

# 中国工业大模型行业发展研究报告

靡不有初，鲜克有终

## ABSTRACT

## 摘要



**工业大模型伴随着大模型技术的发展，逐渐渗透至工业，处于萌芽阶段。就大模型的本质而言，是由一系列参数化的数学函数组成的计算系统，且是一个概率模型，其工作机制是基于概率和统计推动进行的，而非真正的理解和逻辑推理，因此，当前大模型具有不可解释性和幻觉不可消除等主要特征。就大模型落地工业的情况而言，工业互联网、等工作已经让部分工业企业遍历了数据采集-数据存储-数据处理-数据分析-数据资产沉淀-数据应用的过程，部分场景已经准备好了向基础大模型投喂的“数据原料”，当经过简单数据处理、微调、适配后，可以解决部分垂直细分场景问题，具有落地可行性。就工业大模型的发展进度而言，工业大模型与工业互联网一样，都是要挖掘数据资产的价值，而数据准备的阶段性工作在工业互联网时期大部分已经准备好，故我们预计工业大模型的进程在技术不受限的前提下，可能会快于工业互联网。当然，工业大模型是以大模型技术为驱动，其进程快慢很大程度受限于大模型本身能力的进化。**



**工业大模型玩家与工业互联网平台玩家重合度高，其成长路径目前也表现出高度相似的特征，但目前市场产品、服务、落地场景都处于探索阶段，大家都在同一起跑线。就参与玩家而言，大模型技术底蕴、行业know how、运维资源等方面是各类玩家主要锚定的优势，且都是基于自身优势点，围绕具体应用场景摸索大模型在工业的落地性进行市场切入的。就具体产品形态而言，整体还较为稚嫩，目前大模型的能力更多还是依附于已有产品体系，鲜少有独立的产品出现，未来随着大模型流量入口特性明朗，有望独立成产品。就应用场景而言，当前大模型的不可解释性和幻觉等特性，与工业“0容错”的特性相悖，因此当前大模型落地工业的探索更多聚焦于偏运营的、具有一定容错能力的场景（如知识问答、辅助设计/代码生成等），而生产制造等核心场景的探索需要静待模型进化以及CV大模型、多模态大模型的发展。就发展挑战而言，模型、数据、应用、商业变现是无法避开的话题，且各方相互影响，互利共赢。**



**大模型落地工业的探索中，还处于非常早期的阶段，供需双方都在尝试，当然，也有很多问题值得探讨与思考。1) 大模型落地工业的竞争要素：基础能力、模型能力、模型应用是主要竞争点，且在不同行业发展阶段，其相对竞争优势有所不同，具体而言：短期主要看大模型技术，长期则主要看模型应用深度。2) 大小模型间的关系：大小模型间不存在替代关系，是并存且是协同融合赋能的关系。3) 工业大模型服务走向平台化：大模型落地工业的服务平台化特征以开始显现，且逐步形成垂直行业大模型+智能体+小模型+机理模型为主的平台化调用方案。4) 产业数据拉通助力工业大模型能力进化的同时，对大模型落地工业的广度、深度都大有裨益。**

## CONTENTS

# 目录

---

### 01 背景篇

大模型落地工业具有可行性

---

### 02 市场篇

玩家纷纷涌现，但产品和服务形式尚在摸索中

---

### 03 思考篇

模型能力进化带来不确定，合作利大于弊

---

### 04 专家篇

一线企业专家的项目实践经验分享及前瞻的思考

# 01 / 背景篇-基本概念厘清

大模型落地工业具有可行性

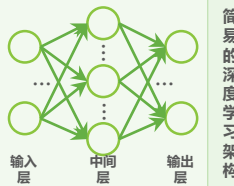
# 当我们在说AI时，在说什么？

AI本质就是一系列参数化的函数所组成的计算系统，并在一定的框架环境下，进行数据计算和参数调优的过程

整体来看，AI通用技术框架主要分为基础层、技术层、应用层三层，不同公司，在整体框架上都保持一致，只不过针对基础层和技术层的部分内容划分，略有区别，但主旨就是为各类函数组成的计算系统提供环境、数据、算力等支撑，最终达到模型运行并输出结果的目的。**具体到不同层级的主要玩家而言：**1) 基础层和技术层的AI框架和算法/架构上，其玩家比较类似，主要是国内外领先的科技巨头、云厂等参与，如谷歌、Meta、微软、亚马逊、阿里、百度、华为等；2) 应用层和技术层的工具组件、通用技术，其玩家更加百花齐放，针对行业、场景、功能等各角度发力，推动AI工具和应用的落地。当前，大部分大模型玩家的侧重点更多聚焦于通用技术的提升、应用的领域的拓展。

## AI的通用技术框架情况

应用层	C端产品及服务	<ul style="list-style-type: none"> <li>硬件类：智能汽车、机器人、XR等智能硬件产品</li> <li>软件类：ChatGPT、智能客服等软件应用</li> </ul>				
	B/G端解决方案	<ul style="list-style-type: none"> <li>金融、工业、教育等行业解决方案</li> <li>智慧城市、自动驾驶等领域或者场景解决方案</li> </ul>				
技术层	通用技术	通过机器学习、深度学习、NLP、计算机视觉、语音识别等算法的组合应用与优化，可以衍生一些相对通用的技术，如生成式AI、NLP大模型、CV大模型、多模态大模型等等				
	工具组件	全流程开发工具、可视化分析工具、安全可信组件、预训练模型应用工具、模型评估工具、云上任务提交工具、并行及优化组件等，目的是便捷AI能力开发				
	算法/架构	<ul style="list-style-type: none"> <li>机器学习，是核心，包含有监督、无监督、半监督学习等类别</li> <li>深度学习，包含CNN、RNN、GNN、Transformer等类别</li> <li>其他，如增强学习、强化学习、回归、分类等类别</li> </ul>				
	AI框架	<ul style="list-style-type: none"> <li>国外：TensorFlow (Google)、PyTorch (Meta) 等</li> <li>国内：MindSpore (华为)、PaddlePaddle (百度) 等</li> </ul>				
基础层	数据集	通用数据	行业数据，包含内部数据整合、外部合规数据等	...		
	软件设施	操作系统	数据库	云计算平台	大数据平台	...
	硬件设施	计算资源，如CPU、GPU、NPU等	存储资源	安全资源	传感器	网络资源



简易的深度学习架构

### 几点说明

- AI框架：**是AI领域的“操作系统”，AI算法的训练及应用等都是通过AI框架进行，主要提供模型训练和模型推理支撑能力。
- 算法/架构：**具体的运算函数，一般由大型公司、学术机构和研究所、开源社区或平台等提供。
- AI框架+算法：**在AI框架的基础上，不同特性算法的组合和变体，在某些能力上相对更加强大，可称为不同的模型类别。如果参数足够多，其推理、泛化等能力更加强，则可以称为大模型。
- 大模型厂商主要服务内容：**通用技术及相关工具组件。

### 本质

- 模型**的本质是由一系列参数化的数学函数组成的计算系统。
- 模型训练**是不断调整参数及其权重的过程，部分可能需调整算法架构。

来源：《深度学习入门》by 斋藤康毅，艾瑞咨询研究院自主研究及绘制。



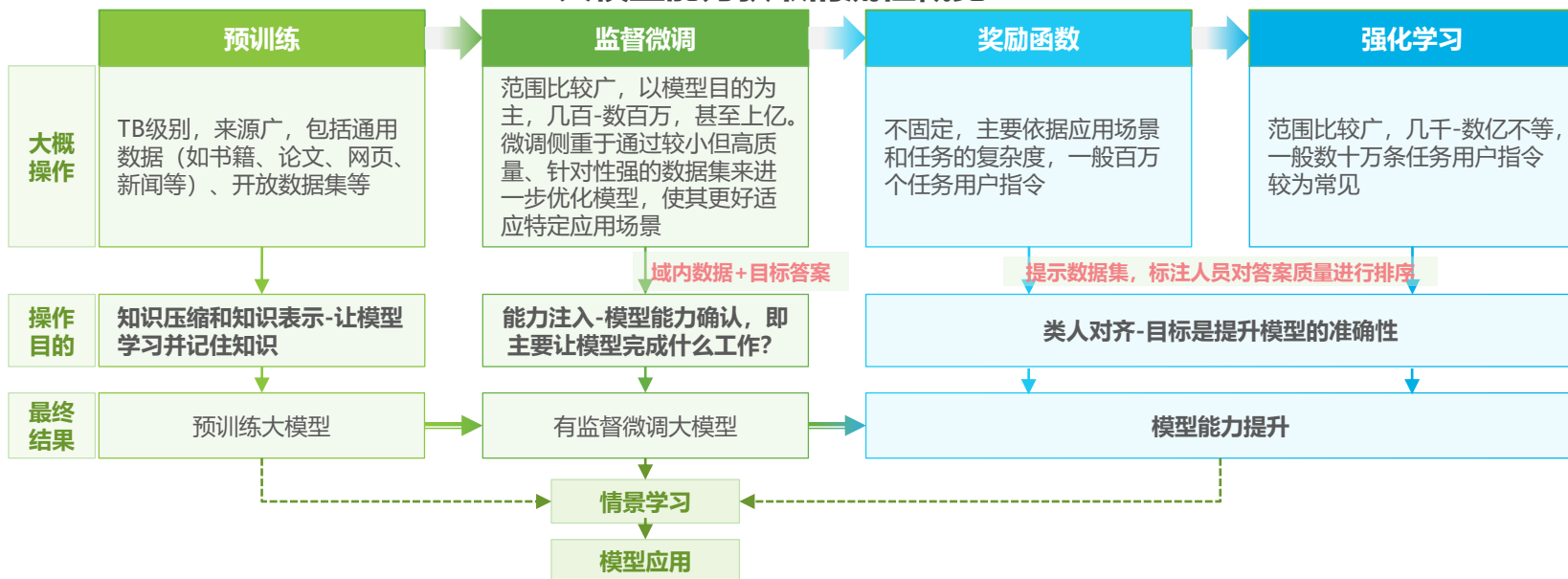
# 大模型是什么？大模型的能力如何获取？

大模型是从海量数据中学习并记住泛知识后，在域内数据+提示数据集的修正下，具备在某类具有相对更准确能力的计算系统，其本质是概率模型

据Andrej Karpathy所言，大模型本质由**巨大的参数文件+运行参数的代码文件**两大核心组件构成。其中：1) 参数文件是通过大量数据训练获得的，代表从数据中提取的有用信息和模式（可称为“知识”），其参数量与对知识的记忆能力，与数据量、数据按照不同说明进行曝光的次数有关。2) 代码文件有点类似“大脑”，主要是指导参数文件进行推理和产出，其中，可根据域内数据、提示数据集等进行调参，不断提升模型的准确性。因此，总结来看，**大模型能力获得主要有四步**：1) 针对大量数据进行预训练，让大模型学习并记住知识；2) 明确大模型能力侧重点，即给一定高质量、针对性强的域内数据集微调，让大模型在某些能力更加突出；3) 通过标注人员对大模型给出的答案进行排序，给予大模型正向反馈；4) 强化学习。以上，也说明大模型的能力不是只要参数量足够大就具备涌现能力，是需要逐项优化才逐步具备某些能力的。需要注意的是，**大模型本质是一个概率模型**，可从2个角度理解：

**1) 具有不可解释性**：虽然可以调参，但是并不知道这些参数在做什么，如何协同工作，表示的是什么意思；**2) 幻觉不可消除**：模型的工作机制是基于概率和统计推断进行的，而非真正的理解和逻辑推理，且与预训练的数据量、曝光次数、微调数据量及数据的选择、奖励函数偏好等息息相关。

