JISUMEN | 极算门科技

让AI基础设施
从可运行走向可运营

把算力、模型与业务系统，交付为可用、可验证、可持续演进的企业AI能力



企业私有AI基础设施与智能化交付

COMPANY PROFILE 2026

当AI进入真实业务，基础设施必须被重新定义

企业需要的，不再是一台设备或一个模型，
而是一套能够在自身数据边界内持续运行、持续验证、持续演进的能力底座。

可控

算力、模型和数据边界掌握在企业自己手中。

可用

能力进入真实工作界面，而不止停留在演示环境。

可演进

架构能够随着模型、负载和业务需求持续升级。

这也是极算门重新理解“交付”的起点。

极算门，面向真实业务的私有AI工程伙伴

以工程化能力连接AI水冷工作站、私有算力、模型运行与企业智能化应用。

JISUMEN
极算门

注册商标 | 第89402856号 | 第9类

东莞市极算门计算机科技有限公司

我们服务企业、教育与科研机构，
提供AI水冷工作站和私有AI基础设施，
并完成模型部署、应用整合与持续运维。

以真实工作负载为起点，把硬件、软件和业务系统组织成一套可长期运营的整体能力。

品牌表达

LOGO以连续波形表达算力、数据与模型的流动；JISUMEN与“极算门”共同构成品牌识别，传达以工程能力跨过AI落地门槛的主张。

从算力到业务价值，中间需要一条完整链路

极算门把分散的技术环节组织成一套有边界、有验证、有持续性的交付体系。



任何一环缺失，AI能力都很难稳定进入生产环境。

工程能力，决定AI能否长期创造价值

产品只是载体，真正形成差异的是跨层协同与持续交付。

基础设施工程	将算力、供电、散热、网络与存储作为一套系统设计	稳定
模型与运行时	围绕模型、精度、上下文与并发选择运行架构	效率
应用与系统整合	让Agent进入文档、知识与跨业务系统流程	协同
验证与持续运维	以可复现测试、验收文档和长期维护保障结果	演进

