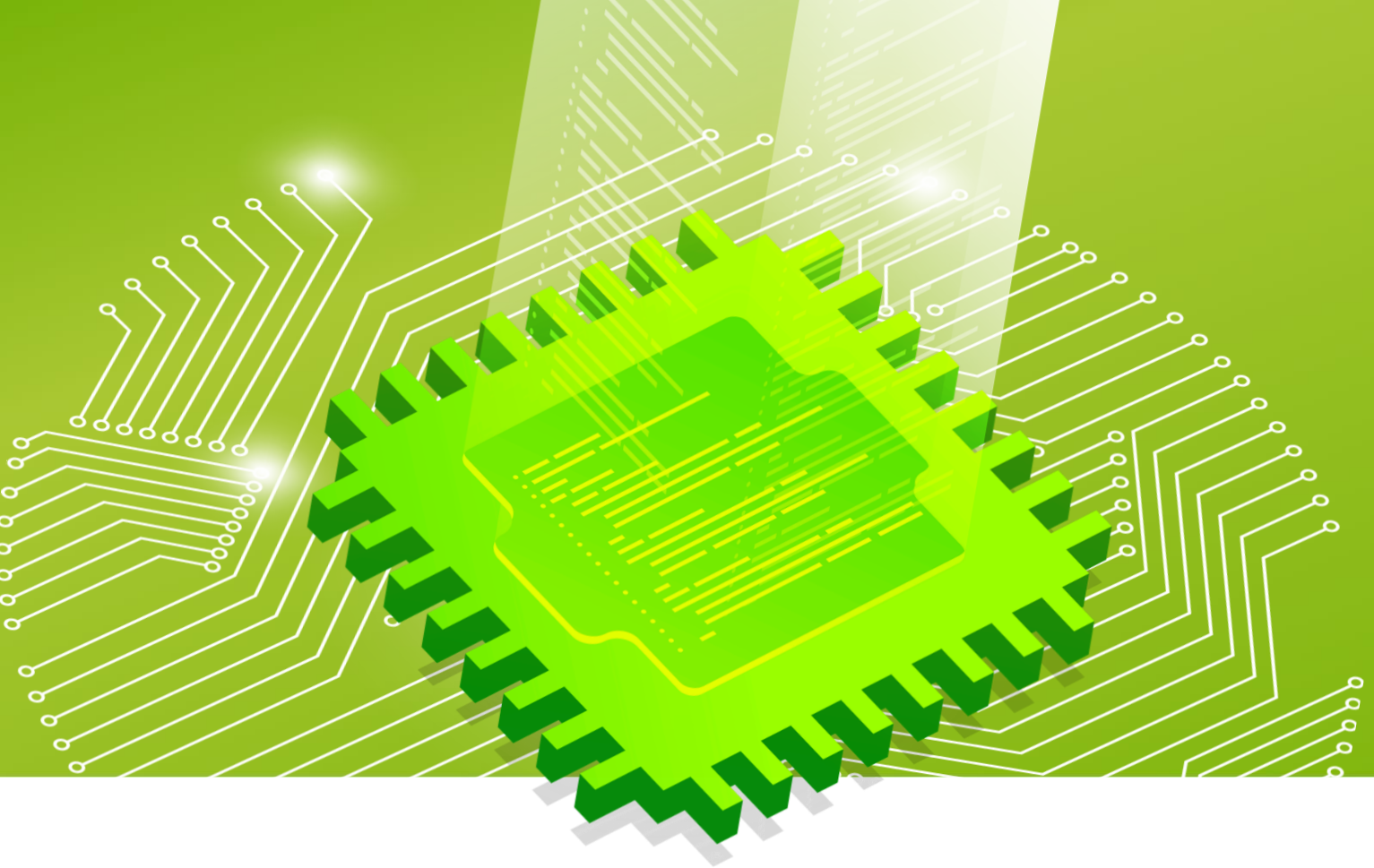


IDC ANALYZE THE FUTURE

数据分析新速度 加速数据科学转变成商业洞察

IDC 中国 2019-12

中国 成为全球人工智能市场规模增速 最大的单一国家

↑30%
加拿大

2%

↑33%
亚太其他国家

5%

↑26%
美国

5%

↑41%
其他

5%

↑41%
西欧

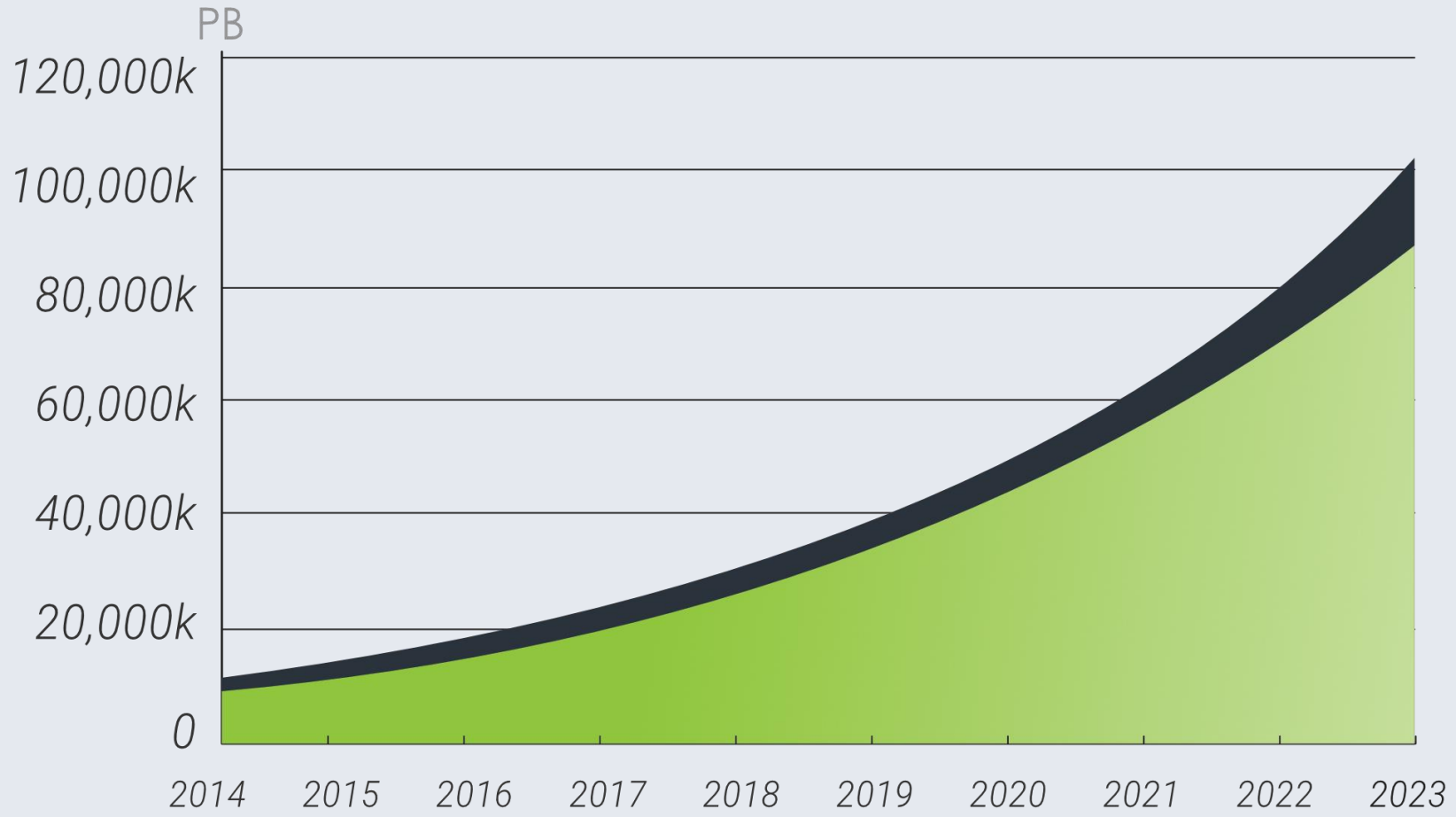
19%

12%

中国
↑64%

来源： IDC's Worldwide Semiannual Artificial Intelligence Tracker, 2019

