



InterGPT

一站式SOTA大模型中间件

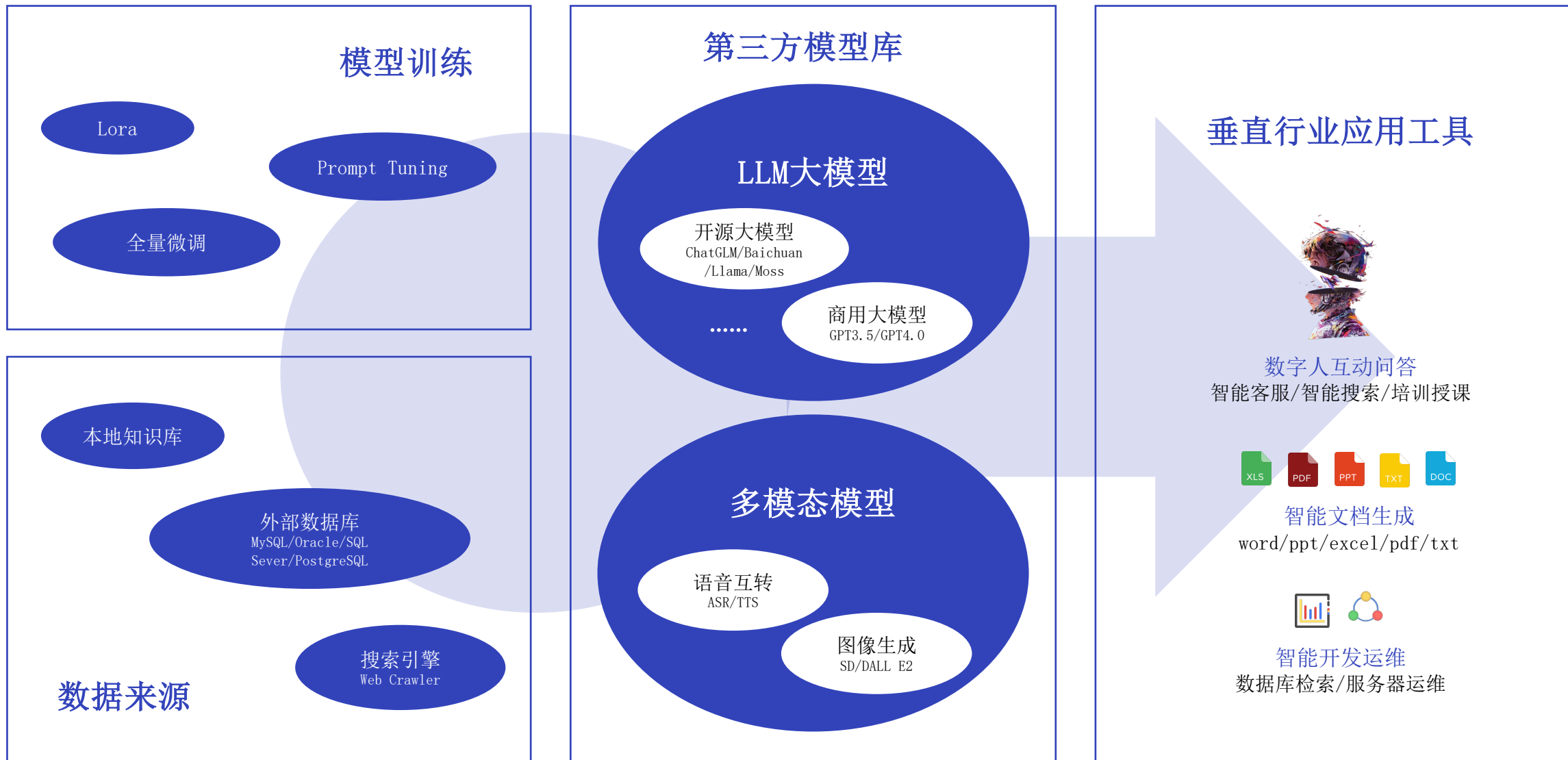
- | 丰富的模型库 | 对接本地知识库 | 一键微调垂直模型 |
- | 本地化部署 | 集成内部系统 | 支持英伟达、华为昇腾生态 |

@ 广州运通链达金服科技有限公司





产品路线图





丰富的大模型库



零门槛调用大模型

在统一平台快速获取大量优质大模型、微调工具及垂直行业应用组件，可视化操作调用模型，一键自动模型微调，少量数据完成场景落地，无需招聘专业AI人员即可实现大模型在企业本地服务器的部署及应用。



