

负责任 技术执行手册

助力企业优化技术决策的
工具与实践指南

2021年8月

© 2021 Thoughtworks



本手册由 Thoughtworks 创建，相关内容已获得《知识共享署名——相同方式共享4.0公共许可协议国际版》授权。

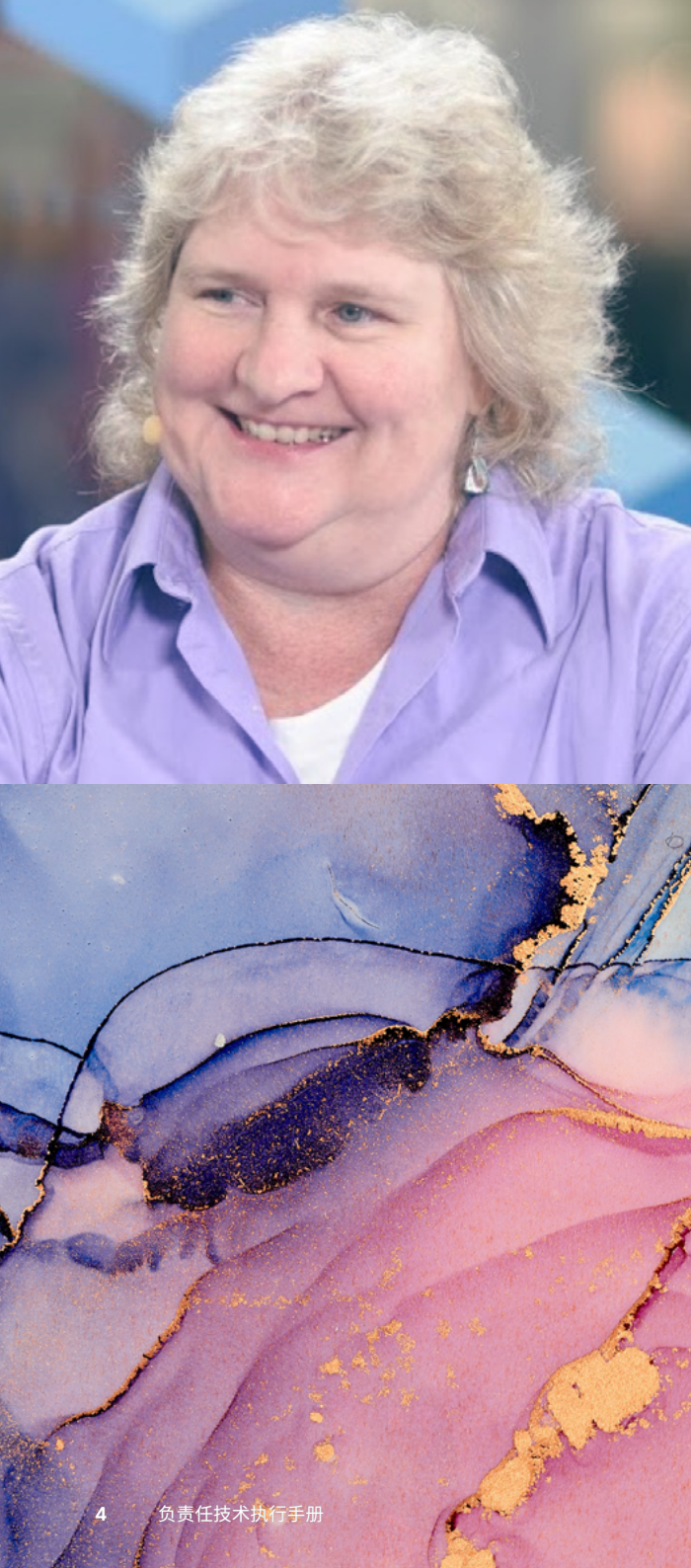
该手册包含了工具创作者创作且拥有的工具和图片，使用已获授权。其他组织或个人如欲使用本手册中的任何工具和图片，请直接联系其创作者/权利所有者，以获取相应的授权确认。

Thoughtworks 并未对本手册中的任何工具、方法或框架给予背书或推荐。



目录

前言	4
我们的使命	5
什么是负责任技术?	6
负责任技术: 为何它这么重要?	8
负责任技术工具与实践	10
开发生命周期中的负责任技术	11
选择适当的工具	12
方法与工具索引	13
起步工具	14
负责任技术的深入工具	15
Agile Threat Modeling (敏捷威胁建模)	16
Consequence Scanning (结果扫描)	18
Data Ethics Canvas (数据道德画布)	20
Ethical Explorer (数据道德探测器)	22
Ethical OS	24
Failure Modes & Effects Analysis (故障模式和效应分析)	26
Flourishing Business Model Canvas (繁荣商业模式画布)	28
InterpretML	30
Materiality Matrix Assessment (重要性矩阵评估)	32
Responsible Strategy (负责任策略)	34
Tarot Cards of Tech (技术塔罗牌)	36
Unintended Consequences (意外结果)	38
继续负责任技术之旅	40
前景光明的方法	41
值得探索的其他新兴方法	46
延伸阅读	47
鸣谢	47
继续对话	48



前言

技术的触角正延伸到更敏感、更复杂的领域，从信贷决策到医疗诊断甚至到量刑等不一而足。它影响着我们与亲朋好友和自身的日常互动；影响着我们作为员工、客户和公众的生活。作为技术专家，我们必须主动担负起相应责任——努力扩大技术的积极影响，减轻负面影响。

本手册所包含的工具和技术试图帮助团队找到更具包容性、更能识别偏见、更加透明且能减轻负面意外后果的战略。这些方法可以帮助技术专家们从多个维度审视他们的产品和技术选择，提高发现有害或意外后果的可能性。

尽管我们的行业没有类似于希波克拉底誓言或者其他行业的职业许可标准，我们仍要考虑，对于技术的用户以及在技术使用过程中受到影响却不被关注的那些人，我们应当承担哪些责任。本手册旨在从源头建立起负责任的技术使用制度。

Rebecca Parsons
Thoughtworks 首席技术官

我们的使命

帮助组织规避日渐数字化的世界中存在的伦理风险

企业越来越意识到他们所发明的技术可能产生巨大的道德（以及相关的名誉和经济）后果。越来越多的企业在开发产品时都试图谨慎行事——但技术是一个宽广、复杂、不断变化的领域，人们通常不知该从何着手去衡量技术的潜在影响。

本手册重点介绍了一些组织和团队可以切实采用的前沿工具。这些工具有助于分析技术和开发工艺，识别潜在伦理挑战，推动更负责任技术实践的进程。需要强调的是，这些参与式实用技术应普遍适用于任何人——而非仅仅技术专家——所有人都可以使用、受益并增加对技术与社会互动方式的理解。

本手册如同一本食谱：我们一开始会使用食谱，但也从其他地方获取灵感，最终找到最适合自身情况的方法并在此基础上发展自己的技能。我们需要从小处做起，时常练习，进而建立起个人、团队、组织专属的文化与能力体系。这个过程和日常锻炼实现增肌是一个道理。

我们非常欢迎更多的建议、工具推荐、使用提示，以共同促成这一领域的进步。

Alexander Steinhart
编辑兼手册负责人代表



什么是负责任技术？

让技术和企业行为符合社会与个人的利益

负责任技术是一种让技术和企业行为更符合社会与个人利益的工作方式。它积极思考并探索技术的价值、不可测的后果和负面的影响，主动管理、减缓和降低技术的风险与伤害。我们思考伦理、个人与人类繁荣、社会结构、包容与公平等概念，竭力追求让技术造福而非剥削大众。

几乎无处不在的数字技术推动了这一进程。数字技术处于我们个人生活和社会的核心。随着人们的日常活动越来越依赖技术，技术所造成的不可测或恶意的后果也随之增加。

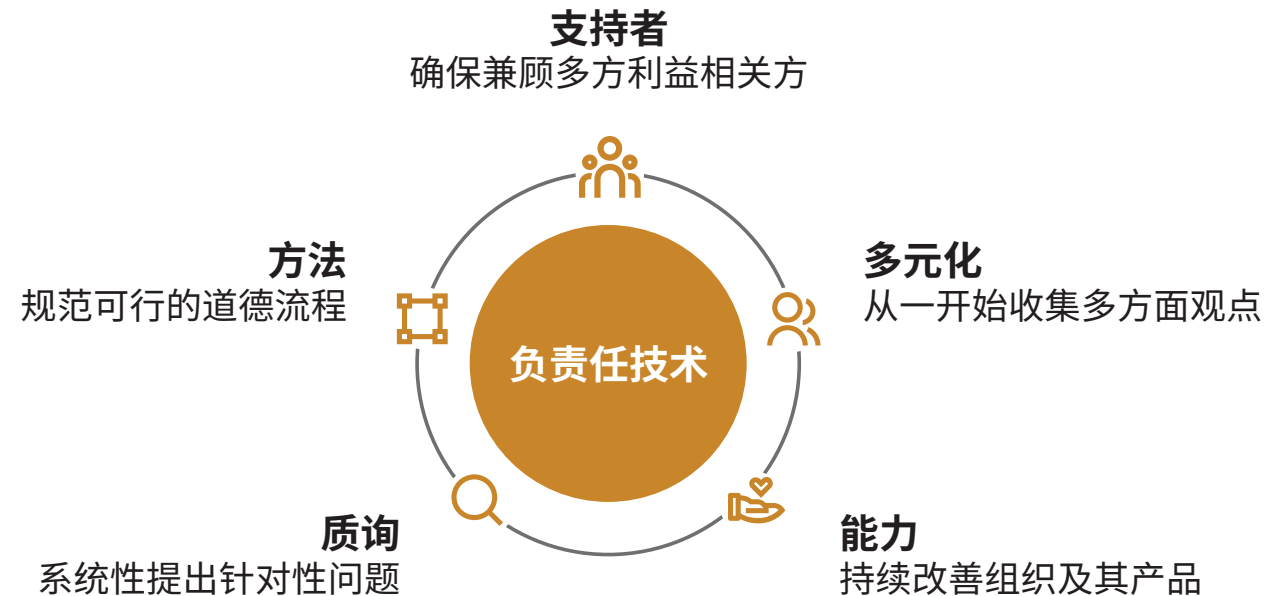
自动化程度提升——没有人类介入但带来真实后果的行为——以及技术系统日益复杂的叠加效果，导致我们对技术的可预测性越来越低的同时受到技术的影响却越来越大。因此，人们迫切渴望成熟的风险管理和安全的实践。

