

ORACLE

数据和 AI： CIO 的成功指南

—
实施全方位企业策略，
提高生产力并解锁创造力



目录

为什么要实施覆盖全公司的 AI 数据策略?	3
提供 AI 就绪的 IT 基础设施	5
AI 就绪 IT 堆栈的五个要素	7
快速成功:使用业务应用中内置的 AI 功能	9
快速成功:确保您的所有数据和文档都可搜索	10
打造 AI 梦想团队	12
您需要考虑的数据和 AI 基础知识	14
让开发人员能够充分利用生成式 AI	15
为什么要选择 Oracle 作为合作伙伴?	16

为什么要实施覆盖全公司的 AI 数据策略？

人工智能 (AI) 激发了人们的想象力。主流大语言模型 (LLM) 的新迭代版本不仅优化了人机交互，还增加了灵活性，让整个企业的领导都在畅想各种使用场景。

首席信息官 (CIO) 可以通过统一的数据平台来支持更安全、更高性能的 AI 实施，从而发现有效的 AI 使用场景，避免影子 IT 问题，为企业带来优势。

随着 AI 就绪的数据平台的推出，企业迫切需要从企业运营数据和领域专业知识中获得更多价值。AI 可以让用户使用自然语言与深入了解组织数据的 AI 模型交互，大大降低了企业上下所有人查询和使用业务和产品知识库的难度。

如果您的 IT 团队和数据分析专家在数据仓库、数据湖和数据湖仓一体中完成了基础工作，并实施了访问控制和数据治理的规则，这意味着您已经成功迈出了第一步。AI 就绪的数据平台包含了一体化的技术和业务流程，可识别、微调和部署复杂的 AI 模型，确保您充分利用所有的数据。

尽管消费者 LLM 和生成式 AI 应用激发了全球领导者的想象力，但您作为 CIO 的责任却十分明确：运用计算能力、技能集以及整个企业的数据流，真正发挥出 AI 对企业承诺的优势。

Oracle AI 堆栈

AI 堆栈使用云技术服务提供了大量的 AI 应用,助您充分利用您的数据。

Oracle AI 堆栈包含了一系列的组件,这些组件与 Oracle 的云基础设施和数据管理平台相集成,可帮助企业构建、部署和管理 AI 驱动的应用和服务。此外,该堆栈也包含了强大的安全和隐私功能,有效帮助企业保护敏感数据并满足监管要求。



提供 AI 就绪的 IT 基础设施

许多企业高管都想知道,复杂的 AI 模型如何提高团队的生产力、效率和创造力。CIO 可以通过从创意构思到生产就绪的清晰路径来帮助高管找到答案。

首先,我们要从可用的开源和专有 AI 模型中选择产品并进行自定义。大多数的企业都会根据自身需求对模型进行优化。从零开始构建新的生成式 AI 模型会产生一定的费用和技术难度,这些挑战可交由该尖端专业领域的科技公司来应对。

幸运的是,自适应生成式 AI 模型的市场正在不断增长和成熟。企业越来越容易找到支持微调和增强的模型,以确保模型与您的数据和业务策略保持一致。如果妥善使用,现成的模型还可以成为量身定制的 AI 中心。

如今,市场上涌现了更具针对性的机器学习 (ML) 模型,支持用户轻松访问计算机视觉识别和语音识别、异常检测、预测和其他 ML 功能。您的内部 AI 计划必须要能够帮助开发人员轻松将 ML 模型应用于应用和业务运营中,并支持开发人员跨服务重用模型、数据集和数据标签。

一个适合的数据科学平台将有助于您推出 AI 功能。这个平台可为数据科学家团队提供一个识别、部署和管理 ML 模型和 LLM 的地方,并通过各种 ML 操作功能(例如自动化管道、模型部署和模型监视)确保系统正常运行。

您可以从小项目开始实施,然后再慢慢扩大规模。类似的使用场景包括将 AI 添加到当前业务应用,以及采用 AI 驱动的工具,让员工更容易访问和使用知识库。

您不可不知的 AI 术语



人工智能 (AI) 是一个广义术语,指的是能够执行曾经是纯人工任务的机器。AI 技术包含了许多子技术,例如机器学习、深度学习和生成式 AI。



机器学习 (ML) 模型是基本的数学模型,允许计算机在没有特定指令的情况下学习完成任务。ML 可驱动异常检测、预测和语音识别等服务。

此外,企业还可以使用针对特定行业的策略。

以[一家使用 AI 的连锁杂货店](#)为例,该企业需要动态调整产品定价,进行个性化促销,使用能够检测客户购物车中的货品的手持式扫描仪,然后在客户离店时向客户收取费用。



