

Presented by

翻译出品



Google Cloud

中国DevOps社区

加速度 DevOps 状态报告 2023



中国DevOps社区翻译小组

译者：冷大鲲 秦团结 任悦 周一行 王晓宁
审校：张扬

社区工作成果非 DORA 官方文档

Premiere Sponsors



Deloitte.

Qarik

v. 2023-10

目录

前言

摘要..... [03](#)

概念与衡量..... [06](#)

第一章

如何进行对比..... [10](#)

第二章

以用户为中心对组织效能的预示..... [17](#)

第三章

技术能力对效能的预示..... [20](#)

第四章

文档是基础..... [27](#)

第五章

可靠性释放性能..... [31](#)

第六章

灵活的基础设施是成功的关键..... [38](#)

第七章

文化是一切的根本..... [45](#)

第八章

如何，何时，为何，您的定位很重要..... [51](#)

后记

结语..... [57](#)

致谢..... [58](#)

作者..... [59](#)

研究方法..... [62](#)

人员和企业统计数据..... [72](#)

模型..... [81](#)

延伸阅读..... [91](#)

附录..... [92](#)

所有引文检索的日期为2023年9月27日

摘要

在过去的十年中，DORA研究团队持续探索高效能技术驱动组织的关键能力和评价标准。我们已从各个行业、各种规模的组织中的超过36,000名行业专家那里获得宝贵的反馈和见解。感谢你们的无私分享和独到的洞见。

DORA致力于探讨工作方法（即所谓的能力）与产出之间的关系：这些产出与整体组织及其员工有着切实有意的联系。本次研究采用了严谨的统计学评估方法，使得报告内容足够中立且平台无关（详见研究方法章节）。

我们希望这些洞见能让领导者和实践者们了解到他们可以在哪里产生影响。今年的研究探讨了三个关键成果，以及有助于实现这些成果的相关能力：

- **组织效能**——组织不仅要创造收益，还需要为客户以及更广泛的社区创造价值。
- **团队效能**——应用程序或服务团队创造价值、推动创新和协同合作的能力。
- **员工幸福感**——组织或团队采纳的策略应该对员工有益：能够减少他们的职业倦怠，为他们营造令人满意的工作体验，并提高他们价值产出的效率（即生产力）。

此外，该研究还涉及了那些我们常常视为目标本身的方法或效能标准：

- **软件交付效能**——团队可以安全、快速、高效地更改他们的产品技术系统。
- **运维效能**——为用户提供可靠的服务体验。



关键发现

>> 建立健康的文化

文化是塑造技术能力、激发技术效能、实现组织效能目标并帮助员工成功的基石。拥有生机型文化的团队可以提高30%的组织效能。

>> 始终以用户为中心

以用户为中心能够指导和驱动研究中涉及的所有技术、流程和文化等能力的改进。团队可能会按照自己的设想迅速且成功地部署产品，但如果忽视了用户的需求和体验，这些努力可能就会付诸东流。真正关注用户的团队能够使组织效能提高40%。

>> 加速代码评审释放软件交付效能

加速代码评审是提高软件交付效能的最有效途径之一。代码评审越快的团队，其软件交付效能有着50%的提升。

>> 借助优质文档增强技术能力

高质量的文档可以显著放大技术能力对组织效能的效果。例如，与低质量文档相比，当有高质量文档配合时，其主干开发对组织效能的促进效果高达12.8倍。

>> 利用云计算提升基础设施灵活性

云计算之所以受到推崇，是因为它能提供高度的基础设施灵活性。例如，与不使用云相比，使用公有云可以使得基础设施的灵活性提高22%。与僵化的基础设施相比，这种灵活性进一步提高了30%的组织效能。要充分利用云的价值，关键在于充分挖掘和应用云的不同特性和能力，也就是基础设施的灵活性。

>> 平衡交付速度、运维效能和用户导向

想要充分激发组织效能的最大潜力，我们不仅需要出色的软件交付效能，还需要强大的运维效能。以用户为中心可以保持这两者之间的平衡，这不仅能够实现最佳的组织结果，还能也提升员工的幸福感。

>> 公平地分配工作

那些被认为是弱势群体、女性或自述性别的人更容易感到职业倦怠，这背后可能涉及多种系统性和环境上的因素。不出所料，我们发现承担更多重复性工作的受访者更有可能经历更高程度的职业倦怠，而弱势群体成员有更大的可能性承担重复性工作。

弱势群体受访者比非弱势群体受访者的职业倦怠程度高出24%，重复工作占比多出29%。女性或自述性别人群的重复性工作比男性多出40%。

将 DORA 洞察应用到您的场景

接纳持续改进思维和实践的团队可能会获得最大的收益¹。DORA能够辅助您制定和推进自身的改进计划。

为了最大限度地利用这项研究，请结合您团队情况和用户场景来考虑。例如，我们前面提到，代码评审越快的团队，软件交付效能就会提高50%。然而，如果您的代码评审已经很快，但在体系其他地方的速度受到制约，那么您的软件交付效能也不太可能得到改善。当从业者就当前的工作完成方式展开沟通时，可以对该研究进行场景化。这些沟通可以促进共情、协作和对每个参与者动机的理解。

改进的工作没有终点。在体系中找到瓶颈，然后解决它，之后再重复这一过程。最关键的对比是观察同一应用团队随着时间推移而变现出的变化，而不是观察其他应用程序团队、组织或行业。

指标和度量

指标和仪表盘帮助团队监控他们的进度并修正方向。

从业者和领导者都在为组织效能、团队效能和员工幸福感而努力。但是度量不是目标，就像交付软件不是目标一样。

过于沉迷于效能指标可能会导致很多无效或不良的行为，而在能力和学习上的投入是走向成功的更加途径。学习得最多的团队，其进步也越明显。

您不必独自改进

我们可以从彼此的经验中互相学习，而 DORA 社区网站 (<https://dora.community>) 就是一个分享和学习改进计划的绝佳论坛。

¹ 2022 加速度 DevOps 状态报告

<https://dora.dev/research/2022/dora-report/2022-dora-accelerate-state-of-devops-report.pdf#page=7>



概念与衡量

本节描述了DORA试图去衡量的一些核心概念，它们是本报告和我们分析模型的基本元素。这些部分构成了制作这份报告的基础成分¹。因此，对于我们作者来说，明确这些概念的含义并在提及它们时保持一致性理解是至关重要的。下表旨在为我们所有人，无论是读者还是作者，提供一个统一的术语参考。

鉴于报告中的许多概念都含有多个方面，我们通常会采用多个指标来刻画它们。为了评估这些概念的度量成效，探索性因子分析和验证性因子分析是我们采用的一种方法。有关该过程的详细信息，您可以在研究方法论章节中找到。

在对度量方法进行评估后，我们将分数设定为从0到10的范围，其中0代表某个概念完全没有体现，而10则代表某个概念最大程度的体现。我们相信，这样的标准化处理将有助于我们在讨论这些概念如何发挥作用时保持一致，并有助于与历年的数据进行比较。

我们讨论的每个概念都包含以下信息：

- 小图标有助于传达其含义，并在希望作为参考时更容易查找。
- 该概念在样本中的平均得分（平均值，Mean）。
- 四分位距（IQR）的边界。给出位于中间50%的数据所在的两个数字（第25和第75百分位数），这些边界应该有助于揭示调查反馈的分布。
- 一组数据的中间数值（中位数，Median）。如果它与平均值差异很大，则表明数据可能存在偏差。
- 概念的描述以及我们如何度量它。

¹ 分析中使用的调查问题发布地址 <https://dora.dev>

关键成果

关键成果是我们认为个人、团队或组织正在努力追求的目标（如组织效能）或尽力避免的结果（如工作倦怠）。因此，我们将这些指标视为评估个人、团队和组织表现的重要方式。

 组织效能 <table><tr><td>平均数</td><td>四分位距</td><td>中位数</td></tr><tr><td>6.3</td><td>5-8</td><td>6.3</td></tr></table> 高效能组织的主要产品或服务拥有更多的客户、更高的利润和更大的相对市场份额。	平均数	四分位距	中位数	6.3	5-8	6.3	 团队效能 <table><tr><td>平均数</td><td>四分位距</td><td>中位数</td></tr><tr><td>7.6</td><td>6.6-9</td><td>8</td></tr></table> 高效能团队能够适应变化、相互信赖、工作高效，并且擅于创新和通力合作。	平均数	四分位距	中位数	7.6	6.6-9	8	 软件交付效能 <table><tr><td>平均数</td><td>四分位距</td><td>中位数</td></tr><tr><td>6.3</td><td>5.1-7.8</td><td>6.4</td></tr></table> 以下四个指标度量软件交付的速度和稳定性。 • 部署频率 • 变更前置时间 • 变更故障率 • 部署失败恢复时间	平均数	四分位距	中位数	6.3	5.1-7.8	6.4
平均数	四分位距	中位数																		
6.3	5-8	6.3																		
平均数	四分位距	中位数																		
7.6	6.6-9	8																		
平均数	四分位距	中位数																		
6.3	5.1-7.8	6.4																		
 运维效能 <table><tr><td>平均数</td><td>四分位距</td><td>中位数</td></tr><tr><td>6.2</td><td>5-7.5</td><td>6.3</td></tr></table> 服务满足用户期望的程度，包括可用性和性能等指标。	平均数	四分位距	中位数	6.2	5-7.5	6.3	 可靠性目标 <table><tr><td>平均数</td><td>四分位距</td><td>中位数</td></tr><tr><td>7</td><td>5-7.5</td><td>7.5</td></tr></table> 服务满足其既定目标（如可用性、性能和正确性）的程度。	平均数	四分位距	中位数	7	5-7.5	7.5							
平均数	四分位距	中位数																		
6.2	5-7.5	6.3																		
平均数	四分位距	中位数																		
7	5-7.5	7.5																		

员工幸福感是倦怠感、生产力和满意度的结合体。

 倦怠感 <table><tr><td>平均数</td><td>四分位距</td><td>中位数</td></tr><tr><td>4.1</td><td>2-6</td><td>4</td></tr></table> 倦怠不仅是心理和生理的消耗，还涉及人们对工作价值和意义的评价。工作倦怠会导致愤世嫉俗的心态。²	平均数	四分位距	中位数	4.1	2-6	4	 生产力 <table><tr><td>平均数</td><td>四分位距</td><td>中位数</td></tr><tr><td>7.5</td><td>6.7-8.8</td><td>7.9</td></tr></table> 高产的个体所从事的工作与他们的技能相匹配，这既为组织创造了价值，也能让他们的工作更高效。	平均数	四分位距	中位数	7.5	6.7-8.8	7.9	 工作满意度 <table><tr><td>平均数</td><td>四分位距</td><td>中位数</td></tr><tr><td>6.08</td><td>5.7-7.1</td><td>7.1</td></tr></table> 这是一个单项问题，要求受访者综合考虑所有的因素，从而对他们的工作的整体感受进行评分。³	平均数	四分位距	中位数	6.08	5.7-7.1	7.1
平均数	四分位距	中位数																		
4.1	2-6	4																		
平均数	四分位距	中位数																		
7.5	6.7-8.8	7.9																		
平均数	四分位距	中位数																		
6.08	5.7-7.1	7.1																		

² Maslach C, Leiter MP. 理解倦怠体验：最新研究及其对精神病学的影响。世界精神病学杂志2016年6月；15(2):103-11. doi :10.1002 / wps.20311. PMID: 27265691;PMCID: PMC4911781.

³ Warr, P., Cook, J., & Wall, T. "工作态度和心理健康的衡量天秤。" 职业心理学报, 52(2), 1979. 129 – 148. <https://doi.org/10.1111/j.2044-8325.1979.tb00448.x>

流程与技术能力

这些代表了团队或组织中可能采取的行动、实践或所处的状态。简而言之，它们描述了团队的工作内容和行为模式。

 人工智能贡献	 松耦合架构	 持续集成
平均数 3.3	四分位距 0.3-6.3	中位数 2.4
平均数 6.4	四分位距 4.7-8.3	中位数 6.7
平均数 6.9	四分位距 5-8.9	中位数 7.8
人工智能对各种技术任务贡献的重要程度。	可独立开发、测试和部署的软件。	自动构建和测试软件变更的实践。
 持续交付	 代码评审速度	 文档编制
平均数 7.0	四分位距 5.7-8.7	中位数 7.3
平均数 6.5	四分位距 6-8	中位数 6
平均数 5.8	四分位距 3.8-7.9	中位数 6.25
安全、快速和持续地将所有类型的变更（包括新特性、配置更改、缺陷修复和实验探索）发布到生产环境或交付至用户的能力。 ⁴	评估从拉取请求（PR）到代码变更评审完成所花费的时间。	组织中人员在日常工作中所创建和使用书面内容的质量。
 可靠性实践	 主干开发	 基础设施灵活性
平均数 5.9	四分位距 3.9-8.3	中位数 6.1
平均数 5.6	四分位距 3.9-7.8	中位数 5.6
平均数 6.6	四分位距 5-8.3	中位数 7.3
团队用来提高服务运维效能的活动和实践。	进行小批且频繁的变更，并定期合并到版本控制系统中代码主干分支的实践。	可伸缩的基础架构，如弹性、易访问和可度量。 ⁵

⁴ “什么是持续交付”<https://continuousdelivery.com/>

⁵ 美国国家标准与技术研究所(2018)，NIST对云计算的定义。<https://csrc.nist.gov/pubs/sp/800/145/final>

文化方面

定义文化并不容易，一些人可能会描述它为工作空间的普遍规范（如灵活性）、主流取向（如用户导向）以及环境氛围（如组织稳定性）。

工作分配 <table><tr><th>平均数</th><th>四分位距</th><th>中位数</th></tr><tr><td>5.8</td><td>3.8-7.9</td><td>5.8</td></tr></table> 帮助员工在团队中公平分配任务的标准流程。	平均数	四分位距	中位数	5.8	3.8-7.9	5.8	灵活性 <table><tr><th>平均数</th><th>四分位距</th><th>中位数</th></tr><tr><td>7.7</td><td>6.6-8.9</td><td>8.3</td></tr></table> 一个人该如何、在何时、在何地处理任务。⁶	平均数	四分位距	中位数	7.7	6.6-8.9	8.3	工作安全感 <table><tr><th>平均数</th><th>四分位距</th><th>中位数</th></tr><tr><td>5.9</td><td>3.3-8.3</td><td>6.7</td></tr></table> 这是一项单项指标，询问人们多久一次担心自己的工作保障。得分越高，忧虑越少。	平均数	四分位距	中位数	5.9	3.3-8.3	6.7
平均数	四分位距	中位数																		
5.8	3.8-7.9	5.8																		
平均数	四分位距	中位数																		
7.7	6.6-8.9	8.3																		
平均数	四分位距	中位数																		
5.9	3.3-8.3	6.7																		
组织稳定性 <table><tr><th>平均数</th><th>四分位距</th><th>中位数</th></tr><tr><td>7.2</td><td>6.7-8.3</td><td>8.3</td></tr></table> 衡量员工工作环境稳定与否的一个单项指标。	平均数	四分位距	中位数	7.2	6.7-8.3	8.3	知识共享 <table><tr><th>平均数</th><th>四分位距</th><th>中位数</th></tr><tr><td>3.7</td><td>1.7-5</td><td>7.9</td></tr></table> 想法和信息如何在组织内部传播。团队成员只需回答一次问题，这些信息就会被共享给其他人，人们不需要再次询问和等待答案。⁷	平均数	四分位距	中位数	3.7	1.7-5	7.9	以用户为中心 <table><tr><th>平均数</th><th>四分位距</th><th>中位数</th></tr><tr><td>7.8</td><td>5.6-8.3</td><td>7.8</td></tr></table> 理解并结合用户的需求和目标，使得产品和服务变得更好。⁸	平均数	四分位距	中位数	7.8	5.6-8.3	7.8
平均数	四分位距	中位数																		
7.2	6.7-8.3	8.3																		
平均数	四分位距	中位数																		
3.7	1.7-5	7.9																		
平均数	四分位距	中位数																		
7.8	5.6-8.3	7.8																		
Westrum 组织文化 <table><tr><th>平均数</th><th>四分位距</th><th>中位数</th></tr><tr><td>7.3</td><td>6.1-8.6</td><td>7.8</td></tr></table> 一个组织如何应对问题和机遇。文化的类型有三种：生成型、官僚型和病态型。⁹	平均数	四分位距	中位数	7.3	6.1-8.6	7.8	⁶ Shifrin, Nicole V., 和 Jesse S. Michel. 灵活工作安排与员工健康: 元分析回顾。Work & Stress 36,no.1, 2022.60 – 85 ⁷ “2022开发者调查” https://survey.stackoverflow.co/2022#overview ⁸ Kersten, Mik. 项目到产品: 如何用流框架在数字颠覆时代生存和发展(IT Revolution, 2018), 54.https://itrevolution.com/product/project-to-product/ ⁹ Westrum R. 《组织文化的类型》。BMJ Quality & Safety 2004. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1765804/													
平均数	四分位距	中位数																		
7.3	6.1-8.6	7.8																		

如何进行对比？

古德哈特定律:

一项指标一旦成为了目标，它将不再是个好的指标¹。

要点

提高效能的第一步是为应用程序团队当前的软件交付效能、运维效能和用户导向设置一个基线。这些度量指标可以帮助团队评估他们的工作，并为事情随着时间的推移如何变化提供良好的信号。

然而这些度量指标本身并不是团队的改进对象。确立了这个基线后，我们就可以进一步评估某个团队在人员、流程和技术方面的实际水平，进而确定哪些因素可能会成为瓶颈²。接下来，团队应给予充足的时间和空间来调整、试验和复盘。持续重复地进行这样的循环，这将促使团队养成持续改进的思维和实践。

¹ Strathern, Marilyn (1997). "提高评级":英国大学的审计系统". 欧洲版. John Wiley & Sons. 5 (3): 305–321. doi:10.1002/(SICI)1234-981X(199707)5:3<305::AID-EURO184>3.0.CO;2-4. S2CID 145644958.

² 该报告和“能力目录”(<https://dora.dev/devops-capabilities/>) 中列出的资源可以提供帮助。

³ Forsgren, N., Storey, M-A., et. al. 《开发者生产力的空间:它比您想象的要多》2021年 <https://queue.acm.org/detail.cfm?id=3454124>

在进行比较时，请注意以下已提及和潜在的陷阱:

- **无意义的比较。** 仅仅基于聚类进行应用程序团队的对比用处可能不大。这样做可能会导致每个应用程序团队特定的背景情况丢失，进而影响对其改善的目标。
- **将指标作为目标。** 忽视古德哈特定律，盲目追求指标。例如，盲目要求“每个应用程序团队必须在年底前展示出其处于‘精英’效能”可能会导致团队为了达标进而操纵数据。
- **依赖单一指标。** 依赖“单个核心指标”来评估复杂的系统。应结合多种指标来获得更深入的理解和洞察³。
- **指标范围过窄。** 往往倾向于度量最容易获得的指标，而不是最有意义的指标。
- **将行业规范作为阻碍进步的借口。** 在受到严格监管的行业中，一些团队可能会利用法规作为维持现状的理由。

想要了解更多关于我们的发现，以及如何推动持续改进的思维和实践的建议，请参阅文章“如何转变”：

dora.dev/devops-capabilities/cultural/devops-culture-transform/

介绍

每年我们都会进行一次或多次聚类分析，目的是探寻应用程序之间的共同趋势。

我们建议您使用这些分析来了解该如何进行对比，但不要只关注这些对比。最有意义的对比是同一应用程序团队随时间表现的对比，而不是不同的应用程序团队之间的对比，因为不同的应用程序团队自然有其不同的背景和环境。

团队为用户开发软件，而用户是软件实用性和可靠性的最终评判标准。那些专注于满足用户需求的团队更能够构建正确的产品。将以用户为中心（用户导向）、软件交付效能和运维效能结合起来，意味着这些团队也有能力正确的构建产品。

专注用户需求的团队才能做正确的事，并且正确地做事。

结果

软件交付效能

我们使用以下方法评估软件交付的效能：

- **变更前置时间**——变更从提交到部署所需的时间。
- **部署频率**——将变更推送到生产环境的频率。
- **变更失败率**——部署后导致需要立即干预的故障的频率。
- **部署失败的恢复时间**——从失败的部署中恢复所需的时间。

对于这四项指标的改进，一种普遍的策略是减少应用程序变更的批次大小。较小的变更更容易理解，也更容易在交付过程中流转⁴。一旦发生故障，较小的变更也更容易恢复。

团队应该确保每次变更尽可能的小，以使得交付过程迅速且稳定。这种工作方式有助于提升变更的速度，并保持其稳定性。

⁴ 通常，一个特性可以被拆分成成许多可独立交付的变更。我们通过评估应用程序或服务的变更来评估软件交付效能。

今年，我们改进了软件交付效能的度量。阅读附录中的“度量软件交付效能方式的优化”内容以了解更多关于这些变化的信息。

以下是今年调查受众的软件交付效能全景：

效能级别	部署频率	变更前置时间	变更失败率	部署失败恢复时间	受访者百分比
卓越	按需	小于一天	5%	小于一天	18%
高	一天一次到一周一次	一天到一周	10%	小于一天	31%
中	一周一次到一月一次	一周到一月	15%	一天到一周	33%
低	一周一次到一月一次	一周到一月	64%	一月到六个月	17%



运维效能

我们评估运维效能时，会询问受访者他们的服务遭遇以下情况的频率：

- 收到终端用户对系统可靠性不满意的上报。
- 服务出现不可用、性能不达标或执行错误的情况。

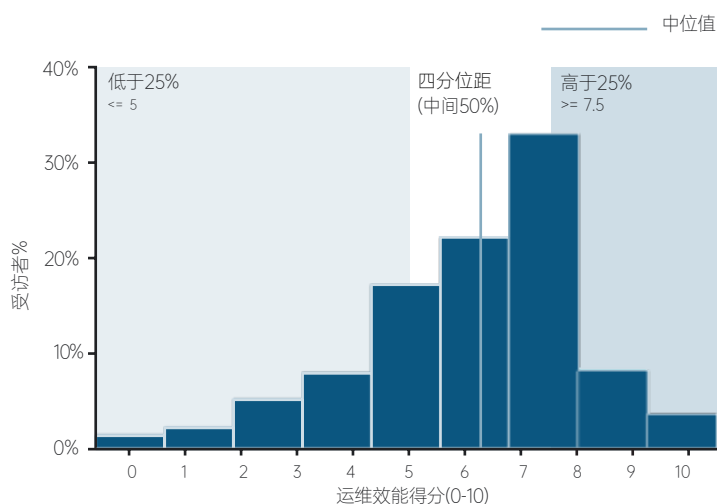
想要深入了解运维效能如何影响组织效能，请参见第5章——可靠性释放效能。

以用户为中心

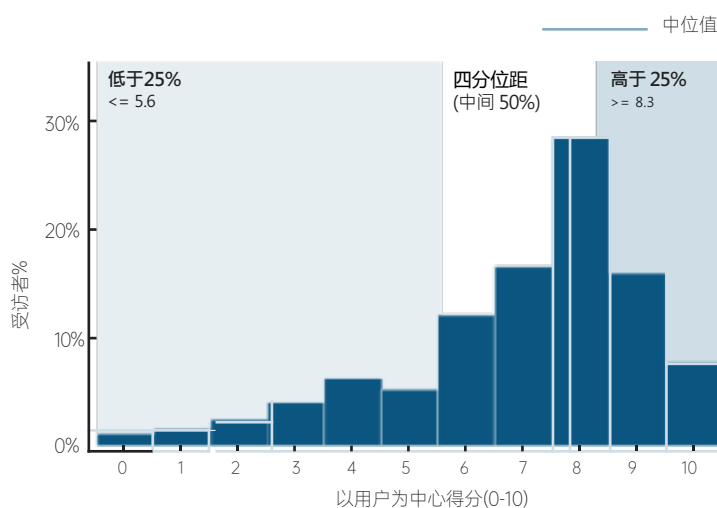
以用户为中心的应用程序或服务是围绕最终用户至上的原则来构建的。打造这样的产品需要对用户需求有深刻的认知，并将其纳入到产品演进路线图中。为了评估受访者用户导向的程度，我们通过询问以下内容在多大程度上为真来进行：

- 团队对用户想要实现的目标有着清晰和明确的理解。
- 团队的成功是根据团队为组织和用户提供的价值来评估的。
- 团队会不断地审查规格说明书（例如需求规划），并根据用户反馈进行优先级调整。

以下是今年调查对象对运维效能的看法视图：



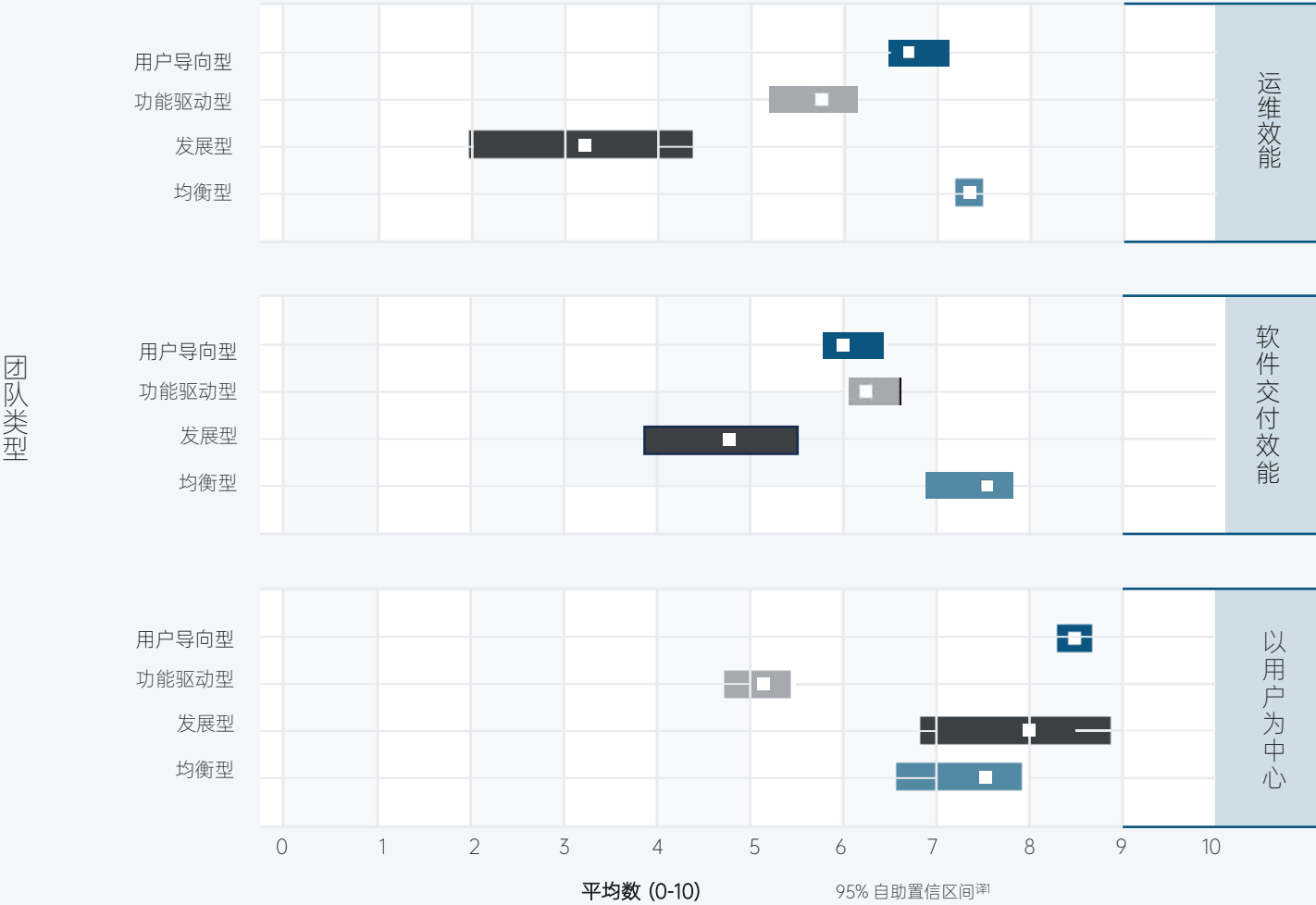
以下是今年调查对象对以用户为中心的看法视图：



聚类团队类型

将软件交付效能、运维效能和以用户为中心作为一个单元进行比较，我们可以看到四种不同类型的团队。像所有的度量标准一样，这些团队的类型是基于应用程序或服务的层面来定义的。