与此同时，“技术”和“社会”正变得密不可分，技术与人类行为相互影响。因此，我们急切渴望用价值明确的实践来打造社区以及我们所期望的未来。



实施负责任技术

如果想要更加深入且广泛的实施负责任技术、保持高度自信的创新能力、获取更多来自利益相关方、消费者和更广泛公众的信任——应该从以下五个方面着手：质询、方法、支持者、多元化、能力。



本手册所含负责任技术工具将提供方法和调查帮助，规范流程，并系统性提出针对性问题。尽管如此，您在具体应用时仍需考虑多方利益相关方（支持者）和多方面观点（多元化），包括伦理专家的观点。最后，还要预留一部分资源，持续构建并改进这一文化和相关能力。

负责任技术： 为何它这么重要？



需求 (或影响) 和好处

- 百分之九十七的消费者希望品牌在创新时采用合乎伦理的技术；([信源](#))
- 百分之七十的消费者表示，他们会更乐于购买在道德方面不断提升的品牌。([信源](#))

挑战与后果

- 仅百分之十九的人认为企业在设计技术产品和服务的时候考虑了他们的最佳利益；([信源](#))
- 百分之八十六的消费者可能断绝与不道德使用数据的企业之间的联系。([信源](#))

不开展负责任、合乎道德的思考,可能给我们的形象、健康、商业、经济、政治系统¹及社会凝聚力带来毁灭性后果。作为技术专家,审慎选择我们使用的工具有助于避免出现这些极端后果。

请在我们的 [《社会影响报告: 技术处于社会中心地位》](#) 中查找更多负责任技术的内容。

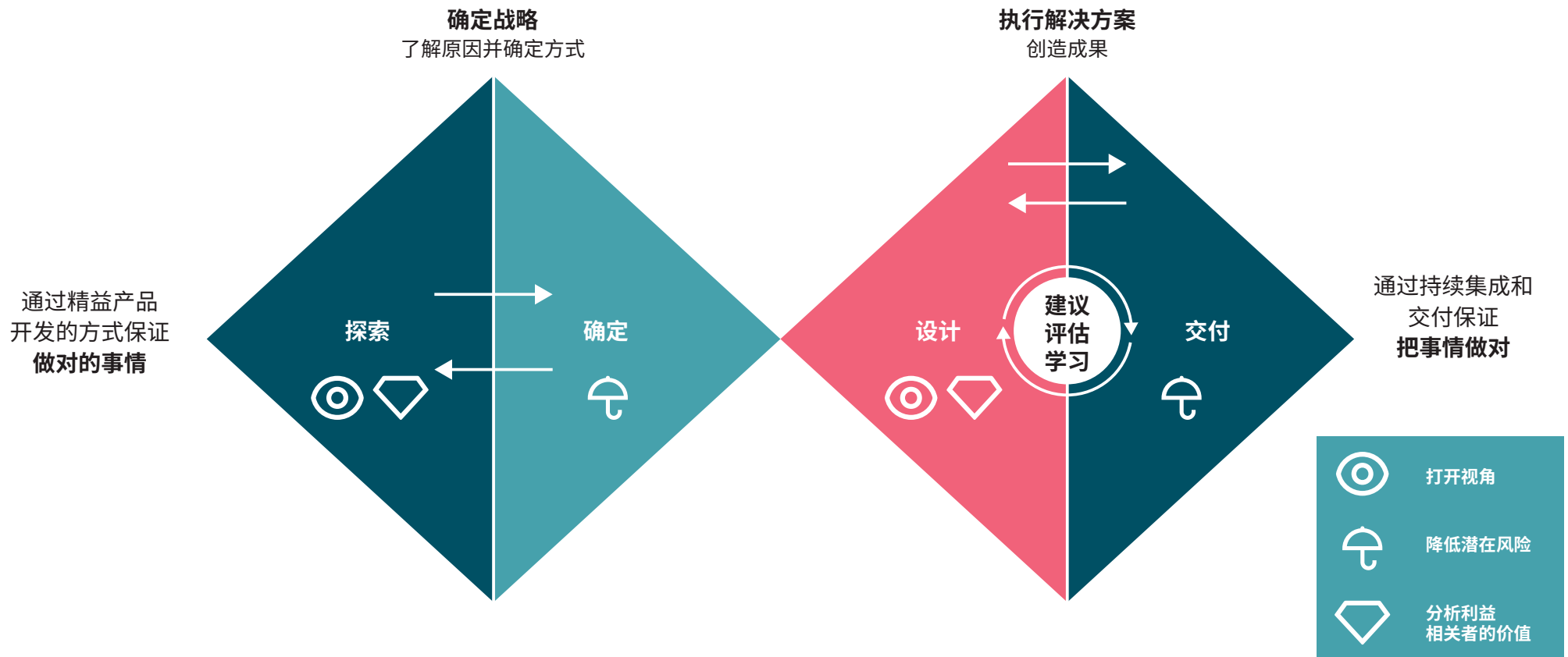
负责任技术 工具与实践



开发生命周期中的负责任技术

您可以在产品研发周期的所有阶段整合负责任技术。在发散阶段, 特别适合选择打开视角和分析利益相关者价值的工具。收敛阶段, 则着重关注减轻风险的工具。

欲了解如何进一步实施负责任技术, 见 [第七页](#)。



选择适当的工具

这是一把可以帮助您判断哪些工具最适合团队的情况和目标的钥匙。



打开视角

征集不同观点, 广泛思考
潜在后果和成果。



降低潜在风险

在道德挑战和弱点演变成更大问题之前,
发现并解决它们。



分析利益相关者的价值

确保技术的设计符合预期服务对象的需求
且支持他们的价值。

方法与工具索引

按英文字母顺序排序



敏捷
威胁建模



page 16



结果扫描



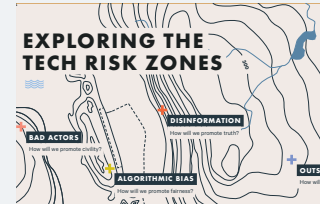
page 18



数据
道德画布



page 20



道德探测器



page 22



Ethical OS



page 24

re	S	P	D	RPN
1 reminds a user of a deceased relative	3	2	7	42
1 recommends a car is lent to someone who does not drive	2	5	3	30

故障模式
和效应分析



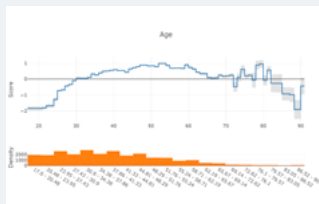
page 26



繁荣
商业模型画布



page 28



InterpretML



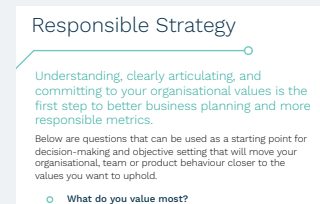
page 30



重要性
矩阵评估



page 32



负责任策略



page 34



技术塔罗牌



page 36



意外结果



page 38

For additional approaches and methods, see pages [41-47](#).

起步工具

不确定从哪里开始?可以试着用这三种工具中的任何一种开始探索,我们认为它们适合起步阶段使用,适用于一项产品的创始或开端。



Ethical Explorer — 道德探测器

Ethical Explorer 为个人、团队和利益相关者提供他们可采用的提升方式与相关指南（大多基于 Ethical OS），帮助您识别旗下产品可能涉及到的伦理问题与风险。

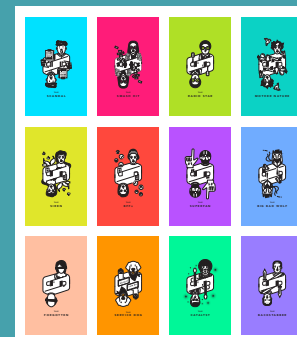
[了解更多 >](#)



Consequence Scanning — 结果扫描

Consequence Scanning 帮助团队正确解决问题，同时赋能他们采取措施做出相应改变。其结构类似于团队执行团队回顾等敏捷流程，非常适用于这些团队。

[了解更多 >](#)



Tarot Cards of Tech — 技术塔罗牌

Tarot Cards of Tech 是一套设计精美的提示卡，可以帮助您打开视角、收集不同观点。在构思过程中，它们能够与 HMW 方法相互配合，提供很好的思路引导。

[了解更多 >](#)

负责任技术的 深入工具



Agile Threat Modeling

— 敏捷威胁建模

来源: Thoughtworks

它是什么?

威胁建模是一种基于风险的，通过识别威胁和缓解措施来设计安全软件的方法。不同于创建和维护一个详尽的“威胁模型”，我们鼓励团队小步、高频地执行威胁建模。

何时使用?

建议团队养成在每次迭代时执行威胁建模的习惯，促使软件交付团队集思广益，在实现之前找到真正的安全威胁。

涉及哪些人?

