A Guide to Solving Social Problems with Machine Learning

by Jon Kleinberg, Jens Ludwig, and Sendhil Mullainathan

It's Sunday night. You're the deputy mayor of a big city. You sit down to watch a movie and ask Netflix for help. ("Will I like Birdemic? Ishtar? Zoolander 2?") The Netflix recommendation algorithm predicts what movie you'd like by mining data on millions of previous movie-watchers using sophisticated machine learning tools. And then the next day you go to work and every one of your agencies will make hiring decisions with little idea of which candidates would be good workers; community college students will be largely left to their own devices to decide which courses are too hard or too easy for them; and your social service system will implement a reactive rather than preventive approach to homelessness because they don't believe it's possible to forecast which families will wind up on the streets.	这是星期天晚上。 你是一个大城市的副市长。 你坐下来看电影和向Netflix寻求帮助。 ("我会喜欢Birdemic吗？Ishtar？Zoolander 2？") Netflix推荐算法通过挖掘数百万的数据来预测你喜欢的电影.以前的电影观察者使用复杂的机器学习工具。 然后第二天你去上班，你的每一个代理机构都会毫不犹豫地做出招聘决定：哪些候选人是好工人; 社区大学生将在很大程度上留下来，他们自己的设备来决定哪些课程对他们来说太难或太容易; 和你的社交服务系统将实施无家可归的反应而非预防方法，因为他们不相信有可能预测哪些家庭会结束 街道
You'd love to move your city's use of predictive analytics into the 21 century, or at least into the 20 century. But how? You just hired a pair of 24-year-old computer programmers to run your data science team. They're great with data. But should they be the ones to decide which problems are amenable to these tools? Or to decide what success looks like? You're also not reassured by the vendors the city interacts with. They're always trying to upsell you the very latest predictive tool. Decisions about how these tools are used seem too important for you to outsource, but raise a host of new issues that are difficult to understand	您希望将您的城市的预测分析用于21世纪或至少进入20世纪。 但是怎么样？ 你刚刚聘请了一对24岁的计算机程序员运营您的数据科学团队。 他们对数据非常了解。 但他们应该是那些人决定哪些问题适合这些工具？ 或者决定成功是什么样的？你也不会被这座城市与之互动的供应商放心。他们总是试图卖给你最新的预测工具。 关于如何使用这些工具的决定似乎也是如此。对你来说很重要，但要提出一系列难以解决的新问题了解。
This mix of enthusiasm and trepidation over the potential social impact of machine learning is not unique to local government or even to government: non-profits and social entrepreneurs share it as well. The enthusiasm is well-placed. For the right type of problem, there are enormous gains to be made from using these tools. But so is the trepidation: as with all new "products," there is potential for misuse. How	这种热情和惶恐的混合机器的潜在社会影响学习并非地方政府所独有。甚至对政府来说：非营利和社会企业家也分享它。该热情是有利的。 对于正确的类型问题是，使用这些工具可以获得巨大的收益。 但是，这情况也是令人惶恐：与所有新产品一样，存在滥用的可能性。 我们怎样才能最大化在减少伤害的同时带来的好处？

can we maximize the benefits while minimizing the harm? In applying these tools the last few years, we have focused on exactly this question. We have learned that some of the most important challenges fall within the cracks between the discipline that builds algorithms (computer science) and the disciplines that typically work on solving policy problems (such as economics and statistics). As a result, few of these key challenges are even on anyone's radar screen. The good news is that many of these challenges, once recognized, are fairly straightforward to solve.	在过去几年应用这些工具时，我们一直关注这个问题。 我们已经了解到一些最重要的挑战属于裂缝之间构建算法（计算机科学）和通常有效的学科的学科解决政策问题（如经济学和统计学）。 因此，这些关键很少甚至在任何人的雷达屏幕上也存在挑战。 好消息是其中许多挑战一旦被注意到，就很容易解决。
We have distilled what we have learned into a "buyer's guide." It is aimed at anyone who wants to use data science to create social good, but is unsure how to proceed.	我们将我们学到的东西提炼成"买方指南"。它针对的是这些想利用数据科学创造社会利益，但不确定如何去实现的人。
How machine learning can improve public policy	**机器学习如何改善公共政策**
