

工业大模型技术应用与发展报告1.0

2023.12

编写单位（排名不分先后）

牵头编写单位： 中国信息通信研究院

参与编写单位：

百度在线网络技术（北京）有限公司

羚羊工业互联网股份有限公司

创新奇智科技集团股份有限公司

智昌科技集团股份有限公司

中国科学院自动化研究所

中工互联北京科技集团有限公司

西门子股份公司

航天云网科技发展有限责任公司

威派格智慧水务股份有限公司

卡奥斯COSMOPlat

阿里云计算有限公司

腾讯计算机系统有限公司

中科云谷科技有限公司

浪潮集团有限公司

中科视语科技有限公司

苏州海赛人工智能有限公司

北京通用AI研究院

树根互联股份有限公司

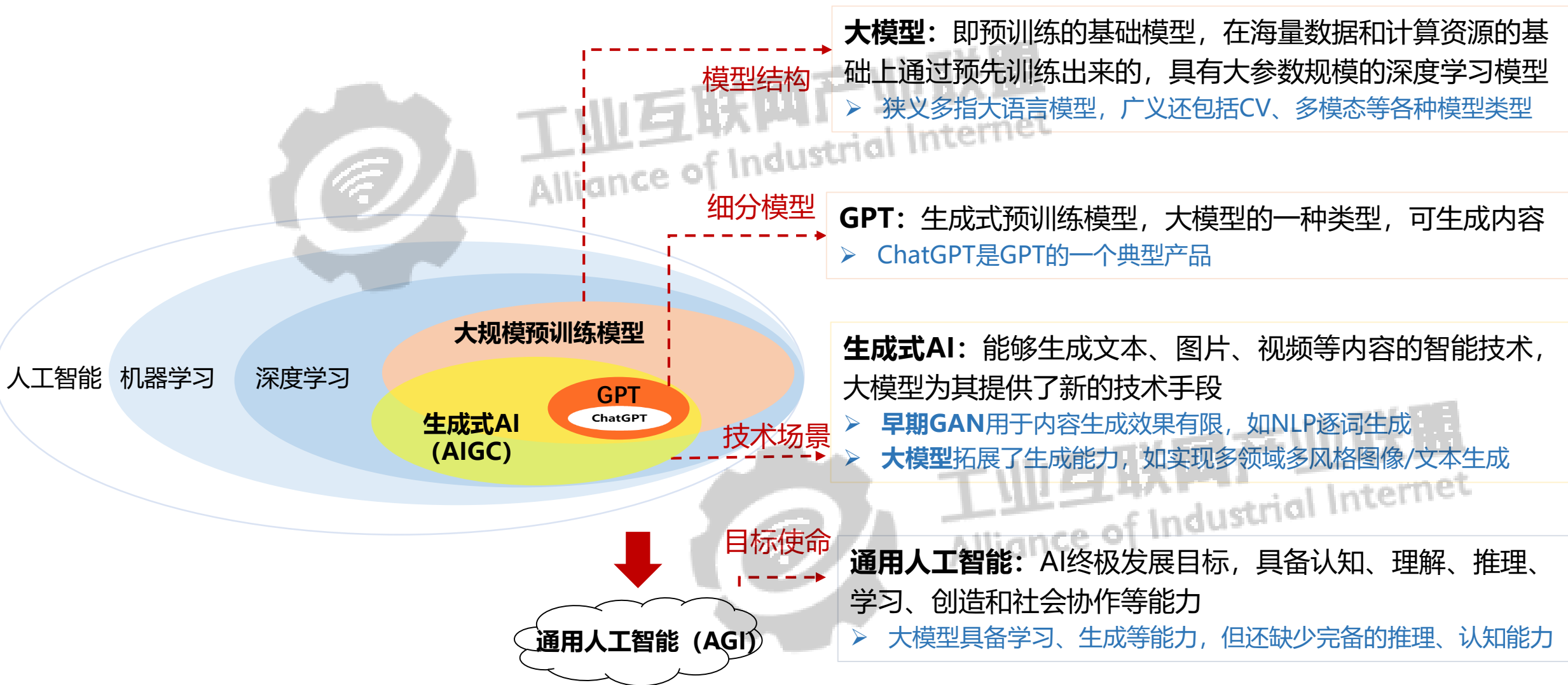
华为技术有限公司

美云智数科技有限公司

工业互联网产业联盟
Alliance of Industrial Internet

工业互联网产业联盟
Alliance of Industrial Internet

人工智能的几个相关概念



1.1 大模型成为通用AI新范式，引发工业应用变革讨论

大模型+大数据+大算力成为
发展主旋律 (ChatGPT)

千亿参数基础模型

GPT-1 : 1.17亿
GPT-2:15亿
GPT-3:1750亿
GPT=4: 1.48万亿.....

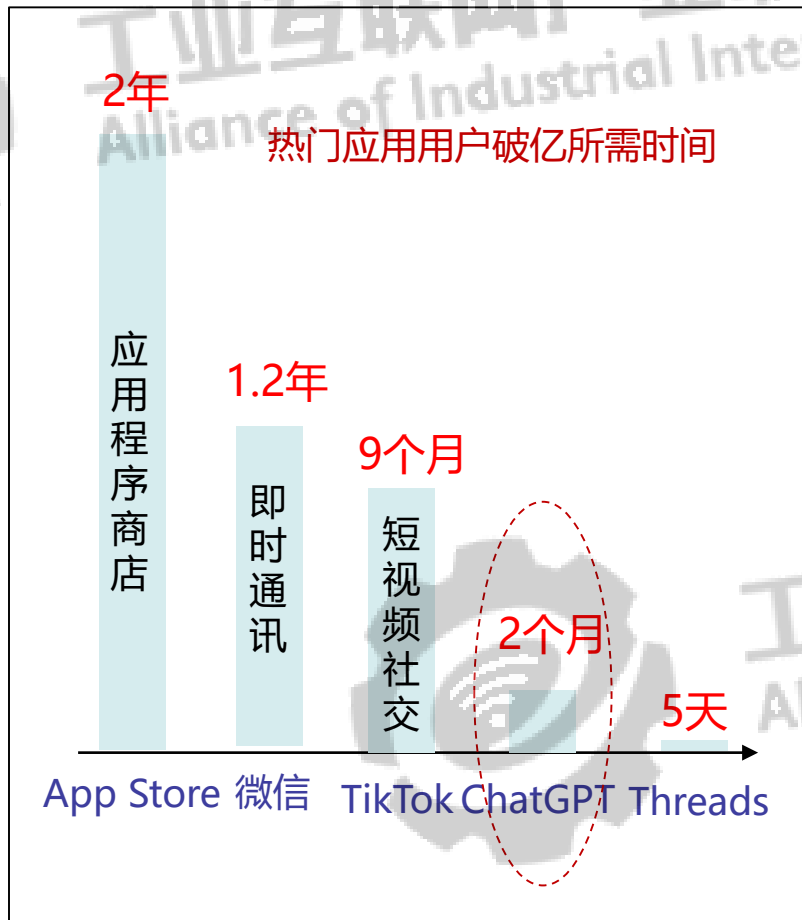
TB级数据

超3000亿单词
830GB代码数据
1000个外包团队标注

E级智能超算总算力

超195PFlops我国最
快超算“神威·太湖之
光” 1.7倍

AI产品渗透速度全球排名第2



引发产业领域应用的热烈讨论与憧憬

- 产业应用场景成为大模型最佳“练兵场”——科技日报
 - AI大模型落地背后，正带来一场智能制造的系统重构——百度
 - 工业大模型将会带来一场新的工业革命，它将来会成为工业领域的基础设施——中工互联
 - 制造业是AI大模型的重要战场，未来10年最大的机会——阿里
- ...