整个交付团队，技术和非技术人员都应当引入其中。比如，让业务分析师和产品经理提高风险意识，这也是团队成员从他人身上获取洞察与能力提升的绝佳机会，但这并不是必需的。

[< 返回方法列表](#)



降低潜在风险



每次活动1-2小时



跨专业团队

我们的观点

- 一次不要尝试太多，选择可实现的目标。
- 用 STRIDE 提示卡集思广益，从其他维度拓宽提示卡（比如本手册高亮显示的卡片）。
- Make sure the actions go straight into the backlog

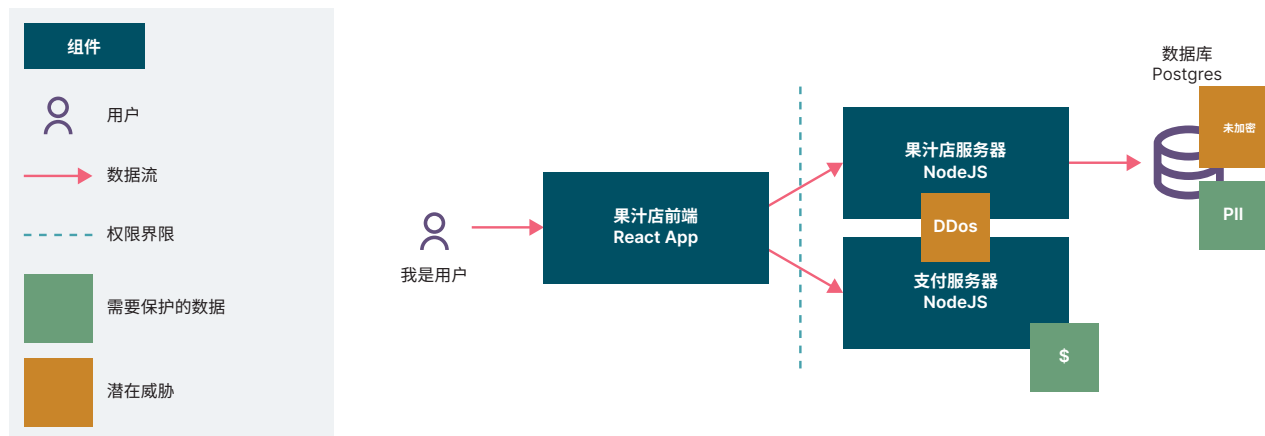
进一步探索：

威胁建模完整指南 (无需下载)

[Jim Gumbley文章, 研讨会启动指南](#)

STRIDE提示卡可以从Thoughtworks英国、德国和印度办公室获取。

Agile Threat Modeling (continued) — 敏捷威胁建模 (续)



工作原则

1. 经常召集交付团队开会，探讨威胁建模，提出以下问题：
 - 我们创建的是什么？
 - 可能出现哪些错误？
 - 我们如何应对？
2. 通过绘制一幅“低保真”技术图，进行阐释和探索，其中包括：
 - 相关组件和用户；
 - 数据的不同流向；
 - 做好网络和界限标记；
 - 对企业具有重要价值的资产。
3. 使用STRIDE提示卡，集思广益发现威胁：
 - 找到攻击、突破或击败软件某个特定点的方法本质上就是威胁建模；
 - 捕捉告示板、便利贴或在线辅助工具（即MURAL）上识别的风险。

4. 共享有助于优化的知识，进行优化和修复：
 - 团队应该就最危险的威胁进行记点投票；
 - 识别最危险的威胁并记录下来；
 - 采取措施，将降低风险措施纳入待办事项。
5. 完成结束：
 - 向团队成员分配任务；
 - 确定修复实施的时间；
 - 首次修复完成后重复一次。

[< 返回方法列表](#)

Consequence Scanning

— 结果扫描

来源: Doteveryone (现开放数据研究所)

它是什么?

这项工具提供了一些工作坊相关的活动方式与材料,帮助团队思考与某项技术或服务相关的预期和意外结果。这项工具适用于所有领域的所有团队。参与者角色应尽可能跨功能:从技术、设计、产品、业务到用户自身。

它也提供一个框架,用以发现在工作坊中无法被发现的问题。

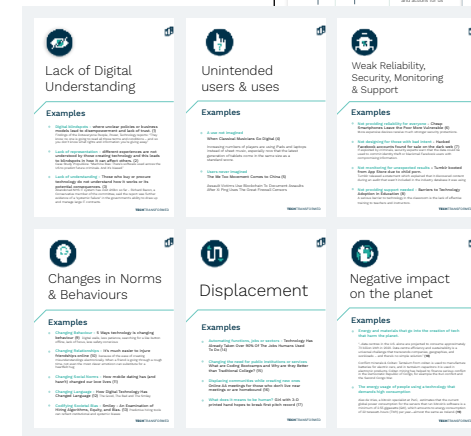
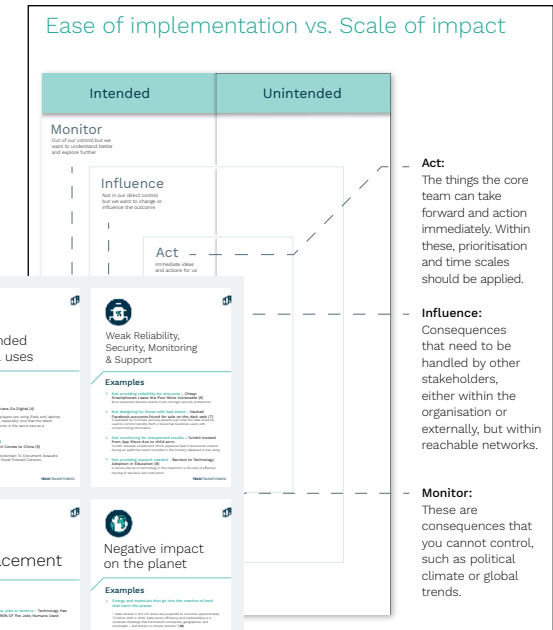
一次安排妥当的会议应涵盖以下问题:

- 本产品或功能的预期和意外结果是什么?
- 我们关注的积极结果是什么?
- 我们希望减少的后果是什么?

何时使用?

其设计是为了配合迭代式敏捷过程,所以适合多个阶段使用,包括初步构想、路线图规划和引入新功能。

- 使用简单,具有非常全面要素的使用说明;
- 一次会议应预留1.5-2小时,否则参会者会觉得很匆忙。



连接愿景、构思、实施、路线图、史诗功能划分和项目收尾(便于销售跟进),使之成为习惯。它可以像回顾一样成为团队日常活动的一部分。这一工具很有代表性,应用起来也非常合理,以行动为导向,可以为团队赋能。

涉及哪些人?

核心团队、用户支持者、技术或业务专家、高级产品赞助商、合作方。

[< 返回方法列表](#)



降低潜在风险



打开视角



每次活动1-2小时



跨专业团队

我们的观点

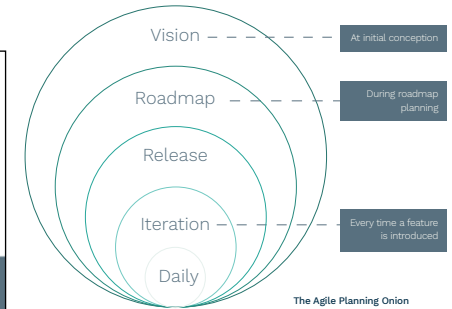
这些提示可能有一些抽象，很难与具体业务关联在一起。使用者需要花时间收集、识别相关度最高的信息，才能有所帮助。您也可以使用其他工具的提示，比如 Tarot Cards of Tech、Ethical Explorer 等。无论如何，这是一款堪称惊艳的敏捷工具，与回顾和检查等敏捷流程配合的很好。

进一步探索：

<https://www.doteveryone.org.uk/project/consequence-scanning/>

Doteveryone原创（现由开放数据研究院保存）《署名——非商用相同方式共享 4.0公共许可协议国际版》（CC BY-NC-SA 4.0）

Consequence Scanning (continued) — 结果扫描（续）



工作原则

该方法分为两个阶段：

第1阶段，构思：借助不同问题和提示来展露预期结果以及意外结果（简介、独立分析时间、排序、对结论分组和讨论归纳）；

第2阶段，行动：将成果进行分类，帮助制定一套行动计划（行动排序、记点投票、讨论和结束）。

思考：

- 如果远程操作，则划分小组，共享墙面或白板，选择可即时协作的工具；
- 针对您的关注点收集不同的观点，比如：持有强烈反对意见的人、有这种“功能”经验的人或者从未遇到这种情况的人。这有助于提高分析的丰富性。

还包括：

- 贵公司的愿景、使命和价值观；
- 支持反思和决策的其他工具或文件。

[< 返回方法列表](#)

Data Ethics Canvas

— 数据道德画布

来源:开放数据研究所

它是什么?

一系列提问,旨在促进讨论,帮助团队思考如何收集、共享或使用数据。

目的:

- 在“惧怕收集数据”和“囤积一切数据”之间找到平衡;
- 思考对人们的影响;
- 鼓励信任,实现透明。

结果:

产生具有优先级排序的、且有明确责任分配的行动计划。

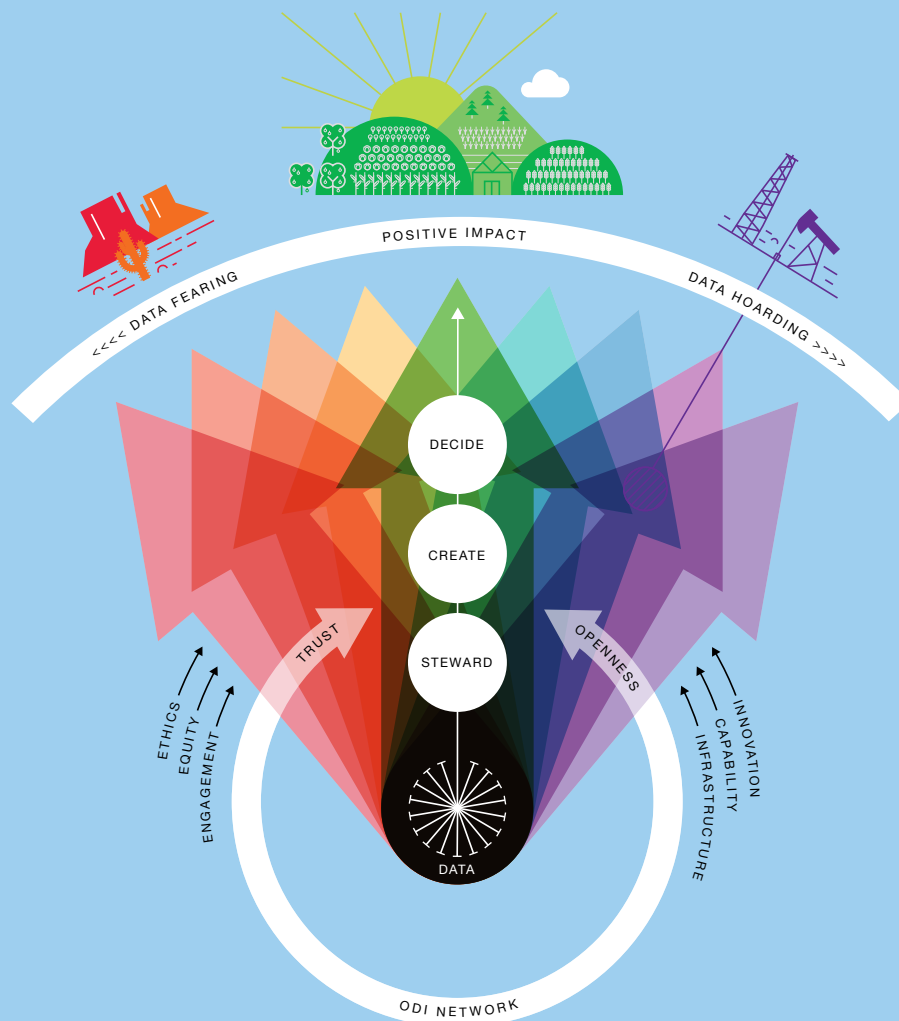
为何用它?