First things first: There is always a new "new thing." Especially in the social sector. Are these machine learning tools really worth paying attention to? Yes. That's what we've concluded from our own proof-of-concept project, applying machine learning to a dataset of over one million bond court cases (in joint work with Himabindu Lakkaraju and Jure Leskovec of Stanford University). Shortly after arrest, a judge has to decide: will the defendant await their legal fate at home? Or must they wait in jail? This is no small question. A typical jail stay is between two and three months.	首先要做的事情是：总有一种 新的 "新事物"。特别是在社会领域。 这些是机器学习工具真的值得关注吗? 是。 这就是我们从自己的概念验证项目中应用机器所得出的结论学习超过一百万个债券法庭案件的数据集（与Himabindu合作Lakkaraju和斯坦福大学的Jure Leskovec）。 被捕后不久，法官必须这样做决定：被告是否会等待他们在家的合法命运？ 或者他们必须等在监狱吗？ 这不是小问题。 典型的监狱逗留时间为两到三个月。
In making this lifechanging decision, by law, the judge has to make a prediction: if released, will the defendant return for their court appearance, or will they skip court? And will they potentially commit further crimes?	为了做出在这个改变生活的决定通过法律，法官必须做出预测：如果被释放，将是被告回到他们的法庭出庭，还是他们会跳过法庭? 他们是否会承诺进一步犯罪?
We find that there is considerable room to improve on judges' predictions. Our estimates show that if we made pre-trial release decisions using our algorithm's predictions of risk instead of relying on judge intuition, we could reduce crimes committed by released defendants by up	我们发现法官的预测有很大的提升空间。 我们的估计表明：如果我们使用我们的算法预测风险，做出预审释放的决定，我们可以减少25%犯罪来自于被释放的人，无需监禁任何额外的人。

to 25% without having to jail any additional people. Or, without increasing the crime rate at all, we could jail up to 42% fewer people. With 12 million people arrested every year in the U.S., this type of tool could let us reduce jail populations by up to several hundred thousand people. And this sort of intervention is relatively cheap. Compared to investing millions (or billions) of dollars into more social programs or police, the cost of statistically analyzing administrative datasets that already exist is next-tonothing. Plus, unlike many other proposals to improve society, machine learning tools are easily scaled.	换句话说，再不提高犯罪率的情况下， 我们可以减少关押42%的人。 在每年有1200万人被捕的美国，这种工具可以让我们最多减少几十万的被关押的人数。 而这种干预相对便宜。 与向更多社会计划或警察投入数百万（或数十亿）美元相比，已经存在的统计分析管理数据集的成本是次要的。 此外，与许多其他改善社会的建议不同，机器学习工具能改善社会，同时也很容易缩放。
By now, policymakers are used to hearing claims like this in sales pitches, and they should appropriately raise some skepticism. One reason it's hard to be a good buyer of machine learning solutions is that there are so many overstated claims. It's not that people are intentionally misstating the results from their algorithms. In fact, applying a known machine learning algorithm to a dataset is often the most straightforward part of these projects. The part that's much more difficult, and the reason we struggled with our own bail project for several years, is accurately evaluating the potential impact of any new algorithm on policy outcomes. We hope the rest of this article, which draws on our own experience applying machine learning to policy problems, will help you better evaluate these sales pitches and make you a critical buyer as well.	到目前为止，政策制定者习惯于在销售宣传中听到这样的声明，他们应该这样做适当地提出一些怀疑。 （政策制定者）难以成为机器学习解决方案的好买家的一个原因是有太多夸大其词的说法。 这不是人故意错误地解释他们的算法结果。 事实上，应用一个已知的机器学习算法到数据集通常是这些项目中最直接的部分。 这部分的困难很多，以及我们为自己的保释而挣扎的原因。 项目多年来，正在准确评估任何新算法对政策结果的潜在影响。 我们希望本文的其余部分能够借鉴我们自己的经验将机器学习应用于政策问题，将有助于您更好地评估这些销售宣传，让你成为一个重要的买家。
Look for policy problems that hinge on prediction Our bail experience suggests that thoughtful application of machine learning to policy can create very large gains. But sometimes these tools are sold like snake oil, as if they can solve every problem. Machine learning excels at predicting things. It can inform decisions that hinge on a prediction, and where the thing to be predicted is clear and measurable.	**寻找与预测相关的政策问题** 我们的保释经验表明机器学习在政策上的周到应用可以创造非常大的收益。 但有时这些工具像蛇油一样出售，好像他们可以解决每个问题。 机器学习擅长预测事物。 它可以为决策提供依据预测，以及预测的事物清晰可衡量的地方。