## 大模型能力获取的流程概览

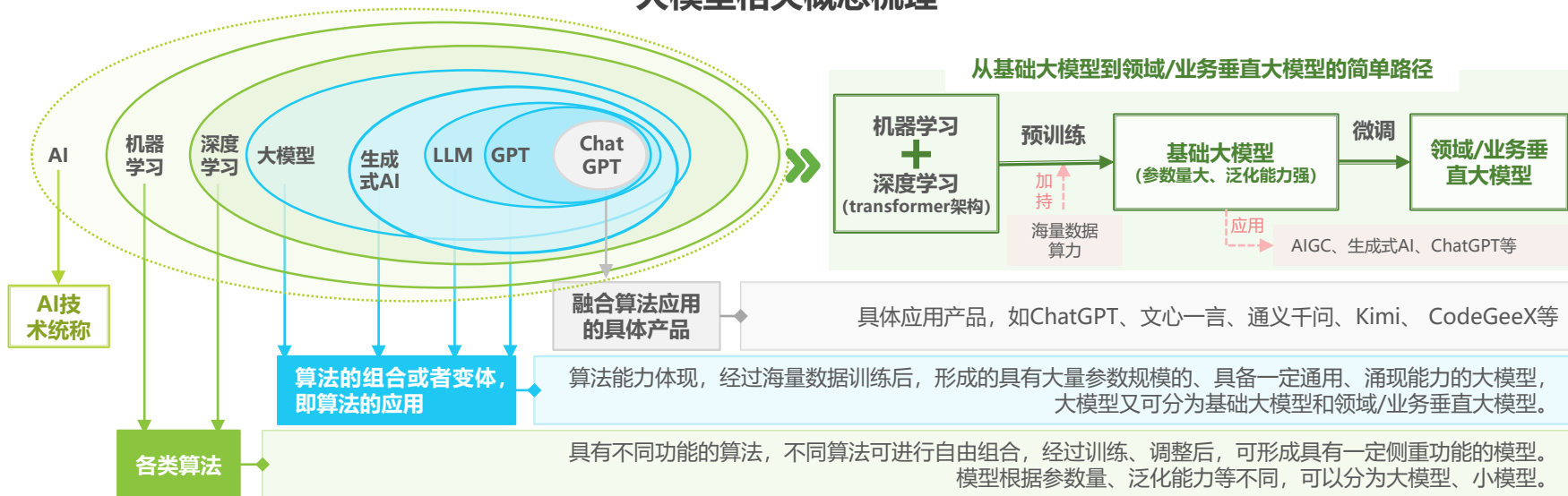


来源：Andrej Karpathy、张奇的演讲，艾瑞咨询研究院自主研究及绘制。

# 大模型及其相关概念的关系是什么？

AI是泛技术的统称，机器学习、深度学习是各类算法，大模型、生成式AI等是对各类算法组合运用的能力的体现，而ChatGPT则是算法应用的具体产品。通常说的大模型是已经预训练完成的具有一定泛化能力的基础模型

## 大模型相关概念梳理



大模型和小模型对比

	大模型	小模型/传统AI能力
算法	均以机器学习和深度学习相关算法为主, 但大模型通常采用transformer架构, 尤其是处理NLP任务时; 而小模型没有统一的架构, 主要取决于任务需求、数据量等	
训练数据量	高, 避免过度拟合	可以在较少数据上训练
参数规模	数亿到数千亿参数	几千到数百万参数
算力需求	高, 需要高性能GPU/TPU和大量内存	可在普通计算机上运行
泛化能力	相对较好, 但可解释性相对低	相对受限, 但可解释性相对高

针对微调的几点说明

- 目的: 在已经训练好的基础大模型的基础上, 结合具体应用场景的特点对其进行调整和优化, 以适应特定的任务或特定领域, 通常会形成领域/业务垂直大模型。
- 内容: 一般, 主要聚焦于权重更新、神经网络层数调整、损失函数调整等方面。
- 特点: 具有训练数据需求少、参数更新有限、训练时间短等特点。

来源: 艾瑞咨询研究院自主研究及绘制。

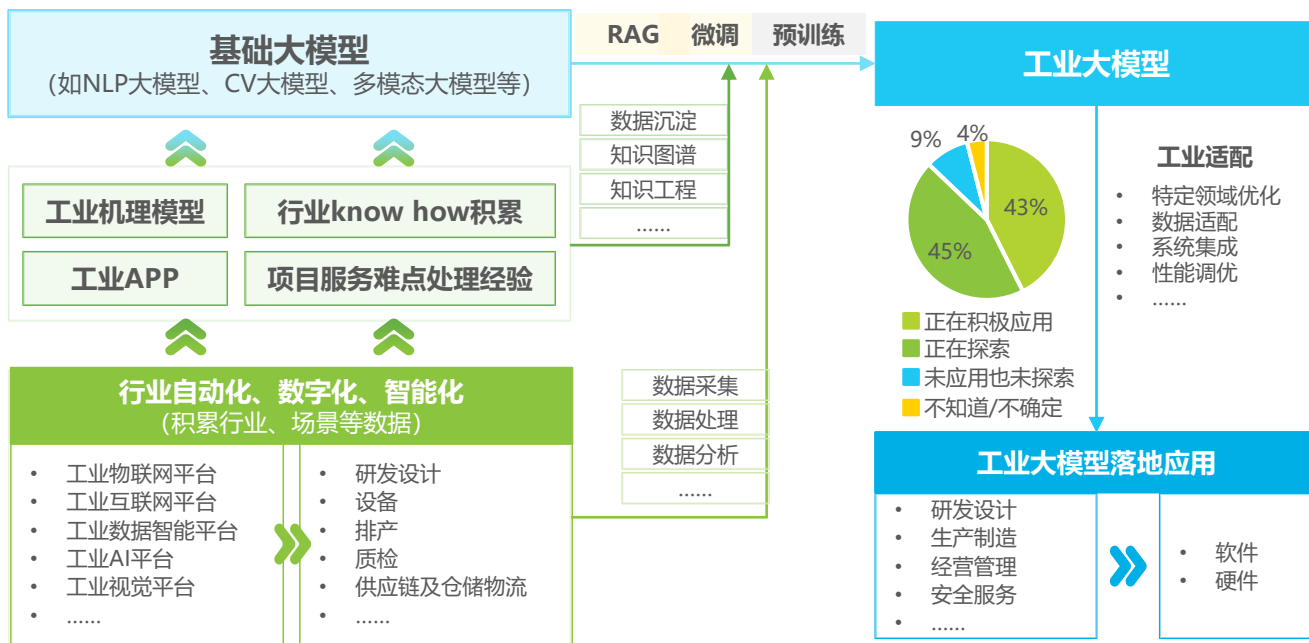
# 大模型在工业应用是否具有可行性?

数据积累、数据资产沉淀等基础以具备，大模型落地工业领域成为可能

大模型的本质是由一系列参数化的数学函数组成的计算系统，是人工智能系统对知识的一种表示和处理方式，本身就具备“数据喂养—认知构建”的“学习”能力。因此，在已经学习海量通用数据、具备一定泛化能力的基础模型的基础上，向其“喂养”特定行业/场景的数据及规则，经过定制开发、调整、训练、调优等操作后，该基础模型将获得特定行业/场景能力的“认知能力”，在一定程度上可称为特定行业/场景大模型。

就工业领域而言，工业互联网、工业智能制造等工作，已经让部分工业领域企业遍历了数据采集-数据存储-数据处理-数据分析-数据资产沉淀-数据应用的过程，部分场景已经准备好了向基础大模型投喂的“数据原料”，当经过数据处理、适配、微调，甚至是训练后，逐步演进为工业大模型，可以解决部分垂直细分场景问题，大模型落地工业成为可能。

## 大模型落地工业领域的可行性



**工业大模型的难点? --- 工业里面的数据通常不具备出现次数多的条件，且这些数据还要以多次不同的说法进行大量曝光。通常需要模型服务提供方基于自己积累的数据训练集去做一些调优或者优化，此时，行业积累就尤为重要。**

**工业大模型通常需要的参数量大概是多少? --- 工业大模型进行应用时，参数量要求通常与应用场景、企业规模等有关，通常8-10亿左右是平衡点，企业基本上可以用了。**

**是否存在万金油的大模型? --- 大目前来看，没有万金油的大模型，脱离场景/问题的大模型暂时不成立。但市场各供给方在追求通过Agent的组合方案去适应不同的场景需求。**



## 02 / 市场篇-初期的产品和服务

玩家涌现, 但产品和服务形式尚在摸索中

# 工业大模型主要玩家有谁？如何切入市场？

厂商类型多样，覆盖软硬件厂商，基因不同所形成的优势各有侧重，但切入市场的角度类似，在市场竞争打法、客户选择、市场策略等方面趋同

根据基因不同可将工业大模型厂商分为8类：1) 互联网大厂、ICT企业、AI厂商，强调AI原生能力和基础大模型能力；2) 传统软件厂商凸显客户服务经验、运营资源优势；3) 制造大厂、机器人厂商彰显行业know how、行业数据、设备应用等优势；4) 安全厂商侧重在企业全局运维中的安全保障。综合来看，虽然各家优势有异，但切入市场的策略相似，即：**锚定自身优势点，围绕具体应用场景摸索大模型在工业的落地性，且目标都是有钱、技术底蕴、有意愿的头部客户。**当前，对比国内外企业的市场切入思路，主要差异有二：**1) 在合作方上**，国外企业具有资源顶端优势，强调从算力底层出发，构建算力-推理的能力地图，而国内企业则强调底层大模型的能力与丰富性；**2) 在服务思路**上，国外企业相对更强调将数据链路服务体系做深，而国内企业更强调各功能模块间的组合。

## 工业大模型主要玩家类型及市场切入点

基因	代表企业	优势	主要切入点	国内外企业切入点对比
互联网大厂	阿里云、百度云等	<ul style="list-style-type: none"> <li>大模型技术积累深、资金雄厚、人才充足，品牌影响力强大</li> </ul>	<b>1-具有通用基础模型的厂商：</b> 以技术为切入， <b>强调自身大模型的能力</b> ，如微调能力、prompt工程能力等。 <b>2-不具有通用基础模型的厂商：</b> 多会引入多种通用基础大模型，他们 <b>主要强调自己工业领域的客户服务经验</b> ，主要分为2类：其一，以RAG服务为切入；其二，自身产品+大模型，为自身产品的性格、功能、效率等赋能，如质检、安全等厂商。	<b>共性</b> <ul style="list-style-type: none"> <li>非常重视人机交互方式、知识问答等基础能力</li> <li>围绕数据的服务</li> <li>关注辅助代码的能力</li> </ul>
ICT企业	华为、浪潮、移动等	<ul style="list-style-type: none"> <li>算力、网络能力、联网能力、整合集成能力强大</li> </ul>		
AI厂商	智谱AI、创新奇智、思谋科技、梅卡曼德等	<ul style="list-style-type: none"> <li>有一定行业know how积淀，并积累部分数据，对大模型的能力理解相对深刻</li> </ul>	<b>两类打法</b>	<b>差异性-国外</b> <ul style="list-style-type: none"> <li><b>1-合作方：</b>多与微软/OpenAI、英伟达合作，重视算力能力+推理能力构建，如西门子、倍福、施耐德等</li> <li><b>2-服务思路：</b>更多从数据采集出发，围绕数据价值挖掘提供服务，但是重视的是数据完整性的服务，即<b>从数据收集-分析-应用的完美闭环</b>，且重视IT和OT的结合，希望能将数据链路服务做深，如Sight Machine、SymphonyAI</li> </ul>
传统软件厂商	中工互联、朗坤苏畅、达观数据、中控技术等	<ul style="list-style-type: none"> <li>多样化的客户资源和服务经验，交付灵活，运维资源强大</li> </ul>		
数据及物联厂商	研华科技、奇云科技、雪浪云、蘑菇物联等	<ul style="list-style-type: none"> <li>数据连接、管理、分析能力强大，对特定场景的分析挖掘经验丰富</li> </ul>	<b>一种选择</b>	<b>差异性-国内</b> <ul style="list-style-type: none"> <li><b>1-合作方：</b>多与国内多家基础大模型厂商合作，如百度、智谱AI等</li> <li><b>2-服务思路：</b>更多的是各能力模块服务平台方的聚合，涉及物联网平台、数据平台、AI平台、大模型厂商等多方</li> </ul>
制造业大厂	卡奥斯、格创东智等	<ul style="list-style-type: none"> <li>行业know how、工业机理积淀深厚，工业领域相关数据丰富，品牌影响力大</li> </ul>		
机器人厂商	达闼机器人、拓斯达等	<ul style="list-style-type: none"> <li>对具体场景对设备的需求了解深，与大模型能力结合时相对更了解其应用范畴</li> </ul>	<b>一种策略</b>	
安全厂商	天融信、启明星辰等	<ul style="list-style-type: none"> <li>安全服务经验丰富，能更好的从安全运维的角度思考系统安全问题</li> </ul>		

来源：艾瑞咨询研究院自主研究及绘制。

©2024.9 iResearch Inc.

# 工业大模型的产业价值流转情况如何？

上游服务方整体集中，且头部效应明显；中游市场参与玩家众，虽服务能力各有侧重，但结合行业know how/行业数据积累进行基础大模型微调、RAG是主要服务思路；下游市场需求广阔，但尚处于探索阶段

## 2024年大模型落地工业领域的产业链及图谱



注释: 1)每个类别企业并未详尽; 2)企业排名不分先后。  
来源: 《2023年中国AIGC产业全景报告》, 艾瑞咨询研究院自主研究及绘制。  
©2024.9 iResearch Inc.

