亚马逊云科技

中国峰会

2026年6月23日-24日 上海 · 世博中心

亚马逊云科技
中国峰会

5 个 AI 智能体运营一座工厂

唐俊杰

资深首席专家

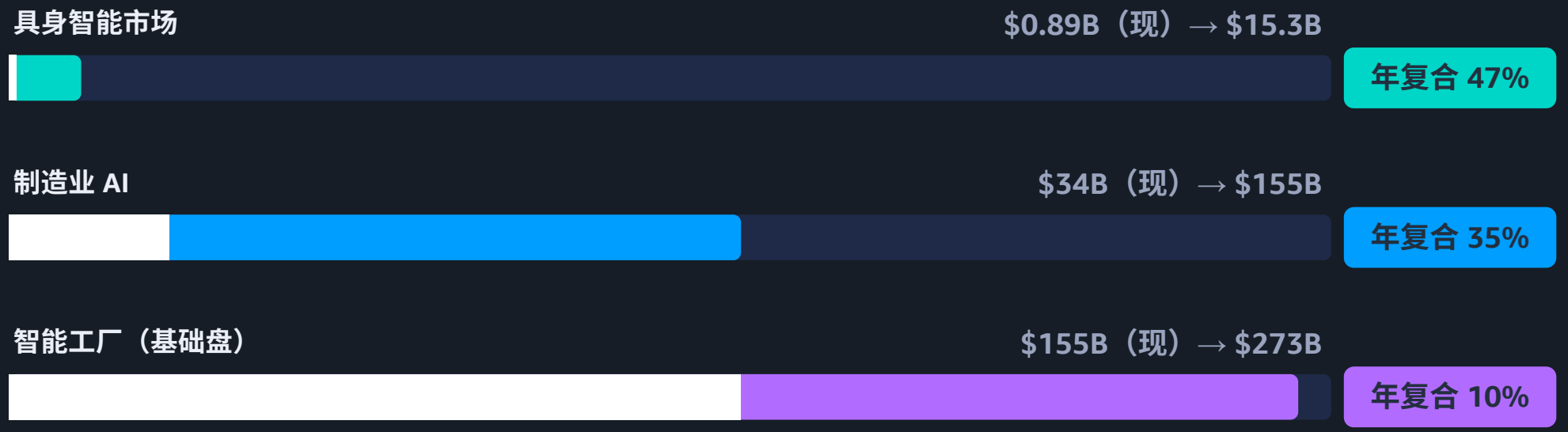
亚马逊云科技

议程

- 市场背景：具身智能为何在中国工厂率先落地
- AgentCorp 架构：五个智能体、三层治理、Cedar 安全闸门
- 现场演示：智能体协同 · 仿真到现实 · 世界模型预测
- 基准结果：质量、成本与可审计的治理证据
- 总结与展望：开源计划、参考蓝图与后续合作

具身智能正在快速从实验室走向工厂车间

三个市场信号，一个复利式拐点



4×

具身智能增速
相对智能工厂基础盘

7 年内，具身智能将从微不足道的零头，变为工厂转型中最大的支出项

*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

中国是工业具身智能的首要试验场

54%
占全球机器人安装量
中国，2024（IFR）

1/10
人形机器人单价
$1.35万 vs 西方 $9–15万

30 min
每下线一台人形机器人
1 万台/年，广东某单一工厂

Where China leads	China	Rest of world
Operational robot stock	~2.0M units — 4.5× Japan	Japan ~440K · US ~370K
Robot exports (Apr 2026)	25,375 units/month, +89% YoY	—
2026 humanoid shipments	Unitree 20,000 · sector 100K+	Rest of world <30K