For Netflix, the decision is what movie to watch. Netflix mines data on large numbers of users to try to figure out which people have prior viewing histories that are similar to yours, and then it recommends to you movies that these people have liked.	对于Netflix，要决定的是要观看的电影。 Netflix挖掘了大量的数据用户试图找出哪些人的观看历史与您的相似，然后它会向你推荐这些人喜欢的电影。
For our application to pre-trial bail decisions, the algorithm tries to find past defendants who are like the one currently in court, and then uses the crime rates of these similar defendants as the basis for its prediction.	对于我们的申请审前保释决定，该算法试图找到和新的被告（经历相似的）一样的曾经判决过的被告，然后使用这些类似被告的犯罪率作为依据进行预测。
If a decision is being made that already depends on a prediction, why not help inform this decision with more accurate predictions? The law already requires bond court judges to make pre-trial release decisions based on their predictions of defendant risk. Decades of behavioral economics and social psychology teach us that people will have trouble making accurate predictions about this risk – because it requires things we're not always good at, like thinking probabilistically, making attributions, and drawing inferences. The algorithm makes the same predictions judges are already making, but better.	如果做出的决定已经取决于预测，为什么不用更准确预测去帮助做出决定？ 法律 已经要求债券法院法官根据对被告风险的预测做出预审释放的决定。 几十年行为经济学和社会心理学告诉我们，人们将难以制作准确预测这种风险 - 因为它需要我们并不总是擅长的事情，像概率思考，做出归因和绘制推论。 算法可以做出与法官已经做出的做法相同的预测，但更好。
But many social-sector decisions do not hinge on a prediction. Sometimes we are asking whether some new policy or program works – that is, questions that hinge on understanding the causal effect of something on the world. The way to answer those questions is not through machine learning prediction methods. We instead need tools for causation, like randomized experiments.	但许多社会部门的决定并不取决于预测。 有时我们会问一些新的政策或计划是否有效 - 也就是说，取决于理解的问题某事物对世界的因果影响。 回答这些问题的方法不是通过机器学习预测的方法。 我们反而需要工具确定因果关系，像随机实验。
In addition, just because something is predictable, that doesn't mean we are comfortable having our decision depend on that prediction. For example we might reasonably be uncomfortable denying welfare to someone who was eligible at the time they applied just because we predict they have a high likelihood to fail to abide by the program's job-search requirements or fail a drug test in the future.	另外，仅仅因为某些东西是可预测的，那不是意味着我们很乐意让我们的决定取决于那个预测。 比如我们会感到不舒服如果拒绝给予那些有资格获得福利金的人的福利申请只是因为我们预测他们很有可能不遵守计划的求职要求或将来未通过药检。
	确保您对预测的结果感到满意

Make sure you're comfortable with the outcome you're predicting Algorithms are most helpful when applied to problems where there is not only a large history of past cases to learn from but also a clear outcome that can be measured, since measuring the outcome concretely is a necessary prerequisite to predicting. But a prediction algorithm, on its own, will focus relentlessly on predicting the outcome you provide as accurately as possible at the expense of everything else. This creates a danger: if you care about other outcomes too, they will be ignored. So even if the algorithm does well on the outcome you told it to focus on, it may do worse on the other outcomes you care about but didn't tell it to predict. This concern came up repeatedly in our own work on bail decisions. We trained our algorithms to predict the overall crime rate for the defendents eligible for bail. Such an algorithm treats every crime as equal. But what if judges (not unreasonably) put disproportionate weight on whether a defendant engages in a very serious violent crime like murder, rape, or