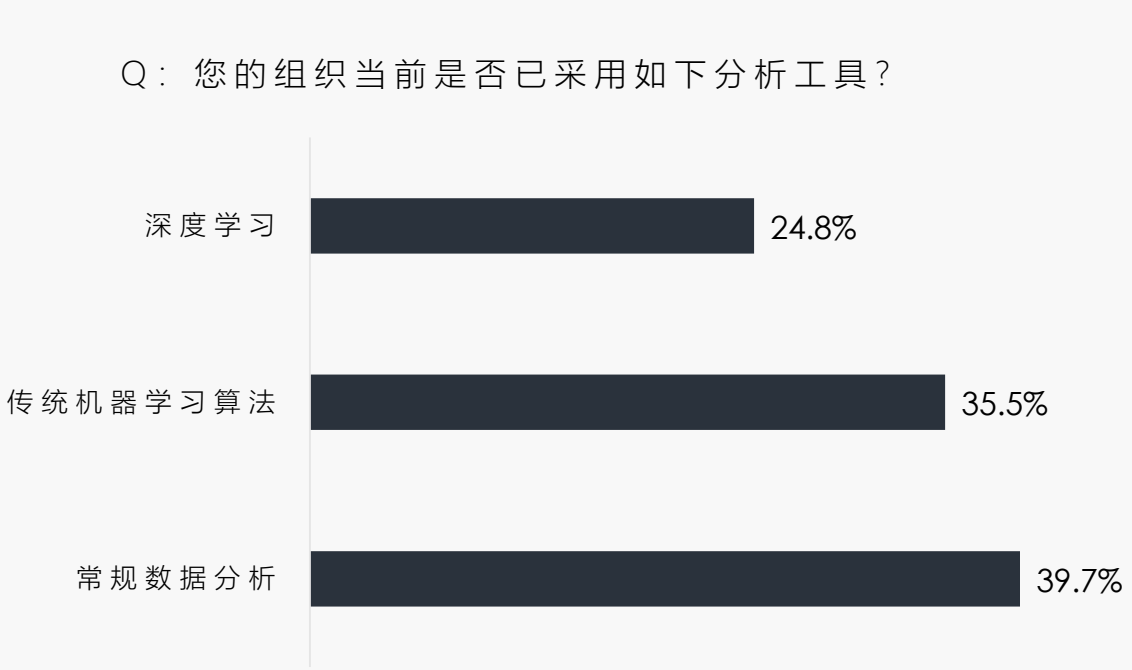
数据爆炸式增长 驱动数据科学新需求

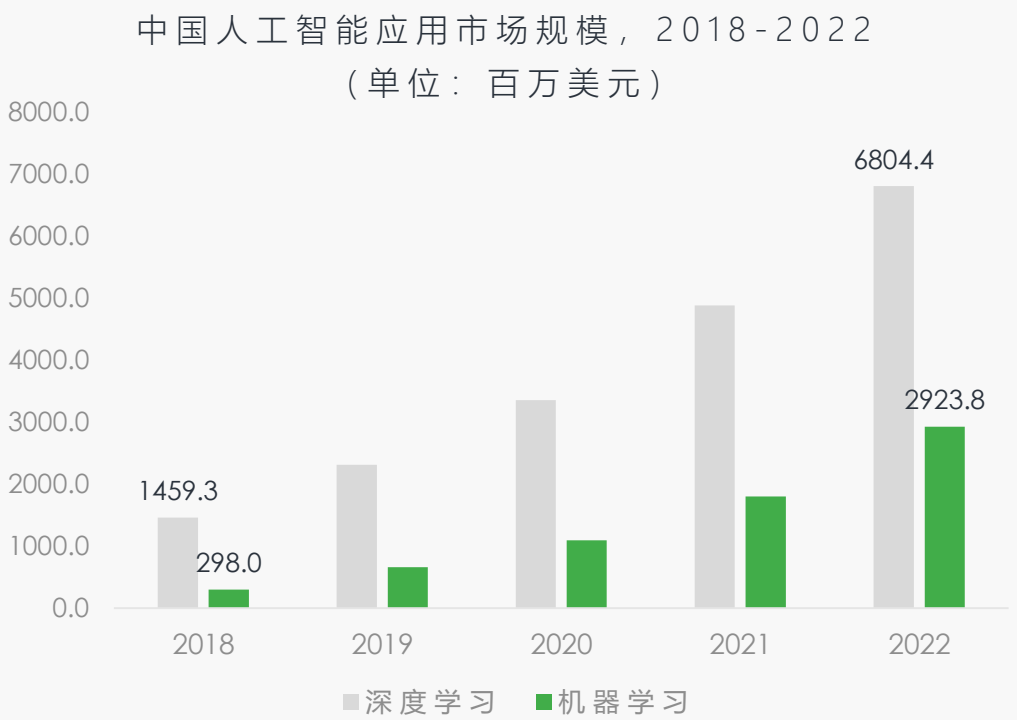


数据科学1.0-3.0的演进

	数据科学1.0	数据科学2.0	数据科学3.0
数据类型	历史交易数据 客户管理系统数据	互联网数据/APP数据/企业数据	物联网/车间设备数据 实时生产/交易数据 流数据/批数据
主流工具	Excel	商业智能 常规数据分析 NoSQL Hadoop	高级预测分析 机器学习/深度学习
分析行为	Excel进行分析	数据清洗/存储	采用快速/敏捷/灵活的数据科学平台建立模型