中国不只是最大的市场 —— 更是具身智能最先从试点走向规模化的市场。成本、产能与部署速度，都领先 5–10 倍。

*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

软件的错误可以回滚，物理世界的错误不能

软件 AI 错误

- 错误提交 → git revert
- 幻觉事实 → 重试
- 写错文件 → 回滚
- 训练 OOM → 扩容

具身智能错误

- 错误运动规划 → 压坏托盘、损伤机械臂
- 幻觉障碍物 → 错误转向指令
- 坐标系错误 → 撞上传送带
- 错误模型 OTA 到车队 → 500 辆车变砖

大多数开源智能体框架都没有「仿真到现实」闸门。AgentCore + Cedar + Harness Engineering 是第一个把「已在数字孪生中验证」作为部署确定性前置条件的技术栈。

如果一个多智能体系统能够构建、治理并部署具身智能工作负载，且带有一道无法绕过的安全闸门，会怎样？

五个专精的 AI 智能体。

一道强制的「仿真到现实」闸门。

三层亚马逊云科技架构。从第一天起就具备生产级治理。

*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

不是一个模型，而是一支五人团队

每个智能体运行在硬件隔离的 Firecracker 微虚拟机中，由 Cedar 限定写权限

数字孪生架构师

仿真编写

Opus 4.6
Omniverse, Isaac Sim
simulation/, digital_twin/

感知智能体

CV + LiDAR 融合

Sonnet 4.6
SageMaker
perception/, models/

执行与物流

ROS2 控制 + AGV

Sonnet 4.6
AgentCore Runtime
control/, logistics/

平台与 IoT

OTA + 边缘推理

Sonnet 4.6
Greengrass, SageMaker Neo
edge/, deployment/

工厂运营评审

仿真到现实验证 + 闸门

Sonnet 4.6
AgentCore + Cedar
safety/, validation/

五个有边界的角色与真实工厂组织一一对应 —— 没有需要调试的「全能」单体智能体。

三层架构，三个亚马逊云科技服务面，一道强制的「仿真到现实」闸门

HARNESS ENGINEERING KIT · 推理 + 安全

仿真到现实闸门：数字孪生必须 PASS
硬件上下文：传感器规格、关节限位 · PretestValidator、审计轨迹、重试循环

AGENTCORE + CEDAR · 治理 + 隔离

每个智能体一个 Firecracker 微虚拟机 · Cedar 限定每个智能体的写权限
符合 ISO 26262 的只追加审计轨迹

5 个 CLAUDE CODE 工作智能体 · 执行

数字孪生 · 感知 · 控制 · 边缘 · 安全验证

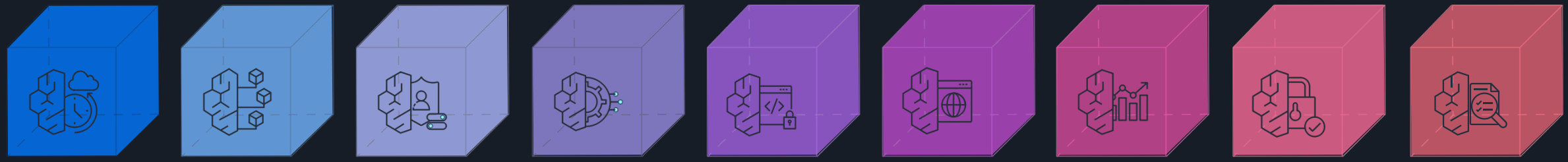
价值

换掉领域包，保留治理。同样的三层架构可运行车身车间、喷涂线或 AGV 车队。

*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

Amazon Bedrock AgentCore

全面的智能体平台：将智能体推向生产所需的一切



运行时 | 记忆 | 身份 | 网关 | 代码解释器 | 浏览器 | 可观测性 | 策略 | 评估

*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

策略是一个文件，而不是一页幻灯片

```
// 范围隔离：控制工程师不能写入边缘部署目录
forbid(principal == Agent::"control_engineer",
       action == Action::"write", resource)
  when { resource.path like "edge/*" };
```

```
// 无法绕过的规则：除非安全验证通过，否则任何东西都不能部署
forbid(principal, action == Action::"deploy", resource)
  unless { resource.safety_validation == "PASS" };

// 无法绕过。用策略引擎取代「信任开发者」。
```

