

Responsible Innovation: Ethics, Safety and Technology

(2nd edition)

负责任创新：伦理、安全和技术

——新时代技术发展的风险和伦理问题应对

第 2 版

Joost Groot Kormelink 编著

何良兴 译

【编者（第二版）】

Joost Groot Kormelink

荷兰代尔夫特理工大学, 技术、政策和管理系

【译者】

何良兴

南京财经大学工商管理学院、创新与创业研究中心主任，研究方向为创业管理、负责任创新。主持国家自然科学基金青年项目、中国高等教育学会课题重点项目；在 *Technovation*、*Technological forecasting & social change*、《科学学与科学技术管理》《外国经济与管理》等国内外知名期刊发表学术论文近 20 余篇，出版学术专著 1 部；省级一流课程主要参与人；指导学生多项创新创业训练项目和学科竞赛，并获得多项省赛特等奖、一等奖，以及“最佳指导教师”等荣誉。

【Translator introduction】

Liangxing HE is currently an Associate Professor at the School of Business Administration, Nanjing University of Finance and Economics. He is Director at the Innovation and Entrepreneurship Research Center, Nanjing University of Finance and Economics. He was also a visiting scholar at the University of Sydney. His research interests focus on entrepreneurship, responsible innovation. His work has appeared in both book chapters and journals, such as the *Academy of Management Proceedings*, *Technovation*, *Technological forecasting & social change*, *Technology Analysis and Strategic Management*, etc. He got the best paper award from the 2019 *Academy of Management Annual Meeting*. He has served as a reviewer for some journals and top management associations.



中文版序言

1. 创新何时对社会有益？

创新给社会发展带来诸多好处，但创新本身并非好事。或者创新未能解决重大问题，无法对社会做出有意义的贡献。历史上有很多创新和新技术的相关例子说明了这一点，例如含有 DDT 的杀虫剂、含有石棉的建筑材料（长期吸入石棉会致癌）。

新技术带来了许多伦理问题、争议和未知风险。想想纳米技术、生物技术、人工智能、自动武器、基因组和大数据分析等等。同时，我们确实需要新技术来应对重大社会挑战，如能源短缺、人口老龄化、水资源管理和粮食安全等问题。因此，为创新和技术界定一个恰当且共同的责任观念至关重要，这甚至可以说是人类的责任。对此，我们只需要想想以下问题：

- 人类的创新能否拯救生命？
- 它们是否会创造更多就业机会？
- 它们能否拯救地球，还是只会产生更多的废弃物和污染？
- 它们对用户是否安全，是否可以防止滥用？
- 它们是否尊重人们珍视的价值观和基本人权，如隐私、自由、自主和平等？
如果不尊重，我们该如何使之实现？如果不是我们，那又是谁？如果不是现在，那又是什么时候？

“负责任创新”这一概念最初于 2006 年在荷兰研究理事会（NWO）的“社会负责任创新”项目中提出。2014 年 11 月，这一政策在《关于负责任研究与创新的罗马宣言》中得到认可和推广。如今，已被纳入欧洲联盟（EU）更大的研发议程中。本书旨在深入阐述负责任创新相关内容，并从伦理角度帮助塑造应对全球和地区问题的社会技术解决方案。

2. 章节内容和学习目标

本书内容结构共 5 部分：负责任创新概述；负责任创新的应用伦理；制度与价值观；管理与创新；风险评估与安全。

第一部分对负责任创新做了总体介绍。具体的，在第一章详细阐述了复杂社

会技术系统的现状，并介绍了负责任创新概念，它是发展创新和技术重要且必要的方面。

在总体介绍基础上，第二部分将应用伦理和社会价值观作为创新的起点。具体的，第二章引入了各种思想实验，用来探索缺乏（或混淆）价值与责任所引起的不同困境（如电车问题、“多面手”问题）。当存在多个需要坚持的价值时，每个价值都有其独特的重要性和可取性，但由于时间和资源限制，无法同时满足所有价值，因此会出现道德过载情况。此外，由于潜在价值冲突，人们情绪会变得激动，而情绪反应可以被视为探索这些价值的机会。此外，利用创新满足多个价值，我们可以对此持乐观态度；毕竟，这是创新的真正意义所在。

第三部分将制度作为创新的背景。具体的，第三章讨论了制度——即嵌入或明确的社会习俗与规则，构建了个体和群体间的社会互动方式——如何深刻影响有利的价值，这些价值的体现方式，以及如何保持这些价值。

第四部分详细阐述了不同创新类型、创新实施过程，以及简约创新。首先，第四章重点关注了企业在竞争和机遇背景下如何思考创新。本章阐述了渐进式和激进式创新如何产生，其影响因素有哪些，以及如何管理这些因素促进相关创新。其次，第五章重点介绍了简约创新，这是一种专门针对“金字塔底层”消费者的创新类型。简约创新并不仅仅意味着使用更便宜的技术，而是这些创新对社区生活和居住条件提供了便利。不过，简约创新并非就是负责任创新，它必须符合相关社会标准。第三，第六章描述了一种新方法，即路线图。企业可以使用它来制定负责任创新战略，并将其纳入企业社会责任政策。该路线图始于伦理型领导，并以负责任创新的四大支柱——预见性、反思性、包容性和响应性——为基础。

第五部分阐述了风险评估和安全管理的相关内容。在本部分，第七章和第八章探讨了任意一种技术的重要价值，即安全。为确保技术安全，第七章讨论了如何评估新技术的潜在风险。之所以评估新技术潜在风险，原因之一在于“科林里奇困境”：当技术处于新兴阶段时，人们倾向用自己期望的方式塑造技术发展，但无法知道所有风险。另外，一旦技术嵌入社会，其危险性就变得显而易见，而要改变技术却非常困难。因此，并非所有风险都能预见，总是存在“未知的未知”的可能性。在这种情况下，预防原则可以成为一个好的准则，因为它允许在新技术开发时预先设置保障措施，以尽可能减轻已知风险。

除了理解和识别风险外，还可以量化风险并设计安全方案。对此，第八章介绍了风险分析和安全工程。首先，介绍了一种最常用的风险分析方法，即成本效益分析。其次，还介绍了全面风险分析框架，以及诸如故障树分析、蝴蝶结分析和危险——障碍——目标模型等工具。这些工具有助于从定量和逻辑角度理解风险及其后果。

第六部分论述了与创新相关的价值设计内容。具体在第九章介绍了价值敏感设计（VSD），并将其作为在技术中保留价值并付诸实践的框架。VSD 可以在价值层次矩阵中表示出来，也可以自上而下和自下而上地实现。价值的可视化和明确表示，允许利益相关者以建设性的方式辩论和谈判这些价值。此外，还可以批判性地解构和质疑操作标准：我们重视的价值是否融入了设计之中，或者，这些标准是否实现了我们期望的价值？此外，本书还引入一些相关案例贯穿主要章节，来加强对负责任创新理论知识的理解。

负责任创新和本书的学习目标如下：

- 理解负责任创新概念及其关键伦理维度
- 了解分析新技术风险的各种方法和工具，包括前瞻性和回顾性工具（例如事故原因）
- 学会如何处理新技术带来的已知风险和未知风险
- 熟悉不同类型创新（如激进式创新、利基场创新、渐进式创新、简约创新）及其成功条件
- 应用价值敏感设计（VSD）的概念
- 学会从伦理和风险角度对新技术进行批判性反思
- 能够展示并思考如何将道德价值观（如隐私、安全、可持续性、包容性）转化为新技术的技术要求

3.致谢

本书得以出版，要感谢代尔夫特理工大学 Jeroen van den Hoven 教授的支持、课程主任 Saskia Roselaar 细致入微的审核校对，以及学校图书馆提供的出版机会和管理支持。

译者序

在数字化浪潮席卷全球的今天，技术的飞速进步和创新活动的日益频繁，为人类社会带来了前所未有的便利与机遇。然而，正如一枚硬币有两面，技术创新在推动社会进步的同时，也伴随着伦理问题、争议和未知风险的涌现。创新并非无条件的“好”，而是需要在特定的社会、伦理和制度框架内进行考量。技术创新若缺乏责任与伦理的约束，很可能带来灾难性的后果。

随着纳米技术、生物技术、人工智能等前沿技术迅猛发展，人类面临着更加复杂紧迫的伦理挑战。这些技术不仅涉及到生命、健康、隐私等核心价值的保护，更要求我们思考如何在创新过程中尊重并维护人类尊严和基本权利。因此，将“负责任创新”的理念融入技术创新全过程，已经成为当今时代的重要课题。

本书从多个角度深入剖析了负责任创新的内涵与实践，包括应用伦理、制度与价值观、管理与创新、风险评估与安全等方面。它为我们提供了一个全面的框架，指导我们如何在数字化时代进行负责任的创新活动。无论是企业、研究机构还是政府部门，都需要在创新过程中充分考虑伦理、社会和环境因素，确保技术创新能够真正造福人类社会。

在启动正式编译工作之前，我们抱着试试看的态度，向原作者 Joost Groot Kormelink 发送邮件以征求他的意见。出乎意料的是，他很快便回复了我们的请求，不仅给予了我们热情的鼓励，还建议我们根据实际情况增补相关内容。我们对此欣然接受并深感荣幸。在编译过程中，我们深感责任重大。我们力求保持原作完整性和准确性，同时结合我国实际情况和读者需求，对部分内容进行了适当调整。比如，在对应章节增加相关谚语、讨论题。鉴于我们在每章节结束加入了讨论思考题，故将原版附录中的章节思考题删去；鉴于原版电子学习资源可在网络链接获取，团队在编译过程中将原文附录 4 的教师和网络课程链接列表删去。一些同行认为，现在人工智能可以完全胜任翻译工作，无需大费周章。然而，我们认为，人工智能对某些内容的理解和识别还比较生硬，而且这与本书主题难免有些冲突。加之考虑原作者的建议，我们还是本着认真负责的态度进行了逐句翻译和校对。本书分工如下：何良兴负责统筹与校对；第 1 章、第 3 章和第 5 章，刘彩珠、何良兴；第 2 章和第 4 章，王佩玥、何良兴；第 6 章和第 7 章，何良

兴、季星月；第8章，何良兴；第9章，刘彩珠、季星月。

最后，感谢所有为本书付出努力和贡献的专家和学者，正是他们的智慧和努力，才使得这本书得以完成并呈现给读者。本书希冀以此增强读者对负责任创新相关内容的了解，期待更多读者能够加入到“负责任创新”的实践中来，共同为构建一个更加美好的未来而努力。

何良兴

2024年小暑，南京

目 录

第一部分 负责任创新概述	1
1. 负责任创新简介	1
1.1 负责任创新的现实背景.....	1
1.2 为什么讨论负责任创新?	3
1.3 界定负责任创新.....	5
1.4 欧盟的负责任创新定义.....	6
1.5 负责任创新实质和过程.....	7
第二部分 负责任创新的应用伦理	9
2.负责任创新的应用伦理	9
2.1 应用伦理：思想实验.....	9
2.2 工程师如何解决电车难题.....	12
2.3 个人道德责任.....	13
2.4 集体道德责任.....	16
2.5 复杂系统中的责任.....	18
2.6 情感与价值观.....	22
2.7 道德困境和道德过载.....	25
案例 1：智能电表与价值观冲突——创新的机遇.....	27
案例 2：人工智能和大数据时代的医学伦理.....	28
第三部分 制度与价值观	31
3.创新的制度背景	31
3.1 实质性价值和程序性价值观.....	31
3.2 制度及其价值观.....	31
3.3 解释和设计公共价值观.....	32

3.4 了解开发人员和决策者的价值观.....	33
3.5 创新中的制度价值观解释.....	34
案例 1：北海风能.....	36
案例 2：自动驾驶.....	38
第四部分 创新与管理	42
4.创新与商业	42
4.1 渐进式和突破性创新.....	42
4.2 创新的决定因素.....	45
4.3 创新管理.....	48
案例 1：电视的发展与传播.....	50
案例 2：冷却剂.....	51
5.简约创新	54
5.1 什么是简约创新.....	54
5.2 简约创新案例.....	55
5.3 简约创新和负责任创新的联系.....	56
5.4 创新和社会标准.....	57
5.5 创新和包容性发展.....	59
5.6 结论.....	63
案例 1：TAHMO 气象站	63
6.企业负责任创新实施：新标准	68
6.1 企业负责任创新实施简介	68
6.2 企业负责任创新实施路线图.....	68
6.3 负责任创新模板：路线图.....	75
6.4 实施负责任创新的 SWOT 分析	76
第五部分 风险评估与安全	78

7.认识风险	78
7.1 风险、不确定性和无知.....	78
7.2 未知的未知：极端不确定性.....	80
7.3 技术评估.....	82
案例 1：关于核能的争论.....	84
案例 2：当大数据遇到“老大哥”	90
8.风险管理和安全工程	94
8.1 风险管理和安全工程简介	94
8.2 风险和安全的定义.....	94
8.3 成本收益分析.....	95
8.4 量化与比较风险.....	98
第六部分 价值敏感设计	110
9.价值敏感设计	110
9.1 价值敏感设计简介.....	110
9.2 价值敏感设计方法定义.....	112
9.3 价值敏感设计的实践应用	113
9.4 如何将道德价值观转化为设计规范？	115
9.5 复杂的过程.....	118
案例 1：自动武器.....	119
案例 2：护理机器人.....	122

第一部分 负责任创新概述

1. 负责任创新简介

“让新技术为社会服务……而不是带来更多超过其解决范畴的问题。”

（“Making new technologies work for society…… without causing more problems than they solve.”）

——希拉里·萨克利夫（Hilary Sutcliffe）

1.1 负责任创新的现实背景

在了解负责任创新定义之前，我们先将相关讨论放在具体背景下，通过反思以下困境和潜在问题，帮助我们思考并讨论本书中的负责任创新问题。

（1）困境 1：应对危险

新技术可以带来危险，并且我们很可能无法控制或遏制结果。我们预期每项创新都有一定程度的风险，有些风险不可避免，但我们能接受这些风险对人类健康、环境和社会的伤害有多大？此外，考虑危险是否可控制至关重要。例如，如果我们发现某些东西有危险，我们是否能从社会中移除特定技术、抑制甚至逆转其影响来限制其不良后果？即使它限制了创新的应用性，我们也应该抑制这些影响吗？

你认为危险在多大程度上应该是可控的？它们应该被完全控制吗？或者，你认为一定的风险或危险是生活的一部分，并且它伴随每项创新而存在吗？

（2）困境 2：对结果的认识

全面可靠的新技术评估需要相应的知识水平。那么，我们如何获得这些知识？我们对危险可能发生或不发生的确定程度有多高？

知识水平可以从无知到对可能性的不确定，再到知道失败的概率或拥有特定知识。如果我们不确定结果，由谁负责发现、监测和采取预防措施来应对危险呢？

在评估新技术时，我们需要掌握多少有关其危险和风险的知识，才能决定将其引入社会？我们应该假设重要的风险和危险会不时发生，而且事先无法预见和

评估吗？或者，我们应该事先确定所有可能的危险和风险，并有能力在一定程度上预防或遏制负面结果吗？

那么，使用有潜在危险的技术又该怎么办呢？我们应该监控这类技术的各个方面吗？或者说，持续监测并不是必要的，因为关键问题无论如何都会变得明显，所以我们只需要找到一种方法来报告和应对任何问题？

(3) 困境 3：风险和利益分配

风险和利益应如何分配？它们应该在所有社会群体和代际之间平等分配吗？或者，正如实际生活中通常的情况，利益和风险的分配是很难平等的吗？什么构成公平分配？

本质上，这条思考线索探索了技术的预期社会利益和危害，以及这些利益和危害如何在利益相关者（包括环境和子孙后代）之间分配。



图 1-1 刚果抗议活动“反对儿童在恶劣环境下被迫在钴矿工作”

(4) 困境 4：反馈和民主影响

普通公民是否对新技术的设计和应用有一定影响力？社会活动家、非政府组织、公民和其他公共团体在多大程度上能够影响技术发展？如果需要的话，他们是否有权阻止潜在有害技术的发展？还是只有生产者和专家才有足够的知识和能力来做出关键决定？



图 1-2 伦敦抗议活动“反对杀手机器人”

1.2 为什么讨论负责任创新？

创新常带来美妙又无法想象的新功能，这些功能的需求旺盛，并可能带来新的企业、新的工作岗位，从而实现经济繁荣。创新不仅带来了金钱利润，还给我们带来了青霉素、清洁的水和卫生设施。由于这些创新，我们的预期寿命大幅提高，数亿人从贫困和疾病中解脱出来。显然，许多类型的创新都是可取的。

但可以肯定的是，创新本身并不是好的。如果我们同意某样东西真的很有创新性并带来有趣的新功能，那么我们仍然完全有理由问：“它是好的吗？”。有很多创新例子，起初看似是一件好事，但后来却引起了严重的道德问题，比如含有 DDT 的杀虫剂、含有石棉的建筑材料（长期吸入石棉会致癌）。这些创新曾作为奇妙的新技术发明出售，但现在却与疾病甚至死亡的风险紧密关联。

联合国可持续发展目标和欧盟的重大挑战为全球范围内的创新和应用科学提供了一份亟待实现的道德目标清单，而且欧盟已拨出大部分预算用于加快这项工作。



图 1-3 联合国 17 个可持续发展目标

因此，在我们这个时代，创新不再是生产更舒适的产品设备，而是要拯救地球，并将其以良好的状态留给子孙后代。我们同时也应该担心气候变化、可再生能源、自动驾驶汽车、大数据和隐私、核能和核武器的扩散。我们已经知道许多创新都对社会产生了巨大影响，它们影响着地球偏远角落的人们、影响着整个地球以及子孙后代。

创新甚至已经开始改变人类的意义，比如人工耳蜗可以让失聪的人恢复听力，先进的假肢和人工器官为疾病患者和残疾人带来了便利，认知神经增强能让一些人变得更加聪明。

这些创新是否可以接受，将取决于它们的确切特征以及我们如何塑造这项技术。这意味着我们必须对创新负责，并意识到技术从来都不是中立的，而是受主观价值影响。

过去许多学者意识到技术继承了制造者的价值观。几个低技术含量的例子可以说明这一点。比如，伯利恒耶稣教堂的入口被称为“谦卑之门”，因为游客必须弯腰才能进入；几个世纪以来，为了防止小偷骑马进入教堂，入口已经变小；坚固但低矮的门与谦卑无关，实际上是为了发挥安全功能。

Langdon Winner 在其著名文章《文物有政治立场吗？》指出，20 世纪初纽约低垂式立交桥被故意设计得很低（图 1-3），目的是防止公交车从贫穷的黑人社区开往中产阶级的白人社区。



图 1-4 低垂式立交桥

随后，技术和设计中表达、体现的这一基本价值观念在科学技术研究领域得到了进一步发展。近年来，软件工程研究让人们注意到信息通信技术成为重要的新价值载体。它表明搜索引擎、金融软件和地理信息系统（GIS）也可能包含有争议的算法和模型，这些算法和模型在我们与之互动时塑造着我们的行为和思维。

如果我们不对技术所支持和体现的价值观进行批判性、系统性评估，那么意图不高尚的人会干涉他们对可持续性、生命财产安全、健康和福祉、隐私和问责制度的看法。因此，创新不仅致力于解决世界上的重大挑战，它本身也必须表达共同的道德价值观。技术太核心了，其背后的科学也太基础了，我们不能忽视。我们不应该等待结果并只在事后反思。这就是为什么我们需要思考和行动来促进负责任创新，无论是明晰现有技术嵌入的价值观，还是将所期待的价值观发展成实际的、可部署的设计参数。

1.3 界定负责任创新

我们同时追求诸多不同价值观，很难在它们之间做出选择或妥协，有时甚至不可能。基于这种事实，我们高度同时重视隐私、健康、可持续性、效率、公平、安全、问责制度等问题。



图 1-5 隐私和安全：哪个更重要？

然而，对于上述问题我们不能同时满足，需要权衡。换言之：我们经常发现道德义务超出了我们所能满足的范围，这会导致道德过载问题。我们将在本书第二章对此进行更详细讨论。这通常被当作一个问题，但它实际上还能激发创造力和承诺，使人们试图通过智能设计和创新来适应相互冲突的价值观。例如：

- Fairphone 是一家使用无冲突金属制造智能手机的初创企业，因此它可以将人权、可持续性、公平和安全都纳入一种设计中。
- 荷兰已经建立了大型防潮堤来保护国家免受洪水侵袭，它们同时也是管理生态系统和产生潮汐能的方法。
- 隐私增强技术让我们享受计算机带来的好处，而不受泄露隐私的困扰。
- 清洁技术让我们在不破坏环境的情况下，实现工业生产和经济繁荣。
- 瑞典对致命交通事故采取零容忍政策，这引发了汽车行业的大量创新，沃尔沃现在成为安全汽车生产领域的领导者。

1.4 欧盟的负责任创新定义

欧盟报告中“关于加强研究和创新的选择”，提供了相关条款来界定负责任创新。该报告认为，要使一个创新组织或过程是“负责任的”，这就意味着那些发起并参与它的人必须是有道德的、负责的代理人。换句话说，他们必须：

- 尽可能多地获得关于行动结果后果的知识，以及向他们开放选择权限的相关知识；

- 根据相关道德价值观有效评估所有的结果和选择（包括但不限于福祉、正义、平等、隐私、自治、安全、可持续性、问责、民主和效率）。鉴于“价值观设计”的概念（本书第9章）和通过设计解决问题的可能性，负责任创新的另一方面是相关道德行为主体的能力；
- 将这两个考虑因素作为新技术、新产品和新服务的设计和开发要求，从而提高道德水平。

从本质上说，**负责任创新是指向新形势迈进，即扩大更多义务和责任，以超越以往对同胞、环境、地球和子孙后代的责任。**

欧盟（EU）和责任研究与创新（Responsible Research & Innovation）

欧盟（EU）总是谈到负责任研究与创新（RRI），并将其定义为社会行动者、研究人员和创新者积极合作，共同定义、设计和构建社会可接受、可持续、能解决重要社会问题的方案、服务和产品。

负责任研究与创新是由欧洲资助项目中的一个优先事项，它鼓励社会行动者在整个研究和创新过程中共同努力，以便更好地使其结果与社会的价值观、需求和期望相一致。这意味着研究人员、科学家和决策者彼此之间应该互动，并与其他社会行动者互动，从而建立相互认识、共同定义、共同设计新的举措，并确定解决社会挑战的方案。负责任研究与创新关注的是研究和创新如何更有利于社会，同时保护环境。

研究和创新有助于找到解决社会主要挑战的解决方案，例如循环经济、防止气候变化、减轻人口改变的影响、改善福祉、能源安全、食品安全和社会安全。欧盟认识到这些挑战，并认为负责任研究与创新是解决这些问题的主要方法之一。此外，负责任研究与创新在科学和政策间建立了很好的桥梁，这使决策者能够更好地了解情况和制定改进措施，并实现生态目标和经济目标。

1.5 负责任创新实质和过程

1.5.1 负责任创新实质

创新也可以被建构为一个道德概念，因为它有助于改变世界，这样我们能够满足的一系列道德义务就会被放大。当然，这并不能保证紧迫的道德问题总会有完美的解决办法，在某些情况下，我们可能需要采取更极端、更基本的方法。不

过，我们确实有义务看看是否有可能利用创新来满足相互冲突的价值观。可以说，这就是负责任创新的结果或实质。

1.5.2 负责任创新过程

负责任创新也是一个过程。为了理解责任如何在复杂系统（多行动者）中分配，我们必须确定追究责任的标准（如，知识、意图、非强迫、贡献过错和能力）。这个列表往往与人们否认责任时提供的借口非常吻合：“我不知道”、“我不是故意的”“我是被迫的”“不是我”“我不理解”。

我们所做的每一件事，都能以这样的方式来扩大责任或削弱责任，进而使其他人更难追求我们的责任。为使否认责任显得合理，可以采取许多策略来保持知或假装无知。比如，想想与新材料和化学物质相关的风险：“我们不可能预见这些风险。我们的竞争对手也使用了石棉。有错的不是我们，而是分包商。我们公司当时没有资源来批判性地考虑这个问题。”

讨论题：

1. 整理其他有关负责任创新的定义，并比较这些界定的异同。
2. 企业或组织在产品和技术创新中面临的现实困境有哪些？
3. 负责任创新的本质是什么？
4. 为什么一定要实施负责任创新？

第二部分 负责任创新的应用伦理

2. 负责任创新的应用伦理

“在面临道德抉择时，我们的决策将揭示我们真正的品格。”

——康德

2.1 应用伦理：思想实验

为了用抽象方式探索道德和伦理的细微差别，传统上来说，哲学家们会采取思想实验的方法。思想实验通常会设置一个精心策划的困境，要求读者选择他们喜欢的行动方案，并证明为什么他们的选择是两害相比较轻的那一个。这样能探索对困境不同反应的哲学含义。当我们谈到负责任创新时，真正理解“负责任”这个词非常重要，即由谁负责？如何负责？何时以及为什么负责？而“电车困境”就是能够达到这个目的的思想实验。

(1) 电车困境

“电车困境”（亦称“电车问题”）由英国哲学家菲利帕·富特（Philippa Foot）于1967年提出，它由一系列假设情景组成，每个场景都呈现了一个极端的环境并考验着受试者的道德能力。1985年，美国哲学家朱迪斯·贾维斯·汤姆森（Judith Jarvis Thomson）在《耶鲁法学杂志》上对富特的思想进行了仔细的研究和扩展。

电车困境是伦理学中的一个思想实验，其一般形式如下：有一辆失控的电车在铁轨上飞驰（图2-1），在轨道前方有五人被绑住无法移动，此时电车正朝他们冲去；你站在远处的火车站里，旁边是一个操纵杆，如果你拉动操纵杆，电车将切换到另一轨道；然而在另一轨道上绑着一个人。

基于上述情景，有两种选择：

- 什么也不做，电车会撞死在主轨道上的五个人。
- 拉动操纵杆，将电车转移到侧轨道上，导致一人死亡。

针对两个选择，你会怎么做呢？

- 拉动操纵杆，最大限度地增加挽救的生命数量（一人死亡，五个人活着）。

- 拉动操纵杆，因为你是一个富有同情心的人，这是正确的做法。
- 不拉动操纵杆，因为这会导致死亡，而死亡本身就是错误的。
- 不拉动操纵杆，从文化角度来说，协助他人死亡不正确甚至违法。

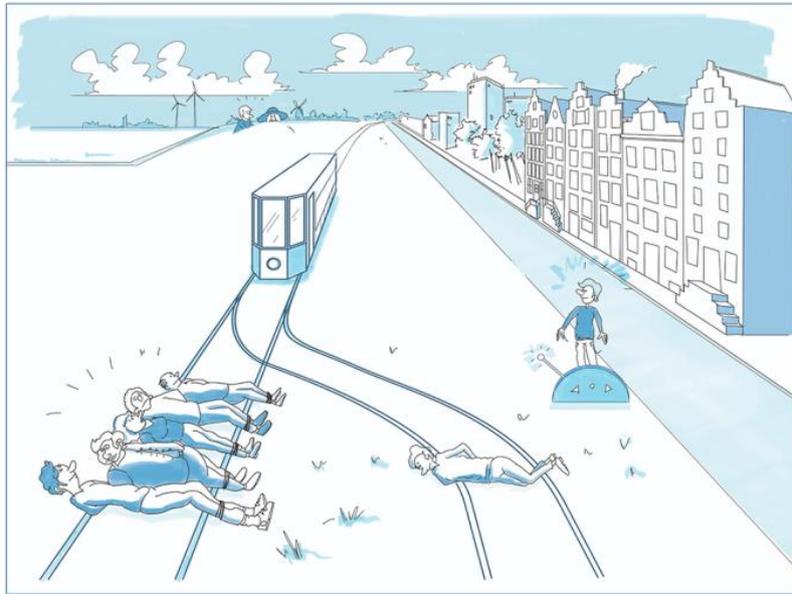


图 2-1 著名的电车困境：五个人，还是一个？

考虑到上述电车困境，你会怎么做？从道德角度来看，拉动操纵杆是被允许的吗？或者你有道德义务这样做吗？在过去三十年里，几乎所有哲学家都在这种所谓的“电车案例”中长大的。如果你想获得电车困境分析的博士学位，如果你能在大量文献中增添新的内容，这将成为世界各地哲学系值得尊敬的话题。

我们之所以讨论这个思想实验，并不是为了介绍与之相关的大量文献，而是为了说明负责任创新需要一种不同于电车困境哲学观点的道德选择和责任观。也许它为我们思考高科技世界中的责任增加了一个有价值的维度。

电车困境案例的简单计算表明，一个人拉动操纵杆可以挽救四条生命。经过思考和计算，大多数人认为拯救五个人是道德允许的，尽管有一人因此失去了生命，但他们中的多数人仍认为自己有道德义务去这样做。

(2) “胖子”案例

现在我们稍微改变电车困境的故事内容，并从故事中切换出来。具体的，仍有五个人被绑在轨道上，电车向他们驶来，但有一个胖子站在轨道的桥上（图 2-2）。把胖子推到轨道上，就可以在电车撞到五人之前使它停下来。我们期望人们对这种情况的反应与原始版本的反应相同，因为两种情况有相同的数字和计

算：通过一人死亡来拯救另外五个人。

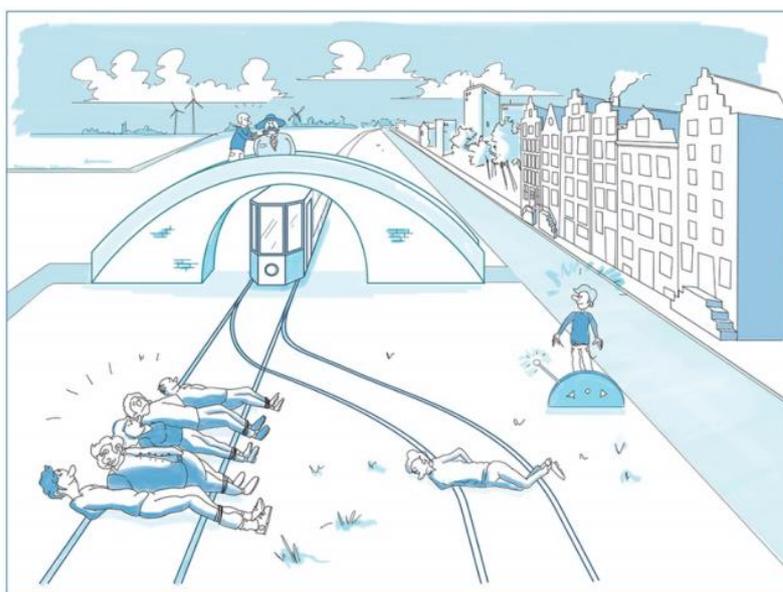


图 2-2 胖子案例

然而，实证研究表明，一些人认为大脑成像研究指向这个方向：尽管数字和计算相同，但人们对两种情况的反应截然不同。在拉动操纵杆的情况下，我们主要依靠冷静推理和对生命损失的计算；然而，当我们可以选择推下“胖子”时，我们的反应往往是厌恶或大笑。因为，利用一个人作为障碍物并杀死他，这总做法似乎是荒谬的。

然而，电车困境不仅仅是一个思想实验。对自动驾驶来说，我们必须做出类似的决定：如果事故无法避免发生什么？它是选择保护司机还是行人？如图 2-3 所示。

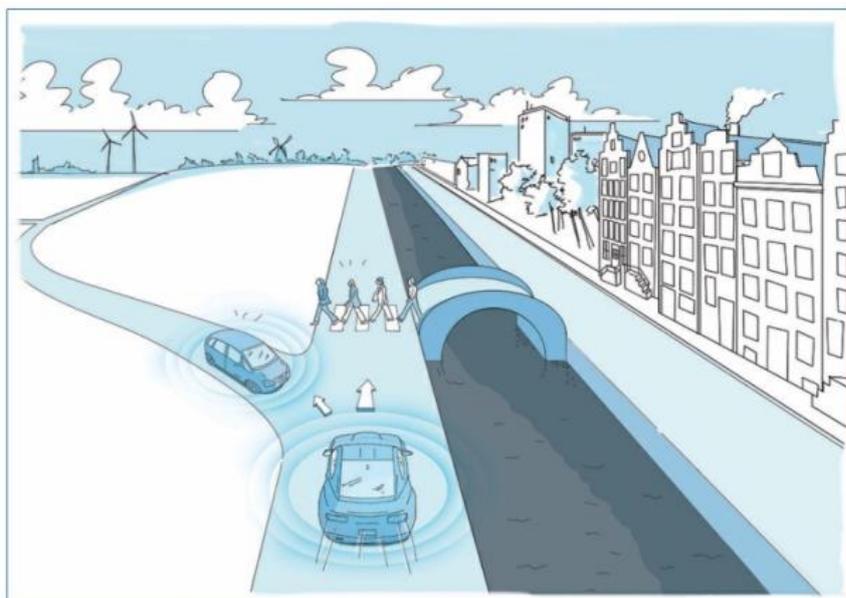


图 2-3 自动驾驶面临“电车困境”

2.2 工程师如何解决电车难题

我们已经看到了电车困境所带来的哲学问题。然而，在电车相关文献中未涉及的是：技术产品的工程师和设计师经常对电车困境持有一种看法，即铁路基础设施设计的很糟糕。基于此，工程师们会立即开始考虑更好的系统设计和创新，从一开始就防止这种悲剧的发生。