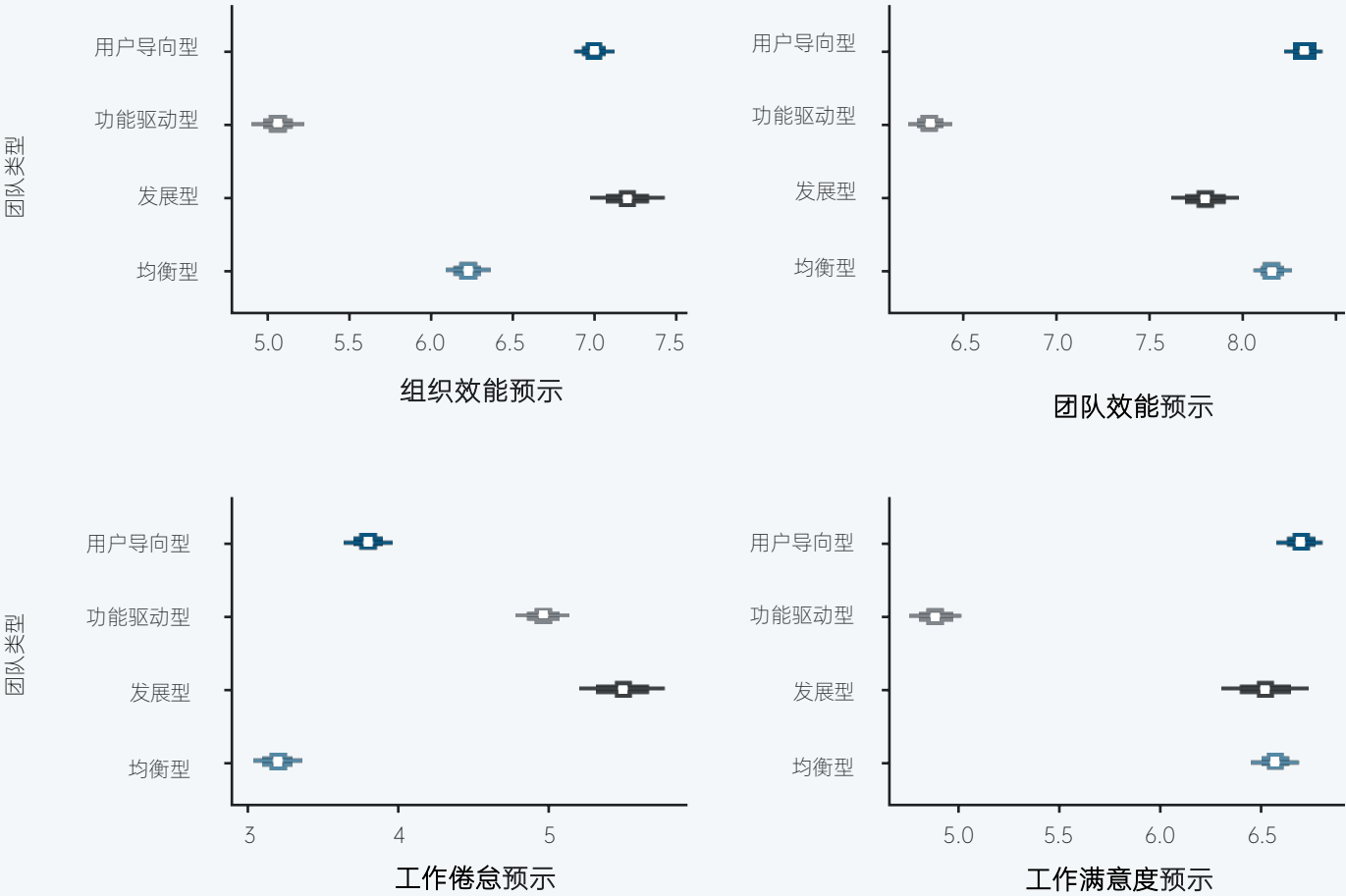
我们将团队类型命名为用户导向型、功能驱动型、发展型和均衡型。



译者注1：置信区间是统计学家用来表述样本结果离整个母群体真实数值之间的差距，这常被表述为95%置信区间。形式上，数值为95%置信区间意味着如果在同样情况下重复样本分析（这会生成不同的数据集），95%的区间会得出符合（母群体）情况的实际结果。

将我们一直在讨论的效能度量指标想象成一个刻度盘，组织可以用其调整组织效能、团队效能和个体幸福感。

下图表呈现了每种团队类型预示的效能结果。



*点表示团队类型平均值的点估计。66%的模拟落在厚区间。窄小的区间89%的模拟都失败了

每种团队类型都有其独有的特性，而且在我们的受访者中占有相当的比例，他们也有各自不同的效能结果。您的团队很可能并不完全符合其中的某一个类型，当然我们也不期望您的团队类型会始终保持不变。



这些结果意味着什么？



用户导向型

这类团队最注重用户需求。结合其出色的软件交付效能和稳健的运维效能，它预示了组织效能的最高水平。但与均衡型团队相比，这类团队确实有稍高的倦怠现象。对于这些团队来说，提高软件交付效能和/或运维效能可能是减少倦怠的最佳途径。



功能驱动型

这类团队的首要任务是交付功能特性。对交付的过度关注可能使团队忽视了用户需求的满足，这一点在以用户为中心和运维效能的较低指标中已得到体现。这种团队类型呈现出最高的工作倦怠和最低的工作满意度、团队效能和组织效能。员工重视的是交付实际的价值，而非仅仅是功能特性。功能驱动型团队可能会从用户需求的反馈中受益，因为这是一种了解从已发布的功能中获取更多价值的方式。



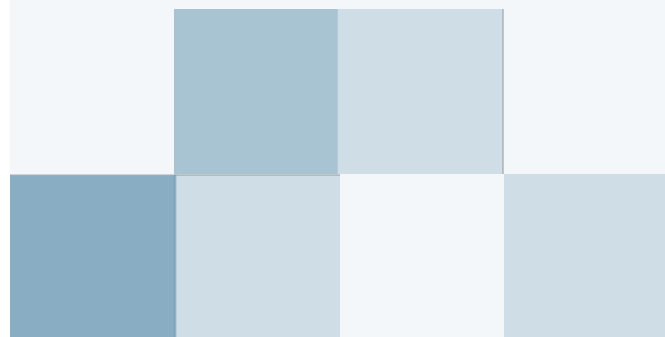
发展型

这类团队通过专注应用程序用户的需求来获得良好的组织效能。然而，他们可能仍在寻找产品与市场的匹配度或者还在补强技术能力。这样的团队在较小的组织中更为常见。他们表现出较低的软件交付效能和运维效能，并且与均衡型或用户导向型的团队相比，在这些应用程序团队的从业者所上报的倦怠程度更高。这类团队可能存在繁琐的流程或繁重的任务，若能够将其自动化，则可能提高其软件交付效能和运维效能。



均衡型

这类团队呈现了一种平衡且可持续的方法。他们以可持续的方式应用技术，从而达到良好的组织效能、团队效能和员工满意度。此外，这类团队的工作倦怠程度也最低。他们已经以一种好的方式调整了自身的策略的能力，以确保在所有三个指标上都有出色表现。提高以用户为中心的思维也许是他们进一步优化组织效能的途径。



以用户为中心 对组织效能的预示

要点

组织在寻求提升开发者生产力和组织效能的过程中，有时会过于沉迷最新的技术和管理趋势。但我们的研究表明，采取以用户为中心的方法来构建应用程序和服务是整体组织效能最有力的预示因素之一。为了提高效能，深入了解您的用户，并反复调整和接纳他们的反馈。

介绍

DevOps运动的初衷是为了鼓励和促进开发团队和运维团队之间更紧密的合作，从而更好地为用户创造价值。这种合作推动了早期的成功，并使得DevOps的理念和能力扩展到这两个部门之外。如今，高效能的技术驱动型组织认识到，为了实现组织目标，所有团队的对齐和协作是关键。

我们调查了以用户为中心的三个关键特征：

- 团队对其用户需求的理解程度有多深入。
- 团队在满足用户需求方面的对齐程度如何。
- 如何通过采纳用户反馈来确定工作优先级。

以用户为中心的团队
组织效能提升

40%

结果

我们发现，采用以用户为中心的软件开发方法可以显著提升效能。当组织把用户置于首位时，他们会获得一系列的好处。用户反馈不仅帮助团队确定项目的优先级，还能指导他们创造真正满足用户需求的产品和服务。这种方法进一步带来了更好的用户体验、更高的用户满意度和更多的收益。

以用户为中心的影响		
组织效能	↑↑	大幅提升
团队效能	↑↑	大幅提升
软件交付效能	↑	小幅提升
运维效能	↑↑	大幅提升
主干开发	↑↑	大幅提升
可靠性实践	↑↑	大幅提升
持续集成	↑↑	大幅提升
持续交付	↑↑	大幅提升
松耦合架构	↑↑	大幅提升
职业倦怠	↓	小幅下降*
工作满意度	↑↑	大幅提升
生产力	↑↑	大幅提升

*减少职业倦怠是一件好事！

这些结果意味着什么？

关注用户需求是预示整体组织效能的强有力因素。想要实现深度的用户关注，需要适当的激励手段、团队拉通以及有效的工作方式。用户的需求和反馈可以为我们在研究中探索的所有技术、流程和文化能力带来深刻的启示并驱动持续的改进。

以下是这些结果对组织中各类团队影响的探讨：

> 产品开发和交付团队

重视用户需求是产品开发和交付团队面向用户构建正确产品的关键，可能的话，还希望能以可持续的方式构建它们。均衡型团队就是这样做的。他们始终围绕用户需求，同时在交付、运维和组织效能方面都有出色的表现。这类团队的成员能够清晰准确地理解客户需求，根据用户反馈灵活调整计划，并最终从该能力中受益。

研究结果显示，功能驱动型的团队未能达到顶级的组织效能。这类团队似乎过度优先考虑交付效能，以致于损害了组织效能和员工幸福感。

> 运维团队

关注运维效能的团队可能会投入大量精力去优化系统指标，如CPU使用率。但是，如果没有深入理解用户对服务的真正期望，那么仍可能会经常收到性能缓慢的反馈。采用网站可靠性工程（SRE）实践，例如确定用户真正关心的服务水平指标和设定目标以确保大多数用户的满意度，有助于运维团队更好地将用户为中心的理念融入到工作中。



平台工程团队

平台工程团队在构建平台时，可能会抱有“建好了自然就会有人来用”的思维。但更为成功的方法做法是将开发者视为平台的真正用户来对待。这种思维转变需要平台工程团队深入了解开发者当前的工作方式，从而有效地找出并解决潜在的问题并消除磨擦和障碍。团队可以将软件交付和运维效能指标作为信号，来检验其在平台上的付出是否真正助力其他团队获得更佳的效果。



领导者

领导者可以通过制定激励机制来鼓励团队为用户创造真正的价值，从而营造一个以用户为中心的工作环境以达到蓬勃的发展。缺少这些机制，团队可能会认为他们的工作只是在交付更多的功能数量或减少服务中断的次数。DORA已经调研了变革型领导力的重要性¹，并为那些希望在这方面提升的领导者提供宝贵建议。

想要了解更多的信息，请访问“DevOps能力：变革领导力”

<https://dora.dev/devops-capabilities/cultural/transformational-leadership>

以用户为中心带来了

20%

的工作满意度提升

入门资源

发展以用户为中心的能力是成功的重要驱动因素。2023年报告的研究结果进一步印证了我们在2018年的发现²，在那次研究中，我们看到精益产品管理能力能够预示软件交付效能和整体的组织效能。想要提升团队的效能，可以尝试采纳以下的用户为中心的策略，如积极获取客户反馈³，确保价值流中工作的可见性⁴，小批量工作方式⁵以及鼓励团队实验⁶。

¹ 2017 DevOps状态报告: <https://dora.dev/publications/pdf/state-of-devops-2017.pdf>, 12-19

² 2018 加速度 DevOps 状态报告：新经济战略
<https://dora.dev/publications/pdf/state-of-devops-2018.pdf>, 49-51

³“客户反馈” <https://dora.dev/devops-capabilities/process/customer-feedback/>

⁴“可视化价值流”<https://dora.dev/devops-capabilities/process/work-visibility-in-value-stream/>

⁵“小批量工作方式”<https://dora.dev/devops-capabilities/process/working-in-small-batches/>

⁶“团队体验” <https://dora.dev/devops-capabilities/process/team-experimentation/>

技术能力对效能的预示

要点

将资源和精力投入到持续集成、松耦合架构以及快速代码评审的过程中，将可能带来众多有益的结果。例如会带来组织效能、团队效能、软件交付效能和运维效能的显著提升。而且，对于从事应用开发或服务运维的团队成员而言，这样的改进不仅不会带来额外的负担和压力，反而有助于提升他们的工作幸福感。

介绍

在摘要部分，我们详细描述了所研究的技术能力，以及这些能力如何影响各种不同的效能和员工幸福感。DORA由始至终的一个核心理念就是探究和量化各种不同的流程与技术能力如何预示效能。

在今年，我们研究了以下技术能力如何在预示效能：

- 人工智能
- 主干开发
- 松耦合架构
- 持续集成
- 快速代码评审

我们研究了它们如何预示以下的效能指标：

- 团队效能
- 组织效能
- 软件交付效能
- 运维效能

此外，我们还测试了这些能力与一些指标之间的关联，以确定它们如何影响从事这项工作的人：

- 职业倦怠
- 生产力
- 工作满意度

结果

我们研究的技术能力和流程对关键效能指标有着各种各样的影响，但总体上而言是积极的。

技术能力和流程	对团队效能的影响	对组织效能的影响	对软件交付效能的影响	对运维效能的影响
人工智能	暂无影响	小幅提升	小幅下降	大幅下降
持续集成	小幅提升	小幅提升	小幅提升	暂无影响
快速代码评审	小幅提升	暂无影响	大幅提升	大幅提升
松耦合架构	大幅提升	大幅提升	小幅提升	大幅提升
主干开发	小幅提升	小幅提升	小幅提升	小幅下降

松耦合的团队或采用松耦合架构的团队，能够独立地对其系统进行重大更改，不会受到其他团队的影响，从而能够更快地采取行动。当团队中有领域专家时，他们能够更快地评审代码，因为他们更理解变更带来的影响。这种松耦合设计使得团队在测试、构建和部署时独立工作，从而避免依赖其他团队成为潜在的瓶颈。

即便在松耦合架构中我们变更的影响较小，但仍然需要确保我们没有引入与团队中的其他开发人员的冲突。采用小批量工作的团队能够有效减少冲突，确保在每次代码提交后软件都会被构建，且触发自动化测试为开发人员提供快速反馈。

代码评审时间较短的团队，其软件交付效能提高了50%。高效的代码评审过程不仅优化了代码，还促进了知识分享、共享代码所有权、团队所有权以及工作透明度。

代码评审是您的瓶颈吗？通过评估代码评审流程及其对变更前置时间的影响，您可以获得对改进契机的洞察。

请思考以下问题：

- 同行代码审查是否已内嵌至您的流程中？
- 编码完成到评审之间的持续时间有多长？
- 您的代码评审的平均批次大小是多少？
- 有多少团队参与您的评审？
- 有多少来自不同地域的参与您的评审？
- 您的团队会通过代码评审建议来提升代码质量自动化么？

已经证明，编码完成和评审之间的较长持续时间会对开发者的效率和交付软件的质量产生负面影响。涉及多个团队跨地域参与会导致较长的评审持续时间，整个过程参与度较低，而且增加成本¹。

在2022年，我们发现技术能力是相互促进和积累的。改进代码评审速度能够有助于提高多种技术能力，包括代码的可维护性、学习型文化（知识传递）和建立生机型文化²。

更快的代码评审也是松耦合团队的优势之一，它可以显著提高您的软件交付效能和运维效能。提高代码评审效率有几种方法。当所评审的代码只涉及团队架构的特定领域时，评审者更能理解和明确代码对整个系统的影响。

评审的代码量越小，评审者越容易理解代码变更的含义。小批量工作模式可以提高团队的反馈周期、效率和专注度³。无论您当前的架构和集成实践是什么，结对编程是一种可以缩短您的代码审查时间的实践⁴。

此外，这些技能和流程对参与工作个体的幸福感没有不利影响。事实上，大多数的预示都是对员工幸福感有所改善。

¹ 基于主客观数据调查分布式软件开发中同行代码评审的有效性 <https://iserd.springeropen.com/articles/10.1186/s40411-018-0058-0>

² 现代化代码评审的期望、结果和挑战。 <https://dl.acm.org/doi/10.5555/2486788.2486882>

³ "小批量工作模式." <https://dora.dev/devops-capabilities/process/working-in-small-batches/>

⁴ "关于结对编程." <https://martinfowler.com/articles/on-pair-programming.html>

技术能力和流程	对职业倦怠的影响	对工作满意度的影响	对生产力的影响
人工智能	 小幅下降	 小幅提升	 小幅提升
持续集成	 暂无影响	 小幅提升	 暂无影响
快速代码评审	 大幅下降	 小幅提升	 小幅提升
松耦合架构	 大幅下降	 大幅提升	 大幅提升
主干开发	 大幅提升	 暂无影响	 暂无影响

*您可能会注意到，对于“职业倦怠”这一指标，颜色方案是颠倒的。这是因为减少倦怠是一件好事！

我们看到，采用松耦合架构、持续集成以及高效的代码评审能够帮助团队在保持甚至提升其成员幸福感的同时，提高了组织级的成果。

当团队拥有自主权去改进和维护一个为用户提供价值的可靠系统时，他们会体验到更高的工作满足度、团队效能和软件交付效能。在团队努力聚焦用户并提升软件交付质量的过程中，架构的选择发挥了至关重要的作用。通过小处入手和聚焦用户，团队在主干开发、松耦合架构、持续集成、持续交付以及SRE等多个方面都取得了显著的进步。若想进一步提高技术能力，应鼓励团队开展实验⁵并持续改进，请为团队提供这样的机会。



⁵ 团队实验: <https://dora.dev/devops-capabilities/process/team-experimentation>

持续交付的收益

作者: Dave Farley

持续交付（CD）的基本原则是确保我们的软件始终处于随时可发布的状态。为了实现这一点，我们必须保持高质量的工作。这意味着，一旦我们检测到问题，我们可以轻而易举地进行修复，从而快速和轻松将软件恢复到可发布状态。

为了确保我们的软件持续保持在可发布状态这一条黄金准则，我们需要努力建立快速的反馈机制，并能从失败中快速恢复。

作为本年度报告的读者，我预想这些观点对您而言并不陌生。**稳定性指标**（变更失败率和部署失败恢复时间）都与质量有关，而**吞吐量指标**（变更前置时间和部署频率）都与反馈的及时性以及轻松检测任何问题有关。

如果您在实践持续交付，那么您在**稳定性和吞吐量**上的表现必定出色。如果您在**稳定性和吞吐量**上表现卓越，那么很难想象您在达到这样的高水平时没有实践持续交付。

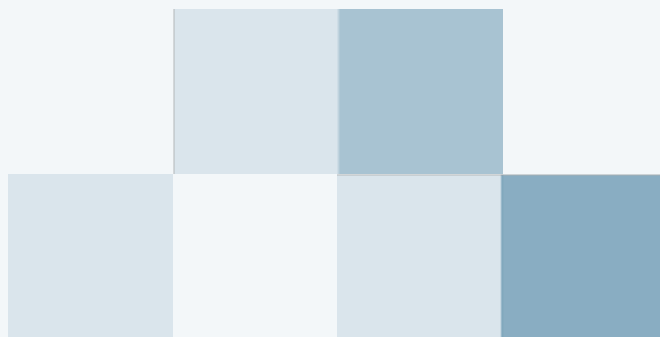
今年的分析深入探讨了通过寻找每个能力的催化剂来了解能力如何驱动效能的研究。持续交付——即快速、安全、可持续地按需发布各种变更的能力，是许多技术能力的重要催化剂。换言之，这些能力之所以能发挥作用，很大程度上是因为它们创造了一个使落地持续交付成为可能的环境。

持续交付实践反过来为这些能力预示更优的软件交付效能提供了机制。

在整体的软件开发中，可发布性是一个非常重要的标准，这也是持续交付强调它的原因。可发布性之所以至关重要，是因为虽然它虽然带有主观性，但确实为特定上下文的质量提供了一个明确的标准声明。开发关键安全系统与开发蛋糕店系统，在定义可发布性的严格程度时有所不同。但在这两种情况下，可发布性都定义了我们认为所必要的一切，以此来说明这段代码已经就绪，足够好且足够安全，并可以发布给最终用户。

因此，通过优化来保持变更的可发布性，实际上也是为我们的系统优化出了一个符合特定上下文的最低可接受的质量定义。

那些将获取并执行高质量、快速反馈的作为高优先级的团队，通常具有更佳软件交付效能。



我有些惊讶持续集成（CI）和主干开发对软件交付效能的影响没有那么大。尤其是CI对我来说似乎相当基础，所以这对我的世界观有些挑战，但分析这些问题是复杂的。例如，如果我们的代码不集成，我们如何在吞吐量上获得高分呢？如果我们没有检查，我们如何确保我们的稳定性很高呢？对我来说，CI就是我们知道这些事情的方式，所以它是软件交付效能的关键中介。这是理解的问题还是更深、更重要的问题？这很有趣！

技术能力与流程	对软件交付效能的影响	* 持续交付作为中介？
人工智能	 小幅降低	 否
持续交付	 小幅提升	 是，全部
快速代码评审	 大幅提升	 是，部分
松耦合架构	 小幅提升	 是，部分
主干开发	 小幅提升	 是，全部

* 中介（或催化剂）是一种评估可能的机制或行动路径的测试。您可以说：“X通过{中介}影响Y”。完全中介意味着整个效应似乎都是通过中介来解释的。部分中介是指只有部分效果是通过中介解释的。

优化组织流程和能力

我们都知道，文化是成功的关键驱动力。那么究竟是什么塑造了文化呢？这是一个饶有趣味的问题，而对于这个问题，似乎所有人都有着一个最普遍的答案：这视情况而异！

从实践者的角度出发，改进日常工作方式对文化中的诸多元素都有积极的影响，如风险共担、增强协作和建立心理安全感。举例来说，定期将变更整合到版本控制系统的主分支能够促进知识共享和团队合作。当安全团队与开发团队并肩合作，共同协作安全策略即代码，这将增加团队之间的信任，以及对部署变更的信心。

从领导角度来说，文化的萌芽首先来自于对其重要性的深刻认识和启蒙教育。变革型领导力⁶有助于塑造一个免责的环境，在这个环境中鼓励创新实验、持续学习，并赋予实践者应有的信任和话语权。工程师的意义是为了解决复杂的问题，而不仅仅局限于响应所分配的任务请求。为了达到这个目标，他们需要对整体业务有清晰的洞察，并拥有根据实际情况采取行动的自主权。归根结底，企业的文化很大程度上是由领导层影响和塑造的。

理想情况下，综合自上而下和自下而上的视角来审视文化会得到最佳的结果。

⁶“变革领导力”<https://dora.dev/devops-capabilities/cultural/transformational-leadership>

人工智能 (AI)

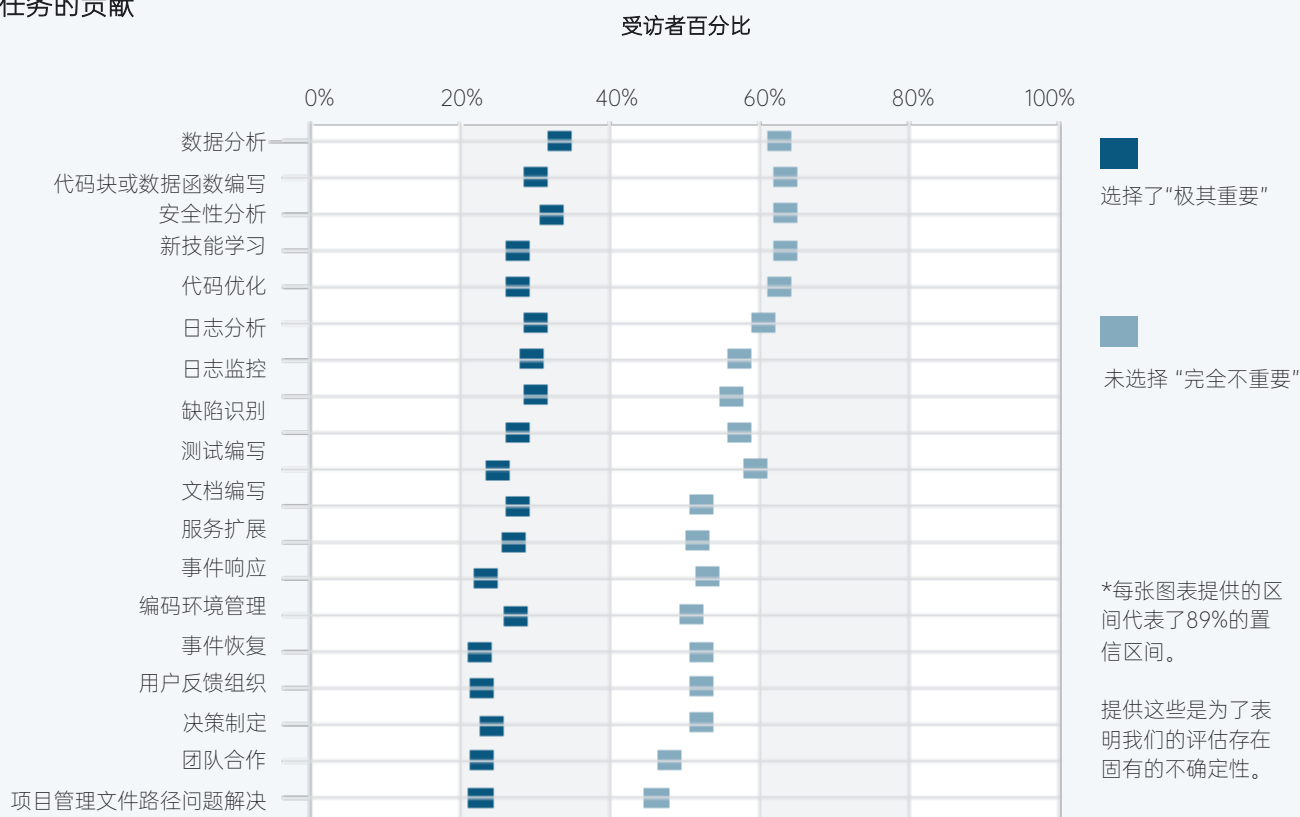
分析师和技术专家推测，人工智能将提升软件团队的效能，同时也不会对专业人员的幸福感产生负面影响。但根据我们迄今为止的调查数据，这一推测并未得到有力的支撑。我们的调查结果表明，AI能够略微改善了个体的幸福感指标（如职业倦怠和工作满意度），但对团体层面的效果（如团队效能和软件交付效能）没有甚至可能会有负面影响。

我们推测，这种混杂的证据或许可以从企业还处于AI工具采纳的早期阶段来解释。这很有可能，毕竟一些大型企业在决定是否广泛采用AI工具之前，会对不同AI驱动的工具开展试用测试。人们对AI开发工具潜力抱有很高的热情，正如调查所证明的那样，大多数人在我们提及的任务中至少融入了部分AI。这在下面的图表中得到呈现。但我们预计，AI驱动的工具在行业中得到广泛和有序的使用还需要一些时间。

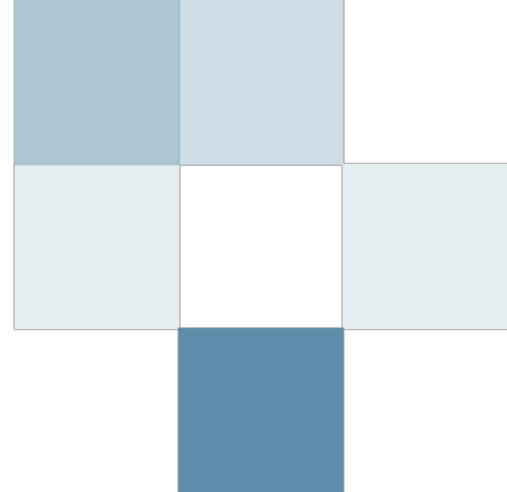
AI的重要性

AI对技术任务的贡献

对于您在应用程序或服务上的主要工作，如今的人工智能（AI）对以下任务的促成有多重要？



我们对看到随着时间的推移采纳会如何增长，以及这种增长对组织重要的效能指标和成果产生的影响非常感兴趣。



要点

高质量的文档是技术能力成功实施的基石，同时也放大了这些能力对组织效能的影响。此外，优质文档还积极地促进团队效能、生产力和工作满意度等方面的成效。然而，需要注意的是，提升文档质量并不一定会使每个人的效能都更出色——随着文档质量的不断提高，部分受访者反映出越来越严重的倦怠感。

介绍

今年，我们更深入地研究了内部文档——组织中人员日常使用的书面知识。我们调查了文档对技术能力和关键成果的影响。

为了度量文档质量，我们度量了文件的可靠性、可查找性、可更新性和相关性的程度。然后我们为整个文档的使用体验计算出一个分数。我们并没有逐页评估文档，而是将其作为一个整体进行评估。

结论

文档是基础：它驱动并增强技术能力

正如我们在 2021¹ 年和 2022² 年发现的那样，文档质量持续推动我们所研究的每一项技术能力的成功实施。

如下表所示，文档质量也放大了每项技术能力对组织效能的影响，这与我们在2023年看到的情况类似。

技术能力	放大对组织效能的影响
持续集成	2.4x*
持续交付	2.7x*
主干开发	12.8x*
松耦合架构	1.2x*
可靠性实践	1.4x*
人工智能贡献	1.5x*