机器学习的采用率 迎来新浪潮



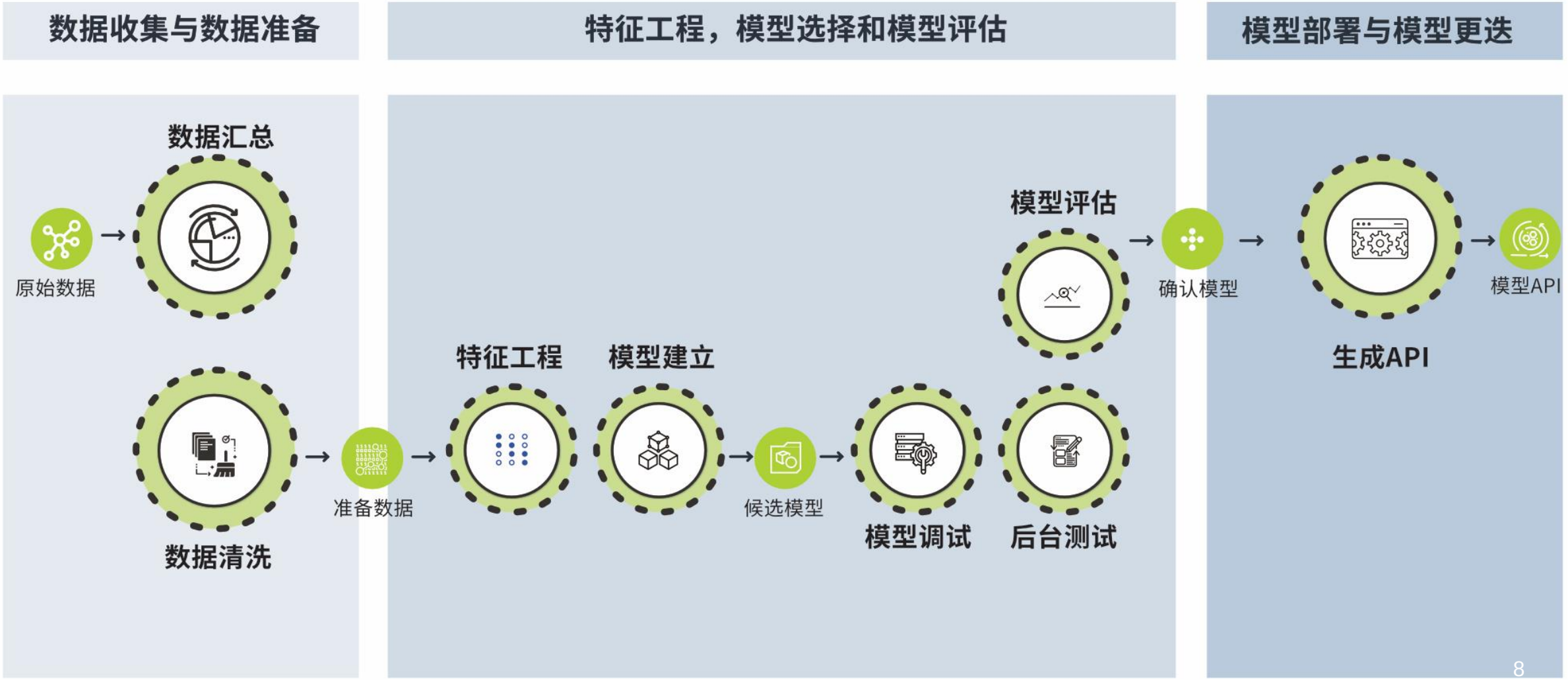


数据科学在 人工智能中的角色

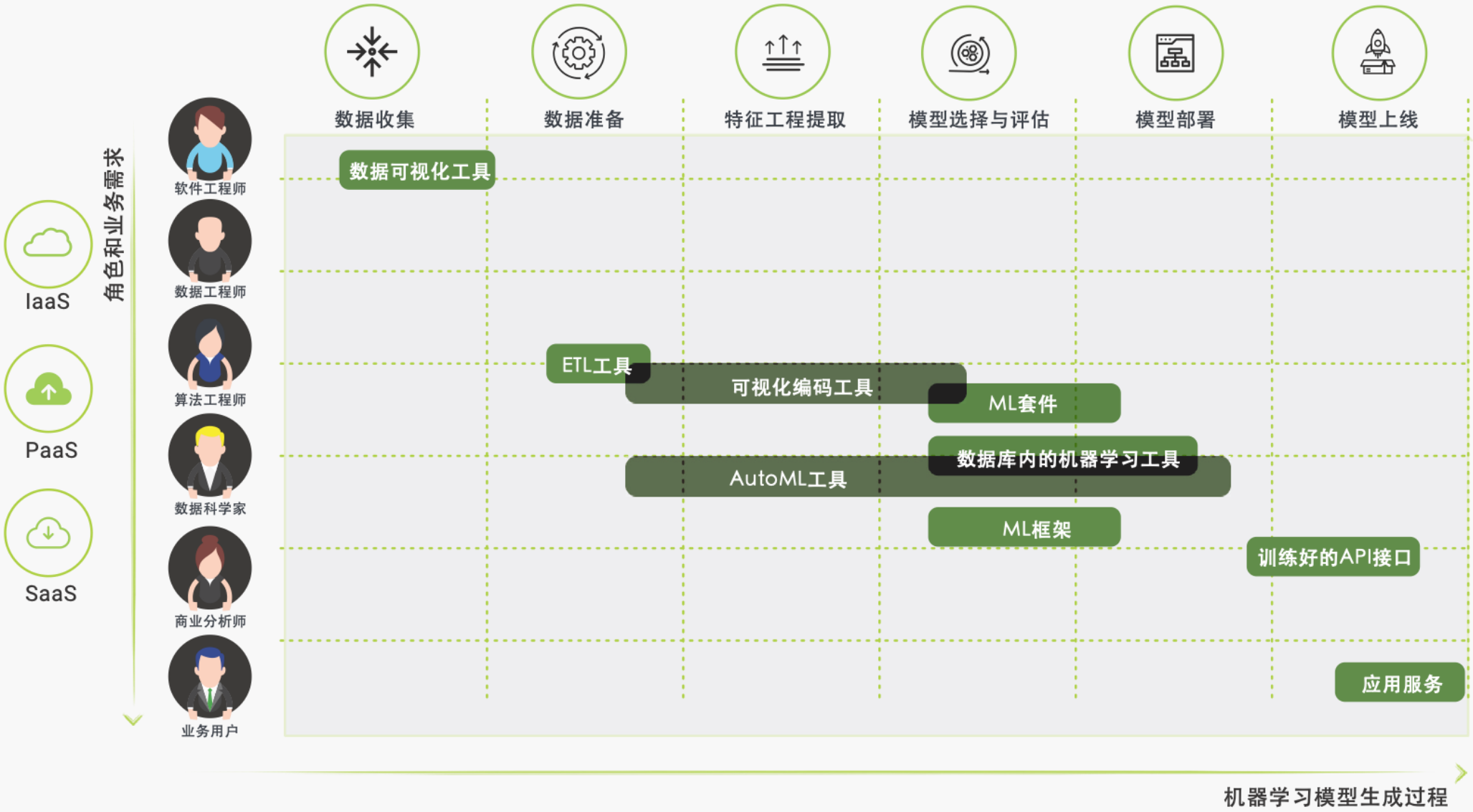
今天的数据科学与以往的商业智能 大不相同

结构化、半结构化为主	各种类型，多模态数据
企业内部系统	内部，外部，社交媒体，移动
分析师	数据科学家
分析趋势，形成报告，展示面板，KPI	发现，调查，可视化，评估，证实，预测，优化，自动化
利用相关性分析，事后发现	发现相关性，事前预测
报表及统计工具	R, Python, Matlab,各种开源工具
统计学，概率论，工程数学	计算机科学，数据科学，人工智能
数据，决策	业务价值
运营层面	战略层面
ETL : 抽取-转换-加载	ELT: 抽取-加载-转换

