

InterSystems IRIS 数据平台加速人工智能

技术简报



前言

人工智能（Artificial Intelligence, AI）完全不是人为制造的仿品。在商业世界，那些积极拥抱 AI、使用 AI 改善自身业务流程从而变得更为灵活的企业与未能采用 AI 而半途而废的企业之间将会逐渐拉开差距。

敏捷性始于主动 现在正是开始改变的最好时机

人工智能、数据存储及处理性能的近期发展正显著改变着当今商业世界。越来越多的企业开始借助 AI 和大数据来交付创新服务、创建新的收入流、改善客户体验、简化业务流程。不仅老牌公司开始实施基于 AI 的新举措，就连初创公司也在借助 AI 打破行业旧格局，老牌公司承受的竞争压力越来越大。

麦肯锡公司记录了 19 个行业 400 多个运用 AI 中获益的案例¹。普华永道公司预测，到 2030 年，AI 将为全球经济贡献 15.7 万亿美元的收益²。Gartner 的报告显示，77% 的高级经理认为数据科学已经为他们的公司提供了显著价值³。

为了应对商业世界中正在发生的变化，InterSystems 已经开发出一个完整的数据平台，协助各领域用户加速、简化数据密集型应用程序的开发、部署和维护。

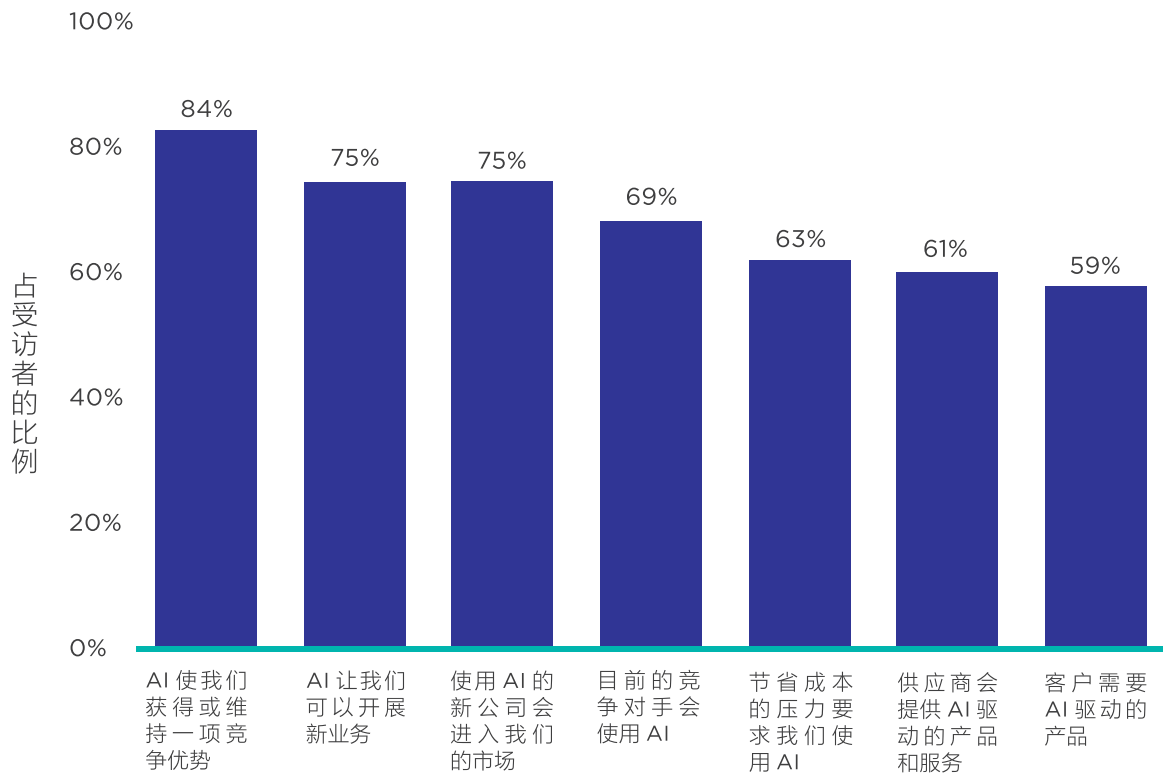
这篇文章详细描述了 InterSystems IRIS 数据平台™ 助力企业利用 AI 加速实施计划、产出价值的多种方式。

