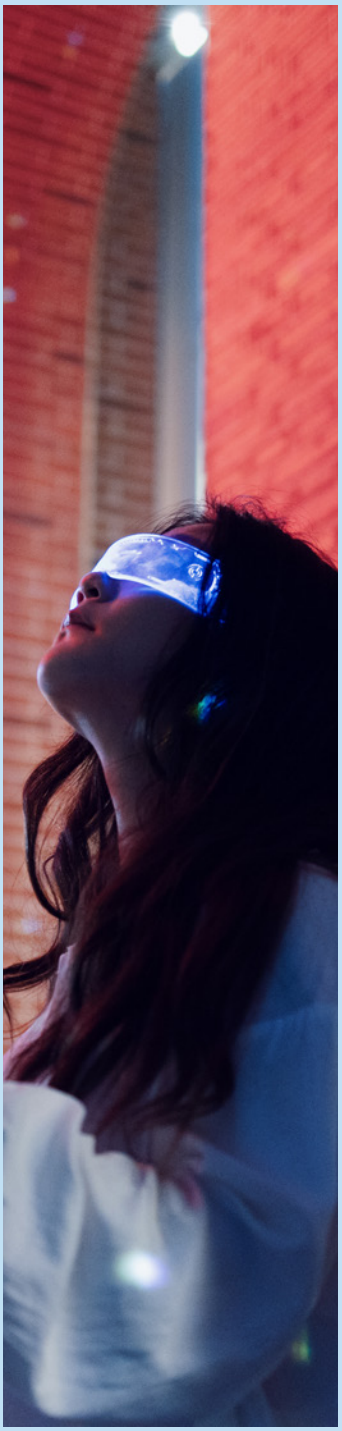
Infosys topaz

VIEW POINT



人工智能在中国：

首席执行官如何驾驭快速演化的生态

摘要

人工智能（AI）正在全球范围内改变各个行业，并且开启了企业释放潜力的新时代。从企业的角度来看，这场变革不仅影响了AI的提供方，也同样影响了使用方。

在全球范围内，不同的AI体系正处于不同的推广和部署周期阶段。深度学习、机器学习和计算机视觉技术已经在企业和公众中已经得到广泛应用，而生成式AI（GenAI）则正在全球的企业中迅速崛起。

对于在中国运营的企业、无论是全球还是本地企业，其商业领袖们必须认识到，中国活跃的AI技术生态、严格的监管环境、行业特定需求等因素对人工智能的叙事、路线图和战略方面产生一些独特的影响，他们因此必须对这些特点予以关注，方才能够在其业务和IT战略方面更好地适应中国的AI生态。

Infosys®
Navigate your next

中国AI生态的特性

我们在中国举办的每一次高层对话中，人工智能（AI）都是议程的首要议题，尤其是那些将其绩效与总部标准进行对齐的全球跨国公司、以及希望以AI驱动生产力的中国企业。对这些企业而言，中国的AI生态独具特色，其特征是：监管环境快速进化、供应商和应用网络日益壮大。这一独特的本地生态系统给他们的AI解决方案同时带来了复杂性和和独特的机遇。

另一个特征在于AI应用的目标市场的不同。例如，在美国等国家，大多数AI应用面向公众，而中国的AI应用则更多地面向细分市场、面向特定的行业应用场景。这就使得这些相关的企业有机会在中国延展他们细分的解决方案，例如，利用AI进行市场需求的预测、并提高其生产协同的自动化程度。

此外，人工智能生态系统的二元性正在出现，这是由于中美主导了世界人工智能的发展、同时参与者所处的不同的法律和经济环境却存在差异所决定的。随着生成式AI用例在各行各业的不断发展，这种分化变得越来越明显。虽然这种二元性增加了在中国运营的外国公司的复杂性，但许多公司已经采取了合适的方法、适应了数字世界的现实、甚至享受了二元性带来的益处。

综上所述，随着中美主导人工智能发展，在中国运营的企业——包括跨国公司和中国民营企业——正在开启一个崭新和独特的机遇之门。

挑战与未来

我们首先要了解有哪些挑战。考虑在中国国内部署AI应用的企业主要面临三大挑战：

1.了解相对复杂的、快速跟进的法律体系

中国早在2017年便开始建立AI治理和监管框架。发布于2017年的《新一代人工智能发展规划》明确了到2030年成为全球AI领导者的目标。此后，政府发布了一系列相关的通知、指南和标准。最近的一次是2024年7月由工业和信息化部发布的《国家人工智能产业综合标准化体系建设指南》。到目前为止，其中引起最多关注的法规是2023年8月发布的《生成式人工智能服务管理暂行办法》。

此外，涵盖网络安全、数据隐私、数据传输等方面的法规陆续出现，其中包括《网络安全法》（CSL）、《数据安全法》（DSL）、《个人信息保护法》（PIPL）以及《促进和规范数据跨境流动规定（CBDT）》。

对于 AI 提供商、包括大型语言模型（LLM）的开发人员，最需要关注AI模型的注册和合规性测试方面的强制要求。合规的 LLM 名单由中央网络安全和信息化委员会办公室（CAC）发布，可在 CAC 网站上查阅。