价值

你的 ISO 26262 流程本就要求的审计，如今自动生成——无需事后追溯。

五个智能体能否在 10 个仿真日内，端到端受治理地，从零构建一座汽车工厂数字孪生？

贯穿全场的示例：一家正在规模化 EV 车身车间自动化的中国汽车 OEM——6 个子系统，一个未经验证的运动规划就可能导致停线。（示意性综合案例，非具名客户。）

19 个用户故事 · 6 个子系统

Omniverse · SageMaker · Isaac
Sim · Greengrass · AgentCore · EventBridge

强制的仿真到现实闸门

由 Cedar 强制执行 —— 无法绕过

真实成本、真实墙钟时间

真实 Bedrock 花费，端到端度量

ISO 26262 ASIL-D 审计轨迹

只追加，契合你既有的审计流程

*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

当系统真正发挥作用的时刻

Cedar 拦截一次跨范围写入

[第3天] control_engineer →
  写入 edge/greengrass_deploy.yaml
[Cedar] 拒绝 — control_engineer
  对 edge/ 无写权限
  规则: forbid(... like "edge/*")
[审计轨迹] 已追加:
  agent=control_engineer
  decision=DENY

Harness 拦截一个会碰撞的运动规划

[Harness] 仿真到现实闸门
 ✓ 关节限位　在 ±5° 内
 ⚠ 碰撞检测　夹爪与传送带
   在 t=3.2s 相交
 ✗ 已拦截 — 拒绝部署

[第4天] control → +15cm 余量 ✓
[第4天] safety → 重新验证 ✓
      仿真到现实闸门 通过

演示 1 — 五个智能体，一次受治理的冲刺

AgentCorp CLI：config → estimate → sprint → report → governance

```
> # 1) Validate the factory use case — 5 specialized agents, 19 stories
> agentcorp config $PA/domains/physical-ai --team
```

Config

pack: /Users/junjtang/Documents/projects/code.aws.dev/agentcorp-usecases/physical-ai/domains/physical-ai
status: **OK**

Company

Physical AI Factory Twin
Build AI-driven robotic training and factory management system with real AWS cloud pipelines and mock physical interfaces for simulation, training, deployment, and multi-agent orchestration.

Team — 5 agents

agent	role	type	model
chief_planner	Factory Digital Twin Architect	opus	claude-opus-4-6-v1
fullstack_sensing	Sensing & Digital Twin Fullstack Engineer	sonnet	claude-sonnet-4-6
fullstack_actuation	Actuation & Logistics Fullstack Engineer	sonnet	claude-sonnet-4-6
fullstack_platform	Platform & DevOps Fullstack Engineer	sonnet	claude-sonnet-4-6
judge	Factory Operations Validator	opus	claude-opus-4-6-v1

Sprint

stories: **19**
duration: 10 days
topology: hierarchical (planner → workers → judge)

```
>
```

演示 2 — 看懂即将播放的仿真

同一工厂单元的两个同步视图

- 物理产线 —— 真实的 6 自由度机械臂（橙色）
- Omniverse 数字孪生 —— 实时镜像关节（青色）
- 红色虚线 = 闸门拦截的未经验证规划
- 视觉 (CV) 工位标记焊接缺陷
- 偏差计 —— 孪生关节角度差（度）
- 传送带 → AGV 物料交接

留意那条红色幽灵轨迹 —— 那就是闸门必须拦截的未经验证规划。

*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

演示 2 ▶ 观看闸门如何阻止托盘被压坏

工厂数字孪生 · 仿真到现实 Cedar 闸门拦截未经验证的运动规划

t=3.2s 碰撞 → 拦截 → +15cm 余量 → 重新验证 → 通过 · 全程无人介入

演示 3 — 什么是世界模型，为何运行在 Amazon Trainium2 上

具身智能的预测层

它是什么

- 世界模型学习场景的动态：给定机器人当前所见 + 一条指令，它预测接下来几秒的视频与动作
- τ_0-WM：一个 55 亿参数的视频+动作扩散 Transformer

