“Good” isn’t good enough 论文阅读报告

1. 文献信息

文章题目：“Good” isn’t good enough

作者：Ben Green (Harvard University)

发表途径：AI for social good

发表时间：2019 年

2. 问题意义

计算机科学家们普遍具有想使社会变得更“好”的愿望，并在为此不断努力。但在社会不断变好的过程中，该领域却缺乏对复杂的社会积极变化动态评价的方法，使得这一善意的举动存在了一些问题。首先，计算机科学缺乏关于“好”到底意味着什么的强有力的理论和论述。因此，该领域通常采用一种狭隘的方法来处理政治问题，这包括了对什么社会条件是可取的提出模糊(几乎是不断重复的)的主张。其次，计算机科学缺乏对技术干预与社会影响之间关系的清晰定义。因此，促进社会变“好”的科学家们往往想当然地认为，以技术为中心的渐进式改革是社会进步的适当策略。因此，从实质平等和反压迫的角度来看，这些想让社会变“好”的行为，是否是“好”的，就变成了不确定的事。为了更好地真正做好事，计算机科学家必须反身性地评估他们的规范性承诺，考虑技术干预的长期影响，评估算法干预与选择性改革，并且不再将技术优先于其他形式的知识。

3. 思路方法

作者首先提出，“好”并没有得到普遍的认同。

作者指出目前计算机领域似乎是基于“眼见为实”的方法，依赖于犯罪=坏事、贫穷=坏事等粗略代替来定义好坏。而实际上，即使在计算机领域之外，何为社会变好也没有准确的定义。其中，金融教育网站的定义是“以尽可能多的方式使尽可能多的人受益的东西，如清洁的空气、清洁的水、医疗保健”。以这种原则为基础会导致行动出现很多矛盾。如以加强警察的问责和促进非惩罚性的替代监禁和使用数据预测和分类犯罪，以帮助警察调查这两种相反的使社会变得更好的行为。我们首先应该明确，如果仅仅以行动定义一切，那行动便将毫无意义。事实上，我们应该承认，对好的定义应该存在不同争论的声音，然而目前最缺乏的，恰恰就是这些争辩。

作者举出南加州大学的社会人工智能中心(CAIS)的计算机科学项目来证明自己的观点。该项目由分析反动组织转变为分析犯罪街头帮派，产生了带有偏见的数据。这些科学家实际上已经在影响整个社会权力、地位和权利分配的规范性立场。这恰恰证明了，如果该领域不公开地反思构成计算机科学基本方面的假设和价值观——例如确定研究问题、提出解决方案和定义什么是“好”——那么占主导地位的群体的假设和价值观就会胜出。那些声称要使社

会变好而不与社会和政治背景相关联的项目，很可能会出现一种以使社会变好为目的的压迫行为。

接下来，作者论述了渐进主义的“好”会导致长期危害。

尽管促进社会变好的努力可能是富有成效的，但计算机科学迄今尚未发展出一套严密的方法来考虑算法干预与长期社会影响之间的关系。该领域理所当然地认为，即使机器学习不能为社会问题提供完美的解决方案，它仍然可以通过改善社会的许多方面来为“好”做出贡献。事实上，一些计算机科学家强调这些即时的改进胜过长期的考虑。例如，他们认为“我们不应该让完美成为好东西的敌人”。这一立场的假设是，因为我们都同意犯罪、贫困、歧视等等都是问题，所以我们应该赞扬任何减轻这些问题的努力。这种生产技术改革的取向把“完美”视为不现实的乌托邦，因为它不可能实现，所以不值得阐明或辩论。

