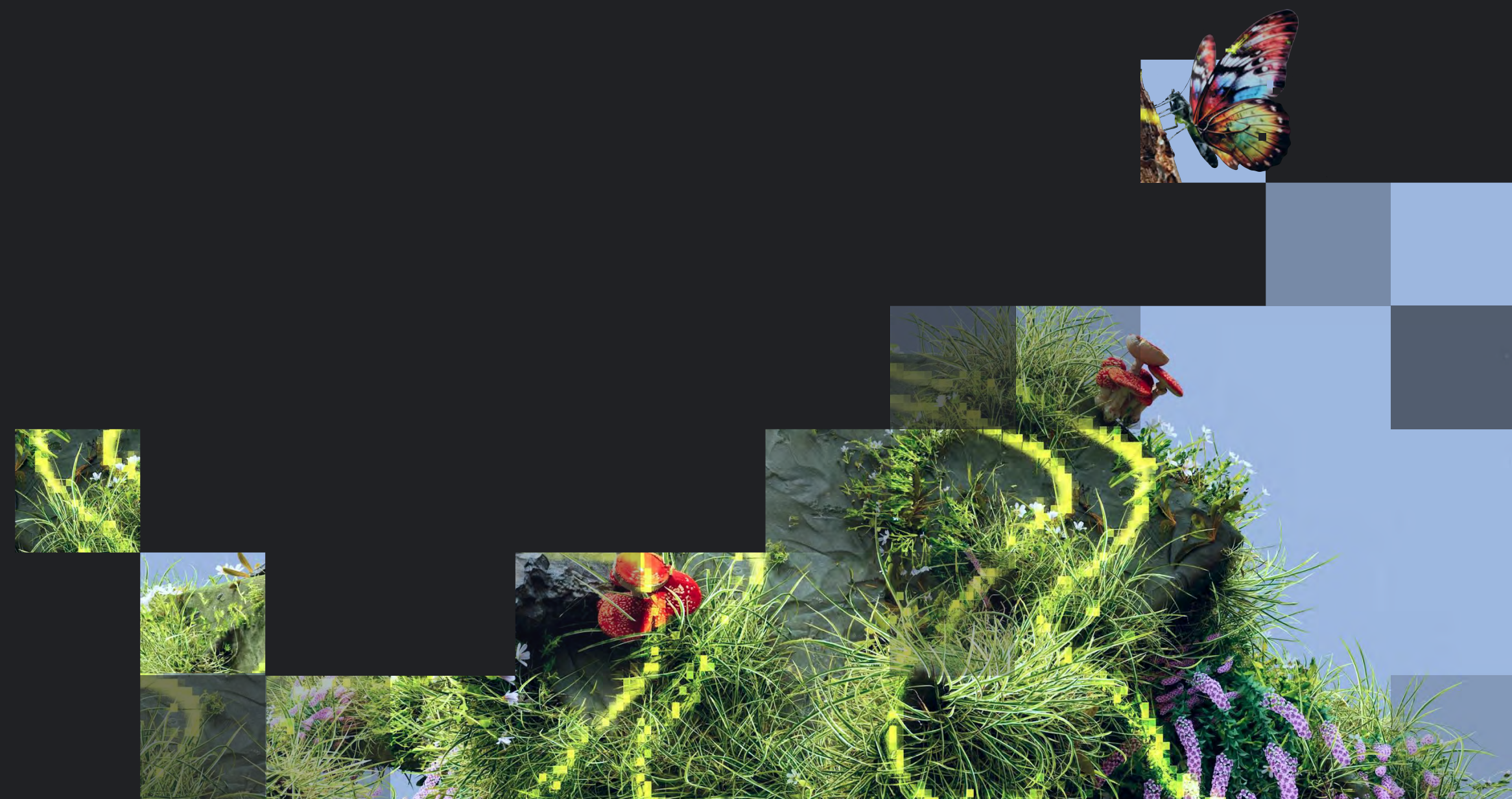


Google Cloud

2024 年 数据力和 AI 趋势报告

生成式 AI 的影响



新机会、 新技术、 新技能。

随着生成式 AI 的问世，全球拉开了革新的序幕。这项革命性技术将以前所未有的方式，颠覆各行各业并改变我们的生活。数据是 AI 发挥强大效用的核心。企业要想充分利用生成式 AI，就需要有能力访问、管理和激活各种系统中的结构化和非结构化数据。

不仅如此，还要使用 AI 和机器学习 (ML) 技术深入了解数据，借此强化模型或改善客户体验。只要做到上述几点、确保数据质量和安全，同时遵循负责任的数据使用原则，企业就能获得成功。

随着新时代的到来，整个数据栈发生了连锁反应。为了进一步研究技术的变化趋势，Google 向数百位企业和 IT 主管展开问卷调查，以了解他们的目标和生成式 AI 采用策略。本报告将深入探讨受访者对 2024 年及未来技术发展趋势的观点，为希望利用生成式 AI 的企业提供宝贵数据洞见。



五大趋势概览：

1

生成式 AI 会加
快数据洞见在组
织内的交付。

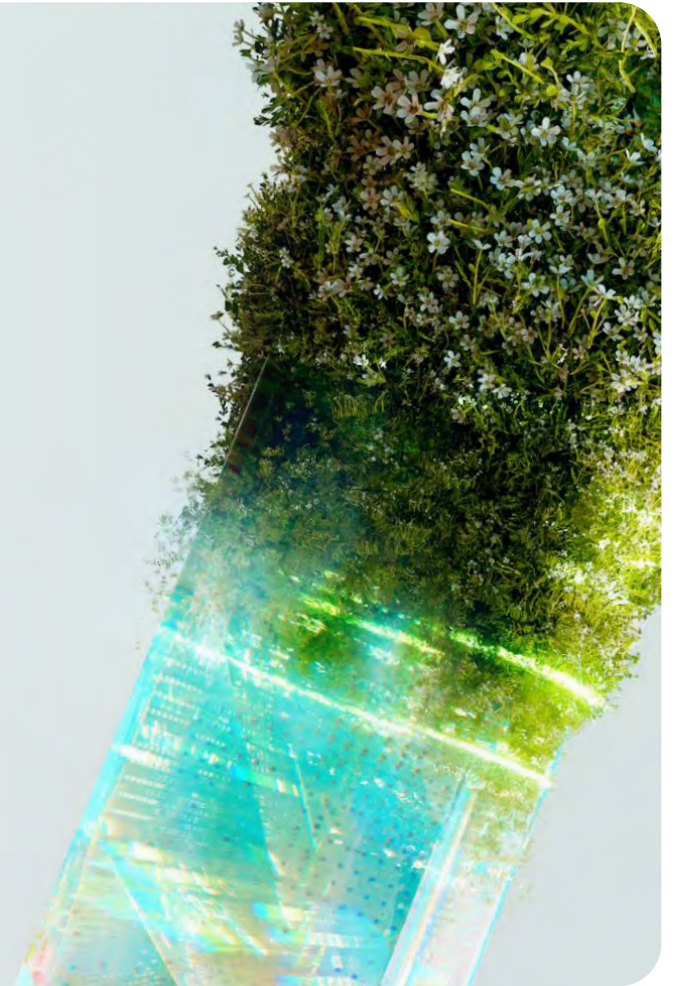
阅读用时：5 分钟



2

数据和 AI 的角色
将变得模糊。

阅读用时：6 分钟



3

强有力的数据治
理机制是实现 AI
创新的关键。

阅读用时：3 分钟



4

运营数据将激发生
成式 AI 在企业应
用方面的潜力。

阅读用时：4 分钟



5

2024 将是数据平台
现代化改造突飞猛
进的一年。

阅读用时：5 分钟



生成式 AI 会加 快数据洞见在组 织内的交付。

近三分之二的数据决策者预计，**到 2024 年，数据洞见将实现民主化。**
84% 的人认为生成式 AI 将帮助他们的组织更快地获取数据洞见。

为什么您应该关注?