1.2 工业大模型与专用小模型成为工业AI发展的两条协同路径

工业大模型=工业+大模型

1 满足大模型技术基本特征

- Transformer为基础框架
- 在大量通用数据上进行预先训练, 以实现良好的通用性
- 模型参数一般达十亿以上 (最大模型参数已达万亿级)

2 具备在工业各环节进行应用的能力, 或与工业装备软件等融合赋能

工业大模型
Transformer结构

工业专用小模型
传统模型结构

应用层面

- **新场景:** 代码生成、CAD生成等新应用...
- **泛化性强:** 单模型应对多任务, 更适合长尾落地
- **特定任务:** 点状场景应用效果更好

工程层面

- **低成本运维:** 低成本开发+维护
- **模型更新快:** 参数量少, 可进行快速迭代
- **轻量化部署:** 所需存储空间和算力更小

1.3 大模型初步形成赋能工业的核心方式与产品形态

3类主要赋能方式

当前(可用于)工业领域大模型超30个

4类模型产品形态

基于通用底座直接赋能行业

- OpenAI ChatGPT
- SIEMENS PLC编程
- 用友 YonGPT
- 科大讯飞 IFLYTEK 星火大模型
- Paradigm 式说大模型

基于通用底座进行场景化适配调优或形成外挂插件工具

- HUAWEI 制造、矿山等行业大模型
- Baidu 百度 航天、能源等行业大模型
- salesforce Einstein GPT
- AUTHENTISE 3D打印GPT

面向工业或具体任务针对性开发

- Google AlphaFold2
- Meta ESMFold
- DPTechnology 深势科技 Uni-Mol
- HUAWEI 科学大模型

大模型API调用或软件方案

- 基于ChatGPT直接开展智能客服等应用

成熟工业产品叠加基础模型能力

- 倍福将大模型融入 TwinCAT XAE客户端，实现基于对话辅助编程

外挂插件工具

- 工业管理软件企业Authentise推出插件，用户可查询最大的增材制造知识库
- 浙大开发用于表格处理的TableGPT

用于私有化部署的一体机

- 科大讯飞推出星火一体机

2.1 大模型赋能工业领域的适用边界与核心能力

适用问题：大模型并非万金油

● 大场景

- 工业场景具备一定通用性
- 涉及关联复杂的智能任务
- 数据边界对决策效果有直接影响

● 大语料

- 工业场景的基础数据/语料/规则约束充足

● 问题边界清晰

- 结果存在于封闭信息环境，不依赖语料外的信息

核心能力

工业领域应用变革

✓ 语言理解

预置型对话 → 与设备/工业系统的自然交互与推理

✓ 生成创作

规则式生成 → 工业代码/图文内容的“涌现式”生成

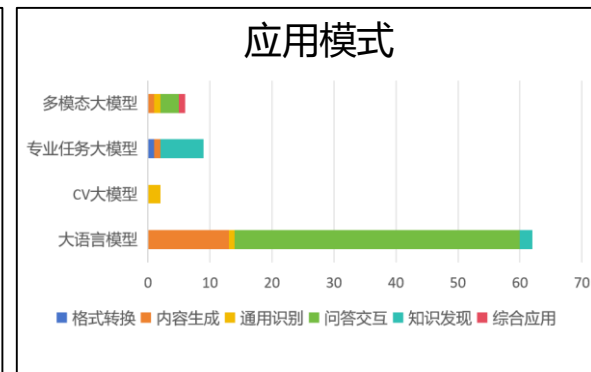
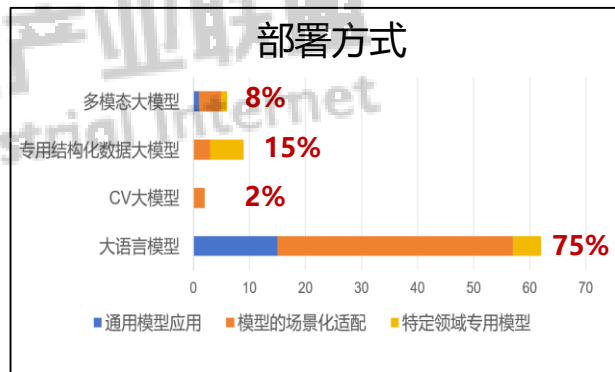
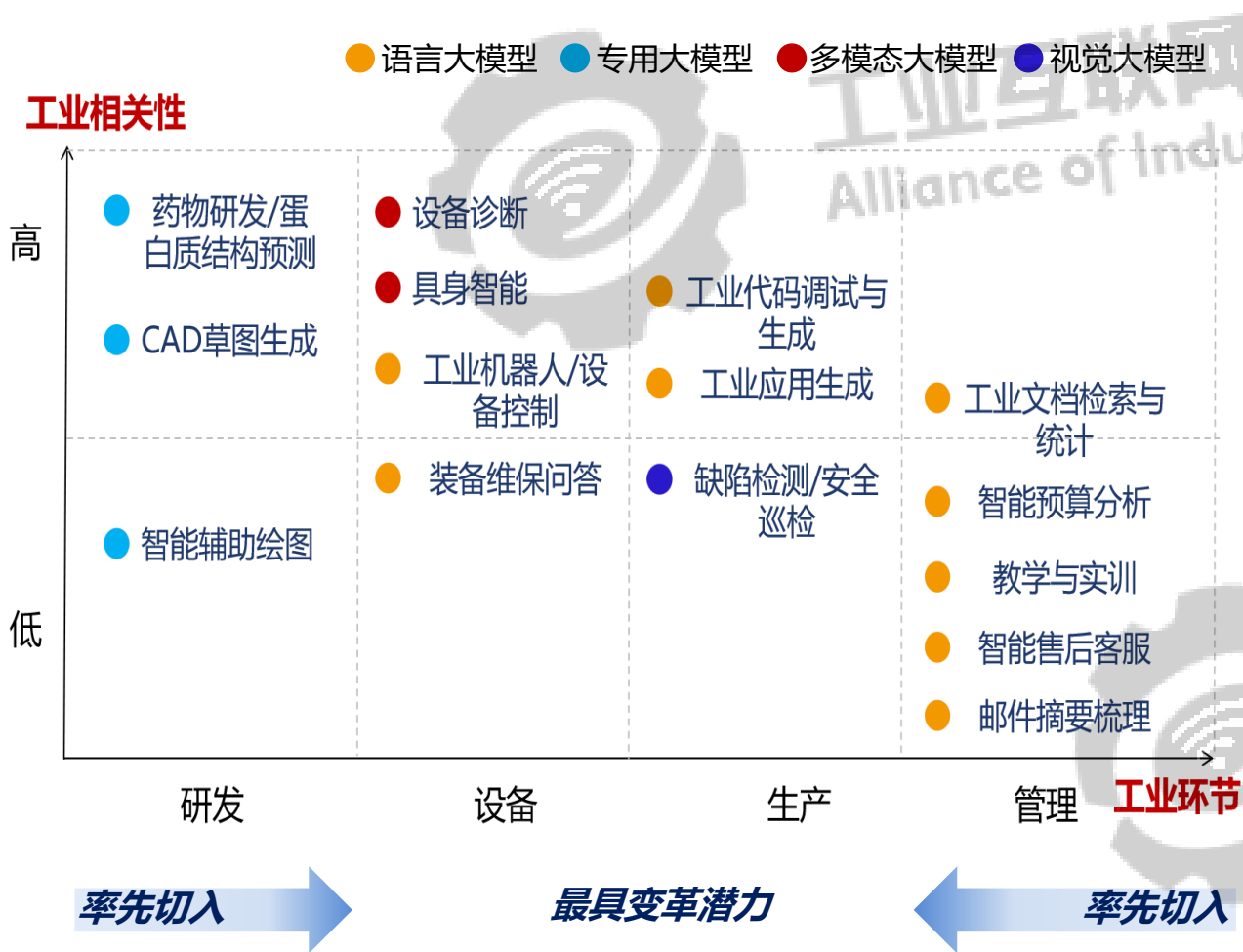
✓ 识别/模拟/预测

局部建模预测 → 基于全局信息高效高精度预测优化

✓ 多模态

单一格式工业数据处理 → 多格式数据综合转换分析

2.2 应用总体视图：4类核心模型、15+应用场景，目前处于初步探索阶段



*信通院统计的全球79个大模型工业应用案例

- **工业各环节围绕语言、专用、多模态和视觉四类大模型开展探索**
- **当前以大语言模型为主，4类模型应用占比：75%、15%、8%和2%**
- **通用模型的场景化适配调优是主要部署方式，问答交互为主要应用模式**