# 工业大模型玩家将如何挣钱？

## 定制化的综合解决方案是当前核心收费模式，更多收费模式可能需探索

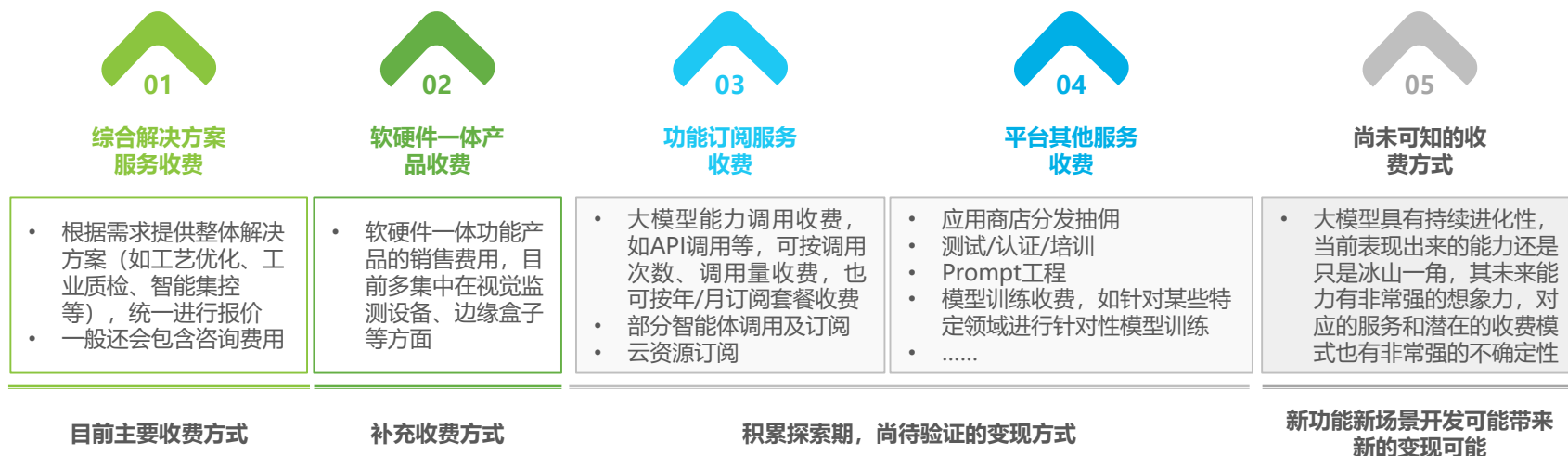
大模型落地于工业领域尚处于萌芽期，其产品形态、服务内容都还在逐步摸索中，其变现方式虽有一些相近服务或行业可以参考，整体不确定性非常高，但一切尝试也都有意义。

工业大模型供给方常规的做法是：与基础大模型厂商合作-测试大模型特点、性能-结合自身产品/能力进行微调-根据客户需求进行定制化服务。因此，**目前相对比较确定的变现方式是类似于“保姆式”的针对某个场景/具体功能的综合解决方案服务收费**。除了定制化的解决方案收费外，部门供给方也开始尝试通过软硬件一体产品进行补充收费。

虽然一些基础大模型的厂商API调用是目前较为普遍的变现方式，但深入到工业大模型服务领域，**API调用、智能体调用及分发抽佣等方式的变现还处于积累阶段，目前市场还需要验证**。

值得注意的是，大模型的能力具有很强的进化性，2024年9月OpenAI推出的最新的o1版本，表现出比较强的复杂推理能力，这又是扔工智能新水平的一个重要进展。因此大模型未来的能力极限在哪里，我们尚未可知，大模型能为工业领域带来哪些类型的赋能和服务，同样也未知。正是这份未知，未来工业大模型的变现将存在诸多可能。

### 工业大模型企业盈利变现可能

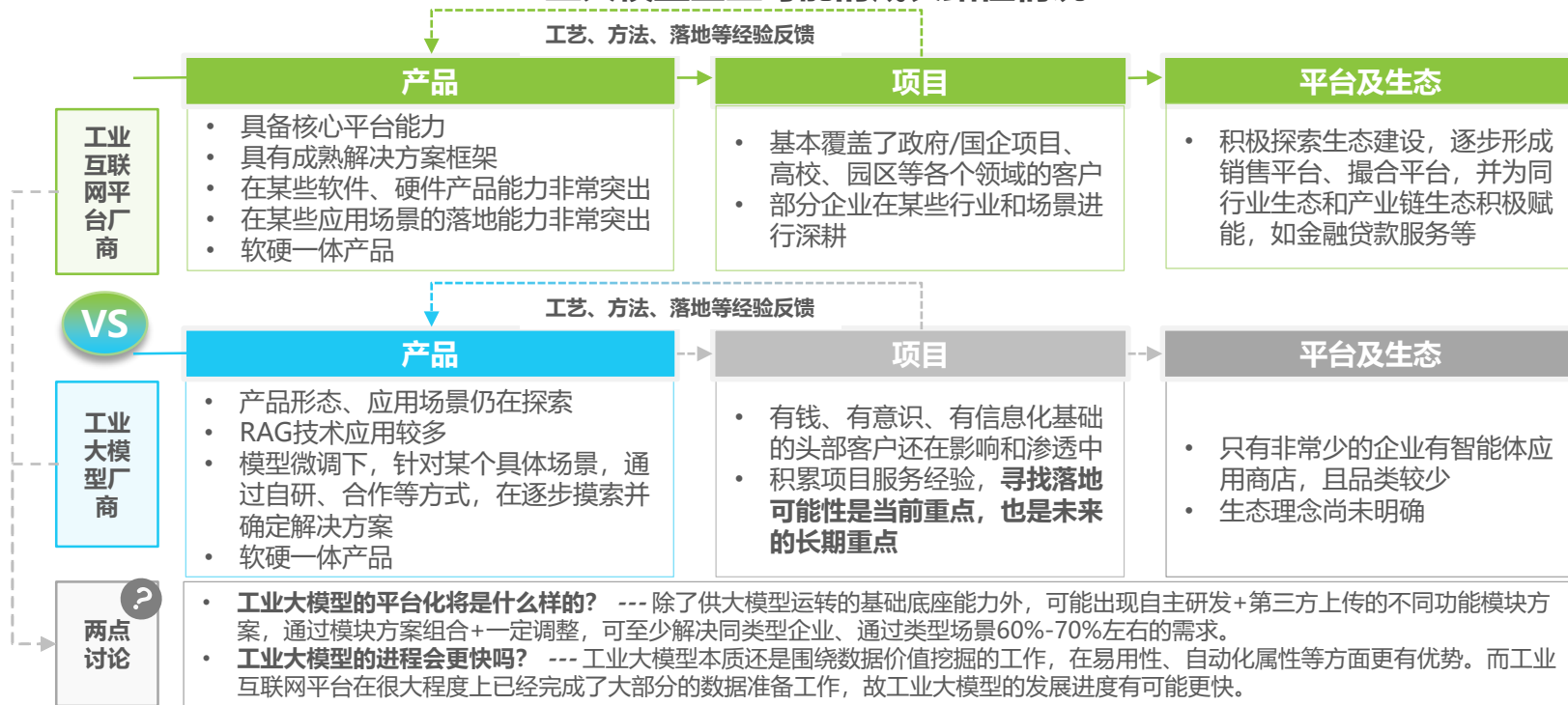


# 工业大模型玩家的成长路径为何？

## 目前尚处于产品雏形期，探讨成长路径为时过早，但很大可能会借鉴工业互联网平台产品-项目-平台的成长路径

工业大模型市场还处于非常初期的阶段，大模型能力更新迭代快，还有非常多尚未未知的潜力，故工业+大模型的应用可能也有无限想象空间。**工业大模型的成长路径可借鉴与参考工业互联网平台的**，主要原因有2点：1) 二者在客户定位、服务内容、服务目标等各方面的重合度比较高；2) 当前阶段，工业大模型表现出来的能力，更像是在工业互联网平台的基础上，对数据信息价值的挖掘进一步深入和易用化，服务思路是一致的。但大模型的成长路径也具有极强的不确定性，因为：1) 大模型能力具有进化性；2) 工业大模型市场仍处于产品雏形期，很多能力仍然停留在产品设想与理论实验阶段，尚未进入真正的工业实践。因此**目前急需解决的是产品-项目的积累与闭环链路构建**，打牢基础后，项目-平台及生态才有机会逐步验证。

### 工业大模型企业可能的成长路径情况



来源：《2023年中国工业互联网平台行业研究报告》，艾瑞咨询研究院自主研究及绘制。



# 大模型落地工业的产品服务架构是怎样的？

## 模型及服务是大模型落地工业的核心，也是大部分工业大模型厂商对外赋能的基础

大模型落地工业领域的服务框架依然分为基础底座-模型及服务-模型应用三层。其中，**模型及服务是核心，主要提供两大类型的服务**：1) 提供预训练模型，可供其他企业通过调用/部署，提供基础的大模型能力；2) 提供预训练模型后服务，如辅助微调、提示工程、训练、压缩、优化等，可供模型能力进一步适配与适用。当前，工业大模型市场的模型应用都是在大模型的基础能力+模型的适配与适用的基础上进行定制衍生的。**2024年大模型落地工业领域的服务体系架构图**



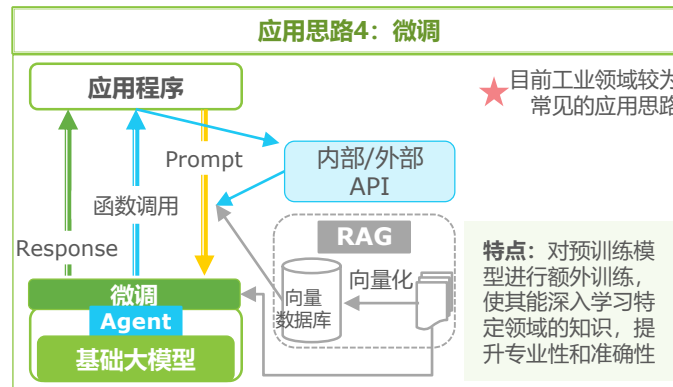
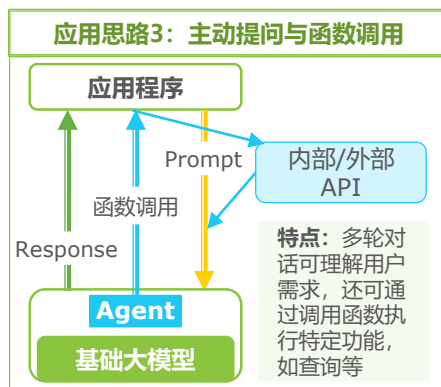
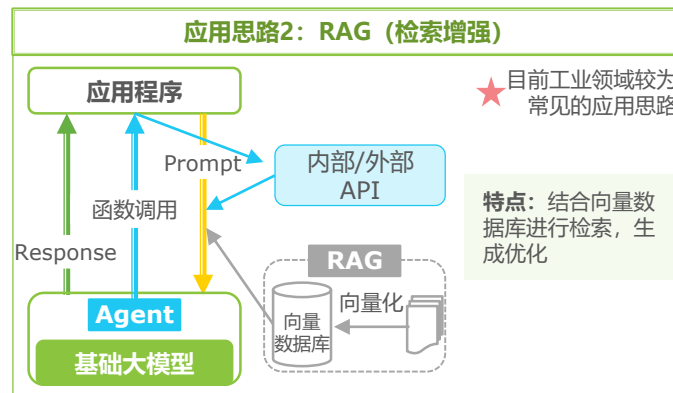
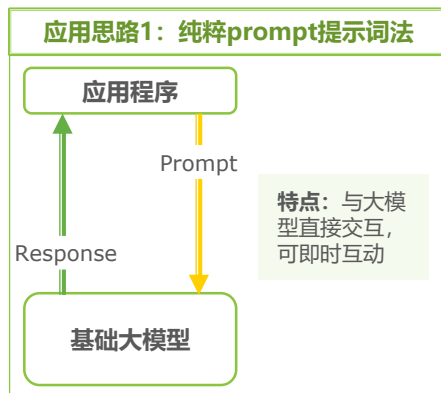
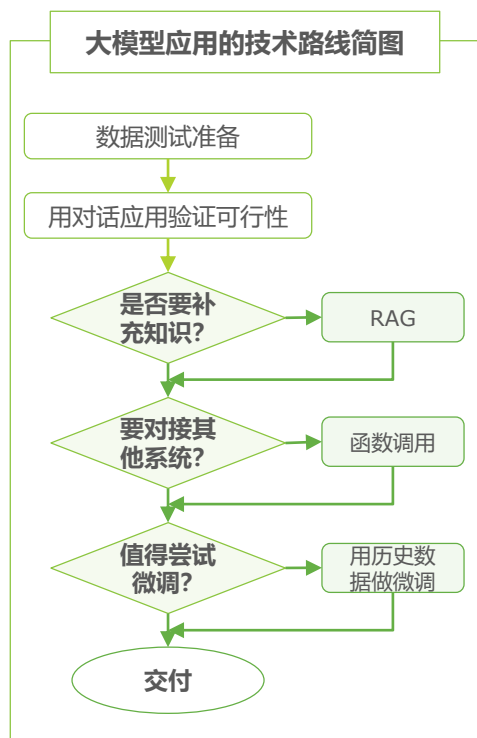
来源：艾瑞咨询研究院自主研究及绘制。