生成式 AI 为技术环境带来的巨大变革怎么强调都不为过。就商业智能 (BI) 而言, 相关工具将变得越来越易于使用, 即便是非技术团队成员也将可以利用这些数据洞见来提高工作效率, 并比以往更快地传播知识。这意味着您的组织可以全面提高数据素养并制定更明智的决策, 最终在市场上取得更重大的成就。

目前 52% 的非技术用户已经在使用生成式 AI 获取深入数据洞见。

现代 BI 工具已经在设法为用户提供所需数据, 一个简单的例子就是内嵌在最相关情境中的数据报告, 例如销售人员所用 CRM 中显示的客户数据洞见。不过, 这些数据洞见始终需要由分析师精心整理。最终用户距离他们的数据始终有一步之遥, 通过将大语言模型与业务数据对接起来, 可以帮助弥合这一差距。团队成员能够以对话方式直观地与数据交互, 还可以通过与数据交互或在整个公司内执行简单搜索来创建报告和信息中心。事实上, 本报告调查的许多组织已经在这样做了。

“Moody's 在财务数据、披露和报告制作方面拥有深厚的专业知识, 这项独特优势让我们能开发经过微调优化的大语言模型。Google Cloud 的生成式 AI 将帮助我们的客户和员工比以往更快地生成新数据洞见。”

Moody's

NICK REED

Moody's Corporation 首席产品官

无论是走在前沿的数据科学家，还是业务线人员，都已开始使用 AI。

从根本上讲，如果工具能让用户直接以自然语言获取所需的关键业务数据，将成为弥合组织当前技能缺口的一大助力。

预计从 2024 年开始，更多企业用户将使用搜索技术与数据“对话”，并利用对话式界面创建报告、信息中心和直观、智能的可视化图表，

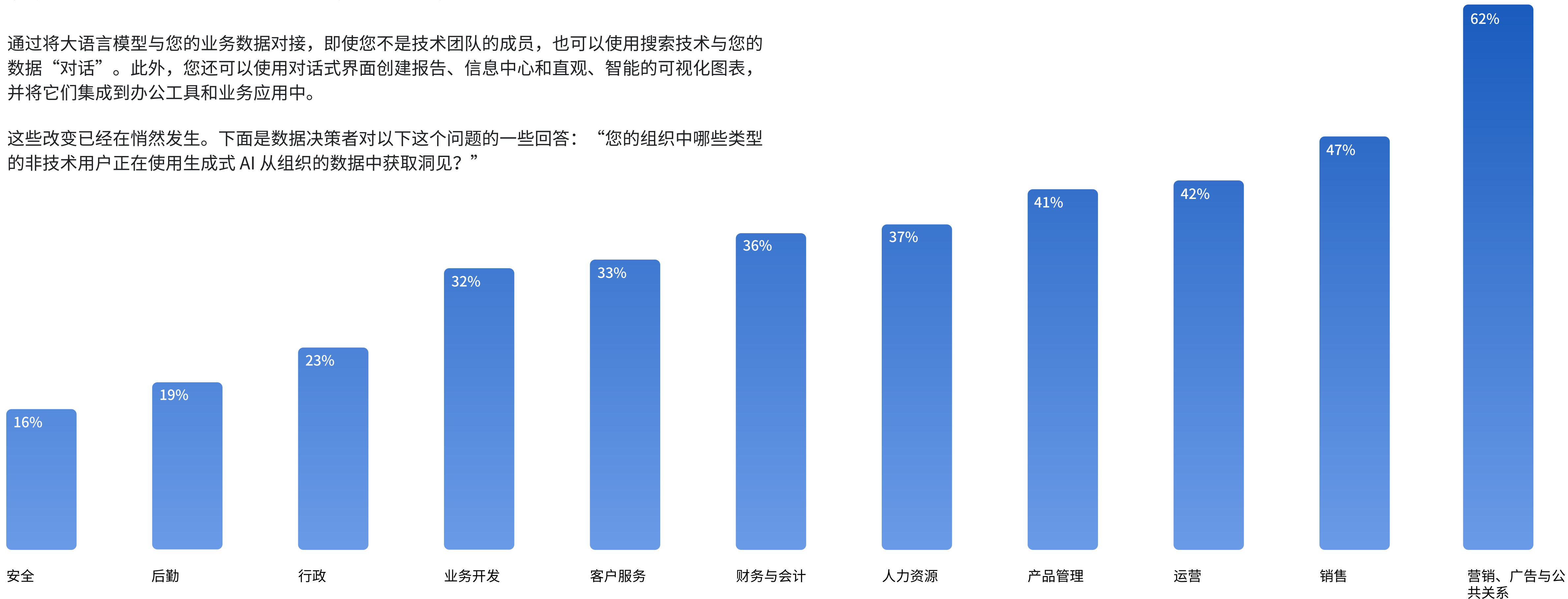
同时将这些集成到办公工具和业务应用中。许多应用都可让用户查看他人是如何找到问题的适当解答，这样，人们不仅可以集思广益，还能深入分析哪些互动在一天、一季度或一年内带来了最大的影响。



所有人都是可以善用 AI。

通过将大语言模型与您的业务数据对接，即使您不是技术团队的成员，也可以使用搜索技术与您的数据“对话”。此外，您还可以使用对话式界面创建报告、信息中心和直观、智能的可视化图表，并将它们集成到办公工具和业务应用中。

这些改变已经在悄然发生。下面是数据决策者对以下这个问题的一些回答：“您的组织中哪些类型的非技术用户正在使用生成式 AI 从组织的数据中获取洞见？”