robbery? It might look like the algorithm's predictions leads to "better outcomes" when we look at overall rates of crime. But the algorithm's release rule might actually be doing worse than the judges with respect to serious violent crimes specifically. The possibility of this happening doesn't mean algorithms can't still be useful. In bail, it turns out that different forms of crime are correlated enough so that an algorithm trained on just one type of crime winds up out-predicting judges on almost every measure of criminality we could construct, including violent crime. The point is that the outcome you select for your algorithm will define it. So you need to think carefully about what that outcome is and what else it might be leaving out.	算法不仅在应用于从过去的大量历史案例中学习，而且衡量一个明确的结果时最有用。 ，因为具体地衡量结果是预测的必要前提。 但是预测算法本身将无情地集中在预测你提供的结果上尽可能准确地牺牲其他一切。 这会造成危险：如果你在乎关于其他结果，它们将被忽略。 所以即使算法做得好，你告诉它要关注的结果，它可能会对你关心的其他结果做得更糟，如果没有告诉它要去预测。 在我们自己的保释决定工作中反复出现这种担忧。我们训练了我们用于预测有资格获得保释的被告的总体犯罪率的算法。 这样的算法将每个犯罪视为平等。 但是如果法官（不是不合理地）设置了不合理的权重在对被告是否参与非常严重的暴力犯罪谋杀，强奸还是抢劫该怎么办呢？ 看起来算法的预测可能会"更好结果"当我们审视整体犯罪率时。 但算法发布的规则可能会实际上，特别是在严重暴力犯罪方面比法官更糟糕。发生这种情况的可能性并不意味着算法仍然无法使用。 提到保释，它事实证明，不同形式的犯罪 是相互关联的，所以只针对一种类型的犯罪进行训练的算法在我们可以构建的几乎所有犯罪行为（包括暴力犯罪）上做出超出预测的判断。关键是你的结果选择您的算法将定义它。 所以你需要仔细考虑结果是什么，它可能会遗漏什么。

Check for bias Another serious example of this principle is the role of race in algorithms. There is the possibility that any new system for making predictions and decisions might exacerbate racial disparities, especially in policy domains like criminal justice. Caution is merited: the underlying data used to train an algorithm may be biased, reflecting a history of discrimination. And data scientists may sometimes inadvertently report misleading performance measures for their algorithms. We should take seriously the concern about whether algorithms might perpetuate disadvantage, no matter what the other benefits. Ultimately, though, this is an empirical question. In our bail project, we found that the algorithm can actually reduce race disparities in the jail population. In other words, we can reduce crime, jail populations and racial bias – all at the same time – with the help of algorithms. This is not some lucky happenstance. An appropriate first benchmark for evaluating the effect of using algorithms is the existing system – the predictions and decisions already being made by humans. In the case of bail, we know from decades of research that those human predictions can be biased. Algorithms have a form of neutrality that the human mind struggles to obtain, at least within their narrow area of focus. It is entirely possible—as we saw—for algorithms to serve as a force for equity. We ought to pair our caution with hope. The lesson here is that if the ultimate outcome you care about is hard to measure, or involves a hard-to-define combination of outcomes, then the problem is probably not a good fit for machine learning. Consider a problem that looks like bail: Sentencing. Like bail, sentencing of people who have been found guilty depends partly on recidivism risk. But sentencing also depends on things like	**检查偏差** 这个原则的另一个重要例子是种族在算法中的作用。 有任何制定预测和决策的新系统都可能加剧的可能性种族差异，特别是在刑事司法等政策领域。 值得注意的是：用于训练算法的基础数据可能有偏差，反映了历史歧视。 数据科学家有时可能会无意中报告误导他们的算法的性能测量。 我们应该认真关注无论其他什么好处，算法是否会使劣势永久化。 但最终，这是一个经验问题。 在我们的保释项目中，我们发现了算法实际上可以 减少监狱人口中的种族差异。 换句话说， 我们可以在算法的帮助下减少犯罪，监狱人口 和种族偏见 。 这不是一些幸运的偶然事件。 适合评估使用算法的效果的第一个基准是现有系统 - 预测和决策已经存在由人类做出了。 在保释的案例中，我们从几十年的研究中了解到这些人类预测可能有偏见。 算法具有的中立形式，这是人类想努力获得的思维方式，至少在他们狭窄的焦点范围内。 这完全有可能 - 就像我们看到的-算法具有作为公平的力量。 我们应该把我们的谨慎与希望结合起来。 这里的教训是，如果您关心的最终结果难以衡量，或者涉及难以定义的结果组合，那么问题可能并不适合机器学习。 考虑一个 看起来像保释的问题： 量刑。 像保释一样对被判有罪的人判刑部分取决于累犯风险。 但量刑还取决于社会的报应感，怜悯感和赎回，这些是无法直接衡量。