2.2 (1) 大语言模型：主要应用于工业问答交互、内容生成，以提升任务处理效率为主，暂未触及工业核心环节

有望形成具有认知智能的数字员工及超级自动化链路，实现从需求理解到规划、自动化执行及结果交付的全链条能力

工业相关性

高

低

研发

生产

设备

管理

工业代码生成：基于输入文本实现PLC代码生成与辅助编程

设备控制与维保助手：从基于对话实现设备指令识别与维保知识获取

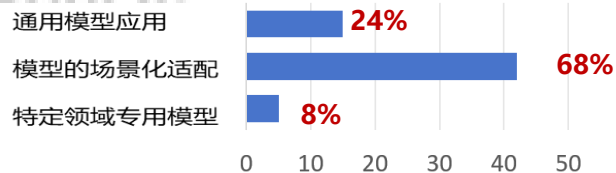
工业管理助手：实现BI数据查询与图表绘制

工业文档外挂与快速检索：实现设备诊断、知识查询、员工培训、设计合规检查

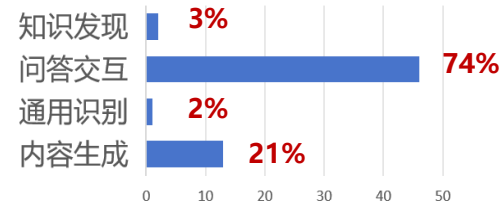
通用文档生成：基于输入文本实现设计方案、报告与邮件编写

● 内容生成 ● 问答交互

部署方式



应用模式



问答交互：应用相对广泛，行业与设备的(半)开放数据是主要语料

运维助手



罗克韦尔：将数字孪生与AIGC相结合，实现设备状态及设备故障原因咨询

管理助手



C3：通过自然语言对话，以文字+统计图完成业务指标的分析与洞察

检索助手



中工互联：基于智工大模型实现跨行业、跨领域和跨岗位工业专用知识检索

内容生成：已能实现通用内容和PLC基本控制逻辑代码的生成

工业代码



西门子：与微软合作，基于GPT开发代码生成工具，可通过NLP输入生成PLC代码

通用内容



BACANCY：基于RPA+GPT实现自动邮件回复等功能

2.2 (2) 专业任务大模型：围绕研发形成辅助设计、药物研发两个重点方向，进一步增强研发模式的创新能力

面向工业设计、蛋白质结构预测及药物研发创新等场景，扩展创新边界、降低创新成本与时间

工业相关性

高

需求导向的辅助设计

智能辅助设计：自动生成大量符合需求的设计，并与既有零部件及加工能力结合优化

药物/材料研发突破

高效研发：整合基因库/结构库，加速材料发现、药物分子结构发现、性能预测效率

低

研发

设备、生产、管理

● 格式转换 ● 知识发现

部署方式

模型的场景化适配

29%

特定领域专用模型

71%

应用模式

知识发现

77%

内容生成

13%

格式转换

10%

智能辅助设计：基于图像或文本进行2D-CAD草图构建



DeepMind：基于图像或文本进行2D-CAD草图构建，受样本数量+生成规范的限制，仅个别企业开展验证性探索

- ✓ 基于470万CAD草图数据训练
- ✓ 每个CAD草图对应构建草图生成规范

药物材料研发：聚焦蛋白质/药物的性质、结构与匹配能力的预测优化



Meta：E SMFold模型能够基于序列输入，实现蛋白质结构和序列的预测，模型参数已达150亿，仅2周完成包含罕见物质的6亿+蛋白结构预测

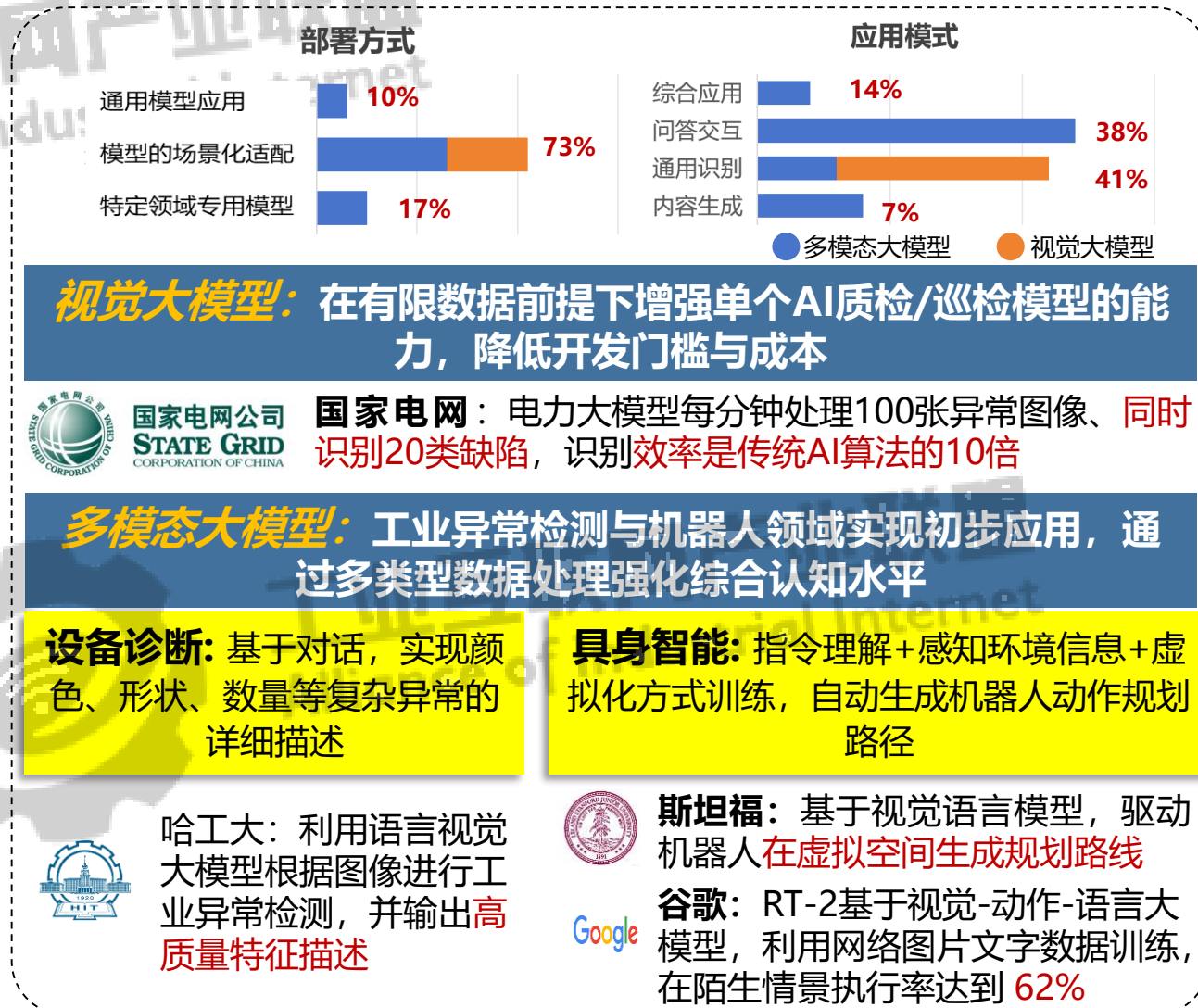
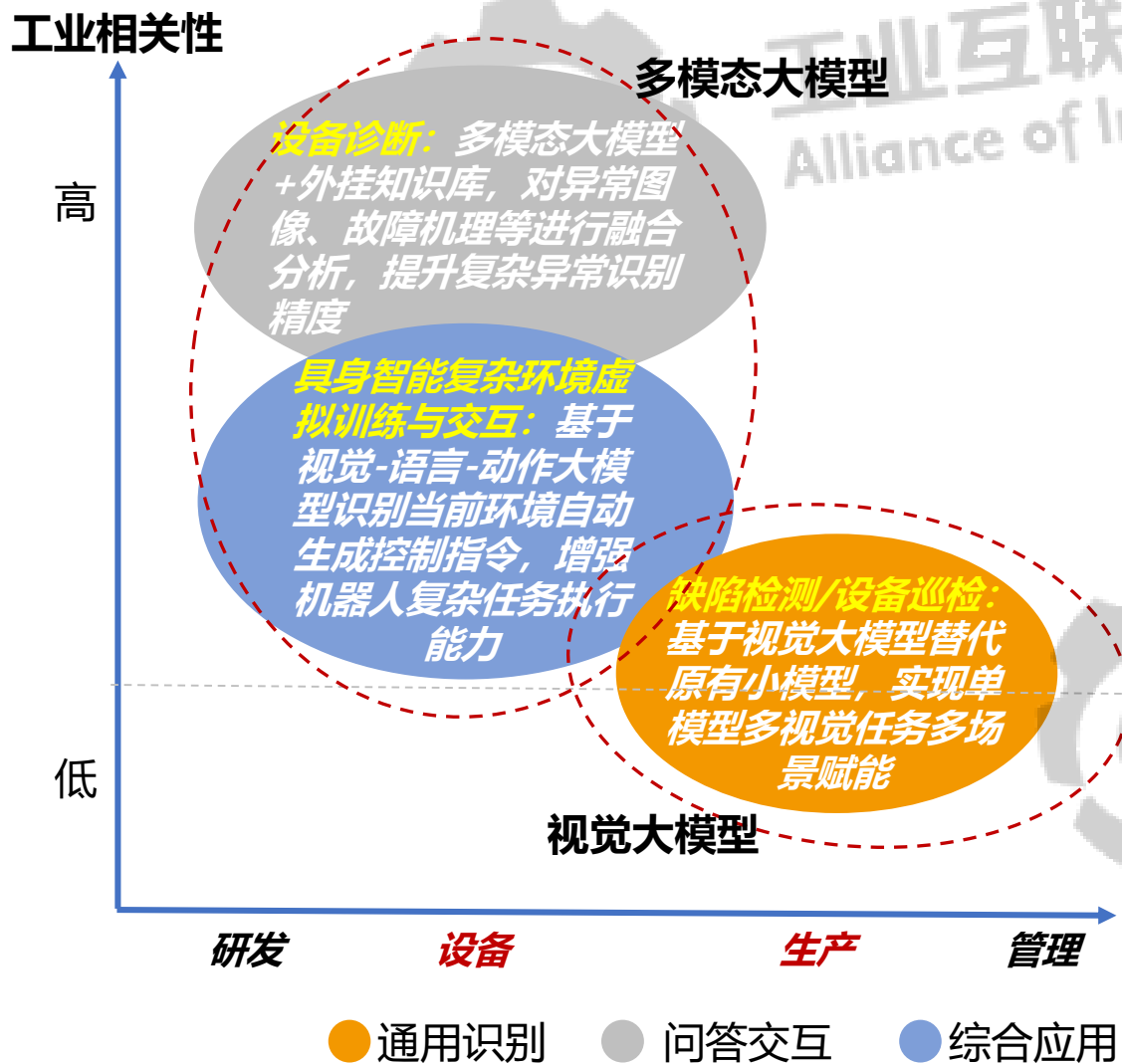