数据科学建模流程 数据准备—模型部署

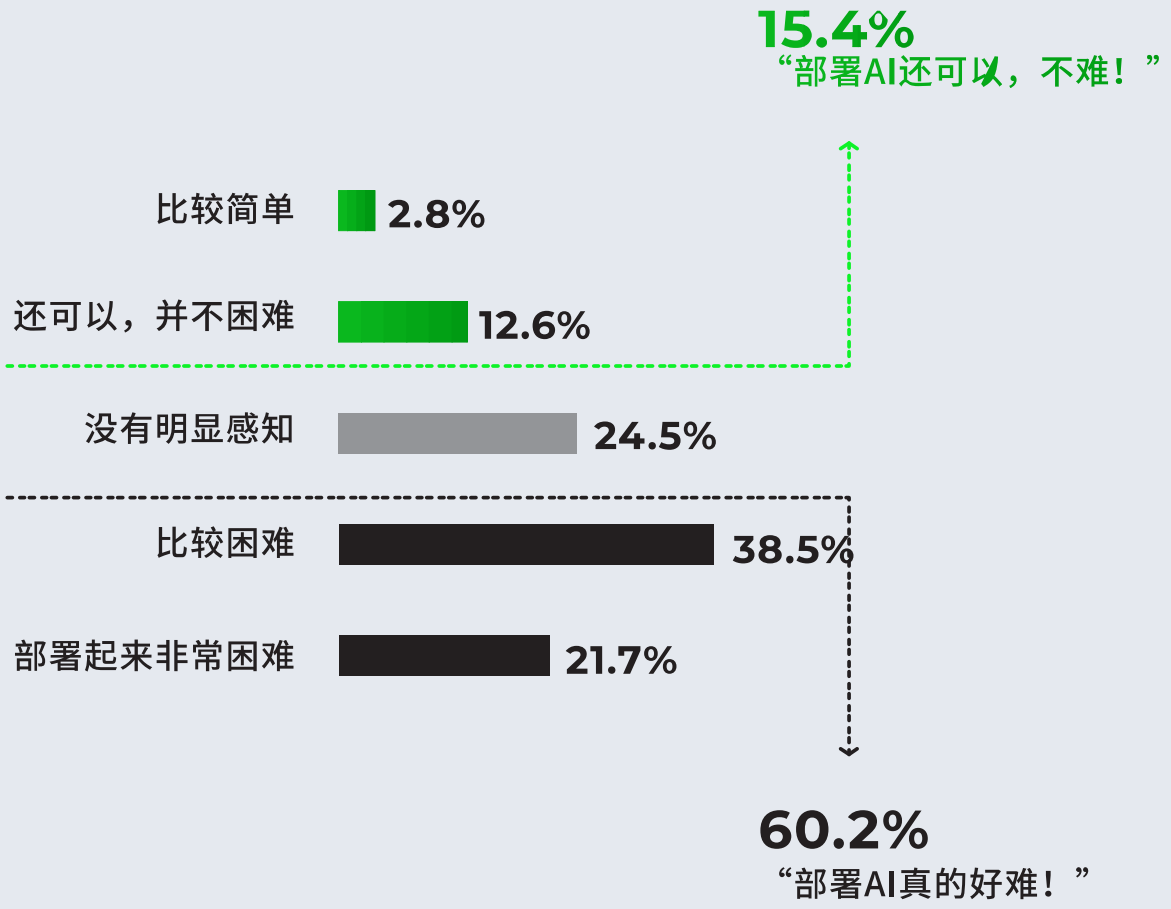


数据科学团队 需要从架构师到业务用户的协同

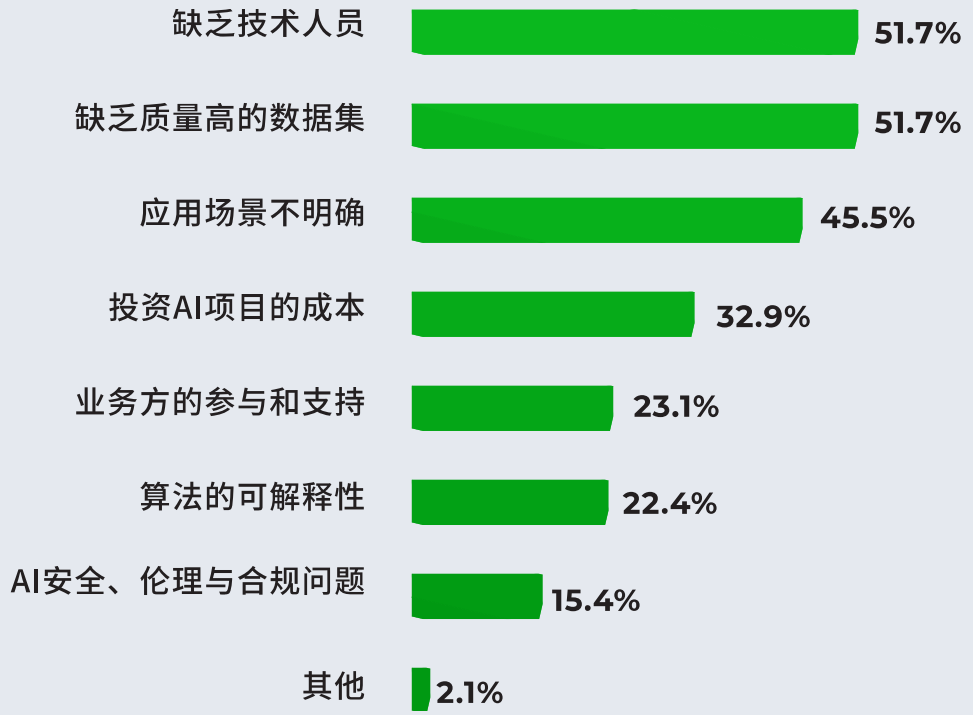


缺人才、缺数据是显著挑战 其次是定义应用场景

Q：您认为2019年部署AI系统的难易程度?