帮助识别数据使用中潜在的道德问题,找到可行的预防措施。

何时使用? 最理想的情况是在新项目开始的时候,但在项目开始之后也是有用的。

与谁使用? 这项活动最好能够引入足够多样的角色群体。

[< 返回方法列表](#)



降低潜在风险



打开视角



每次活动1-2小时



跨专业团队

我们的观点

- 将笔记和行动措施文档化；
- 与多个角色的人一同完成一组画布；
- 注意小组的权力变动可能会影响练习；
- 定期回顾结果和行动措施。

进一步探索：

<https://theodi.org/wp-content/uploads/2019/07/ODI-Data-Ethics-Canvas-2019-05.pdf>

由开放数据研究所制作，经《知识共享署名——相同方式共享4.0公共许可协议英国国际版》授权。

Data Ethics Canvas (continued) — 数据道德画布 (续)



工作原则

您可以从画布的任何地方开始，按照任何顺序完成章节。

1. 在每个章节中，谨慎思考您的答案。有助于在便签纸上做记录，并贴到对应的章节。
2. 如果可以，与他人分享讨论您的想法；

3. 完成画布后，列出后续步骤，确保每一项措施都有人负责；

4. 如果可以，更广泛地分享讨论的内容。

[< 返回方法列表](#)

Ethical Explorer — 道德探测器

来源：Omidyar Network

它是什么？

这是一套卡片，帮助您识别软件产品可能存在的各种风险。用探索家作比喻，风险区域就如同地形，设想您的产品或服务当前所处的情景，并搜寻前方可能出现的风险。它与 Ethical OS 类似，所以对两种方法都应该尝试，并进行对比。

- 用比喻和提示帮您识别产品可能存在的风险并找到可以采取的预防措施；
- 可以在任何领域使用。

为何用它？

提出能够引发“风险识别以及应对措施”方面讨论的问题，鼓励进行常规风险之外的思考。

何时使用？ 可以在产品交付的任何阶段使用，但越早识别风险，越容易降低风险。

涉及哪些人？ 项目早期阶段涉及的重要利益相关者，但也可以在任何阶段与跨专业团队（两人以上）共同使用。



[< 返回方法列表](#)



降低潜在风险



打开视角



每次活动1-2小时



跨专业团队

我们的观点

这套卡片是 Ethical OS 很好的迭代，就如何使用它给予了更多指导。另外，该卡片使用的比喻和隐喻非常恰当，鼓励使用者就此话题展开探索。有关如何使用这些卡片的各种举例，不管是个体行动还是团队活动，都做了很好的引入。

Omidyar Network 与 Artefact 出品的工具

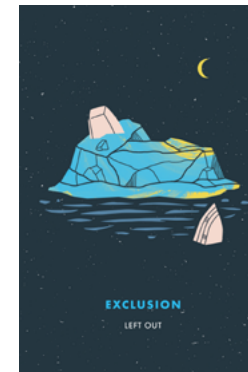
进一步探索：

<https://ethicalexplorer.org/download/>

Ethical Explorer (continued) — 道德探测器（续）

工作原则

1. 开始先做热身练习，让每个参与者单独浏览一下卡片，然后与小组分享一个潜在风险可能横跨一个或多个风险区域的产品例子。
2. 让参与者思考产品或服务可能落在哪个风险区域中，选择其中的三个。
3. 挑选出小组选中的最高优先级区域，根据卡片背面的提示集思广益。
4. 讨论结果，向小组提问“随着时间推移它会怎么影响产品/服务？最有害的是什么？”
5. 通过记点投票决定如何优化选项：您可能会问“哪个风险区域的有害影响最大？哪个更容易改变？”
6. 分组讨论缓解这些问题的方法以及这些措施的具体执行步骤。



[< 返回方法列表](#)

Ethical OS

来源：奥米迪亚网络公司的技术和社会解决方案

它是什么？

Ethical OS 是形成一个公司及其产品和服务的策略、流程和价值观的工具。它与 Ethical Explorer 相似，所以两种方法都应该尝试，并进行比较。

它包括：

- 八大风险区域检查表；
- 让风险区域更加显著的十四个未来场景；
- 在研讨会背景下使用风险区域和场景的说明；
- 符合伦理道德技术的战略措施；
- 自定义用例。

为何用它？

Ethical OS 是在探索或咨询阶段让主要利益相关者参与其中的最佳战略工具。

但它也可以用来：

- 在任何阶段（即便是销售阶段）为项目做准备，明确关键问题，为不理想的未来创建真实的场景；
- 查阅现有解决方案，了解需要注意的风险区域。



[< 返回方法列表](#)



降低潜在风险



打开视角



每次活动1-2小时



跨专业团队



Senior stakeholders

我们的观点

- 将检查表植入在产品设计要求中；
- 将优先问题提交至领导和顾问，寻求他们的建议；
- 将问题发布到论坛或者能被高度关注的地方；
- 收集资源，寻找周围精通您最高风险区域问题的专家来提供意见；
- 只要开始开发或者计划改变一项新产品或服务的规格，应重新对照检查表。

进一步探索：

www.ethicalOS.org

EthicalOS is an initiative by the **Omidyar Network's Tech and Society Solutions Lab**

<https://omidyar.com/responsible-technology-2/>

Ethical OS (continued) – Ethical OS (续)

Remember: Don't get caught up in whether a scenario is likely or even possible. Just pick one and go with it.

“When something of massive consequence happens that no one predicted, we often say it was simply unimaginable. But the truth is, nothing is impossible to imagine. When we say something was unimaginable, usually it means we failed to point our imagination in the right direction.”

—Jane McGonigal, Institute for the Future

Addiction & the Dopamine Economy | **(∞)**

Scenario 2: Are you ready for a world in which ...

Conversation bots have been trained to imitate specific people, using data sets collected from public social media posts. These bots are deployed across social media networks, email, and text messages in super-targeted, super-personalized propaganda campaigns. They are highly effective in changing opinions and driving action, as personalized messages appearing to be from real friends, family, and favorite celebrities are more influential than advertisements.

© 2018 Institute for the Future and Omidyar Network (CC BY-NC-SA 4.0) SR-2005 | ethicalOS.org

工作原则

1. 首先，选择你正研究的一项技术、产品或功能。
2. 识别风险区域，找到与团队和所选技术最相关的各区域问题。先用**检查表**找到要向团队提出的问题。
3. 通过这些问题启动团队或组织的内部对话。
4. 问题：“我们如何开始纠偏，或降低这些风险？”
5. 写下行动措施的优先级。这一阶段还需考虑以下问题：
 - 最艰难或者最大的威胁；
 - 有助于缓解威胁的策略；
 - 推进之前所需要的帮助。

[< 返回方法列表](#)

Failure Modes & Effects Analysis

— 故障模式和效应分析

来源：美国质量学会（ASQ）

它是什么？

FMEA 是设计者和工程师用来设计、创建和运行更有弹性、更可靠系统的风险方法。它采用定性描述和定量评分计算潜在故障的故障致命度和风险优先级（安全、隐私、可用性、性能、道德），用于分析和预防风险，帮助在多个领域识别和预防故障。

为何用它？

人类不太善于应对潜在的灾难，对于思考我们设计的东西会怎样发生故障也不太有想法。FMEA 程序可以通过明确所有潜在“不好”的结果和潜在故障提供帮助，帮我们明确各种故障可能发生的方式，以达到预防的目的。

何时使用？ 开发产品或服务概念以及需要做出改变或决策的时候。初始运行需要4-8小时，更新需要大约2小时。难度中等到困难（需要具有经验的引导者）。

涉及哪些人？ 整个团队（技术、设计、BA、QA、中小企业）；三到六个人最理想，但不超过六个人。

Failure	S	P	D	RPN
System reminds a user of a now-deceased relative	3	2	7	42
System recommends a car advertisement to someone who does not drive	2	5	3	30
System raises interest rates on user based on automated credit scoring algorithm	5	3	1	15

[< 返回方法列表](#)



降低潜在风险



每次活动1-2小时



跨专业团队

我们的观点

这是源于工程行业的一种普遍方法，遵循了许多其他工具（如Agile Threat Modelling）的一般原则。

虽然核心原则是优势，但需要根据自身程序和背景进一步调整以适应工作。工具和模板都有，但在软件环境下工作需要更新语言。

进一步探索：

<https://asq.org/quality-resources/fmea> 提供了可下载资源、详细指南和示例。

美国武装部队原创，后由美国国家航空航天局（NASA）制作使用——该方法现由美国质量学会支持。

Failure Modes & Effects Analysis (continued) — 故障模式和效应分析（续）

Failure	S	P	D	RPN
System reminds a user of a now-deceased relative	3	2	7	42
System recommends a car advertisement to someone who does not drive	2	5	3	30
System raises interest rates on user based on automated credit scoring algorithm	5	3	1	15

工作原则

1. 定义潜在故障模式——描述系统会以何种方式发生故障。
2. 了解故障的后果——即故障对用户的影响。
3. 找到可能的原因和级别——故障发生的可能原理、故障的可能性和可检测性或者原因。
4. 评估风险危急性和优先级——找到适当的设计和控制措施减轻或预防故障的发生。
 - 可用Mural一类的电子画板画布和Google Docs工作表完成；
 - 部分了解系统的整体风险状况；
 - 通过提示和示例促进过程中每一步的思考，这很重要。