HUAWEI

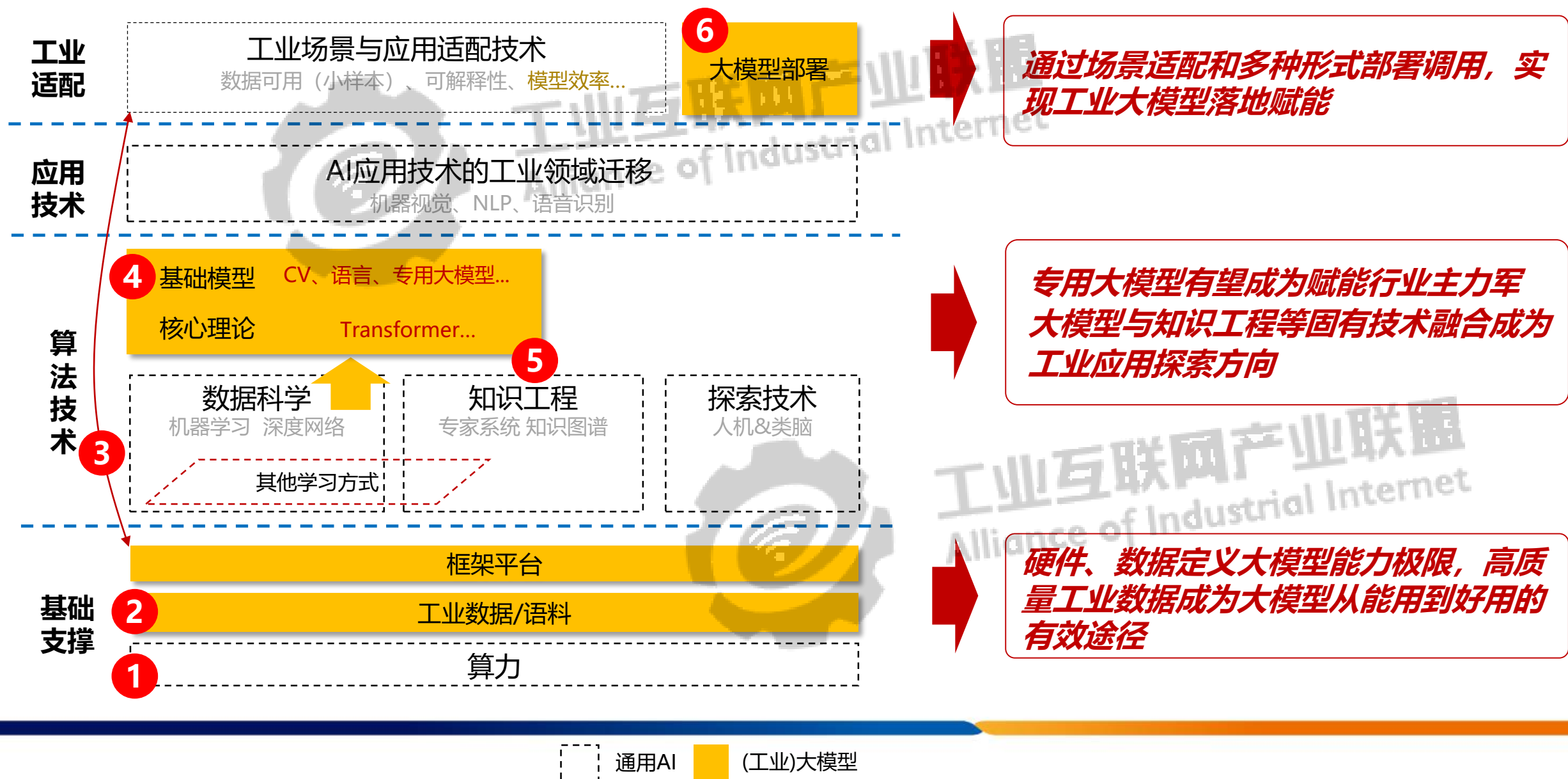
华为：盘古药物分子大模型，能够基于图结构药物分子输入，实现高效的药物分子生成和药物分子定向优化，生成1亿药物分子，新颖性达99.68%

2.2 (3) 多模态大模型与视觉大模型：在装备智能化和视觉识别领域应用获得初步尝试

结合视频、语义、执行等多类型数据综合分析，有望构建认知能力的装备、系统方案及智能工厂



3 技术体系：大模型是工业AI深度学习路径的深化与拓展



3.1 算力：工业领域大模型推理速度为需求关键，未来有望向端/边缘侧推理发展

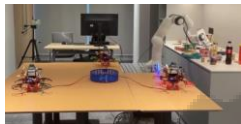
工业大模型推理速度需满足工业应用及峰值QPS等需求

1 大模型训练推理算力需求相对可控

大模型每10亿参数（1G模型文件）所需最低显存需求

模型精度	训练显存	微调显存 (LORA)	推理显存
float32(全)	14G	5G	4G
FP16	7G	2.4G	2G
int8	3.5G	1.2G	1G
int4	1.8G	0.6G	0.5G