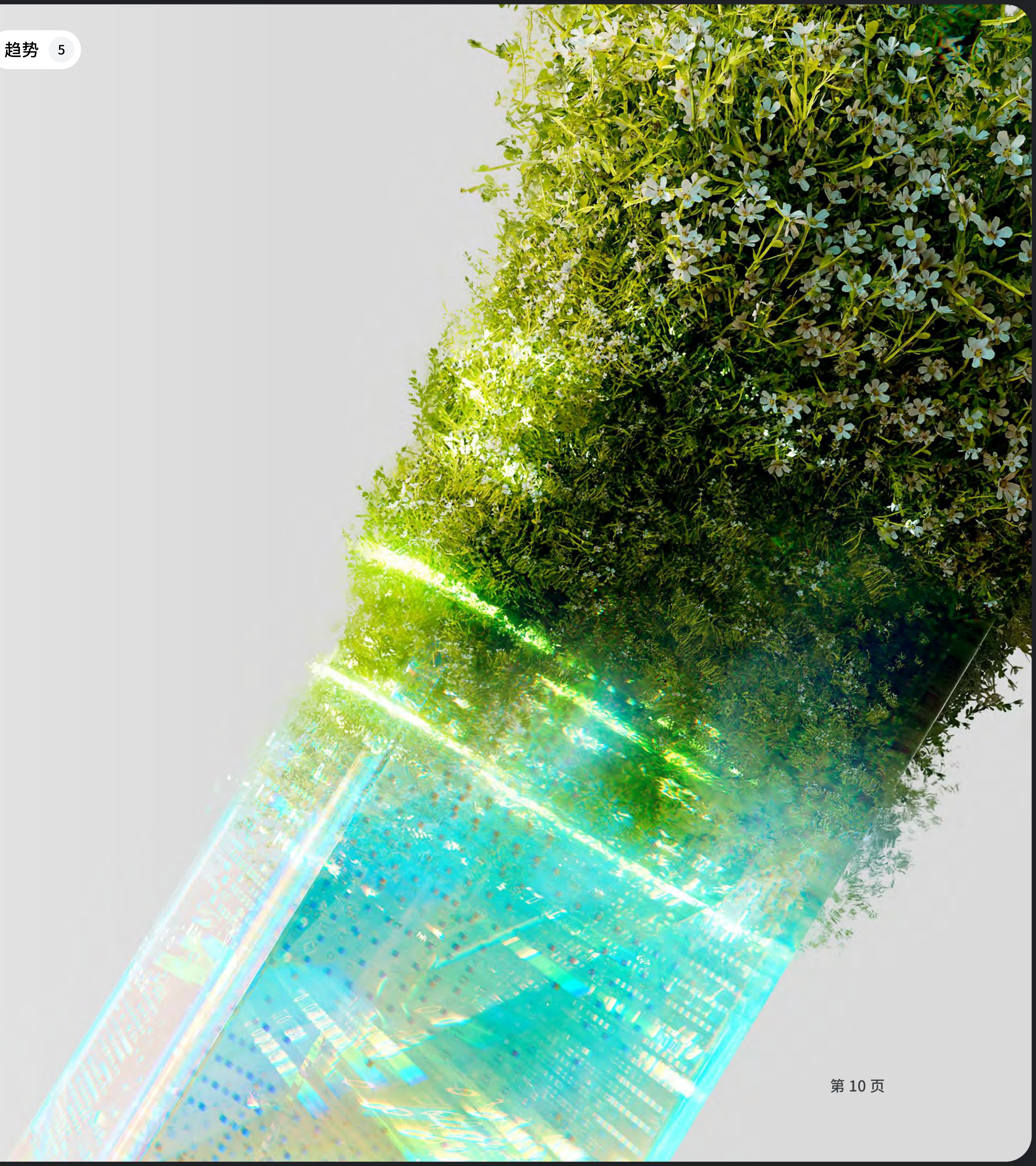


“50 多年前，Wendy’s 推出了业内第一个现代化的餐品自取窗口，我们很高兴能够与 Google Cloud 合作，继续推进我们的事业，为得来速体验带来新一轮创新浪潮。Google Cloud 的生成式 AI 技术为我们创造了重大机遇，让我们有机会为顾客提供真正与众不同、顺畅、更快速的体验，让员工能够继续专注于制作美味餐食并与食客建立联系，让他们不断再次光顾。”

TODD PENEGOR

Wendy’s 总裁兼首席执行官

数据和 AI 的角色 将变得模糊。



80% 的受访者认为各种数据角色之间的界限开始变得模糊。



为什么您应该关注？

随着 AI 的应用范围不断扩大，公司将原始数据用于 AI 的速度将变得日益重要。

如果组织建立成熟的使用流程，将能制定更明智的决策、更快推出新产品和服务，并提供更优异的客户体验。

随着流程的简化，数据和 AI 的角色会变得越来越模糊。这意味着以前相互孤立的团队必须比以往更加紧密地合作。

“在我小时候，我妈妈会花几个小时与旅行社一起规划度假行程。与 Google Cloud 合作采用生成式 AI 后，我们通过聊天机器人就能提供定制旅行规划服务。我们的目标不仅仅是为客户安排行程，还要协助他们打造独特的旅行体验。”

priceline

MARTIN BRODBECK
Priceline 首席技术官

现在，许多数据分析师逐渐开始承担以前由数据科学家负责的业务，反之亦然。

当前，各组织最欠缺的技能有数据分析、数据工程、AI 以及业务分析。生成式 AI 可帮助提升现有数据团队和工作负载的效率，进而弥合日益严重的技能差距。无缝使用数据和 AI 平台可以帮助组织加速从数据到 AI 的历程，从而提高效率并加速创新。

此外，数据和 AI 工具的互联程度越来越高，可以帮助用户简化数据和 AI 工作流。

这种互联让用户可以：

拥有一个为数据工程师、数据分析师和数据科学家提供多种编程语言（如 SQL、Python 和 Spark）支持的共同工作空间。

使用 CI/CD、版本记录、源代码控制等软件开发最佳实践来管理数据资产，以提升协作和交接效率。

研究表明，超过一半 (54%) 的数字化领先企业表示，由于缺乏技能，他们难以跟上变革的步伐。

Nash Squared, 《Digital Leadership Report》
(数字领导力报告), 2023 年。



生成式 AI 还可以帮助员工高效完成更多技术任务。例如，针对更新财务报告系统所需的新代码提供建议、概述营销活动的 A 和 B 版本，或制作初稿供员工用于正式生产环境。

《The organization of the future: Enabled by gen AI, driven by people》（未来的组织：由生成式 AI 辅助，由人类驱动），McKinsey & Company，2023 年。