计算依据：
高文档质量的技术能力的影响
低文档质量的技术能力的影响

文档预示着高产和快乐的个人和组织

除了提高技术能力，我们发现高质量的文档对个人的幸福感有积极的影响：它可以减少倦怠，增加工作满意度，并提高生产力。我们发现，这种影响在一定程度上是因为高质量文档增加了知识共享。

这种影响并不令人惊讶。当您知道怎么做的时候，事情会更容易完成，当分享知识的时候，工作也就变得不那么无聊。

福祉的各个方面	文档质量的影响
职业倦怠	 大幅减少
工作满意度	 大幅增加
生产力	 大幅增加

高质量的文档也会对关键成果产生推动作用，影响团队效能、组织效能和运维效能。

关键成果	高质量文档的影响
团队效能	 大幅提升
组织效能	 大幅提升
软件交付效能	 暂无影响
运维效能	 大幅提升

* 我们仍然惊讶地发现文档质量对软件交付效能没有影响。这是我们连续第二年看到这种反应。

¹ 2021 加速度 DevOps 报告
<https://dora.dev/publications/pdf/state-of-devops-2021.pdf>

² 2022 加速度 DevOps 报告
<https://dora.dev/research/2022/dora-report/2022-dora-accelerate-state-of-devops-report.pdf>

³ 2022年DevOps报告数据深度挖掘:文档就像阳光。
<https://cloud.google.com/blog/products/devops-sre/deep-dive-into-2022-state-of-devops-report-on-documentation>

这三个关键结果背后的积极影响是什么？对于读者而言，使用明确的文档是有益无害的，而撰写的过程也可能是关键。制定高质量的文档要求团队首先明确流程，文档能使组织内的各团队明确地讨论并达成共识，明确自己的工作内容和方式。

高质量的文档同样作为团队的知识库，尽管团队成员的更迭不可避免，但它能够帮助知识和经验跟随组织随时间发展而积累和扩散。

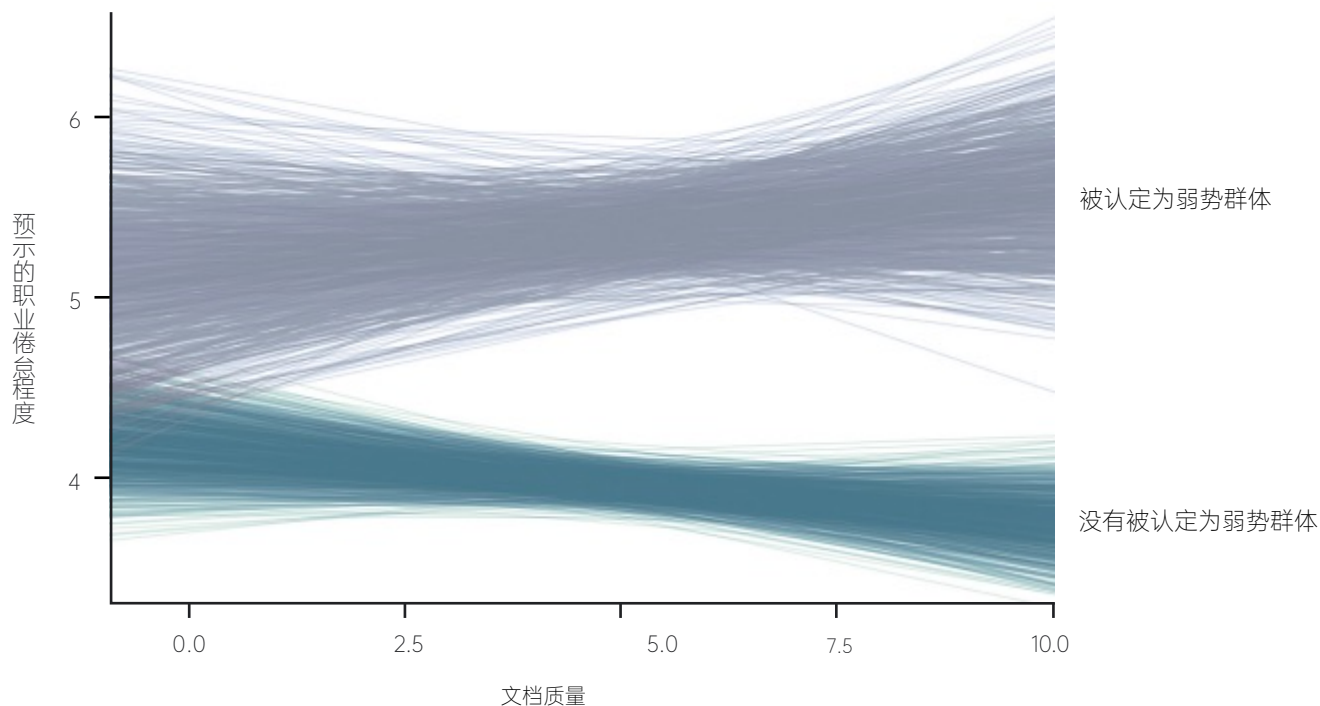
对于某些人，文档是否与幸福感降低有关？

当观察那些被认为是弱势群体的受访者时，我们注意到一个出乎意料的趋势：对于这一群体，文档的质量会增加职业倦怠感。

为了更深入地了解这一结果，我们也考虑了性别因素，但令人惊讶的是，性别并没有造成任何显著影响。无论是认为自己是男性、女性还是自述性别的受访者，他们都认为通过高质量的文档，工作倦怠有了显著的减少。但是，那些被视为弱势群体的人，不论其性别如何，都表示在高质量文档的环境下，他们的倦怠感明显增加。

以下图表基于我们的数据进行了模拟预测。在图表的下半部分，我们可以看到随着文档质量的提高，大多数受访者的工作倦怠程度有所降低。但在上半部分，我们注意到那些认为自己是弱势群体的受访者，随着文档质量的提升，他们的倦怠感显著上升。

这张图为每个群体模拟了 1000 条趋势线。线条越密集，意味着根据我们的数据，这种趋势的可能性越高。



文档质量、生机型文化和团队稳定性中的这一发现很类似：随着这些属性的增加，被认为弱势群体的倦怠感也会增加。那么对于文档场景来说，到底发生了什么？

创建和维护高质量的文档需要付出努力。这是技术性工作，对技术能力、团队生产力和组织效能有重大影响。但这类工作的努力和付出并不总能得到与其重要性和影响力相匹配的认可。被认为弱势群体的受访者是否承担了过多的这类工作？如果属实，那不就能解释文档对倦怠的影响了吗？

过度依赖文档会产生问题吗？难道随着文档质量的增强，对一些受访者而言，知识共享其实没有增加？或者，如果共享确实增加了，但仍然存在不足以抵消导致这个群体倦怠的其他方面或因素？

有可能还存在其他因素在发挥作用，虽然推动了高质量的文档，但也导致或维持了被认为是弱势群体受访者的倦怠。这些还需要更多的研究。

在团队中找到自己的定位似乎很重要（您是谁）。工作环境中的方方面面，比如高质量的文档，对团队和整个组织都有显著的好处。但对某些人来说，它们也可能会带来负面结果。我们会在第 8 章中进一步探讨——如何，何时，为何，找到定位很重要。

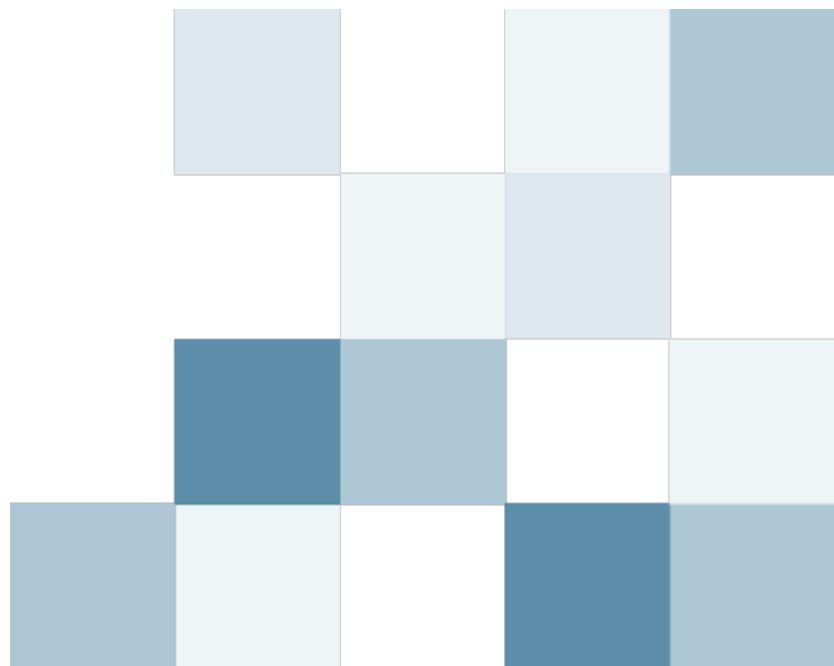
⁴ 2021 加速度 DevOps 状态报告, P22.
<https://dora.dev/publications/pdf/state-of-devops-2021.pdf#page=22>

入门资源

有关推动高质量文档的实践，请参阅 2021 年报告⁴。我们在今年还发现，工作的合理分配，包括文档工作分发的标准流程，显著提高了文档的质量。

有很多关于技术写作的培训资源。您可以从这些资源中学到更多：

- 技术交流协会 (stc.org)
- 工程师技术写作课程 (developers.google.com/tech-writing)
- 文档撰写 (writethedocs.org)



可靠性释放效能

要点

强大的可靠性实践预示着更好的运维效能、团队效能和组织效能。数据显示，改善这些实践的效果遵循非线性路径——也就是说，在组织构建更强大的能力时，效能改善可能会停滞不前。然而，随着时间的推移，坚持这些实践仍然预示着良好的结果。

介绍

可靠性是IT运维领域广泛使用的术语。我们将可靠性定义为服务在可用性、性能和正确性等衡量标准上达到预定目标的程度。实现可靠性成果的常见方法是SRE（网站可靠性工程），该方法起源于谷歌（<https://sre.google>），现在许多组织都在采用。SRE强调经验学习、跨职能协作、高度依赖自动化以及使用服务级别目标（SLO）等测量技术。

许多组织使用可靠性实践但不采用SRE里的术语，包括生产工程师、平台团队、基础设施团队、TechOps等。为了尽可能客观地评估这些实践的广泛程度，我们的调查在调查文本中使用中性的描述性语言。

我们还收集了可靠性工程的成果数据，即团队能够实现其可靠性目标的程度。可靠性实践和可靠性成果（我们称为运维效能）均反映在我们的预测模型和其他能力中。

可靠性实践

我们要求受访者从其运维的三个方面思考可靠性。首先，他们是否为依赖项制定了消减计划？其次，他们是否通过模拟中断、实际故障转移或桌面演习定期测试其灾难恢复计划？最后，当他们未能达到可靠性目标时，他们是否进行改进工作或以其他方式重新确定优先级和调整工作？

我们认为这些措施体现了遵循既定SRE原则（例如“接受风险”和“衡量用户满意度”）的团队的精神。这样的团队设定一个与用户满意度相符的合理目标。然后，他们进行测试以确保能够达到该目标，但如果遇到困难，他们会改变计划。我们以此作为团队成功“实施SRE”的代理指标，而不需要将团队评估与特定的SRE实现绑定起来。

结论

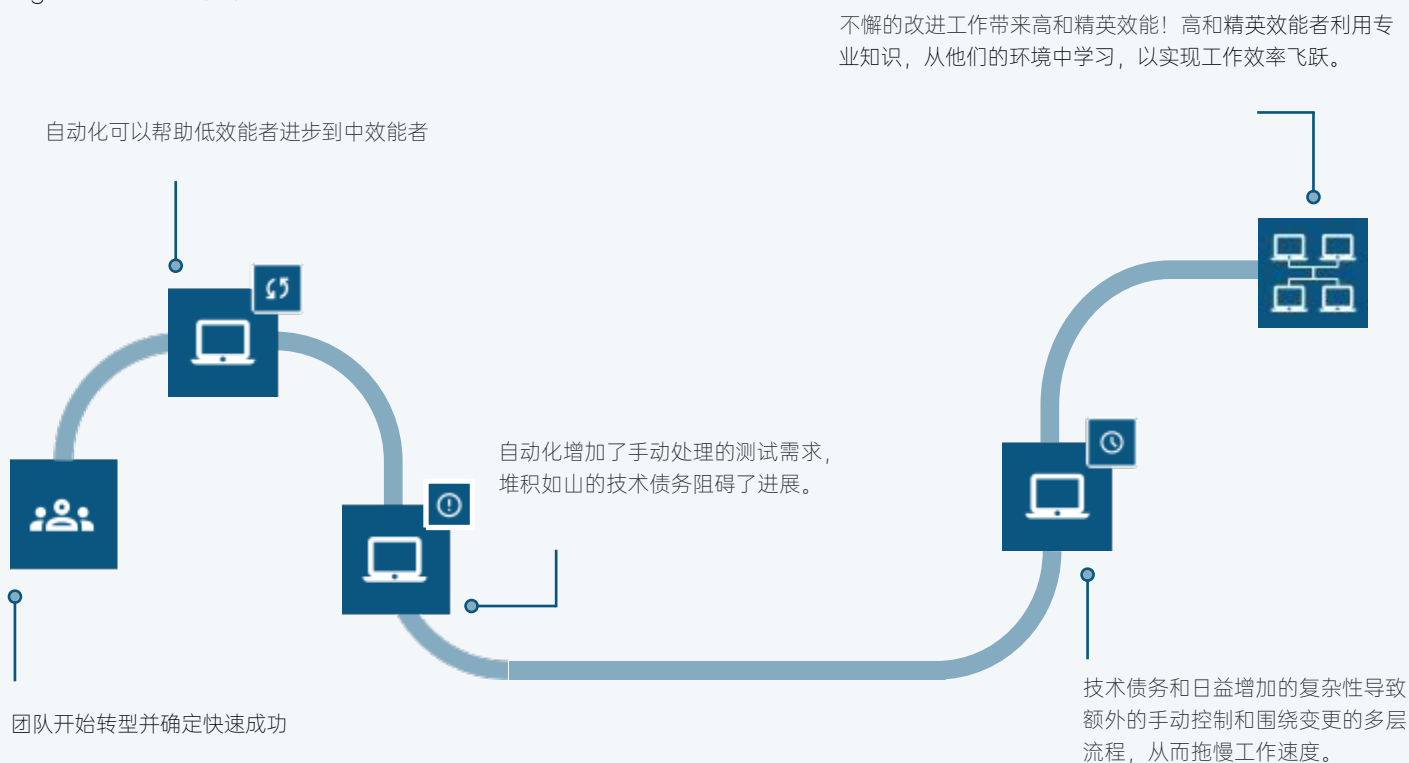
证实可靠性实践的J形曲线

自2018年以来，DORA一直认为运维效能和自动化等实践之间存在非线性关系（图1）。

随着我们对可靠性实践的深入探索，我们已经在调查数据中看到了这种模式的证据。

在 2022 年，我们直接测量了这一点。我们对团队进行了调查，并观察到可靠性实践和可靠性结果之间的关系确实遵循这种类型的非线性曲线(图 2)。这表明团队只有在采用许多可靠性实践后才会看到显著的可靠性提升¹。但是，以这种方式看到的数据并不能让人感觉我们看到了全貌。2022 年的曲线让人觉得 SRE 只适合专家领域，或者不值得投资，这与许多 SRE 团队的经验相冲突。我们需要更多的数据。

Figure 1: 2018年假设的 J 曲线



¹ 2022 加速度 DevOps 状态报告 P27-28. <https://dora.dev/research/2022/dora-report/2022-dora-accelerate-state-of-devops-report.pdf#page=27>

图2：2022年曲线

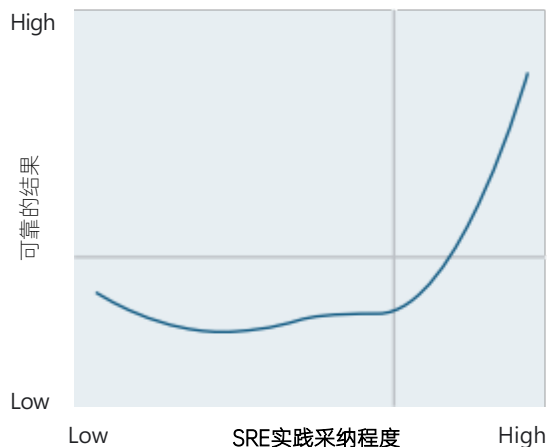
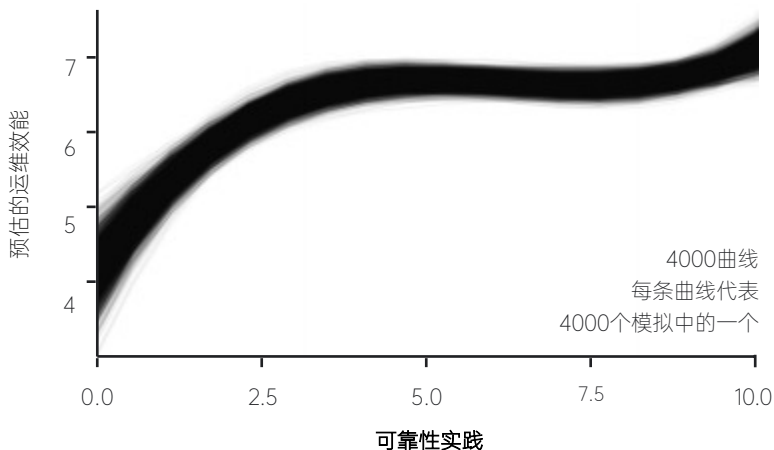


图3：2023年曲线



2023年，我们能够提出更多问题，这有助于更好地定义更接近我们实际经验的曲线。新的曲线更接近2018年报告中描述的假设的转型J形曲线（有关我们如何进行分析的更多信息，请参见方法论部分）。这表明在采用可靠性实践方面确实存在早期的胜利，随后随着复杂性引入新的挑战而出现间歇，然后最终在运维效能上再次上升。这些结果强化了我们在许多团队中看到的情况。

这条曲线之所以重要，有以下几个原因：

- 它有助于公司合理化并投资初始的SRE采用，即使他们不寻求极端的可靠性水平或不期望在SRE上进行大量投资。即便是采用很小水平的可靠性实践也可以促进运维效能的改善，这对团队效能和组织效能产生进一步的有益影响。

- 它为那些预期大力投资可靠性的公司做好准备，使其能够平稳的渡过任何平台期或低谷。从长期投资SRE中期待线性积极结果可能很诱人，但是数据告诉我们并非如此。当团队提前了解该曲线的非线性时，他们可以做出是否进行这项投资的决定，并且可以提前制定计划以确保在实现全部效益之前不会中途放弃。
- 这样的变化可能需要文化上的转型。我们发现成功往往源自于自下而上和自上而下相结合的变革。团队可以采纳可靠性实践并获得即时回报，然后将这些好处展示给其他团队，并得到领导层的支持和鼓励。这些激励措施和结构化的项目在设计时就可以考虑J形曲线。

²“如何转型”- <https://dora.dev/devops-capabilities/cultural/devops-culture-transform/>

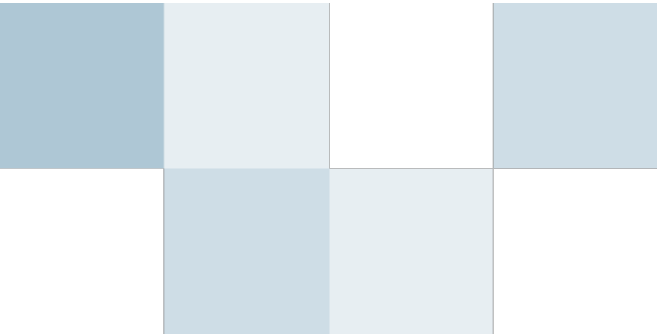
可靠性实践与员工幸福感

传统运维实践通常是高度反应式，而且通常更加关注技术系统的健康状况，却忽视了最终用户的愉快体验。对那些不会影响用户体验的事物进行告警待命，高度重复的手动作业、畏惧犯错的心理以及类似的经历会导致团队成员倦怠和幸福感下降。

而那些采纳可靠性实践的团队，情况却正好相反。从这些团队报告中看到，与未采纳这些实践的同行相比，他们的生产力和工作满意度更高，而且倦怠程度更低。

我们怀疑这些员工幸福感方面的改善是由一些已发布的SRE实践所驱动的：

- 减少无效劳动³
- 免责的事后剖析⁴
- 团队自主决策权⁵
- 团队非线性增长⁶



³ Beyer, Betsy, et al. 《站点可靠性工程：谷歌如何运行生产系统》（O'Reilly, 2016），49–54. <https://sre.google/sre-book/eliminating-toil/>

⁴ 同上, <https://sre.google/sre-book/postmortem-culture/>

⁵ Beyer, Betsy, et al. 《站点可靠性工作手册》（O'Reilly, 2018），<https://sre.google/workbook/team-lifecycles/>

⁶ Brookbank, James, and McGhee, Steve. 《到SRE的企业路线图》（O'Reilly, 2022），11. <https://sre.google/resources/practices-and-processes/enterprise-roadmap-to-sre/>

运维效能

我们还要求受访者描述其服务的运维效能。首先，我们询问他们直接从用户那里收到对关于服务可靠性不满的频率是多少。接下来，我们问他们服务不可用、响应缓慢或以其他方式运行不正确的频率是多少。

➤ 可靠性实践让团队和组织效能得以增强

通过对可靠性实践的采纳，团队能够改善其运维效能。如果一个组织能够有效地运维其生产系统，我们发现这会增强和放大其他方面的成效。如果成效本就很好，可靠性实践会使它们变得更好。如果成效较差，可靠性实践也不会带来帮助，它们只会保持原状。

即使系统可靠，我们仍然需要面向客户提供正确的软件功能，并确保有效地交付这些功能。这么说是有道理的，因为SRE不是为孤立运作而设计。达到可靠性目标是SRE团队成功的关键指标，这也会反映在运维效能上。尽管使用可靠性实践可能会带来其他好处，但数据表明，最关键的好处还是对运维效能的积极影响。此外，提高运维效能不仅仅在于提高服务健康状况，事实上，我们看到的证据表明，采用可靠性实践预示着从业者工作幸福感的提升。



运维效能影响员工幸福感

行业中的一个普遍看法是，高度可靠的服务会对服务维护者的幸福感产生负面影响，例如随时待命或工作时间之外的紧急维护。然而，我们发现高水平运维效能实际上会带来较低的倦怠、更好的生产力和更高的工作满意度。这与 SRE 减少无效劳动的原则一致⁷。将手动运维的部分自动化，既令个人感到满意，又能减少团队的持续负担。



组织效能和团队效能增强运维效能

我们发现运维效能对团队效能和组织效能都有实质性的积极影响。这对DevOps运动的追随者来说不应该感到惊讶。能够有效地操作机器使团队能够取得更多的成就，从而令组织蓬勃发展。



运维效能增强软件交付效能

尽管软件交付效能可以改善团队效能和组织效能，但它们都显著得到运维效能的增强。此外，没有高水平的运维效能，高效能的软件交付团队也无法实现非常高的团队效能和组织效能，这两者都是必需的。事实上，在不具备相应水平的运维效能的情况下提高软件交付效能的团队最终会产生较差的组织级成效。所以，如果您的软件能够快速编写出来，但是并没有按预期方式在生产环境中运行，那么市场就不会给予任何回报。

不足和展望

我们相信，还有更多的度量标准可以帮助我们理解这些能力之间的相互作用。例如，今年经常被提及的一个问题是——成本管理是如何与这些能力和成效结合起来。某些组织对成本的敏感度会高于其他组织，这会影响组织的规划和决策的方式。同样，我们推测，在高度协作的文化中⁸，即使没有被明确提及或规划可靠性实践，它们也可能会自然地发生。我们希望更深入地了解团队如何演进其现有的IT运维实践，以及这种发展如何影响系统可靠性、团队效能和员工幸福感。

更重要的是，我们希望听到您的声音。欢迎您加入我们，与其他实践者在DORA.community网站⁹上交流。SRE仍然是一个新兴领域，每一个采纳可靠性实践的组织，或那些意识到自身已经一直在开展SRE的组织，其所产生的影响是不同的。这些变化的进程是缓慢的，但希望随着时间的推移，我们可以通过持续的度量来展示所取得的进步。作为一个社区，我们可以相互分享有价值的东西，在前进的过程中互相鼓励和提升。

⁷ Beyer, Betsy, et al. 《站点可靠性工作手册》(O'Reilly, 2018), <https://sre.google/workbook/team-lifecycles/>

⁸ Brookbank, James, and McGhee, Steve. 《到SRE的企业路线图》(O'Reilly, 2022), 11. <https://sre.google/resources/practices-and-processes/enterprise-roadmap-to-sre/>

⁹ DORA 社区, <https://dora.community/>

谷歌是如何做 SRE 的

在谷歌内部，SRE 在一个不断发展和资金雄厚的组织中已经发展了 20 年。推动 SRE 的原动力是一种需要，即在不过度影响财务投入的情况下，使谷歌搜索和广告实现超高速增长。早期，这些产品的实时性质揭示了对高可靠性的需求——搜索或广告中的短暂错误意味着立即失去客户，而且没有重来的机会。动态广告必须在数毫秒内计算出来，缓慢的搜索会破坏谷歌品牌的所有承诺和核心价值。

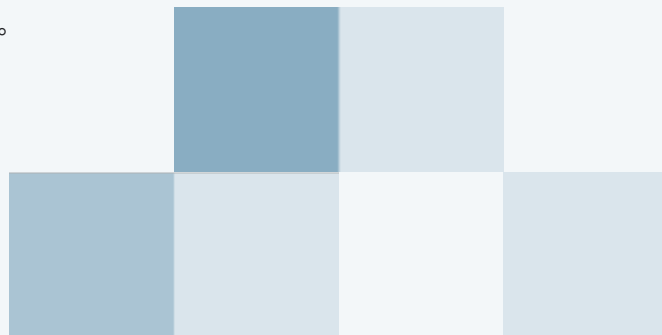
与此同时，SRE 正在一种新型公司的环境中孕育和壮大：这是一个是自下而上、以工程师为主导的公司，它选择自行构建技术而非直接购买。谷歌聘用了一批擅长分布式系统和编译器设计，在计算机科学学术领域表现出色的站点可靠性工程师（SRE）。这种文化基因给 SRE 的崛起和繁荣创造了一个丰饶的土壤，并使其蓬勃发展。SRE 团队不仅是负责生产系统的关键，而且还肩负着寻找扩展系统的新兴和新颖方法。

那么 SRE 是如何随着时间的推移而扩大规模的呢？答案是：次线性。也就是说，谷歌不可能在规模扩大两倍的同时，员工数量也增加一倍。鉴于早期谷歌产品扩大规模以满足全球用户需求的速度，以及 Gmail、谷歌地图、Android、YouTube 和谷歌云等新产品的推出，SRE 这类新兴和稀有的资源不可能以客户规模增长相同的速度扩大。

一种说法诞生了：

- SRE 不应随用户数量线性增长。
- SRE 不应随服务器数量线性增长。
- SRE 不应随群集数量线性增长。
- SRE 不应随服务数量线性增长。

为了保持这种受约束的增长模型，一种特定的管理结构被建立。SRE 之间需要进行对齐和协同，同时与他们合作的产品开发团队也需要密切合作。SRE 通过其自有独立的管理体系向上汇报，其中 Ben Treynor-Sloss¹⁰ 处于最高汇报位置。这些团队将自己按产品领域划分，并与对应的产品开发团队完全拉通。SRE 团队与他们的开发团队共同协作，探讨和决定如何最大化地利用手头的 SRE 资源。



¹⁰ Benjamin Treynor Sloss, 工程副总裁。《站点可靠性工程：谷歌如何运行生产系统》(O'Reilly, 2016).
<https://sre.google/sre-book/part-i-introduction/><https://sre.google/sre-book/introduction/#id-2opuzSiFr>

开发团队有能力直接对新的SRE提供资金支持。并非所有开发团队都会推出面向用户的产品，其中许多开发团队是如Bigtable¹¹（结构化数据存储）、Borg¹²（计算调度）和Colossus¹³（分布式存储）这样的公共基础设施团队。通过这些公共的基础设施服务，面向客户的团队能够在不拥有自己专属SRE团队的情况下实现扩展。

通过将这些团队留在自己的组织中，团队能够保持一致的招聘和晋升流程。一般来说，SRE团队的规模常常只有开发团队比例的十分之一甚至更少，因此确保SRE团队拥有自主权，不向任何违背SRE原则的方向发展是很重要的。

SRE开发自己的内部产品，并配备了自己的内部产品经理。SRE产品的用户是想要更好地运维其生产系统的其他团队。SRE团队的产品围绕事件响应、监控和可观测性、发布管理、容量规划以及故障排除进行设计与开发。