- 西工大：基于大模型的多设备协同，采用云端统一控制，需求为单卡4090



2 工业领域(边端侧)对推理计算速度及满足峰值QPS等需求较大，

- 百亿参数大模型，使用1张英伟达A100GPU进行推理，每秒生成的token数大约为60
- 已有大模型一体机及端侧优化芯片，实现推理加速



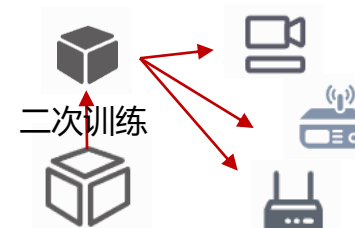
联合华为发布
星火一体机，
提供2.5P算力



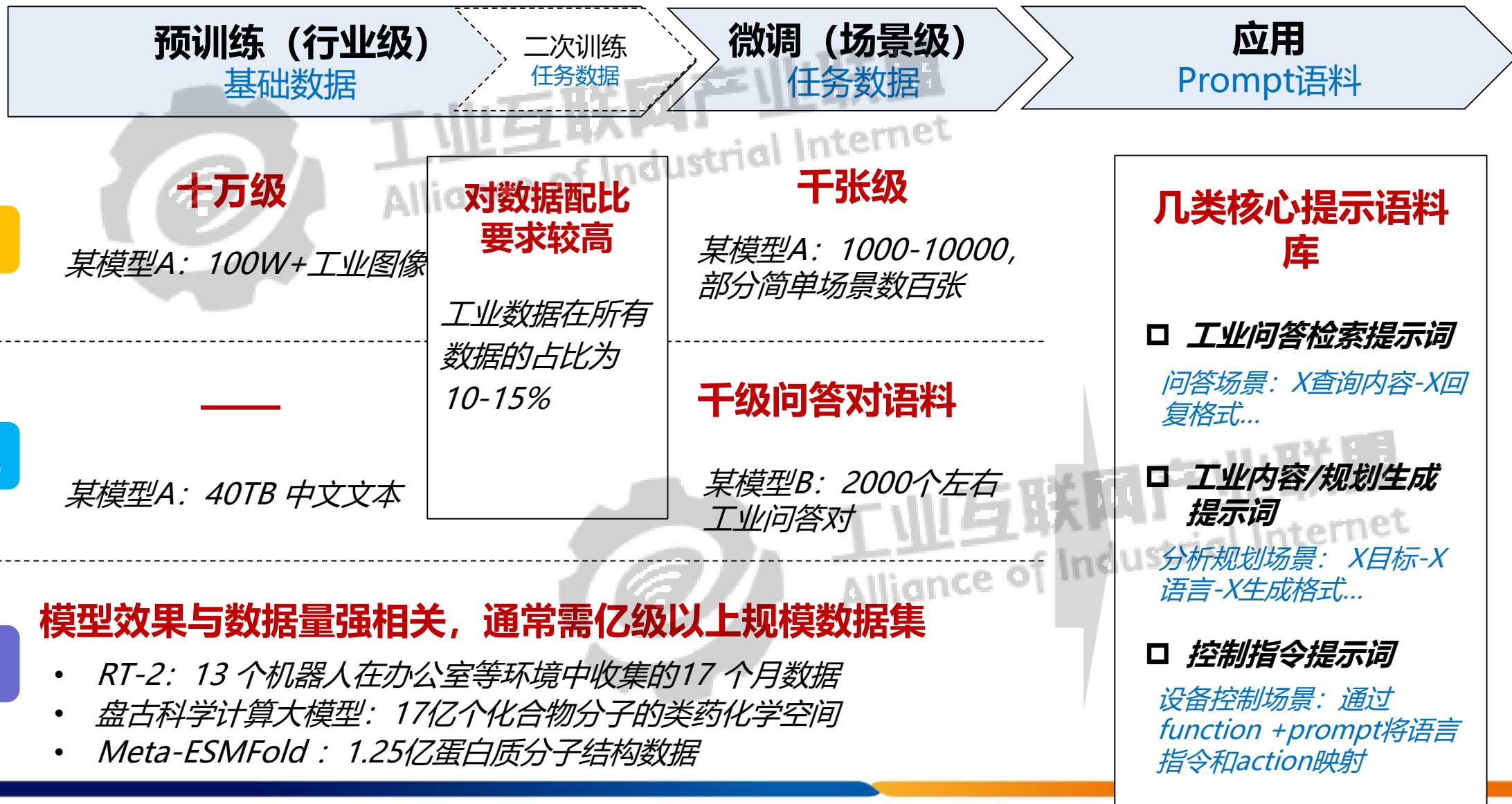
爱芯元智 -
AX650N芯片，
可达361 FPS

3 工业算力智能分配可能成为关键

- 云端大算力和终端小算力的平衡使算力分配和性能达到最优



3.2 数据：海量高质工业数据/语料库将成为落地部署的关键要素



3.3 工具链+模型：低门槛开发和轻量化部署成为工业大模型探索重点

各主体围绕通用大模型开发到部署全流程工具链，多推理后端兼容、半自动微调成为重点

➤ 与多推理后端兼容，实现工业低成本迁移



- 飞桨通过标准化部署接口，实现不同推理后端的零成本迁移



- 兼容20余家芯片厂商硬件设备，实现云边端全场景协同

➤ 大模型微调由手动向半自动化演进

手动调参

半自动化调参



可视化界面
参数设置 → 模型性能
观测



无默认参数
人工调参 → 模型性能
观测

企业	调参方式
	MakerSuite工具迭代prompt 自动合成数据扩充数据集
	基于信息抽取能力，简化训练 数据获取
	集成AIGC能力，基于自然语言 实现调参

知识蒸馏成为模型层面降低工业部署应用成本的探索途径



需兼顾减轻模型体量、维持性能衰减，较具数学挑战性

2个主流路径，工业领域以跟随应用为主

模型压缩

在相同的带标签数据集上指导子模型训练，形成高效小规模网络

模型增强

利用其它数据资源或优化策略(相互学习等)提高子模型性能

对开源大模型进行蒸馏+预训练+指令微调，形成工业大模型Alno-15B (150亿参数)

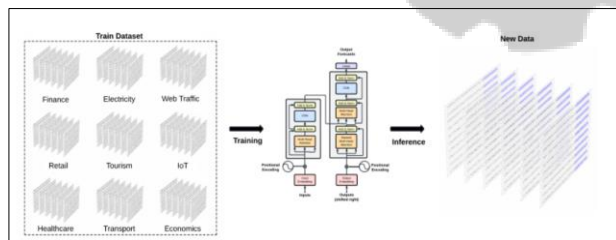
数据、算法要求较高，尚无工业实例

3.4 基础模型：通用大模型的快速演进和专用大模型的能力升级

面向更多样数据类型、更强综合能力的大模型技术迭代创新，为工业领域应用创造更大前景

1、时序数据大模型有望最大化利用海量工业设备与过程数据，赋能流程优化、设备诊断和异常识别

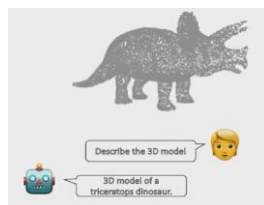
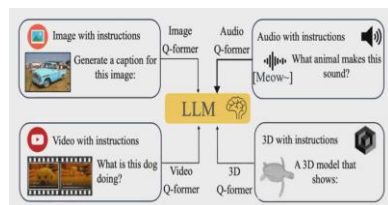
TimeGPT