正如工程师建议的那样，基础设施应该包括：早期预警系统、自动断开系统和开关，以防止操作员在开关附近做出如此悲剧性的选择。但是这在哲学研讨会上可能不太合理，因为解决困境并不是目的。然而，在另一种情况下，即与“防止铁路运输中的死亡和维护铁路基础设施安全有关”的情况下，这种推理方式则非常有趣。

这清楚表明：在这种“电车案例”中，情境是给定的、不可改变的，正如在思想实验中所期望的那样。但是，由于工程师们不愿把现状看作理所当然，他们的目标是通过更好设计来改变世界，避免悲剧性的选择。

这种关于电车困境的主流道德思维模式，其中的条件是既定、不可改变的，转移了人们对以下事实的注意力，即现实中的问题通过不是想象力丰富的哲学家为准备学术论文而努力构想出来的。

道德困境是众多先前设计决策的结果，这些决策由许多人做出，但不一定由最终面临抉择的人做出。日常和职业生活中的道德困境——当然还有那些涉及技

术的——几乎总是由不同行动者在复杂过程中做出成百上千个决定和选择的结果。设计历史在现实世界中的确很重要，因此学习如何防止困境发生与在困境出现后如何看待它们同样重要。

无论思考设计或开发自动驾驶汽车、IT 基础设施、新材料、食品、药物还是能源选择，我们都不可避免地要在塑造未来用户的选择架构(即可供选择的不同设计方式)。工程师们知道，在现实生活中处理道德问题的最好方法，通常是预测失败的情况，并解决这些问题，而不仅仅是等待困境出现。

类似电车困境的道德问题引发了人们的讨论，这些讨论可以充分应用于高科技创新、系统或基础设施设计。此外，这些讨论揭示了人们不同的价值观。因此，让所有利益相关方参与进来，解决他们的价值观问题才是重要的。

负责任创新表示预期道德选择和对他人负责，这些人可以是同胞也可以是子孙后代。它关注的是在理解基础上设计和塑造技术，因为未来用户、消费者、病人、公民和后代都受限于工程师和应用科学家在当下提出的、经过深思熟虑的或早已遗忘的选择。他们的责任承担能力取决于漫长而详细的设计历史。这同样适用于能源选择、互联网协议、智能城市、新材料以及以某种方式或形式应用于社会的任何其他创新。

2.3 个人道德责任

前述章节介绍了思想实验，本节将引入更多复杂场景以便更接近现实世界，因此本节内容将更多灰色地带包含在内。接下来，本节将用这些场景具体考察个人道德责任的各种概念。在负责任创新背景下，区分这些概念非常很重要，因为其重要目标之一就是以促进责任的方式进行设计和创新。不过，在清楚了解如何促进责任之前，需要先清楚地了解什么是责任。

本节把重点关注“回溯式责任”，即人民对过去行为进行评判而不考虑未来需要承担的义务。以下是相关案例内容。

(1) 案例：“错误开关”

这是一个化学品事故的案例，我们称之为“错误开关”。想象一下，一个化工厂的操作人员注意到一个储罐在泄漏，他在试图控制泄漏时，不小心按错了开关。进一步想象一下，这样做的直接后果导致爆炸发生和另一名工人死亡。鉴于这些信息，似乎可以合理地得出结论，操作员对工人的死亡负有责任。毕竟，是

启动开关引起了爆炸。所以，检验行为人是否对结果负有因果责任的方法，是思考行为人如果不采取这种行为能否引发同样结果。

这种理解因果责任的方式似乎没有争议，可以适用于“错误开关”案例。但需要注意的是，如果询问操作者是否对工人死亡负有道德责任则是另一个问题。一个人对结果负有因果责任，不代表他就负有道德责任。操作员的道德责任似乎取决于他为什么按错开关——在这种情况下，这是一场意外。

(2) 案例：“混合电线”

该案例是“错误开关”案例的另一版本，它包含了一些额外信息，可以更详细地解释哪里出了问题。假设开关的接线弄混了，而操作员不知道这一点。由于线路原因，操作人员拨动了他认为正确的开关，但没有阻止泄漏，而是发生了爆炸，导致另一名工人死亡。重要的是，在这种“混合电线”情况下，操作员试图阻止爆炸，但为时已晚。相应地，声称经营者对工人的死亡负有因果责任似乎没有争议。如果他没有按开关，爆炸就不会发生了。但是，因果责任并不包含道德责任，所以同样的问题：该案例中的操作者对工人死亡负有道德责任吗？

找到答案的方法是考虑一个相关问题，即在这种情况下，操作员是否应对工人的死亡负责。考虑到操作人员不可能知道电线被调换了，而且他无法阻止爆炸的发生，尽管他试图这样做，但将工人死亡归咎于操作人员似乎是错误的。也就是说，将工人死亡归咎于操作人员是不恰当的。

(3) 案例：“可恶的操作员”

为了能清楚地理解这种细微差别，将这种情况与操作员明显应该受到谴责的情况进行比较。因此，在接下来的案例中，再次变换一个版本，情况与之前完全不同：操作员故意打开错误开关，杀死工人。

这个案例在两个方面与“混合电线”案例有所不同。首先，操作人员对爆炸中死亡的工人怀有恶意，而在“混合电线”的情况下，操作人员没有这种恶意，实际上有动机试图阻止爆炸。第二个区别是，在“可恶的操作员”案例中，致命的爆炸可以避免。操作人员知道他将按错开关，并故意这样做，造成工人死亡。因此，直觉上认为，在“可恶的操作员”案例中，操作员对杀害工人负有因果责任和道德责任。以上都得出操作人员做了一些导致工人死亡的事情，并且操作员对工人怀有恶意这一事实，意味着操作员在道德上有过错。重要的是，这种道德追责只是我们认为一个人对某件事或结果负有道德责任的一种方式。

(4) 案例：“额外的努力”

接下来的案例表明，对某件事负有道德责任，但不一定为此事负责是可能的。我们把这种情况称为“额外的努力”。这种情况类似于“混合电线”的情况，因为操作员不知道也不可能知道他转错了开关。然而，想象一下，当操作员意识到他按错开关时，只有几秒钟时间来转动另一个开关，以防止爆炸。但是，打开另一个开关不是正常过程，需要一些努力。最后，假设操作成功，工人获救。

在这种情况下，操作员显然对拯救工人负有因果责任，他必须迅速思考，采取非常困难的行动，来挽救工人生命。他为了拯救受害者而付出额外努力，这似乎是在道德上赞扬操作者的很好理由。需要注意的是，这个操作员在道德上有责任拯救操作员的的生命。然而，在道德上值得赞扬是一个人在道德上负责任的另一种方式。

(5) 案例：“例行程序”

化学品泄漏强调了道德责任的另一个重要方面，我们称之为“例行程序”案例。在这种情况下，有化学品泄漏，操作人员按右开关，没有混合电线，转动开关不需要特别努力。此时，你可能不倾向于操作员转动开关是值得赞扬的，因为他的行为非常普通，不需要付出巨大努力。

考虑到运营商没有做错任何事，他似乎也不应该受到惩罚。操作者自由地、有意地做了正确事情，而且他知道自己在做什么。基于这些原因，我们可以认为，操作员在道德上有责任避免工人死亡。

(6) 个人道德责任包含哪些内容？

从这些案例中我们可以学到几种不同的责任概念：

- 在因果责任的情况下存在最低水平的责任。回想一下，在“混合电线”案例中，操作员对工人的死亡负有因果责任，但没有道德责任。
- 责任的第二个概念是道德责任。在“例行程序”案例中，尽管操作人员既不值得赞扬也不应该受到指责，但他对防止工人死亡仍负有道德责任。
- 第三种责任概念涉及行为人的行为值得表扬或谴责。当操作员超越职责要求去做正确的事情时，他值得称赞；当操作员意识到可能导致工人死亡且仍然采取行动时，他则是有罪的。他们确实在道德上负有责任，但我们倾向认为，他们应该受到赞扬或受到谴责。

理解不同责任概念之间的联系尤为重要。第一个联系是，道德责任以因果责

任为前提，即操作员要对死亡负有道德责任，那么他的行为必须导致工人死亡；没有因果责任，就没有道德责任。第二个联系是，可赞扬性和可谴责性都以道德责任为前提。例如，如果操作员无法避免工人死亡，那么他在道德上就没有责任，因此不应受到惩罚。因此，对值得赞扬和应受谴责的判断，都假设所涉及的人对结果负有因果责任和道德责任。

区分不同责任概念，可以将它们应用到负责任创新中。作为一名创新者或设计师，反思影响责任归因的各种因素能有助于在设计过程中适应这些因素，从而减少出现问题的可能性，无需人们为其承担道德责任。行为人需要对流程及其在流程中的角色有清晰、及时的认知，而且系统本身应该设置多个易于访问的故障保护。

2.4 集体道德责任

在某些情况下，仅靠个人道德责任不足以解决关键问题，特别在涉及对结果具有同等影响力的其他各方的情况下。接下来，本节将讨论集体行动的问题，这种问题有时被称为“公地悲剧”。

该问题可能出现在资源共享背景下，如河流、大气和国家公园。下面将集中讨论公地悲剧的一个典型案例，即过度捕捞。

想象一下，海边村庄的小渔民靠捕鱼为生，村子里的每艘渔船都必须与其他渔船竞争来获得更多鱼类。由于这种竞争和对鱼的持续需求，出现了过度捕捞，最终导致渔业资源枯竭。这里的“公地”是海洋中的天然鱼类资源。但悲剧是什么呢？在这种情况下，悲剧与过度捕捞似乎不可避免，也就是说，渔民个人的行为都出于他们自己的理性利益。

重要的是，尽可能多的捕到鱼符合每个渔民的理性利益。如果渔民 A 捕捞的鱼少于他的最大产能，他将赚更少的钱，同时他的竞争对手渔民 B、C 和 D，将捕获他没有捕获的鱼。这表明，如果捕鱼量少于最大捕鱼量，根本没有什么可获得的，实际上只损失一些利润。因此，A 的理性利益是让捕鱼量最大化。不幸的是，同样的逻辑也适用于其他渔民。由于每个渔民的行为都是出于个人理性利益，所以共同的鱼类资源很快就会枯竭。



图 2-4 过度捕捞

虽然个别渔民采取了理性行动，但从长远看，这种行为并不利于集体利益最大化。社区的利益受到损害，因为他们有可能失去重要的收入来源，失去饮食和经济基础。此外，渔民个人利益也因失去生计而受到损害。鉴于耗尽鱼类资源的这些影响，很明显，当作为一个集体考虑时，个别渔民的行为是不合理的。所以，即使个人理性行为也可能成为集体的非理性行为。

避免这一悲剧发生的解决办法是合作。每个渔民都应该实施可持续捕捞，而不是尽可能多地捕捞。可持续捕捞意味着从海洋中捕捞一定量的鱼类，这与鱼类种群的持续健康状况是一致的。这就意味着，个别渔民不得不把一些鱼留在海里，即使他们完全有能力捕获它们。可持续捕捞可以在合作计划中实现，例如渔业配额计划，该计划限制了每艘船的捕捞量。然而，为了使其发挥作用，整个社区，特别是渔民，必须同意这一计划。也就是说，他们必须共同制定与可持续捕捞相一致的鱼类配额，并且坚持这一配额。

(1) 搭便车问题

为什么渔民会坚持这个计划？回想一下个人理性利益并只考虑渔民 A。如果所有其他船只都遵守配额方案，那么渔民 A 的理性利益就是捕鱼超过配额。这叫做搭便车。同样道理也适用于所有其他渔民。因此，尽管集体理性合作计划的目的是避免耗尽共同的鱼类资源，但它实际上会被个人理性利益的搭便车行为所破坏。

那么，让个人遵守集体配额计划有哪些选择呢？在这种情况下，什么能真正激发合作呢？道德可能会激励个体渔民，但在这种情况下，什么样的道德需要考

虑？首先，渔民可能会把维持鱼类资源当做一个共同的、理想的目标，而配额是实现这一共同的、理想的目标的手段，这一事实是他们坚持配额的道德原因。因此，他们可能会被激励采取必要的手段来实现和维持这一共同的、可取的目标。其次，如果合作计划的设计方式既能维持种群，又不给任何渔民不公的份额或优势，那么渔民可能会受到合作计划公平性的激励。即使个人理性鼓励搭便车，受配额制度道德激励的渔民也可能会坚持这种做法。

然而，请注意，尽管道德动机可能是必要的，但它并不足以真正实现可持续捕捞。这是因为，不能指望每个人都出于道德考虑。如果以某种方式迫使他们这样做，许多人只会做他们在道德上应该做的事。为了弥补缺乏足够的道德动机，可以采取制执行。要使配额制度发挥作用，就必须达成一定程度的遵守。执行遵从性有几种选择。例如，社区授权海警执行配额制度，即使没有道德动机的人，也可以避免他们搭便车；通过罚款或处罚，例如吊销捕鱼许可证，这种强制执行使个人理性自利与集体理性保持一致。

(2) 执行的限制

不幸的是，依靠强制措施本身也可能不够。考虑到渔船数量和捕鱼面积，海警实际上不可能确保每个渔民都遵守规定。此外，如果出于个人理性的私利，海警也可能因受贿或偷懒在执法上松懈。

那么，解决公地悲剧的办法是什么呢？到目前为止，合作捕捞配额可能是维持渔业的最佳途径。然而，实现集体利益的道德动机，受利用这种情况的个人自身利益的挑战。这意味着采取某种强制措施是必要的，尽管这本身也可能存在不足。如果渔民和海警都出于道德动机来维持捕捞配额呢？让各方都注意到这种道德考虑，尤其当他们可能受到违反规则的诱惑时，这是一个有趣的设计问题，负责任的创新者应该努力解决这一问题。

2.5 复杂系统中的责任

到目前为止，我们已经分析了一些很容易分配责任的案例。因此，当出现问题时，通过找出谁在因果或道德上负有责任也很容易。不幸的是，现实世界非常复杂，多个利益相关者相互交织、相互影响对方的结果。要确定谁在因果或道德上负有责任，以及谁应该受到指责，则变得更加困难。

我们看到的是，所有利益相关者的行动共同导致了戏剧性的结果，但没有一

个相关个体被追究责任。这种现象被称为“多手问题”（责任分散）。因为牵涉到很多人，所以不可能确定其中一人为此负责。这类问题在工程中非常迫切，因为在工程中经常存在许多参与技术开发的人，即使在高风险的技术领域，如果出现任何问题，也可能造成严重后果。那么，在复杂的社会技术系统中，我们该如何处理责任分配呢？



图 2-5 复杂性

(1) 道德责任的条件

先从工程师的责任说起。工程通常是在多人的团队或网络中进行。在我们讨论这些团体的责任之前，我们首先要问的是：当我们说个人负责时，到底意味着什么？如果满足以下四个条件，通常可以说一个人是有责任的：

1) 自由条件：

人应该有行动的自由，而不受外界压力干扰。如果我拿枪指着某人的头，让他做一些非法或不道德的事情，这个人不能被追究责任。因为此时他不能自由地做其他事情。

2) 知识条件：

一个人应该知道他的行为会导致消极结果。如果这个人不知道这一点，他通常不会被追究责任。举个例子，有人粉刷了他家的门并且没有贴上相关通知，而你碰巧碰到了门，破坏了油漆，那么追究你的责任或责备你是不公平的。你不知道，也不可能知道那扇门刚刚粉刷过，因此你不应该碰它。

3) 因果关系：

个人行为与消极结果之间应该存在因果关系。我们不能对没有因果关系的事情负责。然而，请注意，有时什么都不做就是错误行为：如果一个人能拯救另一个人免受伤害，但是他却什么都没有做，这就是错误的行为。

4) 违反规范：

如果个人行为在某种程度上是错误的，那么可以说他违反了规范。这可以是法律规范，也可以是道德规范或社会规范。

(2) 多手问题

现在让我们看一个有几个人参与的例子，即一种新防火材料的开发和使用，有四个人参与其中：

- 参与者 A 在实验室工作，对这种新材料的原子性质做基础研究；
- 参与者 B 被消防队雇来为消防员设计一套新装备，使用这种有前途的新材料；
- 参与者 C 是消防队队长，是聘请的设计师；
- 参与者 D 在消防队工作，负责清洗消防员的服装；他把衣服带到干洗店清洗。

事实证明，这种有前途的新材料与洗衣粉接触后会致癌。干洗店的一名员工患上了一种致命的癌症，最终死亡。我们能说 A、B、C 或 D 中的一个人对清洁工的死负有道德责任吗？

这四个人都对因果发挥了一些影响。但上面列出的其他条件可能不满足；至少，我们可以说没有一个人满足所有条件。在实验室工作的人可能已经知道这种材料与其他材料发生化学反应，但他无法预见其他人如何使用这种材料。其他人可能不知道这种物质的致癌性。有人甚至会说，负责打扫卫生的人并没有真正地采取不同行动，因为打扫卫生没有其他选择。

所以，这四个人的行为戏剧性地导致了一个不幸的结果，但没有一个人可以被追究责任。这个案例说明了“多手问题”在实践中是如何发挥作用。因为牵涉到很多人，所以不可能认定一个人对此负责。在复杂或危险技术的工程中，这个问题非常紧迫，因为通常有很多人参与技术开发，更不用说当出现问题时可能会产生很大影响。



图 2-6 英国石油灾难

以 2010 年 4 月 20 日发生在墨西哥湾英国石油公司平台的石油泄漏为例(图 2-6)，这是一场工业灾难。这场灾难的影响是巨大的，它立即引发了谁应该为这场灾难负责的问题。

“多手”问题通常在负面事件发生之后讨论。然而，我们也可以从前瞻性的角度来看待它，然后对照道德责任条件，检查一个人是否有能力履行他的责任：这个人有行动的自由吗？他们有必要信息吗？是否有正确的规范？

(3) 将责任融入技术

这给我们带来了一个有趣的话题，即责任和技术之间的关系。飞机上的自动驾驶仪是技术替代人接管责任的一个明显例子。但同样地，技术的发展是否能够使人们承担责任呢？我们认为技术确实可以承担起这个角色，但为了保证这一点，我们应该在技术开发时候注意具体的责任。这里有两个例子：

例字 1: V 型芯片。V 型芯片是一种防止儿童观看成人电视内容的技术设备。电视台播放收视率作为节目的一部分。家长可以设置一个阈值级别来对 V 型芯片进行编程，这样当 V 型芯片打开时，所有高于该级别的节目都会被自动屏蔽。因此，孩子们在看电视时，并不能看到被封锁的节目。有些人认为，通过使用 V 型芯片，父母把责任转移给了电视台，因为电视台决定每个节目的确切评级，从而决定这个节目是否在电视上播出。从这个角度来看，V 型芯片限制了父母的自由。然而，也有人认为，V 型芯片为父母提供了更多关于成人内容的信息；因此，它给了他们更多自由来控制孩子看什么。V 型芯片是否限制或加强了父母的责任仍有待讨论，但这个例子清楚表明了，技术可以而且确实影响了一个人的责任。

例子 2：控制室。控制室是监视和控制大型设施或服务的中心空间。这些房间通常配备有多个监视器和屏幕(如图 2-7)。在控制室工作的人必须在大量信息的基础上做出决定。这也就意味着这些房间布局以及信息呈现方式，决定了人们能够做出正确决策的程度。我们可以说，控制室的糟糕设计可能会阻碍人们承担责任。反之亦然，一个设计良好的控制室可以提高一个人执行责任的能力。

因此，技术可以赋予人们权力，但也会阻碍人们履行职责。负责任创新的一个方面是开发技术，使其能够促进或加强人们履行职责的能力。



图 2-7 控制室示例

2.6 情感与价值观

科技带来的风险给 21 世纪的人们提出了重要的伦理问题，对事故、污染、职业危害甚至环境破坏的可能性和潜在的灾难性后果进行考虑。由于对这些风险的主观认知，有争议的技术可能引发强烈的(负面)情绪，包括恐惧和愤慨，这往往导致专家和外行人之间的冲突。

在讨论涉及风险的技术时，情绪通常被视为一种干扰因素，因为它们是非理性的，对事实和相关信息免疫。然而，我们将论证情感也可以是实践理性的来源。自然的情感，如恐惧和同情，可以帮助我们把握风险技术的道德显著特征，比如公平、正义、公平和自主，这些在传统的技术官僚风险评价方法中可能被忽视。

(1) 风险和风险感知之间的区别

在纳米技术、合成生物学和信息技术等高风险技术背景下，负责任创新尤其

具有挑战性。这些技术经常引起激烈和情绪化的公众辩论。虽然专家们强调所谓的低风险科学研究，但公众往往担心这些技术会对社会有影响。专家们倾向认为，公众的担忧是由于缺乏了解，但这并没有减少公众情绪。

政策制定者对此的反应通常有两种：要么忽视公众的情绪，要么将其作为禁止或限制一项技术的理由。让我们把这两个极端分别称为技术官僚陷阱和民粹主义陷阱。在这两个陷阱中，没有关于情感、公众关注和道德价值观的真正辩论。然而，这种行为应该得到纠正。

几十年来，社会科学家、心理学家和哲学家一直在反对技术官僚。他们认为，风险不仅是一个定量的、科学的概念，也不仅是通过成本效益分析评估不希望出现的后果的可能性。换句话说，对风险的体验与对风险的计算是完全不同的。

风险关系到人类福祉，涉及到公平、公平和自主等伦理考虑。风险学者之间有一个强烈共识，即在任何风险评估中都应包括伦理考虑。有趣的是，正如心理学家保罗·斯洛维奇(Paul Slovic)颇具影响力的研究中提到，这些因素确实会出现在外行人的风险认知中。显然，人们对风险的预设理论内涵包括了伦理考量，而这些通常被专家使用的以定量为主的风险评估方法排除在外。因此，一些风险学者认为，外行人与专家有着不同但同样合理的理性。

显而易见的是，外行人的风险感知在很大程度上受自身情绪影响。社会科学家努力地想去解决这个问题，因为他们理解情绪是非理性的，这似乎影响了这样一种看法，即外行人可能采用另一种有关风险的、合法理性的思维方式。

(2) 情绪是可接受风险的向导

然而，情绪并不一定会威胁到理性。神经心理学家安东尼奥·达马西奥(Antonio Damasio)有一个著名的论断：没有情绪，我们就不可能真正理性。事实上，当前在哲学和心理学中情绪研究的主导方法是情绪认知理论，根据这种理论，情绪是认知和知识的一种形式或来源。这些观点让我们对情绪在风险技术辩论中所扮演的角色有了全新认识。情绪并非与理性相对立，因此在本质上不一定有误导性；相反，在评估风险的道德可接受性时，情绪可以被当作一种宝贵的智慧来源。

公众的情绪，可以让我们深入了解在有关风险技术和负责任创新的道德决策中，应该考虑的合理的道德考量。

专家们可能会觉得有责任，甚至担心他们开发的技术。这种担心和恐惧是对

一项技术不可预见的负面后果的担忧。恐惧则表明了一项技术对人类健康构成了威胁。例如，当我们面对克隆和人兽杂交时，我们常常感到厌恶。这实际上表明，从道德角度看，创造这样的生物是模棱两可的。与此同时，在我们面临违背自己意愿的风险时，愤怒可能是一种侵犯自主权的迹象。

从定义上讲，人们通常认为情感与技术对立。然而事实并非如此。例如，对一项技术的热情可能表明它对健康有益。同情和同理心有助于我们理解风险和利益的公平分配。因此，情绪可以把我们的注意力吸引到重要的道德问题上，否则这些问题可能得不到充分解决。避免技术官僚陷阱和民粹主义陷阱，为处理公众辩论中有关风险的情绪提供了一种不同方式。



图 2-8 抗议核能

这种替代方法，可以称之为“风险的情感审议方法”，即给公众真正的声音，让他们的情绪和担忧得到倾听和讨论，也为我们提供了一种方法了解如何以道德上负责任的方式沟通风险。反过来，道德情感可以为道德约束和负责任创新的理念参数提供重要见解。例如，在辩论中，专家不仅应该关注可能风险的小概率，还应该提供平衡积极效果和消极后果的看法，使个人做出明智的评估。

在考虑和沟通风险时嵌入情感也有助于行为的改变。例如，在气候变化活动中诉诸情感，可以增加目前缺乏的“紧迫感”，同时为促进环境友好行为提供动力。毕竟，情绪是动力的重要来源，因此可以被用来刺激行为发生改变。

在开发有风险的技术时，必须认真对待情感和道德问题，以便进行有充分根据的道德评估。同时，这种方法有助于克服专家和外行人之间的差距，这种差距

经常在风险技术辩论中反复出现。因此，公众会觉得自身关切得到了重视，从而提高自身参与度和促进负责任创新。

2.7 道德困境和道德过载

科学家、工程师和设计师经常感到有责任让世界变得更美好。具体如，让世界更安全、更可持续，在保护隐私和打击恐怖主义的同时创造新就业机会，给未来用户自主权和自由，提高后代生活质量。他们想实现这些目标，但像许多想做正确事情的人一样，他们也遇到了道德过载问题。

道德过载表示有太多好事要做，我们有太多义务无法完成，至少不能一次全部完成。我们希望经济繁荣和就业机会惠及所有人，但也希望可持续发展；我们重视安全，但也重视隐私；我们要求安全，但不愿牺牲自由；我们要求问责制，但坚持保密权利。

科学技术中的这种道德问题，往往以道德困境的形式出现。道德困境的最基本定义是以下三段论：

- a) 行为人应该做 A。
- b) 行为人应该做 B。
- c) 行为人不能同时做 A 和 B。

为了更好地理解负责任创新，了解道德困境的一些特点则非常重要。具体来说，我们想展示创新和设计如何成为处理道德困境的一种方式。

(1) 处理道德困境

在某种程度上，技术体现了我们的一些价值观，也可以同时质疑我们更渴望哪种价值观。例如以下例子：

- 闭路电视摄像头：我们更重视隐私还是更重视安全？
- 核电站：我们是想要能源安全和低二氧化碳排放，还是想要更少的辐射风险？
- 无人机：我们是希望士兵安全还是负责任？

这些困境是典型的道德困境。任何面临困境的人都有许多义务，但不可能全部履行这些义务。那么你会怎么做呢？

处理道德过载或道德困境有多种策略。其中一种方法是寻找综合考虑后最好的选择。虽然这可以通过不同方式实现，但它通常意味着在各种相关价值承诺之

间进行权衡。换句话说，它会产生“道德残留”。这里的道德残留表示，即使我们在道德过载或道德困境情况下做出合理选择，仍然存在未履行的义务和未实现的价值承诺。

然而，道德残留(或罪恶感)并不仅仅是我们不得不忍受的不幸副产品。哲学家露丝·巴肯·马库斯(Ruth Barcan Marcus)提出了一种关于道德残留或残留遗憾的观点。这种观点可以为道德困境提供一个解决方案。她认为，我们有避免道德困境的二阶义务，即所谓的“应该意味着可以”：如果一个人应该做 X，一个人应该做 Y，那么他可以同时做 X 和 Y。这一原则包含了一项集体责任，即创造一个环境，让我们这个社会能够依据道德义务和道德价值观生活。

(2) 新技术是潜在解决方案

我们能够做到这一点的方法是开发新技术(参阅下述智能电表案例)。技术创新可以通过创造新机会来满足相互冲突的价值观，从而促进道德进步。这是负责任创新的本质。当然，并非所有技术创新都能带来道德进步。

技术创新不一定导致道德进步的一个原因是，它可能导致“技术修复”，即采用技术手段解决本质上属于社会性质的问题。例如，世界饥饿主要不是生产能力的问题，而是粮食分配、腐败、收入、灾害和土地的问题，这些问题很难通过技术解决。

还应该意识到，技术创新不仅扩大了选择范围，而且新选择也带来了新的副作用和风险。这可能会引入新的价值维度，在选择情况下应该考虑这些新的价值维度，这些新的价值维度可能会产生新的道德过载形式。核能可能有助于减少温室气体排放，同时提供一种可靠的能源，但由于需要储存数千年的放射性废物，它也给予子孙后代带来了长期风险。技术革新也可能在以前没有选择的情况下引入困难的伦理选择。产前诊断就是一个例子。这项技术创造了一种可能，即预测尚未出生的孩子患上某种严重疾病的概率。这就提出了一个问题，在某些情况下是否需要或允许堕胎。这种选择情境的特点是生命价值(即便这个生命并不完美)和避免不必要痛苦价值之间的冲突。

(3) 对工程师责任的启示

确保应当做的事情得以完成的更高层次道德义务，可以理解为工程师任务责任的一个重要方面。这也被称为元任务责任(meta-task responsibility)。实现这一职责的有趣方法是价值敏感设计。本书将在第 9 章讨论这个问题。

案例 1：智能电表与价值观冲突——创新的机遇

智能电表是一个很好的设计，因为它可以使我们更加可持续，但它会对人们的隐私产生影响。因此，必须设计出满足所有功能需求的电表，从而服务于可持续发展目标。但除此之外，还需要尊重与家庭用电量有关的隐私，以及我们日常出入的信息。智能电表问题的两难结构如下：

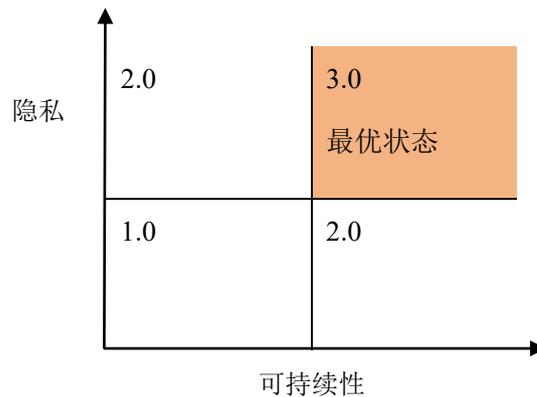


图 2-9 满足不同的需求

图 2-9 中的深色区域就是我们对理想道德解感兴趣的区域。在这一区域，可以满足两个值，超过某个合理的阈值水平。智能电表 1.0 可能既不能带来想要的隐私，也不能带来想要的可持续性。智能电表 2.0 可能会满足其中一个需求，但不能同时满足两个需求。智能电表 3.0，是我们寻求的理想型，旨在满足这两个功能要求，以更有效使用能源和保护个人数据。智能电表 3.0 同时满足了隐私和可持续的需求。在这方面，智能电表的创新是：在智能设计中，协调一系列(相互冲突的)价值观或道德要求。

同样地，如果想从零售中的射频识别技术(自动识别和跟踪附着在物体上的标签)获益，但又担心在购物中心被跟踪，那么也可以探索两全其美的办法。一种带有清晰凹痕的、价格标签形式的“剪切芯片”，允许客户撕下一块标签，缩短标签中的天线，从而限制标签传输数据的范围。除此外，还有很多例子可以说明如何认真对待道德义务——对客户、未来用户、子孙后代、气候，甚至对动植物。在面对艰难的道德选择时，不是通过妥协价值承诺或更多的哲学功课来改变世界，而是通过运用创造力、知识和技能来改变世界。道德困境可以激发创造力和创新，而创新设计则可以帮助我们克服道德过载问题。

案例思考题：

1. 智能电表在提高能源使用效率的同时，如何有效保护个人隐私？请结合图 2-9 中的理想道德解区域，分析智能电表 3.0 应如何设计，以满足隐私和可持续性的双重需求。
2. 请阐述智能电表设计过程中可能面临的道德困境，并讨论这些困境如何影响设计决策，以及针对这些道德困境，提出具体的创新策略或设计解决方案。

案例 2：人工智能和大数据时代的医学伦理

2018 年 3 月 23 日杰伦·范·霍文（Jeroen van Hoven）在塞内加尔达喀尔参加了世界卫生组织“生物伦理、可持续发展和社会”全球峰会，并发表了演讲。本案例以此为基础。

正如本章所讨论的，负责任创新的核心思想是，试图通过设计、改造世界、创新和创造力，尽可能地包容人们的道德将大数据和人工智能应用于医疗领域价值观。这也适用于大数据和人工智能时代的医学伦理。我们应该兼顾道德关切，同时在医疗保健中利用大数据和人工智能。没有人保证这总是可行的。但由于风险很高，我们有义务探索是否存在合适的解决方案。

1. 数字技术及其对医疗保健的影响

数字技术影响着医疗保健的方方面面：研发、临床实践、政策、创新、创业、保险和融资。数字技术改变了一切。它不仅能够实现新的实践，而且也是一种构成技术。显然，大数据和人工智能可以降低医疗保健成本，改善患者安全，赋予患者权力，提高诊断、治疗和患者就医的质量，并创造更高效的计费和物流方式。带有健康应用程序的智能手机、智能手表、可穿戴设备和数字医疗设备是医疗领域正在进行的物联网革命的一部分。可穿戴设备可用于检测心律失常，预测帕金森，并测量一系列生物标志物，如血糖、血压、脂肪百分比、氧气和压力水平等等。它们可以诊断皮肤癌和视网膜损伤，并协助治疗饮食失调、恐惧症、抑郁症、慢性疼痛和创伤后应激障碍。它们甚至可以根据社交媒体上的帖子来判断自杀风险——看帖子发布时间、照片中人脸的数量和颜色。