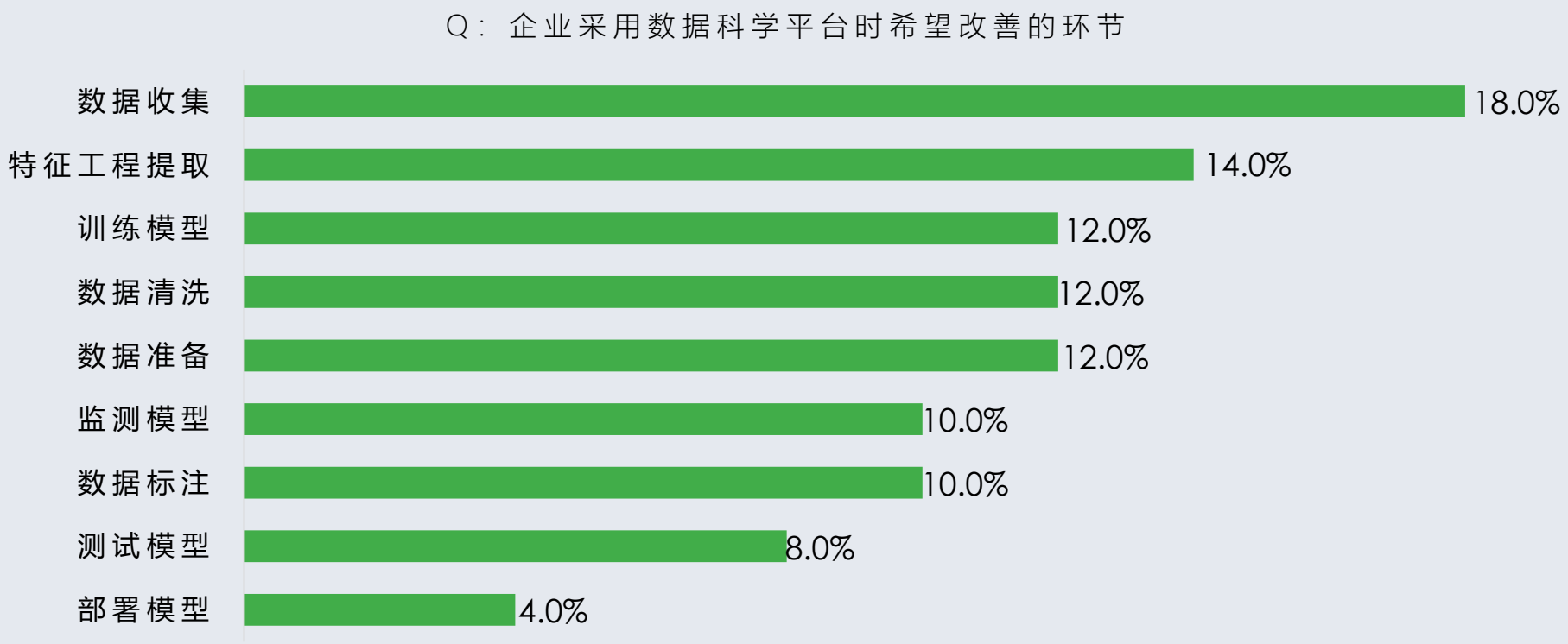


Q：您觉得采用人工智能的挑战有哪些?