深度学习属于 ML 的一个子集,以神经网络为底层技术。这些神经网络是根据人脑工作方式建模的多层网络。深度学习是大语言模型 (LLM) 和更复杂的 AI 系统功能背后的核心技术。



生成式 AI 可以生成新内容。此技术结合了多种 AI 技术,包括 ML 模型、神经网络和深度学习,能够理解和模仿人类的语言、艺术和决策。LLM 是生成式 AI 的其中一个例子。

AI 就绪 IT 堆栈的五个要素

底层 AI 基础设施到位后, CIO 就可以帮助员工充分利用数据, 为企业铺平成功之路。那么, CIO 应该要怎么做呢?

- 1 一切从简。**首先, 您需要了解您的企业已经具备了哪些 AI 功能。某些企业业务应用提供商会将有用的 AI 功能添加到他们的应用中, 您可以充分利用这些功能。此外, 某些数据库提供商也会在数据库中内嵌 AI 和 ML, 从而将 AI 功能整合到您的数据中, 并大大简化所需的架构。这些都是快速取得成功的方法, 因为员工可以继续使用已经熟悉的工具。
- 2 构建单一平台, 支持 AI 模型协同:**确保您的数据科学团队拥有一个能够协作处理 AI 工作流的中央平台。软件或基于云技术的平台有助于引导团队完成从数据准备和清理到模型训练、部署和推理的各个步骤, 同时为技术人员和业务人员提供访问、比较、评估和改进模型的方法。此类平台将利用您的数据仓库、数据湖和数据湖仓一体, 并维护这些资料档案库的所有数据治理和访问角色。
- 3 构建模型库:**AI 模型具有多种大小和不同程度的复杂性, 包括 LLM 和更简单的 ML 模型, 适用于文本分析、计算机视觉识别和异常检测等服务。正如开发人员在构建应用时需要依赖框架和库一样, 您的开发人员和数据科学团队也需要一个存储库, 其中必须包含成熟的 ML 和深度学习算法以及定制的 LLM 版本。
此存储库通常是您 AI 工作流平台的一部分, 团队可以为这些模型创建插件, 并将其连接到数据库和其他后端系统。

4 用您的文档和数据来补充 LLM。围绕着 AI 和企业的技术正在快速发展中。目前,有两种新技术可以使 LLM 的响应更准确并且与您的组织更相关,分别是[检索增强生成 \(RAG\)](#)和[向量搜索](#)。这两种技术可使用新的文档和运营数据来丰富 LLM,从而指导和优化 AI 输出结果。借助这些技术,制造商可以使用生成式 AI,帮助员工和合作伙伴以全新的方式深入了解产品信息,并通过文本或语音提示进行查询和探索。同样地,零售商也可以将仓库和 CRM 数据开放给 LLM,帮助客户服务代表了解每个客户的案例并快速解决问题。

5 实现针对流数据的实时 AI 推断。如果要将 AI 融入日常运营,企业通常需要为 AI 推理进行流数据处理。AI 推断指的是,您的数据科学团队需要使用大量精选数据,针对特定任务对 AI 模型进行训练,然后将该模型迁移到生产环境中,并基于新发现的数据进行推断或总结。许多适用于日常运营(包括物流、客户服务等)的有效 AI 工具都涉及了流数据处理,以便 AI 能够基于最新数据生成结果。视具体使用场景而定,流式 AI 推断的数据一般会来自您的业务应用、技术和网站日志、基于位置的数据或者是其他潜在来源。

什么是训练数据?

“训练数据”是指用于训练 ML 算法的带标签或带注释的数据集。这些数据集通常非常庞大,结合了一个规模较小但高度相关的内部数据集,以及一个来自现成在线资料库的预标记和带注释的数据。

经过格式化的完整训练数据对 AI 模型的开发和性能至关重要。您的企业很可能会选择对预训练的 AI 模型进行自定义,把从零开始的训练过程交给资金充足的内行专家来完成。

重要的是,只有您的企业才能利用您的数据和专业知识来构建独特的 AI 工具或服务。

快速成功:使用业务应用中内置的 AI 功能

Oracle 和某些企业 SaaS 应用提供商(包括 ERP、HR、CRM 和供应链管理应用的提供商)已经为常见使用场景构建了嵌入式 AI 模型。您可以利用这些现成的 AI 实施,并研究它们是如何构建的。

- **在您的 ERP 应用中**, AI 如何在核心业务流程中提高效率并减少人为错误?
- **在您的 CRM 应用中**, AI 使用什么标准来识别高潜力的销售线索并为销售代表提供推荐方案?
- **在 HR 系统中**, AI 如何通过自动执行步骤来缩短招聘时间?或者, AI 如何通过辅助员工完成绩效评估和职业发展计划等任务来改善员工体验?