大型科技公司正在大举进军医疗保健和生物医药领域。在这个领域中可以赚很多钱，收集很多数据。反过来，大数据将推动更强大的机器学习和人工智能发

展，从而在市场上占据优势地位。IBM 将沃森（Watson）技术应用于肿瘤研究，并使用苹果的 ResearchKit 和 HealthKit 收集健康数据。Alphabet 集团通过与 Deep Mind Health 合作，进而专注于人工智能健康领域。该公司已与英国国家医疗服务体系（NHS）和个别医院达成协议，以获得来自英国的大量患者数据，并试图取得突破，堪比在国际象棋和围棋比赛中击败人类世界冠军。Alphabet 集团已经声称，能够比使用传统方法的临床医生更早地预测住院患者的死亡。微软刚刚在剑桥成立了新的医疗保健部门，用于研究医疗算法。苹果选择了硬件路线，并与斯坦福大学合作，专注于智能手机和可穿戴传感器作为医疗设备。亚马逊与沃伦·巴菲特的投资公司 Berkshire Hathaway 公司以及投资银行摩根大通合作，进军医疗保健领域。

然而，纯数据驱动方法和美化统计相关性存在严重局限。在医学和其他具有重大社会意义的领域，数据驱动方法虽然在临床上有用、在道德上负责任，但需要以旨在揭示因果机制的理论驱动方法作为补充。

2. 态度问题

另一个问题是由科技公司对医疗保健的态度造成的。数字行业和硅谷在医疗保健方面采取了一种以解决方案为中心的方法，该方法只关注那些有干净利落技术解决方案的问题。大卫·雷泽（David Lazer）将他们的方法称作“大数据傲慢（Big Data Hubris）”：在非常复杂的医疗保健中，通过简单的数字解决方案来解决复杂问题，医疗保健机构非常复杂，利益相关者众多，道德价值观多元，文化多样性很大。这种态度简直是应该受到谴责的幼稚行为。

3. 道德问题

在卫生领域的数字篡夺中，不仅存在认知上的失败；还有道德方面的担忧。我们知道，在医疗保健数据中存在种族和性别以及许多其他方面的偏见，这些偏见可能会在算法中根深蒂固，甚至被内置到医疗系统中，目的是为了省钱或盈利。例如，美国的 Aspire Health 公司试图通过估计哪些病人将很快死亡来节省治疗的资金。此外，决策支持系统中的算法会影响医患之间的信任关系。此外，在过去十年中，医疗保健领域发生了大规模的安全和隐私侵犯事件。Deep Mind Health 因处理 NHS 患者数据而受到英国信息委员会的谴责。Deep Mind 回答说：“它低估了 NHS 的复杂性和患者数据规则的复杂性，以及对一家知名科技公司在医疗领域工作的潜在担忧。”

4. 一切与信任有关

人工智能和大数据应用到医疗领域的关键问题是，我们为什么要信任 Facebook、优步、谷歌、亚马逊和微软，让它们保管所有的敏感医疗数据呢？他们甚至无法解决假新闻、数据安全、过滤泡沫和偏见等基本问题，也无法防止 5000 万用户的数据被滥用于政治活动。大型科技公司本质上关注的是收入。这些公司以硅谷的创新方式进入医疗保健领域：在灰色地带创新，快速行动，先打破常规，然后再道歉。在医疗保健中，这不是一个很有用的方法。

这就提出了 21 世纪最重要和最困难的问题之一：信任，或者更确切地说是，缺乏信任。如果我们能信任别人，把数据交给他们，我们能从中受益；如果我们不信任或者信任错位，代价也将是巨大的。我们能把健康数据托付给大型科技公司及其子公司吗？我们能保证他们的服务不会其他国家征用吗？我们能确定它们（以及我们的数据）不会与那些不受法治约束的寡头公司和数据库合并吗？

正是在这样的背景下，我们需要对身份相关数据的共享和使用进行讨论。数字世界的伦理应该是关于设计事物、系统、设备、算法、治理系统和协议的组合。如果我们不能自觉地、仔细地设计，在技术时代维护共同的道德价值观，那么隐私、问责、民主、自治、安全和保障等将不复存在。在大数据和人工智能时代，医学伦理如果想要拯救人类的生命和尊严，就需要改变方向，并保持相关性。在利润最大化压力下，我们不能在道德、人类福祉、人类尊严和公共利益上妥协。当其发挥实际作用时，再加入道德就已经太晚了。

讨论题：

1. 你能否想到现实中的一些两难困境问题或关注点吗？我们能否通过引入创新，以某种方式来有效地解决这些困境？
2. 你能想到你所在领域的其他道德困境吗？并简要解释和讨论。
3. 尝试解决思想实验中提到的所有困境，如“电车难题”或“胖子问题”是否有用？为什么？
4. 你能想到其他合作方案来解决“公地悲剧”的其他情形吗？
5. 道德动机为何会起作用，又为何可能失效？

第三部分 制度与价值观

3.创新的制度背景

“不以规矩，不能成方圆；无五音难正六律。”

——《孟子·离娄上》

正在发展和实施一项技术的制度背景非常重要，因为价值不仅嵌入技术，也嵌入制度背景。本节讨论的重点是具有空间影响的技术项目，例如，基础设施项目（如修建道路和堤坝）、能源项目（如风力发电和天然气生产）。由于公众历来对修建铁路、输电线路、碳捕获和存储、废物处理设施等持有不同看法，许多这样的项目都必须处理好公众的接受度问题。

3.1 实质性价值观和程序性价值观

负责任创新的主张之一是，如果这些项目设计方式更容易被接受，并且对相关价值更敏感，这能增加公众对此类项目的支持。然而，基于技术的价值敏感设计不足以开发可接受或已接受的项目。这需要考虑这样一个事实：除了所谓的实质性价值之外——即与技术本身相关的价值（如安全或效率），程序性价值也可以决定一项技术的可接受性。

程序性价值观表示在特定政策环境中执行决策和项目的方式。科技研究领域的文献表明，人们对新技术的反应在很大程度上取决于公众知情度和参与度。

这意味着，新能源项目的可接受性不仅取决于技术本身的特点，还取决于决策程序的特点。比如透明度、公平和程序公正等价值观最为重要。程序性价值观的重要性表明，为负责任创新而进行的价值敏感设计，不应局限于技术的技术设计，还需要考虑更广泛的范围。

3.2 制度及其价值观

价值观不仅存在于技术中，还存在于创新发展及相关规章制度中。因此，将讨论范围扩展到技术或机构以外是有意义的。我们所说的机构是能够限制和促进某些行为的“游戏规则”，而这些规则通常被称为制度。对此，杰夫·霍志森（Jeff

Hodgson) 给出的定义很有启发:

制度是既定和嵌入的社会规则体系, 这些规则构成了社会互动的结构。制度既可以是正式的, 也可以是非正式的。正式制度如法律、标准、法规和合同, 非正式制度可以是习俗、传统和惯例, 两者都体现了一定的价值观。这一点对正式制度来说最为明显。比如, 为了维护环境健康和安全价值, 法律规定项目开发商必须对其计划的项目进行环境影响评估。值得注意的是, 在争议中, 诸如环境影响评估的制度规则往往成为激励辩论的焦点。例如, 人们并不总是同意评估的范围。通过专注一套特定的价值观——在这个案例中是环境健康和安全——这些价值观在决策过程中能被优先考虑, 而牺牲了其他价值观(如可负担性或可获得性)。这表明嵌入制度的价值观是如何与技术的可接受性交织在一起的。

非正式制度可能不那么具体, 但它们同样体现了某些价值观。例如, 惯例代表不需要太多精力就能重复执行的做事方法。非正式制度使有效行为成为可能, 但也存在缺点, 即它们可能偏爱某些不言而喻的价值观。

术语的使用就是一个很好的例子。重复使用特定的单词或缩写可能会变成一种有效的术语, 以便于同伴之间简单快速的交流。然而, 它也作为一种机制, 排除了不熟悉这些术语的人。因此, 这可能会妨碍公众参与决策。由此所产生的风险是某些公共价值观最终没有体现出来。

3.3 解释和设计公共价值观

如果我们想设计价值观, 这意味着我们不仅要考虑技术设计, 还应该考虑如何设计或重新设计制度, 以适应不同的价值观。分析人员的任务是确定根植于正式和非正式制度中的价值观, 以及这些价值观之间的潜在冲突。这暗示了对一个广泛实证领域的研究: 不仅是不同国家层面的法律框架, 也包括公民社会、工业和政策各个领域的战略、文化和惯例。

制度背景并非静止不变, 而是随着时间和地点变化。这意味着技术的可接受性——或者被认为是可接受的东西——也会随着时间的推移而变化, 并取决于技术的开发和实施环境。例如, 在荷兰, 洪水安全的价值目前正在重新制定, 这是对洪水威胁感知的变化以及对抗洪高堤坝接受程度的双重反应。这表明, 无论是价值观还是价值观的表现形式, 都不能想当然地认为是理所当然的。事实上, 价值观在技术的开发和实施过程中会不断出现并发生转变。

如果我们想设计价值观，这首先意味着我们不能只依赖于对相关价值的事前评估。相反，它需要对公共价值进行持续不断的评估，以确保也能考虑到新出现的价值。其次，这意味着设计只有在适应当前的空间、时间和文化等环境时，才能具有价值敏感性。特定技术的价值敏感设计没有固定蓝图。如果我们真的想设计价值观，就必须走出去并使用社会科学强调的方法与人们沟通，以了解这项技术对他们意味着什么、技术如何影响他们、对他们有什么风险、他们想要如何参与等等。

3.4 了解开发人员和决策者的价值观

本章已经谈到了公众及其价值观。然而，为了全面理解如何实施价值敏感设计（VSD，详细内容参阅本书最后一章），考虑技术开发人员和/或相关决策者的价值观和信念也很重要。正如我们所知，因为公众对技术的反应方式在很大程度上取决于技术开发者或政策制定者与他们沟通的方式。

以一个简要例子为例，项目开发者经常使用一个标签来描述公众对新技术的反对，即 NIMBY（Not In My Backyard 的缩写）。NIMBY 的标签表明，人们反对新技术是因为他们把个人短期利益（例如安静和美观的生活环境）置于长远集体利益之上（例如风力涡轮机的安全能源供应），这可能并不总是一个视觉上令人愉悦的现象。这与根深蒂固的信念紧密相关，即公共缺乏信息且规避风险。正是这种信念塑造了项目开发者与公众的互动方式。如果项目开发人员认为公众消息不灵通，那么在他的沟通中，他可能会重点提供技术事实并解释项目安全。

然而，如前所述，公众可能更关注程序性问题，如公平、透明度、成本和利益分配。这些问题无法通过提供更多技术和风险信息来解决。双方假设的不匹配阻碍了沟通过程，从未导致一个悖论，即通过提供“确凿事实”来阻止反对的努力可能引起更多公众反对，因为公众感到他们的担忧——价值观——没有得到认真对待。

因此，为价值观而设计意味着，技术开发人员需要思考并反思自己的信念和价值观，以调查这些假设（信念、价值观等）如何引导其与其他利益相关者互动。接受这种反思的需求势在必行，认可这一点则意味着要认识到科技领域存在一系列不同的价值观和问题定义。

3.5 创新中的制度价值观解释

目前为止，我们应该清楚，价值敏感设计不仅与技术本身相关，还与制度重新设计相关。以下列出 4 个对制度价值观进行设计的行动要点。

- 价值敏感设计与技术和制度有关；
- 价值敏感设计需要持续评估公众价值观；
- 设计应适应特定的环境，而不是以某个标准蓝图为基础；
- 设计应将设计师、技术开发人员和政策制定者的反思考虑在内。

上述内容表明，许多情况都会涉及不同类型的制度。奥利弗·威廉姆森（Oliver Williamson）在四层级制度模型中给出了不同制度类型，如图 3-1 所示。

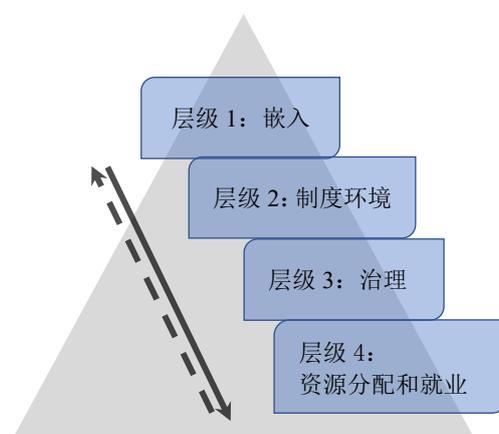


图 3-1 四层级的制度类型模型

（1）层级 1：嵌入

嵌入通常指非正式的机构、习俗、传统、规范和宗教。如前所述，传统上在某些区域捕鱼的渔民，或每年习惯在某些地方露营的公民，即使这些权利没有被特定规章制度正式证实，他们也会这样做。这就是“嵌入”。非正式规则的变动频率通常很低。威廉姆森甚至指出，变化会在一百到一千年中发生一次；这表明这些传统深深根植于个人（或行动者）行为之中。其目的通常不需要计算。非正式规则会自主演化发展并且很难被计划。相应地，试图影响非正式规则也非常困难。另外，如果我们想要追求负责任创新，这一层则非常重要。

（2）层级 2：制度环境

制度环境是正式的“游戏规则”。例如，主权国家的宪法，或在这种法律下建立海上风力发电场所依据的能源政策。该层级的变化频率大约每十年到一百年

一次。所以，这些正式规则相对稳定。我们还可以识别制度环境背后的具体目标。从经济学角度看，其目标可能是创造正确的制度环境，威廉姆森称之为“一阶节约（first-order economizing）”。一旦我们确定哪些制度最能促进海上风能发展，我们就可以为建立风电场设计具体的、正式的规则。

（3）层级 3：治理

治理，也可以理解为“游戏的玩法”。基于到嵌入和制度环境，行动者可以选择什么样的合同或法定组织形式进行正确治理，以实现他们的目标？什么样的合同和组织结构最适合服务于企业的个人目标和其他行动者的目标？这种考虑就是威廉姆森所说的二阶节约（second-order economizing）。这些治理结构可能在一到十年内变化一次，其变化频率比上述层级快很多。

（4）层级 4：资源分配和就业

层级 4 代表与资源分配和就业相关规则、法规的持续变化。这些都是利益相关者和行动者为了获得边际条件而进行的日常活动。这些个体和集体的持续互动共同塑造了制度在实践中的运作方式。这就是三阶节约（third-order economizing）。

（5）四层级制度模型的应用

图 3-1 展示了制度的四个不同层次。有趣的是，这些不同层级存在从上到下和从下到上箭头。图中的箭头表明，不能孤立地分析不同层级。例如，一个国家或地区有某些特定规则或非正式制度，这些规则和非正式制度必须受制度环境保护。这也意味着，“游戏”的正式规则在一定程度上是基于非正式制度的。否则，我们将面临一个严重问题，因为正式制度对相关社区来说将不可信或不相关。我们需要协调非正式制度和正式制度，如果进一步深入研究这四个层级，还可以认为治理和资源分配需要相互协调。

因此，不同层次的制度由特定逻辑构成，并以特定方式建立起来，否则我们就会受到干扰。但要注意的是，图中还存在自下向上的虚线箭头，表明下层对上层会产生反向影响。比如，资源分配或治理发生变化，会影响正式规则和非正式制度发生变化。

如果以能源部门为例，目前人们对分散式能源的生产和消费投入了很大的关注，甚至深入到家庭层面。屋顶的太阳能电池板不仅被家庭使用，多余电量则被传输到电网中。可以认为，这些地方层面的举措足以改变能源部门的治理方式。同样，治理的任何变化都要求制度环境适应能源生产和消费的新做法，而这些新

做法反过来则可能影响非正式制度。家庭可能认为发电不再是国家应该提供的公用事产品，而是他们可以自助去做的事情。这将促使非正式制度以及能源生产和使用相关的价值观发生变革。

如果考虑到负责任创新，不同层级制度之间的相互关系非常重要。这不仅仅是自上而下的活动，也是个体用户自下而上的发展，推动治理、制度环境以及固有制度、价值观和习俗的变革。在这个方面，制度背景对理解和塑造新创新的成败至关重要。

案例 1：北海风能

海上风能是电力系统的一种创新。图 3-2 展示了荷兰靠近北海地区的地图及该地区的具体用途。

图中不同的线条和点表明，北海不仅是一个空旷地区，还具有多种用途，如娱乐、海军运输、渔业、电缆，甚至军事。该地区引进地风能生产日渐增多。由于北海有许多不同用户，这些用户自然要遵守一定的社会规则。这在多数情况下只是实际性的问题：在非常靠近海军运输路线或风力发电的地区进行军事演习是没有用的。

1. 正式和非正式规则的解释

我们需要将某些领域用于特定目的，以确保所有利益相关者能充分利用北海，同时不侵犯其他用户的需求。换句话说，我们需要社会规则来规范如何将这个空间用于不同目的。这些社会规则可以是明确建立的，也可以是隐含其中的。

如果我们想利用北海进行风力发电，就需要明确地将北海某些特定地区用于生产风力发电；如果我们想将北海用于军事目的，就需要明确军队可以进行演习的特定区域。这就意味着，需要有正式规则来规定将这些领域用于特定目的。

不过，这也可能存在一些不成文（约定俗成）的社会规则。北海渔民可能有着世代相传的悠久传统，即在特定区域捕捞鲱鱼或其他鱼类。此外，还有一些非正式规范来明确某些地区被用于娱乐目的。诸如此类的社会规则并非以正式规章制度为基础，而是以人们长期持有的传统为基础（即前述所讲的层级 1“嵌入”）。



图 3-2 荷兰近海（北海）地图及其各种用途

基于这种对社会规则的理解，回到在北海建造海上风力涡轮机的可能性上来。我们可能不得不改变有利于海军运输或渔业的现有规则，并重新分配一些区域用于风力发电。这不仅是为风力涡轮机分配一个特定区域，因为其他利益相关者可能会受到新规则的约束，我们还需要了解如何纠正。例如，如果为某些区域分配风力发电，渔民可能就无法在这些区域很好捕鱼，海军运输也需要规划其他路线，以及该区域附近的军事演习也不允许。这表明，我们不是在讨论单一规则，而是考虑一个由不同社会规则所构成的系统，该系统既规定又描述了社会行为。

案例思考题：

1. 在北海地区实施海上风能发电，如何平衡不同利益相关者的需求和权益，同时确保遵守正式和非正式的社会规则？
2. 面对北海地区的多用途和复杂的社会规则系统，如何设计并实施有效的社会规则调整策略，以促进海上风能发电的发展？

案例 2：自动驾驶

2018 年，出行公司 Uber 的一辆自动驾驶汽车，在美国亚利桑那州 Tempe 市的街道行驶时，撞倒了一位名 “Elaine Herzberg 的行人，后者不幸死亡。这是全球第一起自动驾驶汽车交通致死案。调查结果显示，Uber 自动驾驶汽车内的安全员事发时正在收看电视节目《美国之声》，没有注意到正在横穿马路的行人，于改事故发生。Uber 并没有因地被提起刑事诉讼，但安全员后来被控过失杀人。这一事件引发了社会广泛讨论。人们对其讨论的焦点主要围绕自动驾驶汽车对拥堵、出行时间和安全的影响——亦即效用问题，而对伦理问题的讨论并不多。本案例旨在讨论自动驾驶伦理方面的关键问题。那么，自动驾驶的重要伦理问题有哪些呢？

1. “多只手”问题

“多只手”问题是由不同参与者在车辆设计和部署中扮演不同角色所造成的。这引起了人们对任意一个行动者责任的关注。假如，自动驾驶汽车由于传感器故障引起交通事故，而部分原因与恶劣天气有关。那么，谁该为此负责呢？司机？汽车制造商？提供传感器的公司？提供关键软件的公司？车辆维修工作的经销商？或者允许这些车辆在恶劣天气上路的相关部门？

即使一个行为人被确认负有法律责任，也并不意味着其他行为人没参与其中，他们的责任问题也远未得到解决。为便于讨论，可以先假设司机负有责任。自然地，关于司机的责任将不可避免地引发一场争论，因为司机觉得他们没有任何不当行为。

2. “电车问题”

我们还可以思考，即使最复杂的智能算法该如何处理“电车问题”？如果一辆迎面而来的自动驾驶汽车撞击在轨道上，且只能做出唯一选择：“撞车，危及两个司机”或“突然转弯，危及附近行人”。面对这种情况，自动驾驶汽车应该怎么做？自动驾驶汽车很可能面临这种情况，需要在两种场景中做出选择。无论什么选择，结果都是基于它所编写的指令。

从结果主义角度看，撞到行人甚至杀人，都将会是首选，因为这种选择下只有一个人面临危险。但从康德的道德观点来看，这并非首选，更不用说行人在自动驾驶汽车（无处不在）的情况下可能会改变他们的行为。

3. 效用分配 (Distribution of utility)

第三个问题是行车时间、安全和可持续性之间的潜在权衡。从这三个角度中的任何一个优化系统都可能产生不同结果。从安全角度看，最好保持更长车距，但这种安排会由于更多空气阻力而消耗更多燃料；此外，道路也不能被最佳利用，会导致更多拥堵和更长行车时间。或者，假设自动驾驶汽车以时速 160 km 短距离行驶且没有任何风险。这能缩短行车时间，但也会增加二氧化碳排放量；或者，考虑到自动驾驶汽车更加方便，甚至成为长途旅行交通工具的首选，这会增加二氧化碳排放量，也会间接导致城市扩张和土地需求增加。即使在目前现状下，诸如此类的多样性权衡任然存在。

4. 经济差距

自动驾驶汽车引发的第四问题是公平分配问题。至少在一开始，自动驾驶汽车比普通汽车更贵。专家预测，自动驾驶汽车的成本要比同类普通汽车至少高出 1 万欧元。虽然我们可以认为考虑到自动驾驶的好处，价格不是问题。然而，如果道路管理部门专门为自动驾驶分配专用道路，这会减少普通车辆的可用道路容量，从而导致更严重拥堵。最终，富人购买昂贵的自动驾驶汽车，会间接惠及社会阶层较高的人，但也牺牲了较低社会阶层的利益。

5. 公共交通需求减少

自动驾驶汽车让个人拥有汽车更具吸引力。一方面，公共交通的优势之一是人们能在上下班途中工作、阅读或睡觉，比如乘坐火车；另一方面，私家车提供了按需出行的选择。而自动驾驶汽车可以提供两全其美的服务，既可以作为个性化的按需交通工具，又可以将驾驶员解放出来。从长远来看，人们可能会大规模地从公共交通转向自动驾驶汽车，从而加剧高速公路拥堵、污染和碳排放等问题。此外，公共交通需求下降会对社会经济水平较低的人群造成影响，因为这个群体最依赖公共交通，他们根本无法负担得起自动驾驶汽车。

6. 一些主要优势

(1) 便利性增强

首先，在农村和城郊，由于难以获得教育、经济机会、医疗服务，甚至与远方亲友保持联系，会导致缺乏人际交往的个体被社会排斥。在这种情况下，自动驾驶汽车——也许是通过汽车共享等方案——可以提供基本的交通选择，减少经济、教育和社会的排斥程度。

(2) 安全性增加

从长远来看，自动驾驶汽车不仅可以显著提高使用者的安全性，还可以显著提高行人、自行车（摩托车）和其他司机等非使用者的安全性。另一个被忽视的领域是，减少因酒驾或服用其他药物导致的交通事故，这能够带来立即且显著的安全效益。

（3）降低污染排放

自动驾驶汽车设计的更合理、操作更精确，能直接减少总体能源使用和排放量；人们使用燃油效率高的自动驾驶汽车，也能够间接减少能源使用和排放量。毕竟，拥有一辆功能强大但效率低下的汽车没有任何意义。

7. 过渡问题

传统汽车和自动驾驶汽车之间的过渡期也非常重要。专家认为，自动驾驶汽车首次进入市场后，道路容量可能会立即下降，道路安全性也会下降。由于“成长的烦恼”、技术故障，或人们对真实世界的学习周期，这种情况是可以预见的。不过，最终而言，自动驾驶汽车将会更加高效，并解决高速公路运力问题以及提高安全性。换句话说，自动驾驶汽车运营的初期可能会降低交通系统性能，但随着自动驾驶汽车学习效应和市场渗透率的不断增加，系统可以随时间推移而得到改善，并最终超越现在的性能。当然，这同时也会产生跨时间的伦理问题。

另外，我们还可以认为，与人类驾驶员相比，自动驾驶汽车的相对安全性会促使汽车保险产生一种有趣的权衡。随时间推移，如果自动驾驶汽车始终比司机发生更少、更轻的事故，那么自动驾驶汽车的保险费会变得更低。从长远来看，有效使用自动驾驶汽车会成为首选投资，而传统汽车定价则纯粹出于经济考虑。

总结：保持谨慎乐观

本案例研究重点是提醒人们不要对复杂性和创新持有过于保守的态度，尤其涉及自动驾驶等技术时。比如，假设只有自动驾驶汽车，没有传统汽车，此时如果引入传统汽车并禁止自动驾驶汽车，我们可以轻易提出理由来反对这一举措，比如它可能造成伦理后果，包括司机、行人和其他人的安全降低，效率下降，社会中贫穷群体受到更大排斥，以及更高的排放和能源消耗等等。因此，即使在讨论自动驾驶汽车复杂且引人深思的伦理问题时，也没有理由先验地从伦理角度认为引入自动驾驶汽车不可取。

讨论题：

1. 负责任创新的实质性和程序性价值观有哪些？
2. 结合中国社会价值观，讨论企业技术开发人员和决策者在新技术、新产品开发时，如何解释相关价值观？面临哪些价值观冲突？
3. 制度与价值观的区别？
4. 以新能源汽车为例，分析该产品开发涉及的不同制度类型，以及各层级制度类型间的关系。

第四部分 创新与管理

4.创新与商业

“千里之行，始于足下；万丈高楼，起于垒土。”

——《老子》

4.1 渐进式和突破性创新

前述章节对创新给出了界定。创新是一种活动或过程，能产生以往未知的设计，涉及物质世界(如建筑物或基础设施)，概念世界(如概念框架、数学、逻辑和软件等)，制度世界(如社会和法律制度，程序和组织)，或这些方面的组合，当把它们付诸实施时，可以扩大我们解决问题的选择。

(1) 创新的分类

技术创新有不同种类，其中经常需要区分的是产品创新和工艺创新。产品创新是对产品设计的改进，工艺创新则与生产过程本身的变化有关。手机上的一项新功能是产品创新，而有效地组装手机的新型机器则工艺创新。

在技术创新中，另一种需要区分的是渐进式创新和激进式创新。创新在很多方面可以是激进的：基于新操作原则，基于新科学知识，提供新功能，接触新用户群体，或者服务于新价值观等等。



图 4-1 Abernathy 和 Clark 的创新分类法

Abernathy 和 Clark (1985) 开发的创新分类法，基于两个维度对创新进行分类：创新是基于现有知识还是新知识 VS. 创新是针对现有用户还是针对新用户。这两个维度相互组合，形成了 2×2 的四象限分类，如图 4-1 所以。

首先，常规或渐进式创新。该类创新建立在现有知识基础上，并以现有客户为目标。典型例子是一款新型号手机，它每年都在开发和更新。

第二，利基创新。这类创新建立在现有知识的基础上，但以新客户或新市场为目标。典型例子是为自行车骑手设计的 GPS 设备。

第三，革命性创新。即针对现有客户和基于新知识的创新。电动汽车就是一个很好的例子。

第四，结构创新。该创新以新知识为基础，以新市场、新客户为目标。典型例子如福特 T 型车、电视、第一架战斗机、化肥、互联网、智能电网和城市、纳米技术等等。

(2) 激进创新与负责任创新之间的联系

结构创新具有一些可识别特征。首先，结构创新只是偶尔发生。其次，它们为渐进式创新创造了基础。第三，它们通常由局外人发起，如新企业或成熟企业，这些企业通常会打破现有知识和市场关系。苹果公司依靠 iPhone 进入手机市场就是最典型的例子。那么，负责任创新需要激进式创新吗？为回答这一问题，需

要再次回顾负责任创新的定义。

负责任创新表示，在实施时能增加解决道德问题可用选项集合的创新。

根据这一定义，四种类型的创新都能扩展选项集合。因此，所有类型的创新都能够负责任。然而，负责任创新往往需要激进创新。为什么如此？为确保负责任创新，我们需要在设计过程中将价值观考虑在内。通常我们需要考虑以前未曾考虑过的价值观，而考虑这些新价值往往需要新知识。例如，如果考虑智能电表的隐私，就需要了解什么是隐私，还需要了解如何将隐私转化为智能电表设计。

考虑新价值观也会对用户关系产生影响。它意味着，先前设计所满足的功能需求得到了扩展。因此，负责任创新意味着参与新市场和增加新功能的机会。这表明负责任创新通常与结构创新相似。弄清这一事实还需要更多实证研究，但如果这是真的，将会产生一些有趣含义。例如，市场领先的公司不总是率先提出负责任创新；相反，这种倡议通常会来自外部人士或新进入者。

(3) 激进创新的伦理思考

激进创新是否会带来新的伦理问题？的确会。这是因为，激进创新不仅是用新方式和商业手段做事，还是在总体上做新的事情。比如互联网、智能手机、航空运输或产前诊断，所有这些技术都为创新行动创造了新的可能，因此，它们提出了新的伦理问题。例如，互联网引发的隐私问题。再比如产前诊断。它可以让我们有能力预测孩子患上某种疾病且无法治愈的可能性。这些信息给父母和社会提出了全新伦理问题，父母也将面临堕胎与否的选择。整个社会应该讨论的是，更多患有综合症的胎儿被打掉是否成为问题——我们是否应该致力于消除这些综合症呢？

对于现有技术而言，往往存在道德习俗和规则的约束。在渐进创新情况下，这些规则和习俗通常足够可用。然而在激进创新情况下，同样的规则和习俗往往不足以应对新情况。一个很好的例子就是在民用航空中引入喷气发动机。这是一项突破性创新，喷气式发动机问世不久，两架名为哈维兰彗星的飞机就坠毁了。其问题不在于发动机本身，而在于当时喷气式飞机比其他飞机飞得高很多，为使乘客飞行舒适，机舱必须加压，结果机身的一些点受到比以前更大的压力，最终导致金属疲劳、造成灾难。这为激进式创新带来了警示：负责任创新往往需要激进创新，而这反过来又经常引发新的伦理问题。

4.2 创新的决定因素

了解了什么是创新之后,接下来让我们了解什么因素决定了技术创新能够成功。首先,必须说明创新参与者是谁以及他们的动机是什么。其次,必须了解如何扩大创新。第三,关注哪些决定因素或激励因素影响以盈利为目的的私营企业创新绩效。

(1) 创新参与者及其动机

谁是创新的参与者呢?由于创新是一项人类活动,最直接的答案则是个体发明者。个体发明者是有创造力的人,他们受内在动机激励。也就是说,他们的创新动力来源于解决特定技术问题的个人兴趣。结合个人能力或创造力,他们用新方法或新答案解决技术问题。发明电灯泡的托马斯·爱迪生(Thomas Edison),以及发明柴油机的鲁道夫·迪塞尔(Rudolf Diesel)就是生动例子。

为了扩大创新的生产规模,有必要将有创造力的人聚集在一个组织中,并以最佳方式使用他们的创造力,从而构建整个创新过程。这样做的好处是,组织通常比任何一个发明者都拥有更多资源,因此组织可以用来刺激或扩大创新。例如公共组织(代尔夫特理工大学等)。不过,还需要考虑的是,多数创新都发生在私营的、以营利为目的的公司(如苹果、IBM、飞利浦),以及规模较小、知名度较低的公司,尤其是初创企业。

(2) 创新的经济决定因素

创新过程高度不确定,将创造性想法转化为新颖、有用的东西非常昂贵,而收益则高度不确定。这源于一个事实,即在客户认可新想法之前,新想法应该先证明自己是有用的。许多研究表明,将新的原创想法转化为成功的商业产品机会不到0.1%。这意味着,一千多个新想法中只有一个真正成为成功的商业产品。换句话说,创新活动能否带来更高利润是高度不确定的。成功的技术创新有以下几种决定因素:外部因素,如技术、经济和法律环境;内部因素,则在企业内部发挥很大作用。

首先,企业所处技术环境。即公司所属的行业部门。例如,航空航天行业中一家公司在动态的技术环境中运营,与纺织行业等其他部门相比,保持领先地位则更加迫切、更加必要。

其次,竞争或市场结构组成的经济环境。奥地利裔美国经济学家约瑟夫·熊彼特(Joseph Schumpeter)是最早研究市场结构影响创新的学者之一。他的核心

问题是：在什么样的市场中，企业能实现最高创新绩效？对此存在两种相互矛盾的解释。

- 创新主要在许多竞争激烈的小企业市场中产生，这些小企业被迫创新以保持领先于竞争对手。
- 竞争较少的市场会有更多创新。因为大公司（如飞利浦、联合利华等）有很多可用资源，他们可以参与不确定的创新过程，并且不会在创新失败时立即破产。

实证研究对于这一问题尚无定论。技术环境似乎是一个重要因素。例如，当今的软件业由谷歌、苹果、Facebook 和微软等大公司主导。这些软件巨头有许多可供创新的资源，但这并不意味着他们很容易创新，也不意味着他们的持续利润能得到保证。因为，该行业在技术快速变化的环境中存在激烈竞争。

另一例子则是石油和天然气行业，这一行业也由主要参与者主导，如壳牌，埃克森美孚等。尽管该领域的公司也在创新，但它们之间的竞争没那么激烈，因为该行业的技术环境变化速度比软件行业慢。