为何对具身智能重要

- 先想象后行动：在真实机械臂动作前，先在想象中预演下一个动作
- 它正是为「仿真到现实」闸门提供预测、供其检查的模型

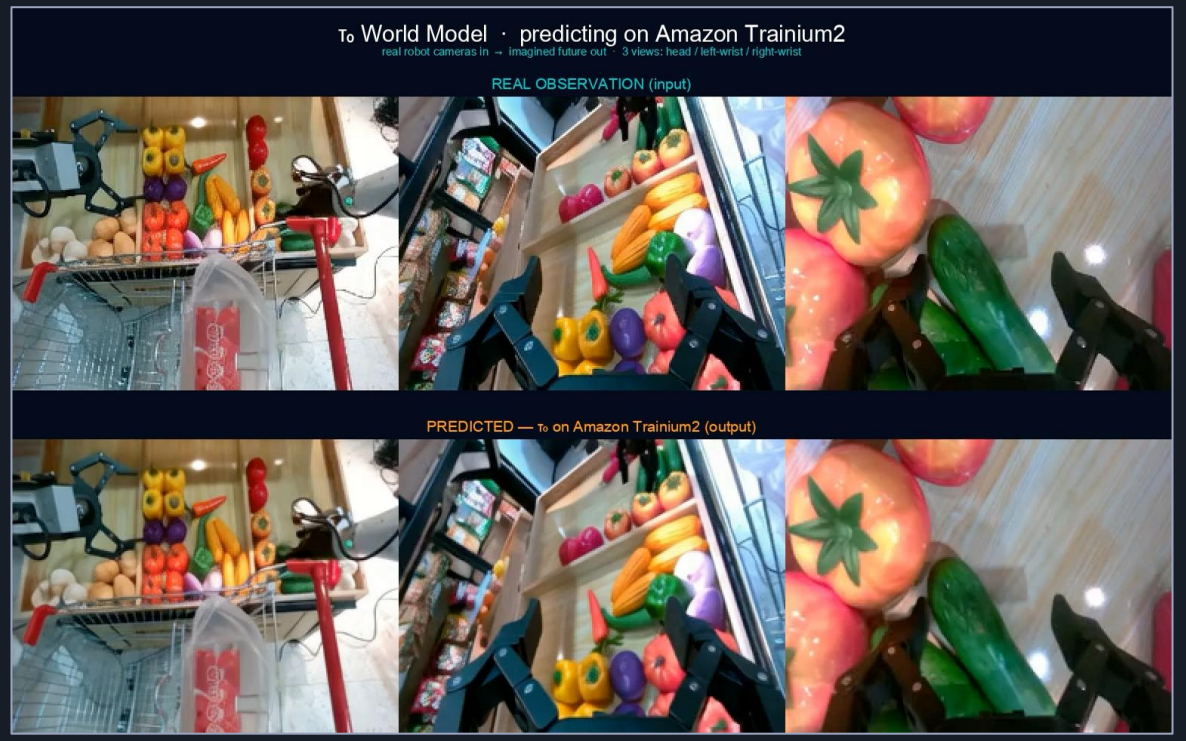
为何移植到 Amazon Trainium2

- 每个机器人一路世界模型，一颗芯片——96GB HBM 中仅占 13GB，留有批处理空间
- 在 Neuron 上 PyTorch 原生 —— 无 CUDA、无需重写内核。车队规模下每路成本更低