借助业务软件中包含的 AI 功能,员工可以轻松了解和信任 AI,相信 AI 能够帮助他们更好地完成工作。



快速成功:确保您的所有数据和文档都可搜索

通过面向消费者的 LLM 和由 LLM 支持的搜索引擎,人们已经习惯使用那些能够从大量在线数据中获得答案的 AI。现在,越来越多的公司(包括 SoundHound 和 Cohere)为企业用户提供了一种安全的方式来访问类似的自然语言界面。因此,员工和合作伙伴可以提出与运营数据相关的问题,例如文档的深度归档和在线通信。以前,只有那些懂得使用精心设计的仪表盘的分析人员或熟练掌握编程语言的开发人员才能回答的问题,现在员工和合作伙伴也可以轻松找到答案。

下面,我们将为您介绍如何安全地结合使用特定域的数据与生成式 AI 的强大功能。

首先,向 LLM 提供数据

如前所述,RAG 和向量搜索有助于使用当前组织数据来丰富 LLM 的输出结果,企业无需支付高额费用来修改底层模型。这些信息可能来自您的结构化数据库和非结构化文档,例如 PDF、网页和博客、Word 文档、新闻源、通话记录和在线通信。但凡是对企业的绩效很重要的数据,您都需要将其添加进来。

这些数据源将存储在知识库(例如向量数据库)中,以便用于快速检索,然后再添加向量,以供 AI 模型使用。向量是一个易于搜索的数字字符串,用于表示内容中的某个细节,例如照片中的人员或文本文档的主题。当某个人进行查询时,系统将通过搜索向量在数据库中查找相关信息,并提供与查询相关的情境。此情境化信息以及查询将发送到大语言模型,该模型将使用相关情境来创建及时、准确且符合情境的响应。

然后, 引用信息来源以提高对 AI 输出结果的信心

向量数据库和知识库包含了元数据, 可清楚显示信息的来源。因此, LLM 生成的结果引用了信息出处, 让您可以轻松快速地删除不需要或不准确的信息源。



您的使用场景, 决定了您的企业如何将 LLM 作为数据的“代言人”。

销售人员可能会询问 LLM 有关某个条件下的销售成功案例, 以及公司最近与追加销售潜在客户互动的详情。客户服务座席则可能会通过 LLM 来查看产品文档, 以找到解决方案。诸如此类的例子还有很多。事实上, 在您阅读本文时, 企业上下可能都在设想各种各样的使用场景。



打造 AI 梦想团队

AI 人才的需求量很大,特别是在医疗卫生、金融和制造业等领域。如果要引入生成式 AI 企业,您可能需要各种技术、组织技能和人际交往技能。因此,您需要与 HR 合作组建团队,确保团队具备以下技能:

- ✅ **关键的编程语言:**包括 Python、R、Java 和 C++。团队需要需要结合使用这些编程语言,从而构建、定制和实施模型以及访问外部 API。
- ✅ **数据库专业知识:**面向 AI 的数据平台涉及了数据建模、数仓和集成
- ✅ **ML 专业知识:**许多 AI 模型使用 ML 算法来处理和分析大量数据、识别模式并根据所观察到的模式做出智能决策或预测。数据分析团队的 ML 专家将专注于定制或增强生成式 AI 模型。
- ✅ **智能用户界面(包括语音)的知识:**设计和实施用户体验,让人们在日常工作流中使用文本或语音提示,即可与生成式 AI 模型直接交互,无需进行编码。



什么是多模态 AI?

多模态 AI 同时使用多种数据类型来获得对环境的细微感知,类似于人类通过五种感官来感知世界。通过将视图、音频和文本等数据类型相结合,AI 能够为一系列实际应用提供准确的情境化结果,其中包括图像描述、文本到图像生成甚至是情感识别。以自动驾驶汽车为例,这种汽车可以结合图像识别和道路标志识别,在城市街道中实现自动驾驶。

对于数据管理团队和数据科学团队而言,这意味着他们需要使用来自业务应用的视频、音频、语音、图像、文本和数字数据集来定制 AI。因此,多模态 AI 训练基础设施需要经过特殊的转型和采用独立的架构,以输入、编码、整合并最终输出有意义的预测结果。

除了这些硬技能之外,您还需要坚定的合作伙伴。为了促进 AI 的采用,您需要确保执行发起人能够帮助团队快速识别和解决生成式 AI 模型带来的问题,并为此提供合适的资源。