这些法规给人工智能的提供商带来了一些挑战：

- **挑战之一**，挑战之一，合规和监管导致解决方案的难度：企业必须进行额外关注以确保遵守新法规。严格的数据安全、处理和处理要求解决方案必须满足严格的数据安全标准，特别是对于那些依赖多样化的、广泛的数据集的企业。例如：一家企业在位于新加坡的总部部署了一套基于OpenAI的智能客户情报分析和智能响应系统，这家企业在中国开展业务时，就必须考虑在中国境内重新部署经过国家权威部门认证的新模型，以避免敏感数据的出境、以及使用合规的AI模型的问题。

- **挑战之二**，选择正确的技术路径：这些要求可能会限制技术选择的范围，要求企业在合规的框架以内进行技术创新的竞争。例如，某家跨国医药公司在国外成功开发了一套先进的肿瘤诊断AI系统，该系统基于第三方的最新的机器学习算法，能够对患者的医疗影像进行高精度的分析，可以识别早期、尤其是细小而难以辨识的癌细胞。当需要在国内应用时，由于模型是第三方所有，该公司无法直接驱动其认证过程，这样就迫使这家公司基于国内合规的模型重新进行开发，这表明，该公司在构建技术治理框架时，没有充分考虑合规性的问题。

尽管如此，对于使用 AI 服务的企业而言，了解这些法规有助于明确每个项目的合规性需求。我们强调：经过大型模型标准合规性评估授权和模型注册是安全合规性的重要保障。企业部署的 AI如果是向公众提供服务，其监管会更加严格，但是对于企业内部使用的限制相对会小一些，这就造成了当前在企业内部 AI 部署增长得更快的事实，其用途包括优化工作流程、提高生产力和重塑内部企业职能等。这是业界的一个显著的特点，同样应当予以关注。

2. AI服务供应商及其LLM的可用性问题

一方面，由于政府设置了严格的网安措施，一些AI服务无法从国内访问，加之 OpenAI等公司对中国等国家的用户也采取了限制，中国用户无法访问许多知名的全球性AI 服务，例如，GPT-4、LLaMA（Facebook）、Gemini（Google）和 Claude 等平台对中国国内的用户是无法使用的，其中也包括了知名的AI源码社区 Hugging Face。

另一方面，中国国内也发展出了足够的模型可供广泛使用，形成了一个与国外平行的生态系统，类似于中美之间更广泛的互联网经济中的双栈结构。众多的通用模型，例如阿里巴巴的通义千问（110B 参数，10M 上下文长度）、智谱的 ChatGLM（130B 参数，128K 上下文长度）、零一万物的Yi-Large、商汤科技的 SenseNova 等都得到了广泛的应用。机构们会根据全球常见的评估模型（例如 MMMU 基准）不断进行评估，我们看到这些模型正在被快速改进。

此外，如前所述，中国在垂直、行业特定模型方面具有明显优势。百度的羚羊（制造业大模型）、商汤的Sensenova（医疗和法律等）和百度的零一（生命科学）等模型现已越来越广泛第用于商业用途。

羚羊是为制造业量身定制的，支持工业文档(如报告和PLC代码)，以及价值链活动(如维护，故障诊断，材料检查和统计分析)。同样，来自京东的九书针对在线零售，提供个性化推荐、产品销售预测、日志和点击行为分析、内容创建、通过图像匹配进行 SKU 分析等功能。这些垂直模型正在获得广泛的采用和使用。

对于人工智能的实施，这些在本地数据上训练的本地化模型通常是更好的选择。它们提供相关的本地环境、更具成本效益，并通过轻量Transformer、检索增强生成（RAG）等实施方法进行部署，能够提供更集中和满意的产出。

另外，像始智社区这样的线上社区正在成为Hugging Face的本土替代品，主要应用于技术研究与应用开发。但在这一领域仍需要更多的进步。

例如，我们曾经为一个大型的电商公司设计了一个跨国的解决方案，在一定程度上解决了合规性的挑战、同时也充分利用了国内的技术生态。这家公司需要采集用户数据优化其用户

体验、供应链效率、通过客户的沟通记录监控服务及产品质量、以及发现产品创新和改进的需求等等。由于其业务、产品、供应链横跨了不同的国家，包括中国在内，在部署AI模型时就遇到了上述多个问题。我们通过分部式部署、分别采取了四项主要的措施（参见图1），既满足了合规性要求、又充分利用了国内外不同AI模型的优势。

3. 硬件和GPU的可用性问题

另一项AI部署中讨论最多的限制之一是高性能AI专用GPU的禁运，同时，高端芯片技术的限制也对自行研制用于先进GPU的能力施加了进一步的限制。像英伟达的A100和H100这样的GPU的缺乏在一定程度上限制了大型语言模型的训练能力。

这些限制迫使企业进行创新和不断适应。在硬件方面，中国制造商正在大力投资研发，新的GPU不断出现，如Ascend 9x0（华为）、Hugon GPU（Higon）和YunFei（芯动科技）。英伟达还开发了中国专用的 GPU，以缓解禁运方面的影响。与此同时，业界也正在探索架构的演化，例如横向扩展较低端的 GPU，以满足 LLM 训练需求。