[< 返回方法列表](#)

Flourishing Business Model Canvas — 繁荣商业模式画布

来源：Flourishing Enterprise Institute

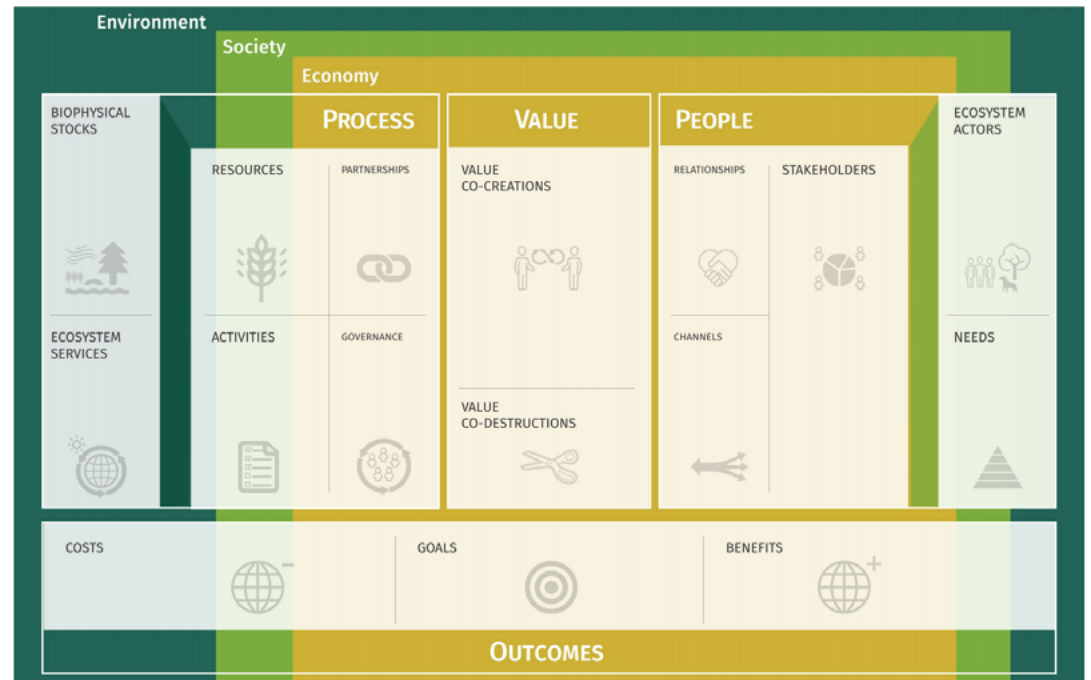
它是什么？

Flourishing Business Model Canvas 是 Business Model Canvas 的延伸。它拓宽了一项业务、产品或服务的探讨内容，不仅包括可行性、生存力和需求性，还考虑了经济、社会和环境问题。

它通过十六个问题，促进了各利益相关者和价值观之间的协调——就关键战略决策达成一致，并创建了一种通用语言。

为何用它？

- 在评估新的业务、产品或服务机遇时，了解更广泛的经济、社会和环境影响；



- 如果您比较适应使用Business Model Canvases，那么这种画布稍微复杂一些，因为涉及范围更广泛。

何时？ 项目或产品生命周期早期阶段。

涉及到谁？ 参与人的数量高度取决于复杂程度和可能受影响的潜在利益相关者。应该咨询国内外利益相关者，比如客户、社区和环保组织。

[< 返回方法列表](#)



分析利益相关者的
价值



打开视角



每次活动1-2小时



Senior
stakeholders

我们的观点

Flourishing Business Model Canvas 提供了一种有用的工具帮助全面思考潜在经济、社会和环境影响。发现并邀请所有受影响的利益相关者验证自己的思想可能是这种方法最困难的一部分。

进一步探索：

<http://www.flourishingbusiness.org/the-toolkit-flourishing-business-canvas/>

Flourishing Business Model Canvas (continued) — 繁荣商业模式画布 (续)



来自 Holonomics 的 Simon Robinson 和 Maria Moraes Robinson 正与巴西的能源部门高管制定新的繁荣业务模式。

工作原则

该工具的使用方法与 Business Model Canvas 类似，它不仅帮助人们把事做好，还要把事做对。

- 在两到三个小时的合作研讨会中即可初步完成画布。然后通过精益实验和利益相关者持续验证不断演变；
- 它作为一个关键工具快速制作原型并迭代业务模式；
- 它允许公司从自身的核心价值观出发，努力成为成功的企业家，然后让他们设计测试潜在模式。

[这篇博客帖子](#)介绍了 Dear Green Coffee 的 Flourishing Business Model 实践案例研究。

Flourishing Business Canvas 由“繁荣企业家创新工具项目”团队成员创建。© Antony Upward / Edward James Consulting Ltd., 2014年所有权利保留。www.FlourishingBusiness.org 图片使用已经授权。如欲使用画布，请联系他们获得授权。研讨会照片所有权 © Holonomics, www.holonomics.co。

[< 返回方法列表](#)

InterpretML

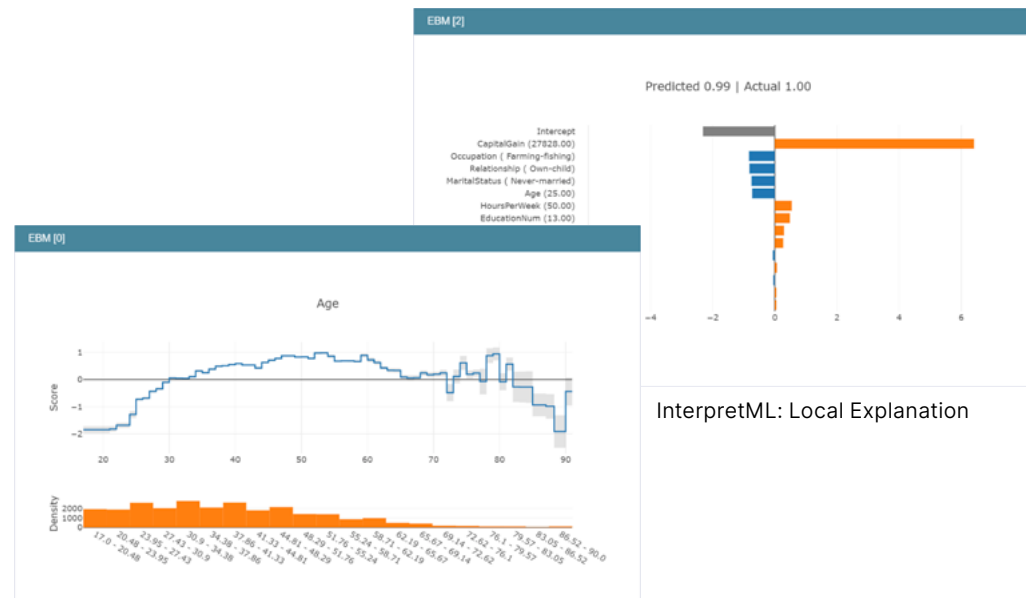
来源： Microsoft Research

它是什么？

InterpretML 是一个让利益相关者了解和解释机器学习模型预测结果的开源库。其目的是打消利益相关者必须在模型预测能力与可解释性之间作出取舍的固有印象。除了创建全新、透明的模型，它还提供了先进方法以阐明已有模型的行为（这些方法优于 [LIME](#) 和 [SHAP](#)）。

InterpretML 的能力：

- 创建可解释的透明的“玻璃盒”式模型，这些模型甚至允许领域专家直接编辑；
- 用户还可以引入自己的黑盒模型，开展事后可解释性分析；
- 解释模型的整体（全域）行为和单个（本地）预测行为；
- 下钻至预测子集，开展反事实（假设分析）实验；
- 与补充框架一起使用，以达到公平AI目标，比如 [Fairlearn](#) 和 [AIF360](#)。



InterpretML: Explainable Boosting Machine (EBM) analysis

为何用它？

对开发者和数据科学家而言，机器学习可解释性对调试模型、确保负责任预测甚至防止对抗攻击至关重要。对管理层和监管者而言，这些工具可以帮助阐明产品可信度，支持合规性审计。

何时使用？ 对高风险应用非常重要（比如涉及安全或者会影响社会不公的应用）

涉及哪些人？ 有Python知识和数据背景的团队成員。

[< 返回方法列表](#)



降低潜在风险



每次活动1-2小时



跨专业团队

我们的观点

InterpretML 让负责任 AI 更加便捷，利益相关者无需掌握先进的统计学知识便可以理解，同时又保持了技术的严密性。我们特别希望看到现代玻璃盒模型走到前列，希望 AI 性能和可解释性不再完全对立。

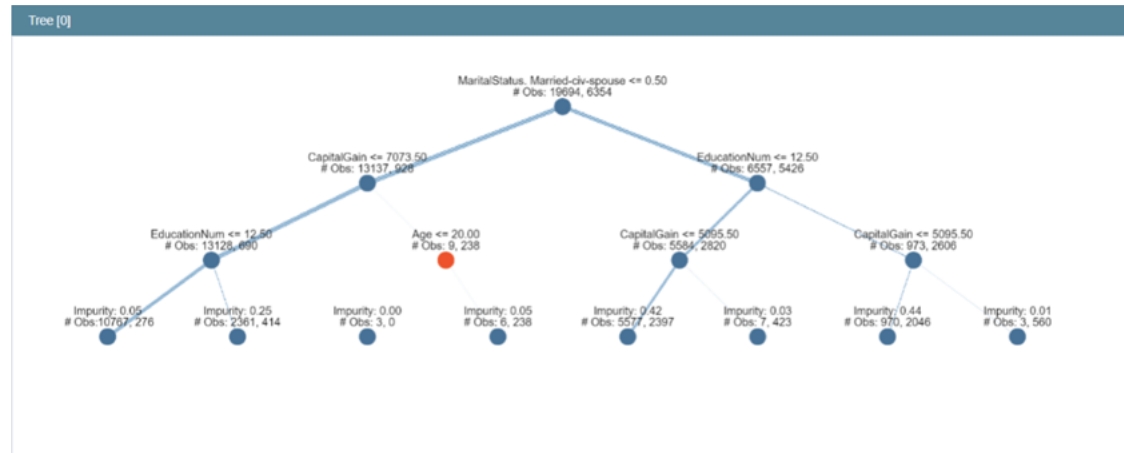
由于可解释性可能关系到公平问题，我们建议增加分析的深度，甚至在适用的情况下整合公平框架，如 [Fairlearn](#)，从算法上缓解不公平性。