# 大模型落地工业的应用思路有哪些？

## 4种应用形式各有千秋，其中RAG、微调是目前较为普遍的用法

目前大模型落地工业主要有4种应用思路，分别是直接使用、RAG、结合函数调用、微调。其中，**RAG结合向量数据库进行检索、微调是目前落地性相对最常见的思路**。RAG的语料库都是比较贴近实际的工业生产运营的know how、流程、制度等信息，能为生成结果做验证，在一定程度上有利于纠正“幻觉”。目前安全、问答系统等场景用的比较多。微调则是会结合一些域内数据或者具有行业know how的提示数据集对模型进行额外训练，使模型能深入学习特定领域的知识，提升专业性和准确性。

### 大模型落地工业应用的几种思路

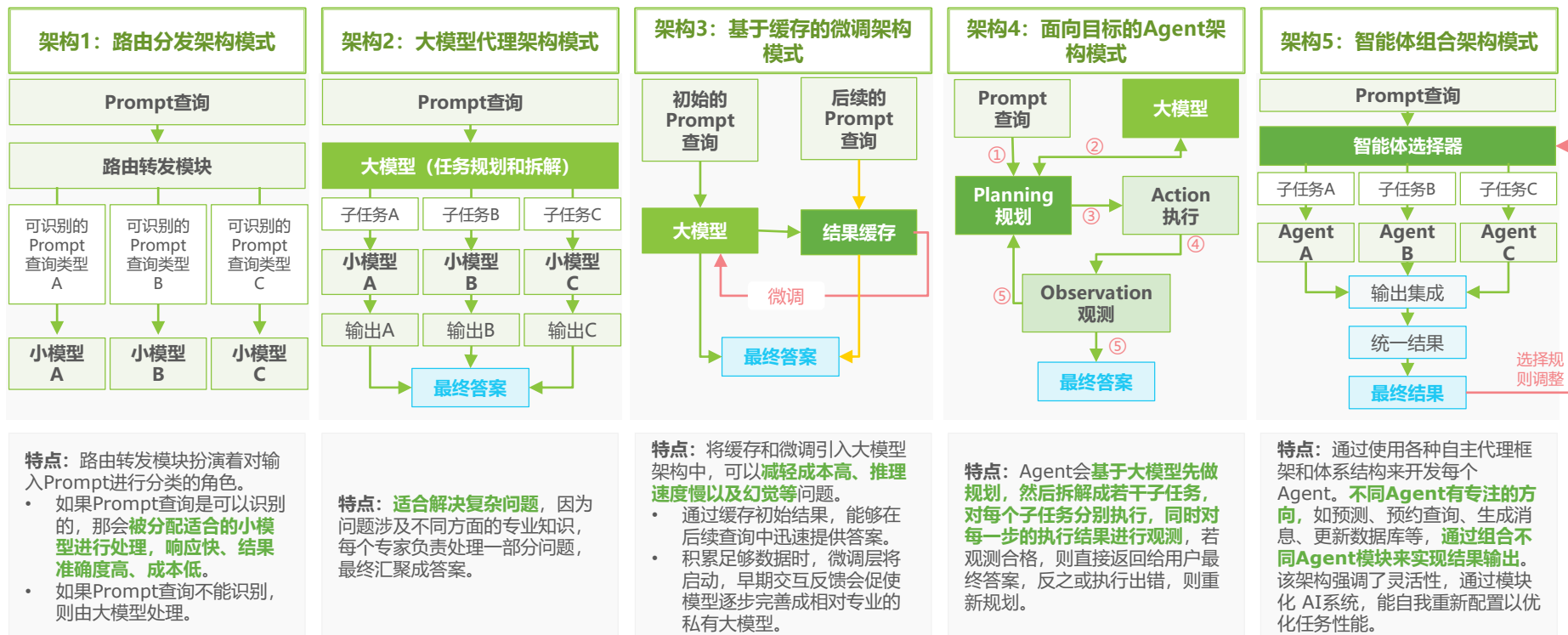


# 大模型落地工业的常见架构有哪些？

## 大模型落地工业时有5种常见架构，没有最好，只有相对最优

目前大模型落地工业主要有4种应用思路，每一种应用思路都可能有不同的架构去实现相应的结果或目的。目前，大模型落地工业时常见的架构有5种，不同架构各有特色。对于供给方或者需求方而言，不同架构没有绝对的优势，更多是根据不同需求，结合效果、性价比等考量后，做出的架构选择。

### 大模型落地工业应用时常见的几种架构



# 大模型落地工业的产品落地模式有哪些？

## 大模型落地工业时主要以软件为主，但具体产品形态还在摸索中

目前大模型落地工业时的**产品切入角度主要有5大类**：基础底座、工具、软件服务、硬件产品、硬件赋能。**1) 基础底座与工具**主要由一些基础模型厂商提供，具有较强的头部效应，未来随着大模型能力的提升，有助于赋能软件开发、测试等全流程，以及硬件端的具身智能。**2) 针对具体应用时**，以一些工业互联网平台厂商推出为主，但各方尚在探索大模型与自身产品及服务的结合点，产品形态还较为稚嫩，多以基于通用大模型进行微调、RAG验证后形成自己的大模型，然后依附于已经存在的产品（如工业APP）上，单独形成具体应用产品的并不多。未来，随着大模型能力稳定性、准确性提升，大模型有望成为新的流量入口，独立成一个新的产品的也并非不可能。

### 大模型落地工业的产品情况

产品属性	主要服务内容	主要价值	收费模式	产品形态	典型玩家	未来畅想
基础底座	<ul style="list-style-type: none"> <li>大模型技术框架服务</li> <li>基础大模型</li> </ul>	<ul style="list-style-type: none"> <li>提供模型开发、训练环境、技术等</li> <li>提供已经预训练好的大模型</li> </ul>	云服务、模型API调用收费等			<ul style="list-style-type: none"> <li>模型能力提升，模型服务的生态越来越健全，工具链趋于完善，为软硬件赋能：               <ol style="list-style-type: none"> <li><b>软件端</b>：赋能软件开发、测试、验证等；</li> <li><b>硬件端</b>：应用场景敏捷调整、具身智能、端到端执行复杂任务等。</li> </ol> </li> <li>通用数据趋于集中，但行业数据还是分散，未来政府主导的数据交易可能会让部门垂直行业数据标准逐步走向流通。</li> </ul>
工具	<ul style="list-style-type: none"> <li>模型全流程开发</li> <li>模型优化、评估、压缩、增强、部署调用等</li> <li>模型安全</li> <li>数据管理</li> </ul>	<ul style="list-style-type: none"> <li>模型开发等服务</li> </ul>	根据具体服务内容进行收费，如部署、精调、评估等	软件调用，一般都会结合工业APP进行功能调用	百度、阿里、腾讯、华为等	
具体应用	<b>软件服务</b> 基于通用大模型的能力，提供具体的大模型产品 基于通用大模型的能力，提供微调、RAG等，形成新的大模型服务产品	<ul style="list-style-type: none"> <li>多轮对话、代码生成、图片生成等</li> <li>知识问答、代码生成、工艺设计/优化、AI检测、设备运检助手、AI客服等</li> </ul>	API调用收费 项目定制服务收费		百度、阿里、智谱AI、新华三等 百度、格创东智、浪潮云洲、朗坤苏畅、天融信、创新奇智等	<ul style="list-style-type: none"> <li><b>加强时序数据的处理</b>，海量工业设备与过程数据得以利用，赋能生产制造等核心环节</li> <li><b>多模态能力创新与增强</b>，从赋能经营管理软件逐步渗透到研发设计（如仿真验证等）、生产控制软件（如人机交互、生产数据分析等）</li> </ul>
	<b>硬件产品</b> 软硬件一体化产品	<ul style="list-style-type: none"> <li>大模型一体机</li> </ul>	硬件产品付费	硬件产品	卡奥斯、华为、科大讯飞、达观数据、智谱AI、开普云、思谋科技等	
	<b>硬件赋能</b> 成熟产品叠加大模型能力，如与机器人、scada、PLC、客户端产品等结合	<ul style="list-style-type: none"> <li>机器人控制指令更改、辅助编程等</li> </ul>	—	软件调用	达闼机器人、ABB、倍福、西门子等	

来源：艾瑞咨询研究院自主研究及绘制。

©2024.9 iResearch Inc.

www.iresearch.com.cn

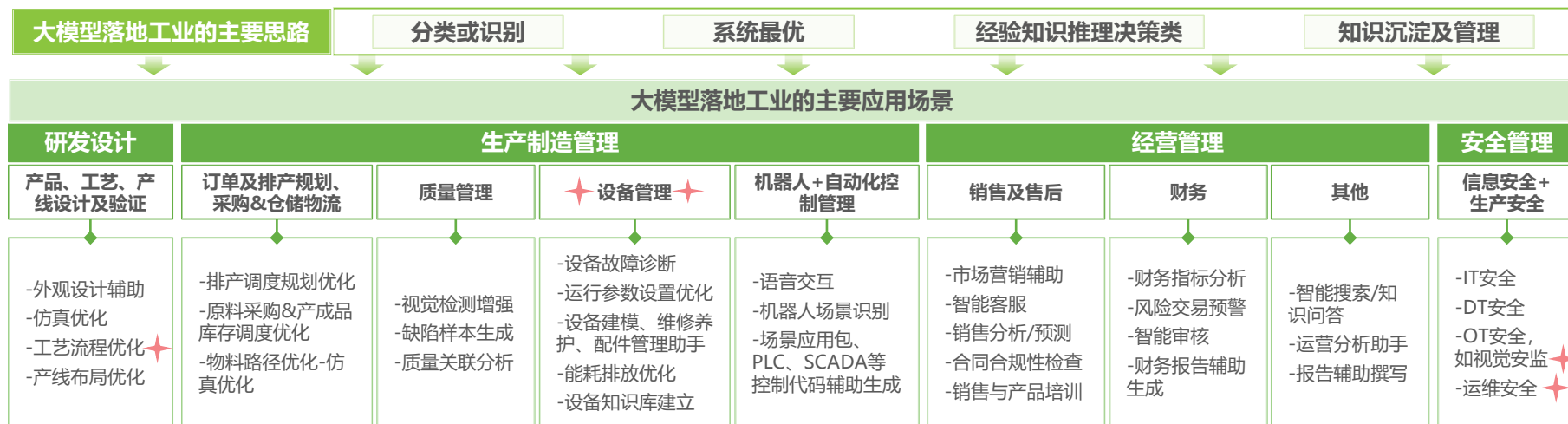
17

# 大模型落地工业的应用场景有哪些？

偏运营的、有一定容错能力的场景是当前重点，生产制造等核心场景的探索需要静待模型进化以及CV大模型、多模态大模型的发展

目前大模型落地工业主要表现为3大特征：1) 大模型无法保证输出的准确性，故**当前有容错能力的场景是各家探索重点**；2) 大模型的应用无法脱离行业know how的加持，基本是都是通过微调或者RAG技术进行调整，故**掌握着丰富工业数据、工业机理模型的企业在工业领域将更有优势**；3) 目前还是以分析和预测为主，决策还是非常初级的辅助决策，未来控制和应用决策能力的发展需要充分发挥时序数据（如设备、产线运转数据、监控数据等）、IT和OT数据相结合后的价值，故大模型能力进化与其他类型大模型的成熟变的非常迫切。

## 大模型落地工业的主要应用场景总览



### 几点说明

- 1) 场景探索依托于大模型的特性展开：**当前以大语言模型为主，生成式、相对强的泛化能力且存在结果不准是其本身特性，各玩家在探索应用落地时，无法避免。各家整体能力输出还是以**辅助性为主**，故偏运营（如销售、售后）、信息比较结构化/标准化的场景（如设备管理）领域是各家尝试重点，在工业领域，除了常规的智能客服、知识问答外，**安全运维、设备管理**是各家推出的比较多的场景，除此之外，也有企业尝试赋能**工艺流程优化、生产整个环节赋能**。（详情见案例）
- 2) 应用价值需要行业know how的叠加：**大模型的应用需要行业数据和工业语料的投喂，需要解析并内化工业专家“看不见的经验”，这些域内数据可对大模型的认知及输出能力进行纠偏，方能在一定程度上保证准确度。
- 3) 潜力场景值得期待：**当前大模型应用时的能力侧重于分析和预测，决策与控制将是下一步的重点，也是渗透核心工业场景的关键。未来随着模型能力进化与其他模型的成熟，赋能软件全流程开发测试、赋能硬件具身智能、赋能操作系统自我进化等场景都是非常值得期待的。