充满热忱的倡导者、数据专家和产品经理将能够帮助员工进行数据管理和利用 AI 功能。

AI 建模和生成式 AI 实施领域在不断发展。因此,除了寻找和招聘特定职能所需的技术专家之外,您还需要培养学习文化并建立变革管理团队,以确保每个人都了解自己在 AI 驱动的新世界中的角色。

您需要考虑的数据和 AI 基础知识

您的数据是否“AI 就绪”?或者说,您的数据是否经过正确的编目和标记?能否跟踪其系属?如果答案是肯定的,您的数据科学家就可以使用这些数据来进行 AI 建模了。标签对 AI 输出结果中的数据隐私和安全也至关重要。

以下是您在制定策略时需要考虑的其他问题。

- **您的数据是否可以跨业务职能实现集成?**数据是 AI 的基础。企业必须不同的职能部门中确保数据收集和存储的一致性,才能获得有意义的成果。各个部门,包括销售、营销、客户服务和采购倾向于创建各自的孤岛数据。这种方法在 AI 出现之前可能具有意义(也可能没有意义),但放到现在,这种做法将会影响 AI 模型的有效性。
- **您的数据治理是否有序?**每个数据域是否都有一个所有者?是否具备自动报告结构,以证明您满足监管机构的合规要求?
- **数据从何而来?**为了推动 AI 模型实用程序的发展,许多企业利用来自在线市场、合作伙伴和公共机构等第三方数据来补充内部数据,这些数据源通常可以提供有关位置、天气、人口统计等方面的有用信息。AI 将会基于所有这些数据输出结果,您又该如何跟踪这些结果的适当性、准确性和实用性?
- **您的数据将会流向何处?**如果您的数据会流向第三方 AI 模型的托管服务,您如何评估并消除 AI 输出结果意外泄漏的风险?就算是数据仍保留在企业内部,您如何防止泄露敏感信息?

让开发人员能够充分利用生成式 AI

开发人员渴望在他们的工作中使用 AI。您可以选择一种数据管理策略, 为开发人员提供低代码工具, 支持他们快速创新。新的平台正在将提示工程提升到新的高度, 进一步简化了数据专业人员、开发人员和用户与数据交互的方式。借助合适的平台, 开发人员可以使用自然语言生成应用或 SQL 查询, 再也不需要手动编写代码了。开发人员只需要描述预期结果, 就能获得应用生成的初稿, 彻底告别了手动编码, 大大提高了生产力。



为什么要选择 Oracle 作为合作伙伴？

利用数据推动业务成功是 Oracle 的重中之重。现在，生成式 AI 为企业打开了新的一扇门，企业可以从组织数据中获得更多业务价值，进而凸显了这一重点与企业的相关性。

Oracle 提供丰富的工具、计算能力和专业知识，可帮助 CIO 及其数据管理团队制定和执行覆盖全公司的 AI 策略，从而实现短期成功和长期创新。

Oracle 提供一系列全面的云技术服务，包括一整套云技术应用、数据库和新一代云基础设施。这种独特的组合，意味着新的生成式 AI 功能已嵌入到企业日常使用的业务应用中。事实上，Oracle 已将 AI 功能集成到 Oracle Fusion Cloud Applications、Oracle Database 和 MySQL HeatWave 中。通过使用 Oracle 新推出的 AI 创新，您将能够提高生产力、实现端到端业务流程自动化、改善决策并降低业务成本。

如果您需要基于数据构建 AI 驱动的新服务，Oracle Cloud Infrastructure (OCI) 数据平台为您提供了一个安全的地方来收集、整理和管理所有事务处理数据和数仓数据，让您可以从数据库内机器学习的速度中受益。Oracle 支持所有类型的数据，包括实时、流式、关系型、JSON、文档、空间、图、非结构化和开源数据源的数据。

无论您是在寻找本地部署、混合部署、监管还是公有云技术解决方案，Oracle 的产品都可以满足您的数据和 AI 策略需求。

了解更多

版权所有 © 2024, Oracle 和/或其附属公司。本文档仅供参考，所含信息可能随时修改，恕不另行通知。本文档不保证没有错误，也不受其它任何口头表达或法律暗示的担保或条件的约束，包括对特定用途的适销性或适用性的暗示担保和条件。我们特别声明拒绝承担与本文有关的任何责任，本文不直接或间接形成任何契约义务。未经我们的事先书面许可，不得出于任何目的以任何形式或任何方式（包括电子或印刷品）来复制或传播本文。Oracle 和 Java 是 Oracle 和/或其关联公司的注册商标。其他名称可能是其各自所有者的商标