society's sense of retribution, mercy, and redemption, which cannot be directly measured. We intentionally focused our work on bail rather than sentencing because it represents a point in the criminal justice system where the law explicitly asks narrowly for a prediction. Even if there is a measurable single outcome, you'll want to think about the other important factors that aren't encapsulated in that outcome – like we did with race in the case of bail – and work with your data scientists to create a plan to test your algorithm for potential bias along those dimensions.	我们故意将工作重点放在保释上而不是判刑，因为它代表了刑事司法系统中的一个观点：法律明确要求进行预测。 即使有可衡量的单一结果，你会想到其他未被考虑在结果里的重要因素--就像我们在保释的情况下对种族做的那样 - 并与数据科学家合作创建一个计划来测试您的算法在这些维度上的潜在偏差。
Verify your algorithm in an experiment on data it hasn't seen	**在实验中验证您的算法没有看到的数据**
Once we have selected the right outcome, a final potential pitfall stems from how we measure success. For machine learning to be useful for policy, it must accurately predict "out-of-sample." That means it should be trained on one set of data, then tested on a dataset it hasn't seen before. So when you give data to a vendor to build a tool, withhold a subset of it. Then when the vendor comes back with a finished algorithm, you can perform an independent test using your "hold out" sample.	一旦我们选择了正确的结果，最终的潜在陷阱源于我们如何衡量成功。 要使机器学习对策略有用，它必须准确预测"样本外"。这意味着它应该在一组数据上进行训练，然后在它以前没见过的数据集上进行测试。 因此，当您向供应商提供数据以构建工具时，请保留一个它的子集。 然后，当供应商返回完成的算法时，您可以执行使用"保持"样本进行独立测试。
An even more fundamental problem is that current approaches in the field typically focus on performance measures that, for many applications, are inherently flawed. Current practice is to report how well one's algorithm predicts only among those cases where we can observe the outcome. In the bail application this means our algorithm can only use data on those defendants who were released by the judges, because we only have a label providing the correct answer to whether the defendant commits a crime or not for defendants judges chose to release. What about defendants that judges chose not to release? The available data cannot tell us whether they would have reoffended or not.	更基本的问题是该领域的当前方法通常关注于此。 对于许多应用来说，性能测量本质上是有缺陷的。 目前的做法是报告一个人的算法仅可以在我们可以观察到结果的案例中进行预测。 在保释申请中，这意味着我们的算法只能使用那些由法官释放的被告的数据，因为我们只有一个 标签提供正确回答被告是否为被告或者法官选择释放犯罪。 法官选择不释放的被告怎么办？可用数据无法告诉我们他们是否会被重新审判。
This makes it hard to evaluate whether any new machine learning tool can actually improve	这使得很难评估是否有任何新的机器学习工具能够真正改进相对于现有决策系统的结果 - 在这种情况下，审判。 如果有些新的基于机器学习的释放规则想要释放某

outcomes relative to the existing decision-making system — in this case, judges. If some new machine learning-based release rule wants to release someone the judges jailed, we can't observe their "label", so how do we know what would happen if we actually released them?	个被判入狱的人，我们不能观察他们的"标签"，那么我们怎么知道如果我们实际释放它们会发生什么？
This is not merely a problem of academic interest. Imagine that judges have access to information about defendants that the algorithm does not, such as whether family members show up at court to support them. To take a simplified, extreme example, suppose the judge is particularly accurate in using this extra information and can apply it to perfectly predict whether young defendants re-offend or not. Therefore the judges release only those young people who are at zero risk for re-offending. The algorithm only gets to see the data for those young people who got released – the ones who never re-offend.	这不仅仅是学术兴趣的问题。 想象一下，法官可以访问关于算法没有的被告的信息，例如是否是家庭成员出现在法庭上以支持他们。 举一个简单的极端例子，假设法官在使用这些额外信息时特别准确，可以由这些信息准确预测年轻被告是否再次冒犯。 因此，法官只释放那些年轻人那些重新犯罪风险为零的人。 而算法只能查看那些不会再犯的被释放的年轻人的数据。