1. 麦肯锡，2018. <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning>

2. 普华永道，2017. <https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html>

3. Gartner, 2018. “数据科学和机器学习平台的魔力象限”

企业采用 AI 的原因











来源：Statista 网站 2017 年对全球超过 3,000 位业务主管的调查⁴

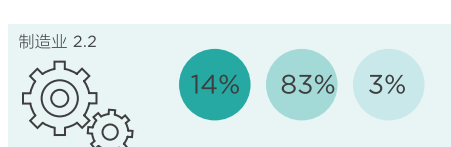
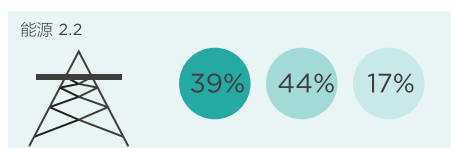
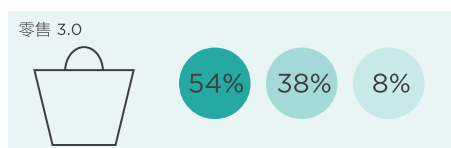
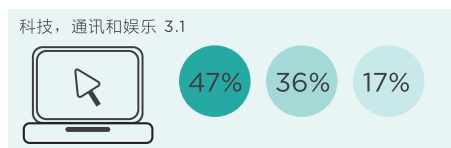
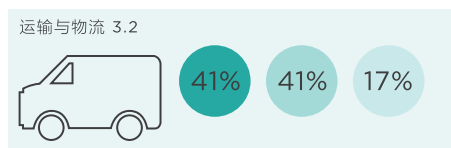
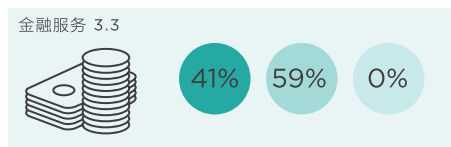
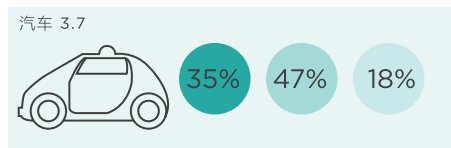
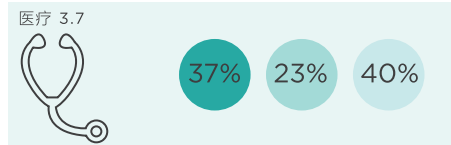
4. Statista. 2017. <https://www.statista.com/statistics/747775/worldwide-reasons-for-adopting-ai/>

各行各业采用 AI 的情况

行业 子行业 AI 对消费的潜在影响

行业	子行业	AI 对消费的潜在影响
医疗		3.7
	医疗机构 / 医疗服务	
	药物 / 生命科学	
	保险	
	用户健康	
汽车		3.7
	售后 & 维修	
	零部件供应商	
	个人移动服务	
	原始设备制造商 (OEM)	
金融服务		3.3
	资产财富管理	
	银行业与资本	
	保险	
运输与物流		3.2
	运输	
	物流	
科技, 通讯和娱乐		3.1
	科技	
	娱乐, 传媒和通讯	
零售		3.0
	消费产品	
	门户	
能源		2.2
	石油 & 天然气	
	电力 & 公用事业	
制造业		2.2
	工业制造	
	工业产品 / 纯原料	
总计		3.1

● % 采用成熟度——短期 (0-3 年)
 ● % 采用成熟度——中期 (3-7 年)
 ● % 采用成熟度——长期 (超过 7 年)



来源基于普华永道 AI 影响指数评价。潜在得分为 1-5, 5 代表 AI 产生了最高的潜在影响, 1 意味着产生最低影响。

已有 80 多个国家数以千计的企业应用程序部署了 InterSystems 的技术，每天处理超过 5 亿患者的电子病历和全球 15% 的金融证券交易。

除了全球性覆盖，InterSystems 还通过 WRC 全球响应中心为客户提供顶级支持。经验丰富的工程师 7×24 小时待命，我们的目标是理解您的整体业务，而不是仅仅关注当前的问题。我们致力于支持您的策略、规划、性能分析、迁移以及最终的成功。

加速 AI

虽然市场上已有许多 AI 工具和框架，但客户们仍然迫切需要加速应用 AI 以满足业务需求。InterSystems IRIS 数据平台可以提供多种能力来满足这一需求。