SRE团队的发展过程需要始终考虑团队健康和可持续性。例如，新设立的轮班SRE团队的最小规模应为12人：两个站点，每个站点6人。这种配置实现了跨时区的覆盖，团队也中有足够的人数，可以在不牺牲任何人的情况下保持良好的工作与生活平衡，避免过度劳累至精疲力竭。

SRE至今仍在持续演变和适应，并非所有团队都完全严格遵循上述模型，有些团队会另辟蹊径。像谷歌内部的SRE这样的大型组织也会受到市场环境的影响，就像任何其他大型组织一样，因此灵活性非常重要。但最重要的是，SRE团队始终秉承自己的原则：勇于承担风险、准确衡量服务水平、努力消除无效劳动、积极拥抱自动化并始终追求简单高效。



¹¹ Chang, Fay 等，Bigtable 面向结构化数据的分布式存储系统。7th USENIX Symposium on Operating Systems Design and Implementation (OSDI), {USENIX} (2006), pp. 205-218, <https://research.google/pubs/pub27898/>

¹² Verma, Abhishek 等，Borg 谷歌大规模集群管理, Proceedings of the European Conference on Computer Systems (EuroSys), ACM, Bordeaux, France (2015), <https://research.google/pubs/pub43438/>

¹³ Hildebrand, Dean等，Colossus: 谷歌可扩展存储系统预览。April 19, 2021 - <https://cloud.google.com/blog/products/storage-data-transfer/a-peek-behind-colossus-googles-file-system>

灵活的基础设施是成功的关键

要点

灵活的基础设施是团队效能、组织效能、运维效能以及软件交付效能的预示指标。云计算是实现灵活基础设施的核心驱动因素，但这种优势并不能够通过自动实现来获得：我们的数据显示，如何使用云才是关键。

介绍

在DORA的大量研究中，我们通过聚焦云计算的基本特征（由美国国家标准与技术研究院定义¹），询问了从业者关于他们的基础设施情况。

- 按需自助服务
- 广泛网络接入
- 资源池
- 快速弹性
- 可计量的服务

我们一直认为这五个特征预示着更好的组织效能和更优秀的软件交付效能。今年，我们进一步探索使用云计算是否预示着更加灵活的基础设施。



¹ NIST特别报告 800-145: “NIST对云计算的定义”。



	百分比
多云	19.6%
公有云	51.5%
混合云	33.6%
本地化	19.8%
台式机	3.2%
其他	2.5%

* 被调查者可以选择多个答案

结果

再一次，我们确认了之前的发现：一个团队**如何**使用云计算是一个比他们**是否**使用云计算更重要的效能预示因素。虽然使用云计算可以成为一个强大的推动因素，但它并不会自动产生效益。事实上，我们有强有力的指标显示，除非团队使用灵活的基础设施，否则公有云可能会导致软件和运维效能的下降。这一发现进一步推动了单纯的“lifting and shifting”（将工作负载从数据中心转移到云端的行为）并无益处，甚至可能有害的想法。使用云计算与员工的职业倦怠大幅减少、工作满意度和生产率大幅提高有关。

计算环境

上表显示了受访者表示他们主要的应用程序或服务运行的位置。



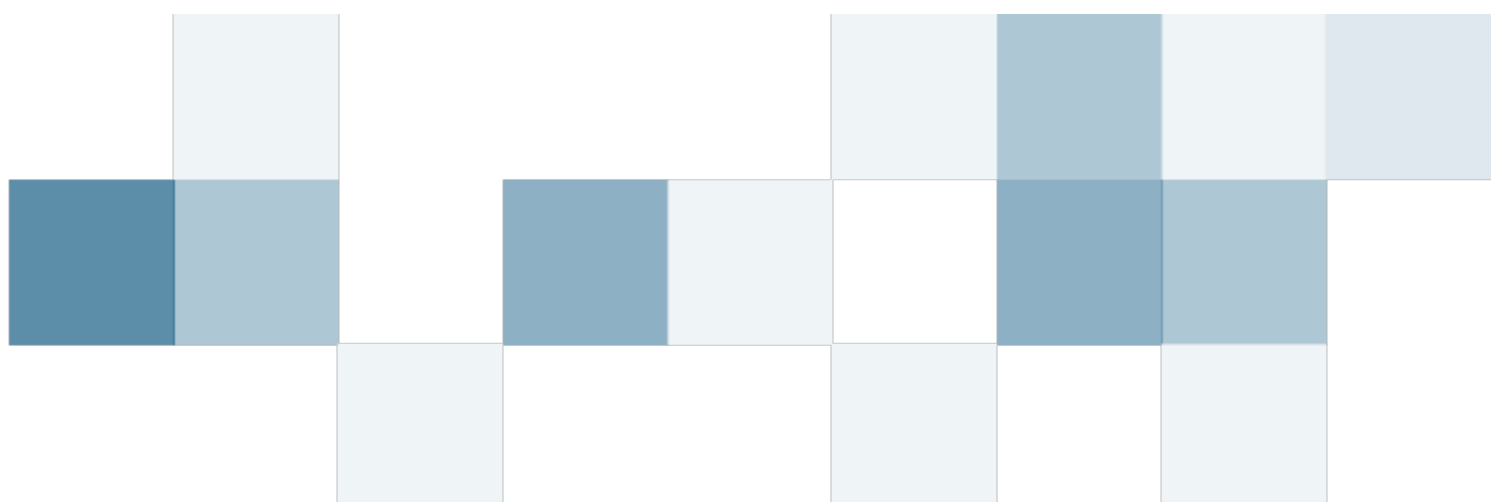
云类型	组织效能	团队效能	软件交付能力	运维效能
私有云	 暂无影响	 使用云计算带来大幅提升	 暂无影响	 使用云计算带来大幅提升
公共云	 使用云计算带来大幅提升		 使用云计算带来大幅下降	 与使用云计算相关的非常显著的下降
混合云				 暂无影响
多云			 暂无影响	 使用云计算带来大幅下降

单纯地“使用云”会带来好坏不一的结果

如表格中的结果所示，单纯“使用云”对软件交付和运维效能要么没有影响，要么有负面影响。这种中性或负面的影响很可能是因为从业人员在刚迈出他们在云计算之旅上的第一步的结果，现在他们面对的是一个全新的环境、全新的工具和不同的方式进行操作。通常，企业只是用他们在自己的数据中心相同的方式来使用云，这样只会增加了对新环境的复杂性和认知负担。

未能适应新环境的失败之处在于，不但不能提高软件交付或运维效能，相反会带来负面的影响。这个发现的唯一例外是在私有云环境中的运维效能。

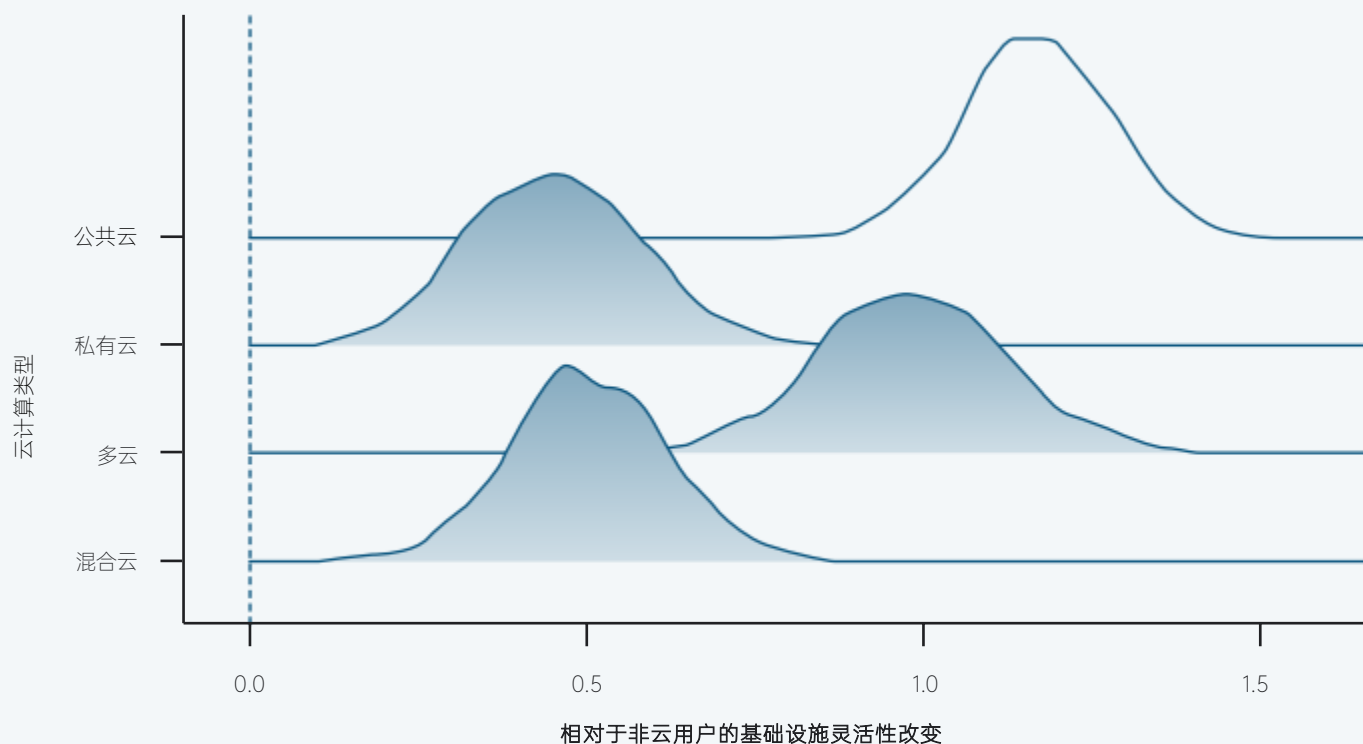
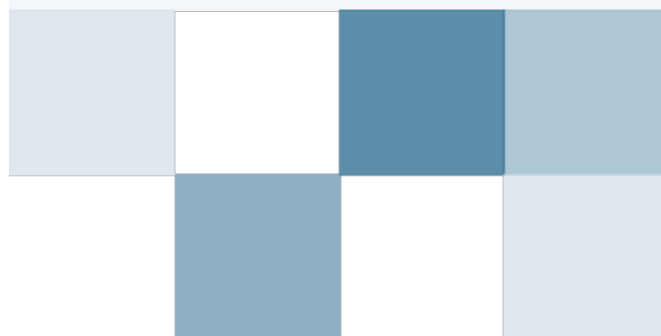
而真正能够提高软件交付和运维效能的，是灵活的基础设施，我们将在后续详细讨论这一点。



云基础设施能够实现灵活性

与不使用云相比，使用公共云促使基础设施灵活性提高了22%。同时，尽管不如单个公共云，但使用多云也同样能够提高灵活性。这里的一个显而易见的问题就是——为什么会这样？我们的数据显示，由云计算赋能的灵活基础设施的影响力远超于单纯对某个云平台的使用。对于大多数人来说，云计算代表了一种新的方式，而要完全掌握它则需要时间。鉴于每个云平台都存在不同之处，这就意味着随着云平台数量的增加，运维每个平台所需的认知负担也会增加。

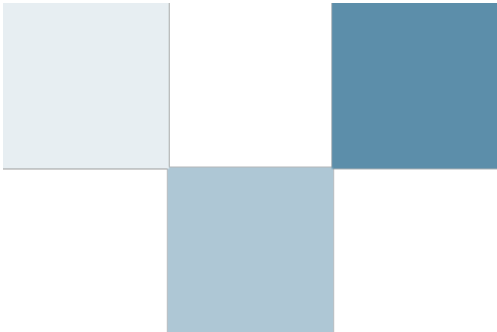
相对于不使用云，使用公共云
会促使基础设施灵活性的增加。



在关键成果上，灵活的基础设施预示着更高的效能。

能力	组织效能		团队效能	软件交付效能		运维效能
灵活的基础设施		更加灵活的基础设施带来大幅提升		更加灵活的基础设施带来大幅提升		更加灵活的基础设施带来小幅提升

重要的是要认识到，灵活的基础设施是推动组织效能、团队效能、软件交付效能和运维效能提升的关键。许多组织选择将基础设施迁移到云端，这已经是一个了不起的第一步，但这也只是旅程的开端。如果您决定迁移某部分工作负载，下一步就是通过重构来将它们现代化，使其真正利用到灵活基础设施的优势。



通过灵活的基础设施，云计算可以对关键成效产生积极的影响。

灵活基础设施的云类型	组织效能	团队效能	软件交付效能	运维效能
私有云	 部分由灵活基础设施促成	 部分由灵活基础设施促成	 部分由灵活基础设施促成	 部分由灵活基础设施促成
公共云	 完全由灵活基础设施促成	 完全由灵活基础设施促成	 部分由灵活基础设施促成	 部分由灵活基础设施促成
混合云	 部分由灵活基础设施促成	 完全由灵活基础设施促成	 部分由灵活基础设施促成	 完全由灵活基础设施促成
多云	 部分由灵活基础设施促成	 部分由灵活基础设施促成	 部分由灵活基础设施促成	 部分由灵活基础设施促成



当我们最大限度地利用灵活基础设施特性的方式使用云计算平台时，它对软件交付和运维效能产生的积极影响是显而易见的。这些影响上的差异也印证了许多实践者和领导者已有的认知——仅仅简单地将工作负载从数据中心迁移到云端并不能带来成功。真正关键在于，如何充分利用云所提供的灵活基础设施。

基础设施类型	结果
云与灵活的基础设施相结合	
没有灵活性的云	

若想充分挖掘这种潜在效益，我们必须重新审视如何构建、测试、部署和监控应用程序的方式。这一重新思考的核心在于充分利用云计算的五个特点：按需自助服务、广泛的网络接入、资源池、快速弹性和可计量服务。



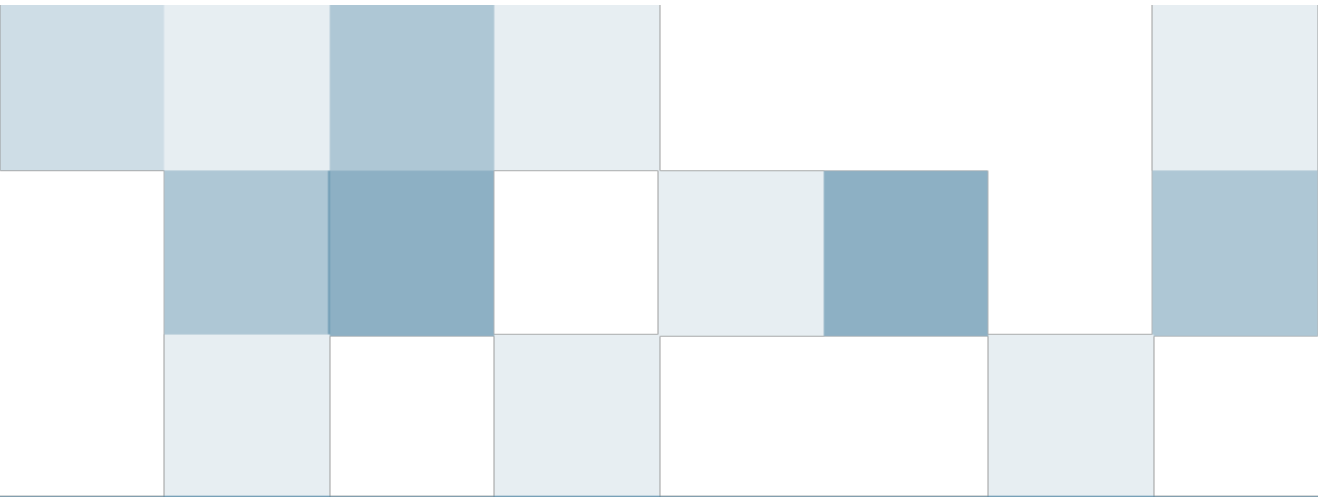
云计算能够提高员工幸福感

云类型	职业倦怠*	工作满意度	生产力
私有云	 暂无影响	 使用云计算带来大幅提升	 使用云计算带来大幅提升
公共云	 使用云计算带来大幅下降		
混合云	 暂无影响		
多云	 暂无影响		

*您可能已经注意到，对职业倦怠的影响后增加了*号。这是由于减少职业倦怠是一件好事！

数据显示，云计算在很大程度上对员工幸福感是有益的。我们也发现工作满意度和生产力都会显著增加，对职业倦怠也没有负面影响，甚至会产生积极影响。换句话说，尽管云计算带来了额外的认知负担，需要学习新的工具和新的工作方式，但云计算并没有对员工幸福感产生不利影响。

作为该领域的实践者，我们推测有以下几个原因解释了为什么会出现这种情况。工程师喜欢学习知识和解决问题，并且也乐于在具有灵活计算特征的环境中工作。学习新技术不仅富有趣味，更是一种卓越的职业发展方式。当工程师发现自己所在的组织日益成功，他们的工作幸福感也会随之上升。



第七章

文化是一切的根本

要点

文化是员工幸福感和组织效能的关键驱动力。一个健康的组织文化有助于减少员工的职业倦怠感、提高生产力，并提升工作满意度。此外，组织效能、软件交付和运维效能以及团队效能也会得到显著提高。一个良好的组织文化有助于团队更成功地运用和改善与成效有关的技术能力。



介绍

文化难以被定义。我们关注的是能够反映与员工工作体验有关的指标。我们采用 Westrum 的组织文化分类学¹，因为它一直是一个强有力的效能预示指标。在今年我们看到，那些属于Westrum 中定义的生机型文化的组织，仍有持续的出色表现。

下表列出了我们认为有助于塑造团队和组织文化的多个方面。

方面	定义
Westrum 组织文化	组织倾向于如何应对问题和机会。 组织的文化有三种类型：生机型、官僚型和病态型。
组织稳定性	员工所处工作环境的稳定性或不稳定性。
工作安全感	员工担心自己工作保障的频率。
灵活性	员工如何工作、何处工作以及何时工作。
知识共享	在组织中，想法和信息是如何传播的。同样问题团队成员只需回答一次问题，信息就可以被其他人获取。人们不必等待答案。
用户导向	在开发软件时以最终用户为中心，并深入理解用户的需求和目标，利用用户反馈来改进产品和服务。
工作分配	帮助团队在成员之间公平分配繁重任务的标准流程。

文化方面、流程能力与技术能力之间的界限并不总是很清晰。我们认为文化是从实践中产生的，而实践又源于文化。我们稍后讨论我们的发现时会涉及这一点。

¹ <http://bmi.co/1BRGh5g>

我们的发现及意义

健康文化改善关键成果

我们发现，健康的文化对所有关键成果都有积极的影响。我们重现前几年的发现：生机型文化能够推动组织效能、软件交付效能和运维效能。它同样也推动了今年新的效能指标——团队效能。

我们发现，以用户为中心的软件开发方法能够显著提高效能，这一点非常值得强调。当组织将用户放在第一位时，它们就能获得一系列的好处。用户反馈帮助团队确定项目的优先级，并帮助它们创建满足用户需求的产品和服务，从而带来更好的用户体验、更高的用户满意度和更高的业务收益。

我们还通过衡量团队之间的工作分配情况来评估组织文化的健康状况。我们发现，公平的工作分配对团队和组织效能有利。但是，我们也发现公平的工作分配与较低的软件交付效能有关。也许围绕工作分配的标准流程会降低完成软件交付流水线中繁重任务的速度。也可能是标准流程会影响团队中应该承担某项任务的人。

另一个看上去不太符合常规的发现是：组织稳定性对软件交付效能呈现出轻微但较为明显的负面影响。这可能是因为更成熟（可能规模也更大）的组织不会感受到像新兴、非成熟（规模较小）的组织同样的加速压力。这些更成熟的组织可能已经拥有成熟和稳定的产品，这给了它们在软件交付速度方面有了更多的灵活性。

拥有生机型文化的团队相比
没有生机型文化的团队
组织效能高出

30%

当信息的流动畅通无阻时，事情就会得到高效解决。我们发现，更高水平的信息共享与提高软件交付效能和运维效能正相关。当信息触手可及且知识孤岛很少时，人们就可以把时间花在关键任务上，而不是花时间去追寻那些完成任务所需的信息。

最后，灵活的工作安排，即允许员工可以决定他们何时、何地以及如何工作，对所有效能指标都有积极的影响。这尤其适用于软件交付效能。即使组织收紧远程工作政策，允许员工保持一定的工作灵活性也可能带来好处。

文化各方面	对团队效能的影响	对组织效能的影响	对软件交付效能的影响	对运维效能的影响
Westrum组织文化	 大幅提升	 大幅提升	 大幅提升	 大幅提升
组织稳定性	 小幅提升	 大幅提升	 小幅降低	 暂无影响
工作安全感	 小幅提升	 暂无影响	 小幅提升	 小幅提升
灵活性	 小幅提升	 小幅提升	 大幅提升	 小幅提升
知识共享	 小幅提升	 小幅降低	 大幅提升	 大幅提升
用户导向	 大幅提升	 大幅提升	 小幅提升	 大幅提升
工作分配	 大幅提升	 大幅提升	 大幅降低	 暂无影响



健康文化提升技术能力

我们的发现表明，良好的企业文化有助于提高技术能力的实现效果。

我们认为，企业文化和技术能力之间存在相互促进的关系：文化从实践中孕育，实践从文化中涌现。

相较于文化的宽泛和难以定义，技术能力通常是有明确的范围和定义的。这对于组织内部成员如何帮助推动变革具有启示作用。

例如，领导者可以构建激励机制来促进生机型文化的产生。无论是领导者还是个体贡献者，都应注重以用户为中心的软件开发方法。个体贡献者可以推动技术能力的实现来改善效能，如主干开发、持续集成、可靠性实践和松耦合架构。

要实现这些技术能力并不容易，成功做到这些需要团队成员的共同努力，持有开放心态，相互信赖并彼此学习。这些都是健康文化的重要组成部分。这样的团队可以成为组织内其他团队的标杆，使得其他团队更有决心和动力，利用他们所掌握的杠杆去推动变革。

对组织文化进行持久且深入的变革，需要结合自上而下和自下而上共同努力、制定变革。

文化各方面	对主干开发的影响	对可靠性实践的影响	对持续集成的影响	对持续交付的影响	对松耦合架构的影响
Westrum组织文化	 大幅提升	 大幅提升	 大幅提升	 大幅提升	 大幅提升
组织稳定性	 小幅提升	 大幅提升	 暂无影响	 暂无影响	 暂无影响
工作安全感	 小幅降低	 小幅降低	 暂无影响	 暂无影响	 暂无影响
灵活性	 暂无影响	 小幅降低	 大幅提升	 小幅提升	 大幅提升
知识共享	 暂无影响	 暂无影响	 暂无影响	 小幅提升	 小幅提升
用户导向	 大幅提升	 大幅提升	 大幅提升	 大幅提升	 大幅提升
工作分配	 大幅提升	 大幅提升	 大幅提升	 大幅提升	 大幅提升

健康文化提高员工幸福感

健康的企业文化通过减少职业倦怠感、提高工作满意度和提升生产力，从而带来员工高水平的幸福感。员工幸福感并不是一个锦上添花的附加品，它是组织整体健康和成功的基石。

当企业不重视不投入更好的文化建设，会出现什么情况呢？职业倦怠的风险会增加，工作满意度会下降。从而员工变得愤世嫉俗，其生产力也会随之下降。

员工的身心健康也会受到负面影响^{2, 3}。职业倦怠是持续性的，它不会因为短暂的休息就会克服或消失。职业倦怠也会导致人员流动率上升——员工会离职去寻找更健康的工作环境⁴。因此，真正要解决员工的职业倦怠问题，需要组织从根源上来改变。

文化方面	对职业倦怠的影响*	对工作满意度的影响	对生产效率的影响
Westrum组织文化	 大幅降低	 大幅提升	 大幅提升
组织稳定性	 大幅降低	 大幅提升	 小幅提升
工作安全感	 大幅降低	 小幅提升	 小幅提升
灵活性	 小幅降低	 小幅提升	 小幅提升
知识共享	 大幅降低	 小幅提升	 小幅提升
用户导向	 小幅降低	 大幅提升	 大幅提升
工作分配	 无影响	 小幅提升	 小幅提升

*您可能已经注意到，对职业倦怠的影响后增加了*号。这是由于减少职业倦怠是一件好事！

² Adam Bayes、Gabriela Tavella和Gordon Parker（2021年），“职业倦怠的生物学：原因和后果”，世界生物精神病学杂志，22:9, 686–698. DOI: 10.1080/15622975.2021.1907713. <https://doi.org/10.1080/15622975.2021.1907713>

³ Maslach C, Leiter MP. “理解职业倦怠的经历：最近的研究及其对精神科的影响。” 世界精神病学，6月 2016, 15(2), 103–11. DOI: 10.1002/wps.20311. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4911781/1> PMID: 27265691; PMCID: PMC4911781.

