

STONE TEK G5208 AIO Qianfan一体机

大模型产业变革带来的挑战

企业端挑战



数据重要
无法上云



政策要求
信创适配



投入巨大
部署困难



应用构建
集成复杂

大模型创新需要一套更便捷，更经济，更有效的冷启动支撑方式

供应端挑战

缺乏AI开发人员

缺乏AI评估能力

缺乏高效算力支撑

G5208 Qianfan

G5208 Qianfan产品特性

百度千帆大模型Air平台

大模型管理与调用

智能体开发

AI 应用预置

AI 原子能力

更多应用...

G5208 Qianfan



G5208 PCIe5 Qianfan



百度千帆 × STONE TEK



私有化部署开箱即用



超强算力全栈优化



场景应用应有尽有

多行业应用预置 场景验证效率提升 **90%**

企业级大模型应用落地 成本能降低 **85%**

中小微企业开箱即用 应用周期缩短 **70%**

百度千帆大模型Air平台

平台能力

使用便利性

一键部署

可将预置模型、用户自导入模型在平台上发布为服务，生成式AI提供兼容OpneAI SDK接口。纳管第三方服务。可进行流控限流，扩缩容，灵活调度。支持用户自定义镜像发布为服务，可各取所需。

安全性

安全算子保障客户使用安全

- 01-回复干预
- 02-输入安全
- 03-输出安全

模型多样性

快速迭代

亮点1: 预置各种开源大模型、向量表征、重排序模型
亮点2: 新模型快速导入
亮点3: 多框架适配，支持PaddlePaddle, PyTorch等
亮点4: 模型类型、框架、芯片组合关联筛选。

企业级软件

最全的企业用户管理体系

以组织/项目维度，建立不同用户，分配不同角色权限，实现企业内部人员组织管理。
以资源池方式管理 CPU、内存以及 GPU 资源，可以把资源合理分配到不同项目组，实现资源有效利用。

应用开发

应用创建

- 应用名确定
- 应用图标AI生成
- 角色指令自动生成

模型选择

- 思考模型
- 问答模型
- 多样性/最大思考次数

能力拓展

- 添加组件
组件广场、创建组件
- 引用知识
知识库、数据库
- 添加记忆
记忆变量、记忆表、记忆片段
- 设定对话
开场白、推荐问、追问
- 形象和声音
数字人形象

调试与预览

- 效果预览
- 调试配置
- 用量用时参考

产品优势

模型多样性

使用便利性

安全性

企业级软件

丰富的生态应用

成熟案例

帮助企业快速迭代

支持一键部署

安全算子保障客户使用安全

支持报表统计等应用

后续支持多种不同生态应用集成

成熟产品，具备各行业落地经验