此外，较小规模的业务集成模型的开发、以及能够减少算力需求的低秩适应（LoRA）等微调技术得到越来越多的应用。这些垂直化模型迎合了特定的垂直领域和使用案例，减少了动辄数千亿个参数的训练开销。得益于广泛的工业数字化、物联网和 5G 部署，企业通过利用内部的企业特有的上下文私有数据可以有效地训练这些模型，而无需高端 GPU。

另一个重要的考虑因素是中国国内模型的成本优势。就训练结果而言，高端GPU的不可用性对人工智能推断的影响有限。而就成本而言，基于中国的模型的人工智能实施的持续推理成本只是全球流行模型的5- 15%！

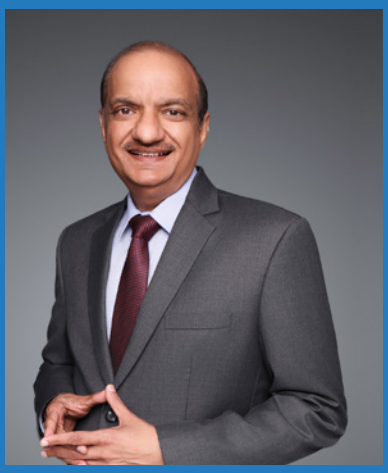
六个要点：对企业的战略建议

以下是对中国企业寻求主流人工智能技术时的一些建议：

1. **寻找机会利用AI生态的差异性进行套利：**在中国的跨国公司可以通过在全球数据中心和跨境数据传输的参数化框架内部署，在合规的前提下，获取两种AI生态系统的优势组合：例如，同时利用中国的垂直模型和国外的通用模型（在合规的前提下）。
2. **采用与生态系统无关的架构：**对于AI实践者和企业消费者来说，架构必须与基础的通用模型无关，允许“即插即用”的方式。这样一来，可以使相对稳定的业务架构能够适应生态的差异性和快速的发展。
3. **投资于人才和变革管理：**为了在竞争环境中成功地实现AI转型，为了适应上述的挑战，持续投资于人工智能的人才至关重要。人才培养的重点方向应该是在整个企业中嵌入人工智能的经验和能力。
4. **为亚太地区和全球南方国家的部署做好准备：**鉴于中国超级云厂商在全球欠发达国家的投资，评估和输出中国AI生态系统是一个有吸引力的话题。
5. **数据准备：**数据的合规性是保证解决方案合规性的前提，因此企业数据的准备工作是AI应用的关键，数据准备不足、存在合规性风险往往导致AI工业化进程的延迟。
6. **负责任的人工智能：**无论法律法规如何发展，无论中美生态的差异如何，负责任的人工智能是一重要的商业责任，是合规的基础。它将道德融入运营的核心，注重维护声誉和提高透明度。在人工智能生态系统中进行适当的检查和保障、维护信任和责任原则至关重要，通过设计实现负责任和安全的指导原则有助于减轻企业的运营风险。

结语

在计算机视觉和深度学习方面，中国企业具有独特的优势，众多企业具有开发和运用相关模型的丰富经验、取得了令人瞩目的成就。通过利用数字化、物联网等技术体系、通过开发和管理数据核心资产，企业可以专注于AI在特定行业的实施，专注于提高生产力并推动创新。特别是跨国公司，在利用生态系统和开发部署负责任的人工智能来可以形成潜在的生成式人工智能方面的优势。挑战是真实的、但找到最先进的解决方案也是可以触及的，关键是要从不同的信息和观点中辨别出有价值的见解，采取果断行动、并开启旅程。



Rajnish Sharma 沙睿杰

Infosys 中国区总裁

Rajnish Sharma沙睿杰是Infosys 全球副总裁，中国区总裁，同时也是Infosys 中国董事会成员，现常驻上海。

沙睿杰负责Infosys 在华整体业务运营和市场营销，领导全球和本地客户服务交付，提升客户满意度。凭借云计算、人工智能和自动化主导的数字服务，他领导的团队致力于为客户制定和实施数字化战略，帮助客户完成数字化转型之旅。此外，他还负责Infosys 大中华区团队的人才管理和能力建设。他重视团队合作，是工作场所多样性和包容性的坚定倡导者。

在Infosys 服务的23年中，沙睿杰曾率领团队为多个行业的全球大型客户提供服务，业务遍及美国、欧洲和亚太地区。沙睿杰拥有印度国家理工学院（National Institute of Technology, Kurukshetra, India）计算机工程学士学位。

Infosys Topaz是一套使用生成式人工智能技术、以人工智能为中心的服务、解决方案和平台。它有助于扩大人类、企业和社区的潜力，创造价值。Infosys Topaz拥有12,000多项人工智能资产，150多个预训练的人工智能模型，以及由人工智能专家和数据策略师指导的10多个人工智能平台；通过“设计负责”的方法，帮助企业加速增长、大规模释放效率并构建互联生态系统。

For more information, contact askus@infosys.com

Infosys®
Navigate your next