Such an algorithm would essentially conclude that the judge is making a serious mistake in jailing so many youthful defendants (since none of the ones in its dataset go on to commit crimes). The algorithm would recommend that we release far more youthful defendants. The algorithm would be wrong. It could inadvertently make the world worse off as a result.	这样的算法可以得出的基本结论是，法官在监禁这么多人时犯了严重错误（因为其数据集中没有一个继续犯罪）该算法会建议我们释放更多年轻的被告。 算法会错的。 它可能无意中使世界变得更糟。
In short, the fact that an algorithm predicts well on the part of the test data where we can observe labels doesn't necessarily mean it will make good predictions in the real world. The best way to solve this problem is to do a randomized controlled trial of the sort that is common in medicine. Then we could directly compare whether bail decisions made using machine learning lead to better outcomes than those made on comparable cases using the current system of judicial decision-making.	简而言之，算法可以很好地预测测试数据的这一事实观察标签并不 一定意味着它会在现实世界中做出良好的预测。 解决这个问题的最好方法是做一个随机对照试验，这在医学上很常见。 然后我们可以直接比较使用了机器学习的保释决定是否比使用现行司法决策制度能得到更好的结果。
But even before we reach that stage, we need to make sure the tool is promising enough to ethically justify testing it in the field. In our bail case, much of the effort went into finding a "natural experiment" to evaluate the tool.	但即使在我们达到这个阶段之前，我们也需要确保该工具足够有希望在道德上公平地在这个领域测试。 在我们的保释金案例中，很多努力都用于寻找"自然实验" 用来评估工具（的有效性）。

Our natural experiment built on two insights. First, within jurisdictional boundaries, it's essentially random which judges hear which cases. Second, judges are quite different in how lenient they are. This lets us measure how good judges are at selecting additional defendants to jail. How much crime reduction does a judge with a 70% release rate produce compared to a judge with an 80% release rate? We can also use these data to ask how good an algorithm would be at selecting additional defendants to jail.	我们的自然实验基于两个见解。 首先，在管辖范围内，它是法官听到哪些案件基本上是随机的。 其次，法官在如何宽容方面有很大不同。 这让我们可以衡量法官在审判新的被告入狱的评判有多好。 具有70%释放率的法官与80%释放率的法官相比能减少多少犯罪率？ 我们也可以使用这些数据来验证在审判新的被告入狱的算法有多好。
If we took the caseload of an 80% release rate judge and used our algorithm to pick an additional 10% of defendants to jail, would we be able to achieve a lower crime rate than what the 70% release rate judge gets? That "human versus machine" comparison doesn't get tripped up by missing labels for defendants the judges jailed but the algorithm wants to release, because we are only asking the algorithm to recommend additional detentions (not releases). It's a comparison that relies only on labels we already have in the data, and it confirms that the algorithm's predictions do indeed lead to better outcomes than those of the judges	如果我们采取了案件量80%释放率的法官并使用我们的算法来挑选另外10%的被告，我们能否达到比70%释放率法官更低的犯罪率？ 那种"人与机器"的比较并没有因缺少标签而被绊倒被告被判入狱，但算法想释放，因为我们只是想让算法建议是否需要额外拘留（而非释放）。 这个比较仅依赖于我们在数据中已有的标签，它确认了该算法预测确实会带来比法官更好的结果。
It can be misguided, and sometimes outright harmful, to adopt and scale up new predictive tools when they've only been evaluated on cases from historical data with labels, rather than evaluated based on their effect on the key policy decision of interest. Smart users might go so far as to refuse to use any prediction tool that does not take this evaluation challenge more seriously.	当他们仅使用带有标签的历史数据对案例进行评估时的工具而不是根据它们对关键政策决定的影响进行评估时采用和扩大新的预测可能是误导的，有时甚至是有害的。聪明的用户可能会拒绝使用任何不严谨地对待评估挑战的预测工具。
Remember there's still a lot we don't know While machine learning is now widely used in commercial applications, using these tools to solve policy problems is relatively new. There is still a great deal that we don't yet know but will need to figure out moving forward.	**记住还有很多我们不知道的事情** 虽然机器学习现在广泛用于商业应用，但使用这些工具解决政策问题比较新。 还有很多我们还不知道但是需要弄清楚前进的方向。