⁴ L.A. Kelly等，“护士职业倦怠对组织和岗位流动率的影响”，《护理展望》，1月，2021, 96–102, 1月–2月 2021, 96–102. DOI: <https://doi.org/10.1016/j.outlook.2020.06.008>

如何、何时、为何 您的定位很重要

要点

你是谁（找到自己的定位）很重要，对这句话的理解如下：我们发现，某些群体的受访者与其他受访者有不同的结果，比如更倦怠或生产力更低。

我们还确定了您可以实施的具体做法，以减轻其中的一些负面结果。

介绍

2022年的分析中普遍存在一个现象：工作的安排方式可能有利于一些人的幸福感，但并不适用于所有人。

2022年，我们发现，被认定为弱势群体的人员报告的倦怠程度更高¹。

在本章中，我们将看到这一情况再次发生，并开始讨论为什么弱势群体更有可能经历倦怠，以及哪些因素可以有助于防止这种情况。

此外，困扰许多行业的形势不稳定也引发了对新员工的质疑。组织担心新员工需要很长时间才能提高工作效率，他们正在寻找方法来帮助新员工更快地适应工作。我们也会在这里深入研究。

¹ 2022加速DevOps状态报告。<https://dora.dev/research/2022/dora-report/>

² Sigalit Ronen和Ayala Malach Pines, “工程师倦怠中的性别差异”, 平等机会国际, 2008年11月7日, <https://www.emerald.com/insight/content/doi/10.1108/02610150810916749/full.html>

³ Dalessandro C; Lovell A; Tanner OC; , “2020年后全球工人的种族、边缘化和压力感知”, 《社会学调查》, 2023年8月3日, <https://onlinelibrary.wiley.com/doi/pdf/10.1111/soin.12505>

我们的发现及意义

有些人比其他人更容易倦怠

去年，在报告中我们发现，妇女或自述性别的受访者，以及在任何方面都处于弱势的受访者，比男子和非弱势群体受访者更容易倦怠。这些发现与之前的一项研究一致，该研究表明，与非弱势群体的同龄人相比，弱势群体经历了更大程度的倦怠²和工作压力³。

由于上述原因，我们很有兴趣了解今年我们研究的数据中是否会再次发现倦怠的差异，事实上的结果也确实如此。女性或自述性别的受访者的报告，与男性的受访者相比，他们的倦怠水平高出6%。与非弱势群体的受访者相比，弱势群体受访者报告的倦怠水平高出24%。

某些类型的工作预示着更多的倦怠感

我们发现，某些在工作环境中看似中立甚至有益的方面，如高质量文档或稳定性团队，并不能减少所有人的倦怠（见第4章）。我们不禁在想，这是否可能是因为一些有利于组织的任务，增加了部分个体⁴的倦怠感。

为深度了解受访者的倦怠体验，我们询问了他们所做的工作内容。我们采用两种方法来对此进行衡量，如下：

- 特定的任务，如编码、参会、支持团队成员。
- 工作的特征，如计划外工作、工作可见性、工作繁重程度。

工作的特征尤为关键，因为同一任务的体验可能会因不同的人或不同的时间而不同。例如，某些代码评审可能是计划外的重复劳动，而另外一些其他的代码评审可能在团队中是高度可见，并能展现领导力和技术专长的⁵。

与非弱势群体的受访者相比，弱势群体的受访者报告他们做的重复性工作（即苦活）会多出24%。女性或自述性别的受访者进行的重复性工作比男性受访者多40%。这两组受访者还表示，他们承担了更多的计划外工作，以及既不被同事人所知，也不与他们的专业技能直接匹配的工作。这些调研结果部分解释了这类群体的倦怠状况。

没有提升的任务

Babcock等人在《The No Club》一书中描述了一种特殊的工作类型：“这种没有提升的任务对组织很重要，但它不会促进您的职业发展。”⁶

在这里，“促进”一词的含义比较广泛，例如，薪酬的增长或适应其他工作的竞争力都算在内。

证据表明，女性更多的从事这类工作。Babcock等人讨论了这种工作分配不平等的背后原因。例如，女性更有可能被安排做这类工作，并且当被要求时，她们往往更有可能答应，因为拒绝往往带来更多的社交成本。

此外，他们还描述了这些任务分配不均的后果，例如一些女性：

- 认识到对其职业或收入的负面影响。
- 花更多的时间来完成与其职业相关的工作。

⁴ Linda Babcock、Brenda Peyser、Lise Vesterlund和Laurie Weingart。《The No Club》（纽约：西蒙与舒斯特出版社，2022），17。

⁵ Murphy Hill, E.等人“审查代码中的系统性性别不平等”，Computer Supported Cooperative Work (2023)（即将出版）

<https://research.google/pubs/pub52204>

⁶ Babcock等人，《The No Club》，17。

标准的工作分配流程减少某些受访者的倦怠感

我们询问受访者，他们是否有标准的流程来均衡的分配工作。这才是我们真正意义上的**工作分配**，我们希望这种做法能减轻一些受访者的倦怠感。

结果显示，工作分配确实降低了男性、女性，以及自述性别受访者的倦怠感。随着工作分配水平的提高，职业倦怠的性别差异就消失了。

然而，我们惊讶地发现，工作分配对弱势群体受访者的倦怠水平没有影响。

这一发现引发了进一步的问题和思考：均衡地分配工作的正规流程是否仍然会导致不平等的工作分配？“平等工作”是否考虑了任务的特征，例如被打扰或可见性？除了工作任务本身，对这一类特定群体来说，是否还有更加重要的因素导致了倦怠，那么我们如何缓解导致倦怠的因素带来的影响呢？



培养归属文化

作者：Jeffrey Winer博士

哈佛医学院波士顿儿童医院助理教授主治心理学家

关键发现和一些背景

该报告的一个关键发现是，将自己定义为弱势群体的个体比他们的同事更容易感到倦怠。

报告探讨了导致这方面的一些可能原因。在本节中，我们希望将这些发现与更广泛的研究联系起来，比如：关于归属感的研究，以及相关的组织实践策略。

在某个群体中被认定为弱势时，这表明易受“归属不确定性”⁷的影响，这是一种公认的心理现象（见Geoffrey Cohen博士的“理解和克服归属的不确定性”）⁸。这种不确定性（例如，“我属于这里吗？”，“像我这样的人能在这里成功吗？”）经由人们持续的体验和对这些体验的解读来强化或重新定义。这些与归属不确定性相关的既定过程可能有助于将本报告的发现联系起来，即被认为弱势群体的个人报告了更高水平的职业倦怠感。

组织可以做什么？

对于组织而言，重要的是要明白，多样性、包容性、公平和归属感意味着不同的东西，要实现它们，需要不同的且相互联系以及持续的策略。归属感的实现需要真诚和持久的承诺。

当组织内有员工面临困境时，我们的第一个问题不应该是：“这个人怎么了？”。首先要问的应该是：“为什么有人会有这样的感觉，而我们的组织中的哪些结构性因素造成了这种感觉（例如，哪些因素让这种感觉持续或甚至更糟）？”

一旦识别出问题，应在组织层面进行变革，同时也在个人层面提供支持——一种“兼而有之”的方法。

支持员工去影响组织内的体系能够确保将变革真正地融入到体系中，并使个体参与者更加持久。采用这种系统性和可持续性的思维方式，可以将变革构建到制度中，并超越个体能起到的作用。这种不断演化的品质，是组织真正迈向归属文化的关键所在。

这里的“迈向”是关键。归属感是通过持续的体验和行动建立的，它从来不会被做到或完成，这就是为什么它对工作环境的健康和生产力如此重要。

有许多工具可以帮助组织开展这项工作。例如，2023年美国卫生部长发布的关于孤独主题的报告指出，社会关系和归属感是对抗孤独和倦怠的关键解药⁹。

⁷ Walton GM; Cohen GL; , “简短的社会归属干预改善少数族裔学生的学业和健康结果”, 《科学》(纽约, 纽约), 2023年9月20日查阅, <https://pubmed.ncbi.nlm.nih.gov/21415354/>

⁸ <https://behavioralscientist.org/understanding-and-overcoming-belonging-uncertainty/>

⁹ 卫生部助理部长办公室 (OASH), “新外科医生总顾问对美国孤独和隔离流行病的毁灭性影响发出警报”, HHS.gov, 2023年5月3日, <https://www.hhs.gov/about/news/2023/05/03/new-surgeon-general-advisory-raises-alarm-about-devastating-impact-epidemic-loneliness-isolation-united-states.html>

新员工面临提高生产力的挑战

新加入员工（团队经验<1年）的生产力得分比经验丰富的团队成员（>1年经验）低8%。这是意料之中的，刚开始在一个新的团队中工作是具有挑战的，即使您在这个岗位上有相应的经验，为了适应新环境起步所需的团队特定知识也可能是令人生畏的。此外，融入团队也不仅仅是技能和知识的问题。据称，社交也是一个对生产力至关重要的因素。归属感、贡献被认可以及心理安全都需要时间来培养。

有什么方法可以帮助到新员工更快适应吗？

我们推测，组织可以通过三种方式帮助新员工：

- 提供高质量的工作文档。
- 将人工智能技术纳入工作流程。根据其他研究中已显示，相比经验丰富的员工，这对经验尚浅的新员工会更有帮助。
- 面对面工作，部分人认为，这在入职阶段可能会特别有益。



今年的研究表明，高质量的文档会对生产力产生重大影响（见第4章-文档是基础），人工智能对个人生产力的提升相对较小（见第3章-技术能力对效能的预示）。我们没有理由期望这些影响不会扩展到新员工。我们所看到的数据显示：这些实践确实有利于新员工，但这些实践对于新员工和其他人的帮助几乎差不多。换句话说，新员工不会从这些做法中获得任何特别的好处。

如果您希望帮助新员工和其他人，考虑到高质量文档对生产力有实质和明确的影响，它会是一个很好的切入点。值得注意的是，文档编写良好团队中的新员工（生产力高于平均值1个标准差）的生产力，与文档编写不良的团队中新员工（生产力低于平均值1个标准差）相比，生产力高出130%。

我们将简短地讨论关于“重返办公室的”话题，以避免引起更多争议。我们在第7章（文化是一切的基础）中强调了灵活性的重要性。此外，我们的数据不是实验性的，尽管我们试图控制可能会使我们的结果产生偏差的因素，但工作安排的好处是一个复杂的、在社会学上具备丰富和深度的议题，因此很难得出明确的结论（这点在阅读关于该主题的研究或思考的文章时值得借鉴）。从我们的数据中可以清楚地看到，工作灵活性对生产力有积极的影响。不过这是否适用于您的工作场景适用，暂时还无法确定。

同样的情况也适用于新员工。我们没有看到面对面工作对新员工有特别好处的证据。如果您试图优化生产力，在如何、何处和何时工作方面为新员工提供灵活性，似乎比强迫他们在办公室工作更有效。当然，组织也许不应该仅仅为了生产力而优化。我们还将生产力视为能够带来价值的工作，而不仅仅是单纯的产出（不是代码行），并且这种工作也不会导致员工感到精疲力竭或过度劳累。



结语

感谢您参与今年的研究并阅读本报告。我们一直致力于寻找更好的方法来探索团队工作方式和可实现结果之间的关联。

多年的研究下来最重要收获是：采用持续改进思维和实践的团队能够获得最佳的结果。我们所探索的各种能力可以用作驱动结果的调节器。其中一些调节器是个人可以实践的，而其他一些调节器只能通过整个组织的协调工作才能达到。请确定哪些调节器需要根据您的组织情况进行调整，然后对这些调整进行投入。

改进工作没有终点，但它能为个人、团队和组织创造长久的成功。领导者和从业者都有责任推进这些持续改进的工作。

如何将研究付诸实践？

在您的组织、团队和您为客户提供的服务的环境中探索这些发现。

通过加入DORA社区，分享您的经验，向他人学习，并从持续改进旅程中的其他人那里获得灵感。 <https://dora.community>

致谢

每年，本报告都得到来自世界各地热情贡献者的大力支持。他们对报告发布所包含的调查问题设计、本地化、分析、写作、编辑和排版等所有步骤都付出了巨大努力。作者想感谢所有这些人的投入、指导和同仁情谊。

核心团队

James Brookbank	Eric Maxwell
Kim Castillo	Steve McGhee
Derek DeBellis	Dave Stanke
Nathen Harvey	Kevin Storer
Michelle Irvine	Daniella Villalba
Amanda Lewis	Brenna Washington

编辑

Mandy Grover	Mike Pope
Jay Hauser	Tabitha Smith
Stan McKenzie	Olinda Turner
Anna Eames Mikkawi	

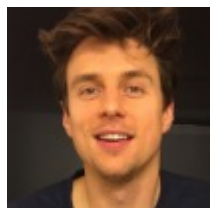
问卷本地化

Daniel Amadei	Mauricio Meléndez
Kuma Arakawa	Jeremie Patonnier
William Bartlett	Miguel Reyes
Antonio Guzmán	Pedro Sousa
Shogo Hamada	Laurent Tardif
Yuki Iwanari	Kimmy Wu
Vincent Jobard	Vinicius Xavier
Gustavo Lapa	Yoshi Yamaguchi

领域顾问和专家

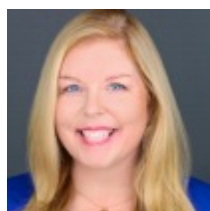
Jared Bhatti	Harini Sampath
Lisa Crispin	Robin Savinar
Rob Edwards	Dustin Smith
Dave Farley	Jess Tsimeris
Steve Fenton	Dr. Laurie Weingart
Dr. Nicole Forsgren	Betsalel (Saul) Williamson
Aaron Gillies	
Dendli Lumma	Dr. Jeffrey Winer
Emerson Murphy-Hill	

作者



Derek DeBellis

Derek是谷歌的定量用户体验研究员，也是DORA的首席研究员。Derek专注于调查研究、日志分析，并找出衡量概念的方法，以证明产品或功能并向人们提供资本价值。Derek发表了关于人类人工智能互动、新冠肺炎发病对戒烟的影响、NLP错误设计、用户体验在隐私讨论中的作用、团队文化以及人工智能与员工幸福感和生产力的关系的文章。他目前的课外研究正在探索信念传播算法和权力传播模型。



Amanda Lewis

Amanda Lewis是DORA.com社区开发主管，也是谷歌云DORA倡导团队的开发人员关系工程师。她在职业生涯中建立了开发人员、运营商、产品经理、项目管理和领导层之间的联系。她曾在开发电子商务平台、内容管理系统、可观察性工具和支持开发人员的团队中工作。这些联系和对话使客户满意，并为业务带来更好的结果。她将自己的经验和同理心带到了帮助团队理解和实施软件交付和可靠性实践的工作中。



Daniella Villalba

Daniella Villalba是谷歌的用户体验研究员。她使用调查研究来了解使开发人员快乐和高效的因素。在谷歌工作之前，Daniella研究了冥想训练的好处，以及影响大学生体验的心理社会因素。她获得了佛罗里达国际大学实验心理学博士学位。



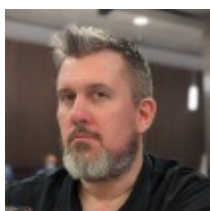
Dave Farley

Dave Farley是Continuous Delivery Ltd.的董事总经理和创始人，《现代软件工程》（Modern Software Engineering）一书的作者，也是畅销书《持续交付》（Continuous Delivery）的合著者。他是《反应宣言》（Reactive Manifesto）的作者之一，并因开源LMAX Disruptor项目获得杜克奖。Dave是连续交付的先驱，思想领袖，CD、DevOps、测试驱动开发（TDD）和软件设计的专家实践者。他在创建高性能团队、塑造成功的组织和创建杰出的软件方面有着长期的记录。Dave致力于与世界各地的软件开发人员分享他的经验和技能，帮助他们改进软件的设计、质量和可靠性。他通过咨询¹、YouTube频道²和培训课程分享自己的专业知识。



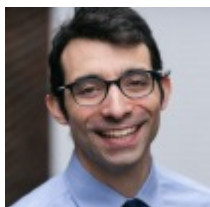
Eric Maxwell

Eric Maxwell领导谷歌的DevOps转型实践，他为世界上最好的公司提供如何通过更快地交付价值来改进的建议。Eric在职业生涯的前半段时间是在担任工程师，将所有事情自动化，并为其他从业者建立共同的标准。Eric共同创建了谷歌的云应用程序现代化计划（CAMP），并且是DORA团队的成员。在谷歌工作之前，Eric在Chef Software工作。



James Brookbank

James Brookbank是谷歌的云解决方案架构师。解决方案架构师通过解决复杂的技术问题和提供专家架构指导来帮助谷歌云客户。在加入谷歌之前，James曾在多家大型企业工作，专注于IT基础设施和金融服务。



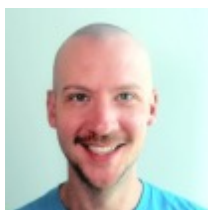
Dr. Jeffrey Winer

Jeffrey P.Winer博士是波士顿儿童医院创伤和社区复原中心（TCRC）的主治心理学家、行为健康系统顾问和心理社会治疗开发人员，也是哈佛医学院的助理教授。与TCRC的同事一起，他的工作主要集中在为具有难民和移民背景的青年和家庭建立、测试、传播和实施文化响应和创伤知情心理社会干预措施。他是《移民和难民青年的心理健康实践：社会生态框架》³一书的合著者。他曾为美国和加拿大的项目提供咨询。他帮助开发或适应的心理社会预防和干预工具目前在世界各地使用。有关更多信息，请参阅<http://www.drjeffwiner.com>。

¹ <https://continuous-delivery.co.uk/engineering-for-software>

² <https://www.youtube.com/@ContinuousDelivery>

³ <https://www.apa.org/pubs/books/4317536>



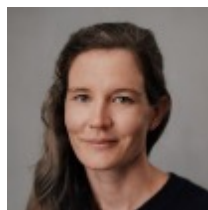
Kevin Storer

Kevin M.Storer是谷歌的用户体验研究员，他领导了一项研究，旨在了解软件开发团队如何与DevOps工具交互，并通过这些工具进行交互。在加入谷歌之前，凯文在加州大学欧文分校获得信息学博士学位。在公共和私营部门，Kevin撰写了关于以人为中心的编程、开发人员体验、信息行为、可访问性和普适计算等主题的高影响力出版物。



Kim Castillo

Kim Castillo是谷歌的用户体验项目经理。Kim领导了DORA背后的跨职能工作，从监督其研究运营到发表本报告。Kim还为谷歌云中的DuetAI进行用户体验研究。在加入谷歌之前，Kim一直从事软件交付，从事技术计划管理和敏捷指导。Kim做的是心理社会研究，专注于她的祖国菲律宾的法外处决、城市贫困发展、社区创伤和恢复力等主题。DORA结合了Kim在软件交付实践、心理安全和组织方面的多种贡献。



Michelle Irvine

Michelle Irvine是谷歌的技术作家，她领导着关于技术文档的影响和制作的研究。在加入谷歌之前，她从事教育出版工作，并担任物理模拟软件的技术作家。Michelle拥有滑铁卢大学物理学学士学位，以及修辞和沟通设计硕士学位。



Nathen Harvey

Nathen Harvey领导DORA Advocacy团队，是Google Cloud的开发人员关系工程经理。Nathen有幸与一些最好的团队和开源社区合作，帮助他们应用DevOps和SRE的原则和实践。过去三年来，他一直是《加速DevOps状态报告》（Accelerate State of DevOps report）的合著者。Nathen还参与编辑并贡献了《97件云工程师应该知道的事情》（97 Things Every Cloud Engineer Should Know）一书。



Steve McGhee

Steve McGee是一位可靠性倡导者，帮助团队了解如何最好地构建和运营世界一流的可靠服务。在此之前，他在谷歌担任了10多年的站点可靠性工程师，学习如何在搜索、YouTube、Android和谷歌云（Google Cloud）中扩展全球系统。他在加州、日本和英国管理多个工程团队。Steve还与一家位于加州的企业共事了一段时间，帮助他们过渡到云。

研究方法

本章概述了本报告是如何从一组最初的想法最终演变为您面前的报告的（具有稍微完整的想法）。我们希望它回答了您关于如何生成该报告的许多问题，并为您提供了一个蓝图，以帮助您开启自己的研究。

步骤1 产生一组我们认为对高效能、技术驱动的组织很重要的成果

这是至关重要的。我们的计划是基于帮助引导人们达到他们所关心的目的。如果我们不知道个人、组织或团队想去哪里，我们的开始就会很糟糕。那么我们如何解决这个问题呢？

我们混合使用定性研究（即询问人们以及他们的团队和他们的组织想要实现什么）、调查、与更广泛的社区互动，以及许多研讨会。我们最终得出如下结果：

- **组织效能。** 组织不仅应该创造收入，还应该为客户和发展社区创造价值。
- **团队效能。** 应用程序或服务团队创造价值、创新和协作的能力。
- **员工幸福感。** 组织或团队采取的策略应该有利于员工减少倦怠，培养令人满意的工作体验，并提高员工创造有价值产出（即生产力）的能力。

我们还听到人们谈论这样的目标：

- **软件交付效能。** 团队可以快速、成功地部署软件。
- **运维效能。** 所提供的软件为用户提供了可靠的体验。



步骤2 假设如何、何时以及为何实现这些结果

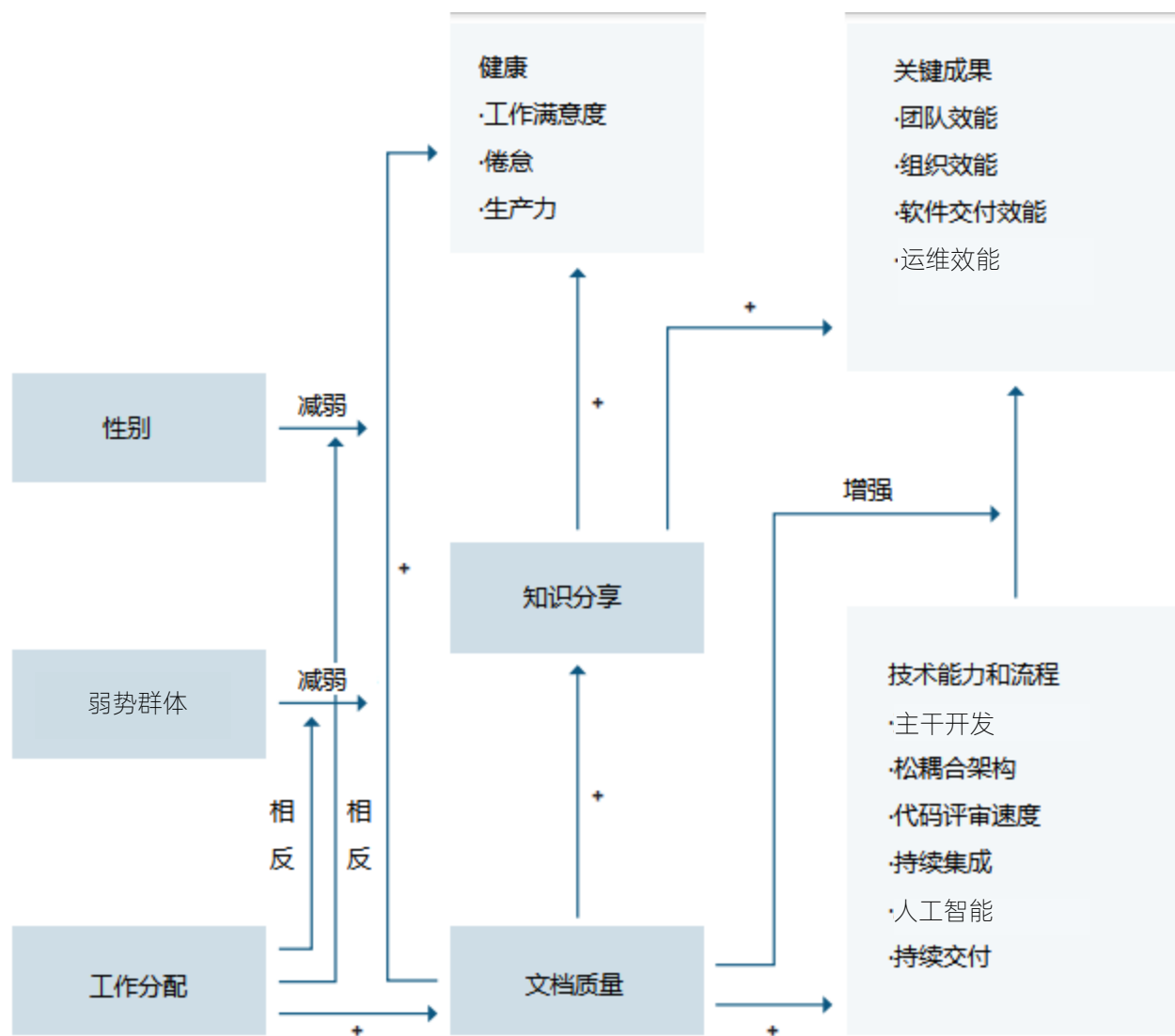
考虑到第1步的结果，我们需要假设可能会让团队到达那种结果的途径。这涉及寻找似乎对结果有可靠影响的因素。我们想说的是，“保持一切相等，x对y有影响”。该信息可以帮助从业者根据数据做出关于要尝试哪种类型的变革的决策。

我们不仅想了解哪些途径有影响，我们还在探索在什么样的条件下这些途径可能或多或少地产生的影响。这相当于询问“何时”和“为谁”。例如，文档质量已被证明可以显著减少工作倦怠——就平均而言。但当我们观察那些报告为弱势群体的受访者时，情况正好相反：文档质量会增加工作倦怠。了解这些影响的条件是至关重要的，因为任何团队或任何人都很难达到平均水平。

此外，我们假设了解释为什么或如何的机制。今年，根据我们去年看到的结果以及关于该主题的大量文献，我们假设那些弱势群体的人经历了更多的倦怠。人们问的第一个问题是“为什么会发生这种情况？”为了回答这个问题，我们假设了要测试的潜在机制。例如，弱势群体的受访者可能会经历更多的倦怠，因为他们承担了（或被分配了）更辛苦的工作。

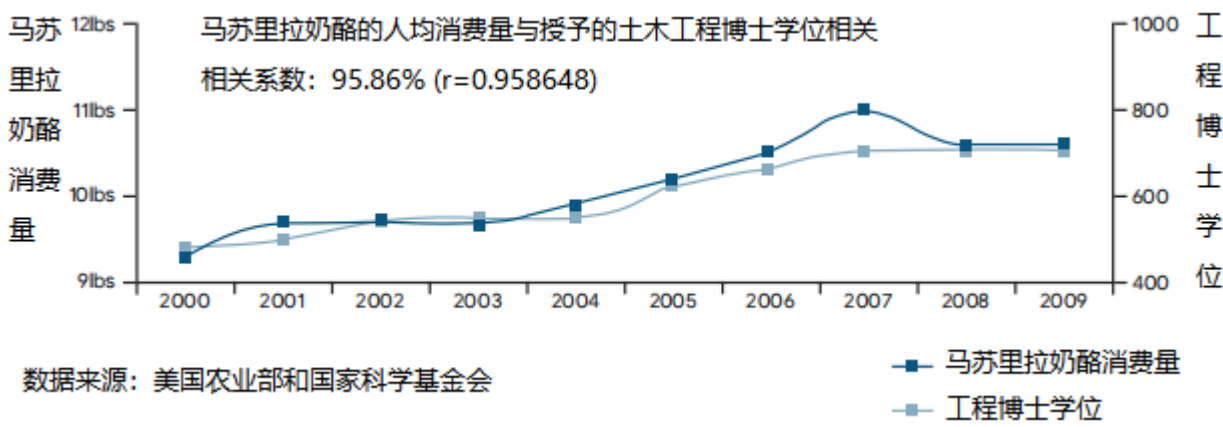


绘制这些假设间的关系，以便我们可以构建调查问卷并进行分析。下面是文档章节的假设模型示例，包括效果、条件和机制：



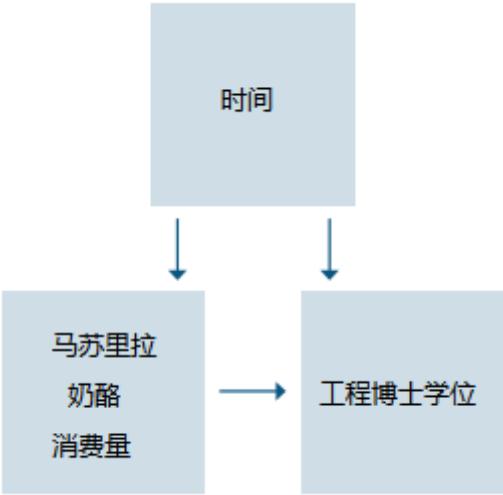
步骤3 假设潜在的意外

如果您曾经讨论过数据，您可能会遇到虚假的相关性¹。您可能熟悉显示许多虚假相关性的网站，例如以下示例：



工程博士学位和马苏里拉奶酪的消费之间不太可能有任何因果关系。然而，在这种关系背后隐藏着一个令人困惑的因素：时间。如果马苏里拉奶酪消费量和工程学博士学位在同一时间段内都呈积极趋势，那么它们很可能具有正相关性。

在模型中包含时间，或减少数据，可能会使关系无效。我们可以如下所示绘制模型：



¹ <https://www.tvlervigen.com/spurious-correlations>

如果我们不考虑时间（第三个变量），数据可能会显示马苏里拉奶酪消费量和工程学博士学位之间的虚假关系。

有一些工具可以帮助研究人员做到这一点，如 Dagitty(<https://dagitty.net/dags.html>)。该工具允许我们指定因果模型。为了帮助我们正确估计X对Y的影响，它告诉我们模型的含义，我们需要解释的内容，以及我们不应该解释的内容。像 Dagitty 这样的工具可以得出这样的结论：相关性可能并不意味着因果关系，但它最肯定地暗示了人们对因果关系的看法。

我们不可能捕捉到影响研究人员评估的所有因素——想想长达50小时的调查和全部知识。尽管如此，我们仍尽最大努力控制偏差路径，以便我们可以为您准确估计各种活动、技术和结构，以及您关心的结果的影响。最后，许多从业者想知道哪些因素将影响这些关键成果。未能考虑偏差的模型将无法为从业者提供他们需要的指导。我们不想告诉别人，马苏里拉奶酪提高了软件交付效能，而且犯这个错误比您想象的要容易得多。



步骤4 编制调查问卷

开展调查有三个方面：可操作性、调查体验和本地化。

可操作性

我们希望度量能够充分捕获我们感兴趣的概念，并且能够可靠地做到这一点。将抽象概念转化为可量化的东西是可操作性的艺术。这些衡量是所有分析的基础。如果我们的衡量没有给我们明确的信号，我们怎么能相信其余的分析？我们如何衡量像生产力这样难以捉摸的概念？工作倦怠或运维效能又该如何呢？

首先，我们查阅文献，看看是否已经存在成功的衡量。如果我们可以在调查中使用以前验证过的衡量，我们就可以在调查与围绕该问题积累的所有文献之间架起一座桥梁。我们正在使用Westrum的组织文化分类学，这是我们重用以前验证过的衡量的一个例子。

然而，许多概念在我们研究的领域之前还没有得到验证。在这种情况下，我们正在进行定性研究，以弄清人们如何理解这个概念，我们正在查阅关于这个概念的复杂性的更多哲学文献。

调查体验

我们希望调查是可理解的、简单的、不超过所必须时间的，并且可以得到广泛的访问。考虑到我们想要问的所有问题，考虑到回答这些问题所需的技术理解，以及特定实践的术语变化，这些都是困难的目标。我们进行远程、快速的评估，以确保调查的执行高于某些阈值。这需要进行多次迭代。

本地化

世界各地的人们每年都对我们的调查做出了回应。今年，我们通过将调查本地化为英语、西班牙语、法语、葡萄牙语和日语。这是一项基础工作，由DORA社区的一些不可思议的成员领导。全世界的谷歌人都为这一努力做出了贡献，同时也是该领域的合作伙伴，这多亏了Zenika(<https://www.zenika.com>)用于我们的法语本地化。我们希望扩大这些努力，使本调查成为真正的跨文化的调查。

步骤5 收集问卷反馈

我们使用多个渠道进行参与者的招募。这些通道分为两类：自然流量和定向来源。

自然流量是使用我们所掌握的所有社会手段，让人们知道我们希望他们进行一项调查。我们创建博客帖子、我们使用电子邮件活动、我们在社交媒体上发帖，并要求社区中的人也这样做（即滚雪球采样）。

我们使用定向来源来补充自然流量。在这里，我们试图招募在更广泛的技术社区中的处于弱势群体人，我们试图从某些行业和组织类型中获得足够的回应。简而言之，我们对招募做的一些控制，控制我们没有使用自然流量的地方。定向来源还允许我们更容易地确保获得足够的参与者，因为我们永远不知道自然流量是否会产生帮助我们所做的分析类型所需的响应。