第三，协作。在过去 25 年里，技术创新变得越来越复杂、越迅速，也更加国际化。这种发展结果是，对任何一家公司来说成功创新变得更加困难，成本也变得更高。为了找到新想法，公司必须超越自身边界，与供应商、客户和大学等其他参与者合作，以提高他们的创新绩效。协作的典型例子是硅谷的半导体和软件公司。硅谷不同公司的软件开发人员定期见面，喝几杯饮料，并根据他们的知识交换意见。其他例子还有瑞士制表师、米兰时装设计师的会议等。

技术集群合作是协作创新的典型方式，如图 4-2 所示。在集群合作中公司共享技术（例如，软件或生物技术），并与买方、供应商建立互补关系以从事生产活动，此外这些公司也进行合作研究。这样做的原因是，复杂知识通常是隐性的且没有被明确记录下来，它们存在于工程师或开发人员的头脑或行动中，因此需要人们之间频繁而密切的互动。技术集群有很多优势。首先，技术集群能促使技术溢出。一个创新企业的研发收益会产生溢出效应，提高另一个创新企业的收益。其次，促使劳动力在技术上接受良好教育。另一个优势是供应商、分销商以及其他支持公司（会计师、律师）互动交流，增加企业创新的成功机会。



图 4-2 硅谷高科技企业集群（部分）

第四，法律环境。创新精神的个人或公司经常花费大量资金来开发想法并将其转化为具体应用。高昂的开发成本会在产品上市后收回来，然而届时可能存在一种风险，即其他个人或公司可能会采取逆向工程，弄清楚新产品或新系统如何工作、由什么组成、功能怎样。然后，他们不需要像原始发明者那样支付高昂开发成本，就能生产并销售自己的创新产品或创新系统。如果发生这种情况，原始发明者的利润和回报机会就会大幅下降。知识产权就是为了避免这种情况发生。专利法规定的专利就是知识产权的典型例子。原始发明人获得的保护期为 20 年，在此期间，除非原始发明人通过许可获得经济补偿，否则他人不得生产或销售其创新成果。实际上，原始发明者获得了暂时垄断，保证他有可能获得源源不断的收入，以收回巨大的开发成本。政府为专利提供法律保护的动机，是鼓励原始发明者继续创新，从而推动产业和经济向前发展。

第五，内部因素，与组织本身有关。美国经济学家威廉·鲍莫尔（William Baumol）强调，在资本主义社会中创新企业的力量取决于创新的常规化。为了使创新常规化，企业会不断寻找新想法并将其转化为产品或系统，并形成标准化的程序流程。然而，程序和创造力是两种对立力量，毕竟，创造力在本质上不可预测，而制定程序则是为了使创新结果可以预测。这意味着，集权式组织在自上而下的严格管理中形成全面控制，而不是刺激创新生产；反过来，一个完全分

散的组织，每个员工都有创造力，可以做他想做的事，这种情况也很难行得通。因此，双元组织（ambidextrous organization）通常被认为两全其美的组织。在这些组织中，不同单位有不同组织结构。例如，“产生新颖想法的分散单位”与“将想法转化为具体产品并成功销售的集中单位”共存。

4.3 创新管理

在高度竞争环境中，企业要实现盈利。因此，为了管理创新，有必要了解企业创新的产生和扩散过程。管理创新和参与创新过程的方式，在很大程度上取决于我们对这些过程的看法。以技术为基础的产品创新有通讯设备（如电话和电视）、材料（如合成纤维）、药物（如百忧解或阿司匹林）。这些产品在推向市场时，都是基于技术的全新创新。那么，公司如何从这些发明中获利呢？这就需要区分创新和扩散过程的两种不同观点。

（1）创新是一个简单项目

第一种观点将创新过程看作始于新技术出现的新产品开发项目。当新产品准备好投入生产、分销，且市场营销也准备就绪时，项目就算结束了。接下来是市场引入阶段，即扩散过程开始。

从这个角度来看，成功完成整个流程需要不同类型的管理。具体而言，技术开发需要研发管理，而且在随后的产品开发阶段，研发活动将继续参与其中；管理新产品的开发轨迹则需要项目管理；营销战略需要制定市场导入计划，并管理随后的扩散过程，如图 4-3 所示。

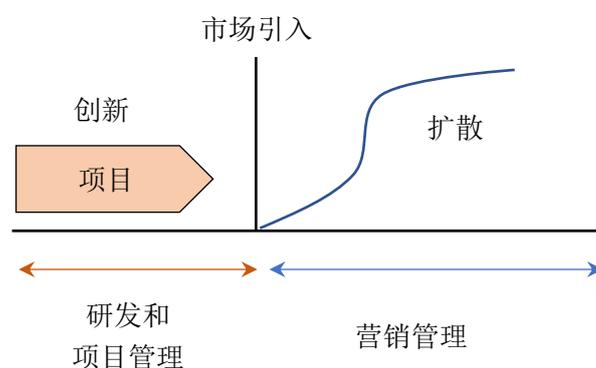


图 4-3 创新是一个简单项目

从 20 世纪 80 年代开始，主流创新管理手册将创新描述为一个项目。从这个角度看，管理工作需要营销和研发团队之间密切互动。两种能力哪一种处于领先

地位取决于公司类型和特定行业，而领导部门通常提供主要项目管理者。这些共同努力的成功，体现在销售大幅增长或产品（技术）引入后的陡峭扩散曲线上。例如，追踪电话和电视扩散趋势，可以发现一条近乎完美的曲线。事实上，这些产品在市场上也相当成功。这种观点在描述趋势、制定创新步骤方面相当敏锐。在创新管理文献中，也存在大量关于营销和研发方面的知识。这种观点一直持续到 21 世纪初。

（2）创新是一个复杂过程

第二种观点则认为创新过程要复杂得多。以下四个方面说明了创新比产品开发项目更复杂。

- 第一，技术开发和产品开发通常同时进行。开发出的第一批产品往往不可靠，只有进一步开发技术才能获得可靠产品。共同开发产品和所需的技术不仅是一个单一项目，而更多是一个复杂且高度相关的项目计划，因此也是一种迭代项目。
- 第二，许多公司同时进行技术和产品开发来相互竞争。一些发明在获得专利后会被其他财团或公司网络使用。在这种情况下，创新过程不仅仅是一个项目；在开放式创新时代，创新是许多独立项目相互关联的过程。
- 第三，创新过程不只是一个项目，因为产品不可能凭空产生。最初市场通常缺乏互补产品和服务，或者尚未获得具有成本效益的生产设施。有时消费者并不真正了解产品，由于所需元素缺失，市场引入则成为试错过程，即使产品已经在市场上扩散，开发也仍在继续。因此，市场开发和产品/技术开发是同时进行的。
- 最后，在某些情况下，一项技术背后的基本科学原理只有在该技术成功应用多年后才变得清晰。例如，在我们了解帆船运动的科学原理之前，帆船就已经被建造了数千年；在我们真正理解并掌握飞行原理之前，飞机也被使用了很多年。

（3）现代创新过程

从创新和扩散过程的更新视角来看，成功完成这一过程需要不同类型的管理。除了研发管理、项目管理和营销管理能力之外，还需要具备新型的能力，具体如图 4-4。

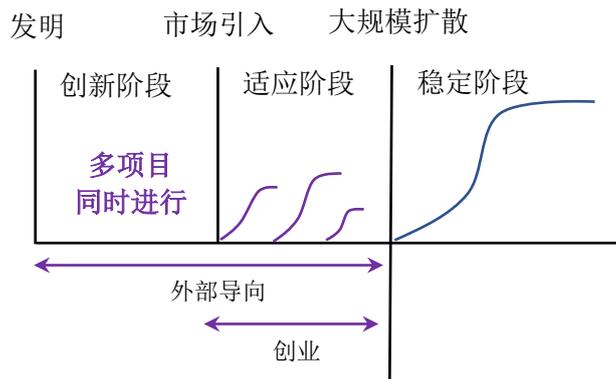


图 4-4 现代创新过程

首先，外部导向能力。公司需要采用外部导向，以便使自身产品开发与合作伙伴公司相关，使互补产品和服务的发展保持一致。这种外部导向也是跟踪竞争对手技术和产品开发活动、了解市场最新发展的必要条件。

其次，创业能力是开发市场的必备能力。第二阶段（即也即适应阶段）通常是许多公司破产或倒闭之前离开市场的阶段。创业者必须对自己的产品有一个愿景，将技术功能与（潜在）市场需求结合起来，从而设计出满足这种需求的产品。他们还必须创造一个可持续的商业模式，使产品商业化。创业者还可以决定引入创新的最佳时机和地点；换言之，他们也可以采取利基战略。

这对负责任创新有一些启示。在适应阶段（即实验和创业阶段，该阶段可以在多个利基市场引入不同产品版本）可能会发生意外，并可能出现意想不到的副作用，此时就需要一种负责任的方法。在稳定阶段，主流应用有时很难预测，这种应用下产品使用的后果也很难预测，同样地，责任地协调利润和后果之间的平衡也非常必要。

案例 1：电视的发展与传播

本案例以一个典型历史案例为主，并将其看作创新扩散过程。目的在于探索，它对有效实现创新扩散过程所需的管理和创业类型有什么启示。

与电视相关的主要技术发明可以追溯到 1925-1930 年。不过，此时并没有立即将电视作为一种产品进行开发。大约过了十年，第一台电视机才问世。显然，将一项发明转化成第一个产品需要几年时间（平均十年）。这是整个创新过程中所谓的第一阶段：创新阶段。

电视刚问世时，它并不是我们如今熟知的那种电器。大约在 1939 年，电视作为一种半公共服务被引入德国和英国酒吧，而且在引入市场后，并没有形成大规模扩散，只能在特定生态位中看到小规模扩散。几乎所有的激进高科技产品创新都存在这种扩散模式。不同产品版本在利基市场的初步小规模扩散，被称为适应阶段。

直到 20 世纪 50 年代（电视机被发明后的 20 多年），电视才真正得到大规模普及。最后这个阶段被称为稳定阶段。

通过分析 1850 年至 2000 年间推出的 100 多个激进高科技产品创新和传播案例，可以认为，就发明和大规模扩散之间的时间间隔而言，电视实际上是一个典型的、平均的例子。

案例 2：冷却剂

制冷的发展历史几乎和人类历史一样古老。在史前时代，人们在深洞里堆冰来冷藏食物，而在中世纪，木墙上的壁龛则具有冷却效果，地窖和冷藏室则成为人们用木桶或陶罐储存食物的空间。从 19 世纪中期开始，氨、二氧化碳、异丁烯等天然物质被当作工业和商业用途的制冷剂。随着氯氟烃（CFCs）在 1929-1930 年被发现，氯氟烃便很快被用作制冷剂，其在全球范围内也取得了显著的市场占有率，见图 4-5。在 20 世纪 60 年代，家用冰箱和其后不久的家用冰柜便成为了欧洲家庭的普通商品。

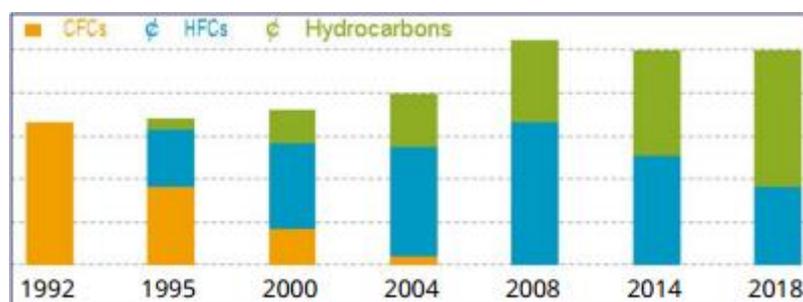


图 4-5 绿色冷冻技术的全球市场渗透率

直到 1973 年，氯氟烃对大气臭氧层的破坏性影响才被发现并得到进一步研究。不过，臭氧消耗和氯氟烃排放之间存在关联的科学证据被否认了。15 年后，维也纳（1985 年）、蒙特利尔（1987 年）、伦敦（1990 年）和哥本哈根（1992

年)的《臭氧议定书》，就通过国际行动逐步淘汰生产和消费臭氧消耗物质达成了政治协议。其中，最主要的是1989年1月1日生效的《蒙特利尔议定书》，推动了工业界对氯氟烃排放的研究。与政治协谈同时进行的是，化学工业开始着手寻找替代氯氟烃的材料和技术。20世纪80年代后期，氢氟碳化物(HFCs)，特别是氢氟碳化物-134a很快被发现和检测，成为最有希望的氯氟烃替代品，因为它们不会损害臭氧层。基于此，人们普遍认为HFC-134a将成为新的通用冷却剂。然而，HFC-134a仍然对全球变暖具有很大威胁，它实际上是一种强大的温室气体，对全球气候变化发挥重要作用。

值得庆幸的是，绿色和平组织在90年代成功推出了一种名为“绿色冷冻(Green Freeze)”的冷却剂替代品。现在许多国家和公司都采用了这种替代品。绿色冷冻使用碳氢化合物，完全不含消耗臭氧层、导致全球变暖的化学物质。这项新技术被迅速传播到欧洲其他国家，它很快便彻底改变了世界范围内的冰箱行业。截止目前，有超过7亿台冰箱采用这种“绿色和平制造(Made in Greenpeace)”技术。不过这一成功来之不易，因为异丁烷虽然对温室效应影响很小，但它易燃，冰箱行业最初对其持反对态度。事实上，现有技术规范甚至禁止使用易燃的冷却剂(如异丁烷)。尽管如此，绿色冷冻还是被接受了，因为德国国际合作组织(GIZ)发现了一些情况：

- 欧洲标准化委员会(CEN)迅速采纳了一项新标准，允许在家用冰箱中使用碳氢化合物。这与对安全价值的重新解读有关。
- 起初，绿色冷冻的安全性是根据这种冷却剂的可燃性来解释的。后来，人们从冰箱的燃点和爆炸角度来重新理解它，发现可燃冷却剂并不像一般认为的那么危险，原因在于现在的冰箱只含少量冷却剂。
- 许多国家经济积极增长，拉动了国内投资，预计家用电器需求将在全球范围内增长。
- 工程师的技术专长，加上管理层愿意冒险，并转向全新且未经充分验证的碳氢化合物技术的决心。
- 早期，各部委和政府机构支持研究和引进碳氢化合物技术。政治措施，例如提高信息和认识(会议和研究)，以及颁发“蓝天使”生态标签，都促进了绿色防冻技术在市场的推广。

- 德国政府的大规模参与进一步促进了全球扩散。
- 20 世纪 90 年代初，消费者具有很高的环境意识，并普遍了解臭氧层的损耗情况。因此，他们愿意接受环保电器，并为环境保护作出贡献。
- 绿色和平组织与利益相关者建立了联系。

讨论题：

1. 你知道有哪些创新从未在市场上出现过，或者出现过但没有成功吗？
2. 除渐进式、突破式创新外，你还知道哪些创新类型？它们的区别是什么？
3. 影响创新的因素还有哪些？
4. 在什么情况下，可以将创新看作一个简单项目/复杂过程/现代过程？并举例。

5. 简约创新

“万物之始，大道至简，衍化至繁。”

——《道德经》

5.1 什么是简约创新

前述章节讲述了公司如何处理创新以便利用新技术，从而进入新市场并获得更多利润。本章将讨论与全球发展相关的一种特定创新形式，即简约创新（Frugal innovation）。简约创新是一个新的全球现象，在理解它之前先要了解“简约”的字典定义。

简约表示“在使用或支出上节省，审慎地节省或吝啬，不浪费，费用少；或只需要很少资源”。需要注意的是，简约并不意味着质量差、不合标准或即兴的解决方案，也不只是让现有产品变得更加便宜。相反，简约创新是一种旨在满足贫困群体需求的创新。

简约创新通常被定义为简化并重新设计产品和服务，从而减少复杂性和成本，为处于“金字塔底层”的人们提供低价的高质量商品，这些人群涵盖世界上近四十亿人口，他们每天的生活费不足 2 美元。

最近的一份产品价格比较报告显示，Rao 在《简约创新的破坏性有多大？》一文中指出，简约创新可以使产品价格从 50%降低至 97%。从经济角度看，简约的产品和服务旨在最小化整个价值链中材料和财务资源的使用，不仅大幅降低产品价格，还降低拥有和使用产品的全部成本，同时满足（甚至超过）预设的、可接受的质量标准。从功能角度看，简约创新通常要考虑客户因素，并能够应对日常苛刻的条件，如灰尘、高温或停电等。因此，相关设计以及设计师的思考方式必须考虑这些因素，以可扩展、可持续的方式为极端约束条件下的用户提供服务。

简约创新通常有两种形式。第一种形式，是对现有产品、服务或系统进行简化，去除奢侈属性，同时保留其基本技术功能，以确保其能够发挥最佳性能（如图 5-1；简约创新在印度又被成为“Jugaad”）。去除这些奢侈属性后，价格会大幅下降，即使发展中国家的低收入群体也能负担得起这种产品或系统。以诺基

亚 1100 手机为例，这款手机主要针对发展中国家的低收入用户，他们只需要打电话和发送短信的功能。

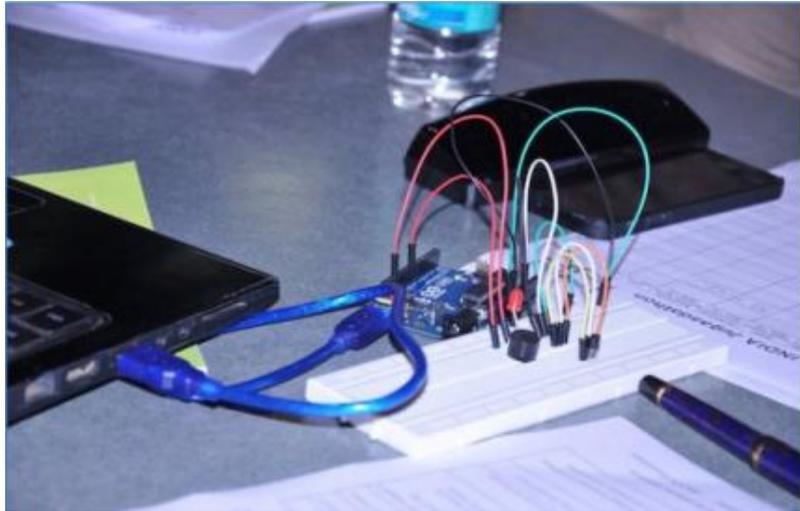


图 5-1 印度举办的简约创新马拉松

第二种形式是从潜在顾客需求出发，创造新产品或新系统。由非洲简约创新中心（CFiA）开发的简约型温度计就是一个典型例子。该中心由莱顿大学、代尔夫特理工大学和鹿特丹伊拉斯姆斯大学合作成立。他们设计的这款温度计具有特殊性能，即使文盲也能正确使用。例如，可以通过扫描额头来测量体温，并用不同颜色显示体温——红色表示需要就诊，绿色表示一切正常。

5.2 简约创新案例

关于简约创新的开发和生产，存在一些误解。在传统的创新和战略管理文献中，研究焦点主要关注高收入市场中创新的决定因素和影响因素。在这些市场中，创新可以追求高单位利润或利润率，为企业提供盈利机会。有关看法普遍认为低收入群体无法产生可观的盈利机会；而涉及为高收入市场量身定制创新时，这一观点则是正确的。然而，当涉及简约创新时，这一观点就不正确了。目前，许多西方跨国公司在简约创新的开发和生产方面仍存在一些战略上的误解：

- 他们认为，由于价格较低，“金字塔底层”消费者的有限购买力无法转化为盈利机会。
- 他们认为，高科技公司在“金字塔底层”市场中没有发展空间，因为这些市场的消费者使用低技术生产过程所生产的简单产品。

- 他们担心服务“金字塔底层”市场会被认为是在剥削穷人。

在过去 25 年里，由于全球化、国际贸易及资本流动自由化，世界已经发生了重大变化。特别是过去 25 年中，新兴市场高增长率催生了一批新的顾客群体，他们对市场提出了新的需求。在这些国家中，中产阶级的人数正在增加。同时，仍有近四十亿的潜在“金字塔底层”顾客，他们超过世界人口的一半，主要生活在发展中国家，尤其是非洲。

跨国公司通过挖掘“金字塔底层”顾客的巨大潜力，为发展中国家的经济和社会发展做出贡献，从而为盈利和经济社会发展创造了机会。刺激经济发展的一个途径是，为“金字塔底层”顾客开发简约的创新产品，从而将混乱分散的当地市场转变为组织有序的私营部门市场，该市场上的产品成本将远低于当前水平。

跨国公司向“金字塔底层”市场提供简约创新的一个例子是通用电气医疗集团。该集团为印度农村地区生产了一种心电图仪，印度农村地区贫困、缺乏电力和训练有素的医疗人员。这种心电图仪的价格约 1000 美元，仅为美国市场开发心电图仪价格的十分之一。

5.3 简约创新和负责任创新的联系

简约创新有两个责任要素。首先，亚洲、非洲和拉丁美洲等新兴经济体的两位数增长率，提高了人们对更高生活水平的渴望。如果日益壮大的中产阶级新客户像高收入国家客户一样消费同样产品，将不可避免地增加全球自然资源压力。这种情况不可能无限期持续下去。因此，各国政府和国际组织正在实施有助于更加可持续发展的法规。同时，低收入和中低收入阶层的顾客，特别是“金字塔底层”的顾客，只能负担得起价格低廉的简约产品。增加社会和可持续法规的成本（提高成本）与世界贫困人口的迫切需求之间存在矛盾。这就需要设计出与不同于高收入国家客户所习惯的设计流程。



图 5-2 通过指尖采血诊断早期疟疾的光学仪器



图 5-3 精悍的磁共振成像（MRI）元件

第二个责任要素与商业模式有关，这种商业模式可以在利润和当地经济发展之间建立联系。传统的产品管理以产品为中心，而在简约创新情况下要设计出一个全新的商业生态系统。这样的生态系统意味着创新型企业必须与外部合作伙伴（如政府和非政府组织）以及当地创业者进行合作。其中，当地创业者尤其重要。这是因为，首先，当地创业者可以为简约创新提供明确的分销渠道，这对生活在偏远农村地区的“金字塔底层”顾客尤为重要。其次，当地创业者更了解潜在顾客的文化和本地偏好，可以在创新型企业设计过程的早期阶段提供重要的信息。值得注意的是，“金字塔底层”顾客可能生活在资源极度匮乏的环境中，在这种限制条件下工作为新的简约创新提供了机会，这可能为当地经济和社会发展做出负责任和“包容性”的贡献。当地创业者也可以成为非常重要的中介，将相关环境知识传递给创新型企业。

5.4 创新和社会标准

简约创新并非自然而然地成为负责任创新。我们还必须关注社会标准问题，看看它们如何共同决定简约创新（在什么情况下）成为负责任创新。接下来本节将论证在生产过程中社会标准过低或过高，简约创新都不是负责任创新。

社会标准，有时也被称为劳动标准，主要有两个要素：

- 第一，确保劳动者享有体面的工作条件。例如，合理的最低工资，适当的健康与安全防护措施，在工厂中灭火器有效可用，在农场中为处理化学品的工人提供防护服。
- 第二，确保劳动者拥有“赋能权利”，包括集会自由和集体谈判权。社会或劳动标准需要确保工人能够成立协会，并且工会领导不会遭受更差待遇，甚至被雇主解雇。

不过，确定社会标准说起来容易做起来难，因为社会标准具有关键复杂特点，通常无法“看到”它。以童工为例，无法通过查看一件 T 恤或喝一杯咖啡来判断该产品加工过程是否涉及童工。我们称这种产品为“信任品”，意味着我们需要信任那些监督这些生产过程的人。

这种监督可以由政府机构（如劳动监察组织）、高水平责任感的企业进行，也可以通过监督自身社会或道德标准来实现，还可以由非政府组织（NGOs）实现，比如公平贸易组织监督小型农户合作社，以确保它们没有使用童工。

当然，监督也面临一些挑战。虽然一些政府能更加有效地监督社会标准，但所有政府在非正式经济中面临生产流程方面的挑战，而许多贫穷消费者正是在这种经济中购买大部分产品。

（1）社会标准如何影响简约创新

在介绍了社会标准的理念和实践后，让我们回到本节的主要论点，即简约创新并非总是负责任创新。社会标准在解释该论点时发挥重要作用。当简约创新以“剥离”现有高价值产品为基础时，生产者首先可能牺牲的就是社会标准，比如工人的最低工资，为降低成本而削减健康和安全方面的考虑。例如，在非正式经济中，简约创新可能无法保护工人免受剥削性的工作条件。在这种情况下，社会标准过低，简约创新就不能被看作是负责任创新。

当然，这不是简单的是非问题，而是努力确保尽可能体面的工作条件的问题。这意味着，我们不能仅从产品技术或伦理维度来看问题，还需要考虑这些社会条件下进行简约创新的情况。

不幸的是，社会标准过高的可能性也存在。有很多例子表明，大型企业成功地向国家政府和国际机构游说，为新企业设置进入壁垒。这些在位企业试图保护他们的既得利益，并阻止拥有新想法的新企业进入市场，他们使用（除其他手段外）更高社会标准作为保护市场主导地位的依据。这种保护主义受到来自新兴经济体企业的强烈批评，因为他们发现自己很难进入欧洲和美国市场。

从原则上看，更高的社会标准是一件好事。毕竟，谁会反对更高的工资或更好的健康安全条件呢？不过，如果大型企业利用这些标准来制造一种印象，即他们的做事方式是唯一合法的商业方式，并成为新企业进入壁垒，阻碍创新。此时情境就完全不同了。这也意味着，过高的社会标准门槛会阻碍简约创新，因为它阻碍了创新，特别是那些在不牺牲用户价值情况下大幅降低成本的创新。

（2）简约创新的注意事项

社会标准设定得过低或过高都会对简约创新产生影响。当社会标准设定过低时，这很容易导致对工人的剥削，从而导致不负责任的创新和生产过程。当社会标准设定过高时，大型的、既有的企业会利用高标准来阻止新企业进入市场，而这并非出于道德原因。

这意味着，在设定社会标准时，需要考虑较高标准（不能过高）以确保体面的工作条件，还要考虑较低标准（不能过低）以允许新的创新机会，这样简约创新才可能促进包容性发展。

简约创新要成为负责任创新，还必须满足另一个方面，即它们是否有潜力将贫困消费者和生产者纳入经济增长和社会发展中。因此，如何实现系统的、包容性的增长对简约创新也非常重要。

5.5 创新和包容性发展

（1）包容性发展的必要性

包容性经济增长和发展是什么意思呢？包容性增长意味着每个人都有足够的机会参与增长过程，同时确保社区内的利益可以共享。为了实现包容性，增长应该惠及每个人，同时减少贫困人群在享受利益和获得参与机会方面所面临的不利条件。

如今，生活费每天不足 2 美元的多数人生活在两个地区：南亚和撒哈拉以南非洲。这些人有近三分之二为极端贫困人口，主要集中在五个国家：印度、中国、

尼日利亚、孟加拉国和刚果民主共和国。然而，自 1990 年以来，这些地区的国内生产总值增长率一直相当高，如表 5-1 所示。

表 5-1 年均 GDP 增长率（世界发展指标，2014）

地区/国家	年均 GDP 增长率（%）	
时间	1990-1999	2000-2013
全球	2.7	2.7
撒哈拉以南非洲	1.9	4.9
南亚	5.5	6.5
中国	9.6	9.8
印度	5.8	6.8

这种经济增长使极端贫困人口数量迅速减少，即生活费每天不足 1.25 美元的人口数量大幅减少。然而，生活费每天不足 2 美元这一贫困线下的人口数量下降得并不显著，2011 年人口数量为 22 亿，而 1981 年则为 26 亿。

在许多发展中国家，贫富差距以及有没有发展机遇之间的鸿沟正在扩大。这意味着，在发展中的经济体中，许多人仍然难以获得优质学校、医疗保健、电力、清洁水以及其他关键商品和服务。

经济增长、贫困和不平等趋势凸显了当前发展特点。在许多国家，许多人仍然无法享受经济增长和发展带来的成果。这主要有两种表现形式：一方面，有些人拥有报酬丰厚的工作或获得土地的途径，却仍然面临实际收入波动大或下降的情况；另一方面，有些人完全不在创收活动范围内，如失业者和无地者。

为什么高经济增长率与贫困人口缓慢减少、不平等现象加剧并存呢？创新的主要轨迹是原因之一。这种轨迹的特点在于资本密集、规模密集、依赖高质量基础设施、依赖熟练劳动力以及通常的产品组合，这些主要满足中产阶级和上层阶级的需求。综上所述，这种创新轨迹系统性使穷人处于不利地位，无论是作为消费者还是生产者。同时，它还将大量人口排除在生产性就业之外。简而言之，主流创新轨迹是全球贫困持续存在的部分原因，但也是重要原因。

（2）通过简约创新实现包容性发展

这引发一个问题，即简约创新能否对实现包容性发展发挥作用？与主流创新

轨迹相比，简约创新是否对贫困人口更具包容性？首先，可以从贫困消费者的角度来回答这个问题。这里有两个重要问题需要考虑：消费单位的身份和贫困消费者的需求特点。可支配收入非常低的人单独购买商品和服务的能力较弱。通常情况下，贫困消费者购买产品或服务时，将以家庭为单位（例如，全家人共用一部手机），或几个家庭共同购买，或整个村庄或社区组织一次性购买（如一头耕牛、一台水泵、一个气象站等）。

（3）为贫困消费者服务

简约创新旨在降低产品和服务购置成本，从而使个人或家庭更有可能做出消费决策。这意味着更多人能负担得起这种产品或服务，因此简约创新将更具包容性。奥妙（OMO）冷水洗衣粉小袋装就是一个生动例子，小份量洗衣粉可以让更多贫困消费者以低成本购买。

如果简约创新能够考虑贫困消费者的需求特点，那么简约创新会更具包容性。图 5-4 展示了贫困消费者所选产品的 9 个方面的特点。这些特点包括产品是否是一次性的或可重复使用、购置成本、使用寿命、维护成本、运营成本、品牌形象、对环境的影响，以及产品或服务在多大程度上反映当地、环保和伦理因素。简约创新通常反映了与低收入消费者相匹配的特点：简约产品和服务具有较低购置成本，因此贫困消费者能负担得起。

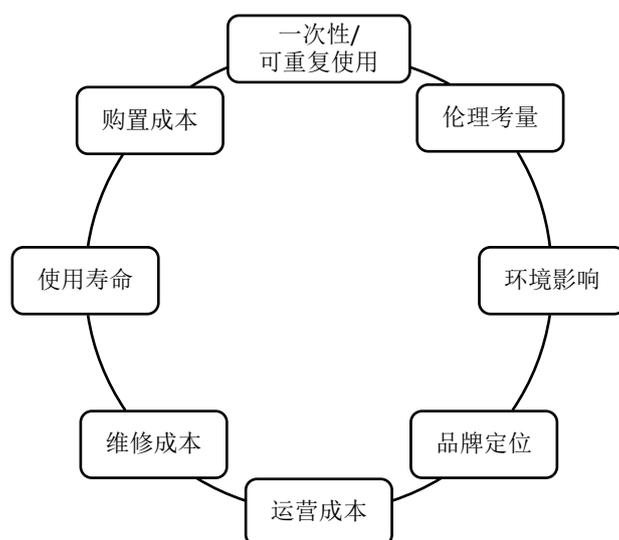


图 5-4 产品特征

到目前为止，简约创新对贫困消费者来说是好的。但是，正如上一节所述，为贫困消费者提供产品或服务可能要付出代价：产品可能无法回收，或者可能涉

及较低的道德、安全、劳动和环境标准。这意味着简约创新的包容性可能与其他方面相冲突，而这些方面能使其成为负责任的创新。

(4) 为贫困生产者服务

简约创新是否比其他类型的创新更有利于贫困生产者呢？大多数贫困生产者存在于非正式部门。贫困生产者通常拥有微型、小型或中型企业，由于可用资本很少，他们经常不得不使用自己的劳动力或家庭劳动力，而且这些劳动力是非熟练或半熟练的，生产规模也比较小。因此，作为一般规律，包容性创新应该产生由小规模或集体生产者所拥有、使用相对劳动密集型技术和非熟练劳动力的生产过程。那么，简约创新是否也符合这一类别呢？答案是不一定如此。

表 5-2 针对贫穷生产者和贫穷消费者的创新（示例）

	贫困消费者	富裕消费者
贫困生产者	<ul style="list-style-type: none"> • 非正式部门的家具 • 非正式部门的服装 • 非正式部门的设备 	<ul style="list-style-type: none"> • iPhone 组装中的非熟练劳动力 • 从事花卉出口的小规模农户
富裕生产者	跨国公司为贫困人口生产产品，例如与能源使用、电子设备、健康和卫生相关的产品	奢侈品，如汽车和手表

如表 5-2 所示，简约创新主要位于左上角和底部象限。通常多国或跨国公司是简约创新的重要推动者；然而，需要质疑的是贫困生产者是否被包括在价值链中。产品分散式营销可以为贫穷的贸易商创造就业机会，也可以在生产阶段雇佣非熟练或半熟练劳动力，诸如此类做法可以产生包容性效应。然而，除非本地生产者能够成为跨国公司价值链的一部分，否则他们的角色通常会受到限制。他们可以通过成为跨国公司营销和分销网络的一部分，或成为当地的投入和信息来源，来实现这一点。