Source：IDC中国，2019.11
N=143

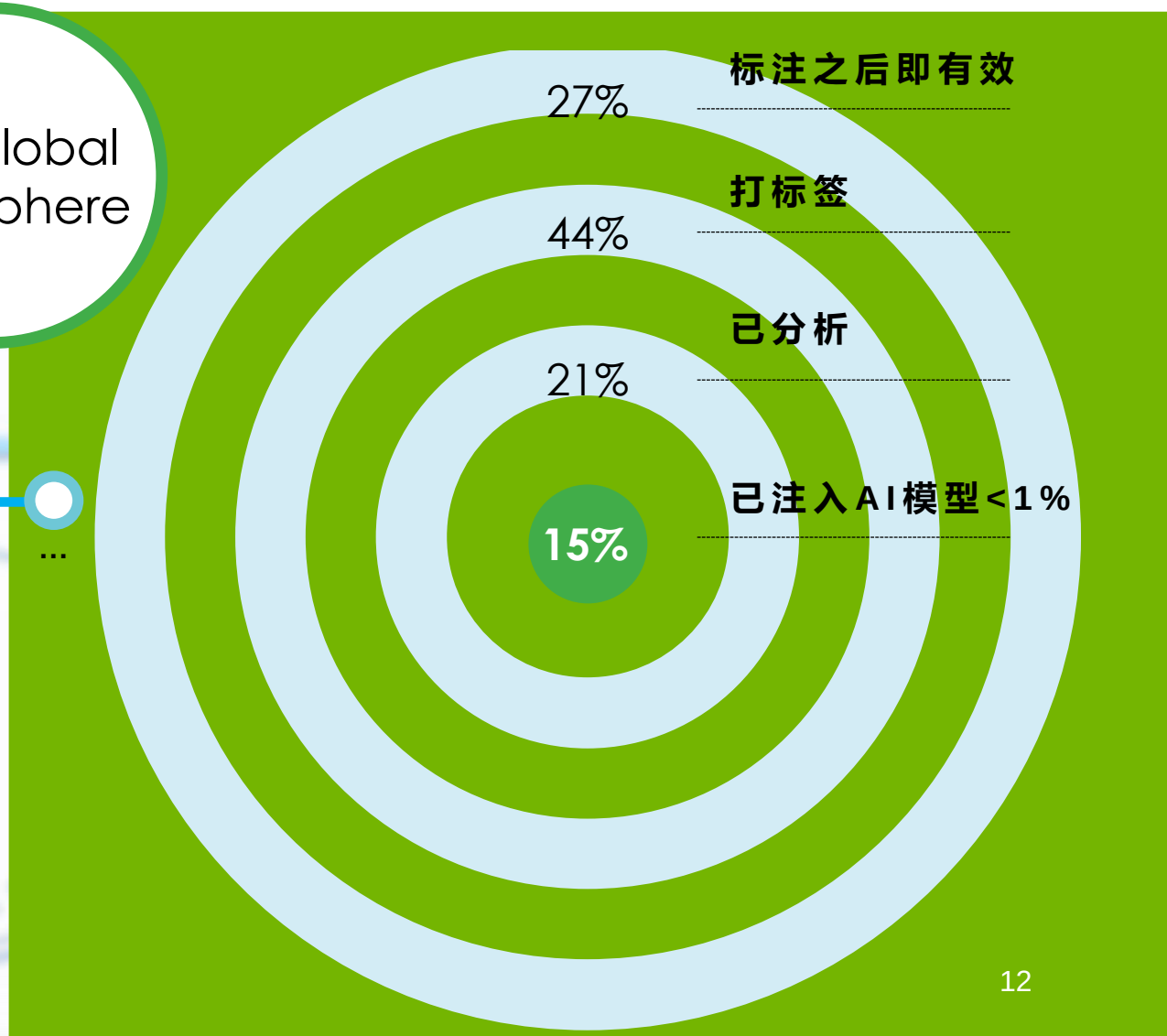
采用数据科学 最希望改善的环节



Source：IDC中国，2019.11
N=143

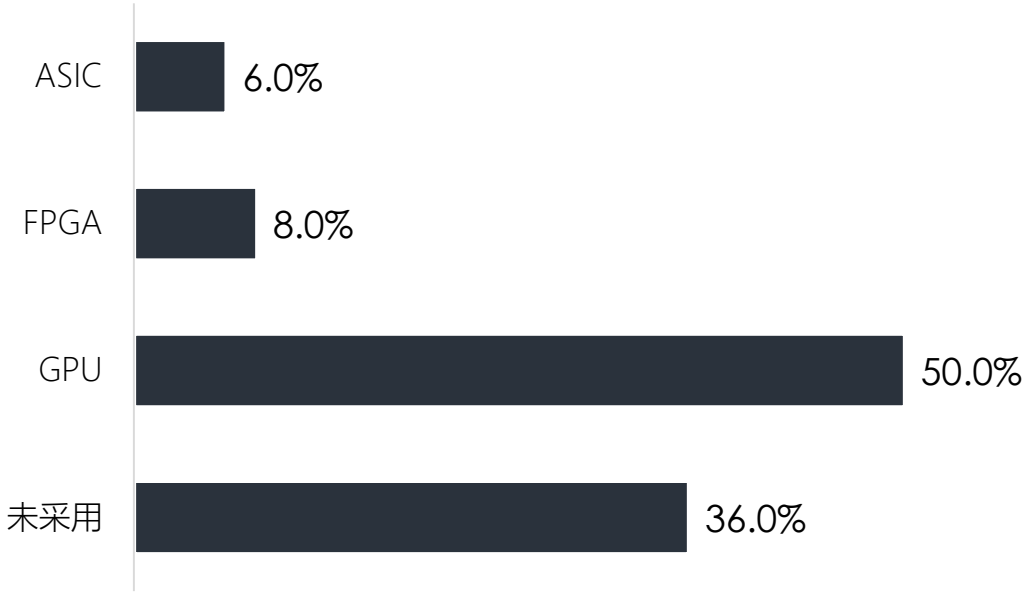
今天的新瓶颈在于严重缺乏可供模型训练的数据

IDC Global DataSphere

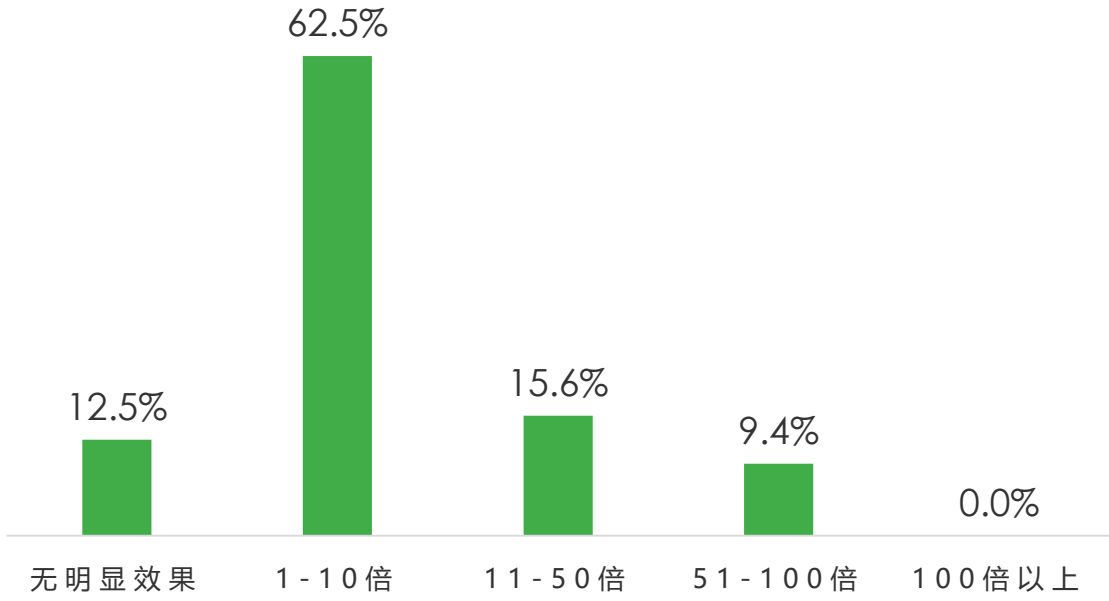


采用加速芯片 加速数据科学成果显著

Q：您的企业当前是否采用如下加速方案？

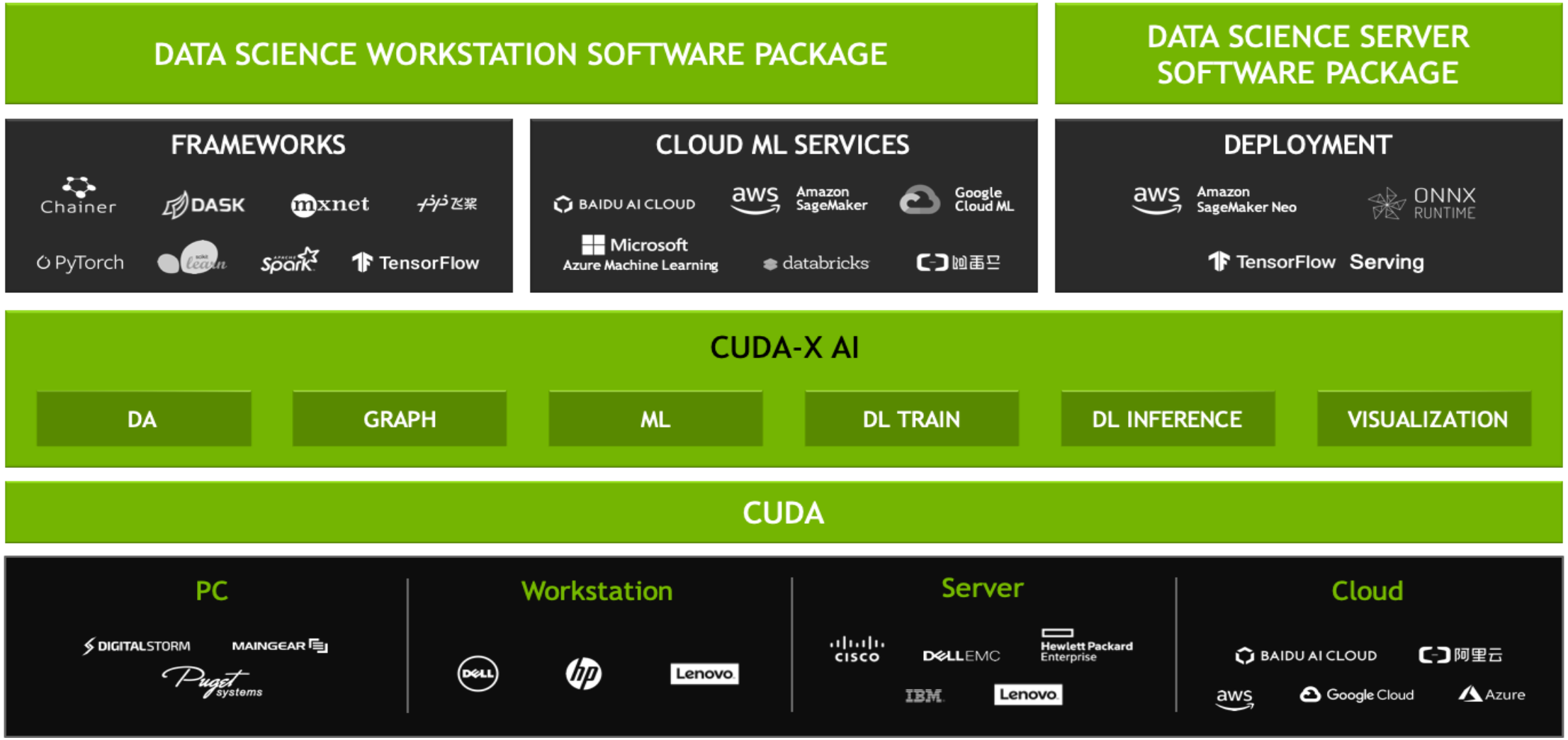


Q：您的企业采用加速方案后是否有加速效果？



NVIDIA RAPIDS

GPU加速数据科学 加速AI工作负载



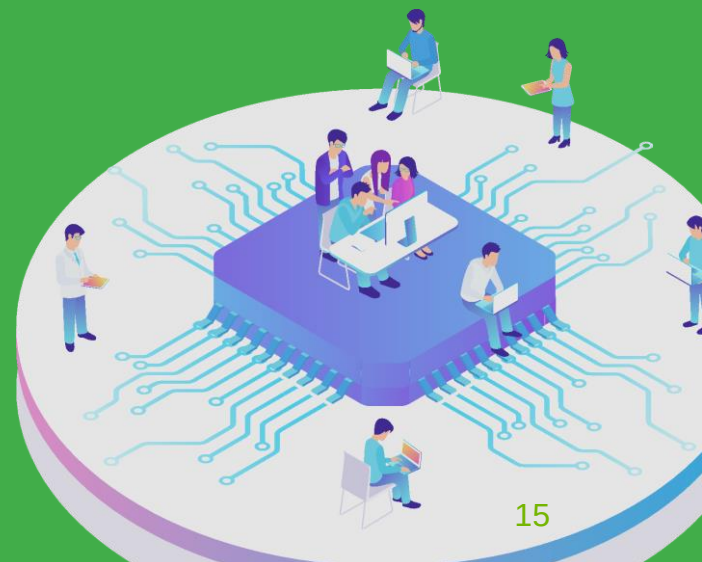
在 NVIDIA RAPIDS 加速数据科学的2个案例

宽邦科技

- 应用场景：在BigQuant平台中，宽邦科技采用随机森林、GBDT、XGBoost等算法对海量、实时数据进行可视化分析，为大型券商、基金公司、银行和10万C端用户每月提供数百万次机器学习模型训练。
- 应用效果：在NVIDIA Rapids平台上，GBDT模型在特征工程提取环节即可实现100倍速度提升，聚类算法可实现200倍速度提升，在提升速度的同时通过增加迭代次数，也提升了系统预测准确度。

平安科技

- 应用场景：银行行业做信用评分场景的时候，通常是使用XGBoost在Spark上进行运算，这一般需要几天的时间来完成一次端到段的模型迭代。
- 应用效果：引入Rapids在DGX-2上进行模型运算，实现了40倍的加速比提升（1 台DGX-2 Vs 5台Spark节点，750G内存，32核CPU）