LLM模型

提供主流开源大模型及商用大模型。



多模态模型

提供语音文字互转、图片生成（文生图、图生图）



模型训练

提供全量微调、Prompt Tuning、Lora微调工具



健全的数据安全

三星电子禁止员工使用ChatGPT等生成式AI 拟推出内部工具 🔊 播报文章



同花顺财经

2023-05-02 11:55 | 财经领域创作者

关注

据界面新闻，出于安全考虑，三星电子已禁止员工使用ChatGPT、Google Bard和Bing等流行的生成式AI工具，正准备推出内部工具。公司内部备忘录显示，三星电子担心传输到生成式AI平台的数据被存储在外服务器上，导致其难以被追回和删除，并可能最终被泄露给其他用户。三星电子的新规禁止在公司所属的电脑、平板电脑、电话及内部网络使用生成式AI系统，但不影响出售给消费者的设备，由用户自行决定。此前，摩根大通、美国银行和花旗等华尔街大行已禁止或限制使用ChatGPT。

本地化、国产化部署

- 支持企业私有化部署，满足企业数据保密需求。
- 数据保留本地，规避企业数据泄密风险
- 同时支持灵活在线、离线多种大模型部署方案
- 无需依赖外部大模型环境，成本可控
- 保护企业知识库、Lora模型等资产
- 支持英伟达、华为昇腾生态