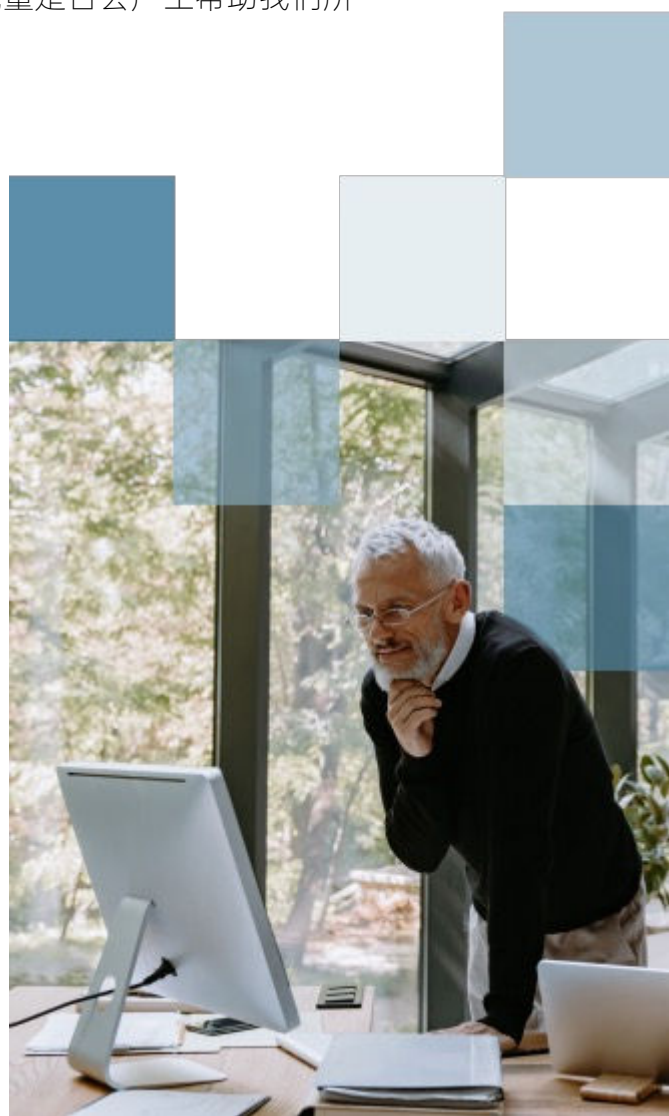
步骤6 分析数据

分析的核心有三个步骤：数据清洗、衡量验证和模型评估。

数据清洗

数据清洗的目标是提高信噪比。由于各种原因，有些问卷反馈会是噪声。一些回答的模式可能表明此人分心、在调查中过快或没有真诚回答。这在我们的数据中非常罕见，但它确实有所发生。不真实表达个人实际经历的情况，这种答卷会被排除在我们的分析之外。

消除噪声的挑战是确保我们不会消除有效信号，尤其是对于以有偏见的方式或以验证我们假设的方式存在的噪声。例如，如果我们得出结论，没有人能在某个值上高而同时在另一个值上低，我们可能会排除那个受访者——这会导致数据与我们的认知更加一致，并增加我们的假设成功的几率。



衡量验证

在本报告的开头，我们讨论了我们试图衡量的概念。我们可以参与很多不同的文字游戏，但有一种观点认为，概念的这种衡量标准被称为变量²。这些变量是模型的组成部分，也是我们研究中包含的元素。有两种广泛的方法来分析这些衡量的有效性：内部的和外部的。

为了理解衡量的内部有效性，我们观察我们认为表明概念存在的内容。例如，高质量的文档是指人们用来解决问题的文档。

我们的大多数变量由多个指标组成，因为感兴趣的似乎是多方面的。为了理解变量的多面性，我们测试了一些条目。如果它们很好地结合在一起（也就是说，它们有很高的公共方差），我们会假设它们背后有一些共性的东西——比如利益的概念。

例如幸福。幸福是多方面的，我们希望有人在快乐的时候有某种感觉，有某种行为，有某种思考。我们认为幸福是某种感觉、思想和行为模式的基础。因此，当幸福存在时，我们期望某些类型的感觉、想法和行动会一起出现。然后，我们会询问有关这些感受、想法和行动的问题。

我们将使用验证性因素分析来测试他们是否真的出现在一起。

在今年，我们使用LavanR软件包进行了此分析³。Lavan返回了各种各样的统计数据，帮助我们理解问卷结构是否真的代表了人们回答问题的方式。如果某个概念的指标不一致，那么这个概念可能需要修改或删除，因为很明显，我们还没有找到衡量这个概念的可靠方法。

结构的外部有效性就是观察这个结构如何融入世界。我们可能期望一个构造与其他结构有一定的关系。有时，我们可能会期望两种结构具有负面关系，比如幸福和悲伤。如果我们的幸福感与悲伤感呈正相关，我们可能会质疑我们的衡量标准或理论。同样，我们可能期望两个结构具有积极的关系，但不是牢固的关系。生产力和工作满意度可能呈正相关，但我们认为它们并不相同。如果相关性太高，我们可能会说，看起来我们在测量同样的东西。这意味着我们的测量没有经过足够的校准，无法发现这两个概念之间的差异，或者我们假设的差异实际上并不存在。

² Moore, Will H.和David A.Siegel。政治和社会研究的数学课程。普林斯顿大学出版社，2013年。

³ 罗塞尔，Y。“lavan:结构方程建模的R包”，《统计软件杂志》，48（2），2012年。1-36。<https://doi.org/10.18637/jss.v048.i02>

模型评估

在第2步和第3步中，我们构建了假设模型。在步骤6为我们提供了干净的数据之后，我们可以看到这些模型与数据的拟合程度。今年，我们采用了贝叶斯方法，以此了解给定数据的各种假设的可信度，而不是给定零假设的数据的可能性（即不存在任何影响）。我们在R中使用的主要工具是blavaan⁴和rstanarm⁵。我们可以测试一个影响的概率，包括是固定的或偶发的，而不仅仅是少数的。为了评估一个模型，我们追求简约。这相当于从一个非常简单的模型开始，增加复杂性，直到复杂性不再合理。例如，我们预测组织效能是软件交付效能和运维效能之间相互作用的产物。

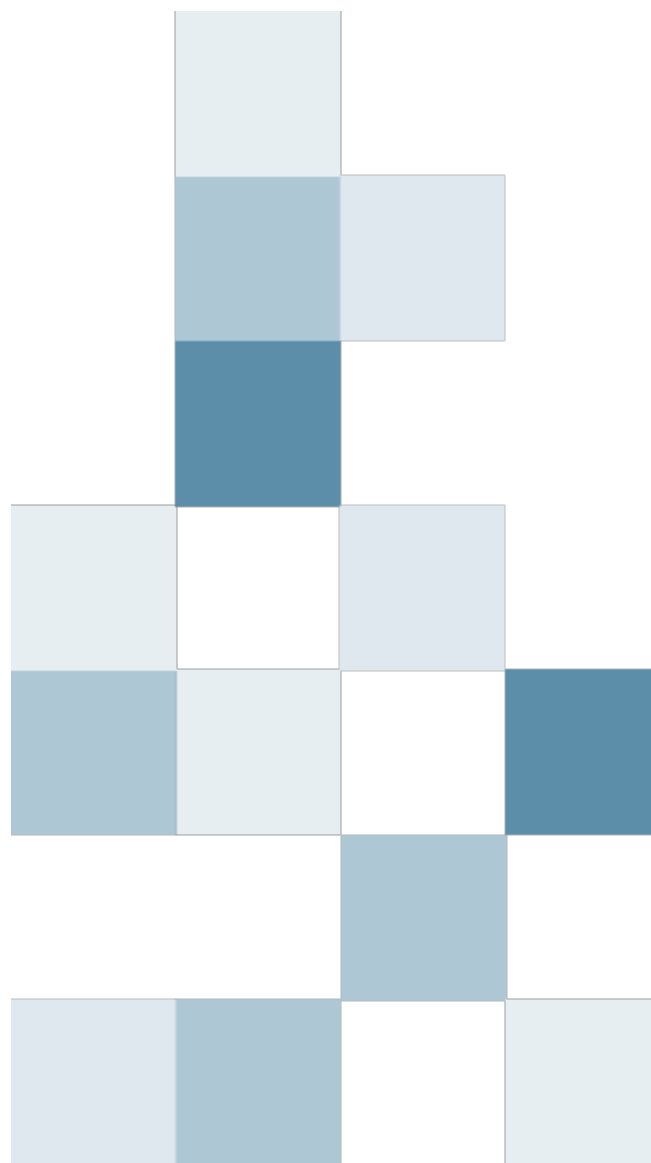
我们不包括其他影响的简单化模型：

组织效能~软件交付效能+运维效能

我们的第二个模型添加了交互项：

组织效能~软件交付效能+运维效能+软件交付效能x运维效能

根据“回归和其他故事”⁶和“统计反思”⁷中的建议，我们使用留一交叉验证（LOOCV）和渡边-赤池广泛适用的信息量准则⁸来确定是否需要额外的复杂性。



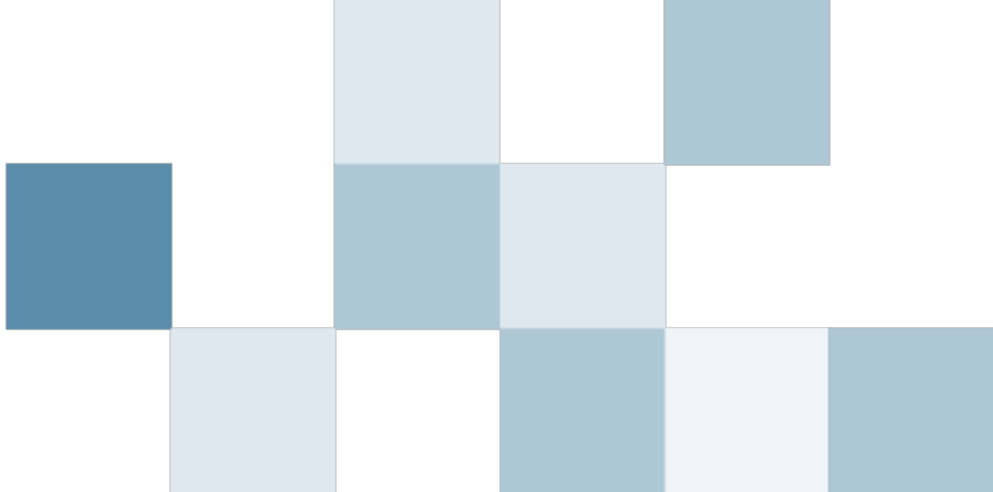
⁴Merkle, Edgar C., and Yves Rosseel. "blavaan:通过参数扩展的贝叶斯结构方程模型", arXiv预印本, 2015. <https://doi.org/10.48550/arXiv.1511.05604>

⁵Goodrich, Ben, Jonah Gabry, Imad Ali 和 Sam Brilleman "rstanarm:通过Stan应用贝叶斯回归建模。"R软件包版本2, 编号1 (2020)。

⁶Gelman, Andrew, Jennifer Hill 和 Aki Vehtari. 回归与其他故事 (剑桥大学出版社, 2020)。

⁷McElreath, Richard. 统计学反思: 贝叶斯课程, 以R和Stan为例。Chapman和Hall/CRC, 2018。

⁸Vehtari, Aki, Andrew Gelman和Jonah Gabry. "使用留一交叉验证和WAIC的实用贝叶斯模型评估", 《统计与计算》, 2017年7月27日。1413-1432. <https://doi.org/10.1007/s11222-016-696-4>



步骤7 报告调查结果

我们作为一个整体团队评审了这些结果。今年，我们在科罗拉多州博尔德市一起度过了几天，将数据与主题专家的经验进行了综合。我们对报告的每一章都这样做，一个假设一个假设地审查。数据解释总是有编造、猜测、无根据臆想和反转的风险。让多个不同背景的人呆在一个鼓励提问、分歧沟通、独特观点和好奇心⁹的房间里，可以减轻这些风险。

有了结果后，报告的作者们回到了各自所在的世界角落开始撰写。在整个撰写过程中，咨询了编辑和主题专家。拥有这些观点对于帮助我们交流想法至关重要。负责分析这些数据的人有责任确保我们所说的一切都与数据所说的相符。

我们才华横溢的设计合作伙伴 BrightCarbon¹⁰ 将这些章节编撰在一起，形成了一个融会贯通的设计。

步骤8 与社区一起整合调查结果

我们依靠社区的参与来想出利用和解释这些发现的方法。我们试图在我们的建议中做到具体，但最终，团队也可以根据我们发现的结果尝试无数的实现方法。例如，基于我们衡量的结果，松耦合架构似乎是一种有益的实践。但是，建立一个松耦合的体系架构肯定不止一种方法。作为一个社区，制定和分享方法是不断改进的唯一途径。我们的世界版本是对您、您的团队和您的组织运作的领域和背景的解释和抽象。

要参与DORA的全球实践社区，请访问DORA社区网站 (<https://dora.community>)。

⁹ Stasser, G.和Titus, W. (1985)。“集体决策中非共享信息的汇集：讨论中的偏见信息抽样”，《个性与社会心理学杂志》，48（6），1985。1467-1478。
<https://doi.org/10.1037/0022-3514.48.6.1467>

¹⁰ <https://www.brightcarbon.com/>

人员与企业统计

调查对象

近十年来，DORA研究项目一直在研究高效能、技术驱动型组织的能力、实践和度量标准。我们听取了来自各种规模和不同行业组织中，超过36,000名专业人士的意见。感谢分享您的见解！今年，来自世界各地各行各业的近3,000名从业人员分享了他们的经验，帮助我们更好地了解推动高效能、技术驱动型组织的因素。

今年的人员统计和公司概况问题利用了Stack Overflow的研究成果。在2022年的Stack Overflow开发者调查中，有超过70,000名受访者参与。这项调查并未涵盖到每一位技术从业者，原因众多，但它几乎可以视为对开发者群体进行普查。有了该调查提供的人口统计信息，我们可以确定我们数据中的回应偏差，并理解我们希望将研究结果推广多远。

那些在Stack Overflow开发者调查中收集到的数据以及所提出的人员统计和公司概况问题是经过精心设计的，并且值得借鉴。与Stack Overflow开发者调查相比，我们的样本中包含比例更高的女性、失能参与者以及在大型组织中工作的参与者。就种族和族裔而言，我们的样本与Stack Overflow的样本相似。

与2022年相比，今年的调查中自然流量回应者的数量增加了3.6倍。

对比2022年，今年的调查
响应者增加了

3.6倍

¹ <https://survey.stackoverflow.co/2022#overview>

人员统计

性别

与2022年相比，今年的样本中女性回应者的比例较小（12%对比18%）。

性别	回应者中百分比%
不愿说	3%
不在此列	2%
女性	12%
男性	81%

失能者

我们按照华盛顿集团短问卷²的指导，确定了失能情况的六个维度。这是我们第五年询问有关失能情况的问题。2023年失能者比例从2022年的11%下降到6%。

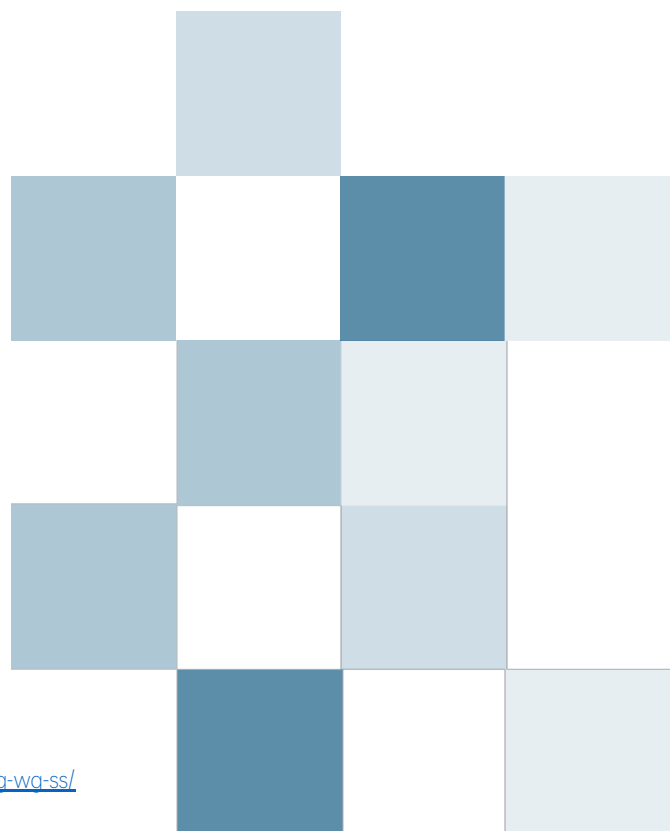
失能情况	回应者中百分比%
没有失能	87%
失能	6%
不愿说 / 没有响应	7%

² <https://www.washingtongroup-disability.com/question-sets/wg-short-set-on-functioning-wg-ss/>

弱势群体

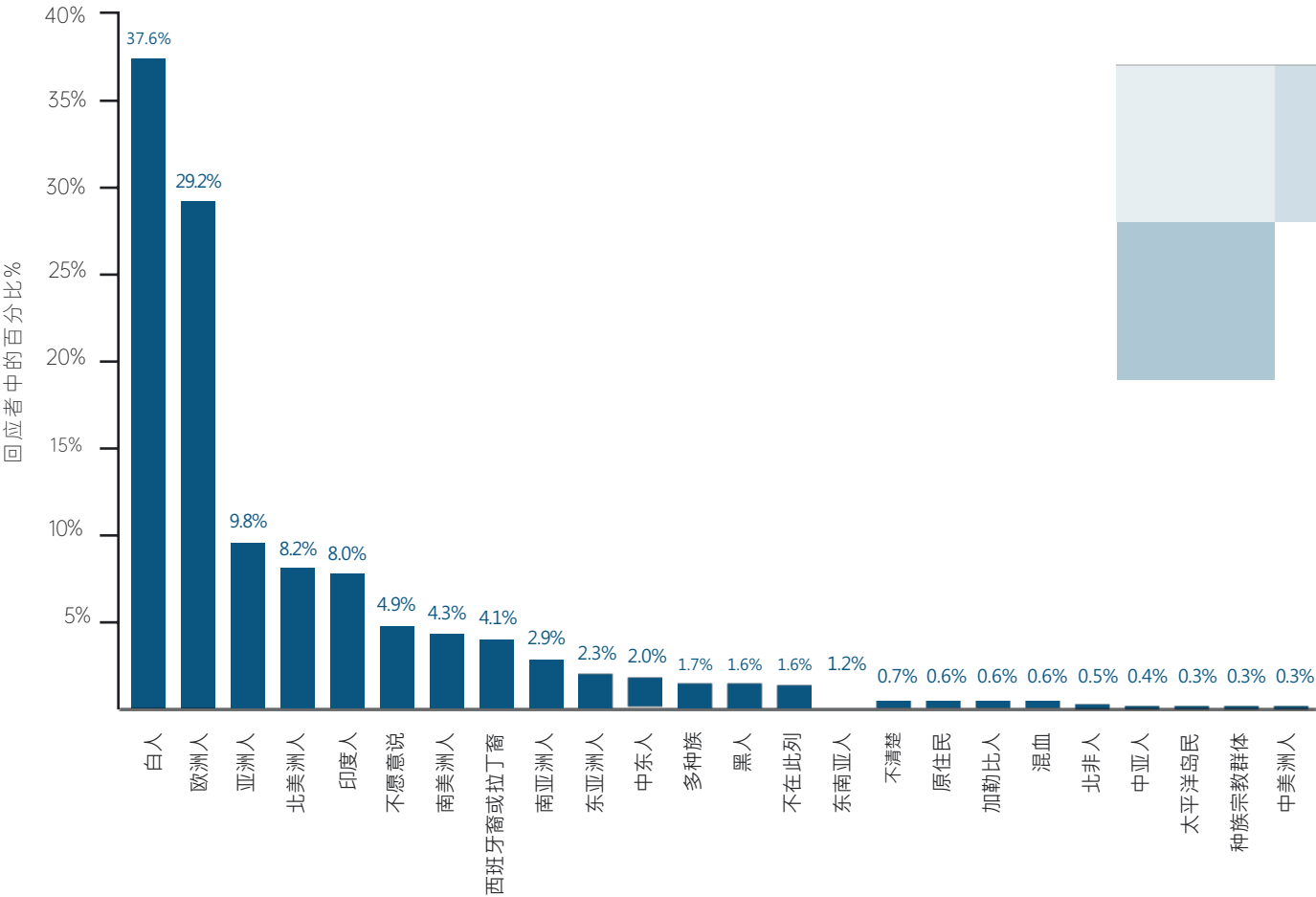
认定自己属于弱势群体可以涉及种族、性别或其他特征。这是我们第六年询问有关弱势群体情况的问题。自2022年的19%以来，自认为属于弱势群体的人数在2023年略有下降，降至15%。

弱势群体	回应者中百分比%
不是	77%
是	15%
不愿说	7%



种族和民族

我们采用了2022年 Stack Overflow 开发者调查中的问题³。正如之前所提到的，我们的样本组与之类似，但有一个显著的偏差：我们欧洲人的比例较低。



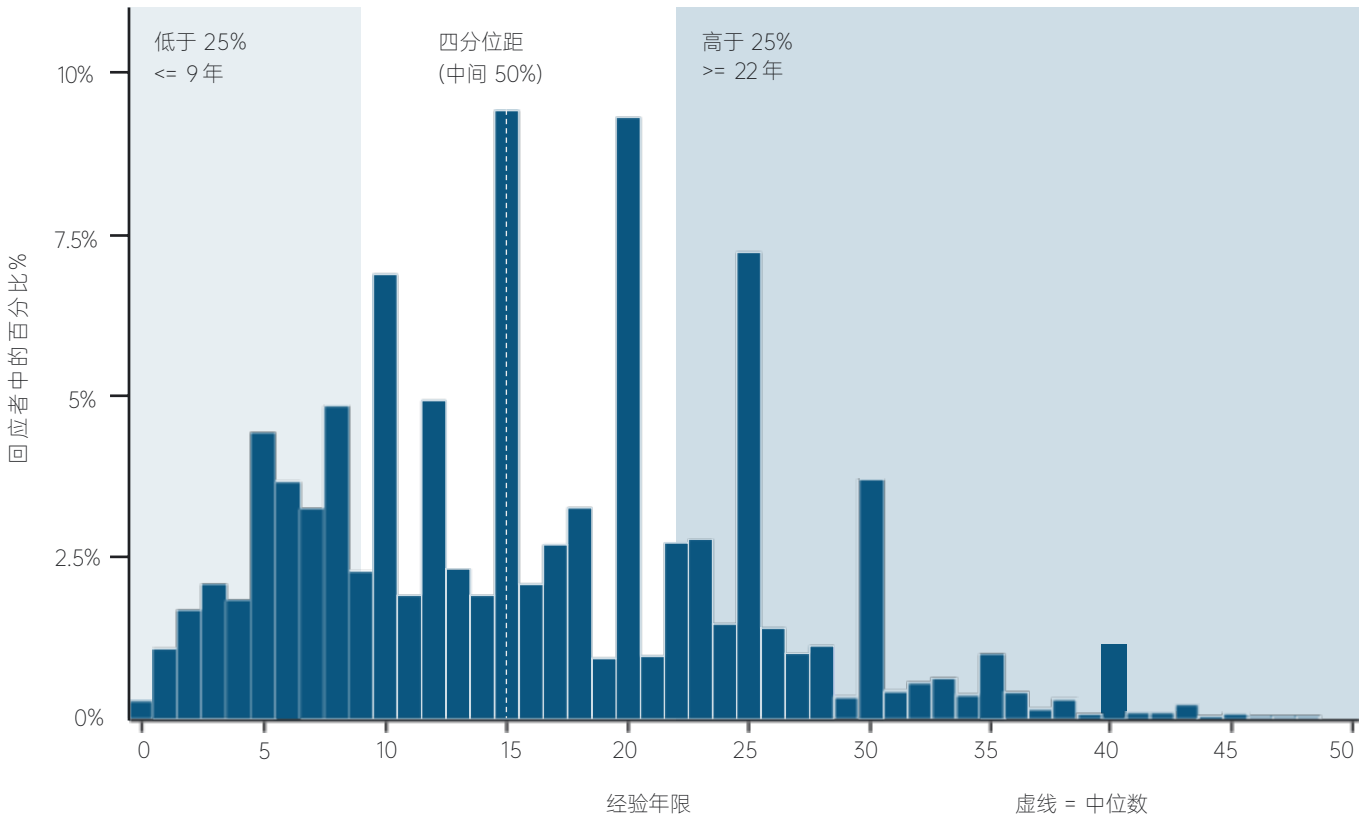
³ <https://survey.stackoverflow.co/2022#overview>

经验年限

工作经验

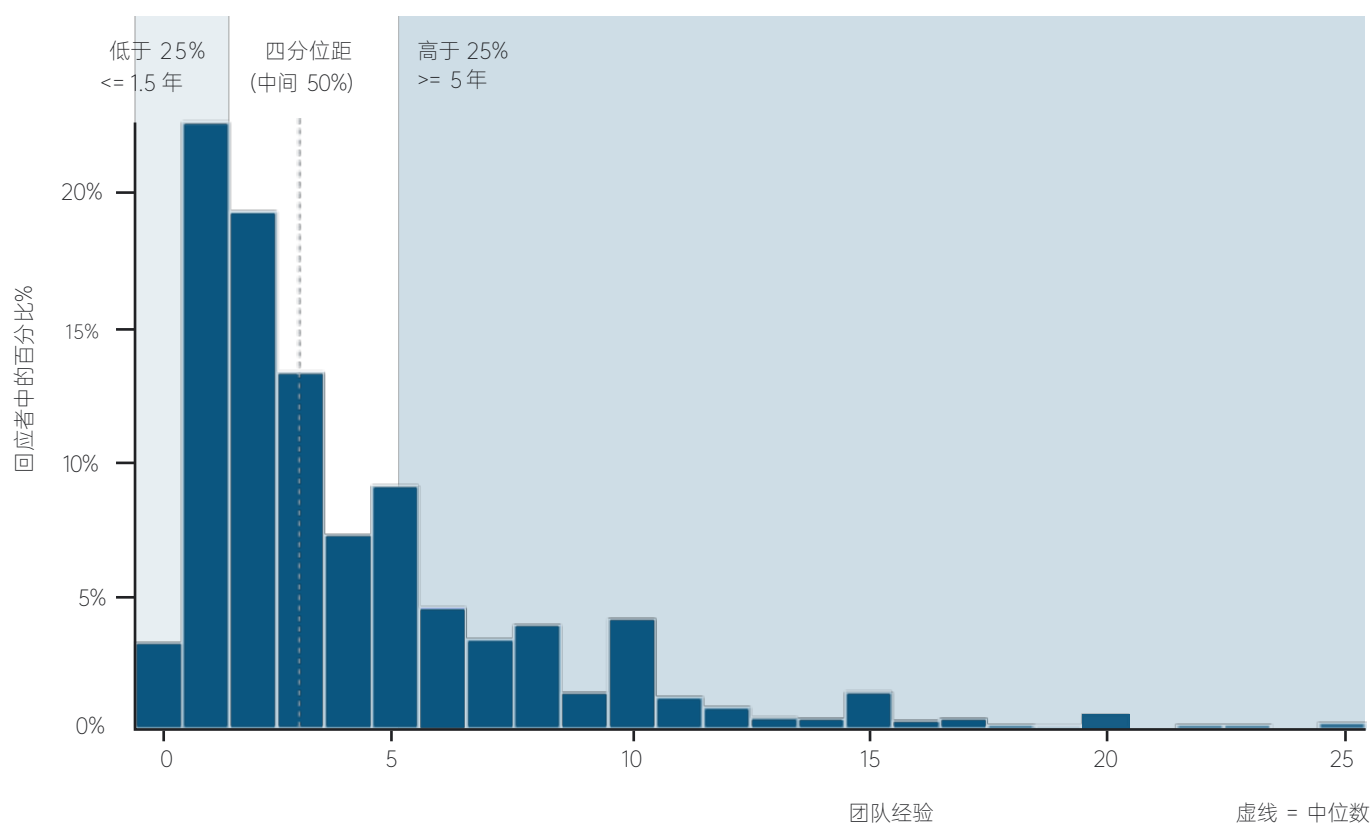
我们希望了解一个人工作多久了。我们问了这个问题：“您有多少年的工作经验？”总的来说，我们正在收集一组相当有经验的从业者的数据。也就是说，我们的调查对象中有50%的人拥有15年或更多年的经验，25%的受访者有超过22年的经验，而25%的受访者的经验不到9年。