然而，作者认为，不考虑长期影响而追求社会利益可能会导致巨大的危害。换句话说，理想化的完美和变得更好之间的割裂是错误的:只有通过辩论和完善我们对完美社会的想象条件，我们才能设想和评估潜在的增加的好处。评估增加的商品和长期的社会变化之间的关系是一项重要的任务，因为并不是所有的发生的改变都是平等的，或者推动社会沿着同样的道路前进。我们必须区分“改革主义改革”和“非改革主义改革”。这两种改革的构想方式截然不同，追求其中一种改革可能导致截然不同的社会和政治结果。计算机科学家提出的解决方案几乎完全是改革主义的改革。算法的标准逻辑是基于准确性和效率，倾向于在现有系统的参数范围内接受和工作，以促进其目标的实现。因此，通常提出计算机科学干预是为了提高系统的性能，而不是实质性地改变它。虽然这些类型的改革在适当的条件下是有价值的，但这种改革主义的精神不足以确定和追求许多社会和政治机构所必需的更大的变革(甚至可能有助于巩固和使现状合法化)。当以这种方式考虑改革时，只有最窄的变化参数是可能的和允许的。从这个意义上讲，该领域目前追求改良、递增的策略就像一个贪心算法:在每一点上，策略都是在局部附近立即做出改善的现状。但是，尽管贪婪策略对于简单的问题是有用的，但在复杂的搜索空间中却不可靠。我们可能很快就会找到一个局部最大值，但却会被困在那里，远离更好的解决方案的广阔领域。

作者对刑事司法系统中变“好”改革的危险做了案例研究。美国的刑事司法系统是一个计算机科学家们不断努力改善的领域，它例证了改革主义者思维方式的局限性。大多数技术上的贡献都是基于现有的犯罪和安全逻辑。即使这些改革带来了渐进式的改善，这些改革也往往会强化和再现刑事司法系统的结构性种族暴力。一部分人们设想一个包括审前拘留在内的公正世界，认为审前拘留的问题不在于它本身不好，而在于它的决定不好，因此，我们应该纠正挑选人进行审前拘留的方法。与此同时，另一方则设想一个没有审前羁押的公正世界，因此，我们应该完全废除这种做法。我们看到，有关风险评估的辩论与诸如公平和准确性等技术问题或关于完美的实用考虑关系不大，而是取决于有关刑事司法系统应如何结构的规范性问题。只有清楚地表达我们想象中的完美，我们才能认识到这两种渐进式改革之间的潜在紧张关系，更不用说就选择哪一种进行适当的辩论了。

4. 实验结论

如果计算机科学要为创造一个更美好的社会做出有成效的贡献，它就必须发展出一套严密的方法，定义什么是好，以及如何在相互竞争的变量中进行权衡。这首先需要一个算法实践的政治方向。计算机科学家应该更明确地考虑并阐明他们工作背后的规范，而不是只强调“好”。第二，计算机科学需要一种实践，这种实践必须与短期和长期的技术干预和社会影响之间的关系密切相关。这就需要参考几代社会思想家和实践人士就如何真正实现积极的社会变革而形成和辩论的经验教训。这样的推理可以帮助计算机科学家考虑算法在改善社会方面的作用，当算法与社会和政治世界相互作用时如何产生意想不到的影响，以及当其他形式的政治行动需要与算法结合或替代算法时如何产生意想不到的影响。第三，该领域必须评估算法对替代改革的干预，而不是假设算法为每个问题都提供了合适的解决方案。这也意味着要找到新的算法干预方式，更好地与社会变革的长期路径相匹配。这些需求中的许多都借鉴了其他领域的专业知识，这就需要一种以跨学科为核心的算法实践，不再将技术放在最高优先级上。

5. 启发思考

选择这篇论文时，首先是它独特的题目吸引了我。当进一步阅读时，这篇论文带给了我很大的震撼。我一直以为，在实现算法时，只需要不断的改善参数、调整算法，不断让结果趋近于完美就是技术人员唯一该做的事。正如作者所说，仅仅把技术放在最高优先级的位置上是很多人的思想误区。通过这篇论文，我第一次将科学技术与社会学联系到一起，并意识到二者之间的关系十分紧密。这篇论文给我提供了全新的思路，也提升了我对于科学技术以及社会变革的认识。“如果仅仅以行动定义一切，行动就将全无意义”给我留下了十分深刻的印象。这种辩证的、全局性的探讨打开了全新的世界。这使我第一次意识到，一味的追求完美，追求改善，而没有明确的原则和标准，最终会使这种改善走向一种新的极端，而背离一切的初衷。以变好为目的的行动，如果只专注于行动本身，这种变好的行为就变成了一种新的压迫。追求完美的算法的过程中，是需要全面的去看待问题，需要辩证的思考，需要多学科的综合学习，需要不同声音的争辩，而不是以无止境的趋近最完美的数值作为唯一的追求。这是我第一次阅读相关类型的论文，它给我带来了全新的感受和浓厚的兴趣。