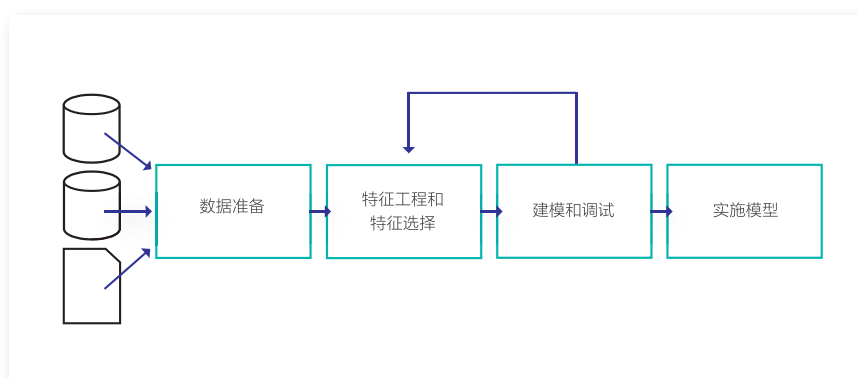
InterSystems IRIS 功能特性：

- 可横向扩展的混合事务 / 分析处理（HTAP）数据库管理功能，可在大型数据集上运行实时分析应用程序。
- 内嵌可运行结构化和非结构化分析和人工智能模型的能力。
- 为规模化 AI 特征工程而优化的灵活多模型数据库。
- 完整的互操作平台，能够集成数据和应用程序，进行下一步分析预处理数据，构建并执行复杂业务流程。
- 统一开发环境，简化复杂分析应用程序的创建及维护。

AI 流水线

开发和部署基于 AI 的方案需要遵循以下步骤：

- 数据准备
- 特征工程和特征选择
- 建模和调试
- 执行和实施 AI 模型



数据准备

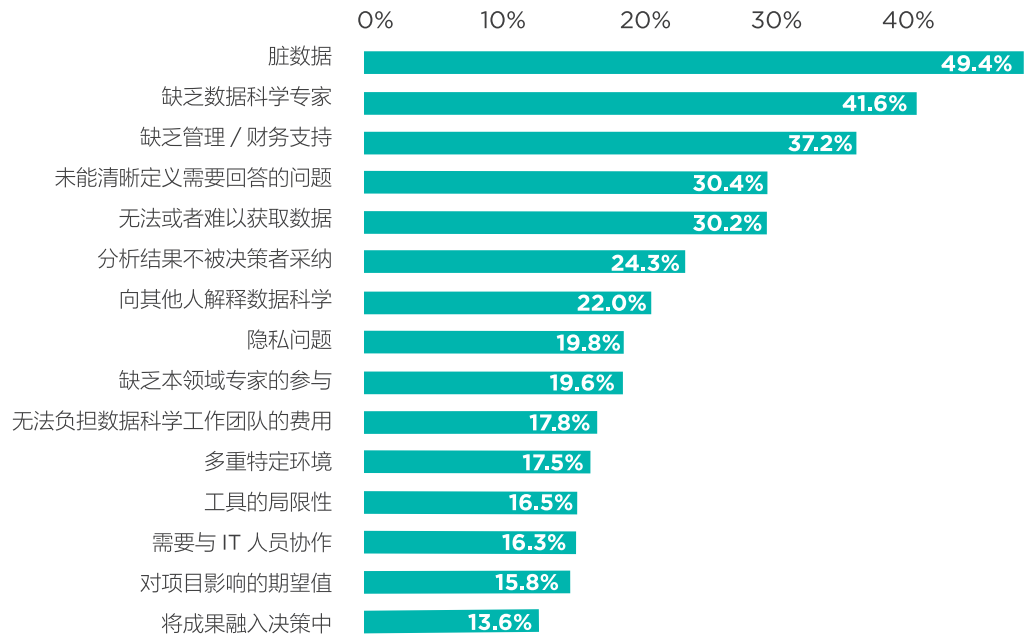
如今，企业累积了大量不同来源、不同格式并具有不同语义表征的数据。分析能纳入的数据越多，从中获取商业益处的机会就越大，而原始数据和脏数据不太可能产生准确结果。因此，为了充分利用 AI，必须对大量异构数据进行收集、集成和规范化，这是一项艰难的工作，IDG 信息显示，数据科学家通常需要在这些事上耗费 80% 的时间⁵。