这种类型的创新与来自非正式部门的简约创新不同。在这种情况下，当地生产者参与了这些创新的设计、生产和营销，但溢出效应也相当有限。贫困的设计者和生产者面临许多限制，这些限制阻止他们扩大规模，或将自身活动与本地、

国家经济中的其他行动者联系起来。

总的来说，简约创新并非定义上的包容性创新，因为它们允许贫穷的生产者“锁定”现状。雷丁（Redding, 2002）将技术锁定定义为一种极端情况——“尽管存在更高效的技术，代理人仍继续使用现有技术”。

然而，在简约产品和服务的设计生产方面，贫困生产者相对于跨国公司而言具有一些比较优势。例如，他们更了解当地贫困消费者的需求和偏好，而且不容易受到因高标准要求而产生的声誉损害。我们需要更多的实证研究来评估简约创新在多大程度上可以成为服务于贫穷生产者的包容性创新。

5.6 结论

前述章节已经表明简约创新不一定是负责任创新，除非它满足包容性的维度。就像在社会标准的情况下，简约创新要成为贫困消费者和生产者的包容性创新，必须满足各种标准。

由于许多简约创新仍然由跨国公司设计、生产和营销，因此其包容性并不一定能得到保证。自下而上的简约创新可能允许更高的包容性，但简约创新的贫困生产者仍然面临各种限制，这些限制阻碍了他们扩大规模并向本地和国家经济产生溢出效应。不过，TAHMO 气象站的案例表明，简约创新仍然具有巨大的包容性潜力，可以作为负责任创新来发挥作用。

案例 1：TAHMO 气象站

制冷的历史几乎与人类历史一样悠久。在史前时代，人类就已经开始探索制冷技术。接下来，本案例将讨论一个简约创新的例子及其背后的不同考虑因素——“跨非洲水文气象观测站（TAHMO）”。TAHMO 气象站项目是一种简约创新，因为它旨在以较低价格复制高科技传感器在气象站中的功能，特别是针对撒哈拉以南的非洲地区。这些简约气象站的最初目标仅仅是收集天气和水文数据。图 5-5 是世界气象组织地图，为我们展示了向全球天气预报提供数据支持的气象站。



图 5-5 全球气象站点

图中的蓝色站点尚在全运行，其余站点则低效或完全停止运行。如图所示，撒哈拉以南非洲地区的气象设备尤为稀缺。这对天气预报的准确性和水资源管理产生了负面影响。在该地区进行简约创新的想法是通过跳跃式发展，建立 2 万个气象站网络，使非洲成为气象监测最完善的大陆。

1. 使功能最大化、成本最小化

荷兰代尔夫特理工大学和俄勒冈州立大学的研究人员正在尝试建立一个自给自足的观测网络。每个 TAHMO 气象站都是人们所熟知的气象站的简化版本，它通过廉价的传感器技术实现简约创新。典型气象站的成本在 5000 美元到 15000 美元之间，这对于撒哈拉以南的非洲地区来说过于昂贵；此外，这样的气象站还需要专业技术人员进行维护。因此，出于种种原因，该地区无法使用标准设备。这也是为什么 TAHMO 项目要安装成本低、坚固耐用、几乎不需要维护的气象站。

在恶劣的环境中，需要考虑的因素有很多。例如，不应该有活动部件。正如图 5-6 所示，昆虫喜欢在这些活动部件内部或周围筑巢，从而使活动部件无法正常使用。再比如，标准气象站有通风良好和防护罩的外壳，用于安装温度和湿度传感器。当研究人员在加纳打开这样一个外壳时，他们发现传感器周围布满了毛毛虫网（图 5-7 所示）。研究人员还希望通过批量传感器而不是专用传感器来降低气象站的成本。例如，ZyTemp TN9 通常用于非接触式医用温度计，能够以官方辐射传感器的小部分成本准确测量长波辐射。



图 5-6 活动部件中的昆虫



图 5-7 毛毛虫网

另一个例子是测量降雨量，降雨量是非洲地区最重要的天气变量。理想情况下，人们不仅想知道降雨量，还想了解雨滴大小的分布，后者对侵蚀研究非常重要。然而，准确测量雨滴大小分布仪器的成本超过 10000 美元。在尝试了多种材料后，研究人员发现了一种简单的压电元件，这种元件可以在任何烟雾报警器中找到，成本约为 1 美元。当该元件受到机械刺激时会产生电信号。换句话说，当雨滴落在它上面时，能产生一个可以被捕获和记录的信号。雨滴越大，信号越强。

2. 利用教育网络获得支持

TAHMO 项目的第二个重要特征是教育帮扶。气象站通常需要围栏和专职看护人员，通过与当地学校建立联系，将观测站设在学校可以保护它们免受盗窃和破坏。作为回报，学校可以获得数据访问权限以及一套完整的教育材料。

TAHMO 在的加纳早期试点项目，旨在确定如何将气象站和水文站纳入学校课程。该项目包括一个校际交流项目，其中发达国家的富裕学校出资购买了两个气象站——一个安装在自己校园内，另一个安装在非洲农村相对贫困的学校。随后，学校进行了一系列关于气候、水文、天气以及姐妹学校之间信息交流的讲座，第一次交流发生在 2014 年，参与者是爱达荷州和肯尼亚的学校。

该项目的另一个组成部分是与非洲大学进行传感器设计竞赛。各团队任务是按照 TAHMO 的设计标准设计新的传感器，比赛最终产生了 23 个设计方案。其中一个例子来自尼日利亚，他们提出了一个想法，即通过称量用于保护电路的干燥剂来反映空气中的相对湿度。这是如何利用现有物品实现其他用途的有趣例

子。这样，我们还能获得相对湿度的额外数据点。另一个例子是来自肯尼亚的吉尔伯特·姆万吉（Gilbert Mwangi）和肯·奥迪安博（Ken Odhiambo）提出的想法，他们试图通过测量旗子运动来确定风速和风向。拥有最有趣设计方案的十三支团队，获得了一套创客套件，其中包括通用的电子设备，如 Arduino 微控制器和许多其他工具，以供他们实现自己的设计方案。

3. TAHMO 项目的商业模式

扩大 TAHMO 项目的规模也是一个重要问题，因此需要合适的商业模式。许多人认为，收集天气和水文数据是政府应该做的事。这可能没错，但在过去几十年，全球环境监测网络却在减少，收集数据并不是政治家能够赢得民心的手段。因此，TAHMO 项目的目标是开发具有财务吸引力的公私合作模式和商业案例。在整个价值链中，从气象站安装运营到数据分析预测，人们都需要某种激励来继续运营 TAHMO 网络。

最初的财务数字虽然可观但并不惊人。可能有必要从政府补助和津贴开始。然而，在政府补助结束后想要继续运营，TAHMO 项目就需要实现财务自给自足。这方面的潜力是存在的。据估计，在美国，天气数据和预测的经济价值每年约 310 亿美元。在非洲，只需要获得该价值的一小部分，就足以维持项目运行。一个可能的商业案例是商品交易商。了解棉花或可可等作物的生长状况将为对冲提供重要的财务优势。这些优势中的一小部分就足以维持 TAHMO 项目运营。

基于指数的天气保险也非常有前景。农民可以利用天气信息做出生产经营决策，例如决定何时播种作物。同样，保险公司也能对这些数据感兴趣，因为它们为保险公司提供了更好的作物歉收估计，这些估计是保险公司计算作物保险费用的基础。因此，当农民投保后，雨水不足，保险公司要进行赔付。该项目目前正与肯尼亚的保险公司 Kilimo Salama 合作，该公司利用移动电话网络销售保险和组织付款。

TAHMO 或许能让非洲在天气和水文监测领域实现跨越式发展。通过将创新设计与教育、商业相结合，TAHMO 网络可以提供出色的信息服务。其目标消费者包括各种实体。在个人层面，气象站的目标是服务当地农民。通过移动信息服务，气象站可以提供及时、可靠和与当地有关的天气数据，使加纳的可可农民更好地管理有限资源，更有效地利用现有水源，并对农场进行投资。其他消费者可

能是较大的贫困农民合作社，还可能是保险公司、地方政府和非政府组织，他们需要气象数据来更好地服务贫困客户，并向贫困消费者提供量身定制的新服务。

目前，对西非的当地创业者来说，他们获得的收益并不在于建立气象站，而在于气象站为数据处理和相关营销服务创造的额外就业机会。这不仅涉及银行、保险或小额信贷提供者，还涉及需要传播信息的通信技术公司。这种类型的就业机会可能无法在所有情况惠及贫困人群，因为它需要半熟练和熟练的劳动力。

因此，气象站最可能产生的包容性影响是，贫困农民受益于易于获取的天气数据，从而改善他们的农场管理。此外，通过天气保险等新举措，贫困群体能减少受到收入冲击的脆弱性。

讨论题：

1. 试举关于简约创新的例子，并讨论它要满足的社会标准有哪些？
2. 还有哪些简约创新符合包容性和社会标准的条件？
3. 简约创新不一定是负责任创新，在什么情况或条件下，它是负责任创新？

6.企业负责任创新实施：新标准

海纳百川，有容乃大。”

——林则徐

6.1 企业负责任创新实施简介

研发工作不断推动新技术开发，这些新技术对人们的日常生活、社区、地区、整个经济和社会都产生深远影响。这些技术越能推动范式改进或变革，其影响也就越大。负责任创新为应对人们和社会的需求与关切提供了途径，并有助于开发能够发挥积极影响的流程、产品和服务，引导创新向可持续发展目标迈进。

对于企业而言，负责任创新有助于预测社会或市场的发展趋势、需求、技术场景、监管变化，从而为企业整体战略提供信息，并有助于节省资金和时间。负责任创新也是企业建立信任、获取合法性的关键。企业对社会需求与关切的早期识别，有助于设计出更符合社会期望的产品，从而获得最终用户和其他参与者（如供应链合作商、监管机构、政府或认证机构）的认可。不过，目前企业实施负责任创新的举措仍然有限。欧盟推动的 PRISMA 项目（即产业负责任研究与创新试点项目，<https://www.rriprisma.eu>）帮助 8 家企业在创新和社会责任战略中实施了负责任创新。

6.2 企业负责任创新实施路线图

PRISM 项目聚焦不同行业和技术领域，根据这 8 个试点项目的经验，PRISMA 项目制定了一套实用指南，即负责任创新实施路线图。该路线图旨在帮助企业在技术和产品开发中加强伦理、法律和社会影响（ELSI）方面的考虑。

负责任创新实施结果取决于行业、技术、公司类型和业务类型，而该路线图详细描述了企业实施负责任创新的过程，有助于企业根据实际情况进行调整。这份路线图已成为欧洲制定企业负责任创新新标准的起点，它与现有社会责任标准（ISO26000）、风险管理标准（ISO31000）、质量管理标准（ISO9001）和创新管理标准（ISO5600）相一致。本节将详细讨论这一方法。

（1）负责任创新原则

负责任创新原则是其路线图的基础。负责任创新具体原则包括预期性（anticipation）、反思性（reflexivity）、包容性（inclusiveness）、响应性（responsiveness）。表 6-1 汇总了负责任创新相关特征的行动方针。

表 6-1 负责任创新原则及其行动方针

负责任创新特征	行动方针
预期性（anticipation）和反思性（reflection）	在产品开发早期阶段整合对伦理、法律和社会影响（ELSI）的分析。
包容性（inclusiveness）	利益相关者参与产品开发的所有阶段，为产品开发提供信息。
响应性（responsiveness）	在产品开发过程中整合监测、学习和适应机制，以应对公众和社会价值观、规范性原则。

注：reflection 侧重于对事物的深入思考和反省，reflexivity 则强调自我反思与验证的过程。

- 反思性

仔细审视每项活动、承诺和假设，以便将它们与道德价值体系和科学的良好实践联系起来，同时要考虑到知识的局限性，以及特定问题框架并非普遍适用。

- 预期性

系统推断研发结果应用的所有可能场景；在这些场景中识别潜在的风险、机遇、不确定性、关键问题，并提出预防、管理或利用这些因素的可能方法。其目的不仅在于预防不良事件，还要塑造理想未来，并围绕这些未来组织活动和资源。预期未来应该基于现实，并避免高估创新的好处。

- 反身性

它表示一种制度实践，公共事务和组织外部人员都可以作为反身行动的一部分。就产品价值链或组织内部其他职能（除研发外）而言，该原则非常重要，因为它们会受到负责创新行动或结果的影响。

- 包容性

在研发过程早期阶段就引入参与式方法，吸引那些对创新过程或结果感兴趣的人参与其中。包容性涉及内外部利益相关者的参与，还与负责任创新其他维度相关，因为不同利益相关者广泛参与可以改进反思、预期和响应的能力。

- 响应性

改变创新过程的方向，以回应利益相关者和公众的建议、需求和价值观，或对环境变化做出反应。在认识到知识和控制不足，或需要对新知识、新视角、监管要求做出回应时，调整创新行动。整个负责任创新过程都应尽可能地呈现出响应性特征。

(2) 负责任创新的实施步骤

表 6-2 汇总了负责任创新的具体步骤：高管承诺和领导→背景分析→重要性分析→试验和参与→验证→路线图设计。

表 6-2 负责任创新的实施过程

步骤	目标
高管承诺和领导 (top management commitment and leadership)	确保组织对负责任创新价值观和方法的认可
背景分析 (context analysis)	分析组织以及负责任创新产品技术；识别产品对伦理、社会和法律的影响，以及产品创新生态系统的利益相关者。
重要性分析 (materiality analysis)	识别并确定优先级：负责任创新的驱动因素和挑战；需要克服的风险和障碍；需要合作的利益相关者；需要追求的重要负责任创新行动。
试验和参与 (experiment & engage)	采取探索性/试验性行动，与利益相关者互动，为负责任创新路线图提供信息支持。
验证 (validate)	评估路线图对产品开发和组织（关键绩效指标）的影响
路线图设计 (roadmap design)	整合并可视化负责任创新的长期战略，覆盖整个负责任创新价值链和产品生命周期。

第一步：管理承诺和领导

负责任创新实施的先决条件是最高管理层的承诺。这一承诺非常有必要，但又不足以实现负责任创新的预期结果，因为自上而下的方法应与自下而上的方法相结合，涉及其他提供领导力的角色。最高管理层可以用以下方式，展示其在负责任创新方面的领导力和承诺：

- 确保负责任创新路线图、相关行动、目标和愿景得以建立，并与组织秉持的价值观、身份和利益相关者相一致。
- 识别并保持企业参与负责任创新的动机。

- 确保负责任创新原则融入组织管理体系和治理，以确保负责任创新实现预期结果。
- 确保能够（长期）获取路线图设计及其未来实施所需要的资源。
- 传达有效实施负责任创新的重要性，支持负责任创新指导原则的应用。
- 支持其他相关角色实施负责任创新，比如支持负责任创新发起人。

第二步：背景分析

负责任创新与一系列因素有关，这些因素与企业的类型和管理政策、企业从事的技术产品、行业和市场、相关监管框架和利益相关者有关。为了有效吸收负责任创新，必须识别符合组织运行现实因素和限制因素的战略和做法。

- 负责任创新产品的伦理、法律和社会影响，以及技术、战略、组织和经济影响。
- 负责任创新的路线图设计重点关注特定技术、产品以及相关负责出创新项目。
- 负责任创新产品的开发阶段，从开始分析到预期产品上市时间。
- 在整个创新生态系统中，对负责任创新产品开发感兴趣或参与其中的利益相关者，并且包括初步理解利益相关者的需求和观点。

第三步：重要性分析

负责任创新的关键方面是预期。在负责任创新价值链早期，识别负责任创新产品和组织的重要性，对于预测负责任创新的影响至关重要，从而获得充足时间改变和调整流程，以确保价值创造（例如，最大限度提高积极影响、最小化负面影响）。这一阶段的目标如下：

- 识别负责任创新产品的相关道德、社会和法律影响，并从驱动因素（价值创造、积极影响）和组织实现影响面临的挑战两方面对其进行描述。
- 识别要解决的风险和障碍（不确定性）以发挥负责任创新的影响作用。在确定风险和障碍时，应考虑科技、战略、组织、经济、伦理和社会方面的因素。
- 在负责任创新产品的创新生态系统中选择利益相关者参与其中。
- 选择重要的负责任创新行动，以发挥其积极作用，解决风险和障碍。
- 设定路线图的初步愿景，应对机遇和挑战。

重要性分析的一个典型例子，是与自动驾驶汽车相关的安全和隐私问题（参见第3章的案例2）。设备在不同工作条件下实际存在的、利益相关者所感知到的安全问题有哪些？如何管理设备在运行过程中可能或必须收集的数据？从社会、伦理和法律角度来看，收集这些数据是否至关重要？这些车辆在运行过程中的自主决策可能涉及哪些伦理问题？所有这些方面都很重要，但具体取决于特定设备（例如自动驾驶汽车或无人机）、技术（例如收集的数据类型、设备管理这些数据的方式等）、使用场景（例如设备在建筑物、城市、农场等中的使用）以及利益相关者。

第四步：试验和参与

利益相关者参与是负责任创新的支柱之一，也是评估重要性分析和路线图设计的关键。在前述步骤基础上，确定组织应该采取的负责任创新试点行动，从而确保负责任创新路线图的适当性和可行性。在这一阶段，需要解决以下方面的问题：

- 至少执行一项包容性行动，让创新生态系统内的利益相关者参与讨论和分析项目的关键伦理、社会影响，并审查路线图草案。
- 执行额外的负责任创新行动，作为路线图中计划好的活动实践和试点。

正如社会责任领域经验所承认的那样，确定需要解决的重大问题并不是一项简单的工作。从经济和金融实际角度制定的方法，虽然只涵盖了短期影响业绩或风险，但从负责任创新角度看，除考虑短期影响外，还应考虑中长期影响（包括有形和无形的方面）。重要的是，要始终考虑利益相关者的观点，并将其适当纳入组织内部进行反思。利益相关者分析包括识别相关群体、组织和人员，以及他们的观点和相关性。表6-3展示了利益相关者分析示例，用于利益相关者分析的工具还有重要性矩阵和利益/影响网络。

表 6-3 利益相关者参与的包容性行动

行动	收益
制定并实施关于伦理、法律和社会影响的沟通对话战略	加强与所有利益相关者的关系和信任，建立网络； 调和对立观点，弥合对立价值观； 创造新价值观； 预测潜在的监管变化；
与分享价值观的企业和社会利益相关者合作，建立积极的道德网络	
与政策参与者、政府和规范机构（欧盟、地区和地方等）对话共同设计产品	
组织公众对话，建立或利用公共平台表达需求和关切	

使用参与式方法连接或组织生活实验室和社会实验	提高产品质量、可取性和可接受性
建立基于用户的实践社区	
促进社会包容性举措，让消费者在创新过程发挥官方作用	
促进价值链中弱势利益相关者的能力建设	

本阶段的主要目标是与创新生态系统的利益相关者（在第3步所选定的）进行对话，讨论他们对负责任创新产品及其伦理、法律和社会影响的观点，以及路线图包含的具体元素的看法。可采取的方法如焦点小组、全体会议、多方利益相关者研讨会、世界咖啡馆和鱼缸练习。该阶段的结果是根据需要解决的重大伦理、社会和法律影响，以及创新生态系统的利益相关者和路线图的综合版本，实现完整的重要性分析。

第五步：验证

负责任创新能否被成功与接受，很大程度上取决于具体环境，并受多个因素影响。如背景分析中强调的内容：公司规模、组织复杂性、技术特点、创新水平和相关风险等。负责任创新行动能产生有形和无形的影响，这些影响涉及从公司层的长期战略因素（如公司声誉）到产品开发的短期因素（如与用户需求和利益相关者价值观的一致性）。因此，在该阶段，以下方面非常重要：

- 确定需要测量和监控的内容，选择标准评估负责任创新行动的影响。
- 选择方法来测量、监控和评估路线图对负责任创新产品和组织的影响。
- 根据选定的标准，评估路线图中负责任创新行动的影响，重点关注有形和无形的附加价值观。
- 探索路线图是否以及在多大程度上嵌入组织的常规创新、风险、质量和社会责任政策。包括识别关键绩效指标以衡量相关影响。

表6-4提供了一些响应性行动示例，这些示例涉及产品开发中的社会价值观。

表 6-4 产品开发中应对社会价值观的响应性行动示例

行动	收益
整合以用户为中心的设计，用户创新，灵活和适应性设计，共同创造	创造价值，增加研发的社会价值或影响力； 树立企业形象和声誉； 符合合格的规范和标准； 便于获得财政支持
筛选供应商的积极做法	
制定针对关注事项或不当行为调查报告的程序	
确保非歧视的招聘流程	

采用适应性风险管理	
让伦理学家参与负责任创新过程	
建立道德、社会和法律监督委员会	
将伦理、法律和社会影响（ELSI）标准纳入内部程序进行研发项目质量监测	
确保研究数据的伦理管理和公平的数据管理	
定期进行道德审查并获得道德认证（通过独立机构）	
在公司和供应链层面获取社会责任感和质量认证	
上市后监测对伦理、法律和社会的影响	
将研发和创新产品的伦理、法律和社会影响纳入企业社会责任/可持续性报告	
支持和投资可持续供应链	
根据道德/责任要求选择投资机制	

第六步：路线图设计

基于上述步骤的结果，负责任创新路线图旨在指导组织将负责任创新付诸实施。这意味着：

- **预期与反思：**从产品开发早期阶段开始，对道德、法律和社会影响进行综合分析。
- **包容性：**利益相关者参与，为产品开发所有阶段提供信息。
- **响应性：**整合监测、学习和适应机制，应对产品开发中的公共和社会价值观、规范原则。

该项目列出了 10 项负责任创新关键绩效指标以用于路线图监测，见表 6-5。

表 6-5 监测路线图的负责任创新关键绩效指标

负责任创新关键绩效指标	量化参数示例
1.道德价值观意识	为学习、反思与创新战略和核心业务相关的道德价值观，企业每年举办的培训或会议次数
2.伦理问题意识	为反思社会和伦理价值观融入具体研发项目的情况，企业每年举办的培训或会议次数
3.企业是否在创新中考虑道德价值观？	负责任创新原则正式纳入公司使命和愿景（例如道德规范）； 每年积极将道德价值观纳入创新战略和技术设计的研发项目数量
4.企业否预见创新对社会的影响？	每年从产品开发早期阶段就邀请内外部利益相关者参与研发的项目数量； 为讨论和确定研发项目的社会影响，与其他创新者和外部顾问开展咨询活动的次数

5.利益相关者参与	公司每年组织利益相关者参与活动的数量； 每年在负责任创新/研发计划中，利益相关者积极参与项目的数量； 每年终端用户参与负责任创新/研发项目的数量
6.性别多样化	企业研发/负责任创新部门或团队中的男女比例
7.负责任创新相关选择的透明度和问责制	在公司层面制定正式沟通策略，以确保关键文件或网站解释与负责任创新最相关的选择； 每年整合非财务价值的专利数量； 开放获取出版物的数量
8. 应对产品开发中公共和社会价值观的学习机制	每年正式纳入企业创新模型中（如用户中心设计、共同创造），以用户为中心的方法数量； 每年为响应（新的）社会发展需求所开发的用户体验工具数量
9. 解决产品开发中的公共和社会价值观的学习机制	每年针对社会或道德导向的产品与服务的研发项目数量
10. 积极监测负责任创新的影响	每年使用影响分析策略（如风险管理、伦理/社会影响分析等）的负责任创新/研发项目百分比； 已设立正式的外部审计程序（至少每年一次）， 监控企业的非财务价值

6.3 负责任创新模板：路线图

基于上述负责任创新步骤，可以形成以下路线图模板，如图 6-1 所示。

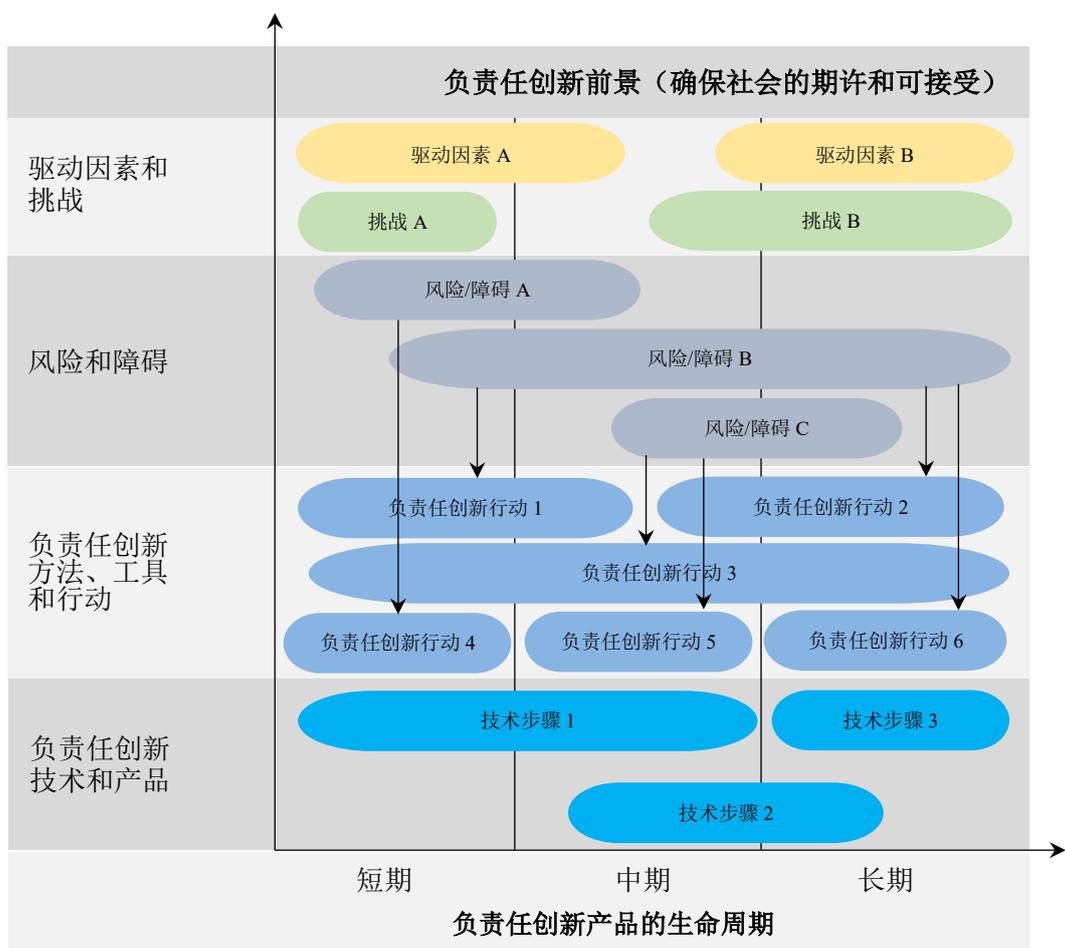


图 6-1 负责任创新模板：路线图

负责任创新路线图模板有 4 条“行动线”（与上述 6 个步骤相一致）：

（1）基于对组织和特定负责任创新产品的重要伦理、社会和法律影响，以及相关战略、组织和经济问题的考虑，来定义驱动因素和挑战。

（2）识别负责任创新行动要解决的风险和障碍。

（3）识别行动计划，以在产品开发各个步骤中实施负责任创新，这是路线图的核心部分。

（4）识别能实现研究与创新目标的创新技术或产品。

6.4 实施负责任创新的 SWOT 分析

如前所述，在企业开展负责任创新需要付出相当大的努力，因此，企业需要对内外部环境面临的优势和劣势、机会和威胁形成较为清晰的了解。表 6-6 例举了欧盟 PRISMA 项目进行的 SWOT 分析。

表 6-6 SOWT 分析

	优势	劣势
组织内部环境分析	<ul style="list-style-type: none"> • 创造价值 • 激励员工 • 提供竞争优势 • 强化与所有利益相关者的关系 • 增加利益相关者之间的信任 • 提高研发的社会价值/影响力 • 加强产业层面的创新质量 • 确保符合规范和标准 • 识别新的市场需求 • 传达产品收益和风险的潜力 • 增强产品开发的透明度 	<ul style="list-style-type: none"> • 负责任创新概念的认识和技能有限 • 额外的官僚负担，缺乏资源（尤其对中小企业而言） • 对产品开发的影响力感知不足 • 负责任创新缺乏跨公司职能的整合 • 公司某些职能部门的内部抵制 • 衡量相关成本存在困难 • 产品开发增加过多额外成本 • 知识产权 • 概念滥用
	机会	威胁
组织外部环境分析	<ul style="list-style-type: none"> • 提高产品质量、可取性和可接受性 • 提高产品的可持续性、安全性和可靠性 • 提高客户满意度 • 解决现有社会需求，提高对客户生活质量和健康的影响 • 提高效率(如使用资源、决策过程)，降低中长期成本 • 树立企业形象和声誉 • 提高市场渗透率、利润 • 为获得财政支持提供便利 	<ul style="list-style-type: none"> • 与利益相关者互动存在困难 • 创新可能放缓甚至过早停止 • 行业中可用的实际案例(案例研究、应用程序)较少 • 价值链和供应链缺乏参与 • 缺乏合作伙伴和供应商的认可 • 被利益相关者认为是“粉饰门面” • 缺乏激励措施（政策和监管层面）

讨论题：

1. 负责任创新不同于企业社会责任，还有哪些相似概念与负责任创新存在明显不同？
2. 据你所知，现在有哪些企业在应用负责任创新？他们面临的主要障碍是什么？
3. 在你所从事的职业中，负责任创新的主要关键绩效指标是什么？
4. 根据负责任创新路线图模板，选择一个创新项目，并尝试进行 SWOT 分析，以及设计路线图使该项目落实负责任创新理念。

第五部分 风险评估与安全

7.认识风险

“盖明者远见于未萌，而知者避危于无形，祸固多藏于隐微而发于人之所忽者也。”

——司马相如

7.1 风险、不确定性和无知

到目前为止，我们已经了解了负责任创新是什么，以及为什么它应该在新技术开发和传播中发挥重要作用。现在让我们讨论技术中的风险、不确定性和未知，以及如何减轻它们。首先，我们需要清楚如何使用“风险”一词。当我们通俗地使用这个词时，我们会在“吸烟增加癌症风险”或“地上的坑洞是一种风险”等陈述中提到。虽然第一句可以用可能性、概率来代替风险，但第二句这样做就没有意义，在这里，风险是实际伤害或直接危险的同义词。现在，在科学或哲学中，风险一词包括两种含义：某种伤害和可能的伤害。

(1) 风险和不确定性的不同

有时，我们可以用概率来表示危害发生的不确定性。例如，自然科学、社会科学或工程科学提供了核电站中某个阀门破裂的可能性。然而，构成危害的东西总是来源于超越科学的规范性概念，并需要一些伦理专业知识。例如，为了理解为什么气候变化实际上是一种危害，我们需要一个规范性概念，告诉我们为什么这是事实，以及为什么我们需要关心环境。因此，风险本身是一个跨学科概念，这具有重要影响。

首先，以人类为中心的伦理告诉我们，气候变化本身并不是一种伤害，而是气候变化对人类的影响可能是可怕的。这些影响并不是通过狭义的气候模型来确定的，而是通过经济学家建立的福利经济影响模型确定。因此，建议如何应对气候变化威胁的政治决策可能不需要更好的气候模型，而是需要更好的环境影响模型。

其次，伦理评估，即规范性评估，至少在一定程度上应该先于预测某种危害不确定性的实证科学。在不确定性可以用概率量化的情况下，风险通常被定义为

平均危害——也就是说，危害乘以其发生概率。

如果没有危害发生的概率估计，我们在技术评估中对气候变化的反应并不构成风险决策，而是不确定性决策。对于风险决策，我们知道决策的所有可能结果，并为这些结果分配有意义的概率。然而，不确定性是指我们知道整个概率空间，但不能将概率分配给所有结果。

(2) 不确定性和无知的不同

在技术评估中，需要进一步区分无知状态下的决策，即连决策概率空间都不知道的情况。这种情况众所周知，一些人称之为“未知的未知”或“黑天鹅”。缓解气候变化是在不确定条件下做出的决策，而 20 世纪 70 年代氟氯烃上市则是在无知状态下做出的决策。在氟氯烃投放市场时，人们还不知道它会对臭氧层造成破坏。

(3) 应对风险、不确定性和未知

风险、不确定性和无知之间的区别，通常基于将概率作为相对频率的某种解释。这些客观概率在技术评估以及工程、科学的许多方面都很常见。然而，有人则坚持更为主观的观点，即概率是信念等级，而不是相对频率。这有时也被称为贝叶斯方法。理论上，这会模糊风险、不确定性和无知之间的区别。然而，在实践中——以气候变化为例——分配主观概率仍然很困难，因为我们无法更新信念，并且在贝叶斯公式中分配先验概率分布也很困难。

那么，我们该如何应对风险和不确定性呢？

对于风险，可以使用预期效用最大化分析，或从反面说，是风险最小化或风险分析。这种分析属于功利主义范式，即为大多数人谋求最大利益。然而，由于我们不知道确切结果，只能最大化预期效用，或从反面说，最小化预期损害，即风险。对此，核能政策就是最典型的例子。在没有合适概率估计的情况下（如气候变化），这种方法当然不适用。我们可以退回到一种更基本的决策方法，这种方法不需要任何概率估计。在环境和工程伦理学中，这种方法最典型的例子就是预防原则。预防原则有多种应用方式，但大多数是以下两种版本的变体。