Perhaps the most important example of this is how to combine human judgment and	也许最重要的例子是如何结合人类的判断和算法判断，以做出最佳的政策决定。 在政策领域，很难想象世界是由算法实际做出决定的;

algorithmic judgment to make the best possible policy decisions. In the domain of policy, it is hard to imagine moving to a world in which the algorithms actually make the decisions; we expect that they will instead be used as decision aids.	我们希望它们将被用作决策辅助工具。
For algorithms to add value, we need people to actually use them; that is, to pay attention to them in at least some cases. It is often claimed that in order for people to be willing to use an algorithm, they need to be able to really understand how it works. Maybe. But how many of us know how our cars work, or our iPhones, or pace-makers? How many of us would trade performance for understandability in our own lives by, say, giving up our current automobile with its mystifying internal combustion engine for Fred Flintstone's car?	对于增加价值的算法，我们需要人们实际使用它们; 也就是说，要注意他们至少在某些情况下。 人们经常声称，为了让人们愿意使用算法，他们需要能够真正理解它是如何工作的。 也许。 但我们有多少人知道我们的汽车是如何工作的，还是我们的iPhone或者是步伐制定者？ 我们中有多少人会那 放弃我们现有的汽车，为Fred Flintstone的汽车配备神秘的内燃机（的做法）来换得我们对性能可理解性。
The flip side is that policymakers need to know when they should override the algorithm. For people to know when to override, they need to understand their comparative advantage over the algorithm – and vice versa. The algorithm can look at millions of cases from the past and tell us what happens, on average. But often it's only the human who can see the extenuating circumstance in a given case, since it may be based on factors not captured in the data on which the algorithm was trained.	另一方面，政策制定者需要知道何时应该覆盖算法。对于需要知道何时覆盖的人来说，他们需要了解他们在算法上的比较优势-反之亦然。 该算法可以查看过去的数百万个案例并告诉我们平均发生的事情。 但往往只有人，才能看到在特定情况下情有可原的情况，因为它可能基于未捕获的因素，但这些因素并未包含在训练算法的数据中。
As with any new task, people will be bad at this in the beginning. While they should get better over time, there would be great social value in understanding more about how to accelerate this learning curve.	与任何新任务一样，人们会对此不满意在一开始的时候。 虽然他们应该随着时间的推移而变得更好，但是更多地了解如何加速这种学习曲线会有很大的社会价值。
Pair caution with hope	**谨慎对待希望**
A time traveler going back to the dawn of the 20 century would arrive with dire warnings. One invention was about to do a great deal of harm. It would become one of the biggest causes of death—and for some age groups the biggest cause of death. It would exacerbate inequalities, because those who could afford it would be able to access more jobs and live more comfortably. It	一个回到20世纪初的时间旅行者会带着可怕的警告到达。一项发明即将造成很大的伤害。 它将成为最大的死亡原因之一- 并且对于某些年龄组而言是死亡的最大原因。 它会加剧不平等，因为那些能够负担得起的人将能够获得更多的工作和生活更舒适。 它会改变我们生活的星球的面貌，影响实体的景观，污染环境和促进气候变化。

would change the face of the planet we live on, affecting the physical landscape, polluting the environment and contributing to climate change. The time traveler does not want these warnings to create a hasty panic that completely prevents the development of automobile transportation. Instead, she wants these warnings to help people skip ahead a few steps and follow a safer path: to focus on inventions that make cars less dangerous, to build cities that allow for easy public transport, and to focus on low emissions vehicles. A time traveler from the future talking to us today may arrive with similar warnings about machine learning and encourage a similar approach. She might encourage the spread of machine learning to help solve the most challenging social problems in order to improve the lives of many. She would also remind us to be mindful, and to wear our seatbelts.	时间旅行者不希望这些警告完全产生仓促的恐，这将完全阻止汽车运输的发展。 相反，她想要这些警告帮助人们跳过几步，走更安全的道路：专注于那些使汽车危险减小发明，建造便于公共交通的城市，并专注于低排放车辆。 关于机器学习，从未来过来的时间旅行者会跟我们带来类似的警告并鼓励采用类似的方法。她可能会鼓励机器学习的传播，以帮助解决最具挑战性的社会问题，以改善许多人的生活。她还会提醒我们注意并系好安全带。