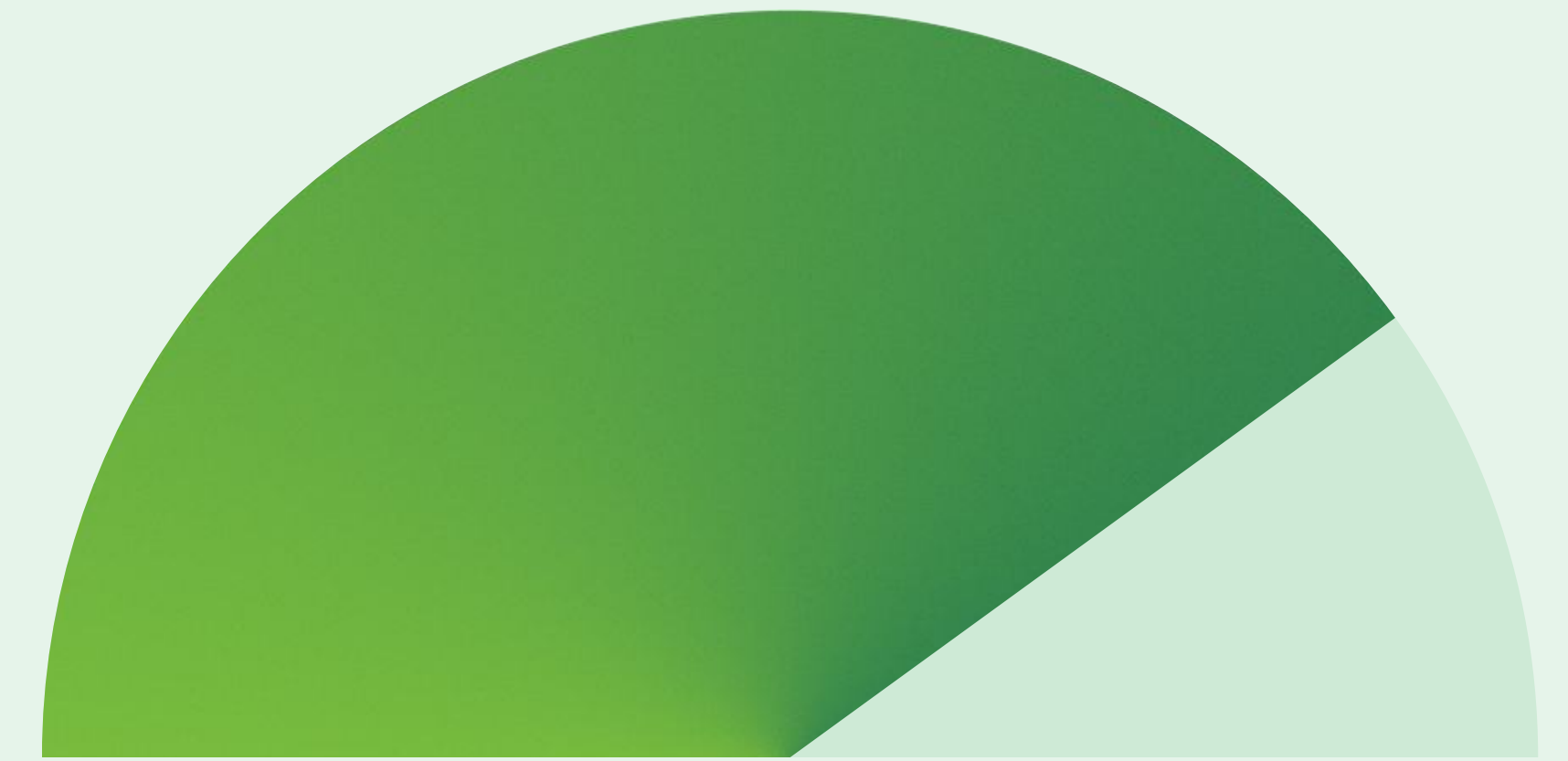


将 AI 直接应用于数据可以充分激发其潜力， 无论数据的格式如何。

导致组织未能充分发挥数据潜力的最大阻碍是当今生成的大量未得到利用的非结构化数据。这包括图片、文档和视频等格式。据估计，约有高达 80% 的数据尚未被组织利用。

由固定字段和列组成的结构化数据（例如电子表格或数据库中的数据）可以使用传统方法轻松处理和分析。而非结构化数据，例如社交媒体帖子、电子邮件、客户通话录音、临床文档和传感器测量结果，通常很复杂且难以解读，因此很难提取有意义的数据洞见。

先进的 AI 和机器学习 (ML) 技术的问世，彻底改变了组织利用数据的方式。这些先进技术提供了独特的机会，可充分激发所有数据的潜力，无论是结构化、半结构化还是非结构化数据。此外，多模态 AI 也解锁了无限可能，让组织能以更高的效率和准确性，针对他们的企业数据执行模型调优及标准答案关联。文本嵌入让您可以直接对数据执行矢量搜索，无需复杂且耗时的预处理步骤。这简化了查找相关信息、识别模式和趋势以及对文档等来源中类似的非结构化数据进行聚类的过程。



**预计到 2025 年，非结构化数据
将占全球数据量的 80%。**

VentureBeat, 2022 年。

强有力的数据治理机制 是实现 AI 创新的关键。



66% 的组织表示，其内部至少一半的数据都是暗数据，这构成了重大的风险。

《Five Factors For Planning A Data Governance Strategy》(制定数据治理策略的五大要素)，Forbes，2023 年；Gartner 术语表，暗数据，2024 年。



为什么您应该关注？

生成式 AI 好处多多，但也有其缺点。许多组织发现这项技术会带来新的漏洞和弱点，特别是在数据质量方面。仅仅将 LLM 应用于数据是不够的。这些模型必须基于优质企业数据，否则就有产生幻觉的风险。在数据治理、质量和信任方面采取务实做法的组织将在运用 AI 交付切实业务成果方面处于有利地位。

近半受访者对其组织的数据质量还算有信心 (45%)，另有 11% 的受访者则不太有信心。

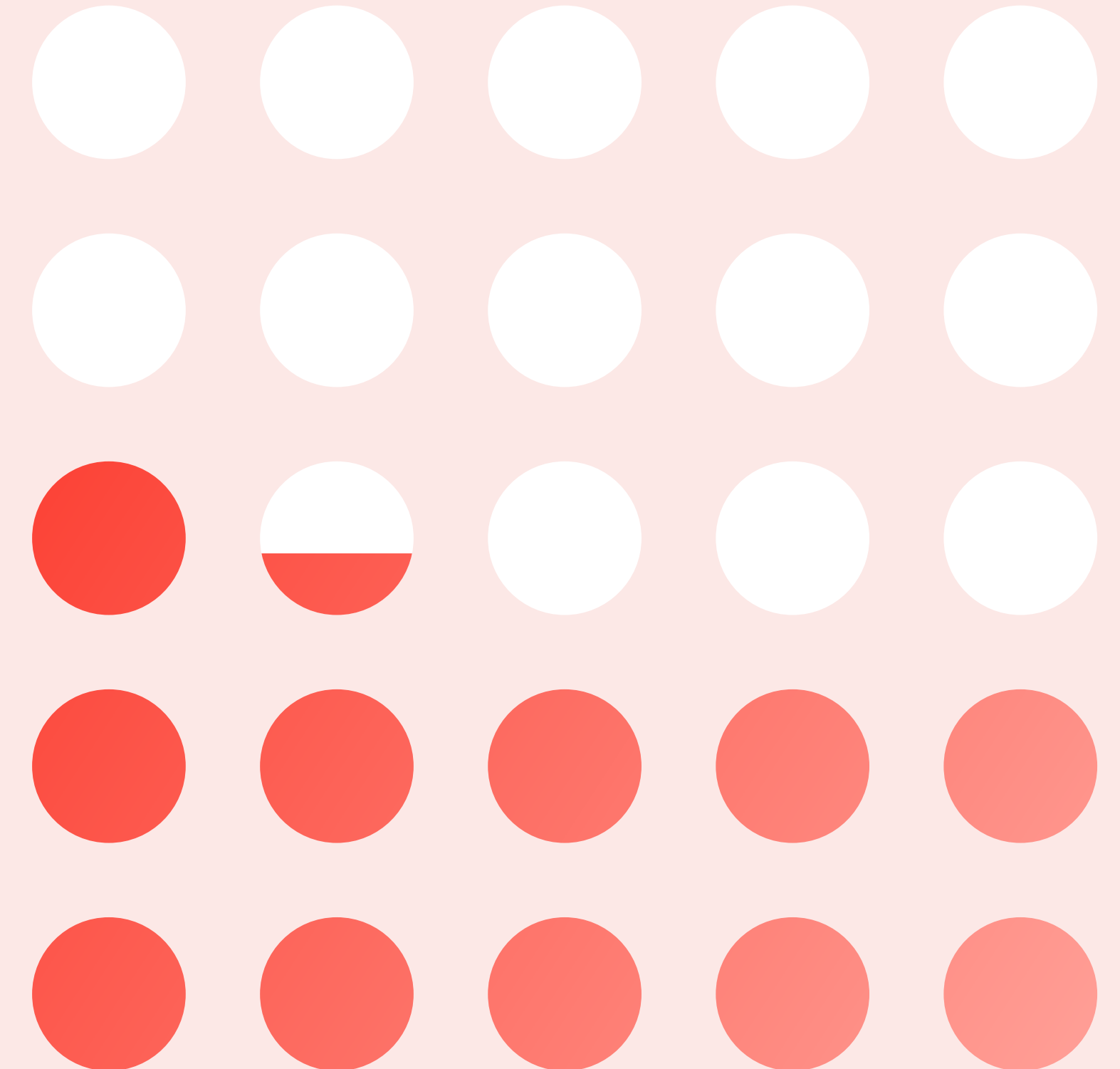
类似的，大多数受访者 (54%) 认为他们的组织在数据治理方面只是较为成熟，而只有 27% 的受访者认为他们的组织极为成熟或非常成熟。