第一个版本出自 1992 年发表的《环境与发展宣言》：“遇到严重损害或不可逆损害的威胁时，不得以缺乏充分的科学证据为理由，延迟采取符合成本效益的措施以预防环境恶化。”这个版本被认为是较弱的表述，因为它建议我们在缺

乏充分的科学证据时，应考虑技术的任何潜在影响；然而，它并没有明确说明如何应对这种不确定的情况。尽管如此，我们还可以进行风险分析，至少可以确保将不确定的影响考虑在内。

第二个版本：“在最简单的表述中，预防原则具有双重触发机制，即如果某项活动有可能造成危害，且对影响大小或因果关系不确定，那么应采取预防措施以避免危害。”这个版本——也被称为强版本——确实向我们提出了行动建议。它甚至告诉我们，当危害不确定时，应采取预防措施来避免它。因此，无论负面影响多么不可能，甚至我们不知道影响有多么严重，我们都需要采取行动来避免这些负面结果。

(4) 预防原则和道德过载

将这一公式应用于气候变化问题，意味着我们需要缓解气候变化可能造成的任何负面影响。因此，减少人为温室气体排放的成本并不高，正如一些经济评估所表明的那样，这非常有用。例如，2007年的斯特恩报告（Stern Report）指出，我们发现每年只需投入全球GDP的1%，就可以避免全球变暖造成的主要损害。这相当于每年约需要4500亿美元。

这个数字很难把握，但我们或许可以将其与其他数字进行比较。想想看，要实现联合国千年发展目标之一——为非洲80%的农村人口提供安全饮水和卫生设施，每年仅仅需要13亿美元。这一比较表明，仅仅应用预防原则并不足以充分考虑这种比较。如果我们每年投入全球GDP的1%来应对气候变化，就必须接受这笔资金无法用于其他目标，那么，我们该如何抉择呢？

这不是可以用短篇幅回答的问题，而是需要政治讨论，而且更需要跨学科方法来应对风险和不确定性，其中不仅要考虑危害，还要考虑危害发生的可能性——无论这种可能性是否可以量化或以概率的形式表达。

7.2 未知的未知：极端不确定性

1943年，IBM公司主席托马斯·沃森（Thomas Watson）曾说：“我认为全世界大概只需要五台计算机。”当然，他当时想的是像ENIAC这样的大型主机。然而，他不可能预见仅半个世纪后，就会出现个人计算机、笔记本电脑、平板电脑、智能手机等产品，而且几乎每个人都有属于自己的电脑，甚至不止一台。这个故事也表明，预测未来是很困难的。

(1) 科林里奇困境

科林里奇困境指出：“在技术发展的初期阶段，技术本身仍然可以改变，但其影响却难以预测；在后期阶段则出现了相反的情况，即技术的影响已经明朗，但技术已经深深植根于社会之中，更加难以改变。”

当前大多数应对科林里奇困境的方法都侧重于预见性，即试图让技术变得更具有可预测性。本节将在此讨论两种预见性的方法：风险方法、预防原则。

风险方法的具体操作如下：首先，确定新技术的风险；其次，判断这些风险是否可接受。风险在此处被客观地理解为可能性乘以严重性。然而，问题在于，我们通常不知道具体概率，这导致了不确定性。有时我们甚至不知道所有可能的后果——最终陷入无知。因此，我们实际上无法确定风险。

另一种方法是预防原则。这一原则主张当某项活动对环境或人类健康构成威胁时，即使某些因果关系尚未得到充分的科学证实，也应采取预防措施。这一原则不要求确定概率，可以处理我们所说的不确定性。

(2) 预防原则的缺点

预防原则存在两个缺点。首先，它可能会给出相互矛盾的建议。以将预防原则应用到二氧化碳捕获和储存为例，在荷兰有人提议在靠近鹿特丹的巴伦德雷赫特镇下方储存二氧化碳，然而这一提议引发了激烈反对。如果我们应用预防原则，可能会有人说：是的，我们应该捕获并储存二氧化碳，因为这会加剧温室效应并造成明显的危害。但根据同样的原则，我们也可以说不行——因为如果二氧化碳从储存设施中泄漏出来，也可能很危险。这两种观点都涉及到了可能但不确定的危险，因此，仅凭这一原则我们无法做出决策。

预防原则的第二个缺点是无法应对无知，无知可能导致“未知的未知”。图7-1很好地说明了这一点。这张图片表明，当我们把新技术引入社会时，总会有惊喜和意外情况发生。

欧盟科学与治理专家组于2007年表达了如下观点：“我们不可避免地处于实验状态中。然而，这通常被公众视野和公众讨论所忽略。如果公民未经协商就被当作实验对象，参与那些未明确命名的实验，那么，我们就必须解决一些严重的伦理和社会问题。”因此，我们建议将新技术引入社会看作一种社会实验。这将引导我们提出以下问题：在何种条件下，此类实验在道德上是可接受的？

现在，我们可以问，为实现负责任创新可以采取什么方法？应如何使用现有工具？这一系列提问属于技术评估领域。总体而言，负责任创新方法是从更广泛的技术评估实践中发展而来的。

7.3 技术评估

(1) 负责任创新的先驱

技术评估的两大先驱，即 ELSI 和影响评估。首先，ELSI 表示伦理(Ethical)、法律(Legal)和社会影响(Social Implication)。该项目作为人类基因组计划的一部分，于 1990 年正式启动。它旨在确定人类基因组图谱的伦理、法律和社会影响。项目年度预算的 5% 被用于解决该项目产生的伦理、法律和社会问题。其次，影响评估是负责任创新的另一个重要先驱。它的历史可以追溯到 60 年代末和 70 年代初。其目的是确定当前或预计行动的未来后果。影响评估有很多种类，其中在某些国家是法律要求，要求某些项目在开展之前必须进行评估。影响评估可以包括环境影响评估和风险评估，也可以包括健康影响评估、社会影响评估和性别影响评估等。

技术评估也是影响评估的一种形式。它表示客观预测新技术社会后果的尝试，旨在为政府制定政策提供依据。在美国，技术评估办公室(Office of Technology Assessment, OTA)于 1972 年成立，并一直作为官方机构运作至 1995 年。其宗旨是为美国国会提供复杂科学技术问题的客观分析。虽然 OTA 现已解散，但仍有几个国家拥有这类机构。

虽然技术评估最初作为一种尝试，旨在客观预测技术给政策制定者带来的后果——这本身就是一项艰巨任务，但随着时间推移，它已经得到了进一步发展。技术评估的范围从客观预测预期后果扩展到预测政府、企业和研究机构可能面临的后果；从被动方法发展到主动方法，甚至影响研发和设计。因此，凭借其广泛的范围，技术评估涵盖了负责任创新背后的许多价值观。

(2) 技术评估种类

技术评估方法多种多样。这里，我们将主要介绍三种方法：建设性技术评估(Constructive Technology Assessment, CTA)、中期调整(Midstream Modulation)、道德评价的网络方法(Network Approach for Moral Evaluation, NAME)。

第一种方法是建设性技术评估。建设性技术评估由艾瑞·里皮(Arie Rip)

和约翰·斯高特（Johan Schot）在 20 世纪 80 年代的荷兰发展起来。建设性技术评估旨在通过预见未来发展和影响，减少（人类）反复试验学习的成本，其目标还包括将这些见解反馈给技术的设计过程。

建设性技术评估还有一些特定目标：了解未来发展的社会后果；反思性——对其他参与者的认识；预见潜在的技术发展及其社会后果。在追求这些目标的过程中，建设性技术评估还旨在纳入更多方面、更多参与者来拓宽技术发展。

建设性技术评估使用的一个工具是情境构建。这种情境的目的不是预测未来，而是进行预期。这种可能的未来有助于避免最坏情况，以及制定应对各种可能未来的稳健策略。

第二种方法是中期调整。这种方法主要由美国艾瑞克·费雪（Erik Fisher）提出。该方法主要针对开发新技术的研究实验室。中期调整的目的是提高实验室工作在更广泛社会层面的响应能力。“中期”用于强调该方法侧重于在研发中间阶段调整或改变研发实践。之所以这样做，是因为与研发开端决定资助哪些研究、研发结束决定如何使用特定技术相比，指导研发过程更加可取。该方法主要应用于纳米技术，这是一种新兴技术，通过操纵纳米尺度（ 10^{-9} 米）的材料特性来发展。

美国的国家纳米技术提倡关注 ELSI（伦理、法律和社会影响）及所谓的负责任纳米技术发展。同样，荷兰的 Nanonext 计划也关注与纳米技术相关的风险评估和技术评估。中期调整意味着纳入人文视角，或者让社会科学家或伦理学家参与研究实验室的日常工作。他们可以负责以下活动：参与观察，向实验室同行提出深思熟虑的问题，讨论各种问题，并从不同角度提供反馈。

第三种方法是道德评价的网络方法，即 NAME。这一方法在荷兰代尔夫特理工大学发展起来。它始于这样一个假设，即创新发生在社会网络的背景和现实中。这些网络由公司、研究实验室、大学、用户、供应商、客户等组成。所有这些参与者都在某种程度上发挥作用并影响创新的发展和扩散。NAME 背后的想法是追踪这些网络动态，以辨别道德问题。该方法还包括用网络规范评判不同利益相关者组成的网络，而这两个主要网络规范是学习和反思性、开放性和包容性。

就学习和反思性而言，人们可以区分一阶学习和二阶学习。一阶学习是关于如何实现目标的，例如学习如何改进技术；二阶学习则是关于实现什么目标的，

例如技术的发展和设计应该考虑哪些价值观。开放性和包容性也可以进一步定义。开放性意味着重新阐述网络的核心问题；包容性意味着所有参与者和相关考虑因素都被纳入网络之中。

最后，正如我们在第 6 章所述，责任创新有四个主要组成部分。它们描述了责任创新应该具备的特质：

- 预见性：预见创新可能产生的社会后果；
- 反思性：反思背后的目的、动机和潜在影响，以及已知和不确定的内容；
- 审议性：包括广泛的利益相关者和观点；
- 应答性：通过回应社会和道德关切来影响技术发展和设计的方向。

我们可以清楚地看到，这四个组成部分都受到了早期技术评估方法的启发。

案例 1：关于核能的争论

制冷的发展历史几乎与人类历史一样悠久。在史前时代，为了将刚才讨论的内容放到现实世界的视角中，我们将讨论核能生产及其过程中涉及的价值观。首先，本案例对核燃料循环产生核能所涉及的几种价值观进行分析，即事后分析——对已经存在的技术进行分析。其次，还进行了事前分析，即对尚未存在的技术进行分析。这种分析在负责任创新方面更为重要，因为它试图在新技术开发之前和开发过程中容纳重要的价值观。

1. 可持续性作为道德框架

在可持续性的定义中，有几个价值观非常重要。基于此，我们将进一步讨论，可持续性应该被当作与另外五个价值观相并列的道德价值观。在这些价值观中，每个价值观都具有时间和空间维度。这些价值观有哪些（如图 7-1）？在讨论可持续性和伦理时，需要先了解什么是可持续性。

首先，我们先区分两个不同方面。一方面，可持续性与维持环境和人类安全有关。因此，我们需要讨论环境和公共卫生与安全。另一方面，可持续性与维持人类福祉有关。在这里，我们讨论的是资源耐用性和新技术的经济方面。与之有关的价值观，每一个都具有空间和时间维度。讨论新技术在改变这些价值观以及价值观之间关系中所扮演的角色非常重要。

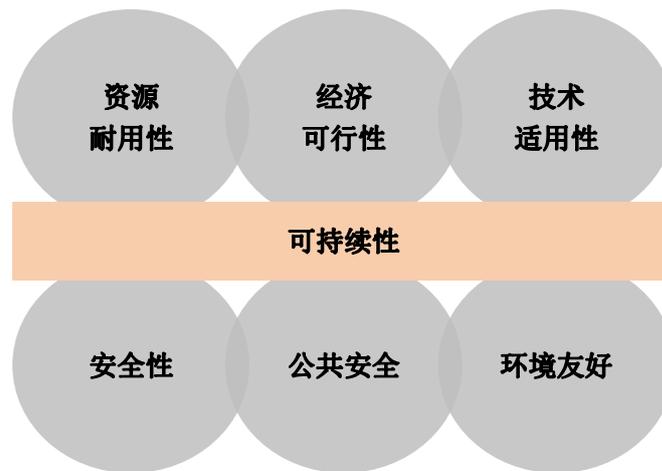


图 7-1 可持续性作为道德框架

2. 可持续性的五个关键价值观

首先，维持环境，从价值观角度看即为“环境友好性”。为什么我们要关心环境？对于这一问题，可以从两种思想流派来回答。一种是人类中心主义，它将人类置于关注的中心。这一思想流派认为，环境本身没有价值，它对人类只有工具性价值。第二种思想流派是非人类中心主义，认为环境具有内在价值，这种价值不一定与环境对人类的意义相关。

第二，公共健康和安全。这个价值观认为，无论现在还是未来，都不应该危及人们的安全。这是一个崇高的目标但问题是：我们应该在多大程度上保护未来，又应该如何提供保护？这个问题与具体政策密切相关。

以尤卡山核废料储存库为例，这是世界上第一个也是最大的核废料储存库，在过去几十年由美国建造。在引入尤卡山储存库的辐射标准时，美国环境保护局（EPA）提出了几项标准。美国环保局认为，在接下来的 10000 年里应该提供完全相同的保护水平。从实际意义上讲，接下来 10000 年的人类生活，每年接触的辐射约为 15 毫雷姆。（REM 是衡量放射性对健康影响的单位，即放射性毒性。）超过 10000 年以后，美国环保局保证的保护水平将大大降低。最初提案是每年 350 毫雷姆，在经过大量的公众辩论后，这一数字才被调整为每年 100 毫雷姆。这意味着，10000 年后的保护水平是现在的六分之一。这就涉及一个基本问题：我们应该在未来多久关心这个问题，以及我们能在多久时间内提供相同的保护？

第三，安全性。在针对核能的讨论中，我们先要区分安全和保障。安全涉及

无意外伤害，而保障则与有意伤害相关。在讨论保障时，一般涉及破坏活动和不扩散，例如制造脏弹的可能性；不扩散涉及制造炸弹，或传播制造此类炸弹的知识等问题。因此，安全和保障是两个不同的概念。

第四，资源耐用性。这涉及到自然资源的可用性。耐用性是对可持续性的一种非常普遍的理解（如图 7-3）。当然，我们不可能立即停止使用不可再生资源，必须有一个过渡期。因此，这里存在的道德问题是：我们能在多大程度上为已经使用的资源提供补偿，而这些资源是后代无法获得的？

第五，经济可行性。使一种能源具有可持续性，则需要在经济上持久耐用。这同样会引发许多道德问题。如对谁来说是持久的？谁的利益受到威胁，以及在道德分析中我们需要认真考虑谁的利益？未来利益是否与当前利益同等重要？如果不是，我们如何权衡未来利益与当前利益？

在经济研究中，未来评估的概念变得重要起来。未来利益的价值将按一定百分比折现成现值。折现是成本效益分析（CBA）的一个非常重要的方面，是为了未来后来利益而进行的。

3. 开放式和封闭式核燃料循环

现在，让我们具体看看核能生产中的关键反应，即核燃料循环。讨论燃料循环是必要的，以便了解处理核废物的可用选项。图 723 展示了目前使用的两种主要燃料循环。黑色箭头表示开放式燃料循环，这种循环在美国、瑞典和其他一些国家普遍使用。首先开采铀并进行铀矿的磨碎，然后进行提纯、转化和富集。处理过的铀氧化物经过核反应堆，从反应堆中出来的是乏燃料，乏燃料可以被视为核废料，但由于其高放射性，需要被埋在地下 20 万年至 100 万年之间。

然而，乏燃料也可以在封闭的燃料循环中进行回收，这种回收过程即为再处理过程。再处理的最大好处是，通过提取和重新插入燃料循环来回收仍然可用的材料——铀和钚。回收或再处理的第二个好处是，极大地缩短核废料的寿命，从最低 20 万年缩短到大约 1 万年。然而，再处理是一个化学过程，也会产生废料。更重要的是，在再处理过程中提取的钚是一种很容易用于制造核装置的材料。这意味着再处理会产生非常重要的安全风险。

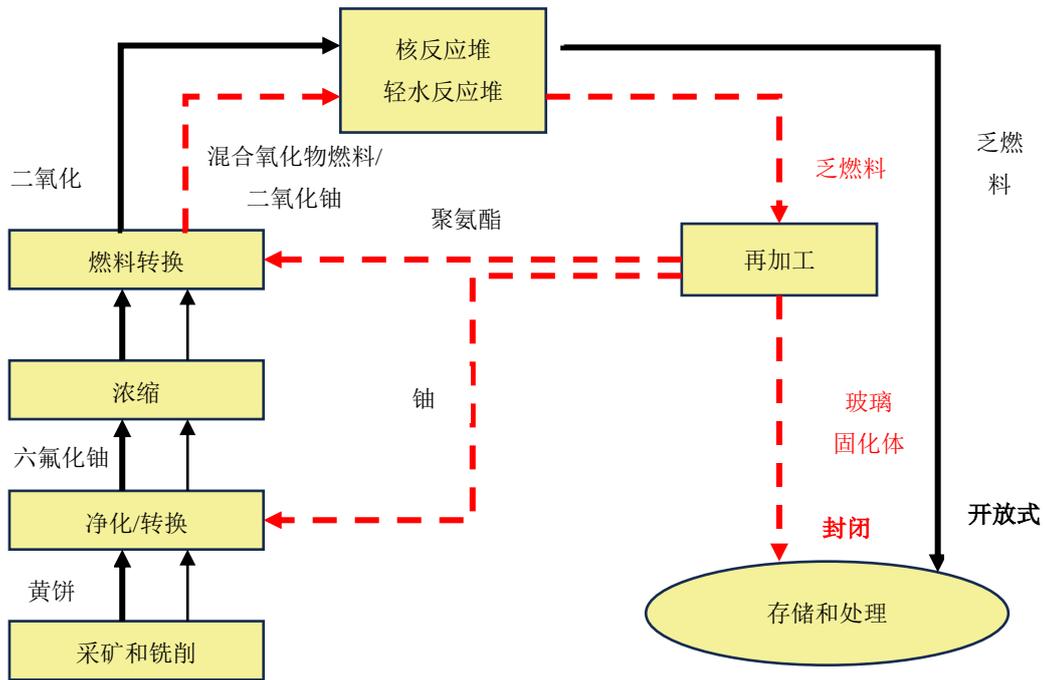


图 7-2 核能生产的两种主要方法

简而言之，我们可以将上述提到的每一种价值观与核燃料循环的两种基本形式——开放式循环或封闭式循环——联系起来。本文的论点是，开放式燃料循环对当代人特别有利，因为它给当代人带来的负担最小；封闭式燃料循环对后代更为有利，因为它能显著减少废料的寿命。不过，封闭式燃料循环也会带来了各种额外风险，尤其是安全风险以及当前再处理工厂的安全风险。表 7-1 汇总了核燃料循环的相关价值观。

表 7-1 核燃料循环相关价值观

	资源耐用性		环境友好		经济可行性		公共卫生安全		安全性	
	短期	长期	短期	长期	短期	长期	短期	长期	短期	长期
开放式	+	-	+	-	+	-	+	-	+	-
封闭式	+	+	-	+		+	-	-		+

4. 核反应堆设计中的安全性

在负责任创新中，我们可以尝试在设计阶段预见并适应相关价值。让我们简要地看一下核反应堆，重点关注核反应堆设计的历史——特别安全性作为该设计的主要标准。此外，还有其他价值也处于利害关系中，因为安全性并不是唯一重要的价值。有时，我们必须为冲突的价值进行重新设计。需要强调的是，负责任创新的一个主要问题是理解这些价值，并在开发新技术之前解决这些冲突。

安全性是核反应堆设计中最重要标准之一。每次发生臭名昭著的核反应堆事故后，安全性都会再次成为必要条件。例如，著名的三里岛（TMI）核事故、切尔诺贝利核事故等。

概率风险评估用于降低反应堆熔毁的概率。这些风险评估实际上在三里岛事故前几年就被引入了。概率风险评估试图描绘可能导致熔毁的事件，并为防止或缓解这些事件分配行动点，优先处理概率较高的事件。最终，我们将熔毁结果的概率作为最终事件，并尝试降低该概率。

概率风险评估由拉斯穆森集团（Rasmussen Group）在 1975 年提出，反应堆熔毁的风险为 5×10^{-5} ，即每 20000 个反应堆年（即反应堆运行的总年数）预计发生一起严重事故。这个数字不是实际年数，而是反应堆运行年数，例如有 500 个反应堆，每 40 年就可能发生一次事故。这种说法在当时被认为是一个相当可接受的风险。

由于反应堆数量可能会大幅增长，因此人们决定对反应堆进行调整。从当时的 500 个反应堆增加到预期的 5000 个反应堆，这意味着反应堆年数量增加了十倍。反过来，这意味着任何事故（如果发生）发生的可能性将增加十倍。因此，根据相同计算，从 500 个反应堆增加到 5000 个反应堆，意味着每四年就会发生一次事故。这是绝对不可接受的，也促使核反应堆设计发生了重大变化。

有两种不同方法可以使反应堆更加安全。首先，对安全性进行渐进式的改进。这意味着以当前设计为出发点，然后添加安全功能或删除不安全元素。其次，进行彻底的设计变革。这意味着需要从头开始，以安全性为主要标准重新设计。

5. 设计安全性的悖论

在不同年代的核反应堆设计中，第一代是原型，现在已经不存在了；第二代核反应堆是目前全世界都在使用的反应堆；在第二代之后的第三代、“第三代+”，

以及第四代反应堆。第三代反应堆现在已经投入运行，而“第三代+”和第四代反应堆仍在开发中。

如前所述，有两种不同的方法来提高安全性：安全性的渐进式改进（绿色表示）；设计的根本性变革（蓝色表示）。在设计根本性变革中，不仅安全性需要考虑，其他相关价值也需要在设计时进行考虑。第四代反应堆应该具有高度的经济性、安全性，减少核废料并具备防扩散性。实际上，它们被设计用来适应多种价值。

这意味着反应堆安全性的悖论如下。目前运行的大多数反应堆是第二代反应堆——沸水反应堆（BWR）和压水反应堆（PWR，也称为轻水反应堆）。轻水反应堆，尤其是压水反应堆，最初是为潜艇设计的，但随后被放大并用于商业核生产。当压水反应堆被用于大规模商业生产时，规模上的扩大意味着安全性随着规模增大而降低。因此，人们添加了各种安全特性，如阀门、泵等。然而，这些额外的特性使设计更加复杂，这种复杂性又使反应堆面临额外的风险。这就是安全性的悖论。

6. 核反应堆设计中的价值与创新

第三代反应堆是以第二代沸水反应堆（BWR）为基础进行设计的。设计先进沸水反应堆（ABWR）的主要原因是使沸水反应堆更安全。先进沸水反应堆还增加了许多安全特性，如反应堆容器底部有单独的内泵、厚厚的纤维增强混凝土安全壳。通过这些特性，先进沸水反应堆大大降低了熔毁的风险。

7. 核能生产的负责任妥协

上述反应堆（及其构成的创新）中的每一个都帮助我们实现了某些价值，同时也牺牲了其他价值。在设计反应堆时，我们是为了实现各种价值而进行设计的。表 7-2 汇总反应堆设计中的价值权衡取舍。

表 7-2 核反应堆设计中的不同价值取舍

	PBMR (球床模块高温气冷堆)	GFR (冷快堆系统)	MSR (熔盐堆)
安全	++	-	0
保障	+	--	-
资源耐用性	-	+	++
经济可行性	+	0	-

PBMR 是一种采用球床燃料元件的高温气冷堆设计，它使用三层各向同性包覆 (TRISO) 燃料以及氦冷却剂，这种设计使得反应堆具有很高的非能动安全性。PBMR 可用于供热和发电，并具有高温出口温度，这为其在多种能源应用中的使用提供了可能。

GFR 指的是气冷快堆系统。这是一种快中子谱氦冷反应堆，采用氦气作为冷却剂，通过闭式燃料循环进行工作，有助于实现可转换铀的有效转化并控制铀系元素。该系统能够在中心或地区燃料循环设施中实施完全再循环铀系元素的燃料循环。

MSR 全称为 Molten Salt Reactor，即熔盐堆。这是一种采用液态燃料进行核反应的特殊类型核反应堆 MSR 的堆芯采用盐冷却，这与传统的水冷反应堆不同，其使用熔盐作为冷却剂和/或燃料，具有更高的灵活性和效率。MSR 设计还包括了较小的高放废物足迹和被动安全功能等优点。

根据表格汇总情况，可以看出，PBMR 是最安全的选择，因为它在物理上不可能发生熔毁；如果我们寻求资源耐用性的最大化，MSR 则是最佳选择，因为它提供了使用钍的可能性，钍在自然界中的储量比铀更为丰富。

总结而言，可持续性最好从上述几个关键价值观的角度来理解，这些价值观既有空间维度也有时间维度。在构想负责任创新时，需要从伦理角度对利害关系进行事后分析，事后分析是事前分析的第一步。为此，我们讨论了核燃料循环的技术复杂性。我们认为，在选择特定的核能发电方法之前，要先评估每种类型的燃料循环，并了解各种反应堆设计的优缺点。这些反应堆中的每一个都可以实现某些价值，但不可避免地会在其他方面做出妥协。例如，最安全的核反应堆不一定最有效或最可持续。为实现核反应堆的负责任创新，必须了解和解决这些权衡取舍问题。

案例 2：当大数据遇到“老大哥”

通过画像和大数据分析，互联网可能比我们自己更了解我们，包括我们的弱点。我们使用互联网时每点击一次都会被 Cookie 记录下来，而这些信息大部分是在我们未曾知晓和同意的情况下收集的。对普通消费者来说，想在使用互联网

的同时不被数字化曝光，几乎是不可能的。目前，我们每个人被互联网收集到的相关信息，已经比秘密警察或极权主义国家情报机构所能收集的信息还要多。我们的隐私已经成为一种商品。这样的情况还会持续多久？

专家们一致认为，现在将大数据、人工智能、智能设备、物联网和量子计算结合起来的技术可能性，已经超越了乔治·奥威尔在《1984》以及阿道司·赫胥黎在《美丽新世界》中所描述的场景。

扩大政府、企业和其他组织对个人信息的获取权，以及提高预测人类行为的能力，将会带来什么后果？这是本案例要讨论的问题。

1. 公民评分系统

假设一个国家拥有公民评分系统。该系统是利用大数据分析技术的大众监控形式。公民评分是从统治者视角来衡量公民“价值”或“实用性”的数字，它将决定公民在未来享受到的产品和服务。按照系统规则，公民所做的每件事都能使他们加分或减分。为了确定公民评分，算法会对个人在互联网上点击的内容、观看的视频、收听的音乐，或者点击的链接进行分析。个人意见是否与政府立场一致，决定得分还是减分。该系统促使公民对他人进行监督，因为朋友和邻居的行为也会影响个人评分。你做什么、你买什么，全都会影响公民评分，并决定着你是否能取得贷款、能否获得某项工作或到某些地方旅行。更重要的是，在资源短缺的情况下，公民评分将决定谁能获得怎样的资源和服务，谁不能获得这些资源和服务，包括医疗服务和能源。对一些人来说，这可能个非常糟糕的消息。

这样的系统非常具有侵入性：“老大哥”正在密切监视你。这种系统已经在许多国家进行测试。比如，在英国，由秘密部门负责的“网络因果警察”项目被揭露投入使用。在下一个灾难出现或某个地区出现危机时，这种监控系统将会被启动，变成可能持续数十年的极权主义系统。此时，自主决定权、民主权和人权将在很大程度上丧失。

2. 无知和虚假新闻

很多人呼吁大家一起对抗科技极权主义。美国前总统奥巴马也告诫大家：我们的自由民主正受到无视科学和事实的势力攻击，这些势力如此强大，我们无法与之抗衡。公众舆论正越来越多地受到智能计算机程序的控制，如社交机器人。因此，我们或许已经进入了“后事实社会”，在这样的社会里，假新闻越来越多

地引导着公众舆论。比如，德国前外交部长施泰因迈尔将现代操纵技术产生的“后事实社会”称为民主的“致命危险”；埃隆·马斯克（特斯拉创始人）则称人工智能可能是人类所面临的最严重的威胁，甚至比核弹还危险。

3. 预测性算法

一些城市已经在使用可预测性警务。这意味着，人们在真正实施犯罪之前就会被逮捕甚至入狱。这些人在监狱里待的时间取决于算法对他们未来行为的预测。另一个例子是基因检测，不同公司会产生不同的检测结果。一种算法可能会对未来行为做出合理预测，但另一种算法也许不能。这无疑是一种我们应该关注的技术使用方式。

4. 其他热点

首先，网络安全。在网络安全方面，网络犯罪正在呈指数级增长。Cybersecurity Ventures 指出，网络犯罪给全球造成的损失，将从 2015 年的 3 万亿美元增至 2021 年的 6 万亿美元。尽管这些数据不完全准确，但是中国相关政府部门、美国五角大楼都曾遭遇黑客攻击，同样的情况在许多大型公司也发生过。显然，我们没有足够网络安全的系统。即使我们拥有强大的信息通讯技术系统（如超级情报系统），仍然可能被有组织的犯罪分子、恐怖分子或极端分子袭击，或者被独裁政治势力滥用。事实上，随着物联网的发展，人们对网络安全的担忧也越来越多。

其次，数据处理。计算机的处理能力正在指数级增长，而计算机产生数据量正在以更快速度增长。在短短一年时间内，我们生产的数据量相当于整个人类历史的数据总和。我们很难想象这意味着什么，但可以确信的是，我们将永远无法处理完或查看完这些不断增加的数据。这也表明，我们需要用科学的方法来判定哪些数据值得我们关注。此外，还有一个重要因素，即万物互联的增加使世界的相互依赖性和复杂性呈指数长，即使庞大的数据量也很难赶上复杂性增加的速度。这就产生一对矛盾：虽然我们拥有比以往更多的数据、更强的数据处理能力，但我们仍然面临对这些系统失去控制的风险。

总的来说，我们正处于一个十字路口，需要用心决定未来整个社会该何去何从。我们需要反思新的控制范式，并在不同层面进行检查和干预。所以是时候停下来三思而后行了。我们需要确保以民主的方式来操控这些技术，从科学和跨学

科的角度对这些技术进行研究，并确保这些技术合乎道德规范，而且必须维护使用过程的透明度和问责制，以及赔偿潜在受害者所遭受的损失。

讨论题：

1. 在创新管理中，风险性决策和不确定决策的区别是什么？
2. 在创新管理中，如何应对未知的未知，即极端不确定性？
3. 在你所从事的职业中，负责任创新的主要关键绩效指标是什么？
4. 如何看待有关核废料的代际正义问题？
5. 预防原则如何影响你对新技术生产的看法？

8. 风险管理和安全工程

“宜未雨而绸缪，毋临渴而掘井。”

——朱柏庐

8.1 风险管理和安全工程简介

在上一章，我们学习了风险以及如何在新设计之前或设计过程中预测可能的风险。但是，对于当前已部署的技术，也需要持续进行风险管理。此外，当面临多种选择时，我们需要决定发展哪项新技术，因为这些选择在不同程度上保留了不同价值，也做出了不同妥协。

让我们以安全价值为例。从抽象角度来看，安全价值是我们非常重视且不愿妥协的内容。安全技术的好处是显而易见的：我们能够有效地降低人员和财产受损的风险。然而，安全也需要付出一些成本——包括实施安全特性的机会成本，以及因剥夺其他机会成本而产生的连锁效应，还有实施其他价值维护特性所需的成本。同样，实施安全特性的好处是假设事故发生时能够节省下来的成本。因此，负责任创新研究应包括对新技术安全方面进行适当的经济评估，量化实施某种设计的净收益和成本，以及每种选择带来的风险。

在考虑大型项目时，风险和安全很快就会成为核心议题。风险和安全这两个词大家已经相当熟悉，但在本章节，我们需要精确定义并仔细考虑它们，以便使它们对我们有用。接下来，让我们从一些定义开始，或者重温一下前述章节出现的术语。本节的引导问题是：什么是风险？什么是安全（safety）？或者概括而言，即什么是保障（security）？鉴于这两个概念在大多数情况下含义非常接近，本章节不会深入探讨他们的概念区分。

8.2 风险和安全的定义

(1) 风险的定义

风险可以被定义为“某事件发生的概率乘以该事件发生后产生的成本或利益”。需要注意的是，这一定义是中性的，并不涉及我们对实际结果的判断。然而在使用中，风险通常仅与负面结果以及我们想避免的结果相关。

我们可以将这个定义简化为一个公式，即风险三元。风险三元是定义风险的三个问题集合：

- a) 可能发生什么错误？
- b) 发生错误的概率是多少（即可能性）？
- c) 产生的后果是什么？

当然，风险有不同类型，因此我们可以使用风险的两个维度——概率和后果——来做出一些区分。我们可以将风险描述为具有小概率和小后果的事件，比如被蜜蜂蜇住或被雷击中；或者将其描述为具有大概率但同样小后果的事件，如交通事故、从各种高度坠落等。目前，我们忽略了风险的第三个维度，即场景。诸多不同方式可能会产生各种后果。因此，当我们谈论一个后果时，明确这个后果是如何发生的也很重要。