同一团队理解硬件、模型与应用边界，才能减少跨供应商之间的责任断层。

同一套能力，适配不同规模与不同阶段

架构不按产品简单分级，而是按工作负载、数据边界与演进路径设计。

01 部门级

专业团队与研发部门

高性能工作站与本地开发环境
强调快速启动与专业负载

02 企业级

跨团队共享与私有部署

多节点推理与统一模型服务
强调协作、权限与内部复用

03 数据中心级

高并发与关键业务

数据中心AI工程与性能验收
强调规模、稳定与长期演进

统一原则 以模型与业务需求倒推资源配置，让今天的投入为下一阶段保留空间。

两类产品形态，承载同一套私有AI能力

从桌边多GPU算力到团队级私有基础设施，核心交付标准保持一致。



AI水冷工作站

单卡 / 双卡 / 四卡

本地大模型
AI开发与AIGC
长时稳定负载

从桌边场景快速启动



私有AI基础设施

单机 / 多节点

模型部署
Agent与系统整合
验证和持续运维

从算力交付走向运营

多GPU架构

全水冷设计

私有模型部署

持续交付支持

把高密度算力，收进可长期运行的桌边系统

水冷不只是散热方式，而是持续AI负载下的系统工程。



实机内部结构 | 双GPU全水冷平台

极算门AI水冷工作站

面向本地大模型、AIGC与专业渲染，
以多GPU架构、CPU/GPU水冷和整机热管理，
支撑高密度算力的长时间稳定运行。

从硬件选型、定制水冷到软件环境与模型部署，形成一套可直接投入工作的交付形态。

工程逻辑

水冷不是单独的散热部件，而是围绕功耗、热密度、扩展和维护形成的整机设计；
配置由模型规模与工作负载决定。

让AI从回答问题，走进企业工作流

AI Agent不是独立产品线，而是私有部署与交付中的业务编排层。

它连接文档、业务系统、私有模型与算力环境，
让AI能够理解任务、调用工具，并在企业权限边界内完成闭环。

01 业务场景

02 Agent编排

03 私有模型

04 算力与运维

文档自动化

材料归集、识别、审核、生成与归档

跨系统协同

减少复制、导出与重复录入

数据闭环

在权限边界内查询、处理、回写与追踪

从一个高价值场景开始，沉淀可复用的平台能力

以真实问题验证价值，再把连接器、流程与运维能力逐步标准化。

01

场景验证

从报价、合同、报告、验收资料或会议纪要等高频文档切入。

先证明价值

02

系统连接

接入 OA / ERP / CRM，
连接文件服务器、数据库与 API。

再形成闭环

03

平台沉淀

把共性连接器、流程模板、模型路由、监控和权限能力产品化。

最终可复制

尤其适合希望低成本启动、逐步扩展，并重视数据边界的中小企业。

把AI项目的不确定性，转化为可管理的交付过程

每个阶段都有清晰输入、输出、验证标准和决策依据。

01

DISCOVER

业务与边界

目标、数据、模型与上下文
并发规模与预算边界

02

DESIGN

架构与标准

算力、系统、运行时、集成方式与验收口径

03

DELIVER

部署与验收

安装、联调、测试、文档与真实工作界面交付

04

EVOLVE

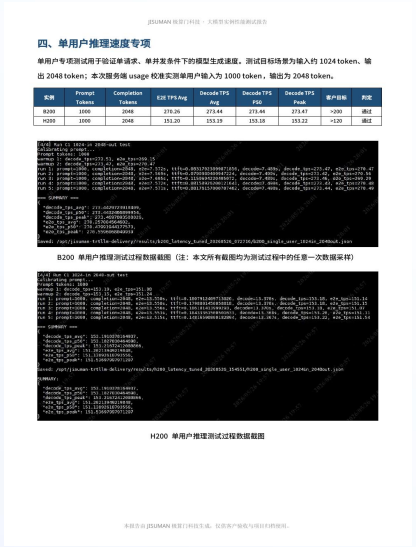
运营与演进

监控、优化、升级、扩容与持续业务接入

交付不是“把系统装好”，而是让客户在真实工作中稳定使用，并知道如何验证与持续维护。

代表性验证：规模不同，交付标准一致

具体设备和模型只是场景差异，真实环境、明确口径与可复现结果始终不变。



四、单用户推理速度专项

单用户专项测试用于验证单请求、单并发条件下的模型生成速度。测试目标场景为输入约 1024 token、输出 2048 token；本次服务端 usage 校准实测单用户输入为 1000 token，输出为 2048 token。

实例	Prompt Tokens	Completion Tokens	纯推 TPS·Avg	Decode TPS Avg	Decode TPS P50	Decode TPS Peak	客户目标	判定
B200	1000	2048	278.26	273.44	273.44	273.47	>200	通过
H200	1000	2048	151.20	153.19	153.18	153.22	>120	通过

B200 单用户推理测试过程数据截图（注：本文所有截图均为测试过程中的任意一次数据采样）

H200 单用户推理测试过程数据截图

数据中心
大模型验收

B200 / H200

固定Prompt
服务端usage计数
真实运行数据

完成性能验收与报告交付



多节点私有
算力验证

单机 / 多节点

模型部署
集群联调
并发与边界验证

形成可复现的工程结论

真实环境

明确口径

可复现结果

可交付文档

选择极算门，不只是选择一套设备

真正重要的，是一套能够陪伴企业从第一次部署走向长期运营的工程体系。

01

从真实负载倒推架构

先理解模型、数据与业务，再决定算力、节点、散热与系统。

02

用验证结果定义交付

把指标、参数、边界和原始结果写入报告，减少模糊承诺。

03

把持续运营纳入第一天

从方案阶段考虑监控、维护、升级、扩容与后续业务接入。

我们把复杂留在工程里，把清晰、稳定和可持续留给客户。

长期合作，建立在真实能力与持续投入之上

品牌、工程资产、测试环境与服务体系共同构成可信基础。

注册品牌

JISUMEN

第89402856号 | 第9类
有效期至2036年6月27日

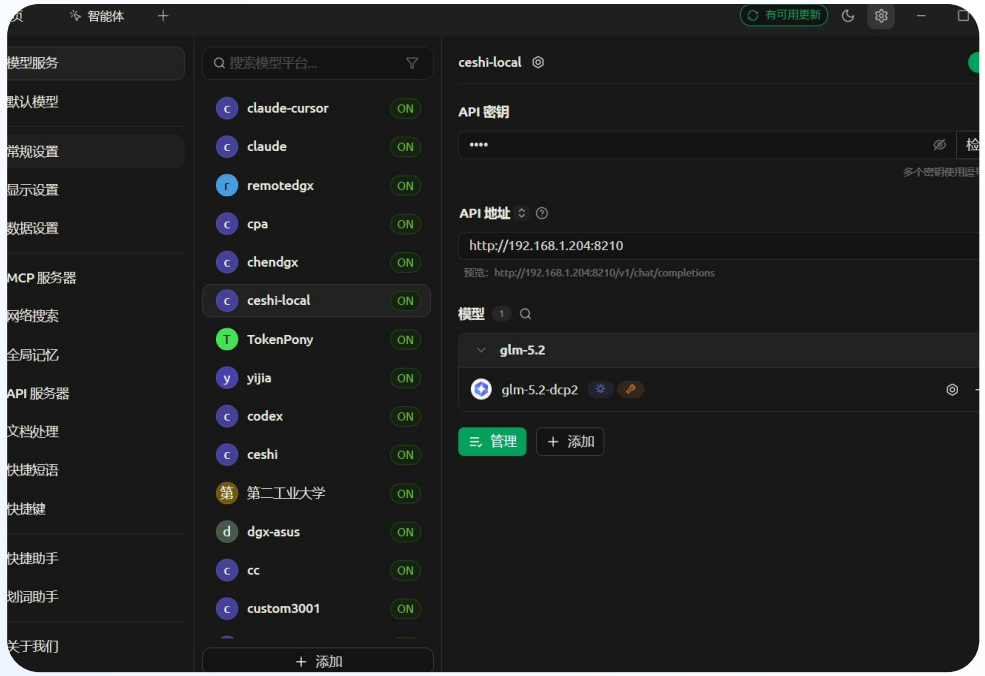
工程资产

真实测试环境与开源工具

持续积累模型部署、性能验证、交付文档与历史结果管理能力。

服务基础

东莞服务基地 | 支持上门服务 | 全国远程协作



真实运行环境 | 模型服务、工具接入与应用整合

持续投入，让项目能力走向组织能力

把已经验证的工程经验，沉淀成稳定的产品、流程与服务体系。

01 **产品化**

把液冷、工作站、私有算力与交付组件沉淀为清晰的方案边界。

02 **交付体系**

完善需求评估、架构设计、实施、验收和持续运维的标准流程。

03 **团队与服务**

加强客户支持、运维、物料管理与生产装配，提升稳定交付能力。

目标不是扩大产品清单，而是让每一次交付更稳定、更清晰、更可复制。

从一个明确的业务问题开始

我们会先理解五件事

01 业务目标

02 数据边界

03 现有系统

04 模型与规模

05 交付要求

以极算之力，启智慧之门

从私有算力、模型部署到AI Agent，让AI真正成为企业能力。