真实摄像头输入 → 预测未来输出 · 首个已知的 τ_0 Amazon Trainium2 移植

演示 3 ▶ 在行动之前，先想象未来

τ_0 55 亿参数世界模型 · 在 Amazon Trainium2 上真实推理 · PyTorch 原生



真实机器人摄像头输入 → 预测视频 + 动作输出 · 与运行智能体相同的亚马逊云科技芯片

*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

已基准测试，已治理，已度量

具身智能工厂冲刺（AgentCore + Claude Code SDK，经 Amazon Bedrock）

0.92
最佳综合得分

16/19
可部署故事数（84%）

100%
已完成故事测试通过率
运行时配置选择

7
Cedar 策略拦截
（默认拒绝）

6
安全闸门捕获
（未受控执行器）

$1.57
Bedrock · 16 分钟墙钟

AgentCore 原生
200 MB
最小车队占用，ASIL-D

AgentCore + Claude Code SDK（经 Amazon Bedrock）
500 MB
生产（默认）

本地 SDK
无
预生产迭代

*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

从基准测试到工厂车间

汽车 OEM 与一级供应商

仿真到现实闸门即你已要求的 ISO 26262 ASIL-D 纪律 —— 如今由 Cedar 策略自动执行。

人形机器人厂商

从 1,000 扩展到 100,000 台，只有在运动规划永不未经测试就到达真实机械臂时才成立。Cedar + Harness 让闸门无法被绕过。

工业自动化构建者

三层架构可移植：AgentCore 可在任意区域，Cedar 策略就是文件，审计轨迹只追加。

*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

这不是理论 ——
企业已经在运行这一模式。

西门子 Xcelerator：基础服务治理上构建的所有产品。

*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

制造软件行业面临的问题.....

系统孤岛

PLM、MES、ERP 与 OT 系统彼此割裂，阻碍贯穿产品生命周期的端到端无缝执行。

人工编排

过度依赖跨工具、软件与团队的人工协调，导致错误与延迟。

非计划停机

软件故障不可预测或发现过晚，维护与运营中协调开销巨大。

变更执行延迟

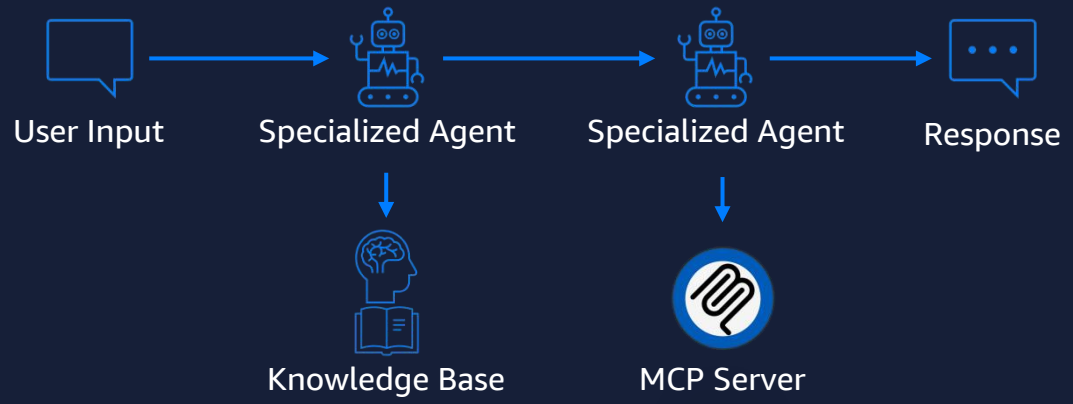
工程与流程变更从设计流转到车间往往耗时过长。

智能体 AI 编排让软件产品得以应对上述挑战.....