注释：1) 应用场景的梳理主要是按照工业核心流程展开；2) ★ 代表在该场景下，有相应的落地案例展示，且排名不分先后。

来源：艾瑞咨询研究院自主研究及绘制。

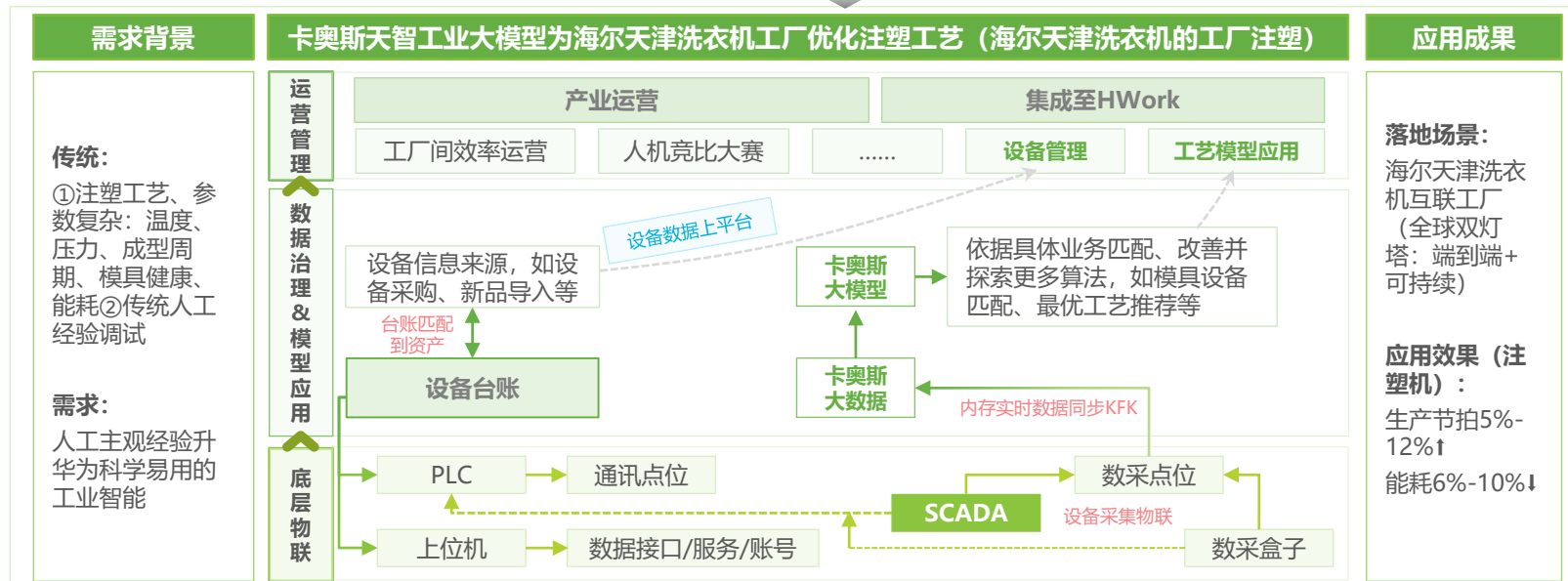


# 工艺优化落地案例案例-卡奥斯

卡奥斯基于1个数据底座+3大能力引擎+N个应用场景的整体架构为企业赋能，实现原始生产数据到模型能力转化的闭环

天智工业大模型架构：以大连接、大数据技术为核心的卡奥斯数字工业操作系统为数据底座；以天智工业大模型、机理模型库、专家算法库为3大能力引擎；以为企业量身打造的天智工业智能体应用平台为核心赋能N个应用场景。

## 卡奥斯天智工业大模型落地实践案例

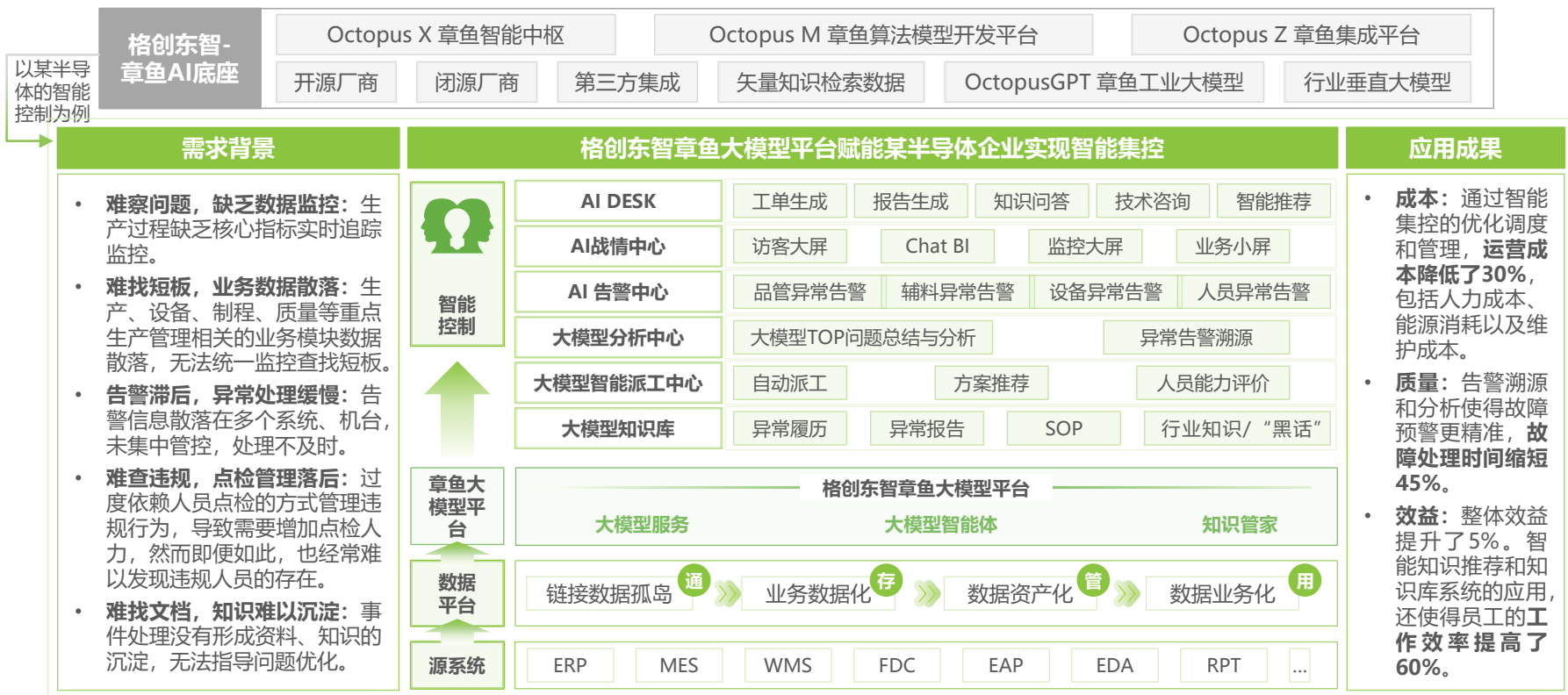


# 智能集控落地案例-格创东智

OctopusGPT是章鱼大模型平台的基座，向上基于数据资产和行业理解提供文本、控制、检测等各类服务

基于章鱼大模型平台的智能集控系统，实现了智能知识推荐、知识库系统，报表智能生成、告警内容深度分析、智能派工等核心功能，全面提升了工厂运维管理的智能化和高效化水平，为企业持续创造价值和赢得竞争优势。

## 格创东智工业大模型赋能智能集控落地案例



# 维修辅助落地案例-智谱

## 智谱通过提示语工程-外挂实时库-微调-二次训练-预训练等流程推动大模型落地于汽车维修辅助，提升故障排查效率

智谱GLM系列大模型赋能维修助手，一线的售后咨询人员，车辆维修人员可以通过耳机+麦克+触屏的智能终端方式，实时与维修人员交互，提供需确认问题/处置方案/预诊断结论等。

### 智谱大模型赋能汽车维修辅助落地案例



# 工业安监落地案例-中移上研

依托中移动智算能力及九天·众擎基座大模型能力，构建“1个基座大模型 + 1个工业大模型（含工具链平台）+ N个应用场景”的“1+1+N”工业人工智能技术能力体系和应用生态系统

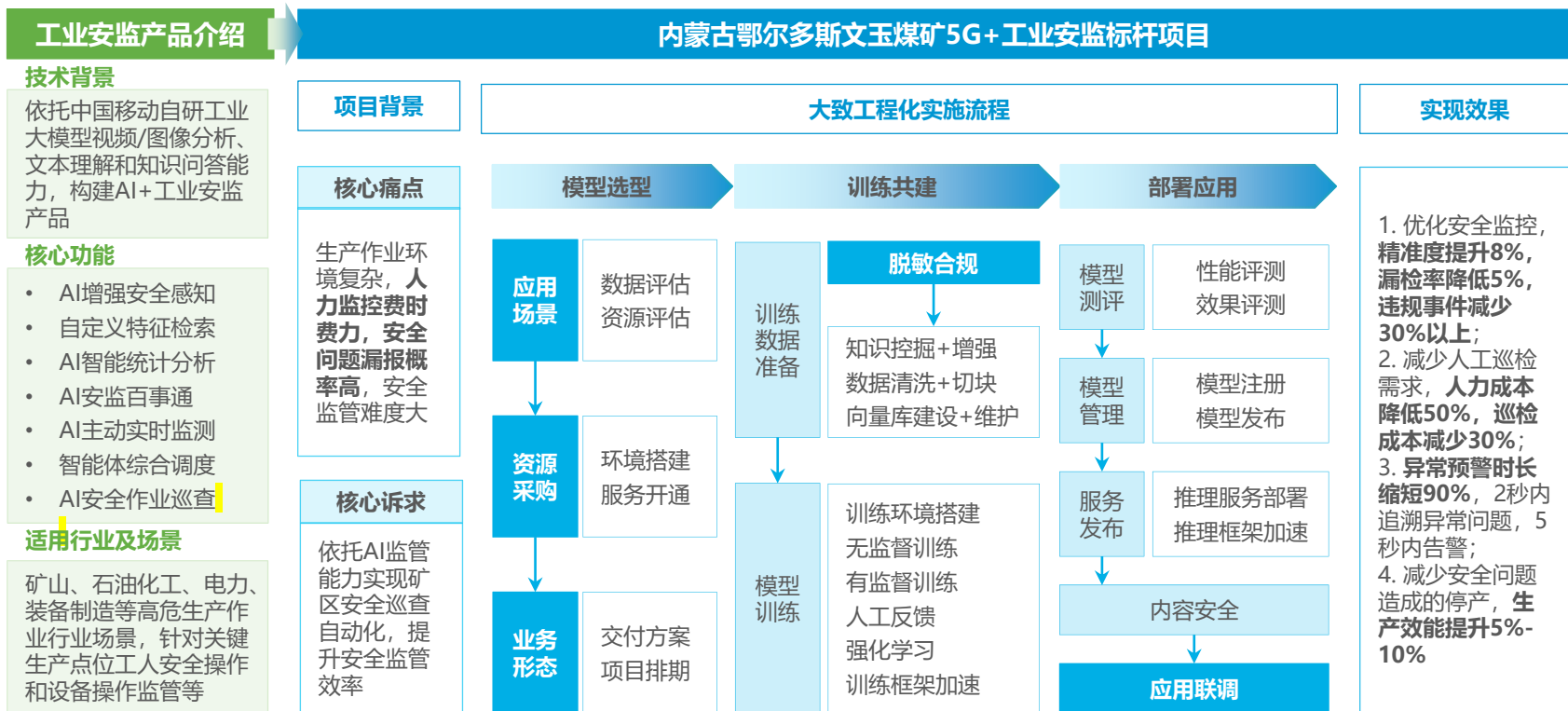
## 中国移动九天·工业大模型及AI应用体系架构



# 工业安监落地案例-中移上研

针对客户需要提升安全监管效率的需求，依托中移工业大模型，经过模型选型-训练共建-部署应用等流程完成工业安监落地，实现全方位、7\*24小时视频违规分析，提升安全监管效率