完善的应用生态组件

搜索引擎

打破开源大模型不能链接搜索引擎的局限。

Office文件生成

融合打通office生态，如Word、excel、ppt、pdf、txt，助力提升业务效率

数据库查询

智能检索MySQL、Oracle、SQL Sever、PostreSQL等数据库

丰富生态工具

服务器运维

简单输入即可完成服务器运维工作。

知识库

对接本地文件及数据库，提升大模型对垂直行业知识的理解及处理能力

图文生成

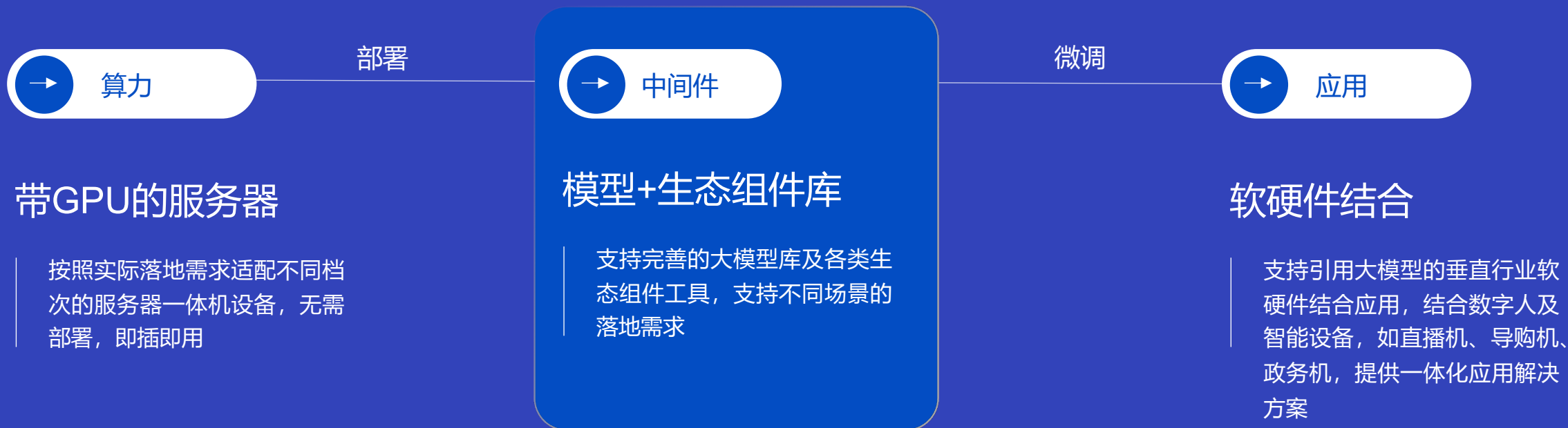
支持文生图、图生图、图片解说，简单的文字或产品图片输入，生成行业适用的图片。



端到端的落地能力

联合多家合作伙伴实现端到端的应用落地

提供集“算力设备→（部署）→大模型中间件→（微调）→软硬件应用”于一体的解决方案






全场景覆盖


2G

政务


7*24小时政务咨
询服务



+



教育



+

2B

企业

中大型企业内部
提效



+



金融



+

2C

电商

大模型+数智人
直播、问答



+



媒体



+



产品服务

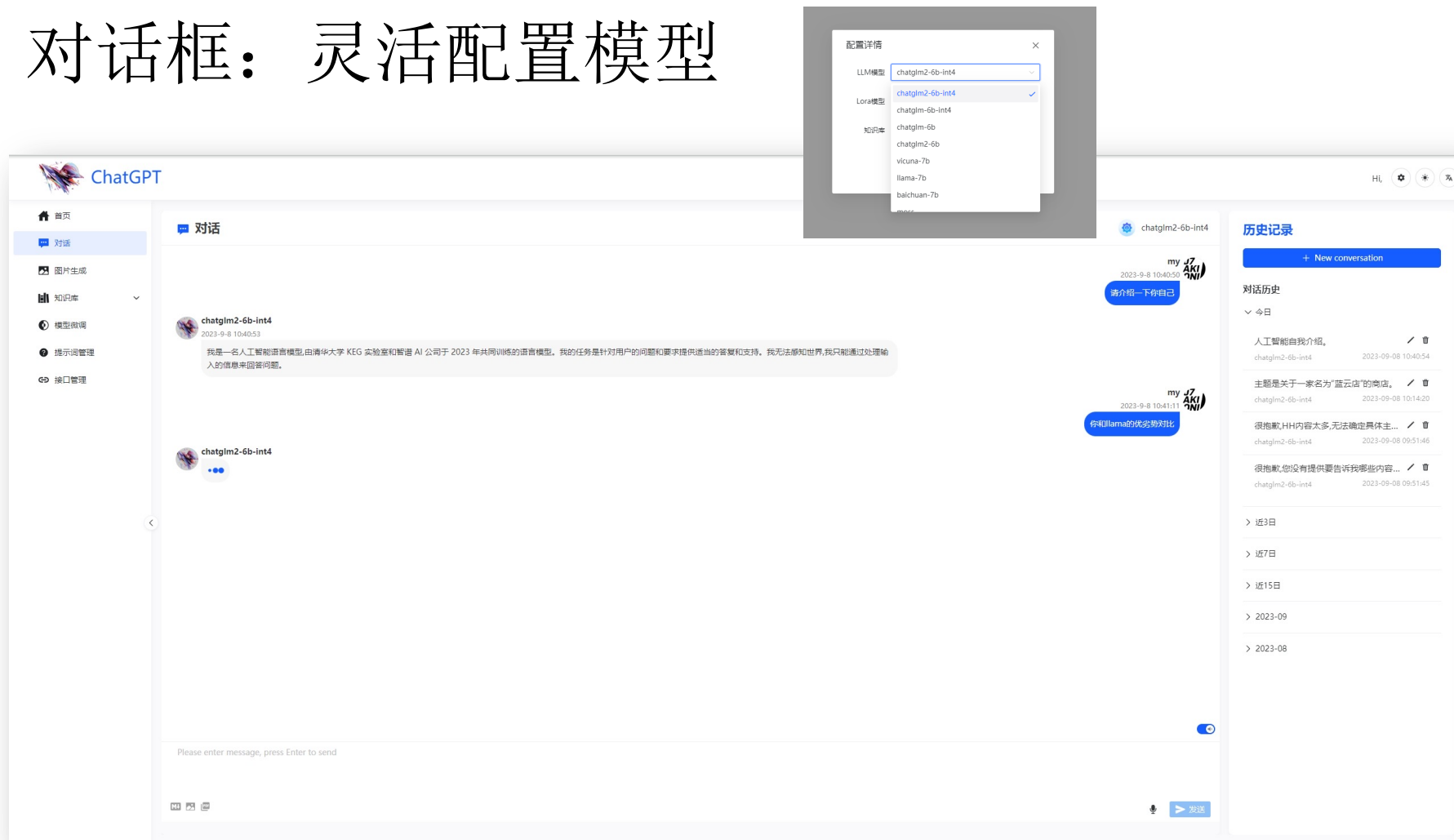
InterGPT中间件平台

InterGPT服务器一体机

InterGPT智能客服

InterGPT数智人一体机

对话框：灵活配置模型



知识库：快速梳理知识

产品服务

InterGPT中间件平台

InterGPT服务器一体机

InterGPT智能客服

InterGPT数智人一体机

The screenshot displays the 'ChatGPT' knowledge base management interface. On the left, a sidebar contains navigation options: 首页 (Home), 对话 (Conversation), 图片生成 (Image Generation), 知识库 (Knowledge Base), 知识库管理 (Knowledge Base Management), 知识库查找 (Knowledge Base Search), 模型微调 (Model Fine-tuning), 提示词管理 (Prompt Management), and 接口管理 (API Management). The main area shows a table of knowledge base entries with columns for '知识库名称' (Knowledge Base Name), '说明' (Description), and '创建时间' (Creation Time). A search bar and a '+ 新增知识库' (Add New Knowledge Base) button are at the top of the table.