回过头来看，一个人认为什么算作“工作”并不明显。



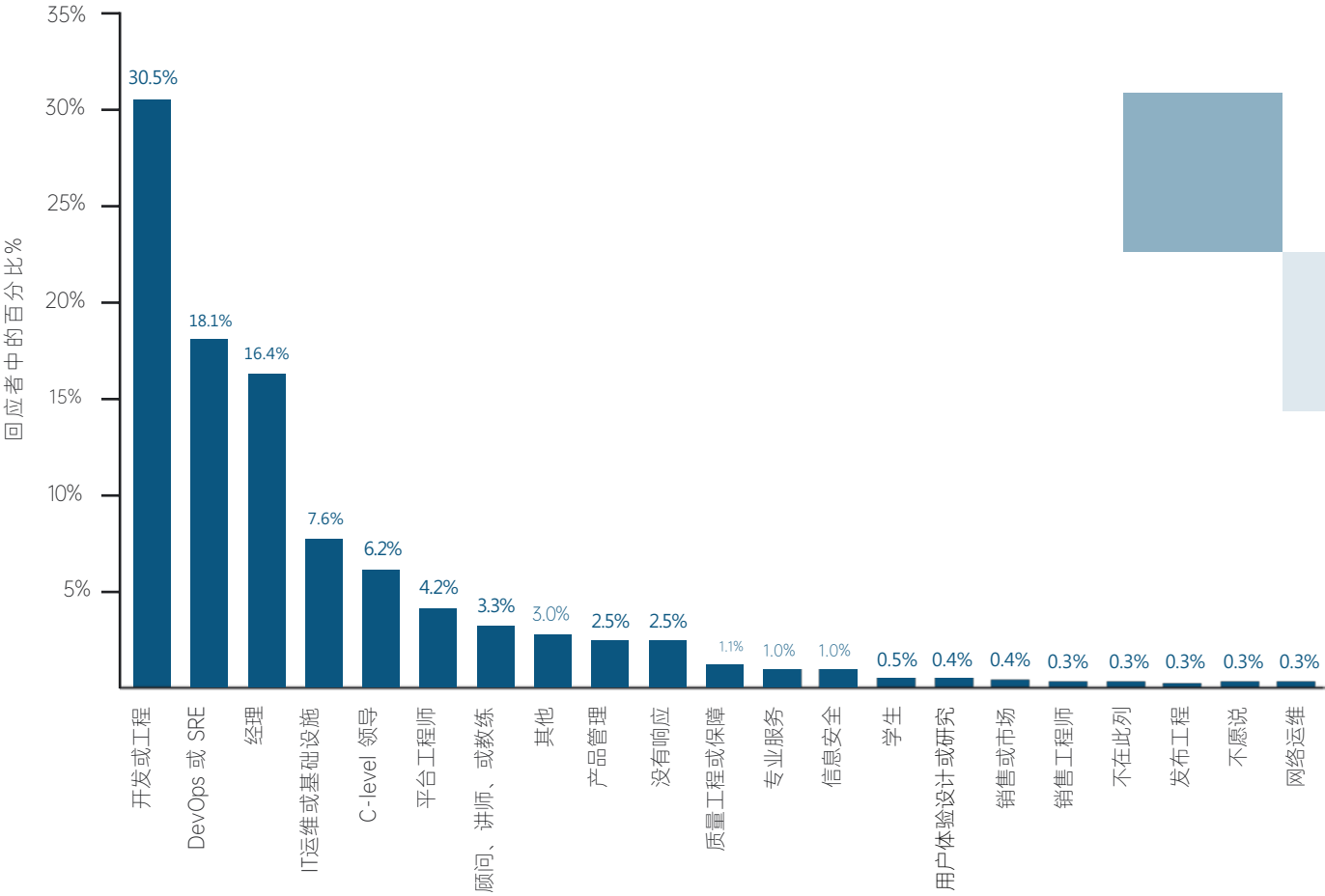
团队经验

有时候加入一个新团队感觉就像重新开始一样，有点回到原点。为了探索这个问题，我们问了这样一个问题：“您目前所在的团队工作了多少年？”尽管工作经验似乎很丰富，但很多受访者对他们所在的团队还是新手。50%的受访者在团队工作不到3年。25%的受访者在团队工作不到或等于1.5年。只有25%的受访者在团队里工作了5年或更长时间。这是否反映了我们受访者中的持续改进的心态？这是否反映了经济的波动和不稳定性？



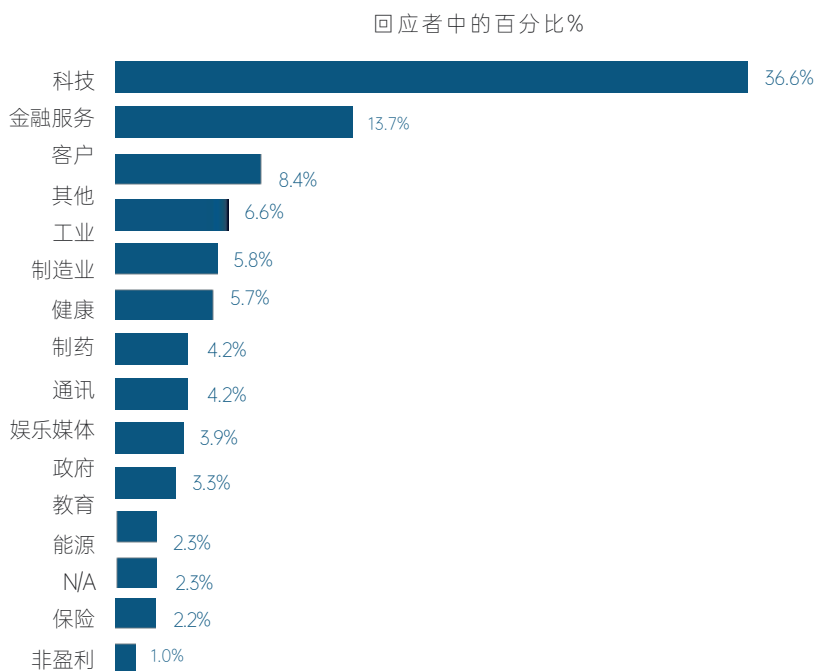
岗位角色

72%的受访者来自于开发或工程团队（30%），DevOps或SRE团队（18%），IT运维或基础设施团队（8%），或者是管理人员（16%）。在2022年，这些角色的受访者占总体的85%。这四个角色的受访者减少，意味着我们能够接触到更多不同角色的人。IT运维或基础设施团队的比例（8%）回到了2021年的水平（9%），而在2022年有所下降（19%）。



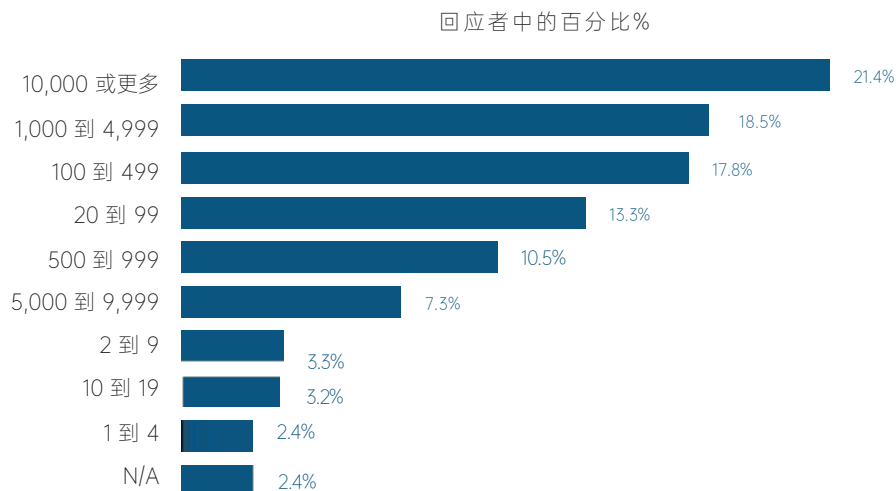
企业统计

行业



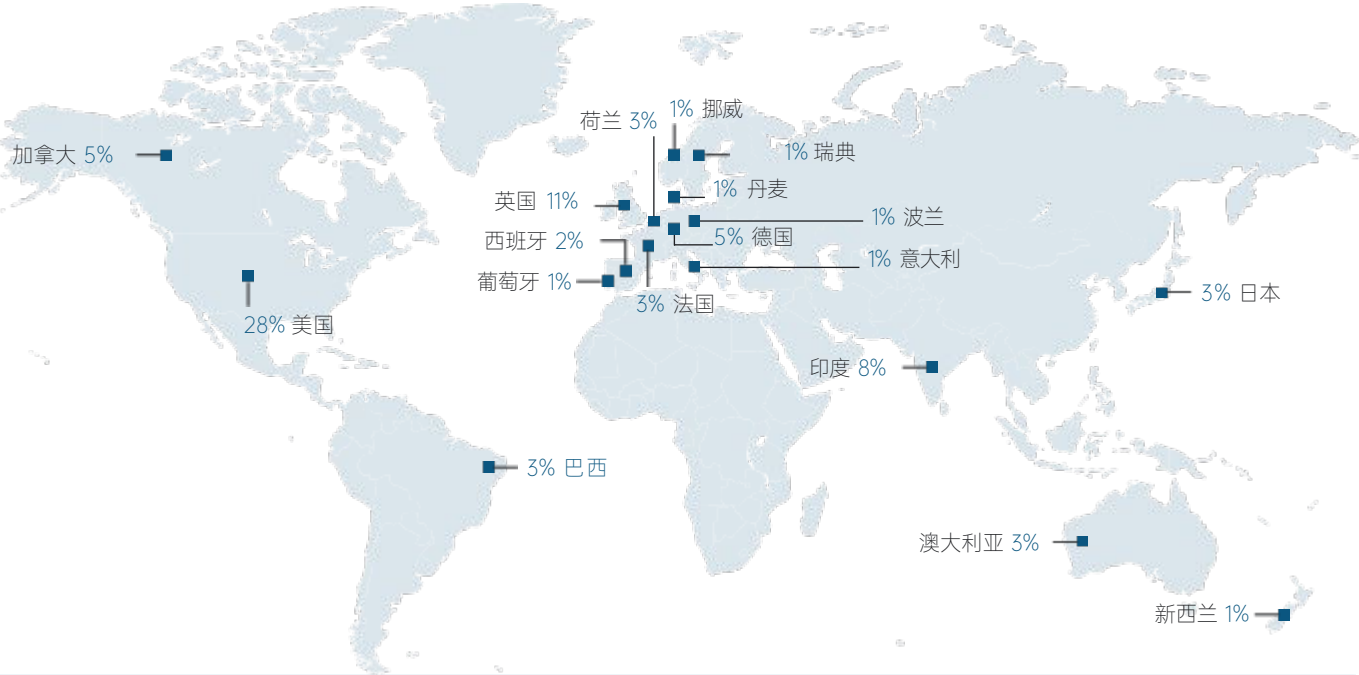
雇员数量

在您的组织里有多少雇员?



国家

我们总是非常激动地看到来自世界各地的人们参与这项调查。谢谢大家！



国家					
美国	丹麦	立陶宛	突尼斯	孟加拉国	危地马拉
英国	瑞士	泰国	乌拉圭	多米尼加共和国	洪都拉斯
印度	奥地利	匈牙利	阿富汗	加纳	拉脱维亚
加拿大	肯尼亚	以色列	阿尔及利亚	香港特别行政区	黎巴嫩
德国	南非	越南	埃及	哈萨克斯坦	卢森堡
澳大利亚	阿根廷	阿拉伯联合酋长国	爱沙尼亚	缅甸	马尔代夫
巴西	捷克	保加利亚	冰岛	沙特阿拉伯	马耳他
不适用	比利时	克罗地亚	伊朗	索马里	毛里求斯
荷兰	哥伦比亚	厄瓜多尔	尼日利亚	苏丹	蒙古
日本	芬兰	印度尼西亚	秘鲁	乌干达	摩洛哥
法国	爱尔兰	菲力宾	斯洛伐克	阿尔巴尼亚	尼泊尔
西班牙	中国	亚美尼亚	斯洛文尼亚	巴哈马	卡塔尔
瑞典	罗马尼亚	格鲁吉亚	韩国	白俄罗斯	
意大利	新加坡	希腊	斯里兰卡	玻利维亚	前南斯拉夫的马其顿共和国
新西兰	墨西哥	马来西亚	安道尔	柬埔寨	特立尼达和多巴哥
波兰	土耳其	巴基斯坦	安哥拉	哥斯达黎加	坦桑尼亚
挪威	乌克兰	俄罗斯联邦	安提瓜与巴布达	吉布提	联合共和国
葡萄牙	智利	塞尔维亚	巴林	萨尔瓦多	津巴布韦

工作安排

雇佣状态

88%的受访者是全职雇员，10%的受访者是合同工。一些合同工报告的经历与全职雇员有很大的不同。

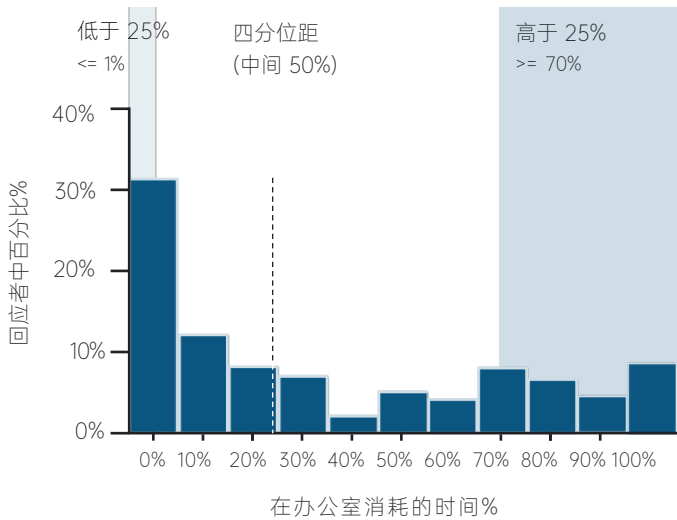
合同类型	回应者中百分比%
全职雇员	88%
全职合同工	8%
兼职雇员	2%
兼职合同工	2%

*在您的主要工作中，那种雇佣类型最能描述您的状态？

不同的经历可能源于他们如何融入团队。一些合同工报告称自己被融入到他们所在的团队中。这意味着他们每天与团队成员紧密合作，认为与全职员工之间的差异微乎其微。70%的合同工受访者要么非常同意，要么同意这样的说法，即他们被融入到他们的团队中。

地点

今年的回应模式表明，尽管推动回到办公室工作，但对许多员工来说，居家办公仍然是现实。近33%的受访者几乎完全在家工作（在办公室的时间不到5%）。63%的受访者在家工作的时间超过在办公室工作的时间。对于其余的受访者来说，混合工作可能是最常见的安排。这是因为75%的受访者在办公室的时间不到70%。对于办公室的强烈依赖并不多见，只有9%的受访者在办公室的时间超过95%。

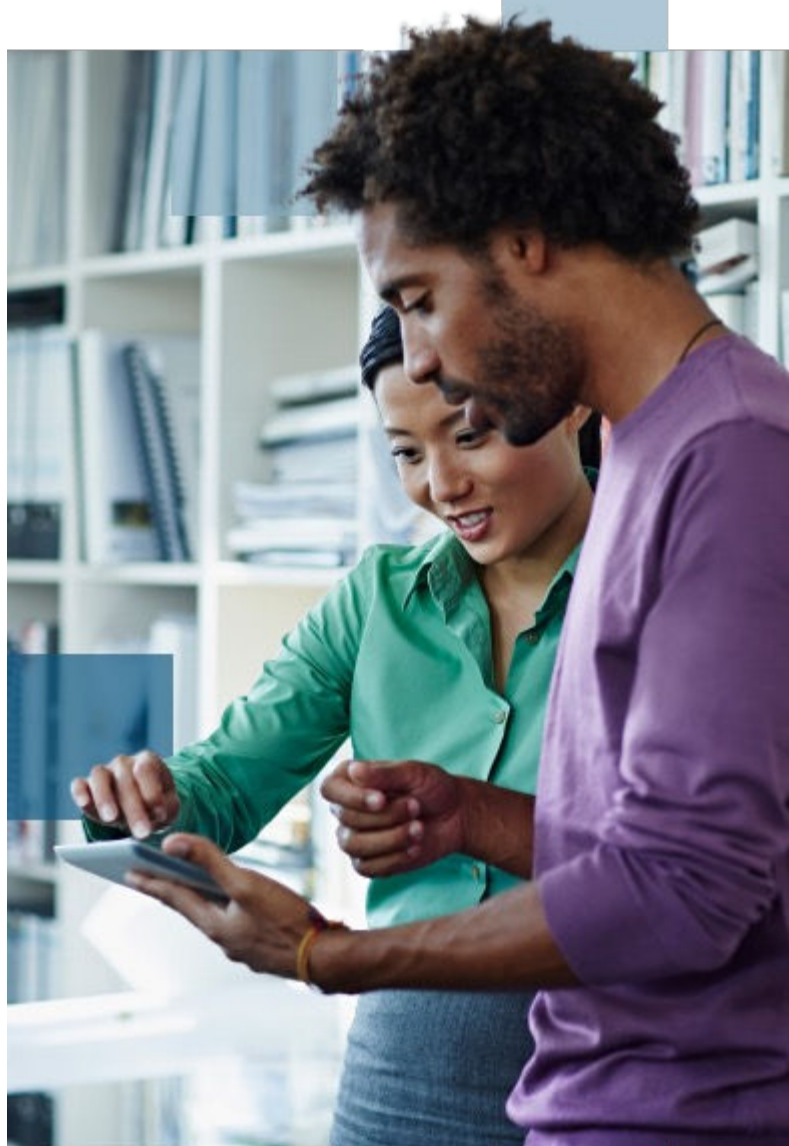


模型

介绍

就传统而言，我们创建了一个巨大的模型。但今年我们决定将其拆分为多个模型，原因如下：

- 巨大的模型可能会变得难以操作。每增加一个变量都会改变模型的运作方式。这可能导致估计不准确，并且使得找出变化原因变得困难。
- 今年我们逐节创建了我们的假设。因此，为每个部分创建一个模型是有意义的。
- 对于估计 X 对 Y 影响的巨大模型的好处并不明显。为了理解 X 对 Y 的影响，我们使用有向无环图来帮助理解应该和不应该在模型中包含哪些协变量。
- 我们今年解决的假设数量对读者来说，将这些假设整合到一个巨大的模型中将会非常难以理解。想象将下面所有的可视化结果合并到一个可视化中。



如何阅读这些图表？

一旦您学会了如何阅读这些图表，您会发现它们是传达大量信息的有效工具。

变量Variable

变量是一个我们试图衡量的概念（例如文档质量）。

变量分类

- 变量
- 变量
- 变量
- 变量

"变量类别"只是表示我们将其视为一个类别，但与分析无关。也就是说，我们并没有对其进行统计评估以确定它是否是一个高阶构建。

+

加号，仅仅表示增加，并不意味着好。

-

减号，仅仅表示减少，并不意味着坏。

假设，而数据并未成立。

中介路径的一部分，我们明确分析的。

警告：这些模型是一般性总结！

为了便于阅读，我们将一些变量进行分类。这种分类策略使得从一个变量类别到另一个变量类别的箭头表示了一般结果的模式，但这并不适用于类别中的每个变量。例如，知识共享对大多数关键结果有积极影响。因此，我们会在知识共享和关键结果变量类别之间绘制一个带有加号（+）的箭头。然而，知识共享对软件交付效能没有积极影响。要了解详细信息，请访问相关章节。

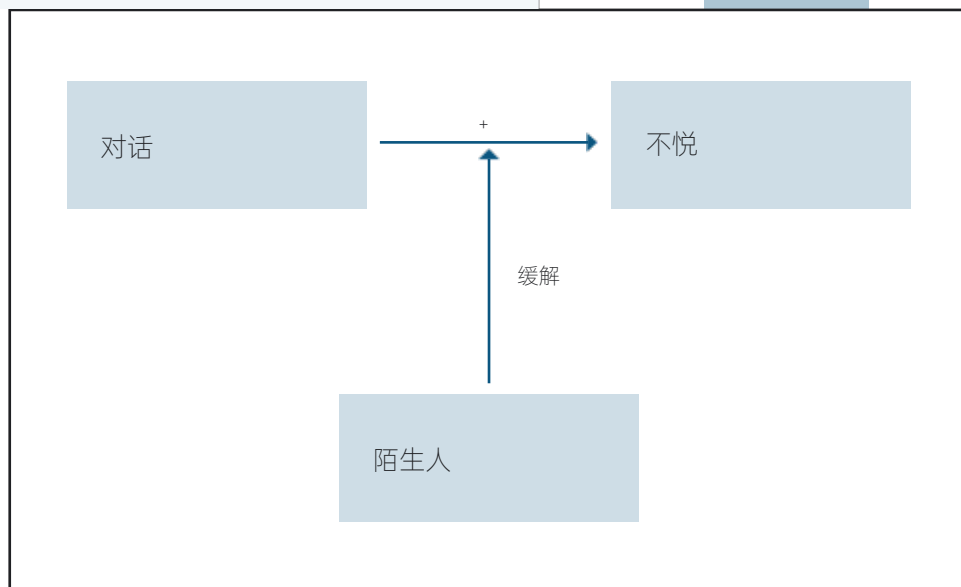
调节参数的示例

在统计学中，调节是一个难以把握的概念，但在现实世界中，调节意味着“它取决于情况”。让我们通过一个快速的示例来阐明在本报告背景下调节的概念。

在美剧《抑制热情》第三季中，拉里·戴维说：“我不喜欢和认识的人交谈，但陌生人我没有问题。”这为我们提供了一个讨论的示意图。

“我不喜欢和认识的人交谈，但陌生人我没有问题。”

Larry David



这个图表显示，对于Larry来说，对话对不悦的感觉有积极的影响。这里的“积极”仅仅是指增加，而不一定是好事。这通过连续的实线表示对话与不悦之间的关系，并用箭头指向不悦来展示。箭头表明我们认为因果流是从对话到不悦。根据我们的了解，对话通常会让Larry感到不悦。

第二点需要注意的是，陌生人（代表布尔值陌生人与否）并不指向另一个变量，而是指向一个效果，即箭头。这意味着我们认为陌生人是在修改一个效果，而不算一个变量。这就是为什么我们将箭头从陌生人指向另一个箭头，而

不是另一个变量。我们的意思是，Larry与陌生人交谈与对话对不悦的效果有关。换句话说，我们说对话对不悦的影响取决于Larry与之交谈的人是否是陌生人。当对话对象是陌生人时，对话对Larry来说是“没有问题的”。我们可以说陌生人减轻了Larry在交谈中感到的不快。

有几种不同的方式可以调节或影响另一件事情:

- **放大(Amplifies)** - 使积极影响更加积极, 负面影响更加负面。
- **减弱(Attenuates)** - 削弱影响。
- **缓解(Mitigates)** - 使积极影响变得不那么积极, 负面影响变得不那么负面。
- **颠倒(Reverses)** - 使积极影响变为负面, 负面影响变为积极。
- **修改(Modifies)** - 有时效果只是发生改变, 但无法用一个词来简单概括。这通常发生在以因素为原因的分类变量中。例如, 某些行业在不同条件下可能会有不同的行为, 比如经济波动。

调节示例

与调节类似, 我们经常考虑中介分析。分析中介的本质是“为什么”或“如何”。为什么太阳会使植物长得更高? 为什么吃草莓会使我尝到甜味? 按下油门踏板是如何使我的车动起来的? 止痛药如何缓解疼痛? 我们可以使用统计学来测试中介效应¹。这让我们能够说: “看起来 X 对 Y 的影响可以被 M 所解释或部分解释。”例如, 太阳对植物的高度影响是通过光合作用解释的。



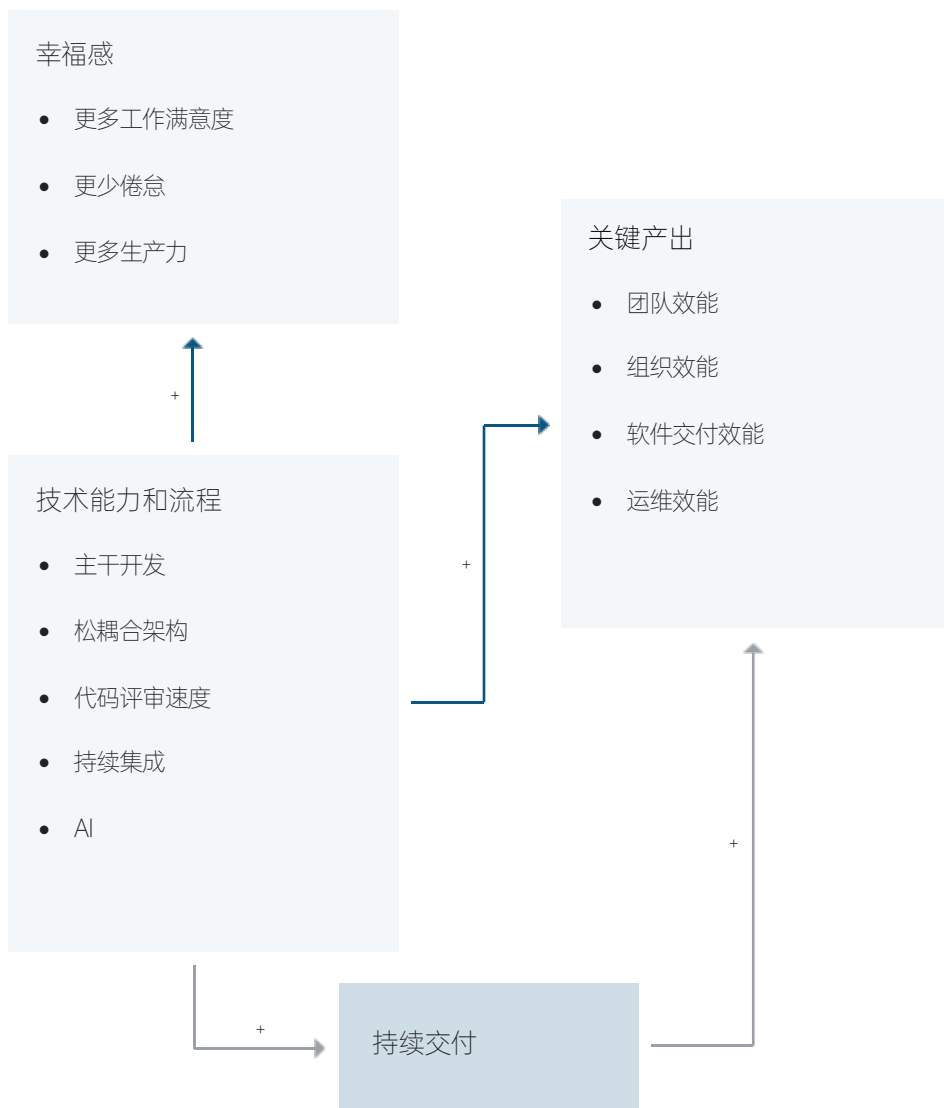
¹ <https://lavaan.ugent.be/tutorial/mediation.html>



第三章的模型

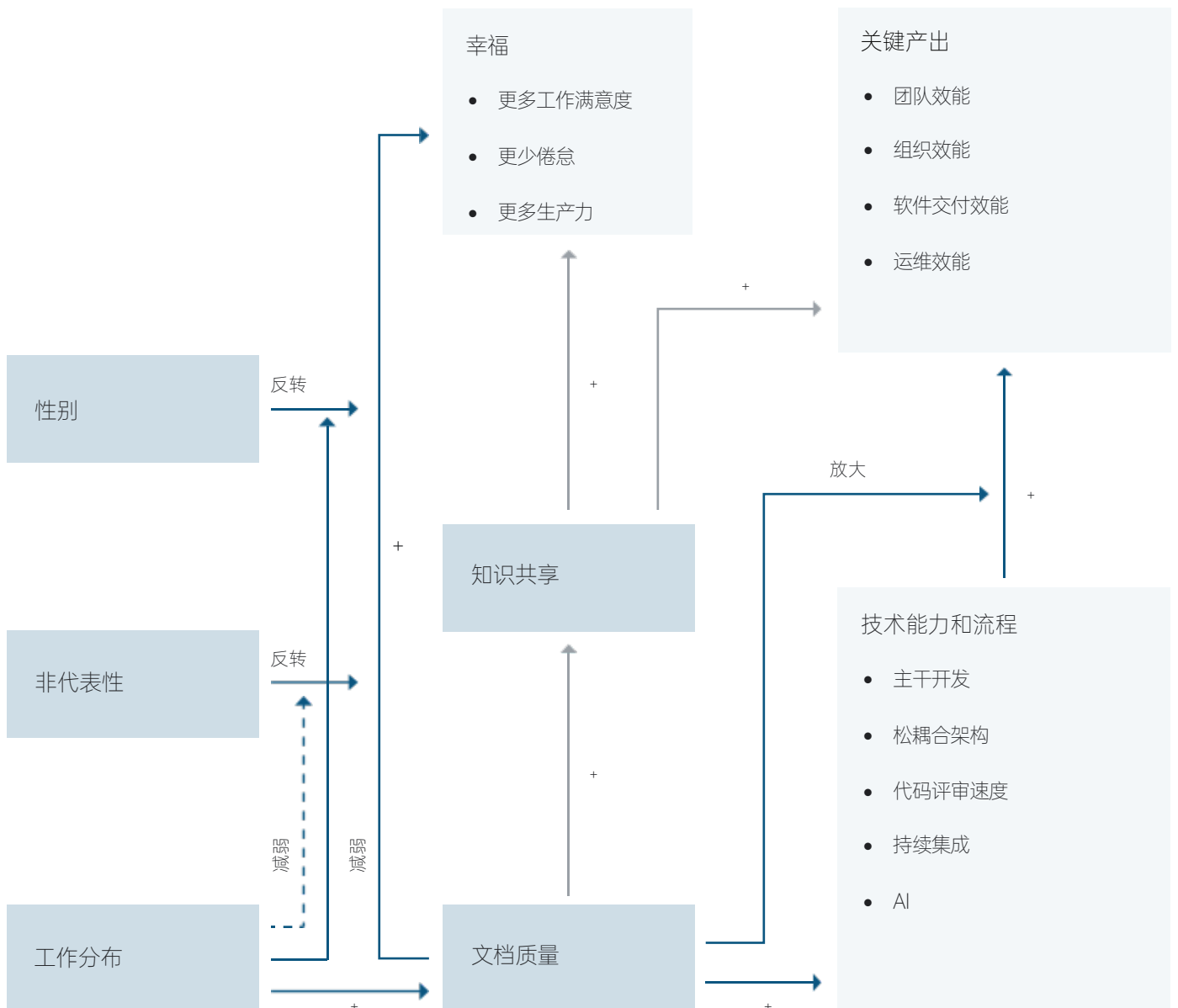
技术能力预示效能

技术能力和过程对幸福感和关键产出有积极影响。技术能力对关键结果的影响在一定程度上通过持续交付来进行中介，这些技术能力创造了一个持续交付的环境，对这些关键结果产生了下游影响。