## 中移（上海）产业研究院九天·工业大模型落地于工业安监的案例

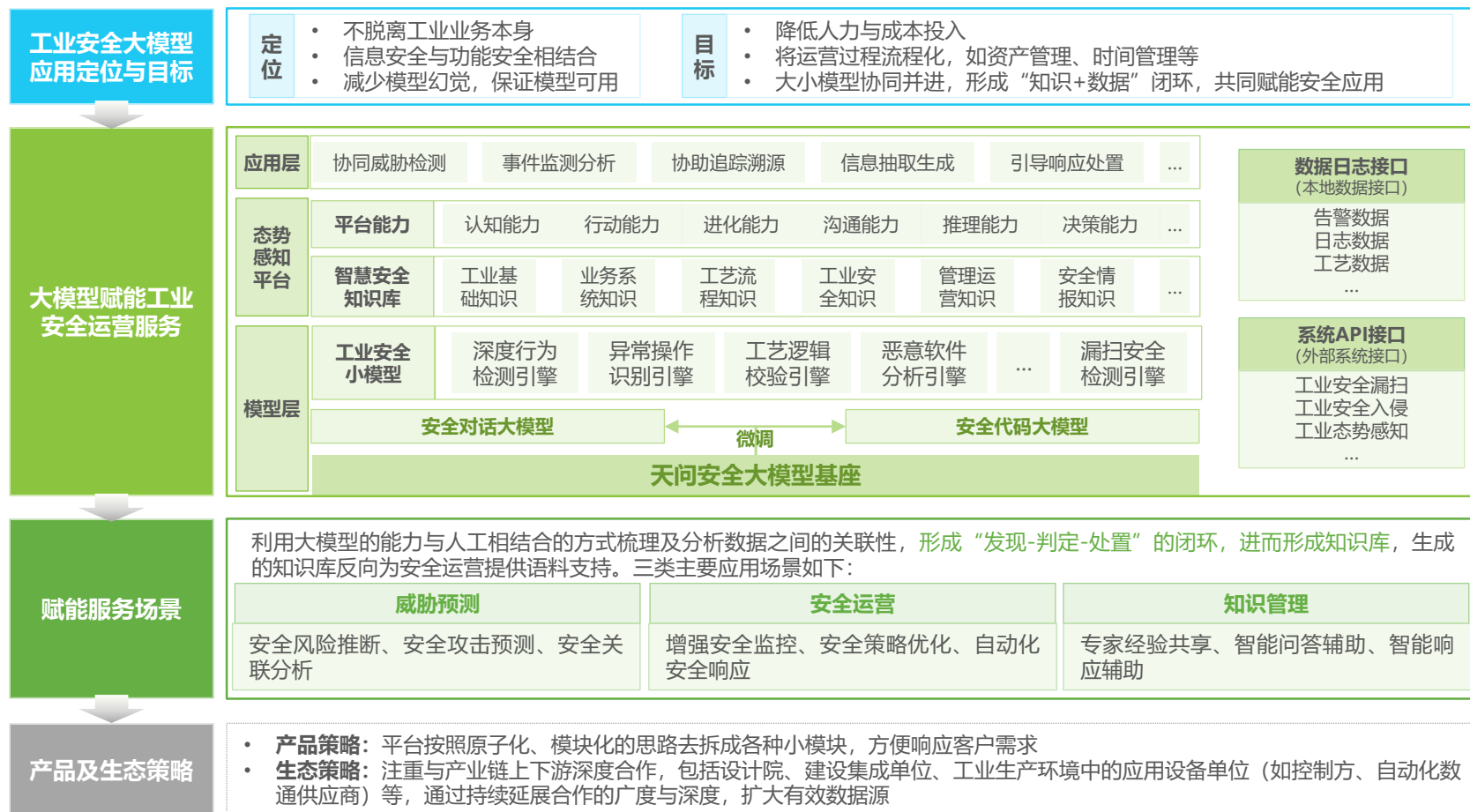




# 安全运营落地案例-天融信

以“双安融合”为基础，结合工业业务本身，致力于为客户提供“发现-判定-处置”的闭环服务

## 融合工业安全大模型的工业安全运营服务体系



来源：天融信，艾瑞咨询研究院自主研究及绘制。

# 安全运营落地案例-天融信

天融信基于工业安全大模型的安全运营平台提供监控管理生产流程，智能优化安全策略，构建资产、威胁、脆弱性的安全风险运营管理闭环

天融信基于项目业务特点与安全需求，聚焦安全事件深度治理与智能分析，整合多源安全数据，利用工业安全大模型构建安全运营管理平台，通过AI模型实时监测工业控制领域的安全威胁，预测潜在风险并智能制定应对策略。同时，通过安全策略引擎驱动决策辅助，结合安全知识与威胁情报进一步优化告警降噪，不断加强通报预警与事件处置的管理，大大提升安全运营效率，保障生产业务的连续性、稳定性及安全性。

## 某电力装备制造集团安全风险运营管理案例



来源：天融信，艾瑞咨询研究院自主研究及绘制。

# 大模型落地的主要挑战有哪些？

模型本身有缺陷、数据准备不足、应用不够深入等是当前遇到的主要挑战，且各方相互影响

大模型落地工业的挑战主要聚焦于**模型、数据、应用、商业变现**几方面。**针对模型**：其根本是模型本身的问题，即存在幻觉、不可解释性、成本高等问题。**针对数据**：更多是数据共享性、数据质量等问题。**针对落地应用**：则关乎决策成本和应用深度等问题。**针对商业化**：则与前面三大挑战息息相关，即模型与数据相互促进，模型进步将推动应用落地，应用的推进能带来更好的商业化变现。

## 大模型落地工业的主要挑战



### 模型相关

- **可靠性问题**：大模型的幻觉问题不可忽视，这有悖于工业场景容错率低的问题
- **可解释性问题**：当前是基于概率和统计推断生成的，而不是明确的因果关系，具有不可解释的问题
- **实时性问题**：目前大模型还不能利用工业的时序数据，无法快速响应生产过程中的变化
- **成本问题**：私有化部署成本高，动辄百万级别起步，企业难以承受
- **中文语料问题**：中文语序灵活、语义内容复杂，且偏重语境，目前中文标注语料相对欠缺



### 生态和商业化相关

- **生态方面**：1) 目前处于需求市场主导合作的阶段，供需合作推进相对难；2) 供给市场整体解决方案尚未明晰，生态伙伴网络尚未建立
- **商业化方面**：定制化为主，其他变现可能还在萌芽



### 数据相关

- **数据数量问题**：1) 企业数字化程度参差不齐，很多数据没有采集；2) 行业内数据开放共享机制尚未建立，数据共建困难，工业语料匮乏；3) 时序数据尚未得到有效利用
- **数据质量问题**：1) 工业数据复杂多样，关联性强，生产/设备/环境等数据的准确性、完整性不足；2) 工业数据标注需要咨询专业人士帮助，解析“暗数据”的实质，耗时又费力
- **数据安全问题**：若要深入到核心应用环节，数据往往涉及企业核心竞争力，安全性和合规性要求高

### 落地应用相关

- **ROI**：企业需要明确可量化的评估方案，这与当前大模型应用很难评估量化价值相悖，客户决策成本高
- **应用场景**：1) 目前还是以浅层的工业知识问答为主，工业执行应用层面尚未渗透；2) 往往存在客户数据量小但结果要求高的悖论

# 03 / 思考篇-持进化思想看待行业

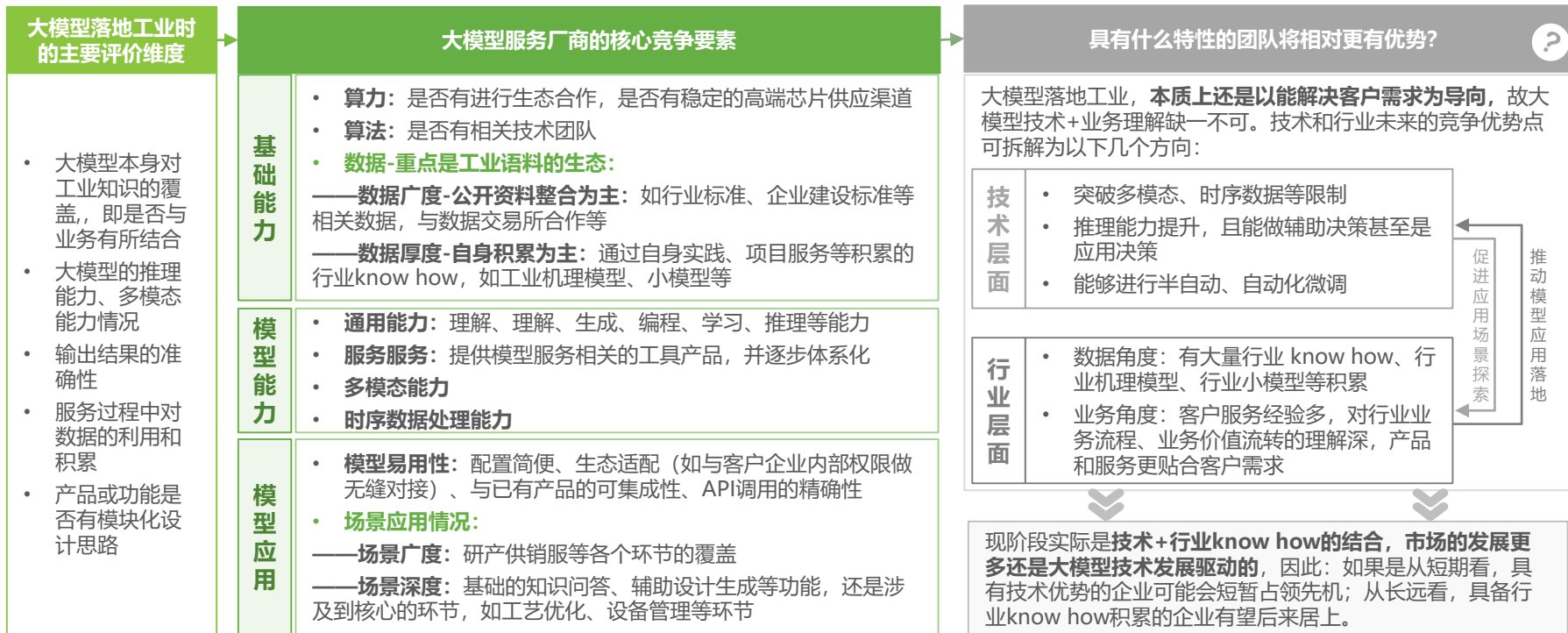
模型能力进化带来不确定，合作利大于弊

# 思考1-大模型落地工业的竞争要素有哪些?

基础能力、模型能力、模型应用是主要竞争点：短期看，具有模型技术优势的企业占优，长期看，则是模型应用更深的企业有望后来居上

大模型落地工业，本质上还是以能解决客户需求为导向。从需求方角度看，客户主要关心两点：**结果准确性+服务于业务**。因此，对于供给方而言，模型技术与业务理解缺一不可，只不过具有不同基因的企业，在**基础能力、模型能力、模型应用**这三方面所具备的优势不同，但这三方面是大模型落地的主要竞争要素，且不同发展阶段其相对优势也有所不同。具体而言：**从短期来看**，工业大模型行业是大模型技术发展带来的，以模型技术推动行业发展为主，此时，具有技术优势的企业将占领先机；**从长远来看**，技术本身是为行业服务的，故具有行业know how积累的企业反而能推动行业走的更远，其相对优势将会更加明显。

## 大模型落地工业服务的核心竞争要素





# 思考2-工业需要的是大模型还是小模型?

## 工业大模型与工业专用小模型并存，且是协同融合赋能工业应用的关系

在进行工业应用时，大小模型各有所长，其能力都不可忽视，当前也不存在谁替代谁的情况。大模型落地工业应用时，主要依托于强大的生成能力和针对复杂信息的捕捉与构建能力，因此在知识问答、文本/图片生成等以创造见长的场景应用比较多。而小模型则凭借高性价比、预测结果相对准确等优势，在工业质检、设备维护等场景高度成熟。

整体来看，大小模型融合赋能工业应用是必然。**原因主要有二**：其一，大模型除了能有效促进小模型推理能力、鲁棒性性能提升外，还能通过调用的形式充分利用小模型，进一步发挥小模型应用的敏捷性；其二，通过小模型归纳总结业务/场景的运转原理后，可帮助供给方预判针对项目设计的大模型行动方案的可行性，进而降低项目时间和金钱成本。