可喜的是，许多组织正采取行动来确保数据质量良好，且准确可信。大多数受访组织认为自己正在：

- 打造数据驱动的文化。
- 将数据治理监控机制集中化。
- 制定集中式政策管理、监控及审核机制。

不到一半 (44%) 的受访者对其组织的数据质量充满信心。

《Google Cloud Customer Intelligence Trends Research Survey》
(Google Cloud 客户智能趋势研究问卷调查)，2024 年。





Snap Inc 使用 Google 的数据云来整合分散各处的数据，提供一个特定业务领域专用的自助式数据平台，实现集中式治理和可见性，同时保持对分散各处的数据的所有权。随着数据效率的提高，他们得以专注于改善用户体验和提高互动度。



家乐福使用 Google 数据云，通过提高数据安全性和强化对业务关键型应用的安全访问，实现零信任网络保护。他们以数据为中心的基础设施让他们能够灵活地快速调整，并为客户提供最高质量服务。

现在组织最需要什么？

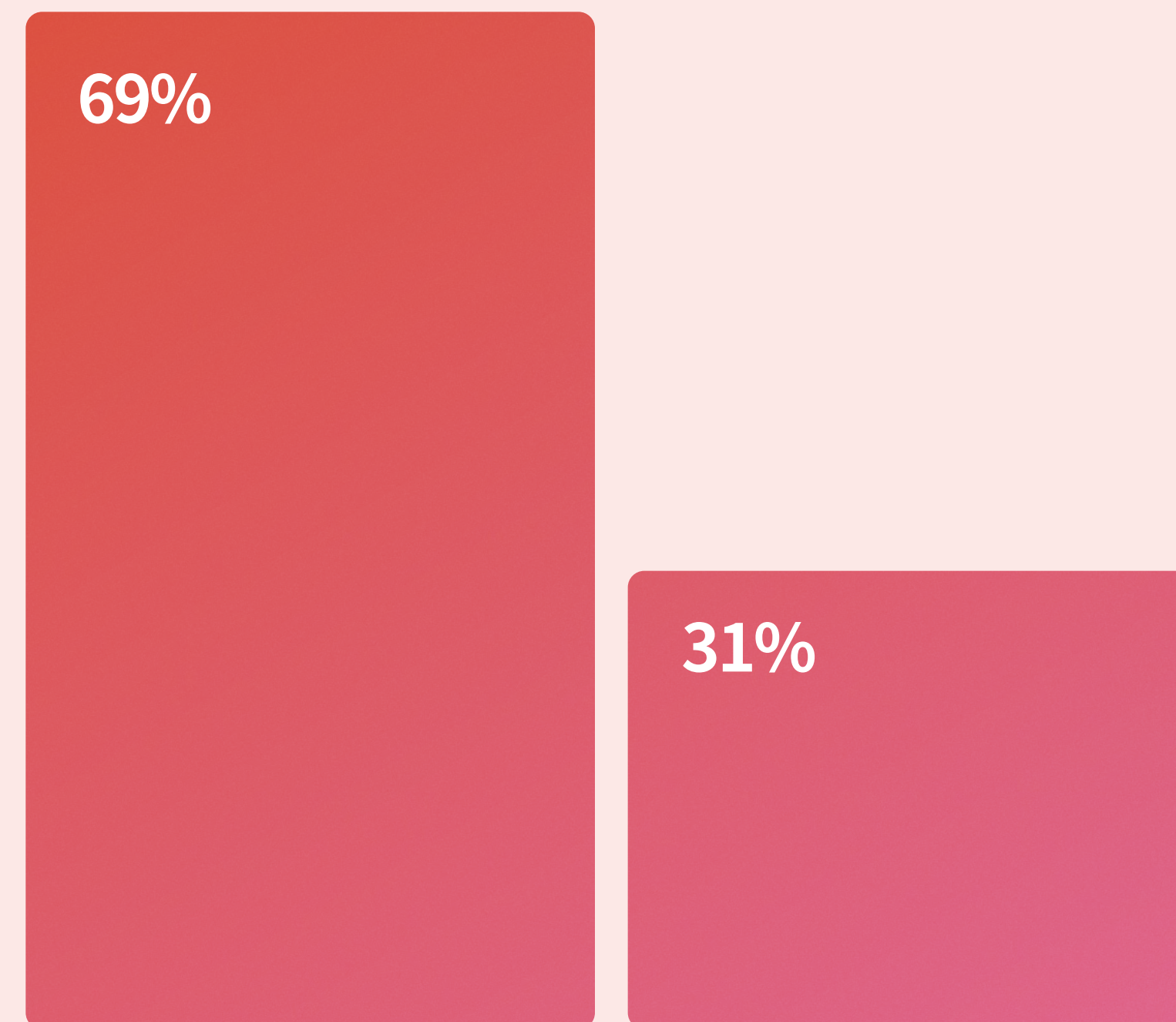
现在，组织最需要的就是从设计上保证安全且充分集成数据加密机制的数据平台。理想的平台应能够自动建立组织自有数据目录，并让组织能使用元数据有条理地整合及整理数据。这样一来，组织就可以根据业务情境，集中保护及治理数据，并使用内置自动化和智能功能来剖析数据、提高质量及沿袭数据等，从而更高效地大规模管理数据。这可以确保：

数据质量。 自动生成数据质量规则来衡量数据的完整性、准确性和有效性。

端到端数据沿袭。

自动生成沿袭，以便您可以跟踪数据流、执行影响分析，并使用沿袭作为跨数据和 AI 模型的治理和合规基础。

数据和 AI 资产的统一治理。 对数据授权、保留和分类进行集中的政策管理、监控和审核。



69% 的员工在过去 12 个月曾绕过组织的信息安全规章。

《5 Gartner Predicts Nearly Half of Cybersecurity Leaders Will Change Jobs by 2025》
(Gartner 预测近半数信息安全主管会在 2025 年前换工作)，2023 年。

运营数据将**激发生成式 AI** 在企业应用方面的**潜力**。



71% 的组织计划使用与生成式 AI 功能集成的数据库。

《Google Cloud Customer Intelligence Trends Research Survey》(Google Cloud 客户智能趋势研究问卷调查), 2024 年。



为什么您应该关注?

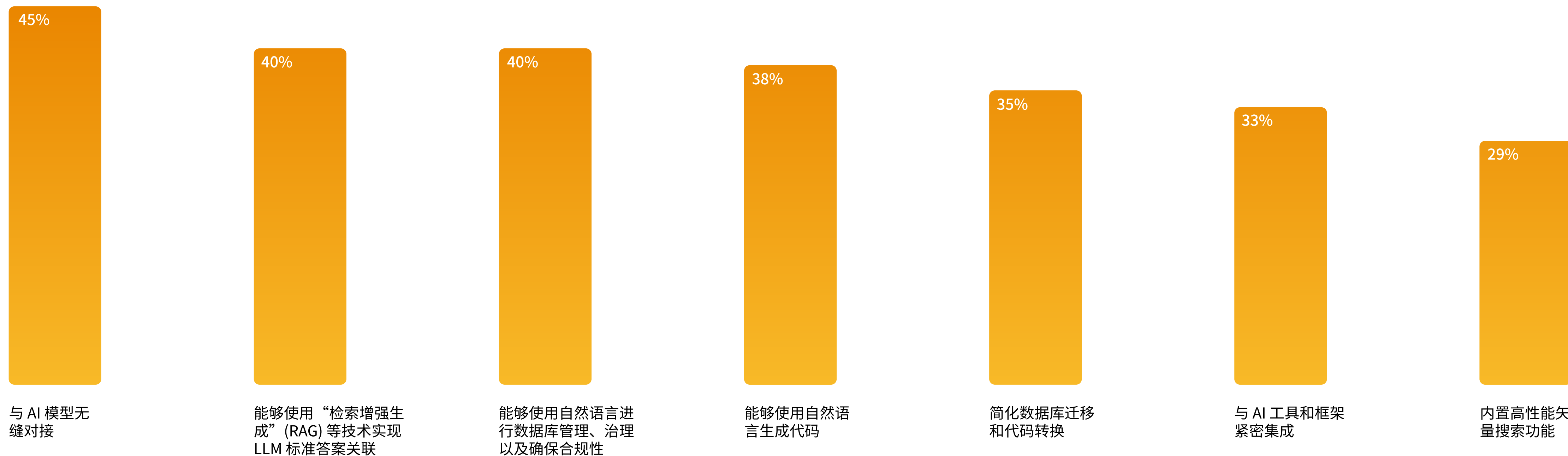
企业非常看好大语言模型 (LLM)。尽管他们都体验到了 Gemini 和其他大语言模型等工具的强大功能，但也意识到这些工具的创造性本质并不适合大多数企业应用场景。企业生成式 AI 应用面临着单靠 LLM 本身无法解决的诸多挑战。例如在不耗费大量成本的前提下，提供准确、最新的信息和符合当下情境的用户体验。