互补工具：[Fairlearn](#)、[IBM AI Fairness 360](#)、[PRESC](#)、[Google What-If Tool](#)

进一步探索：

<https://interpret.ml>

Interpret ML (continued) — Interpret ML (续)



InterpretML：调查一项功能的影响（这个例子中的“资本收益”）用于预测个人收入。每条规则自动呈现数据样本的比例。

工作原则

1. 遵循 [InterpretML](#) 的入门指南。
2. 考虑从 InterpretML 的玻璃盒模型，如可解释增强机（EBM）开始。部分基准已经证明，EBM 甚至可以媲美颇受欢迎的高性能模型，比如 XGBoost！
3. 或者，引入自己的任意模型，运行事后黑盒解释器（比如SHAP Kernel解释器）分析模型行为。
4. 分析本地和全域解释。
5. 用 InterpretML 面板交互性开展假设调查，编辑数据点，让新结果和解释形象化。

InterpretML 受到微软公司 [MIT License](#) 授权。

分析和图片基于 Dua, D.和Graff, C.的成人收入数据库（2019年）。UCI机器学习库 [<http://archive.ics.uci.edu/ml>]加州尔湾：加州大学信息与计算机科学学院

[< 返回方法列表](#)

Materiality Matrix Assessment

— 重要性矩阵评估

它是什么？

重要性是发现对公司和利益相关者至关重要的社会、经济、环境问题的一个原则。重要性矩阵有助于具像化重要性评估结果和见解。

组织矩阵的一种常见方法就是绘制经济、社会和环境主题图表，思考它们对外部利益相关者的重要性以及对企业的影响（内部利益相关者）。

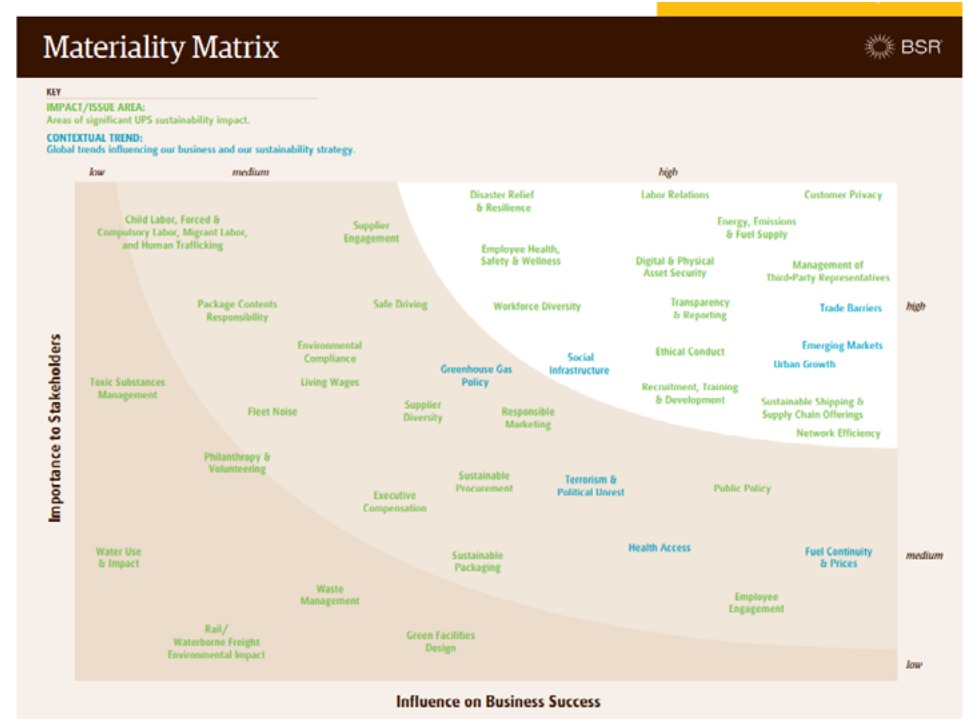
为何用它？

这项练习旨在引导利益相关者了解特定经济、环境和社会问题对他们的重要性。然后用所获得的见解指导其策略和交流，从而讲述更有意义的可持续性故事。

何时使用？ 在新项目或产品开始时使用非常有价值，也可以用于反复评估，比如季度或年度评估。

涉及哪些人？ 利益相关者和内部跨专业团队。

帮助全面概括企业在经济、社会和环境方面的影响。



UPS Materiality Matrix

[< 返回方法列表](#)



分析利益相关者的
价值



Up to 4 hours
initially



跨专业团队



Senior
stakeholders

我们的观点

当公司能从内外利益相关者都获取见解时，该工具是最有价值的，因此需要制定一份名单，包含能给公司可持续性战略提供有意义见解的利益相关者团体，兼顾内部联系人（行政领导、区域经理和雇员）和外部联系人（贸易协会、关键客户、非政府组织），有助于评估价值链上的各种观点。

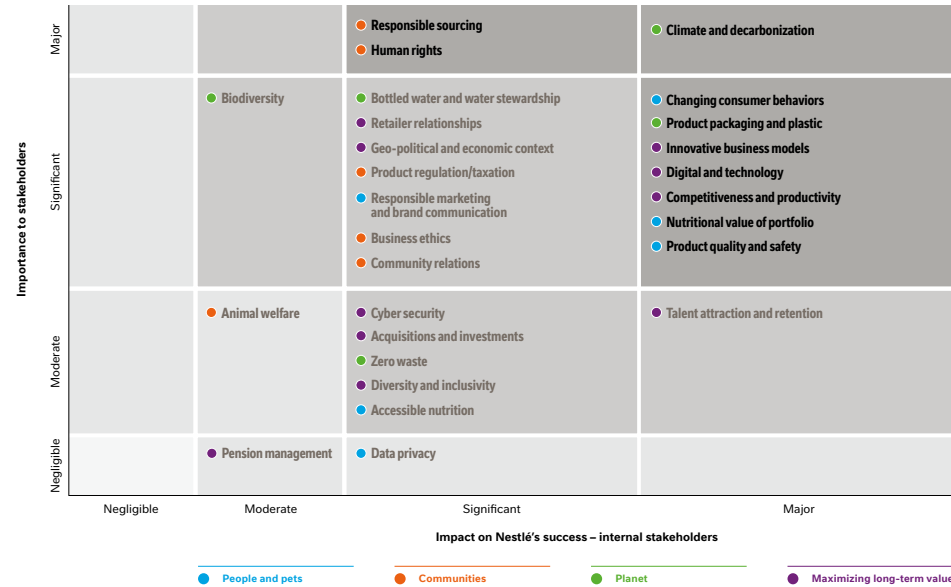
进一步探索：

可用详细指南

<https://us.anteagroup.com/en-us/blog/7-basic-steps-conducting-successful-materiality-assessment>

https://www.stern.nyu.edu/sites/default/files/assets/documents/ternCSBSustainabilityMateriality_2019_0.pdf

Materiality Matrix Assessment (continued) — 重要性矩阵评估 (续)



Nestlé 重要性矩阵

工作原则

1. 明确能就公司可持续性战略提供有意义见解的内外部利益相关者。
2. 确定并优化想用的可持续发展指标。比如经济（利润、收益、公司营业额）、社会（劳工统计、社区影响）和环境（耗水量和废弃物管理）。
3. 设计重要性调查。利益相关者用数值表评定各可持续发展指标的重要性和影响力（如1-5或者1-10）。
4. 启动调查，开始征集见解。
5. 基于调查响应结果创建重要性矩阵。
6. 分享您的结果并将您的见解付诸行动。

[< 返回方法列表](#)

Responsible Strategy — 负责任策略

来源: Doteveryone (现开放数据研究院)

它是什么?

负责任策略包含了七个问题用于作为制定决策和确立目标的起点,这七个问题促使组织、团队或产品行为更加接近想要秉持的价值观。通过将价值观转换成原则和可评估的KPI,反复确认意外结果,有助于推动价值观成为现实。

这是一份有用的调查问卷,可以为 Lean Value Tree 或 OKR 等其他方法开拓基本构建单元。

为何用它?

了解、清晰表达并致力于组织的价值观是更好地进行企业规划和制定更负责任指标的第一步。这些原则和KPI为组织内人员提供了他们对所造产品的集体责任指南。

何时使用? 项目开始阶段 (比如加入一般愿景和使命会议)

涉及哪些人? 团队和管理层。

作为做出决策和制定目标的起点,以更贴近组织和/或项目希望秉持的价值观



分析利益相关者的
价值



每次活动1-2小时



跨专业团队



Senior
stakeholders

我们的观点

- 深入了解组织的根本原则和价值观；
- 充当进一步推进的开拓者；
- 确定能够反映从愿景到价值观进程的KPI；
- 最好尽早使用，最好有关键利益相关者的参与；
- 请注意，开放式问题的格式应该规整扼要。

进一步探索：

问题表：<https://www.tech-transformed.com/download/1008/>

请检查您的价值观和指标

<https://www.tech-transformed.com/strategy/>

Doteveryone制作（现由开放数据研究院保存）。

署名——非商用相同方式共享4.0公共许可协议国际版（CC BY-NC-SA 4.0）

Responsible Strategy (continued) — 负责任策略 (续)

工作原则

1. 最有可能融入初始阶段；
2. 实施一对一面谈或者给团队布置“家庭作业”；
3. 在团队会议中收集并归纳结论；
4. 以透明的方式推动进一步探索，并对其进行文档化记录，以支持决策；
5. 在项目中途进行审查。

Responsible Digital Metrics

Your vision and organisation. for the collected the organisat the world.