### 大小模型融合赋能于工业应用

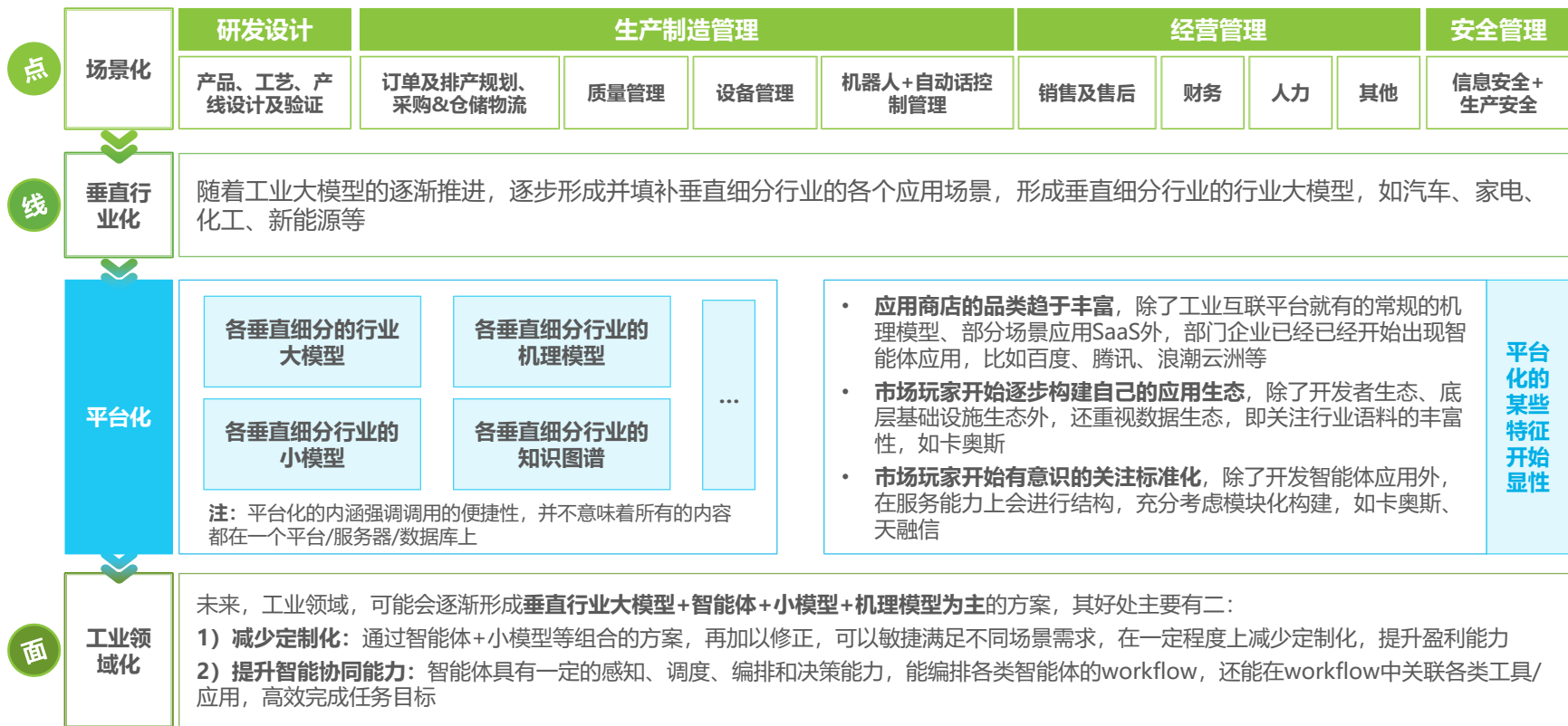
	大模型	小模型
<b>大小模型优劣对比</b>	<p><b>优势</b></p> <ul style="list-style-type: none"><li>通过构建庞大的参数体系来深入理解显示世界的复杂关系</li><li>泛化能力强</li></ul> <p><b>不足</b></p> <ul style="list-style-type: none"><li>数据量需求大</li><li>对算力有要求，成本高</li><li>有幻觉，且具有不可解释性</li></ul>	<ul style="list-style-type: none"><li>通过学习输入和输出之间的关系，进行判断、分析和预测</li><li>成本低、训练速度快、预测结果相对准确</li></ul> <p>训练出来的模型适用范围有限，通常模型A适合于A领域，模型B更适合B领域，通用性不高。新场景通常需要走一遍“数据收集-标注-训练-部署-应用”等一些列流程</p>
<b>大小模型在工业的应用特点</b>	当前更适合创造类的工业场景，或者一些需要复杂计算的场景 <ul style="list-style-type: none"><li><b>创造类场景</b>：主要依托于强大的内容生成能力，如数据/图片/文案/报告等生成、知识问答/搜索等</li><li><b>复杂计算场景</b>：主要利用捕捉并模拟多个系统/流程/要素之间的复杂关系，如汽车模拟碰撞测试、材料性能预测等</li></ul>	已经有深厚的应用基础和经验积累，在生产过程中均有应用 <ul style="list-style-type: none"><li><b>小样本数据</b>：能够基于有限数据支撑精准的判别和决策</li><li><b>对准确性和稳定性要求高的场景</b>：能够针对具体场景进行精细化的调整和优化</li></ul>
<b>大小模型的相互影响</b>	<ul style="list-style-type: none"><li><b>大模型促进小模型性能提升</b>：<ol style="list-style-type: none"><li><b>辅助语料生成</b>：通过大模型生成的数据、图片等信息，可以助力小模型训练，进而提升小模型的性能和鲁棒性</li><li><b>知识转移/转移</b>：知识蒸馏促进了从大模型到小模型的知识转移，增强了小模型的推理能力</li><li><b>辅助调用</b>：大模型通过Agent等方式调用小模型，可负责全局的调度与决策，而小模型可负责具体执行和控制</li></ol></li><li><b>小模型促进大模型的应用</b>：主要利用小模型归纳总结业务/场景运转原理，预判大模型方案的可行性</li></ul>	

# 思考3-工业大模型的服务将会走向平台化?

## 大模型落地工业的服务平台化特征以开始显现，且逐步形成垂直行业大模型+智能体+小模型+机理模型为主的平台化调用方案

大模型虽然在一定程度上具有泛化能力，但具体到应用时，需要进行微调等操作，其具有很强的定制属性，如何减少重复性工作，甚至减少定制化是各供给方共同探讨的命题。而**行业大模型+智能体+小模型+机理模型为主的平台化调用方案**似乎成为大部分企业的共同选择。但**有两点需要注意**：1) 平台化强调知识沉淀、工具等调用的便捷性，而不拘泥于存储形式/路径；2) 大模型工业应用时，其演进具有一定阶段性，基本遵从场景-垂直细分行业-工业领域的路径。

### 工业大模型的落地服务将走向平台化



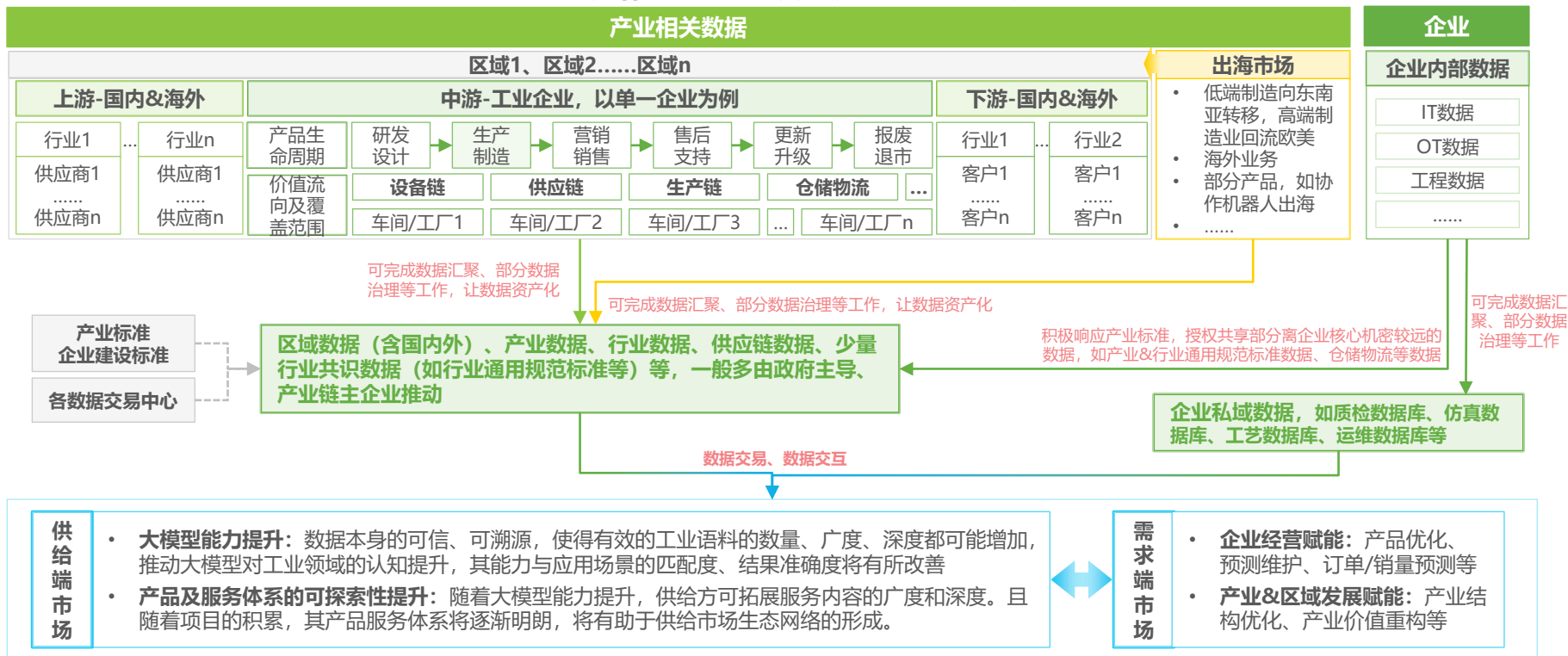
来源：企业访谈，艾瑞咨询研究院自主研究及绘制。

# 思考4-数据拉通对大模型落地工业的影响?

## 产业数据拉通，有利于产业数据资产化，助力工业大模型能力进化的同时，对大模型落地工业的广度、深度都大有裨益

数据成为资产的共识逐步形成，但如何安全规范的集成并使用数据是当前亟待研究并解决的问题。对于工业而言，企业数据往往关系企业核心机密，很难共享，因此**数据拉通的思想主要聚焦于两点**：1) 分离出远离企业核心机密的数据，助力对区域/产业/行业数据链体系化、资产化的形成，如行业产品标准、物流等数据；2) 构筑可信赖的第三方数据交易+数据交互平台，促成供需双方基于某种共同利益/价值，让数据进行流转，如关键设备供需双方数据交互，共同推动设备性能提升。未来，随着产业数据的拉通，工业领域将有望积累更多有效的工业语料，助力工业大模型学习与训练，更好赋能工业场景。

### 产业数据拉通对大模型落地工业的影响



来源：《2023年中国制造业数字化转型落地实践报告》，企业访谈，艾瑞咨询研究院自主研究及绘制。

# 04 / 附录篇-各家观点及鸣谢

Inspiration



秦承刚

卡奥斯

工业智能研究院院长

## ■ 卡奥斯客户服务思路：尽可能降低客户使用大模型的成本

卡奥斯总结过往客户服务的共性能力后，主要做了3件事情：

- 在模型使用界面上，持续优化、标准化，并在生态上做了适配，卡奥斯提供了web端，能与企业内部权限做无缝对接、能与卡奥斯平台上的智能体能力做对接；
- **对接企业已有办公工具：**卡奥斯对接了钉钉、飞书、企微等已经有的办公工具，与企业内部的办公环节就可以无缝集成；
- 中小企业的场景需求相对聚焦，卡奥斯提供一体机产品。

面向客户服务时，卡奥斯做了两个端的服务：

- **客户-研发人员：**企业的IT人员可在平台上开发自己的智能体，具备一定可扩展性。
- **客户-应用端：**可直接使用IT人员开发出来的智能体。

## ■ 前瞻趋势思考

- **大模型+多个智能体的智能平台是企业所需要的：**现阶段，大模型落地工业更多是技术+行业know how的结合，更多的是强化企业自身的产品，给行业带来的变化，可能是点状的变化，还没有面的变化。当行业所有的产品都集成了大模型后，或者都被大模型集成以后，整个行业可能会出现人、软件、设备间交互方式等发生变化。未来，还是需要一个大的以大模型为核心的平台，可以覆盖面的需求。
- **基于大模型形成行业垂直场景方案/工具，推动行业智能化：**在每个行业中，根据行业自身特点，结合大模型，形成专业的工具，让研发设计、生产制造越来越智能化。
- **大模型+智能体快速组合工业各领域/场景解决方案：**基于场景、行业去通过大模型改造后，将逐渐形成一个有各类大模型、各类工具的生态，能将工业领域所有的服务性的企业都整合到生态中，可以很快捷衍生出基于大模型做智能化改造或者转型的整体方案。





陈高平

格创东智

CIO

## ■ 市场服务主要难点

### 1、数据问题：

尽管工业大模型对数据量的需求不如基础模型那般庞大，但为了构建一个可商用的行业大模型，我们仍需整合整个行业大量的背景知识和业务数据。然而，我们面临着一些现实挑战：一是训练数据集往往都是小样本数据，单一企业也难以单独覆盖全面的行业数据，而且存在数据权限等数据安全问题；二是数据准备度不足，大量数据尚未经过有效的治理，质量不尽如人意。

### 2、ROI问题：

虽然大模型应用技术的门槛在不断降低，但大模型应用的业务价值依然难以量化。对于那些不希望公开数据的客户，我们只能本地部署大模型，那么昂贵的AI基础设施势必导致前期的投入巨大，短期内也难以看到明显的业务效益的话，大多数应用项目过不了ROI这关。