在 Kaggle 网 2017 年“数据科学与机器学习现状”的调研中，7,000 多位从事数据科学工作的受访者称，他们最常遇到的问题就是处理脏数据⁶。

5. <https://www.infoworld.com/article/3228245/data-science/the-80-20-data-science-dilemma.html>

6. <https://www.kaggle.com/surveys/2017>

在工作中运用数据科学的阻碍因素



来源：Kaggle 网 2017 年对 7,276 名受访者进行的“数据科学与机器学习现状”调研。

虽然大多数数据科学家更愿意花时间在分析数据上，但现实却是绝大多数与 AI 相关的工作都离不开“数据整理”，即将原始数据转换成可用于 AI 的数据。

InterSystems IRIS 能帮助数据科学家更快速、更便捷地收集、集成、规范那些构建和调试 AI 模型所需的数据。InterSystems IRIS 提供一套完整的嵌入式集成功能，包括广泛适用于打包应用程序、数据库、业界标准、协议和技术的“开箱即用”的连接功能。对于来自多种数据源、应用程序和服务的数据存在的语义及数据模型差异，灵活的图形数据转换功能可以简化其解决过程。平台也支持广泛的行业标准（例如在医疗、金融服务、零售和通讯行业所使用的行业标准）、REST 架构和 Web 服务标准（例如 JSON, XML, XPATH, XSLT, SOAP 和 DTDS）。

此外，InterSystems IRIS 提供复杂的企业级安全保障，包括基于角色访问控制的安全性，具备单行粒度，可保障敏感数据的安全访问。这一特性使得将每一行与授权用户列表关联成为可能，确保只有获得批准的用户能够访问特定的数据项。

特征工程与特征选择

以良好性能生成准确结果的关键是识别和创建用于分析的最佳特征集。特征工程是将原始数据转换成一组与所需分析最为相关且格式可用于 AI 算法的单个属性的过程。

特征工程包括从原始数据中创建新变量或聚合变量，以及转换变量结构。相比于仅使用原始数据，处理后数据可以提供更为有用的信息。例如，一个新特征可以将每一位客户过去一个月的收入与每个客户过去 12 个月的收入之比关联起来。

创建新特征必须转换或编码一部分变量，诸如将数据结构从分类型转换成数值型。例如，美国每一个州（例如纽约州或加州）的出现频率也许与某一项分析相关，但算法可能不接受表示为具有多个字符串值的一个属性的数据。因此，转换或编码任务就是将包含 50 个不同字符串值（美国每一个州对应一个字符串值）的单个属性（美国下属的州）转换成 50 个不同的特征（每个州对应一个特征），每个特征具有与该州出现频率对应的数值。

另一个与特征工程相关的常见需求是从自由文本中提取意义和情绪。譬如，一份文档中某个概念出现的次数也许意味着这个概念的相对重要性，情绪分析可以判断文本的语气是积极的还是消极的。

InterSystems IRIS 能加速和简化与特征工程相关的各项任务。用户可以创建新的计算变量，它提供嵌入式商业智能（多维度 OLAP）功能和分布式 SQL 处理能力，从而支持高性能的大型数据集的复杂计算。一旦数据存储于数据库中，数据平台同时支持对数据进行关系型和非关系型访问，从而实现灵活性。如果处理成行成列的数据更为简单，开发者就可以使用 SQL 指令与数据互动。或者，可以用文档、对象或键值数据的形式来访问和操作数据。

InterSystems IRIS 提供内置的自然语言处理（NLP）功能，用于从文本中推断含义和情绪。它能自动识别文本中的概念和关系，无需前期准备工作或专业领域知识。这些功能使数据科学家们能够从文档、注释字段、社交媒体和其他来源中提取信息，为分析创建新特征。

“获取特征的是费劲、费时且需要专家知识。‘应用机器学习’基本上就是特征工程。”