(2) 安全 (Safety) 的定义

安全被定义为一种状态：即处于安全状态，免受伤害或损伤。这并不是一个客观的状态，因为人们也需要感到安全。由于安全是一种状态，对安全的感受可能从这一刻到下一刻会发生巨大变化。正如我们将要讨论，安全可能与其他需求或利益相冲突，如经济考量。

安全只有在存在威胁时才有意义。我们将这些威胁称为危害，它们被定义为“对生命、健康等构成威胁的情况”。之前，我们利用风险三元来定义风险，即场景——可能出什么问题，概率——发生的可能性有多大，后果——结果是什么。同样，危害也可以看作三元中的一部分，即场景和后果。危害的例子包括：驾驶汽车、运营化工厂或核反应堆，以及驾驶飞机等。后者不仅是对飞机上人员的一种危害，也是对地面人员和财产的一种危害。

(3) 保障 (Security) 的定义

保障又是什么呢？安全与保障之间的区别在于行为背后的意图。在保障方面，重点是可能的场景和一系列控制措施。然而，在安全方面，重点是旨在产生严重后果的故意行为。我们如何分析风险、安全与保障？在这一章，我们将讨论各种分析方法，包括成本收益分析，以及一些有效的工具和手段，如蝴蝶结图、故障树等。

8.3 成本收益分析

成本收益分析（CBA）是一种经济评估方法，它将某一决策的所有成本和后果用相同单位表示，通常是货币。成本收益分析无法直接证明某一安全投资是否在本质上优于另一种投资。然而，通过为决策者提供预防或缓解投资成本、获取适当收益的信息，成本收益分析有助于决策者改进决策。在资源有限情况下，为了达成多个目标和价值，成本收益分析在确定哪些投资最能有效利用资源方面非常有用。

（1）预测各种类型的事故和事件

在某些情况下，可以直接做出决策，但并非总是如此。例如，可能存在许多人们不希望发生的各种事件。

- 第一类事件可视为“职业事故”——例如，导致无法工作数天的事故或需要急救的事故。
- 第二类事件可归类为“重大事故”，涉及多人死亡或巨大的经济损失。第二类事件因此具有更多不确定性。
- 第三类事件可视为“黑天鹅”事件。对于第三类事件，没有任何信息可用，因此无法对这类事件进行经济分析。

第一类事件和第二类事件之间的经济考量有所不同。特别是对于后者，可以使用不均衡因子分析。

（2）净现值

成本收益分析的一个重要概念是净现值（NPV）。与安全相关的投资项目表示对当前资源（如金钱或时间）的分配。这些资源在当前投入，但在未来会产生回报。净现值考虑到了时间对货币价值的影响，通过计算项目的预期收益与成本的差额，并调整至同一时间点（通常是现在）来评估项目的经济吸引力。如果NPV为正，则意味着项目的收益超过了成本，是一个值得投资的项目。

分析的作用在于提供更为客观的评价可能性，但并不主张支持或反对任何一项安全投资，因为在做决策时还需要考虑许多其他因素，如社会接受度、伦理问题和监管事务等。如果决策者决定使用成本收益分析，接受或拒绝投资建议将基于以下过程：

- 确定成本和收益
- 计算所有成本和收益的现值

- 比较总成本和总收益的现值，从而确定净现值（NPV）

为了比较由不同时间点可能产生的总成本和总收益，需要在计算中使用折现率来表示真实的现值。本质上，是将所有现金流（包括未来可能发生的成本和收益）转换为现值。折现率指将未来有限期预期收益折算成现值的比率，表述愿意放弃当前消费以换取未来额外消费的速率。折现率越高，未来现金流的现值就越低。NPV 的计算公式如下：

$$NPV = \sum_{t=0}^T \frac{X_t}{(1+r)^t}$$

X_t 表示第 t 年的现金流， T 是考虑的时间段（通常以年为单位）， r 是折现率。NPV 计算非常有用，因为人们通常对（抽象的）未来经历的重视程度远低于（具体的）当前经历，他们对当前事件比未来事件更有把握。当所有现金流的总 NPV 为正时，可以对该项目进行投资。将 NPV 应用到安全领域，它表达的是通过投资安全项目带来的长期收益与初始投资和后续运营成本相比较的结果。如果 NPV 为正，则表明该安全投资从经济角度来看是可行的，并可能带来净收益。

（3）安全措施的成本与效益

与安全投资相关的成本种类繁多，我们可以将其分为几个明确类别，如初始成本、安装成本、运营成本、维护成本、检查成本等。这些成本显然是负现金流。其中一些成本（如初始成本和安装成本）发生在眼下，因此无需进行贴现处理；而其他成本（如运营、维护和检查成本）则贯穿设施的整个剩余寿命期，因此需要贴现至当前价值。

同样地，与安全投资相关的收益也分为不同类别。但如何解读这些收益呢？我们可以说，安全投资的目的在于降低当前和未来发生事故的风险。这意味着这些收益是假设性的，因为在大多数情况下，事故——或更准确地说，事故场景——实际上永远不会发生。因此，收益的定义是通过比较采取特定安全投资措施前后可能产生的后果差异来确定的，并且在适用时还要考虑事故发生可能性的差异。

由于大量的事故场景通常可以避免，因此，在计算任何单一事故场景的收益时，假设性收益将远大于成本。可以这样认为：现实中只会发生最可能的事故场景，许多其他可能的事故场景都将被避免。

这些收益表现为正现金流，都发生在未来，因此需要贴现至当前价值。与成本类似，我们同样可以将收益分为几个明确类别。如假设收益（或规避事故成本）可以像供应链利收益、保险收益、人类和环境收益、声誉利益等一样多样化。

（4）不均衡因子

最后，我们来探讨一下不均衡因子。现金流、预防成本和假定收益都可能相当不确定。为了应对这一事实，可以采用多种方法。例如，将现金流表示为预期值，以概率的形式考虑不确定性；或者提高贴现率以抵消不利结果的可能性。这对于第一类事故风险可行。

然而，第二类事故风险——即重大事故风险发生的频率极低，且不确定性水平很高。为了考虑这一点，成本收益分析最好包含一个不均衡因子，以反映安全与成本之间有意为之的偏向。在第二类风险情况下，可以使用情景分析对不同情景案例的现金流进行估算。例如，可以考虑最坏情况或最可信情况，并据此使用不均衡因子。如果该方法得出负净现值，则说明正在考虑的安全投资并非合理可行的，因为安全措施的成本与假定收益不成比例。

不均衡因子的大小在理想情况下很少超过 10。风险越高，该因子也越高，从而强调这些风险在成本收益分析中的重要性。这也表明，在风险非常高的情况下，使用大于 10 的不均衡因子也是可以接受的。

8.4 量化与比较风险

量化风险时的一个主要挑战在于风险本身缺乏统一的度量标准。这一点至关重要，因为稍有不慎，我们可能会陷入错误的类比，如同将苹果与香蕉相提并论。因此，在面对不同形式的量化风险时，我们必须保持高度警惕。为了公平比较不同活动的风险，务必确保始终采用一致的度量单位。

让我们通过一个例子来解析为何统一度量如此关键。在以下情境中，我们面对一组数字，但尚不清楚其具体含义。不妨假设这些数字代表的是个人一生中因某种特定原因而死亡的总概率。以吸烟为例，其致死概率是 5×10^{-3} ，这意味着在一千名吸烟者中，有五人因此丧命。同样，交通事故和雷击也存在一定的致死风险，而化学事故的致死概率则相对较低。此外，我们还可将彩票中奖的概率纳入考虑。实际上，中彩票的概率并不高，甚至远低于被蜜蜂蜇伤致死的概率。

人们常常疑惑为何我们选用死亡人数作为衡量标准。简单来说，死亡是一个

明确且易于达成共识的事实，而在评估不同伤害程度时，人们的看法往往难以统一。多年来，死亡人数确实成为了衡量总损失的有效指标。然而，在面对那些伤员众多、物质损失严重的灾难时，仅依赖死亡人数显然不够全面。

(1) 风险分析

现在我们探讨风险分析的主题。其核心问题在于：可能遭遇哪些风险事件，以及这些事件是如何发生的？（请注意，每个问题都可能不止有一个答案，每个答案都代表一个潜在的风险场景。）接下来，我们还需要进一步追问：这些风险场景发生的概率有多大？最后，当这些风险事件真的发生时，我们将会面临怎样的后果？

关于风险分析，存在两种主要方法：确定性方法和概率性方法。确定性方法通常聚焦于预防那些最可能发生的严重事故，比如核反应堆堆芯的暴露，或者装有危险化学品的卡车撞入建筑物。这些都是我们极力避免发生的情况，因此我们会竭尽所能去预防它们。而概率性方法则是通过分析特定事故发生率来评估风险。

风险分析是基于风险的决策分析的重要组成部分，如图 8-1 所示。

基于风险的决策分析包含八个步骤。在初始阶段，我们明确了环境和系统，旨在控制其潜在的危险；而在后续步骤中，我们实际进行了风险分析。

首先，我们明确了风险分析的背景和评估标准。为什么要进行这一步骤？可能是因为我们正在考虑一项新技术，并希望了解其是否可行。或者我们已经决定引入新技术，但想了解其中存在的风险。我们也可能想知道需要涉及哪些人，以便更好地控制风险，或者需要说服哪些人接受这一新技术的选择。无论我们的具体行动是什么，都需要确立决策的基础标准。这看似简单，但在实际操作中却相当复杂。

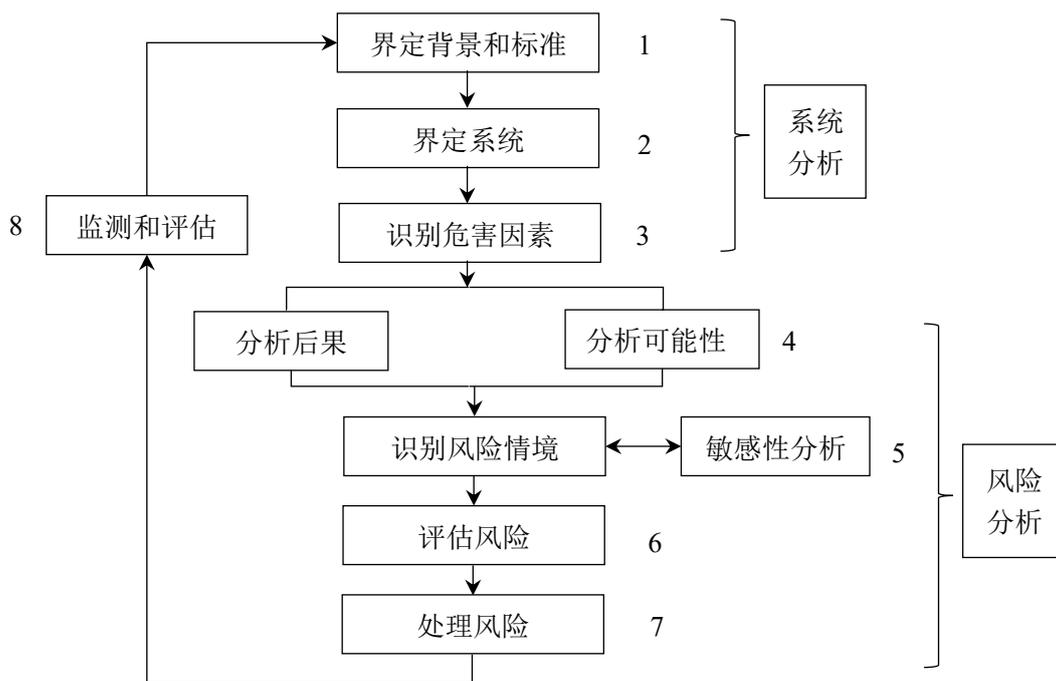


图 8-1 基于风险的决策分析

在荷兰以及其他国家，我们都区分了内部安全和外部安全。内部安全主要指的是职业安全，即工厂或现场工作人员的安全。外部安全则涉及工厂、反应堆或活动周围的所有环境。在此，我们主要关注外部安全。尽管这是一个主观决定，但考虑到每年发生的众多职业事故，职业安全同样至关重要。然而，这些事故更可能是每个行业的特定情况，因此，我们需要对外部安全给予特别的关注。

在外部安全领域，可以定义多种风险类型。首先是个人风险，后来也被称为局部风险，它指的是由于某种危险活动，在特定年份和特定地点导致一个人死亡的概率。其次是群体风险或社会风险，它关注的是每年以一定概率导致特定数量的人死亡的风险。最后，还有期望值，即每年平均死亡人数的预测值。

(2) 风险轮廓图

一年内一个或多个人在特定地点死亡的概率，可以被绘制成我们所说的风险轮廓图。这些轮廓图在地图上清晰可见，如图 8-2，它们环绕着潜在的危险区域，如化工厂或核电站。同一条轮廓线上的所有点具有相同可能性。按照惯例，这些轮廓线是根据 10 的（负）指数来绘制的，如 5×10^{-5} 、 5×10^{-6} 等。

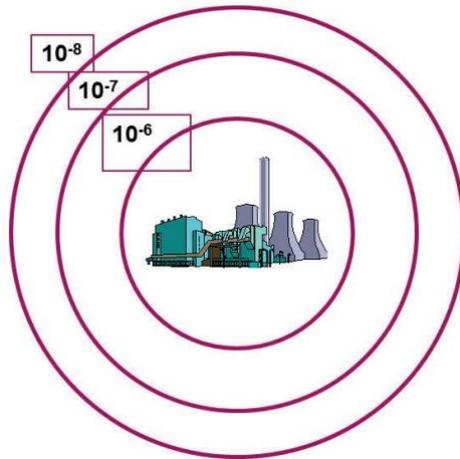


图 8-2 风险轮廓图

群体风险则通常通过 FN-3 曲线来表述，该曲线以年份为横轴，死亡人数为纵轴。在下图中，我们可以看到某项特定活动（例如某个特定化工厂的活动）以及该活动导致一定数量伤亡的频率。以特定装置为例，其事故发生的频率大约是每 80,000 年一次，我们预期每次事故将导致 6 至 7 人伤亡。在图表中，两条线代表了参与方之间达成的共识，界定了“合理允许”的伤亡人数以及这些伤亡事件可能发生的频率。从图 8-3 中我们可以看到，该装置的伤亡情况低于这一标准。

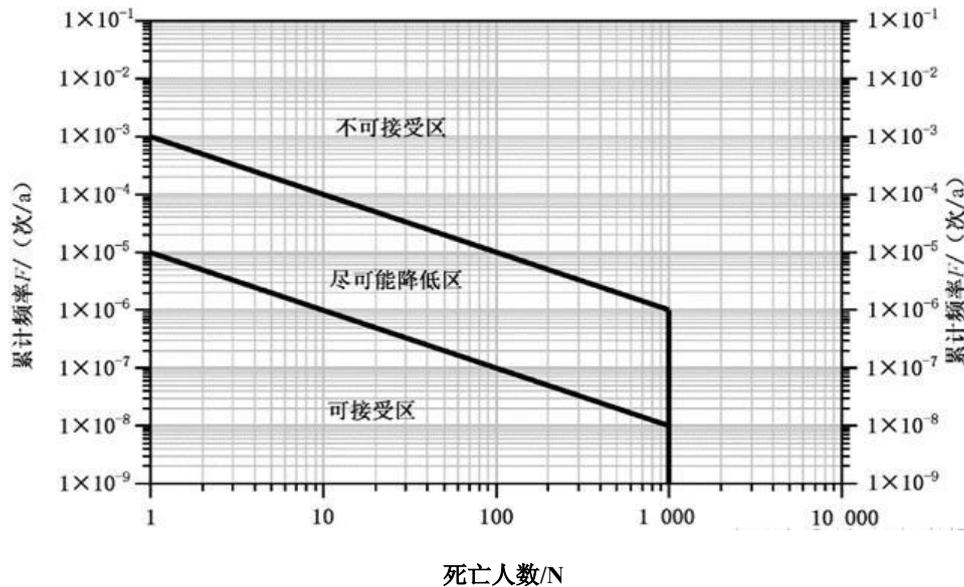


图 8-3 FN 曲线

有一个有趣现象，它清楚地表明局部风险和群体风险这两个概念之间存在某种张力。比如，房屋建在风险轮廓线 5×10^{-6} 之外。这意味着每 1000 万年，可

能会因为化工厂泄漏气体而导致一定数量的人员伤亡。而这些房屋位于百万年一遇的风险轮廓线之外，因此各方都认为这是一个可接受的风险水平。

现在，化工厂决定采取进一步的安全措施。由于这些新措施以及基于风险轮廓线达成的共识，房屋可以建得更靠近化工厂了。然而，如果真的发生气体云泄漏，预期会出现更多伤亡。因此，局部风险和群体风险之间可能会产生冲突，我们需要商定哪一个更为重要或优先考虑。

(3) 界定系统和边界

在确定了情境和标准后，我们必须进一步定义系统的组成部分以及我们要控制的危害。对于正在审视的系统边界，我们需要深思熟虑。我们聚焦的是某个工厂中的特定设施，还是仅关注整个工厂或一个包含多个工厂的大型工作场所？我们的目标是内部安全——即工人面临的风险，还是外部安全——系统外的所有潜在威胁？此外，我们期望探讨的细节层次如何？是深入到每一条管道、每一个容器和挡板，还是只关注特定的风险和活动？

明确系统及其边界具有重要现实意义。从某种意义上说，系统边界内发生的事件是我们可以预防或控制的结果；对于系统边界外的事件，我们只能预测和管理，无法直接控制结果。这可能是因为它们超出了我们的控制范围，或者我们可用的资源有限。此外，我们必须清楚地说明哪些内容在分析范围内、哪些不在分析范围内，以确保所有利益相关者能准确了解分析所涵盖的内容。

(4) 危害分析

在明确了背景和系统之后，接下来需要考虑可能存在的危害。在这个过程中，我们将围绕三个核心问题展开思考：可能会出现什么问题？这些问题会如何发生？我们该采取什么措施或控制手段来应对这些危害？通过这一步骤，我们可以更深入地理解我们所研究的系统。在全面研究并解答这些问题之后，才能进一步分析识别出的危害。不言而喻，这一步骤对后续分析以及整个过程的有效性至关重要。

在识别危害时，可以采用的方法有：

- 初步危害分析检查表或标准列表 (Standard list or checklist Preliminary Hazard Analysis, PHA)
- 危害识别研究 (Hazard Identification study, HAZID)
- 危害与可操作性研究 (Hazard and Operability study, HAZOP)

- 故障模式与影响分析 (Failure Mode and Effect Analysis, FMEA)
- 故障模式影响及危害度分析 (Failure Mode Effect and Criticality Analysis, FMECA)
- 故障树分析 (Fault Tree Analysis, FTA)
- 过去经验分析 (如事件、事故报告或数据库)

这些方法在危害研究中经常出现，每种方法都具有独特的优势和局限性。本节将重点关注故障树分析和蝴蝶结模型 (Bow-Tie Model) 两种方法。

首先，故障树分析方法。故障树分析的是一系列逻辑关联事件，这些事件最终可能导致某个不希望发生的顶层事件 (top event) 发生。故障树具有逻辑结构，可以利用它来量化风险。尽管故障树以特定的顶层事件为中心，但我们也可以前瞻性地使用故障树来分析尚未发生的事件。在故障树分析中，我们可以从顶层事件开始，系统地向下追溯各个事件，直到我们决定停止的地方。理论上，故障树可以无限延伸。

接下来，以房间里有两个灯为例。当按下两个灯的开关时，灯没有亮起，此时房间黑暗就成为顶层事件，导致这种情况可能有两个原因：要么是电源故障，要么是两个灯都出现了故障。连接两个子事件到顶层事件的小符号，被称为“门”。在这个例子中，它是一个“或门” (OR-gate)，“或门”意味着只要其中一个子事件发生，顶层事件就会被触发。还有一种“与门” (AND-gate)，可以用平底表示，“与门”意味着两个子事件同时发生，顶层事件才会被触发。

现在，仔细观察这个系统 (图 8-4)，可以看到一个简单的路线，包括电源、保险丝、开关和两个灯具。开关已经被尝试过，但灯仍然没有亮起。在此基础上，可以进一步完善故障树。由于两个灯都没有亮起，因此这两个灯必然出现了故障，可以添加一个“与门”。然而，灯没有亮也可能是电源供应问题。在这个系统中，电源供应有三个潜在故障来源，即保险丝、开关和电力来源。另一方面，单个元件的故障也足以导致房间黑暗。因此，可以添加一个“或门”。

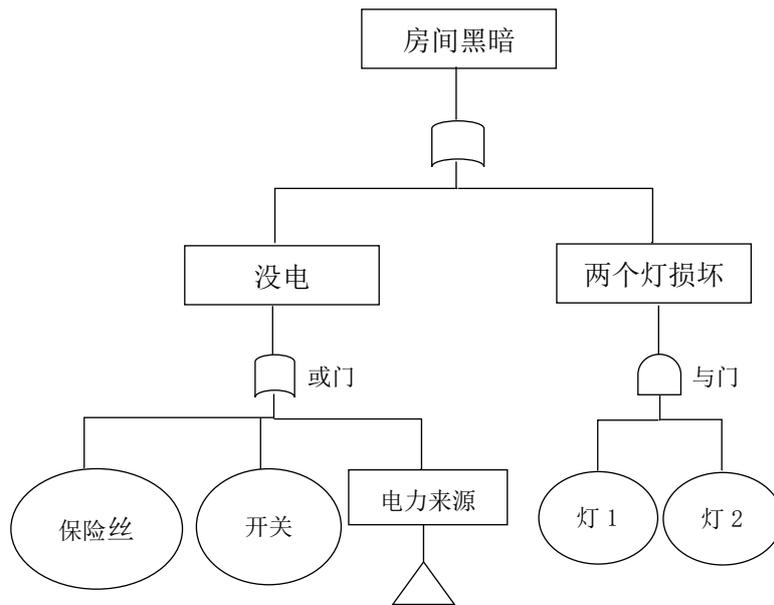


图 8-4 故障树分析

这个小例子清楚地阐明了几个要点。首先，需要清晰地定义背景和系统。我们本可以包括整条街道或邻里的电源供应。我们还可以包括电源供应可能出现故障的许多其他情况。本质上，我们必须根据有效预防或控制的内容来选择系统边界，并且只能适应或管理边界之外的事件。这些决策对于我们的分析至关重要。最后，如果我们知道相关概率，就可以将它们添加到故障树中。那么，保险丝烧毁或灯泡失效的概率是多少？这些数值通常可以从行业通用的数据表中查询得到。利用特定计算方法，可以计算出当房间有两个灯时，房间变暗的概率。故障树不仅直观地展示了各种可能的场景，这些场景也可以被看作是通过故障树的不同路径。通过故障树，能够清晰看到不同故障模式如何相互关联并影响最终结果的。这有助于更全面系统地理解潜在风险，并为制定风险管理策略提供有力支持。

其次，蝴蝶结模型图。蝴蝶结模型图（Bow-Tie Diagram），又称“领结模型图”，是一种风险评价方法，特别适用于分析高风险情境中的因果关系。由于绘制出的图形形状酷似蝴蝶结，因此被称之为蝴蝶结模型图。如图 8-5 所示，故障树占据了蝴蝶结左侧的部分，呈现出不同的形态。

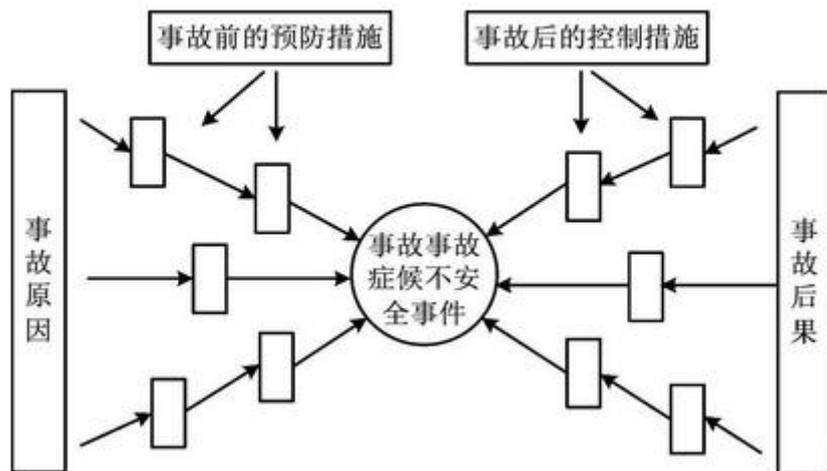


图 8-5 蝴蝶结模型图

蝴蝶结图仅包含“或”门（OR-gates），主要用于定性分析，并不涉及概率的添加。图中的中心事件通常被定义为失去控制的节点。在图左侧，事件按照导致中心事件发生故障的顺序排列；在图右侧，事件则按照顶层事件及其相关连锁事件的后果顺序排列。简而言之，左侧展示了事件失败原因，右侧则描绘了事件后果。

（5）后果分析

风险分析的下一步是探讨某一情境可能带来的后果。通常，会从死亡、受伤或经济损失等角度来定义这些后果。

首先，需要就评估标准达成共识，否则就是在比较不同事物。此外，还要确定时间范围是怎样的？大型事故往往会产生深远后果，这些后果可能会持续很长时间。例如，死亡事故往往会对家庭和公司产生巨大影响。评估这些影响，需要具备与研究领域相关的专业知识。

表 8-1 不同损失类别示例

	轻微损失	较严重损失	严重损失	灾难性损失
设备损坏 和生产损失	短期生产损失	机器损坏 短期内可修复	重大维修成本 严重生产损失	设备严重损坏
环境损坏	排放水平暂时超标	大量泄漏 需要清理	一年内造成 生态损害	超过一年 的生态损害
对人的伤害	非致残性伤害	致残或重伤	严重伤害	一人或多人死亡

为了简化分析过程，可以将后果分为有限的几类，一些企业也通常采取这种做法。在表 8-1 的示例中，针对三个不同目标群体（设备、环境、工厂员工）定义了四个损失类别——轻微、较严重、严重和灾难性。每个类别都附有简短描述。这样的矩阵可以辅助我们讨论损失和目标群体。同时，需要注意的是，这些后果可能与多种情境相关。

在列出所有潜在后果以后——这些后果至少应在所考虑的系统范围内是全面的——需要为这些结果分配概率。这并不是必要步骤，一般在能够量化或应该量化某些结果时才这样做。为了对比不同后果，还必须明确使用哪个基准并对此达成共识，以及明确分析的细致程度和希望量化的后果类型。

(6) 预测风险场景

在风险分析的后续步骤，需要识别风险场景。这一步主要提出各种不同问题，以便尽可能多地识别潜在的薄弱环节和结果。具体来说，根据概率来量化并排序这些风险场景。然后决定是全面预防这些场景，还是通过概率方法来应对它们。

此外，还可以考虑进行敏感性分析，以检验分析是否合理并识别潜在的问题环节。此时可以深入审查数据源，比如，是否考虑了系统中存在的各种变量？是否捕捉到了所有相关因素、相关作用和彼此关系？所使用的数据质量如何？同时，还会需要细致审视分析结果。如果输入变量发生变化，或重新构建关系模型，结果会如何？在每个模型中做出一些假设，如果这些假设不成立会产生什么后果。在这种情况下，风险会如何增加？

(7) 风险评估

风险分析的第六步是风险评估。为了简化这一过程，可以采取类别和分类方法。当然，这并不是一个硬性规定，所有利益相关者应就这些分类达成一致。具体可以将频率和结果组合成一个矩阵。

在这个特定矩阵中（表 8-2），可以区分五个频率类别和四个结果类别。然后，给每个单元格标上特定结果，并将其与特定频率相关联。此外，还有五个严重程度类别。这个矩阵通常指出了如何处理这些风险。一些风险在矩阵中是被接受的，而其他风险则需要预防或使用保险。

我们用四种风险颜色表示。从红色开始，这表示继续开发之前必须在设计中进行更改；到深绿色，这表示风险可以忽略或容易用现有措施处理；介于两者之

间的颜色，即浅绿色和黄色。计算结果处于黄色或浅绿色单元格的风险可能会随着新见解、新技术的出现而降低。相反，未预见的事件或新知识可能会使这些风险变得更加严重。

表 8-2 风险评估矩阵

发生频率	不同风险类别			
	I 灾难性风险	II 较严重风险	III 边际风险	IV 微不足道的风险
频繁 (A)	1A	2A	3A	4A
可能 (B)	1B	2B	3B	4B
偶发 (C)	1C	2C	3C	4C
少有 (D)	1D	2D	3D	4D
不太可能 (E)	1E	2E	3E	4E

在决定每个风险类别使用哪种颜色时，通常可以采用 ALARA 原则，即“尽可能合理的低水平”原则 (As Low As Reasonably Achievable)。它界定了可接受或可容忍与不可接受或不可容忍之间的界限。红色风险，即不可接受的风险，需要被明确处理掉，因此应该消除、预防或充分投保。哪些风险是可以容忍的，则取决于成本和实际性的考虑。这些考虑因素当然可以改变，并需要经过一些协商。

(8) 安全措施

最后，本节讨论如何处理已经识别出的风险。在这个环节，安全性至关重要。基本上，存在四种应对策略来保证安全：风险规避、风险降低、风险转移和风险接受。这些措施可以一个形象事例来说明——“一个人正面临鳄鱼的威胁”。对于这种情况可以选择：①彻底消除风险，杀死鳄鱼，这是一种风险规避措施，安全效果最明显；②与鳄鱼保持安全距离，减少接触危，这是一种风险降低措施，安全效果次之；③将鳄鱼关进笼子，限制其活动范围，这也是一种风险降低措施，但风险效果次于前者；④穿上防护装备来保护自己，这种一种降低风险后果的措施，安全效果最次。

类似上述应对策略的框架被称为哈顿矩阵。威廉·哈顿是一位医生，他尝试思考并总结出所有可能的策略，并根据效果进行排序，最终以矩阵的形式进行可

视化展现。在这些策略中，消除危险源是最为有效的策略，其次是尽量减少与危险源的接触，从而降低其发生的概率；此外防止危险发生，也能降低其出现的概率；或者调整危险发生的方式，以最大程度地减少损害。哈顿共定义了十种不同的策略，其中最后几种与特定方式的后果应对相关。

有趣的是，我们还可以将哈顿矩阵给出的策略与蝴蝶结模型相结合。在失去控制之前，我们有时间采取预防措施；一旦失去控制，便可以采取缓解策略以应对后果。具体来说，可以尝试以下几种方法：①避免负面结果发生；②降低负面结果出现的可能性；③最小化负面结果的后果；④转移风险，例如通过保险手段；⑤接受并学会与风险共存。

另一个有助于思考风险应对措施的简洁模型是“危险——屏障——目标”模型。该模型描述了我们经常面临的情况：某个目标可能受到某种危险的威胁，但该目标受到一个或多个屏障保护。这些屏障可以是物理性的、程序性的，或者是两者的结合。它们的作用是阻止不希望的能量流向目标，从而防止造成伤害。在蝴蝶结模型中，这些屏障可以被看作被阻塞的路径或场景。然而，遗憾的是，总会有一些路径会保持开放状态。当发生事故或破坏行为成功时，这些路径就会被暴露出来。因此，我们需要对这些开放路径进行密切监控。

(9) 风险分析实践

前述章节已经深入阐述了风险分析的主要步骤。需要注意的是，这并非一个线性、有固定终点的过程。实际上，风险分析是一个永无止境的循环，需要不断地进行预测、准备、预防和缓解。

系统不断变化、调整，人员及其行为不断改变。对此，我们必须时刻保持警觉，并将这些变化纳入考虑范围。事故无法避免，但可以从中吸取教训。随着时间推移和个人努力，相关知识和资源库会不断增长，因此，根据“尽可能合理的低水平”原则，风险分析和缓解措施也需要持续做出相应调整和完善。

讨论题：

1. 在复杂的社会技术系统中，组织的任何一次失败都可能造成巨大的潜在负面影响。这种复杂性和不断变化对风险评估意味着什么？
2. 成本收益分析尽管提供了系统性框架，但该实践仍存在伦理方面的担忧。如何对无法估价的事物进行定价？你怎么看？

3. 成本收益分析方法论背后的基本价值观是什么？这些价值观是我们想要强调的吗？
4. 结合蝴蝶结（Bow-Tie）模型，尝试讨论该模型在机场安全风险中的应用。

第六部分 价值敏感设计

9. 价值敏感设计

“设计不仅关乎外观和感觉，更关乎功能和价值。”

（“Design is not just what it looks like and feels like. Design is how it works.”）

——史蒂夫·乔布斯

9.1 价值敏感设计简介

（1）价值敏感设计的发展阶段与核心思想

在负责任创新的深入探讨中，尤其要聚焦价值敏感设计（Value Sensitive Design, VSD）所扮演的核心角色。价值敏感设计旨在构建一个更具实践指导意义的框架，将抽象的道德价值观融入技术与创新实践，进而转化为具体可感知的技术指标与参数。值得一提的是，价值敏感设计常被等同为价值观设计，二者概念相通。

初期阶段：追溯历史，VSD 起源于计算机科学的一个分支，但其影响力远远超出了技术范畴。20 世纪中叶，随着计算机的出现，研究焦点多集中于技术本身，人们将其看作解决多学科问题的通用工具。然而，在初期阶段，人们往往忽视了计算机技术的社会背景及其使用者的需求。