## ■ 前瞻趋势思考

### 1、大模型对市场的影响：

大模型技术和其他新兴的信息技术（如前期的大数据）一样，从落地应用探索期到成熟期，我认为基本上是一致的，真正有研发投入且能持续有效地把技术应用落地解决好客户问题的企业将在市场竞争中胜出。

### 2、市场趋势：大模型将不断进化

- **与企业的各种数据做紧密结合：**一方面，业务领域智能知识库和实时智能检测等方向将会加快发展；另一方面，大模型将会深入到工业智能里面的核心生产制造环节，如智能调度机器人/智能排程、AI+BI实时决策等场景会逐步落地。
- **模型的可信度越来越高：**大模型的自我进化、自学习能力会越来越强，在自学过程中，它可能动态的提供一些推理能力和解释能力，最后它的可解释性和可信度可能会越来越高，逐步赢得B端市场的信赖。

# iResearch – 工业大模型专家之声



马霄

天融信科技集团

助理总裁

## ■ 大模型落地服务的难点

- **减少幻觉工作繁琐：**安全垂域模型依赖于大规模的训练数据，然而企业往往缺乏与外部知识源的整合，导致在需要特定领域安全知识时产生幻觉，在此过程企业需长期标注与对齐模型的训练数据，保证数据高质量，从而避免模型输出安全知识不准确。
- **需要供需双方共同主导项目：**企业用户需要将自身涉足行业领域进行know-how知识化，便于推进安全运营建设。

## ■ 前瞻趋势思考

- **输出能力：**垂域安全模型能够提高安全处置的辅助决策，甚至是安全策略应用决策。过去依靠自动化响应+人工研判的方式处置安全问题，未来，大模型可以完成全自动化辅助决策，甚至代替人工研判，直接做出安全策略应用执行。
- **输入能力：**大模型在安全应用中需要考虑如何将安全威胁评估与风险预判结合，形成综合性的解析能力，通过输出的结果集作为风险管理流程化、处置自动化的输入，代替原本人为判定的输入方式。
- **预测性的统计分析能力：**大模型可以代替人工完成复杂繁琐的统计分析工作。统计分析是基于原始数据，目前，该种统计分析已经成熟。未来，大模型将基于海量的业务数据、安全数据、外部威胁情报数据进行整合，形成预测性的统计分析应用。

## ■ 给从事大模型安全服务的建议

- 安全建设阶段，能否在行业场景**具有落地化能力**至关重要，如果单从安全视角出发，而放弃业务的本质因素，将难以形成安全能力的有效应用，与安全建设的初衷背道而驰。
- 垂域安全大模型**可以充分考虑将不同安全能力进行模块化、原子化，全面满足客户需求**，帮助企业降低成本。
- 大模型应**结合工业机理模型**赋能于行业应用，将行业知识积累到本地，细化内容层面，可以在某个特定的细分垂域中更加贴合用户需求，同时要注重对客户服务的积累，才能构筑竞争壁垒。

# iResearch – 工业大模型专家之声



## 周威

中移（上海）产业研究院，工业  
能源产品部，支部书记、总经理

### ■ 工业大模型落地的难点思考

- **数据挑战：**工业数据分散在不同系统中，存在数据孤岛问题，且数据的质量和完整性不足，导致大模型训练效果受限。
- **算力限制：**算力需求大，而高性能算力成本高且可能难以获得，限制了模型的训练和部署。
- **技术泛化性：**工业领域细分行业众多，应用场景复杂多变，大模型需要具备较强的泛化能力以适应不同的工业环境和需求。
- **专业知识缺乏：**工业大模型的开发和应用需要深入理解行业知识和技术，目前缺乏既懂AI技术又熟悉工业领域的复合型人才。
- **成本和投资回报：**工业大模型的研发、部署和维护需要大量资金投入，而短期内可能难以看到明显的投资回报，影响企业的投资决策。

### ■ 工业大模型市场服务难点总结

- **客户私有化需求高：**客户出于安全考虑，往往希望模型能部署在自己的私有云或本地环境中。  
——**解决思路：**提供定制化的私有化部署方案，确保数据安全和隐私保护；与客户合作，建立数据共享和保密机制。
- **对时延要求高，推理速度快：**工业场景中的自动化生产线，对模型推理速度有极高的要求。  
——**解决思路：**优化模型结构，采用轻量化模型设计；使用高性能计算硬件，如GPU或FPGA；实施模型压缩和加速技术。
- **幻觉容忍度低：**在工业应用中，对模型的准确性和可靠性要求极高。  
——**解决思路：**通过高质量数据的收集和清洗，提高模型训练的质量；采用多模型融合技术，提高预测的准确性；实施严格的模型测试和验证流程。
- **工业行业门类多、应用五花八门：**不同行业和应用场景对模型的需求差异大，难以一概而论。  
——**解决思路：**开发模块化和可配置的模型架构，以适应不同行业的需求；与行业专家合作，深入理解各行业特点，提供定制化解决方案。
- **工业样本难获取：**高质量的标注数据，限制了模型的训练和优化。  
——**解决思路：**利用迁移学习，将在大规模数据集上预训练的大模型应用到特定工业场景；开发小模型，专门针对数据稀缺的工业应用，通过少量数据实现快速部署和迭代。

# iResearch – 工业大模型专家之声



叶军

上海电力设计院有限公司 副总

经理 总工程师

## ■ 生产经营过程中遇到的主要难点

- **日常管理方面：**内部众多的规章制度需要员工耗费大量时间精力学习，大模型的出现，让困难出现了转机，通过大模型对于文本处理方面强大的能力，搭建“办公助手”，为员工提供问答式流程辅助、提升日常工作效率。
- **设计业务方面：**电力设计作为高度知识密集型工作，对于人员的有较高的要求，不同工龄、行业背景的人员在能力方面存在较大差异，如何将院内专家的经验性知识以及过往成功设计案例、图纸等形成知识库，进而通过大模型，搭建“专家型数字员工”，是我们下一步想要进一步研究的方向
- **在总承包业务方面：**院内承接过大量总承包项目，总承包项目中设计阶段与实际施工过程中发生的差异，如工程造价等，对于公司而言是非常宝贵的数据，如何通过大模型将这些数据用起来，指导院内新项目的建设至关重要，这也是我们的研究方向之一。

## ■ 工业大模型落地的主要难点思考

- **数据安全和隐私方面：**电力行业涉及大量敏感数据，数据安全和隐私保护是必须考虑的重要因素。如何在保证数据安全的前提下，有效利用数据训练大模型是一个挑战。
- **模型的泛化能力方面：**电力行业具有高度专业化和复杂性，大模型需要具备良好的泛化能力，以适应不同的应用场景和需求。
- **法规和标准方面：**电力行业受到严格的法规和行业标准约束，大模型的应用需要符合这些规定，可能需要时间和努力来适应。
- **人才培养和技能升级方面：**应用大模型需要相应的专业知识和技能，电力行业需要培养或引进具备相关能力的人才。

## ■ 工业大模型应用展望

大模型在工业领域的应用正朝着集成化、自动化、个性化定制、供应链优化、能源管理、智能质量控制等方向发展。它们通过预测性维护、增强现实集成、跨领域知识迁移等手段，提高生产效率和产品质量，同时促进可持续发展。随着技术进步，大模型还将与边缘计算结合，实现更快速的数据处理，同时注重数据安全和法规适应，以确保技术的健康发展。

# iResearch – 工业大模型专家之声



唐定

通答AI

高级产品经理

## ■ 通答AI大模型在工业领域的应用方向情况

### 1、主要服务内容

- **微调**：项目不多，主要是集中在设备销售场景，结合设备参数+企业设备话术展开。
- **RAG**：基于RAG做知识问答是当前主要方向。

### 2、大模型落地过程中主要难点：主要是数据准备的问题

- **数据集成问题**：工业企业部门多，数据零散分布在各个部门，在项目执行过程中，将数据整理汇总起来，在企业内部有一定门槛，且会花很长实践。
- **数据采集问题**：若是需要的一些数据，企业没有收集，则会面临“企业是否愿意改造？ROI是什么？”等问题，影响最终决策。

## ■ 对新进入者的建议

- **技术积累**：要考虑技术积累，因为直接关系到客户体验
- **深入业务**：尽量要深入到行业里面去，要对行业的业务有非常清晰的认识，才能知道大模型怎么去跟他做结合，才能满足客户需求。

## ■ 对下一步大模型落地工业场景的重点的看法

- **工业大模型+小模型相融合**：底层可能会有一个比较通用的，由行业组织协会牵头做的几类或者大类的通用的工业大模型，上层有很多小模型，比如之前说的预测维护、质检等小模型，可以实现小模型的多种组合。
- **潜在机会点**：对于大部分做应用的企业，可能机会更多的是做小模型。因为国家或者协会会去做引导，不会让我们重复造轮子，进而减少投入，最终期望形成合力。



## iResearch – 工业大模型专家之声-智谱相关专家

### ■ 智谱针对大模型在工业领域的应用方向的思考

可以从**任务或者模型的角度+生产制造流程**的环节这两个角度去理解大模型在工业领域的应用场景，整体可以分为：

- **识别类的任务**，用的比较多的技术是多模态大模型。
- **优化任务**，可能用知识图谱+大模型结合，当然会加一些垂直领域，就像之前的机器学习的方式去做一些相关的工作。
- **辅助决策任务**：这个是基于最终的目标，中间使用一些基础大模型和知识图谱相关的一些技术，这个是从算法上的角度去思考的。

如果是从工业流程环节去理解，还有研发与规划、生产过程管控、经营管理优化、产品与服务这四类。因此，大模型在工业领域的应用其实是在3×4的矩阵框里面可以做一些相关的工作。

### ■ 大模型对工业应用市场的影响

- **价格**：相比于之前的CV小模型，现在价格可能有所下降
- **合作对象**：之前的CV小模型，更多是软件厂商与硬件厂商合作；现阶段，更多是大模型厂商与购买硬件的软件厂商合作，把大模型的能力结合软件加在对方购买的硬件里边，来协助软件公司去服务客户。

### ■ 对下一步大模型落地工业场景的重点的看法

- **隐形资产的沉淀**，即怎么能把老师傅的经验积累沉淀下来，不管是工艺还是设计，可能是所有工业企业现在急需的，语言大模型、CV大模型、多模态大模型等都需要。
- 在工业的场景下，除了人以外，**设备是最重要的**，可能更多的工作是在设备级别的，包括这种设备的维修、预测性养护等相关的工作是都是要做的。

# 单位及人员鸣谢

本报告离不开业内企业及从业者的支持，特此鸣谢

单位	人员名称
卡奥斯	秦承刚
格创东智	陈高平
智谱	专家不愿透露姓名
天融信	马霄
中移（上海）信息通信科技有限公司	周威
上海电力设计院有限公司	叶军
中国移动通信集团有限公司	文静
中移（上海）信息通信科技有限公司	熊诚锋
中移（上海）信息通信科技有限公司	张泽宇
中移（上海）信息通信科技有限公司	陈旭
中国移动通信集团有限公司	黄迪
中移（上海）信息通信科技有限公司	熊冠楚
通答AI	唐定

注释：排名不分先后。  
 来源：艾瑞咨询研究院自主研究及绘制。

BUSINESS  
COOPERATION

# 业务合作

## 联系我们



400 - 026 - 2099



ask@iresearch.com.cn



www.idigital.com.cn      www.iresearch.com.cn

官 网



微 信 公 众 号



新 浪 微 博



企 业 微 信



LEGAL STATEMENT

# 法律声明

## 版权声明

本报告为艾瑞数智旗下品牌艾瑞咨询制作，其版权归属艾瑞咨询，没有经过艾瑞咨询的书面许可，任何组织和个人不得以任何形式复制、传播或输出中华人民共和国境外。任何未经授权使用本报告的相关商业行为都将违反《中华人民共和国著作权法》和其他法律法规以及有关国际公约的规定。

## 免责条款

本报告中行业数据及相关市场预测主要为公司研究员采用桌面研究、行业访谈、市场调查及其他研究方法，部分文字和数据采集于公开信息，并且结合艾瑞监测产品数据，通过艾瑞统计预测模型估算获得；企业数据主要为访谈获得，艾瑞咨询对该等信息的准确性、完整性或可靠性作尽最大努力的追求，但不作任何保证。在任何情况下，本报告中的信息或所表述的观点均不构成任何建议。

本报告中发布的调研数据采用样本调研方法，其数据结果受到样本的影响。由于调研方法及样本的限制，调查资料收集范围的限制，该数据仅代表调研时间和人群的基本状况，仅服务于当前的调研目的，为市场和客户提供基本参考。受研究方法和数据获取资源的限制，本报告只提供给用户作为市场参考资料，本公司对该报告的数据和观点不承担法律责任。



# THANKS

艾瑞咨询为商业决策赋能