| 知识库名称 | 说明 | 创建时间 |
|----------|-----------|--------------------|
| 鸡翅 | 1111 | 2023-09-07 15:04:4 |
| 大文件测试123 | 666666666 | 2023-09-07 16:04:1 |
| 123 | 测试 | 2023-08-31 16:44:2 |
| 123 | 123 | 2023-08-31 16:38:1 |
| 123 | 物权法 | 2023-08-31 16:37:3 |
| 123 | 123 | 2023-08-26 11:15:5 |
| 147 | abc | 2023-08-30 14:20:3 |
| 测试 | 测试 | 2023-08-30 17:05:1 |
| 测试文件上传 | pdf | 2023-08-24 14:29:2 |
| 分为范围few | 显示 | 2023-08-22 12:18:1 |

On the right, the '知识库文件管理' (Knowledge Base File Management) section includes a '文本入库分片长度上限' (Text Ingestion Chunk Length Limit) set to 100, a '知识库文件' (Knowledge Base File) input field containing '2022蓝云点系统介绍详细版.pdf', and a '删除文件' (Delete File) button. Below this is a large upload area with a cloud icon and the text '点击或者拖动文件到该区域来上传' (Click or drag files to this area to upload). A '确认' (Confirm) button is located at the bottom right of the interface.

微调：一键微调模型

产品服务

InterGPT中间件平台

InterGPT服务器一体机

InterGPT智能客服

InterGPT数智人一体机

The screenshot displays the ChatGPT interface with a sidebar on the left containing navigation options: 首页, 对话, 图片生成, 知识库, 知识库管理, 知识库查找, 模型微调 (highlighted), 提示词管理, and 接口管理. The main content area is titled '模型版本管理' and features a search bar with 'lora名称' and '适配模型' filters, and buttons for '查询' and '重置'. Below the search bar is a table listing Lora models:

| lora名称 | 适配模型 |
|-------------|------------------|
| hhh | chatglm2-6b-int4 |
| 大师博士大夫士大夫 | chatglm2-6b-int4 |
| 123 | chatglm2-6b-int4 |
| 123 | chatglm2-6b-int4 |
| 测试 | 6b |
| hhhhffffggh | chatglm-6b |
| hhhhffff | chatglm-6b |
| hhhh | chatglm-6b |
| wfe | chatglm2-6b-int4 |
| fwe | chatglm-6b-int4 |

On the right, a '模型版本管理' panel shows '基本信息' for the selected Lora: 'lora名称: hhh', '微调模型: chatglm2-6b-int4', and '描述: 朝实打实的'. A '+ 新增模型版本管理' button is visible. The '版本历史' section shows a table with columns for '版本名称', '开始日期', and '结束日期', and buttons for '查询' and '重置'. A '测试' button is also present, with a '训练结果: 训练完成' status. Below the test button, training details are shown: '训练描述: 6666', '训练参数: 训练点次数=1000,检查点次数=500', and '使用数据集: data_lora_test.json, 855411375.mp3'. At the bottom of the panel are buttons for '开始训练', '设置训练集', '设置训练参数', and '删除'.

产品服务

InterGPT中间件平台

InterGPT服务器一体机

InterGPT智能客服

InterGPT数智人一体机



S212G



S412G



S816G

该系列提供三个档次的大模型服务器一体化产品，满足了高端工作站对空间部署和高性能的不同要求。基于第三代英特尔®至强®可扩展处理器，支持2/4片GPU加速卡，内置InterGPT大模型系统，是性能、扩展性和密度的完美融合。