Principles provide in an organisation overall vision. For ultimately transla are accomplishin

- 1 Does your v articulate th
- 2 What are th purpose, or decisions? (values, what
- 3 From what values into

- How do y (what are
- What wo product d
- How will successf

Responsible Strategy

Understanding, clearly articulating, and committing to your organisational values is the first step to better business planning and more responsible metrics.

Below are questions that can be used as a starting point for decision-making and objective setting that will move your organisational, team or product behaviour closer to the values you want to uphold.

- **What do you value most?**

Start internally, and pragmatically. This might be your craft, your ability to innovate, your colleagues and collaborators, the investment, or the opportunity you have.

And externally, what do you work the hardest to maintain? This might be trust, sustainability, reliability, or security. Or it might be reputation, profit margin, or pipeline.

How do these translate into values for your organisation that you can tell stories around?
- **What behaviours do you value?**

What are the adjectives that describe the ways you or your team achieve things when you're at your best?

What do you look for and champion in others?
- **Who do you value most?**

Who is most important to you — is it your shareholders, investors or stakeholders, your workforce, your customers or your clients?

If it's your customers or clients, then be specific about which segment: is it the most vulnerable, the most engaged, or the most profitable?

When you need to compromise, who do you feel comfortable benefiting, and who do you feel comfortable disadvantaging?
- **What value do you create?**

If you know what you value, and who you value, then you need to ask what benefit you provide to those people.

TECHTRANSFORMED

[< 返回方法列表](#)

Tarot Cards of Tech — 技术塔罗牌

来源: Artefact

它是什么?

技术塔罗牌是一种头脑风暴练习, 以鼓励创作者思考技术和产品可能产生的真实结果, 这些成果既有意外的结果也有可以积极改变的机遇。每张卡片包含一个冲突性提问, 涵盖类似以下主题:

用途。您的产品将怎样影响文化习惯和关系? 您的产品将以哪些您没有考虑过的方式 (或者被哪些您没有考虑过的人) 使用?

公平和使用权。哪些用户被排除在外? 他们的体验会是怎样的?

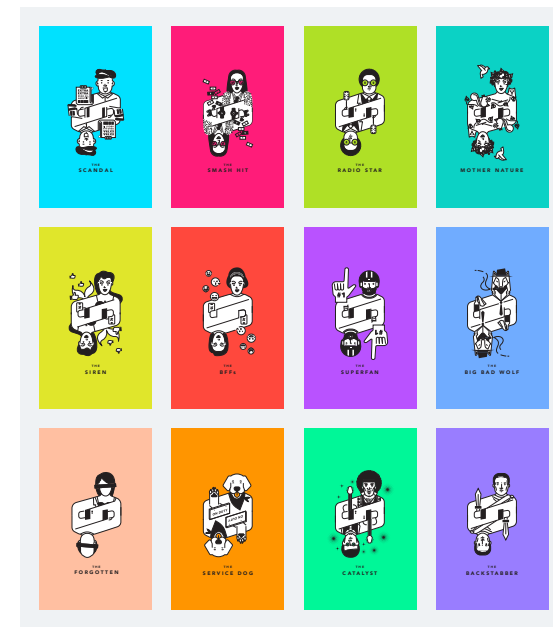
Scale and disruption. What would using your product 'too much' look like? What would you change if your client was Mother Nature?

为何用它?

拓展您对您的产品/技术不同维度的思考, 推动团队做出有意识的决策。

何时使用? 最好的时机是产品构思的早期阶段 (比如创始时), 也可以在“初始”和“概要”阶段使用。

涉及哪些人? 整个团队——包括利益相关者。



[< 返回方法列表](#)



打开视角



每次活动1-2小时



跨专业团队

我们的观点

这完全可能是一次简短的会议，卡片中包含各种各样的冲突性提问。有时卡片帮助大家更清晰、更充分的参与讨论。需要在使用时保证“最坏情况”和“良好机遇”之间的平衡——尽管“暗算他人者”等“邪恶”卡片似乎能激发更多思考和讨论。

进一步探索：

（卡片可以获取。）

<http://tarotcardsoftech.artefactgroup.com/>

Artefact制作。

Tarot Cards of Tech (continued) — 技术塔罗牌 (续)

工作原则

这是一种头脑风暴练习，旨在激发讨论，让团队从更宏观、更长远的角度思考他们制作的产品。

1. 下载并打印卡片；
2. 分成小组——按主题将卡片分发给小组；
3. 选择一张卡片，读给小组——卡片包含一个冲突性提问，帮助团队从不同的视角看待产品；
4. 讨论卡片上的问题，做好笔记（以便回顾）；
5. 与更广泛的团队分享，记录讨论得出的一切行动措施。

您可以考虑：

- 分成小组，进行回顾讨论；
- 根据讨论结果制定行动计划；
- 挑选最相关的卡片进行深入研究。

如何应用？

- 在设计或者角色创作过程中；
- 对于策略或新点子，特别适用于设计冲刺阶段；
- 集思广益发现负面结果；
- 跳出来思考/与众不同；
- 留意盲点；
- 发现极端例子/行为



Workshop at the Humane Tech Berlin Group

[< 返回方法列表](#)

Unintended Consequences — 意外结果

来源：Hyper Island

它是什么？

在这一工作坊中，团队测试新技术是否会产生意外结果，并在此基础上激发潜在的商业机遇。工作坊从建立“意外结果”的观念出发，驱动新点子产生。它看起来超出了对新技术的一般理解，激励参与者去发现意外潜力以及加以利用。例如，Instagram是为智能手机创建的虚拟相册，但通过用户之间的联结，它有潜力产生更深的意义，比如参与悲伤和治愈的过程，正如近期[这位丧偶作家展示的例子](#)。

为何用它？

本工具用于演示当产品的建立没有考虑责任机制时会发生什么事。本工具将说服您的团队思考伦理道德问题，建立紧迫感，考虑解决方案的潜在问题，让伦理道德成为组织的优先考虑点。

[< 返回方法列表](#)



打开视角



每次活动1-2小时



跨专业团队

我们的观点

- 举行这个工作坊，揭示您的新商业计划或点子可能存在的一些后果；
- 为此，分两步走，先大略列出新技术或业务，然后向团队解释您的解决方案，包括已经发现的挑战；
- 工作坊结束后，会呈现很多关于技术或业务如何给社会造成潜在伤害的模式。参与者一般都审慎思考过他们的新商业模式，但没有考虑这些新商业模型的意外结果。由于激发这种讨论的工具不多，特别是与商业模式有关的工具，所以这是一个很大的优势。

进一步探索：

<http://toolbox.hyperisland.com/unintended-consequences>

Hyper Island制作。

Unintended Consequences (continued) — 意外结果（续）

工作原则

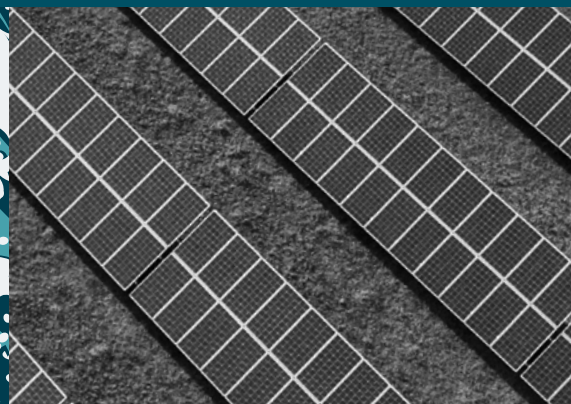
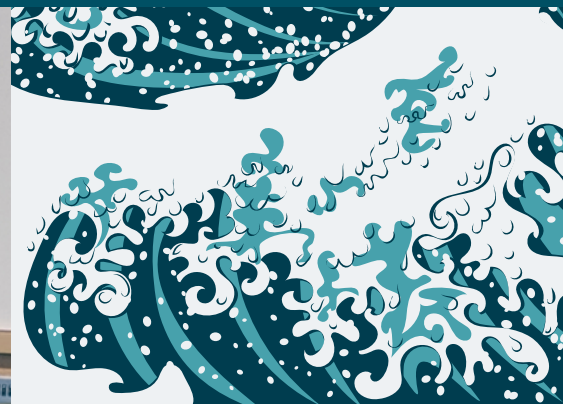
1. 将团队分成三至五人小组。分享几例与技术意外结果相关的例子，进而阐述活动的目的。
2. 给小组五分钟时间，制定一份清单，至少包含五种新技术或者业务，比如Netflix、Instagram、Google Maps等。为了增加挑战性，让参与者只列出除了自己以外人们广泛使用的技术。把它们写在便利贴上（每张便利贴写一项）。
3. 接着，让参与者集思广益：对每一种新技术或业务，他们应推测出至少三种意外结果。比如，Netflix本来只是一种简单便捷的流媒体服务，但它会导致“刷剧”等新行为。鼓励参与者的创意思考，强调问题没有正确答案。让他们将答案写在便利贴上（每张便利贴写一项）。10分钟后，让每个小组分享他们的想法。
4. 每个小组对每种意外结果给出新的商业模式建议。比如Netflix可以创造一种新的商业模式，用户只能每天免费看一集电视，若要刷剧他们必须加价。构思时，应该欢迎所有主意，因为没有所谓的坏主意。12分钟后，小组集合并分享他们最有趣的新商业模式。
5. 让参与者仔细思考最意外且最可行的新商业模式，并给出理由。也可以思考这次练习教会他们有关构思和新商业模型的哪些内容。

[< 返回方法列表](#)

继续负责任 技术之旅

我们没有深入探讨的方法还有很多。

下面我们举一些例子。



前景光明的方法

对立思考

大部分伦理道德实践都将导致竞争价值之间的矛盾。这些对立（有时候也称为棘手问题、复杂问题、窘境、悖论、矛盾）不能被当作传统“问题”来解决，即弄清哪个选项是“解决方案”，然后加以实施。相反，需采用“不同的”方式进行管理。九十年代巴里·约翰逊广泛推广的基于“对立管理”的方法和观念模式可以解决这个问题。

进一步探索：

对立思考 * Neesa Sweet对立映射简介

<http://braidedrivergroup.com/wp-content/uploads/2011/04/Polarity-Thinking-Overview.pdf>

Barry Johnson (2014). *Polarity Management: Identifying and Managing Unsolvable Problems*. HRD Press.