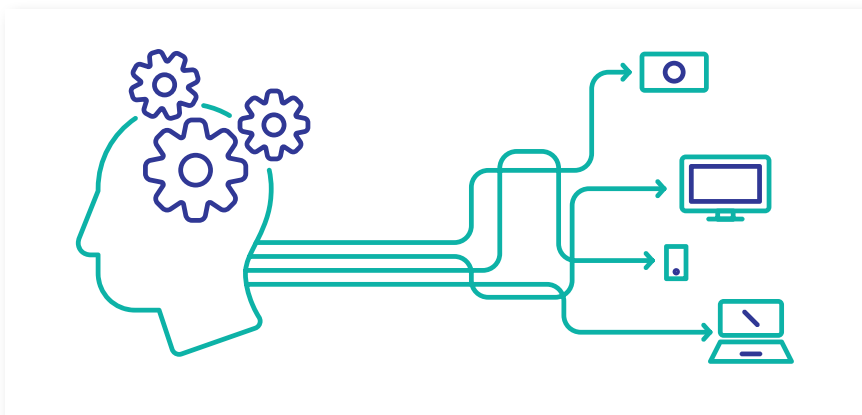
斯坦福大学，
Andrew Ng. 教授

特征工程可以生成数以万计（或者更多）的数据稀疏的大型矩阵。InterSystems 有一位来自医疗行业的用户，已经创建出 113,000 个不同的特征来支持 AI 患者风险模型的开发。为了以最高效率处理这类任务，InterSystems IRIS 将数据库中的数据存储为多维数组，并使用了横向扩展技术。因此，在特征工程任务中，InterSystems IRIS 利用极少的硬件资源便可提供高性能表现。

建模和调试

虽然 InterSystems IRIS 提供高性能 SQL 功能，但很多建模和调试任务都是使用专业工具来完成的。在 2017 年的调研中，Kaggle 网列出了 49 种常用建模工具并将 Python 列为最常使用的数据分析语言，R 和 SQL 紧随其后⁷。

除了支持 xDBC 协议和其他标准，InterSystems IRIS 还通过一个可识别分区原生连接器与 Apache Spark 直接集成。InterSystems IRIS 应用程序因此能纳入 Spark 处理流程，Spark 应用程序也可以从 InterSystems IRIS 中整合分布式数据。连接器充分利用 InterSystems IRIS 数据库的分区性能，这样一来，Apache Spark 工作节点可以自动直接连接到数据分区并对不同数据块进行平行处理。这些并行的、直接连接支持高数据吞吐量，并支持对水平分布集群的高速数据摄取，在数据输入或从应用程序中流入时，这一架构也使得模型可以进行实时增量式开发，使数据科学家能够随着业务条件和环境变量改变保持模型最新。



7. <https://www.kaggle.com/surveys/2017>

实施 AI 模型

一些应用程序只需要以批量处理模式独立运行 AI 模型。然而，很多其他应用程序需要将 AI 模型融入对事件、交易或用户行为做出实时响应的业务流程中。譬如，欺诈检测模型通常会被纳入实时信用卡交易处理系统中，用于识别和中断潜在的欺诈交易。

业界创造了“实施”（Operationalizing）这一名词来描述将 AI 模型纳入业务流程这一过程。此外，Gartner 在其最近发布的 2018 年度“数据科学和机器学习技术成熟度曲线”中创造了一个新术语——“持续智能”。Gartner 将“持续智能”定义为“已被集成到业务操作中、用于处理当前及历史数据并指导行动的实时分析。这一创新能实现决策自动化或提供决策支持。它使用多种技术，例如增强分析、事件流处理、优化、业务规则管理以及机器学习”。持续智能被 Gartner 评为“具有变革性”的少数几个 AI 子域之一⁸。

InterSystems IRIS 提供一整套工具，用于开发和执行富含实时数据和分析功能的业务流程，这些业务流程可以与内部和外部数据源、应用程序和服务相连通。它同时也提供图形工具，使流程、规则和工作流的图形化过程变得更为轻松，从而支持智能业务流程。

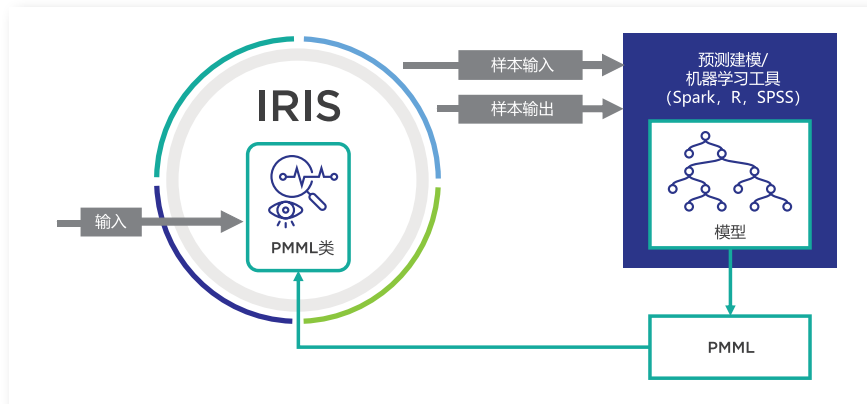
此外，内置的 HTAP 功能使得事件和交易能实时触发 AI 模型，并实现毫秒级响应。这是关键任务应用程序所必需的，例如在购买时检测欺诈消费者的行为、或者在呼叫中心员工处理客户需求时评估客户流失风险。