具有矢量支持的运营数据库和仓库有助于将 LLM 和企业生成式 AI 应用融和在一起。这就是为什么矢量搜索和矢量数据库受到如此多的关注，也是为什么“检索增强生成” (RAG) 是改进和增强 LLM 及生成式 AI 模型的重要技术。此外，我们看到业界涌现了很多创新技术，而且大量都是开源社区推动的，包括开发者最常用的数据库之一 PostgreSQL。



组织希望从依托 AI 技术的数据库中 获得什么？

在数据库中使用 AI 时，最重要的功能包括与 AI 模型无缝对接、使用 RAG 等技术实现 LLM 标准答案关联，以及使用自然语言进行数据库管理。



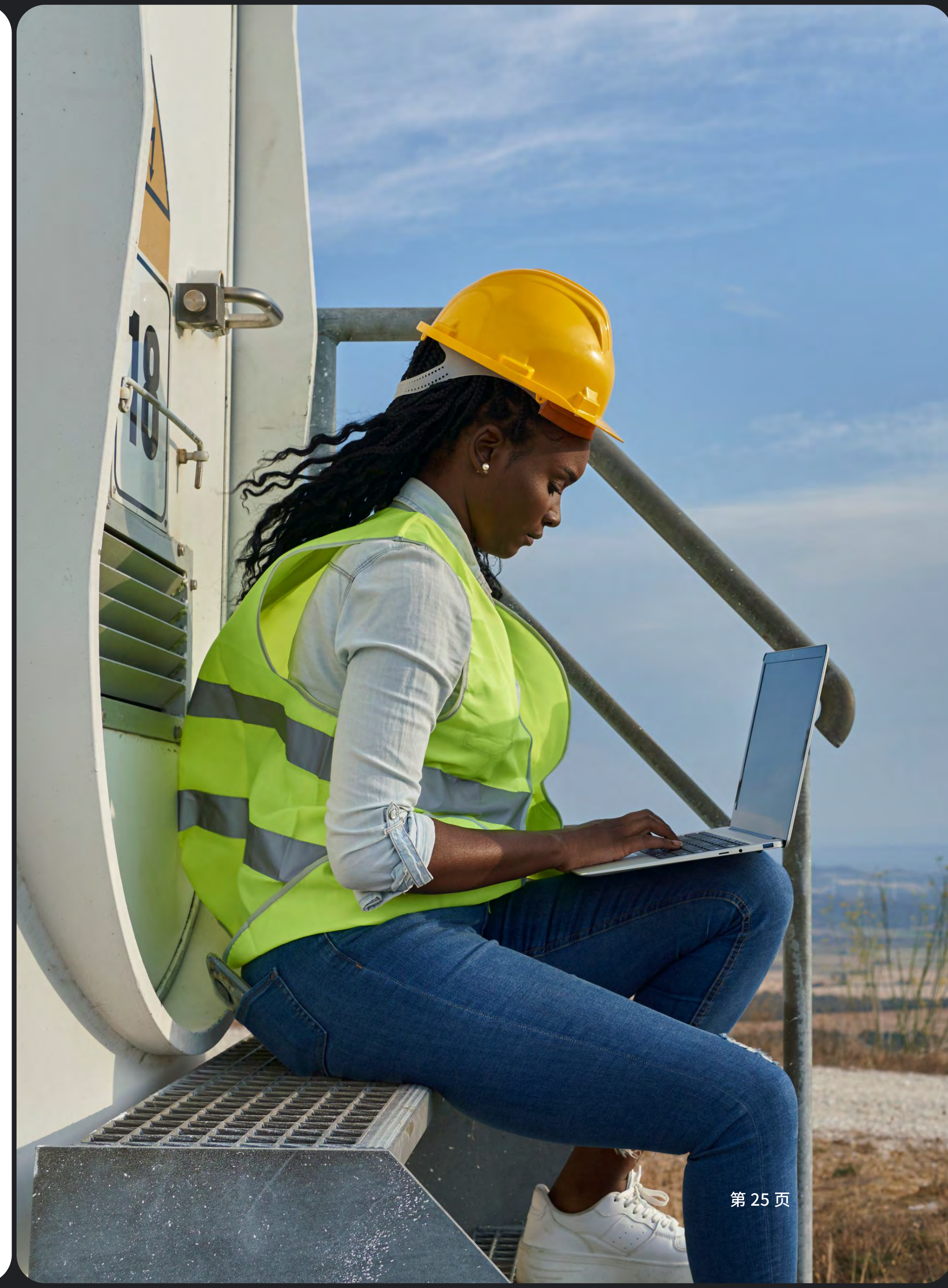
合理运用，获得丰厚回报。

当运营数据与生成式 AI 紧密集成时，生成式 AI 的真正力量就会被释放出来，在企业应用中实时提供与情境相关的超个性化体验。简而言之，依托生成式 AI 技术的运营数据库以及所存储的相关业务数据将是在企业中发挥生成式 AI 优势的关键。

成功的数据库将以 AI 优先的方式发展，并深度整合包括下列技术在内的诸多技术：矢量搜索、与 AI 模型的无缝对接、对 SQL 的自然语言支持以及与 AI 工具和开源框架的紧密集成。所有这些技术都将成为内置于运营数据库中的标准功能。

未能集成生成式 AI 的数据库可能会被淘汰。

将 AI 与运营数据进一步整合，也让开发者能够运用所有可用数据实现快速迭代，并改善数据处理体验。由于现有数据库已在驱动组织的所有应用，组织可以直接在目前存储数据的位置执行这些操作，无需另外设置或学习使用全新系统。这可以大幅提高成本效益。此外，得益于 PostgreSQL 等开源技术，开发者可以使用熟悉的工具和功能快速上手。