训练：超1000亿个时序数据点

测试：超30万个时序数据集，开展时、日、周、月的预测评估

2、多模态能力持续创新，加速实现工业图像几何、机理、文档等各类数据模型的综合感知和认知推理



GPT-4.5融合了处理3D模型和视频的能力

专业任务大模型的数字支撑能力提升是主要方向，在药物创新发现与产品设计形成初步成果

1、依托大规模结构化专业数据嵌入表示能力，训练数据样本的全面性是提升模型性能核心

2、由一维序列、二维拓扑图转向三维表征的先进表征技术是探索方向

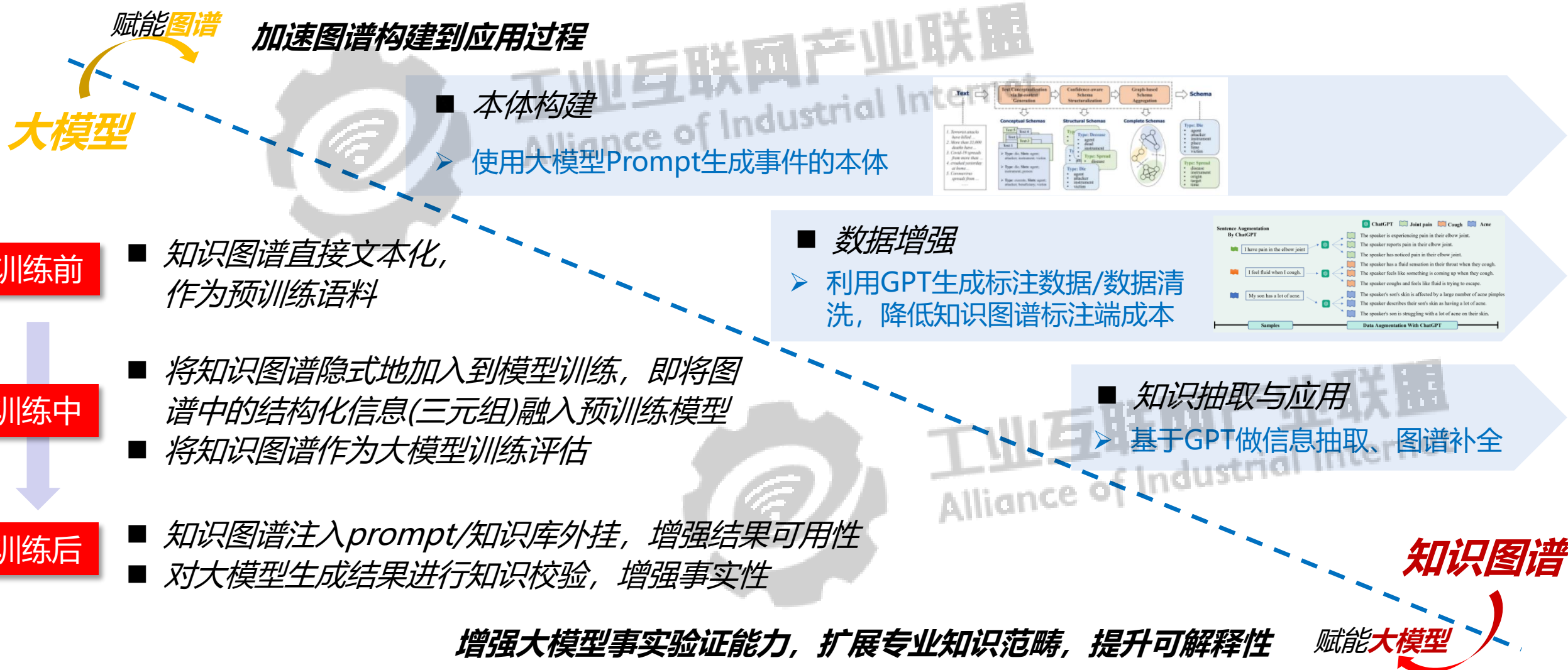
机构	细分方向	大模型	数据集大小
Meta	蛋白质结构预测等	ESMFold	UR50/D 1.25亿 (1维)
华为	药物分子生成优化	Pangu	药物分子 17亿 (2维)
深势科技	分子性质预测等	Uni-Mol	多数据集混合 2.09亿 (3维)

基于序列数据开展优化，为现阶段主流

基于分子特性开展探索，在领域93%数据集中表现最优

➤ 深势科技发布Uni-Mol，直接将分子三维结构坐标信息作为模型输入输出

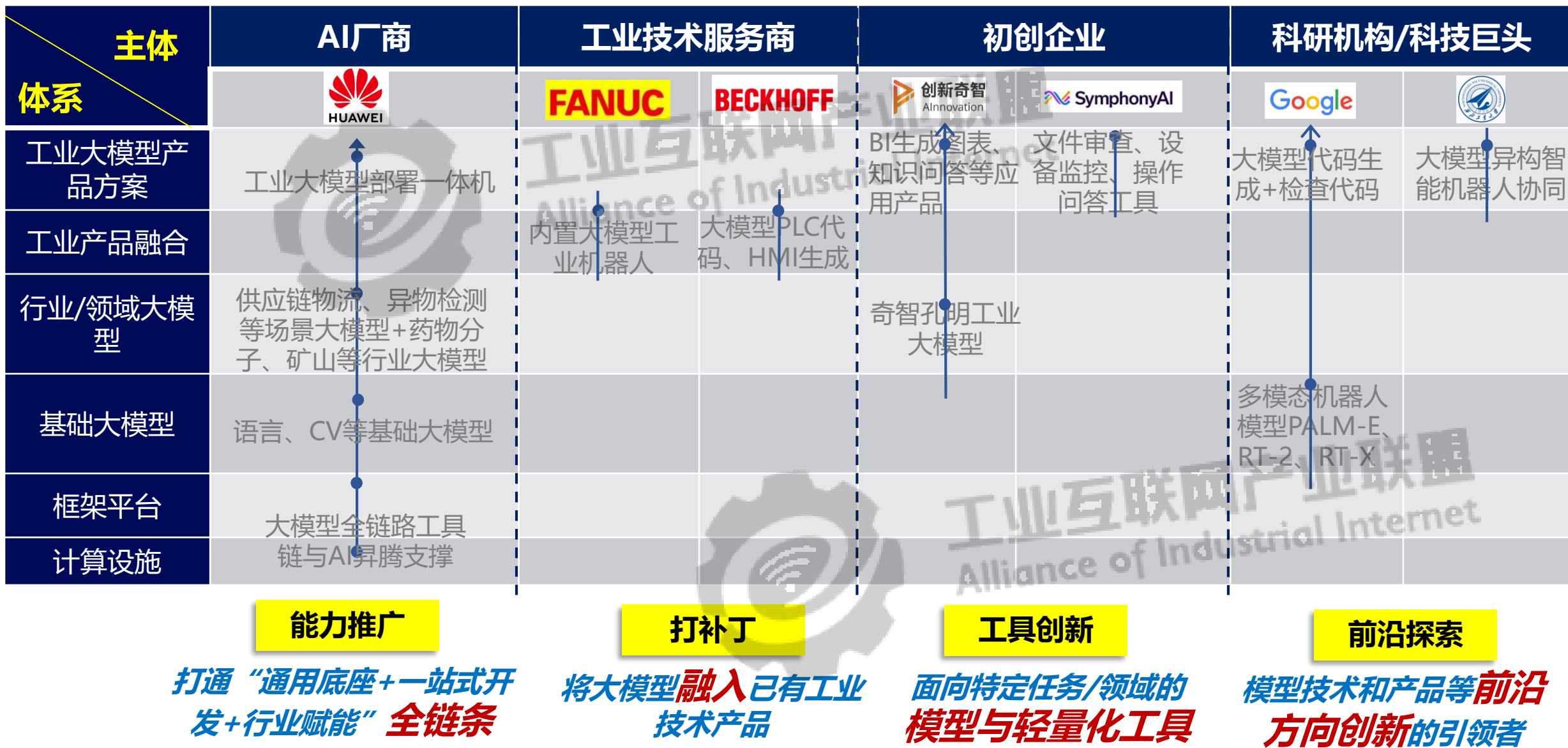
3.5 大模型+工业知识图谱：大模型可能对通用知识图谱产生一定冲击，但与领域知识图谱将融合共生，工业领域暂未出现典型探索