通过预测模型标记语言（PMML），InterSystems IRIS 应用程序也能纳入使用外部工具和应用程序创建的 AI 模型。PMML 是一种 XML 标准，它全面定义了使用外部 AI 应用程序或框架开发的 AI 模型所含的所有参数。大多数领先的 AI 工具和框架都支持导出 PMML 标准模型。InterSystems IRIS 包含一个内置的 PMML 运行引擎，在无需部署任何其他工具或跨系统传输数据的情况下，就能执行使用外部工具开发的模型。

利用这一结构方法，能够在复杂的实时业务流程中无缝执行 AI 模型。

8. Gartner, 2018 年度“数据科学和机器学习技术成熟度曲线”

实施 AI 模型，在 InterSystems IRIS 应用程序内执行模型



为规模化任务而构建

对于企业关键任务应用程序中的事务型工作（OLTP）和分析型工作，InterSystems IRIS 数据库管理功能已被证实能规模化地实现非常高的性能。它支持极高速的数据摄取，能使用商用硬件、基于实时数据和非常大型的非实时数据集（例如历史数据和参考数据）处理分析型工作。因此，对必须通过规模化识别模式和异常现象来进行实时决策的应用程序而言，InterSystems IRIS 是理想选项。即使工作负荷意外飙升，这一架构也能可靠地执行规模化任务。

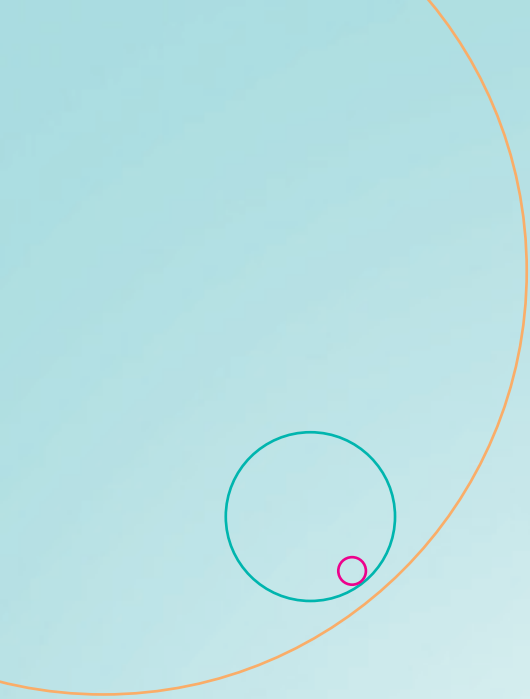
总结

InterSystems IRIS 是一个全面的数据平台，能加速和简化数据密集型实时分析型应用程序的开发、部署和维护。它提供数据预处理的关键功能以支持 AI 建模和调试任务，也提供用于将 AI 模型纳入实时业务流程的关键功能。通过上述功能，InterSystems IRIS 正帮助企业利用 AI 和大数据技术来创新服务、创建新的收入流、改善客户体验和简化业务流程来获得竞争优势——以及快速实现价值。

请选择 InterSystems IRIS 数据平台，快速启动并加速贵公司的人工智能项目。

InterSystems 是驱动世界上众多重量级行业应用的信息技术引擎。在医疗、金融、政府和诸多涉及国计民生的关键领域，InterSystems 为成就各项事业的未来而赋予其科技原力。InterSystems 创立于 1978 年，是一家总部位于美国马萨诸塞州剑桥市的私人控股公司。该公司在全球设立分支机构，其软件产品每天都在 80 多个国家被数百万用户所使用。

更多详情，敬请登录：<https://www.interSystems.com/cn/products/interSystems-iris>



InterSystems 中国

系联软件（北京）有限公司
北京市朝阳区建国门外大街乙 12 号
双子座大厦（东塔）1902 室
电话：+86 10-8524-9700
传真：+86 10-8524-9755

InterSystems.com/cn

为成就未来，赋予科技原力