未来展望 加速数据科学可扩展至更多应用场景

风控防欺诈
个人征信评估
智能投顾
产品推荐
用户画像
用户行为分析
客户关系管理
客户评级

优化基站和WLAN热点设计
网络流量和流向资源调整
客户关系管理
企业运行管理
垃圾短信分析

舆情管理
智能警务情报系统
公共安全响应及预警
精准扶贫

质量管理和推荐系统
维修及生产检测自动化
供应链管理自动化
AVG小车路线规划
生产线产品良率优化

辅助临床决策、用药
药品研究及发现
基因检测
健康管理与服务

景区客流管理

客流分析
店铺选址
会员管理
客户行为偏好分析

自适应学习
个性化辅导

精准广告投放
知识库
智能选题

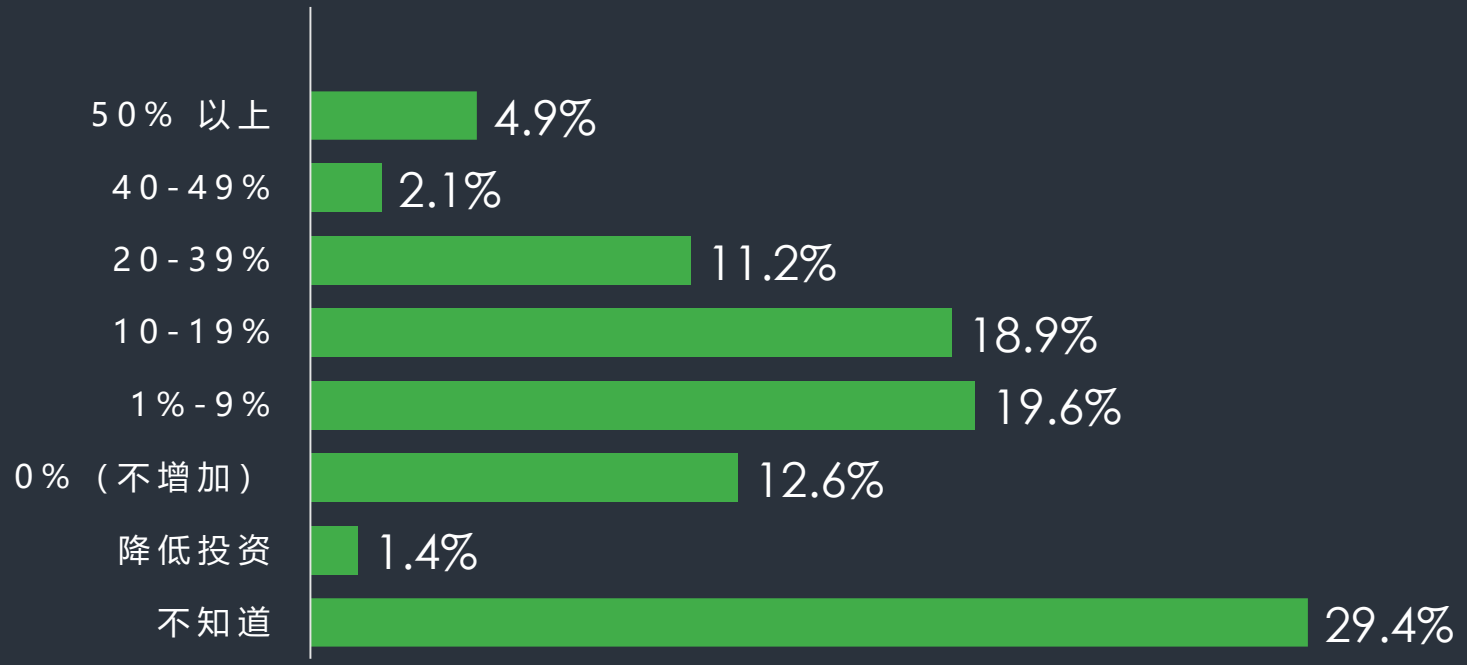
设备维护及生产检测自动化
能源勘探及开采

车流量分析
高速监控

园区选址与规划
人流量分析

回归到行业应用 逐步演进的应用智能化

Q：相比2019年，贵单位未来1-2年是否会增加AI投资？



2019需尽快优化的能力

数据智能标注 | 图算法 | 自动化机器学习 | 商业化机器学习 | 知识图谱 | 流数据批数据处理

中期值得关注的领域

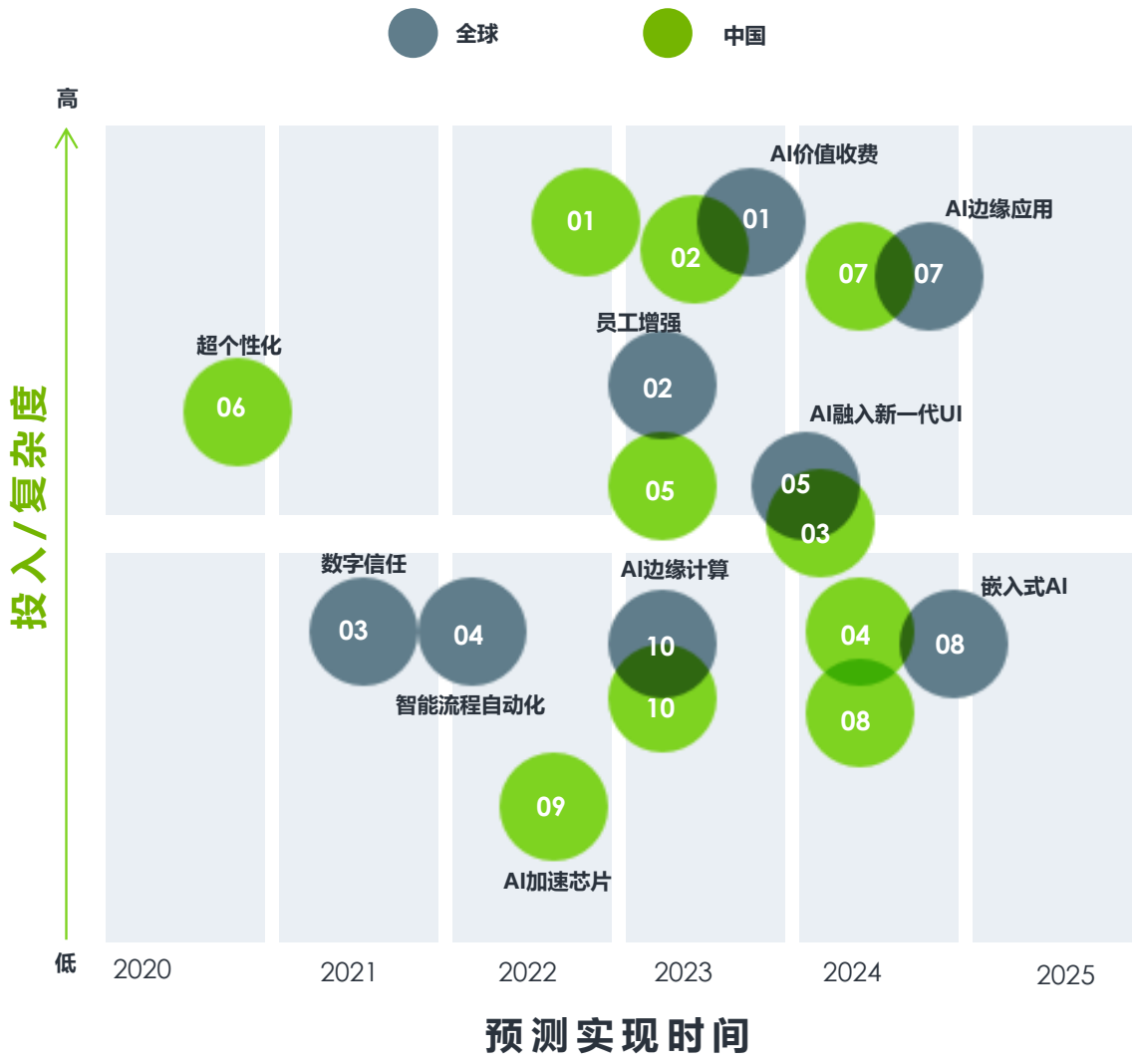
基因检测 | 量化投资 | 气象预测 | 交通大脑 | 石油勘探 | 图像渲染

长期应持续投资的领域

可解释的深度学习 | 生成对抗网络GAN | 自然语言理解 | 多模态交互分析 | Ai安全 | 引入社会因素的AI模型

IDC FutureScape

全球人工智能市场预测及中国启示



Source: IDC, 2019

01	到2023年，人工智能将成为各个业务环节不可或缺的一部分， 25%的人工智能解决方案支出将按价值收费，可推动大规模创新并实现巨大的商业价值。
02	到2023年，将有70%的企业投资于员工再培训和发展，包括第三方服务，以满足AI采用带来的新技能需求和工作方式转变。
03	到2024年，超过60%的中国1000强公司将通过正式项目监控企业内数字信任情况，因为数字信任已成为公司的重要资产。
04	到2024年，将有75%的企业将智能自动化嵌入技术和流程开发，并借助AI软件生成运营和体验洞察，引领创新。
05	到2023年，人工智能将因重新定义用户体验而成为新一代用户界面的关键技术，超过50%的用户触达将通过计算机视觉，语音，自然语言和AR / VR等技术实现增强。
06	到2021年，将有15%的客户体验应用将通过持续结合各种数据和创新的强化学习算法实现超个性化。
07	到2024年，将有50%的计算机视觉和语音识别模型将在边缘端（包括端点）运行，卷积神经和递归神经网络会成为深度学习的重要组成部分。
08	到2024年，至少90%的新企业级AI应用版本将融入嵌入式AI功能，但真正具有颠覆性、以AI为主导的应用将仅占总数的10%。
09	2023年，技术买方在专门用于加速AI训练和推理的半导体（包括GPU，FPGA，ASICs，ASSPs）上的支出将增至近90亿美元。
10	到2023年，将近20%用于处理AI工作负载的服务器将使用AI优化的处理器和辅助处理器，并部署在边缘侧。

非常感谢！