3.6 应用部署：三类核心部署方式，当前以通用模型场景化应用为主

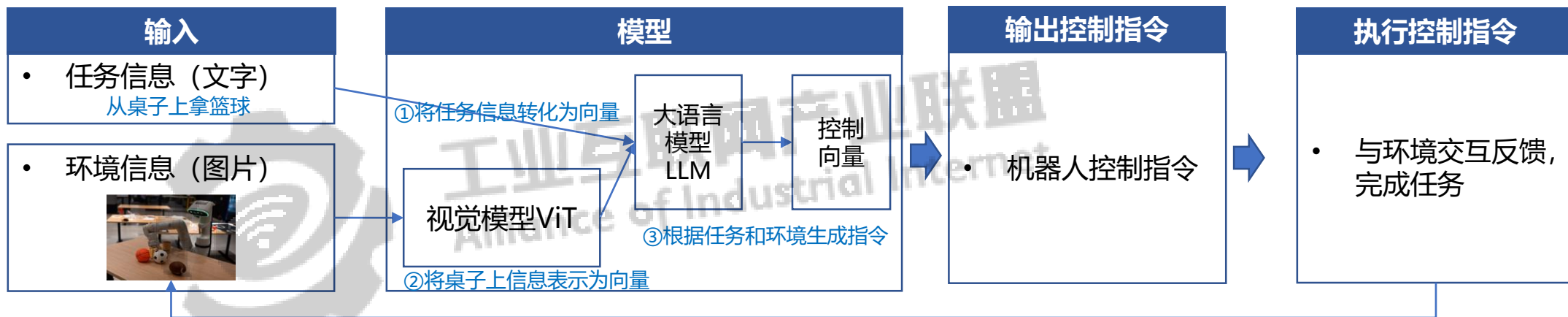


4 产业体系：不同主体布局及技术产品升级



4.1 大模型+装备：增强具身智能水平，有望提升工业设备灵活性和协同性

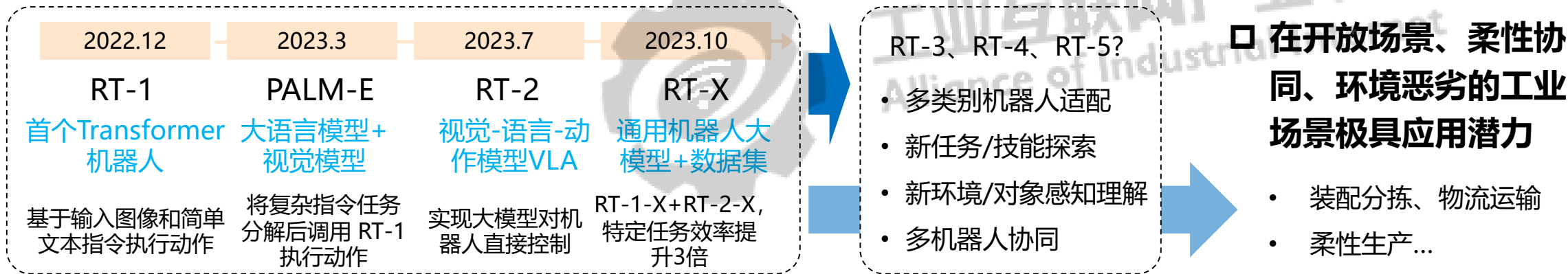
原理



能力

- 1. 综合分析能力** 综合视觉、语言、空间、理解、决策能力，从被动感知向主动认知跨越
- 2. 任务执行能力** 无需控制代码预设，实现模型对机器人直接控制
- 3. 强泛化能力** 对于陌生场景，无需针对性训练或模型微调，通过多步推理和知识迁移控制机器完成任务

实例



4.2 大模型+自动化：探索极为初步，距离实际应用还比较遥远

需求分析与系统设计

程序开发

参数调整

调试与集成

基于大模型实现简单控制代码生成，但当前语言匹配度和生成准确度仍有待提升

ABB □ ABB：通用大模型工业代码生成能力验证

能力现状

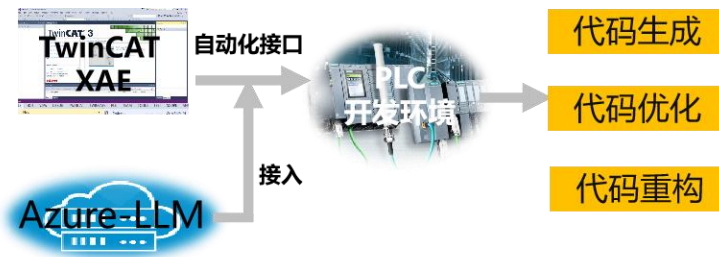
- 生成工业代码**逻辑正确率64%**，**执行成功率39%**
- 能够准确生成计数器、定时器等**标准算法**，交通控制等控制代码，前馈、压力控制等**流程代码**
- 基于**多轮对话**能够实现代码**优化与错误修正**

面临挑战

- 包含**控制逻辑、接口需求**的复杂Prompt设计
- **标准化**的工业控制**函数库构建**
- **大模型接收**文档**长度与模态限制**
- 生成代码**工业场景测试与应用**

