

Cynthia Dwork
Gordon McKay Professor of Computer Science,
Harvard School of Engineering
and Applied Sciences
Radcliffe Alumnae Professor,
Radcliffe Institute for Advanced Study

Urs Gasser
Professor of Practice,
Harvard Law School

Kira Hessekiel
Project Coordinator,
Berkman Klein Center for Internet & Society

Joichi Ito
Director,
MIT Media Laboratory

Ronald L. Rivest
MIT Institute Professor

Madars Virza
Research Scientist,
MIT Media Laboratory

Jonathan Zittrain
George Bemis Professor of International Law,
Harvard Law School
and Harvard Kennedy School
Professor of Computer Science,
Harvard School of Engineering
and Applied Sciences