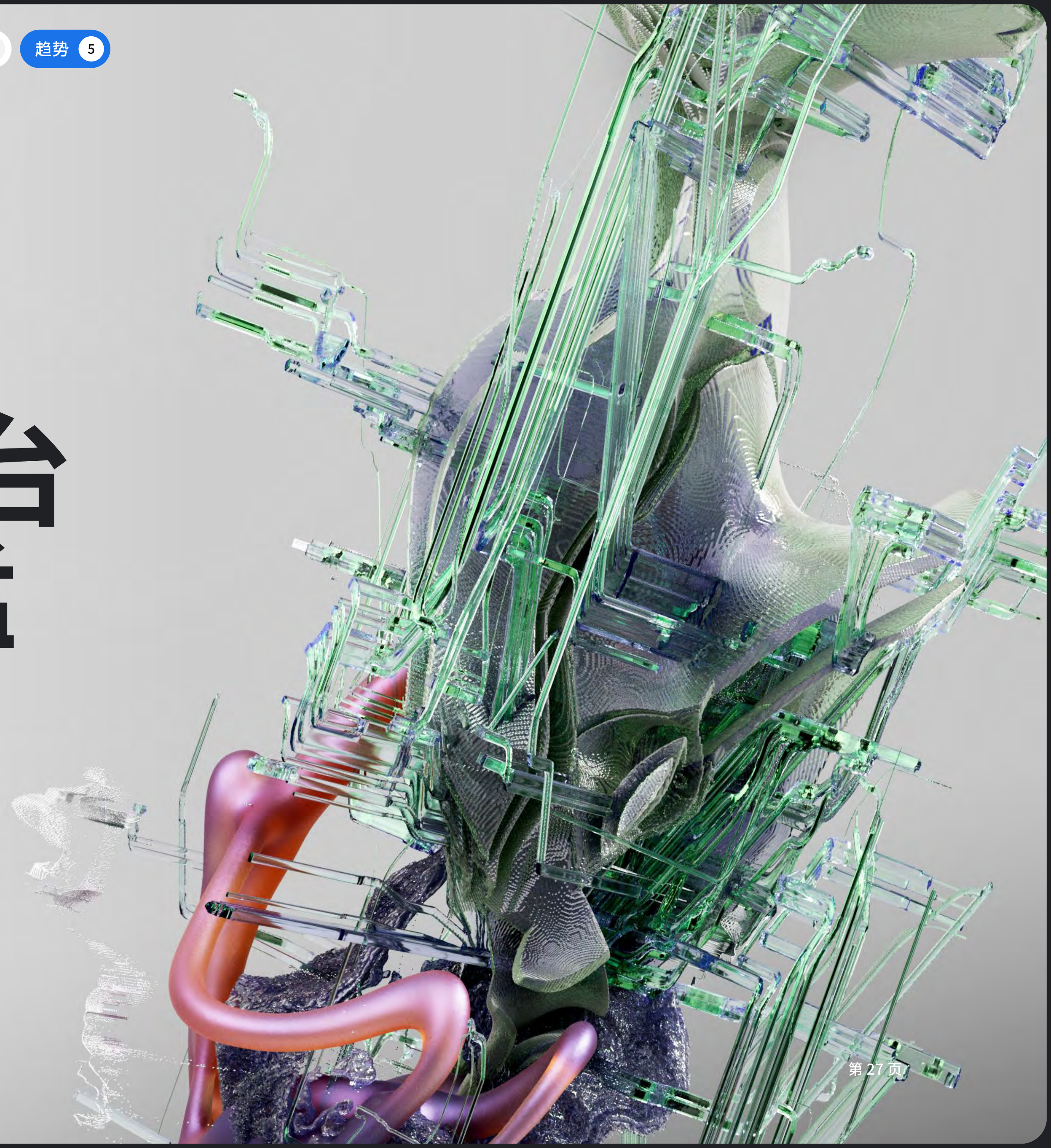




“我们研究了数据库市场上一些专注于矢量存储的新产品，然后试用了其中的一些产品。考虑到 Linear 现有的数据量以及我们寻找经济高效的解决方案的目标，在 Cloud SQL for PostgreSQL 增加了对 pgvector 的支持后，我们选择了使用这个产品。它的可伸缩性和可靠性令我们印象深刻。而且该产品也与我们目前使用的数据库、模型、ORM 等兼容，因此我们的团队很快就上手了。”

TOM MOOR
Linear 美国工程部主管

2024 将是数据平台 现代化改造突飞猛 进的一年。



只有 1% 的组织对其旧数据库的 AI 支持状况非常满意，这表明还有很大的改进空间。



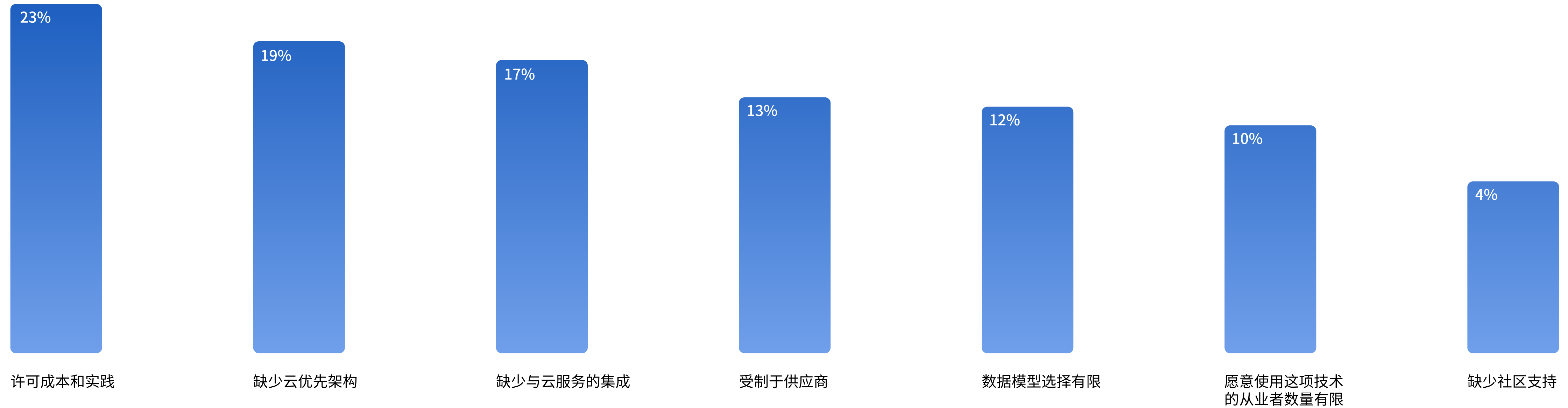
为什么您应该关注？

许多组织纷纷尝试把握生成式 AI 带来的机会，却发现其旧数据库成为了绊脚石，因为这些数据库的技术已经落后，且用户体验不佳。除了过时的技术和不佳的开发者体验，高管层也注意到，由于价格高昂、许可不便、受制于单一厂商，这些老旧数据库导致每年产生不必要的成本，金额往往高达数百万美元。

随着生成式 AI 的蓬勃发展，对数据库进行现代化改造的新需求迫在眉睫。这是因为用于处理矢量、模型和数据的流行 AI 工具都在云端运行，并且基于开源数据库技术，例如 PostgreSQL。此外，最先进的 AI 模型只能在主要云平台上运行。



旧数据库的主要问题。



平稳过渡的可能性从未如此之大。

所幸，随着数据库迁移工具和计划的不断改进和成熟，从旧数据库迁移变得越来越容易。我们还注意到，AI 有助于增强这些工具。受益于 AI 辅助的代码转换、代码补全和更高的效率，现在组织更容易挣脱旧数据库的束缚。

IT 决策者现在可以放心地批准大规模的现代化改造项目，因为他们希望在创新过程中采用开放式技术（包括生成式 AI）。



“数据迁移工具已经存在很长时间了，但最近，它们变得更加智能，能够进行 AI 辅助的代码转换和代码补全。迁移过程中最困难的一环是转换数据并让新应用适应新数据库。有了生成式 AI，这些问题就可以迎刃而解。您可以使用模型来分析源数据库，并找出如何将数据转换为目标数据库的格式。您能够快速取得成果并最终提高开发者的效率。虽然仍有大量遗留数据，但生成式 AI 正在降低门槛、简化迁移。”