对立思考实践——利用差异化进行创新，Coreative。

<https://www.wearecocreative.com/pot-of-gold>

前景光明的方法

Humane Design Guide（人性化设计指南）

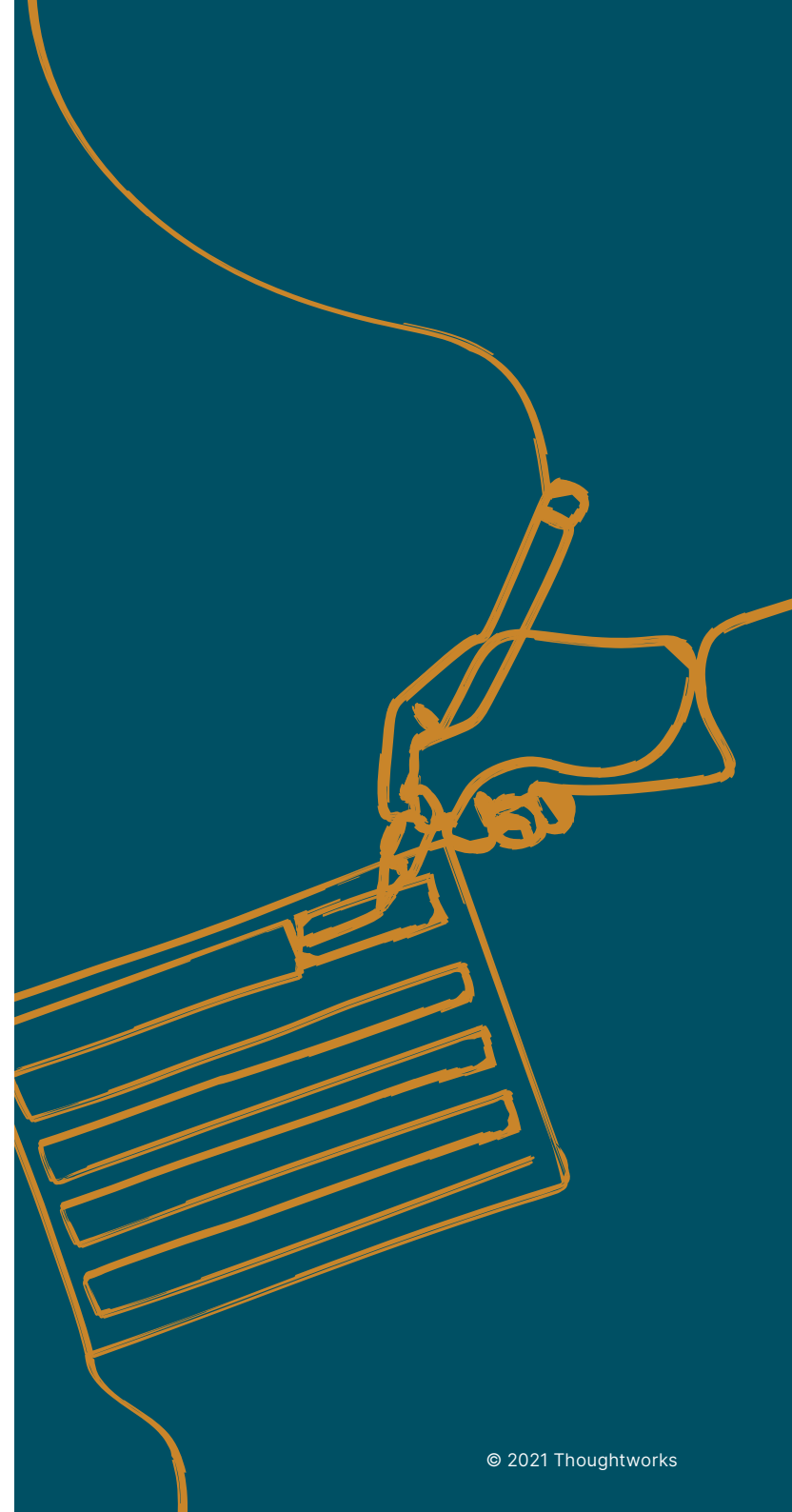
Center for Humane Technology的《人性化设计指南》探索了不剥削大众而是真正支持、服务大众的技术构建方式。这个工作表可以帮你采取合理的步骤来设计一个更加人性化的产品，认识到对人性更深层次的投入将会带来的更多收益。这个指南将引领你评估六种人类感官，帮你和团队采取行动。工作表分两页，一页用来评估产品的当前状态，另一页用于构思。

当然，这份工作表并非解决所有问题的灵丹妙药，人性化设计是一个重要领域，可以让人更清楚地意识到我们想要以什么方式定义人类与技术的关系。

进一步探索：

<https://humanetech.com/designguide/>

Humane by Design (<https://humanebydesign.com/>) 提供了一个资源，在这一指南中，通过那些专注用户福祉的模式来设计合乎伦理的人性化数字产品。



前景光明的方法

思辨设计/激发试验

对设计者来说，这个方法特别有趣。通常来说，设计和设计者是问题解决者，他们的关注点是让一个产品更好、更漂亮，或让一个流程更高效，但是在这个方法论里，设计和设计者不是来解决问题，而是来提出问题的。

在这里，设计的目的是想象一个完全不一样的未来，创造一个可体验的场景，帮助阐明（或展现）道德、伦理、政治和审美问题。这个想法黑人男性会喜欢吗？白人女性呢？孩子或坐轮椅的人士呢？一个非常富有的家庭或低收入家庭会怎么看待它？这个创新还会感动哪些人？

进一步探索：

<https://uxplanet.org/can-speculative-design-make-ux-better-design-trend-4-4-ce8d13148e5d>

前景光明的方法

分组

通常来说，分组是定性研究中的一个方法，这种方法可以减少研究者的假设前提、偏见、假定、理论或之前经验带来的潜在有害影响。

这种方法也可以用在开发流程中。设计者和开发者可以像在架构决策记录（ADR）中那样，任何时候，只要意识到他们自己的偏见可能影响到决策，就会把其记录下来。这个清单随后可以用于广泛讨论，也可以用来检查和改变算法数据。

Explore further:

https://www.researchgate.net/publication/257924681_Bracketing_in_Qualitative_Research



前景光明的方法

参与式设计法

过往的案例大部分都是工具，不过参与式设计更像是一个方法和设计理念。这个方法的核心要点是让参与者有更高的参与度—这不是简单地问参与者需要什么，在PD法中，用户是内容和流程的所有人。在这种认知里，用户就是专家，我们只是协调员，帮助用户寻找答案，最重要的一点是，用户在产品设计上应该有更多权力。

进一步探索：

https://en.wikipedia.org/wiki/Participatory_design

值得探索的其他新兴方法

这份指南并非最终清单，会有越来越多的方法和工具涌现。如果对负责任技术和符合伦理道德技术感兴趣，可以通过探索下列更多工具和资源获得更深入的了解。

点击链接探索更多

- [AI Blindspot Cards](#) , MIT Media Lab —在规划、构建和部署AI产品时发现盲点
- [AI Ethics Cards](#) , IDEO — 汇集了一系列问题、方法和样例，帮助跨专业团队实现以人为本的理念
- [Bracketing in research](#) — 通过定性研究减少可能干扰研究进程的潜在有害影响
- [Ethics Canvas](#) , ADAPT Centre for Digital Content Technology — 改编于Alex Osterwalder的Business Model Canvas
- [Equity-Centered Community Design Field Guide](#) , Creative Reaction Lab — 重心在社区联合创造，以公平为核心的问题解决流程
- [Future Wheel](#) , Jerome C. Glenn — 对产品的非计划内方向或未预见影响进行头脑风暴
- [Inverted Behaviour Model](#) , Katherine M. Zhou — 反转BJ Fogg的Behaviour Model，设计意想不到的行为
- [Ethical and Rights-Based AI Approaches](#) , BKC for Internet & Society — 概述了各种 AI原理和指导方针
- [Needs & Collateral](#) , Katherine M. Zhou — 帮助用户发现负面附带影响和不想要的副作用
- [Privacy by Design](#) , Ann Cavoukian — 呼吁在整个工程/设计流程的生命周期中都要考虑隐私问题
- [Green Patterns: Helping Users Make More Sustainable Choices](#) by Tim Frick
- [Systemic Design Framework](#) by UK Design Council

延伸阅读

下列在线资源提供了负责任技术和符合伦理道德技术领域的更多有价值洞察，其中许多观点值得我们思考吸纳。

点击链接探索更多

- [Ethical and Rights-Based AI Approaches](#)，BKC for Internet & Society，对Ethical AI 方法进行了整体性概览
- IEEE的 [Ethically Aligned Design: Prioritizing Human Wellbeing with Autonomous and Intelligent Systems](#)
- [Guide to Responsible tech: How to Get Involved & Build a Better Tech Future](#) by All Tech is Human
- [Teaching Responsible Computing Playbook](#)，火狐
- [The Business Case for AI Ethics](#) by All Tech is Human
- [Ethical & Humane-Tech Reading List](#) 概述

鸣谢

这份手册汇集了许多个人、组织和合作伙伴的成果、知识和专长，感谢你们。

此外，我们还衷心感谢与我们反复讨论、测试，帮助我们增加和审核负责任技术工具和相关文档的Thoughtworks同仁。这份手册是名副其实的许多人经验和智慧的结晶，衷心感谢你们！

延伸对话

如果你已经尝试过这些工具中的任何一种，我们很想听听你的感受。请给我们发送电子邮件，或使用表格来建议你认为我们应该在这一手册中补充的其他方法。如果你需要一些帮助，或者有问题需要解答，我们很乐意安排一个沟通，以更有针对性的方式帮助你。

我们的目标是进一步支持和促进这一领域的成功。

通过电子邮件联系我们：

[**RTteam@Thoughtworks.com**](mailto:RTteam@Thoughtworks.com)

使用此 [表格](#) 提交更多方法或工具。