第二发展阶段：20 世纪 70 至 80 年代，计算机应用的广度与深度显著增强，并渗透到各类现实组织，服务于不同工作环境、社会及机构背景下的多样化用户群体。这一背景下，社会与行为科学领域（如人机交互、参与式设计、社会信息学等）与信息技术应用的融合日益紧密。在该阶段，人们对社会背景、使用模式及用户行为的关注，初衷在于识别并消除系统实施中的潜在障碍，以保障成功和规避投资风险。但这一趋势最终推动了用户友好性、可用性及用户接受度研究的兴起。

价值敏感设计的第三发展阶段：进入 21 世纪，人们愈发认识到信息技术的成功应用不仅关乎用户友好，更需要契合广泛的人类价值观。无论身份如何，人们均持有各自的道德观念、偏好与理想。社会对于平等、财产、隐私、可持续性、

自主性及责任性等议题的道德讨论持续进行，计算机网络和系统也应该在可能且适当的情况下，以某种方式或形式容纳这些价值观。

近十年来，价值观已从单纯的技术实施限制因素，转变为信息技术发展的核心目标与驱动力。例如，2001 年加利福尼亚州成立了“以社会利益为核心的信息技术研究中心（CITRIS）”，标志着信息技术发展迈入新阶段，即人类用户（无论其身份为雇主、消费者、公民或患者）的需求与价值观成为关键考量。信息技术被视为服务人类——作为道德个体——在个体道德和社会事物中的工具与技术支撑。

因此，我们需要转变视角，从单纯的技术细节讨论转向在社会语境中构建技术。同时，道德价值观也应该从被动遵守约束向技术积极服务于社会需求与目标转变。这一转变在“价值敏感设计”中得到了深刻体现，并在过去十年获得了广泛认同与应用。价值敏感设计的核心在于，道德价值观能够被转化为工程语言，使我们在设计与制造过程中融入有关可持续性、安全性、隐私等伦理思考的具体成果，从而创造出更加符合人类道德价值观的技术产品与解决方案。

（2）价值敏感设计的思想线索汇集

价值敏感设计这一概念汇聚了多条相互交织的思想与研究线索：

① “人造物是否蕴含政治色彩？”

首先，这一思想脉络的标志性进展是 Langdon Winner 于 1980 年撰写的开创性论文《人造物具有政治性吗？》（参见第 1 章）。该论文深刻指出，价值观能够在现实世界的物体和技术中得以体现，并深远地影响人类行为。Langdon Winner 通过鲜活实例，阐述了技术及其内含的价值观、政治观点和权力如何塑造并限制人的行为，对技术设计伦理问题的思考产生了深远影响。

纽约桥梁的低垂式立交桥便是一个引人深思的案例（见第 1 章 1.2 节）。著名建筑师与城市规划师罗伯特·莫斯（Robert Moses）故意将纽约高速公路上的立交桥设计得比较低，以便使汽车顺利通过、公交车无法通行。这一设计的社会文化影响在于，拥有汽车的白人中产阶级能够轻松前往对岸的琼斯海滩，而更多依赖公交出行的贫困黑人社区居民则被无形阻隔。间接地，这座立交桥成为了种族隔离的边界机制，将富人与穷人、白人与黑人分隔开来。

尽管这一案例的历史真实性存在争议，但我们一旦了解这一实例，便能即刻

体会到价值观是如何深刻而又不易察觉地融入我们周遭的事物之中，从而对我们的生活产生深远影响。

②科学与技术研究

20 世纪 80 年代的其他研究同样聚焦于技术的哲学与社会学层面，这一领域被称为科学和技术研究（Science and Technology Studies, STS）。科学与技术研究揭示了众多实例，并通过详尽的案例研究证明，社会政治偏见（特别是种族、性别和收入方面的偏见）能够深刻地烙印在技术产品、系统及基础设施之中。Geoff Bowker、Susan Leigh Starr 和 Lucy Suchman 等研究人员在此领域做出了卓越贡献。

③工程师的担忧

与此同时，一些设计与工程的专业领域也开始采用“价值设计”或“价值敏感设计”的基本概念。以下两个例子可以说明这一点：

首先，让我们简要回顾隐私增强技术。20 世纪 80 年代，一些隐私学者开始探索如何设计 IT 系统和应用程序，以提高用户遵守隐私规范的可能性。这种方法不再仅仅依赖于用户的自觉性，而是通过产品设计本身，使遵守隐私规定成为用户的自然行为。

另一个例子来自于 20 世纪 80 年代的建筑与环境研究。在建筑设计与城市规划中，已开始采取措施预先规划安全性并预防犯罪。研究表明，照明、建筑多样性、建筑间距及视线等因素均对降低犯罪率具有积极影响。因此，这些因素被精心融入建筑和社区的设计参数之中。

9.2 价值敏感设计方法定义

价值敏感设计这一概念最清晰、最精确的阐述，起源于 20 世纪 70 至 80 年代斯坦福大学计算机科学领域的一场运动，该运动由特里·维诺格拉德（Terry Winograd）大力推动。如今，这一概念已被多个研究团体广泛接纳，并统一称为价值敏感设计。

价值敏感设计是 20 世纪末期由巴提亚·弗里德曼（Batya Friedman）等人发展出的一种系统开发与软件工程方法论。他们汲取了人机交互（HCI）领域的智慧，旨在提升人们在设计过程中重视社会与道德维度。价值敏感设计的核心在于将广泛的人类价值观与道德考量融入信息技术的设计之中。

弗里德曼曾提到：“最终，价值敏感设计要求我们拓宽对技术系统质量评判的视野与标准，以纳入那些能够弘扬人类价值观的元素。”尽管价值敏感设计并未预设特定的规范框架，但据弗里德曼所述，该方法的核心关注点在于以人类福祉、尊严、正义、福利及人权为核心的价值观。它旨在连接设计系统界面的专业人员与那些能够深刻理解并反映利益相关者价值观的人员。

代尔夫特理工大学将价值敏感设计视为一种应用伦理学方法，其目的在于将道德价值观融入技术设计、研究及发展的全过程。价值敏感设计依托的主要方法论结构是一个综合迭代的三层次框架，涵盖概念、实证和技术调查。这三者均呈迭代关系，相互提供信息并受其他调查结果的影响。更多具体详情可以参见《价值敏感设计与信息系统》一文^①。

价值敏感设计展现了一系列与负责任创新相契合的特性。它要求在设计过程的可塑性阶段，明确阐述所有利益相关者的价值观与道德关切；这些关切需要以引导设计决策的方式明确界定；同时，设计与产品还需要根据秉持的价值观与道德关切进行评估。值得注意的是，尽管价值敏感设计起源于信息技术和计算机科学领域，但其应用范围非常广泛，它与各类新技术创新设计，以及技术产品推广部署均紧密相连。

9.3 价值敏感设计的实践应用

上述章节已经深刻阐述了价值敏感设计理念的重要性。在当今错综复杂的社会技术体系中，我们面临着前所未有的挑战。作为个体与社会成员，我们各自坚守并珍视着安全、可持续性、公正、隐私及人类福祉等核心价值观。在过去，这些价值观主要通过人类行为、法律框架及政府政策等手段得以维护与彰显。

然而，随着科技的飞速发展，我们已经置身于由技术深刻塑造世界的情况之中。我们只需要简单回顾一下，就不难发现今天的生活方式与过去几代人的生活方存在天壤之别，尤其是互联网、计算机及智能手机的广泛普及。面对这种情境，我们的新挑战在于：如何确保这些日新月异的技术能够忠实反映并践行我们所珍视的价值观。

因此，我们需要将价值观与理念世界转变为技术与物质世界——这是一项非

^①Friedman, B., Kahn, P.H., Boring, A., Hultgren, A. (2013). Value Sensitive Design and Information Systems. In: Doorn, N., Schuurbiers, D., van de Poel, I., Gorman, M. (eds) Early engagement and new technologies: Opening up the laboratory. Philosophy of Engineering and Technology, vol 16. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-7844-3_4

常艰难的任务，毕竟这两个领域曾长期处于割裂状态。这自然而然地引出了后续的思考：如何在设计过程中有效体现价值观？

(1) 技术是否体现了价值观？

针对这一问题，技术如果能体现价值观，那么是如何体现的？接下来我们将从三种立场来回答这一个问题：工具主义、实体主义和交互主义。

- 工具主义

工具主义认为技术是价值中立的，因为技术只是人类手中的工具。一种技术是否服务或阻碍某种价值，取决于人民如何使用它。例如，一把面包刀可以用来切面包，但也可以用来杀人。美国全国步枪协会的口号“枪不会杀人，但人会杀人”便是对工具主义的生动诠释。然而，不应该忽视的是，技术的存在往往能轻易放大或改变人的行为模式。比如，用枪杀人比不用枪要容易得多；家中遭盗贼入侵时，你手上有枪与没有枪的反应会截然不同。

- 实体主义

实体主义认为技术本身是具有价值观的，并且人类无法对其产生影响。例如，有人认为技术体现了诸如效率之类的价值观，也有人担忧技术会导致环境恶化、真实感缺失以及降低人际互动。然而，这一观点忽略了人类可以通过使用技术、设计来施加影响。

- 交互主义

交互主义是我们要捍卫的立场。该立场认为，价值是在人与技术的互动过程中被创造和嵌入的，它既体现在技术的使用上，也体现在技术的设计中。在本书中，重点关注设计方面。

(2) 技术设计应该包含哪些价值观？

技术设计应该融入哪些价值观，这是一个复杂而深刻的问题。在工程设计中有许多重要价值观，它们来源广泛，比如设计理念（说明项目动机）、设计师及其专业团体、用户与利益相关者、法律和政府政策、技术规范与标准、伦理道德准则。然而，即便列出所有这些价值观，我们也无法确定应该包含哪些价值观，因为这是一个规范性问题——一个关于我们应该做什么的问题。

价值多元主义进一步加剧了这一问题的复杂性，因为不同人对价值观的重要性排序可能大相径庭，从而导致在设计中决定包含哪些价值观变得更加困难。但

即便如此，我们仍有理由相信在设计中融入价值观也并非不可能：

首先，尽管价值多元，但对设计中需要整合的某些核心价值观，人们往往能达成共识。

其次，尽管人们就价值观的重要性存在分歧，但他们仍然可能考虑一系列广泛而相关的价值观。例如，人们就技术设计中的安全性和可持续性产生分歧，但大多数人都会同意，无论如何，安全性和可持续性都应该被纳入新汽车的设计中。

第三，在设计技术时，也存在尊重不同群体与利益相关者多元价值观的方式。

9.4 如何将道德价值观转化为设计规范？

(1) 工具价值观和内在价值观

当涉及哪种价值观最重要时，哲学家们通常在工具价值观和内在价值观之间做出区分。工具价值表示对其他事物具有重要性的价值。例如，金钱通常被认为具有工具价值，因为它有助于我们实现生活中的其他重要目标和价值观。另一方面，内在价值是对自身而言非常重要的价值，不是用来实现其他目标。典型的内在价值包括幸福、正义、美丽、诚实和真理。

那么，我们应该如何将抽象的道德价值观转化为具体有效的设计要求。为了回答这一问题，本节将讨论一个价值层次结构。价值层次结构有三个层级：价值观、规范 and 设计要求，如图 9-1 所示。

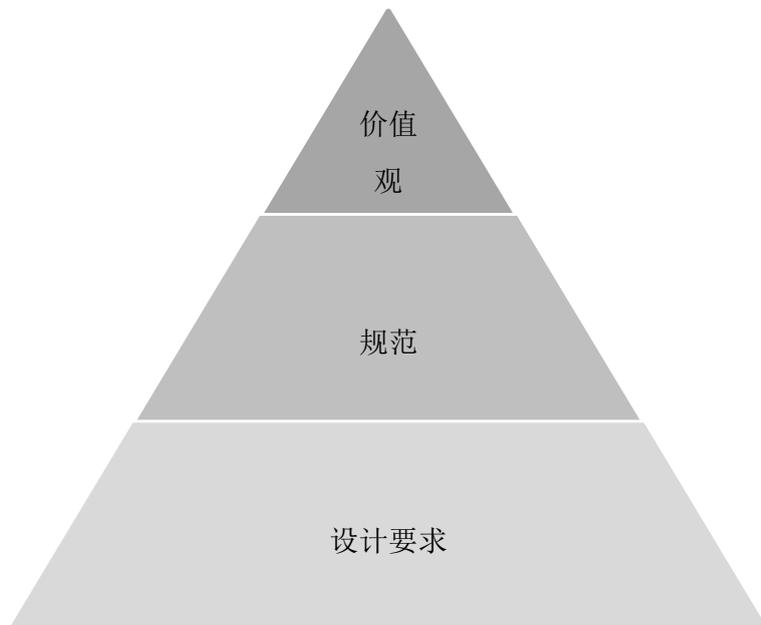


图 9-1 价值层次结构

(2) 价值敏感设计的两个案例：蛋鸡笼子和生物燃料

案例 1：动物福利

图 9-2 是基于欧盟指令对蛋鸡笼进行设计，旨在保证笼子的设计能充分考虑动物的福利价值。可以看出，这个价值可以转化成几种具体规范。例如，规定鸡应该有足够的生活空间，这种规范然后被转化为更加具体的设计要求——生活空间至少要有 450 平方厘米的地板面积。需要说明的是，本案例中的价值层次结构以欧洲法律为基础，我们自己也可以根据实际情况构建相应的价值层次结构。

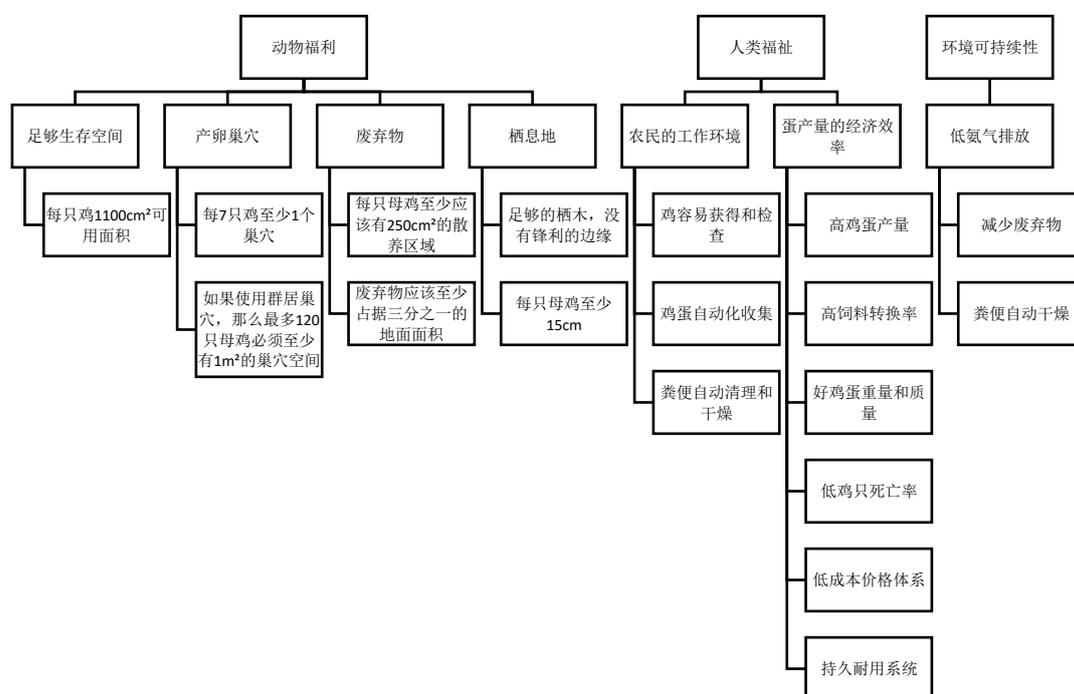


图 9-2 蛋鸡笼的价值敏感设计矩阵

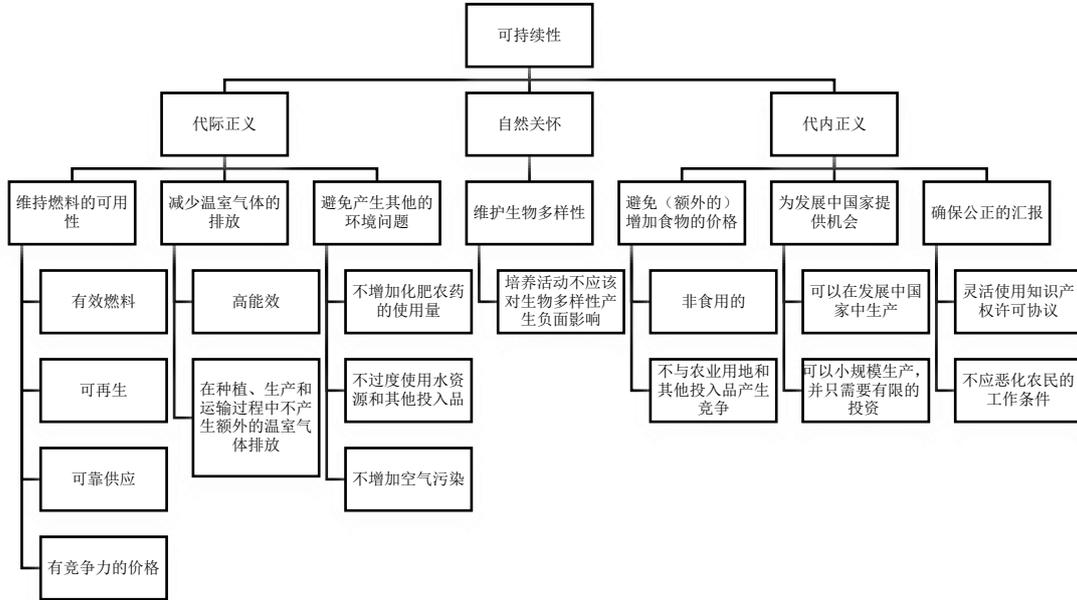
案例 2：生物燃料

另一案例是为生物燃料的使用制定价值观层级结构。生物燃料以相对较新的无生命生物或有生命生物为材料。使用生物燃料的目的是应对化石燃料短缺、减少温室气体排放。然而，它们却因为对环境、食品生产、食品价格的影响而受到严厉批评。类似于纳菲尔德生物伦理委员会（Nuffield Council on Bioethics）^①的组织，已经制定了使用生物燃料需要满足的伦理原则，以便其在道德上可以

^①<https://www.nuffieldbioethics.org/>

被接受。图 9-3 展示了生物燃料所有相关规范和设计要求组织成的价值观层次结构。

图 9-3 生物燃料的价值敏感设计矩阵



该价值观的顶层是可持续性价值观，可持续性是生物燃料发展的主要价值观。这一价值被分解为三个象征可持续性的具体价值观，即代际公平、自然关怀和代内公平。

与每个价值观相关联的是一系列规范。首先就代际公平而言，与其相关联的规范有三种：需要维护燃料的可用性、减少温室气体排放、避免其他环境问题增加。每个规范依次被转化为一系列更具体的设计要求。例如，燃料可用性规范意味着燃料应该有效、可再生、可靠，并且价格具有竞争力；避免其他问题增加的规范如避免食品价格上涨，意味着生物燃料不可食用，不应该与农业用地产生竞争。然而，在实际使用中，目前还不存在符合所有这些要求的生物燃料。当前大多数生物燃料都属于第一代或第二代，它们是可食用的或者与食物作物争夺土地。目前正在开发第三代生物燃料，据说可以解决这些问题。

价值观层次结构可以从某个价值观开始进行自上而下地构建，比如动物福利或可持续性。然后，在层次结构中向下详细说明这些价值观。价值观层次结构也

可以自下而上地构建，从给定的设计要求开始。在这种情况下需要解决的关键问题是：这些要求实现了什么样的最终目标？

一个重要问题是，价值观在价值层次中的具体规定是否充分。我们可以质疑符合低层次的设计规范是否有助于满足高层次的规范或价值观（即自下而上的方式：从设计要求到价值观）？重新回顾动物福利和蛋鸡笼子的例子，其问题在于，满足有关设计要求是否能够实现动物福利的价值观。许多人会怀疑这是否成立。事实上，欧盟已经修改了法律并制定了更严格的设计要求，从根本上禁止了使用蓄电池笼饲养。

通过上述方式，价值观层次有助于将抽象的道德价值转化为具体可见的设计要求，并将这些要求结构化。

9.5 复杂的过程

上述案例深刻揭示了价值观向设计要求转化的多维度过程，这一过程往往错综复杂。

首先，将价值观（尤其是新兴价值观）转化为具体设计要求，往往是一个漫长繁琐的过程，需要时间与耐心的双重投入。

其次，这一转化过程需要特定的专业知识，有时这些知识可能源自工程领域之外。正如前述讨论的案例中，动物行为学提供了相关专业信息；在环境价值观的案例中，则需要环境科学或生态学的专业知识；对隐私与信任等价值观而言，哲学分析有助于更好地理解这些价值观，并将它们转化为更具体的规范。即便工程领域所熟知的安全性、可用性等价值观，也离不开安全科学、安全工程及人体工程学等专门领域的深入探索。

第三，转化过程并非全部局限于设计流程之内，有时也会发生在特定设计过程之外。多数情况下，工程师往往会依据更广泛适用的规范开展工作，这些规范主要源自立法、技术规范和标准，它们通常由标准化委员会的工程师起草，并为处理安全性、兼容性等价值考量制定要求或指导方针。

第四，价值观向设计要求的转化是一个充满价值判断的过程，可以通过不同方式完成。不同学科为明确价值观提供了不同视角和方法，而规范地明确价值观取决于当前技术的可行性或对诸多相关价值的权衡。例如，欧洲经济共同体 88/116/EEC 指令只涉及了四条一般化行为规范中的一条，其原因在于，从

经济角度来看，制定严格要求以禁止常用的蓄电池笼是不可取的。从哲学角度来看，明确价值观需要关注的主要问题是特定规范在何时是适当的，或者至少是可行的。

第五，将价值观转化为设计要求依赖于具体背景。尽管动物福利是大家普遍认可的价值观，但在不同应用背景下其设计要求还存在很大不同。例如，在设计养殖系统时，其具体要求与毒性测试或医学实验等情境下的要求是不同的，欧盟理事会 1999/74/EC 指令规定了三种不同规范以适用于三种养鸡系统的设计要求。

第六，这个例子说明价值观和设计要求具有层级结构。在这种情况下，动物福利这一普遍价值首先由动物行为学家转化为一系列养殖规范，然后由政府将这些规范转化为具体要求。

总之，价值观是设计各种技术时不可或缺的组成部分。包含或排除哪种价值观，以及程度如何，并不总是清晰和明显的。面对价值选择的模糊地带，我们不应回避或忽视价值冲突，而应该采取策略性、深思熟虑的方法加以解决。首先，通过技术创新与制度创新，主动应对或规避价值冲突。其次，通过设定必须满足的最低阈值，并优化设计以达到最佳平衡，从而权衡相互冲突的价值观。具体可以采用“满足性”策略——即“满足”与“足够”相结合。这意味着两个价值都需要达到所需阈值来确保处于足够的水平。比如，在设计汽车时，工程师必须遵守一系列关于安全性、可持续性的监管限制和标准。除了这些阈值之外，还需要研究每个价值观在最低水平上能够得到满足的程度。换句话说，要解决安全性与可持续性之间的冲突，需要首先为每个价值设定最低可接受（或足够）的水平，然后在最低水平上优化平衡每个价值。

案例 1：自动武器

什么是自动武器？这实际上是一个颇为复杂且颇具争议的问题。一部分原因在于机器自主性概念非常复杂，另一部分原因在于它本身就备受争议。首先，我们尝试以最简单的方式定义它：自动武器，即能够在无人直接操作的情况下执行特定任务的装备。一旦经过预先编程，它便无需操作员的进一步输入或指导。

然而，这样的定义可能略显笼统，因为我们早已有了能在无人操作下执行任

务的自动化武器，比如导弹防御系统已经实现了自动化。在拦截敌对导弹时，速度至关重要，而这样的系统一旦被编程，系统就可以拦截导弹，无需操作员做任何事情。

在关于新武器系统的讨论和相关学术文献中，一个核心议题是自主性与自动化是否应当被视作两个独立的概念。尽管有人认同这一观点，但并未形成广泛共识。

在讨论中，我们经常听到的一个论点是：自动武器可以自行做出选择目标的决定。但这里存在一个关键问题：当机器自主决策时，它是否真的能够理解在特定情境下什么样的决策才是有意义的？这正是我们需要通过政策研究与深入探讨来解决的问题。

如果人们真的允许机器进行决策，那么理论上，它们将能够基于智能标准自行行动。我们甚至可以对它们进行部署，让它们在特定环境下适用规范武装力量所使用的标准。然而，现实情况是，这仍是一个遥远的科幻愿景。

我们更倾向于将自主性视为一种高级的自动化形式。我们设想的机器并不能像人类那样运用智能标准做出决策。尽管如此，它们仍然与自动化系统有所不同。例如，巡航导弹就是一个自动化系统，我们可以将 GPS 坐标编程进去，让它们自行搜索并击中目标。

然而，对于自主系统，我们可以期待它们执行更为复杂的任务。例如，一个自主系统可以部署在非常复杂且充满挑战的环境中，在这种环境中，它可以自主导航定位目标，而无需操作员提供 GPS 坐标。

从某种程度上说，自动武器是自动化武器发展连续体中的一环，是军事自动化水平不断提升的下一个阶段。作为决策者，当然也作为公民，我们自然关心这些系统带来的风险和优势。

一方面，它们在军事上可能为我们带来微弱的优势。例如，通过自主系统，我们可以在更远的距离上以更快的速度进行战斗。想象一下，一架由操作员远程控制隐形无人机，为了方便进行远程控制，操作员与无人机之间需要建立通信连接。但是问题在于，一旦无人机飞入敌人领土，操作员和无人机之间的联系可能会被敌方追踪，从而暴露行动。如果有了高度自动化的隐形无人机，我们就可以预先编程，让它自行飞入敌方领土，追踪并攻击某些目标，操作员除了最初进

行预编程外，无需再做任何操作。这样的隐形任务对军方来说无疑极具吸引力。

另一方面，使用这类系统还可能存在道德和法律优势。一些机器人学家认为，通过增加机器的自主性，我们或许可以防止战争罪行的发生。如果这些作用是真的，那么这些系统无疑是可取的。谁不希望减少战争罪行呢？尽管如此，这一观点仍然停留在假设阶段。

因此，那些坚信这些系统在战场上能发挥巨大作用的人，有责任提供充分的证据。越来越多的文献开始强调这些系统可能带来的风险。例如，一个显著的风险是这些机器可能无法准确识别被指定的目标。我们谈论的是那些能够在复杂战场和极端环境中运作的机器，人们担心它们可能难以找到并攻击正确的目标。这无疑是一个巨大的风险。此外，还有技术风险，例如我们的系统可能被敌人入侵并重新编程，然后重新部署，从而犯下战争罪行或攻击我们自己的部队。

综上所述，自动武器确实在某些方面具有优势，特别是在军事层面。同时，它们也可能带来道德上的益处，但同样伴随着显著的风险。在使用这些武器时，我们必须认真对待人们的担忧，并确保它们的安全与可靠性。

在探讨机器人武器的使用问题时，人们最难以接受的点是什么呢？不难发现，许多关于机器人武器的讨论都聚焦于其潜在风险。对于这个问题，我们可以从两个角度进行思考。

首先，人们往往对新技术抱有恐惧。然而，机器人武器或自动武器在某种程度上并非全新概念。导弹防御系统的自动化就是一个被广泛接受的先例，人们对此并无过多担忧。那么，问题究竟出在哪里？是否因为人们对特定技术持有负面看法，从而放大了这些问题？或者说，这些技术虽然有所创新，但本质上仍是在已有基础上的发展？显然，人们对技术的感知在其中起到了关键作用。

除了感知问题，我们还需正视人们对这些武器的担忧。例如，人们可能担忧其可靠性，以及这些武器是否能以安全的方式部署。在复杂的战场上，这些武器能否准确找到目标？这是整个讨论中的核心问题。这些系统的可靠性如何？它们的部署是否在可接受的风险范围内？或者，它们是否会给无辜者带来过大的风险？显然，人们有这样的担忧是有充分理由的。

在开发武器和新武器系统方面，早已有明确的指导方针。这些方针不仅适用于自动武器，也适用于其他类型的武器。当然，我们还需要牢记这些武器的部署

指导方针。这意味着在开发这类武器系统时，我们并非在法律或道德的空白地带行事。

然而，问题在于这些指导方针是否会被遵守。人们担心那些热衷于开发这些技术的国家可能会忽视这些指导方针。以美国在反恐行动中使用无人机为例，这一行为在法律和道德层面都颇具争议。对于这种行动和特定情况，法律如何适用并不明确。合规性无疑是引发进一步担忧的重要原因。

因此，我们需要社会的强烈回应。同时，联合国、红十字会等国际机构也应发挥积极作用，明确向各方传达一个信息：新武器系统的开发和部署必须严格遵循国际法。我们期望各国军队能够高度尊重法律。虽然许多军队已经在践行这一点，但严格的国际监督仍然是必要的支持。

在这场讨论中，透明度是一个至关重要的主题。各国和军队应在一定范围内透明地公布其正在开发的武器系统，以及这些系统的使用情况。他们应积极确保这些系统依法使用，并为此承担相应责任。

案例 2：护理机器人

我们目前面临着一场医疗保健的革命，这场革命围绕着是否应在手术室和整个护理部门引入机器人展开。护理机器人的定义尚未有普遍共识，但我们可以将其理解为一种融入护理实践中的技术设备，旨在协助医护人员扮演护理者的角色。

然而，我们何时、何地以及如何引入护理机器人，这并非简单的问题。它们的角色与定位，取决于我们自身的选择与判断！

为了判断护理活动引入机器人是否可取，除了考虑成本效益和健康效益等因素外，我们更需要了解这些活动背后所涉及的价值观。这些价值观包括但不限于人的尊严、隐私、关注力、责任、能力以及互惠关系。要识别这些价值观，关键在于理解护理活动的真正目标。只有如此，我们才能评估机器人的引入是否会推动或阻碍这些目标的实现。

让我们通过两个具体例子来深入探讨。

例子 1：患者搬运

若仅从外部目标来看，搬运患者即是从病床移至轮椅的过程，以便于进行各种活动如上厕所、就医等。从这个角度来看，搬运活动只包括以一定的角度、速

度和力量安全地将患者从床上抬起，然后安全地放在轮椅上。

然而，从实践角度来看，这一过程则显得更为复杂。在搬运过程中，患者处于脆弱和敏感的状态，他们需要信任护理者，而护理者则需扮演可信赖的角色。这种搬运不仅是简单的移动，更是护理者与患者建立治疗关系的关键时刻。这种关系对于患者未来的护理至关重要，因为它能激励患者诚实地描述症状、按时服药并遵守护理计划。因此，在这种实践导向下，护理者不仅需确保患者的安全移动，同时还需要评估其状态并满足其重要的社会和医疗需求。

面对这样的行为活动，我们不禁要问：是否应该将搬运任务完全委托给机器人？实际上，这一问题涉及了更为广泛且可能相互对立的价值观。若仅将其视为运输过程，则机器人无疑能提供更安全、高效的服务；但若将其视为建立信任与关怀的时刻，则人类的参与与专注则显得不可或缺。

例子 2：尿样收集

在儿科肿瘤学中，定期收集尿液样本是监测患者体内化疗药物毒素的重要手段。从护理伦理的角度出发，尿样收集体现了护士的专业技能与人文关怀：在整个活动中，护士始终对患者保持关注、对患者负责并胜任护理提供者的角色。此外，尿液收集过程也是患者与护理者进行短暂互动的时刻。然而，这一过程也面临着挑战与风险。样本收集有可能让患者感到尴尬，也有可能对护士的健康造成危害。由于时间紧迫，护士往往无法穿戴足够的防护服来阻挡化疗药物毒素的侵袭。

这种情况与病人搬运活动相比，护理活动不仅可以在概念上加以区分，而且还可以在实质上进行分离：我们不应简单地考虑是否用机器人完全替代护士收集尿样；相反，我们应该思考是否可以将护士从部分有害的、目标导向的活动中解放出来，同时确保护士与整个样本收集过程保持紧密联系。

这种以护理为中心的价值敏感设计方法为未来护理机器人的设计提供框架。它要求我们在设计过程中系统地考虑护理价值观的识别与融合。通过深入理解护理活动及其背后的价值观与角色分配，我们可以更明智地选择是否将某些任务委托给机器人，从而确保机器人与护理价值观的兼容并进。为此，设计建议如下：

机器人可以设计成自主前往患者房间。到达后，机器人从护士那里获取信息以确定护士是否在场。这种设计考虑确保了护士必须在现场进行尿液测试。然后，

机器人自主进入患者浴室收集样本；它从患者的马桶或垃圾桶中收集样本，然后离开浴室前往护士站，并再次确认护士是否在场。机器人传输已获得的样本信息，有了这些信息，护士可以选择将携带样本的机器人发送到肿瘤科实验室，或者让机器人完成分析并将结果发送给肿瘤科医生。无论护士做出何种决定，护士都知道样本已经收集，并将分析结果传送给肿瘤科医生。这一设计建议旨在确保有人类代理人为成功收集样本负责。

以上建议可以安全收集尿样，同时避免了对护士的伤害，同时又让护士参与样本采集并与患者保持联系，以及确保护士对样本采集过程的责任性。

讨论题：

1. 如何为自动武器制定一个价值敏感设计的层次结构？
2. 如何处理总是会出现的价值冲突？