□ 倍福、西门子：将大模型融入客户端，实现辅助编程

BECKHOFF 倍福：将大模型融入 TwinCAT XAE客户端，实现基于对话辅助编程



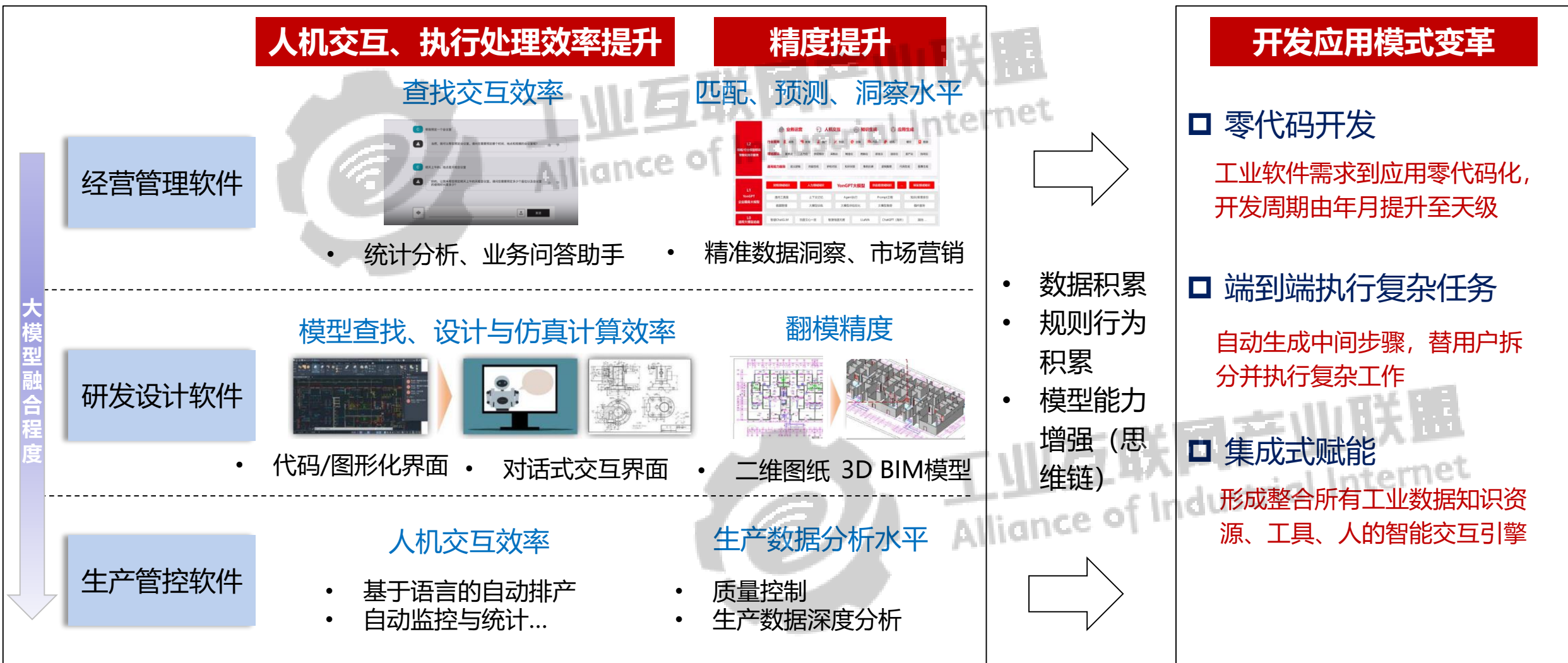
SIEMENS
西门子

西门子：联合微软开发工业Copilot工具并将其集成于自身工程框架，通过语言交互实现自动化代码的快速生成、优化和调试

基于大模型的控制参数整定已有实验性探索

□ 西门子：利用GPT-4开展非线性多因素PID控制算法整定，并完成某行业具体工况下的模拟验证

4.3 大模型+工业软件：从效率精度提升到应用开发模式重构



5 当前挑战

应用局限性

场景选择难

大模型如何应用于生产或开展模式创新还不清晰，且无法直接判断ROI

机器视觉 VS 大模型

机器换人，可解问题及ROI十分清晰 效率提升收益较难量化计算

低时效性

大模型的认知决策取决于历史训练数据，在解决动态工业问题的应用效果较差

 ChatGPT
抱歉，截止到我知识截止日期2021年9月，我无法提供2023年的事件信息。

ChatGPT训练数据集截止在2021年9月，导致无法回答训练数据以外的问题

低可信度

具备广博知识，但信息精确度低，制约工业核心环节/直接决策场景应用

模型幻觉 VS 工业场景容错率

大模型准确率80% 工业场景准确率需求99%+甚至100%

工程化局限性

工业语料匮乏

工业场景复杂，导致高质量工业语料难以收集，制约大模型性能

100张 VS 10000张+

针对具体场景，小模型仅需百张图像即可完成训练，大模型微调可能需万张

私有化成本高

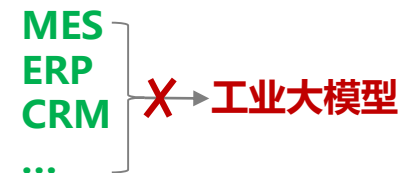
私有化部署大模型的算力成本+人工较高，多数企业难以承担

十万级 VS 百万级

AI专用小模型成本 国内AI企业私有化设备售价100-200W

系统集成难

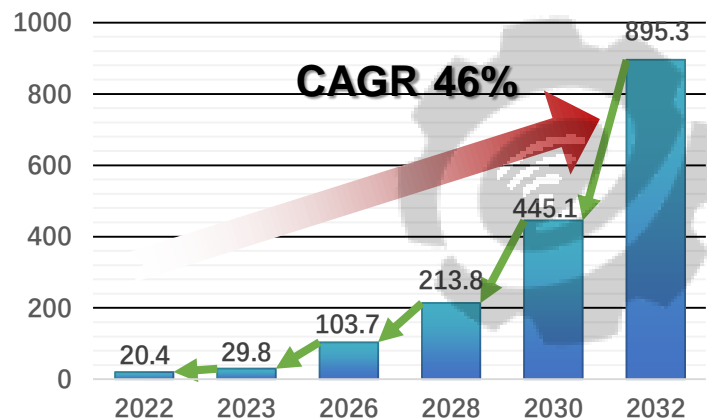
业务系统差异性导致工业大模型难以由统一口径集成系统数据



展望：AI与大模型加速赋能新型工业化

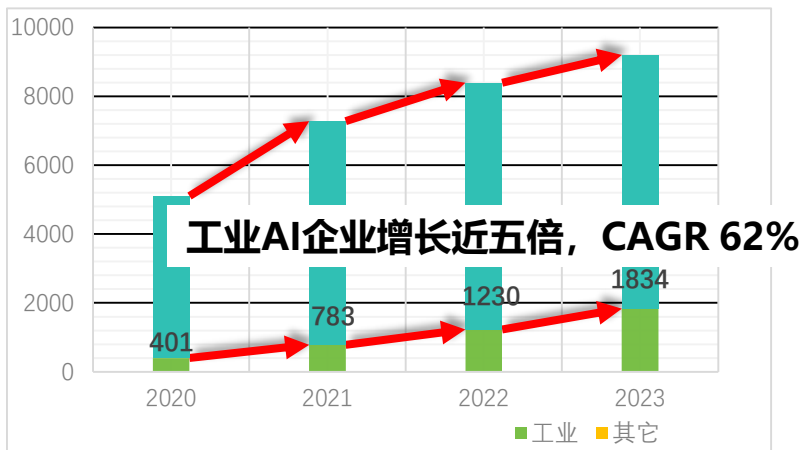
AI与工业融合展现强劲产业增长势头

2022-2032工业AI市场规模 (亿美元)



*数据来源: marketresearchfuture、MMR

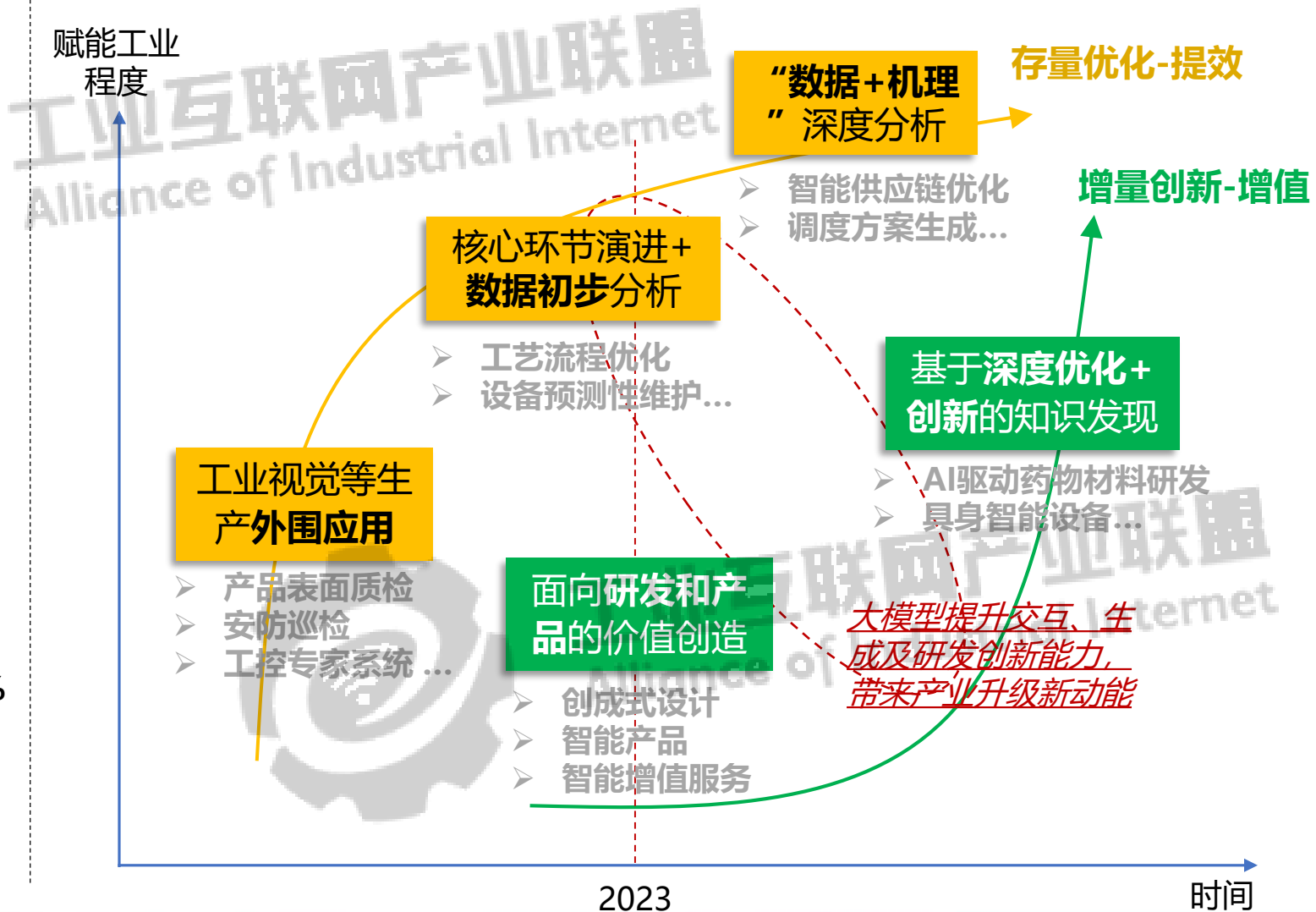
AI及工业AI初创企业数量 (个)



*数据来源: coresignal、VC、CB Insights

工业人工智能探索日益活跃, 存量优化+增量创新并行推进智能升级

赋能工业
程度



融合·协作·共赢

共同把握工业互联网的历史机遇



联盟公众号：工业互联网产业联盟

联盟网址：<http://www.aii-alliance.org/>

联盟邮箱：aii@caict.ac.cn

工业AI特设组、工业AI业务交流：18612054717