第四章的模型 文档是基础

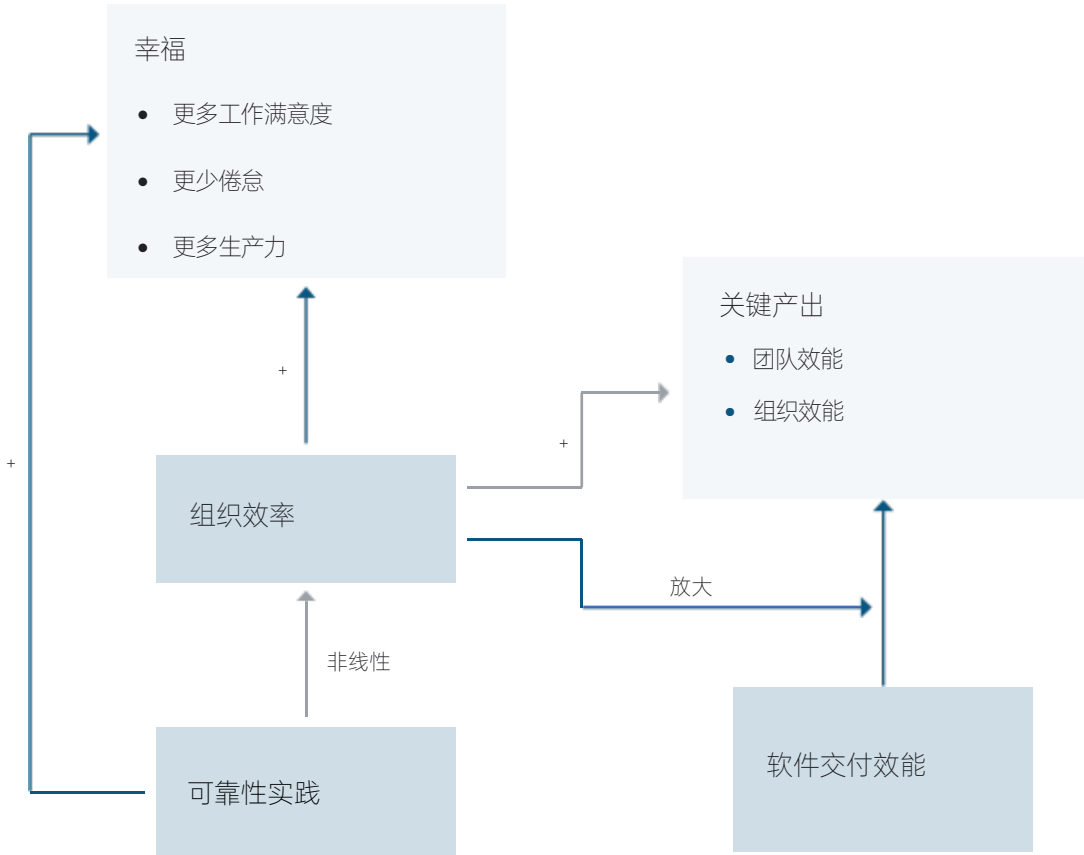
简而言之，高质量的文档对幸福感和关键产出有积极影响。这在一定程度上是因为它创造了一个可以进行知识共享的环境。高质量的文档还帮助团队建立技术能力和过程。此外，它还帮助技术能力和过程对关键结果产生更强的影响。最后，文档质量并不能使每个人的幸福感变得更好。我们建议阅读该部分以了解这一复杂发现的细节。



第五章的模型

可靠性释放效能

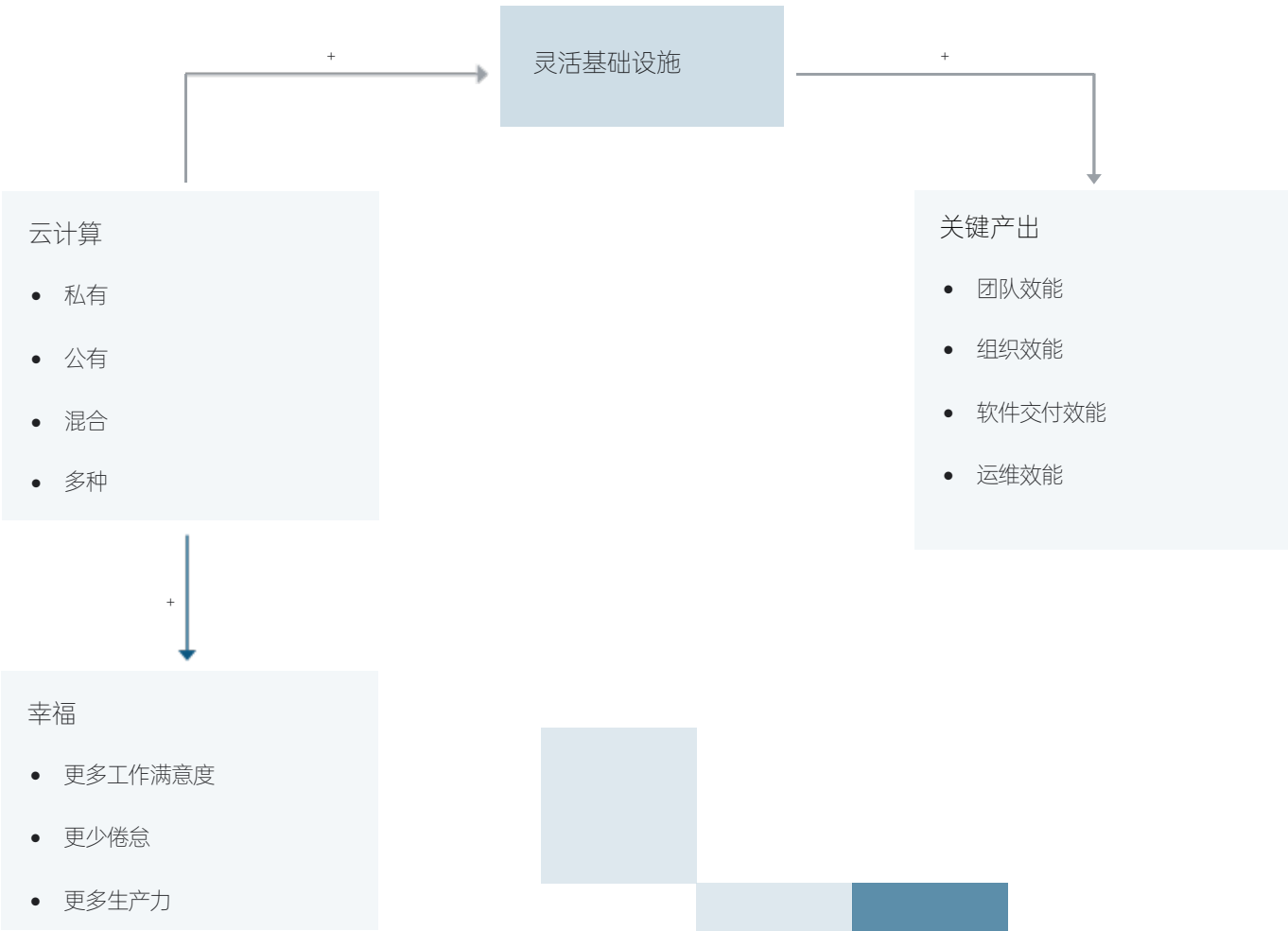
在这里，我们探讨了运维效能的核心作用。它不仅提高了幸福感和关键结果，还放大了软件交付效能的影响。可靠性实践与运维效能之间存在非线性关系。我们建议阅读本章以了解更多细节。



第六章的模型

灵活的基础设施是成功的关键

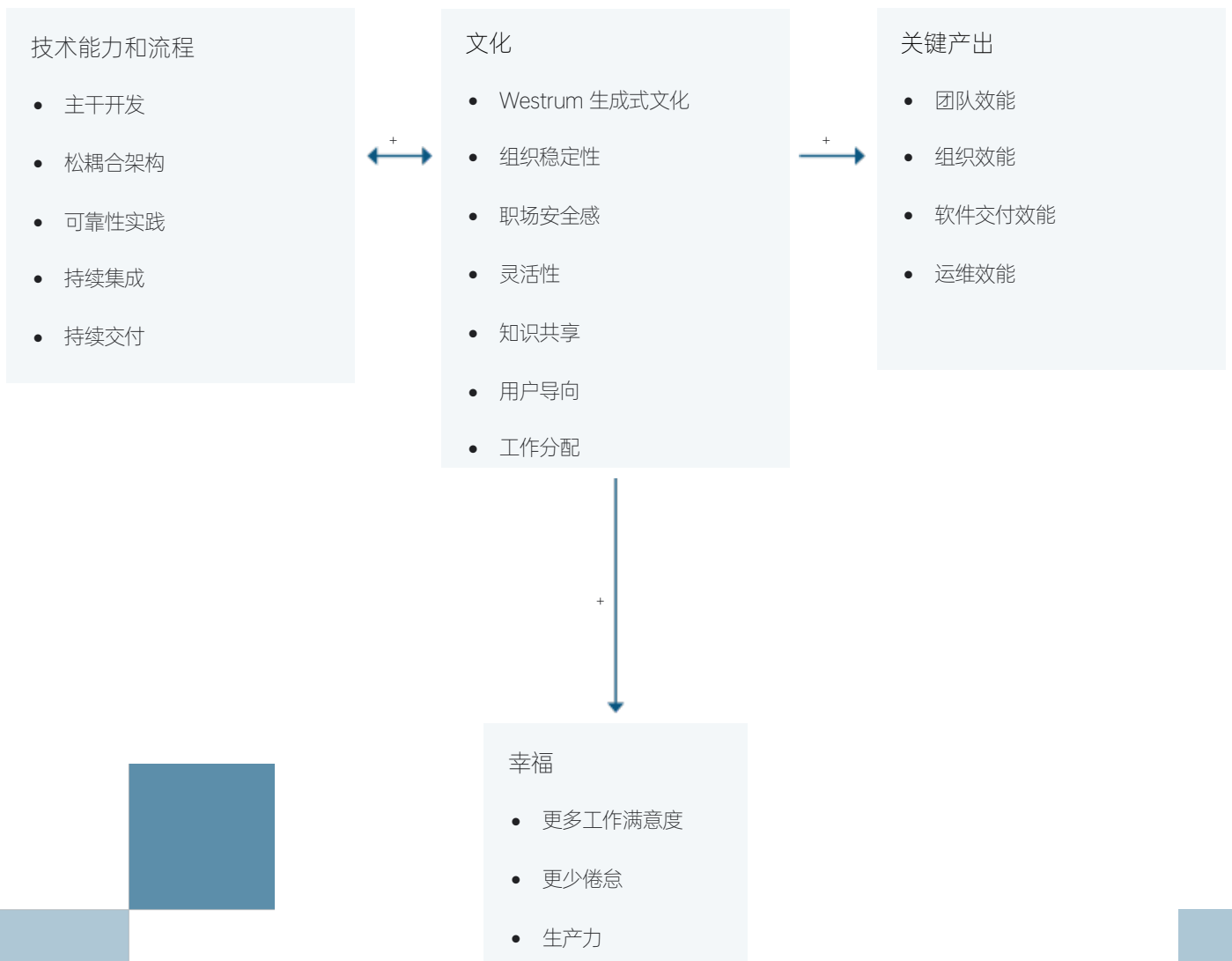
云计算对关键结果有影响，因为它提供了更灵活的基础设施。云计算也能提高幸福感。



第七章的模型

没有对文化的投入，这一切都无从谈起

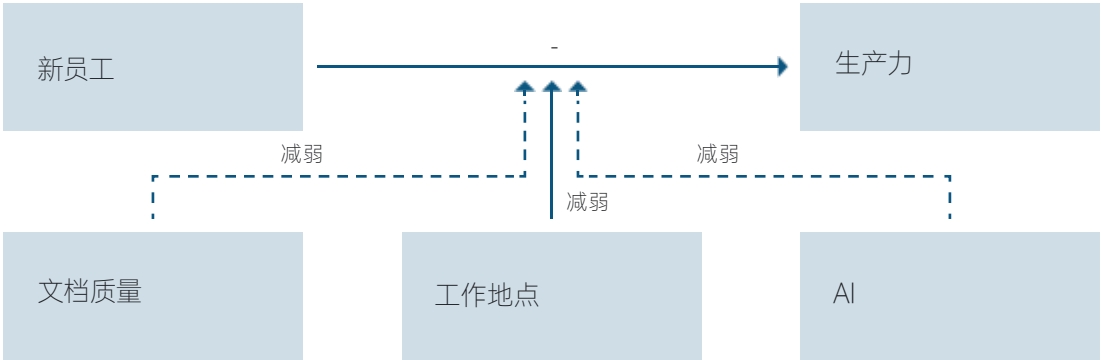
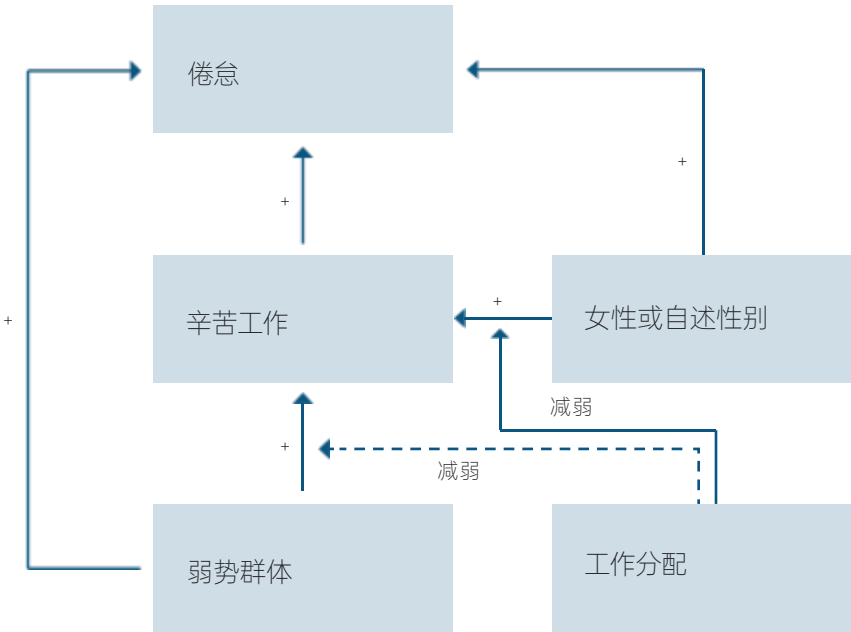
我们可以看到在这个模型中，文化是如此的核心。我们发现文化与技术能力、关键结果和幸福感有着积极的关系。



第八章的模型

在何时、何地以及为何你的定位重要

本章节包含两个模型。其中一个探讨为什么和何时，那些自认为处于被较少代表的群体中以及那些不自认为男性的人更容易经历高程度的倦怠感。另一个模型则研究了文档质量、工作地点或人工智能是否能够帮助新员工提高工作效率。



延伸阅读

加入DORA社区，一起讨论、学习并合作改进软件交付和运维效能。[DORA.community](https://dora.community)

参加DORA DevOps快速检索 <https://dora.dev/quickcheck>

探索推动更高软件交付和组织效能的技术、流程和文化能力。<https://dora.dev/devops-capabilities/>

在 SRE 上寻找资源

<https://sre.google>

<https://goo.gl/enterprise-roadmap-sre>

阅读书籍：《Accelerate: The science behind devops: Building and scaling high performing technology organizations》IT Revolution.

<https://itrevolution.com/product/accelerate/>

使用SPACE框架为您的团队发现适当的度量指标组合。
“The SPACE of Developer Productivity: There’s more to it than you think.”

<https://queue.acm.org/detail.cfm?id=3454124>

有一些关于现代代码审查的研究报告可以探索，请参考以下几份报告：

- “Expectations, Outcomes, and Challenges of Modern Code Review”
<https://dl.acm.org/doi/10.5555/2486788.2486882>

- “Code Reviews - From bottlenecks to Superpowers”
<https://learning.acm.org/techtalks/codereviews>

- “Modern Code Review- A Case Study at Google”
<https://dl.acm.org/doi/10.1145/3183519.3183525>

- “The Choice of Code Review Process: A Survey on the State of the Practice”
https://link.springer.com/chapter/10.1007/978-3-319-69926-4_9

- “Investigating the effectiveness of peer code review in distributed software development based on objective and subjective data”
<https://iserd.springeropen.com/articles/10.1186/s40411-018-0058-0>

还有一本书可供阅读：《The No Club: Putting a Stop to Women's Dead-End Work》。Simon & Schuster.

<https://www.simonandschuster.com/books/The-No-Club/Linda-Babcock/9781982152338>

您还可以查阅DORA的研究项目发表的出版物，包括之前的《Accelerate State of DevOps Reports》。

<https://dora.dev/publications/>

请访问常见问题解答了解更多关于研究和报告的信息。

<http://dora.dev/faq>

勘误表 - 您可以阅读和提交有关此报告的更改、修正和说明。请访问 <https://dora.dev/publications/errata>

附录

度量软件交付效能方式的优化

今年，我们对评估变更失败率和恢复能力进行了改变。

我们改变了受访者报告变更失败率的方式。在过去的几年中，我们给受访者提供了六个选项（0-15%，16-30%等）。今年，我们提供了一个滑块，使受访者可以选择0%到100%之间的任何值。

我们做出这个改变有两个原因：

- 变更失败率一直与其他三个衡量软件交付效能¹的指标有所不同。我们推测这可能部分原因是桶的大小。提供更精确的答案可能会产生更好的统计结果。我们已经验证了这个推测。
- 我们假设团队对于他们的变更失败率有比近十年前开始研究时更好的了解。我们没有真正的方法来验证这个假设。

今年，我们以两种不同的方式询问了关于恢复能力的问题：

- 之前，我们询问了这个问题：“对于您负责的主要应用程序或服务，在服务中断或影响用户的缺陷（例如，计划外的停机，服务受损）发生时，通常需要多长时间恢复服务？”
- 今年，我们在问题中添加了以下说明（以下是加粗的差异，但在调查中没有加粗）：“对于您负责的主要应用程序或服务，在生产中更改或向用户发布导致服务下降（例如，导致服务受损或服务中断），并随后需要纠正（例如，需要热修复，回滚，修复前进或打补丁）时，通常需要多长时间恢复服务？”

¹ Forsgren, N., Humble, J., 和 Kim, G. 《加速：精益软件与DevOps 背后的科学：打造和壮大高效能技术组织》（IT Revolution Press, 2018），37–38。

以前询问恢复时间的方式无法区分由软件变更引起的故障和由其他因素（如地震）中断数据中心服务的故障。我们有一个假设，更精确的语言将使我们能够将类似的故障类型进行比较，并且该语言在统计上与其他三个衡量软件交付效能的指标更加一致。

我们现在使用术语“失败的部署恢复时间”来区分我们的度量指标与过去使用的更通用的“恢复时间”（有时缩写为“MTTR”）不同。

MTTR在社区中引起了一些困惑：这个“M”是表示均值还是中位数？此外，寻求从故障中学习更多的从业人员（如弹性工程领域的从业人员）正在逐步放弃MTTR作为可靠的度量指标来指导学习和改进。²

这个新增的问题和新的度量指标，即失败的部署恢复时间，更符合衡量软件交付效能的目的。

²“超越简单的事件指标：Courtney Nash 谈 VOID”
<https://www.infoq.com/articles/incident-metrics-void/>



对比背后的数学计算方法

在整个报告中，有统计数据表明，某个特定变量的较高测量值会导致另一个指标得分高出N倍。

这是什么？它相对什么而言？以下是我们创建这些公式的配方：

- 1. 创建一个由回归技术评估的模型的公式，考虑到潜在的偏倚路径：

Happiness ~ 5.64 + 0.19 * sunshine + 0.14 * temperature

- 2. 对于兴趣变量，比如阳光sunshine，找到高值 high 和低值 low 进行比较：

mean = 6.3; sd = 1.4

High sunshine = 1 sd above mean = mean + sd = 6.3 + 1.4 = 7.7

Low sunshine = 1 sd below mean = mean - sd = 6.3 - 1.4 = 4.9

- 3. 计算协变量的均值。在这个例子中，协变量是温度temperature，其值为0.24（标准化温度）。
- 4. 根据第一步的公式，填入高阳光和低阳光的值。您会注意到公式中只有一个数字不同。这样做可以使其他条件保持不变，只关注兴趣变量的差异。

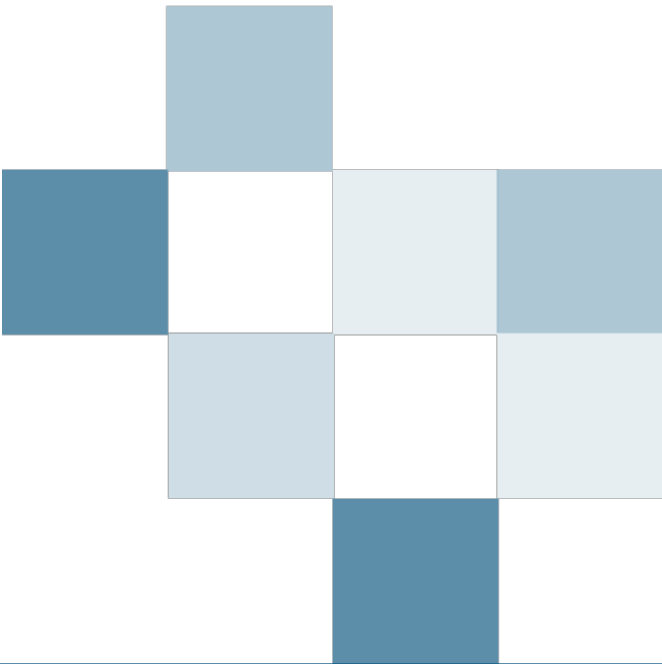
Predicted high sunshine happiness = 5.64 + 0.19 * 7.7 + 0.14 * 0.24 = 7.1

Predicted low sunshine happiness = 5.64 + 0.19 * 4.9 + 0.14 * 0.24 = 6.6

- 5. 计算比率：

predicted high sunshine happiness		7.1	
		=	
predicted low sunshine happiness		6.6	
		=	1.1x

- 6. 这个比率表明，相对于阳光较少的情况，阳光较多的情况下会导致幸福感happiness 水平提高10%。



“模拟”是什么意思？

并不是说我们捏造了数据。我们使用贝叶斯统计方法计算后验概率，后验概率捕捉“不同参数值出现的预期频率。”³而“模拟”部分是从这个后验分布中抽取超过1,000次，以探索给定数据的情况下，某个参数（均值、 β 权重、 σ 、截距等）最可信的取值。“想象一下，后验分布就是一个充满了参数值的桶，比如 0.1、0.7、0.5、1 等等。每个值在这个桶里存在的概率与其后验概率成比例，这样靠近峰值的值比较常见，而在尾部的值比较少见。”⁴

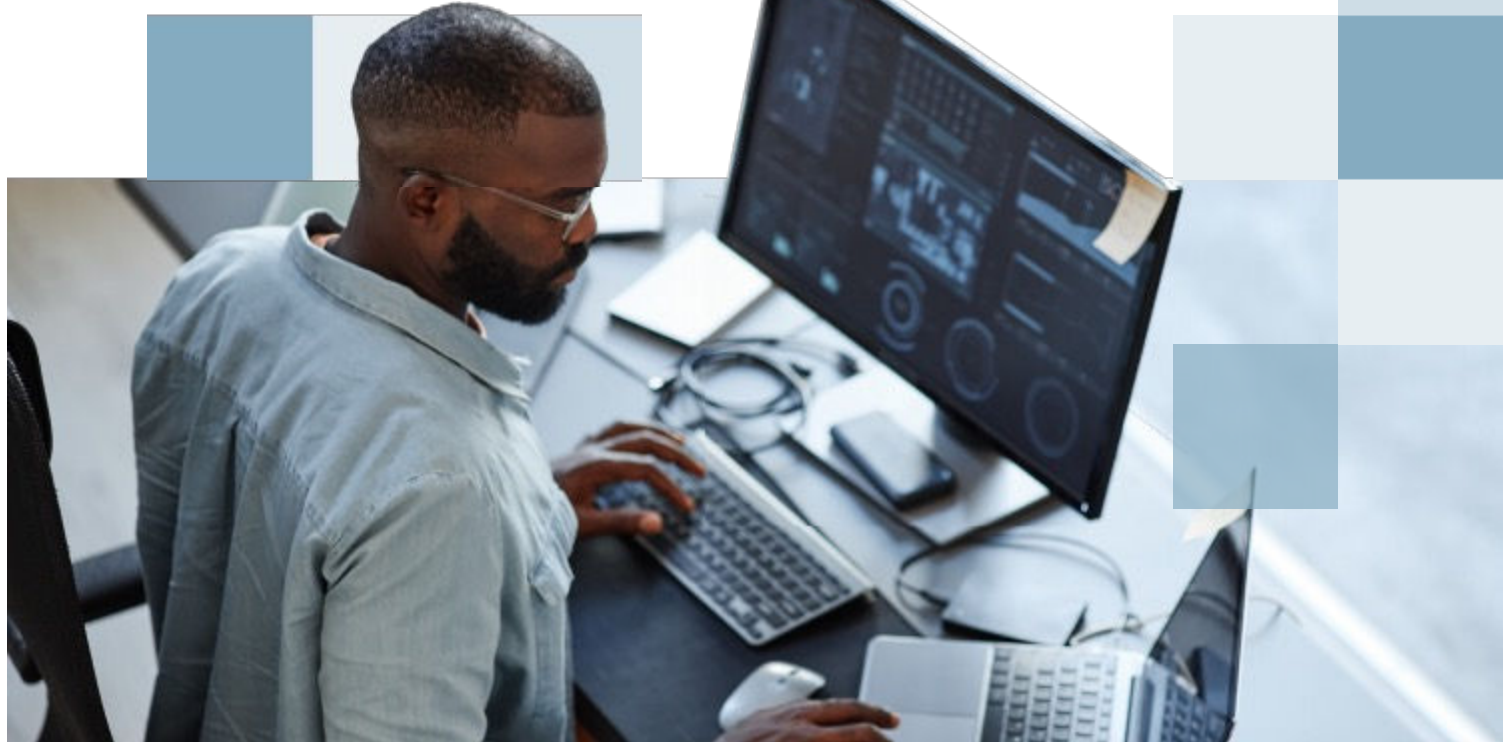
总结起来，这就是我们使用模拟来探索数据的可能解释，并了解其中的不确定性程度。您可以把每次模拟看作是一个小AI，它只知道我们

的数据和一些规则，试图利用知情推测来填充一个空白的参数。您这样做了4,000次，就得到了对于给定参数的4,000个小AI的推测。从这些推测中，您可以掌握很多东西。您可以了解推测的平均值是什么，89%的推测值落在哪些范围内，有多少个推测值高于一定水平，这些推测值的变化幅度有多大等等。您甚至可以将多个模拟结果组合起来做一些有趣的事情。

当我们展示一张有很多线条或潜在数值分布的图表时，我们试图向您展示基于我们的数据，什么是最有可能的情况以及其中存在多少不确定性。

³ McElreath, Richard. Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC, 2018, 第50页

⁴ McElreath, Richard. Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC, 2018, 第52页



“Accelerate State of DevOps 2023” 由 Google LLC 发布，并在 CC BY-NC-SA 4.0 许可下授权。

译者

中国 DevOps 社区翻译小组



张扬
DevOps解决方案架构师
社区志愿者



冷大鲲
产品设计
社区志愿者



王晓宁
交付经理
社区志愿者



任悦
联通西部创新研究院
软件工程师，社区志愿者



秦团结
联通西部创新研究院
软件工程师，社区志愿者



周一行
程序员
社区志愿者

反馈收集社区官网: <https://DevOpsChina.org/contact-us/>