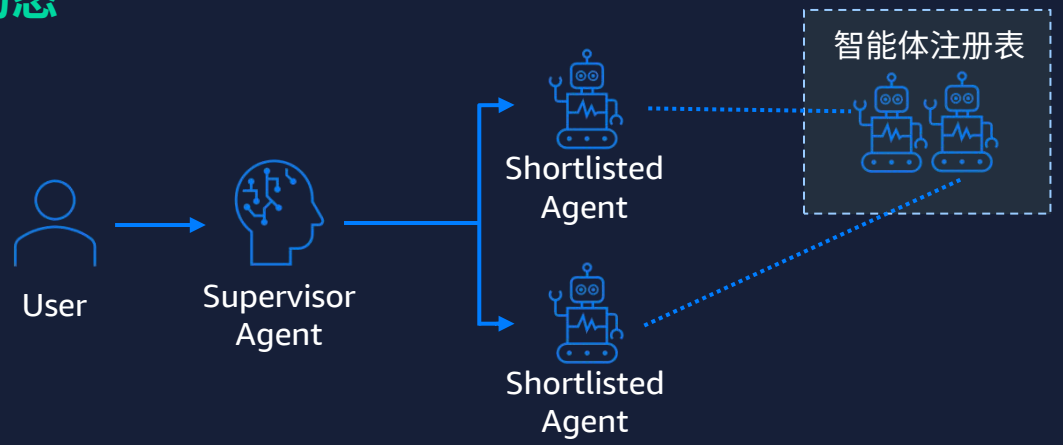
*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

Envoy 支持的智能体架构模式

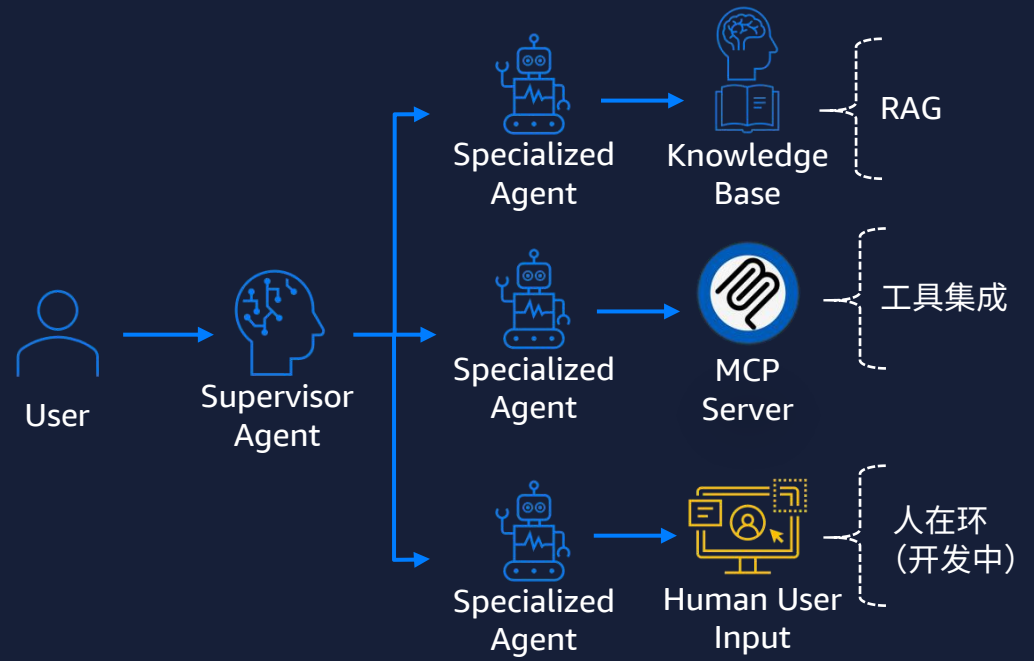
自定义



动态



层级式



*前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

在具身智能里，治理不是一句 PPT 上的口号，而是一个 Cedar 策略文件。它要么触发，要么不触发。这就是标准。

AgentCorp 开源 —— 即将推出 · 关注以获取发布信息 ·

合作请联系 Amazon ProServe

亚马逊云科技
中国峰会

Thank you