领先架构

- * 支持全新100Gb网络卡，提高数据传输效率；
- * 支持NVIDIA NVLINK互连技术，能够提高显存和性能扩展。

灵活配置

- * 支持3片3.5/2.5”硬盘，支持RAID数据保护；
- * 支持1+1 冗余CRPS电源或塔式电源，有效降低成本投入。

应用场景

*适用于深度学习训练、推理、视频分析、大数据、虚拟计算等超高算力应用需求

产品服务

InterGPT中间件平台

InterGPT服务器一体机

InterGPT智能客服

InterGPT数智人一体机

| 型号 | S212G | S412G | S816G |
|--------|--|--|--|
| 规格 | 2U机架式 | 4U机架式 | 8U机架式 |
| 处理器 | 2颗英特尔®/至强® | 2颗英特尔®/至强® | 2颗英特尔®/至强® |
| 内存 | 支持16条DDR4 ECC RDIMM/LRDIMM内存 支持频率 1866/2133/2400/2666/3200MHz 支持可达4.5TB内存 | 支持16条DDR4 ECC RDIMM/LRDIMM内存 支持频率 1866/2133/2400/2666/3200MHz 支持可达4.5TB内存 | 支持16条DDR4 ECC RDIMM/LRDIMM内存 支持频率 1866/2133/2400/2666/3200MHz 支持可达4.5TB内存 |
| GPU | 支持2张双宽GPU卡 | 支持4张双宽GPU卡 | 支持8张双宽GPU卡 |
| 存储 | 2个M.2 NVMe硬盘 3块3.5英寸SATA硬盘 | 2个M.2 NVMe硬盘 3块3.5英寸SATA硬盘 | 2个M.2 NVMe硬盘 3块3.5英寸SATA硬盘 |
| 网络 | 集成双千兆网络 | 集成双千兆网络 | 集成双千兆网络 |
| PCIe扩展 | 4个PCIe扩展插槽 | 6个PCIe扩展插槽 | 6个PCIe扩展插槽 |
| 电源 | 铂金级1600W 1+1 冗余 | 铂金级2200W 1+1 冗余 | 铂金级2200W 2+2 冗余 |
| 工作环境 | 工作温度：10℃~35℃，工作湿度：8%~90% | 工作温度：10℃~35℃，工作湿度：8%~90% | 工作温度：10℃~35℃，工作湿度：8%~90% |

产品服务

InterGPT中间件平台

InterGPT服务器一体机

InterGPT智能客服

InterGPT数智人一体机



知识库

数据集

大模型



提升问题匹配率和回答率

无需做大量数据处理，在相似问/同义词/缩写名称等方面，利用推理能力和多轮问答即可实现毫不逊色的匹配和回答效果

多语言自动切换

无需做大量数据翻译处理，即可实现多语言自动识别和翻译

延申问答

能够自然寒暄，回答各类延申问题，提高用户交互体验，让智能客服更拟人化

产品服务

InterGPT中间件平台

InterGPT服务器一体机

InterGPT智能客服

InterGPT数智人一体机

智能直播一体机是一款基于人工智能与计算机图形学技术，集内容创作、素材编辑、AI实时互动、画面导播、人声回复、智能问答于一体的“大模型+数字人”内容创作平台，可广泛应用于电商、政务、文旅、教育等垂直领域，提高内容生产效率，助力品牌打造和营销增长。

文案解说

支持自定义解说文案，支持音频、文本解说，数字人解说延时<1s

平台推流

可向抖音、快手、微信、拼多多等六家平台推流，同时支持真人/数智人直播。

互动问答

支持设置话术匹配问答、大模型自主问答，提升回答匹配率和解决率

定制形象

主播可以定制自己的数智人形象和音色，模拟真人直播及问答

智能直播一体机



智能导购一体机适用于数字柜台、数字门迎等场景，虚拟数智人化身导购、门迎、客服角色，植入IoT硬件、结合InterGPT+数智人，提升导购设备的交互体验，可展示商城、门店、商品或活动信息，降低运营成本，提升获客能力。

产品服务

InterGPT中间件平台

InterGPT服务器一体机

InterGPT智能客服

InterGPT数智人一体机

智能导购一体机





关于链达

上市国企背景

顶级专家领衔

毕马威全国金融科技双50企业

德勤广州明日之星

广州运通链达金服科技有限公司(简称“运通链达”)是广州无线电集团旗下长期专注于区块链技术+AIGC大模型应用的高新科技公司,矢志于利用数字化技术增信,助力商业生态公平与效率协同发展,成为数字经济时代领先的数字资产服务商。

多年来,运通链达由“WEB3.0 AIGC专家—邹均博士”领衔,全栈攻克数字化创新技术,坚持自主研发“Web3.0区块链+AIGC”基础技术平台和应用系统,包括区块链鉴云BAAS平台、大模型中间件InterGPT、链达智分身SaaS平台等,创新性地将区块链和AIGC技术应用于资产数字化和数字资产管理领域,为政企实现高质量发展、乡村振兴提供数字化转型解决方案。

欢迎合作咨询

@ 广州运通链达金服科技有限公司

@ 广州市黄埔区科学城科林路9-11号

@ 郑经理（13922446823）