ANDREW STORRS
Aritzia 数据工程副总裁



“Character.AI 是开放式对话应用设计与开发领域的引领者。我们的生成式 AI 平台利用我们自己先进的神经语言模型，来生成类似人类的文本响应并进行符合当前情境的对话。发现 AlloyDB for PostgreSQL 的时候，我们正在困境中苦苦挣扎。我们服务的使用量呈指数级增长，给我们基础设施的各个部分（尤其是我们的数据库）带来了极大的压力。Google Cloud 的 AlloyDB 和 Spanner 提供了坚实的基础，可为从互动和运营到 AI 和分析的各种工作负载，带来可靠性、可伸缩性和性价比。”

JAMES GROENEVELD
Character.AI 研究工程师

**整整一代开发者都在构建 AI 应用，并利用 AI 更高效地编程、改进数据库性能数据分析并改善安全状况。
您是否也是其中的一员？**



Google Cloud 能提供怎样的帮助。

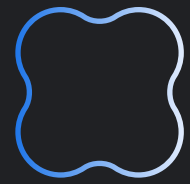
Google Cloud 可帮助组织整合数据并将其与突破性的 AI 技术对接，以提供变革性的数据洞见和个性化体验。利用 Google 在数据和 AI 方面的统一方法所具有的简易性、可伸缩性、安全性和智能，企业可以通过单一、简化的解决方案释放数据的全部潜力。

Google 数据云可整合工作负载并管理整个数据生命周期。这样，数据团队就能使用常见的开源引擎和模型开发现代数据驱动型应用。

此外，我们还提供先进的、随时可用于处理数据的 AI/机器学习和生成式 AI 功能，帮助组织中的每位员工快速、轻松地访问他们所需的数据，以发挥数据的真正价值。所有这些功能都具有企业级效率。这种独特的融合使得 Google Cloud 成为了卓越的合作伙伴，可助力将原始数据转化为组织价值。



深入了解我们的独特方法。



数据与 AI 全面对接。

随着与系统和数据交互的新方式的出现，组织显然需要基于高质量企业数据的 AI 模型，以便提供数据分析洞见和增强的体验。

Google 的数据云让数据团队能够使用生成式 AI 工具跨 BigQuery 和 AlloyDB 激活企业数据，并使用内置功能轻松将 AI/机器学习应用于数据。例如，数据团队可借助 BigQuery ML 使用 SQL 直接通过 BigQuery 数据构建机器学习模型，甚至可以从 Vertex AI 调用基础模型。AlloyDB 的内置矢量嵌入功能还允许用户在其数据存储区中存储和生成嵌入，以帮助增强 LLM 并支持生成式 AI 应用场景。



统一的数据基础。

Google Cloud 的统一数据基础基于 BigQuery 构建，可将数据集中到一处，并将结构化和非结构化数据与 AI 集成，以提供涵盖所有数据资产的数据洞见。这个统一的数据基础可管理整个数据生命周期并让组织中不同类型的用户轻松进行数据访问、管理、治理和分析，从而帮助您有效消除数据孤岛。

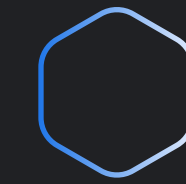
高度可扩展的架构可以整合事务和分析系统，以支持跨 BigQuery、AlloyDB 和 Spanner 紧密集成的数据服务。这样您就可以轻松地执行从 Spanner 到 BigQuery 的数据分析，并且对底层事务工作负载几乎没有影响。



助力现代化改造的开放数据平台。

Google Cloud 致力于成为最开放的云提供商，帮助客户构建现代化的数据驱动型应用，无论其工作负载位于何处。我们支持开源和开放标准，并且提供与广受欢迎的开源引擎和模型完全兼容的托管式数据库服务。

AlloyDB Omni 和 BigQuery Omni 可帮助您消除与数据迁移相关的成本、安全风险和治理疑虑，并利用您的数据跨 Google Cloud、AWS、Azure 和 Google Distributed Cloud 实现应用现代化改造。现在，在数据平台上开始使用生成式 AI 比以往任何时候都更容易，无论您处于现代化改造之旅中的哪一环，都能获得相应支持。



大规模企业级效率 and 安全性。

Google 数据云在效率、安全性和规模方面处于行业领先地位，能够为各种规模的组织提供服务并满足最严格的企业要求。

我们可帮助组织跨组织边界安全地分享数据，以超快的速度查询 EB 级数据，并以普遍较低的成本轻松处理数十亿项事务。

如何迈出 下一步？

显然，2024 年对于许多组织来说将是令人兴奋且非常重要的一年。能够让人员和平台做足准备以充分拥抱生成式 AI 带来的新能力的组织，不仅可以实现短期生产效率的提高，还可以开始高效地针对未来未雨绸缪，以适应日益激烈的竞争。

当然，许多新机遇需要新技能并优化现有流程才能从容掌握。认识到需要提高员工技能并为其做好充分准备的组织，将很快意识到这项投资将以近乎无限的潜力获得回报。



准备好行动了吗？

如果您对本报告的内容有任何疑问，或者想详细了解 Google Cloud 能为您的组织提供哪些支持，可以随时与我们的专家联系。

 [联系我们](#)

调研方法。

Google Cloud Customer Intelligence 团队于 2023 年 12 月 18 日至 2024 年 1 月 17 日对 410 名数据决策者开展了一项关于数据和 AI 趋势的全球研究。冬季假期期间，即 12 月 23 日至 1 月 1 日期间暂停了招募活动。受访者包括数据、IT 和业务领导层，职位范围涵盖高管层到经理。所有受访者均受雇于拥有 1,000 名或以上员工且目前使用数据产品和服务的组织。受访者不知道 Google 是这项研究的赞助方，并且 Google 并不知晓参与者的身份。

区域

北美地区	180
欧洲、中东和非洲地区	104
日本和亚太地区	76
拉美地区	50

行业

金融服务	74
零售（例如杂货店、商店、精品店、特许经营店、餐馆等）	42
科技	85
其他	103

公司规模

1,000 至 4,999 人	19%
5,000 至 9,999 人	20%
10,000 至 49,999 人	30%
50,000 人以上	31%

职位级别

高管层	14%
副总裁或同等职位	20%
总监	43%
经理	7%
主管/负责人	4%

与数据产品和服务的交互

实操	40%
战略/监督	60%

角色

业务开发	2%
IT 或 IS（信息技术、计算机工程、安全等）	40%
软件开发	1%
技术战略或产品开发	11%
营销/广告/公共关系	16%
运营	3%
产品管理	3%
研究/数据分析/战略规划	6%
销售	1%
数据科学	15%



现在，许多数据分析师逐渐开始承担以前由数据科学家负责的业务，反之亦然。

当前，各组织最欠缺的技能有数据分析、数据工程、AI 以及业务分析。生成式 AI 可帮助提升现有数据团队和工作负载的效率，进而弥合日益严重的技能差距。无缝使用数据和 AI 平台可以帮助组织加速从数据到 AI 的历程，从而提高效率并加速创新。

此外，数据和 AI 工具的互联程度越来越高，可以帮助用户简化数据和 AI 工作流。

这种互联让用户可以：

拥有一个为数据工程师、数据分析师和数据科学家提供多种编程语言（如 SQL、Python 和 Spark）支持的共同工作空间。

使用 CI/CD、版本记录、源代码控制等软件开发最佳实践来管理数据资产，以提升协作和交接效率。

研究表明，超过一半 (54%) 的数字化领先企业表示，由于缺乏技能，他们难以跟上变革的步伐。

Nash Squared, 《Digital Leadership Report》
(数字领导力报告), 2023 年。

