

# LANChat 产品介绍文档

版本

产品概述

核心优势

功能介绍

界面展示

v0.4

v0.5

v0.6

其他资源

LANchat：面向边缘设备的本地大模型助手



LANCHAT

## 产品手册V0.5

2025/08/21

### 版本

版本号	时间	说明
-----	----	----

v0.4.4	2025/06/04	每种 agent 都是独立的界面
v0.5.6	2025/08/21	把 agents 进行合并，界面更加简洁
v0.6.X	2025/08/28~今	优化rag，在路上

## 产品概述

LANChat是一款专为高保密需求场景设计的本地化大语言模型部署平台，搭载 NVIDIA Jetson AGX Orin（64GB）边缘计算设备，支持6-8人同时低延迟交互。通过完全离线的模型部署，LANChat确保敏感数据无需上传至云端，彻底杜绝外部泄露风险，为金融交易、科研机密、企业战略等场景提供安全高效的AI对话能力。

## 核心优势

### 1. 极致数据安全

- **完全本地化运行：**所有数据处理与模型推理均在Jetson设备内部完成，无需网络传输，避免云端泄露或第三方拦截风险。
- **合规性保障：**满足金融、医疗、法律等行业对数据本地存储的强制监管要求（如GDPR、HIPAA）。

### 2. 低功耗高性能

- Jetson AGX Orin的能效优化架构，仅需30W功耗即可支持6-8人并发使用，降低长期运营成本。
- 内置Tensor Core GPU与12核ARM CPU，提供40 TOPS算力，响应速度低于500ms。

### 3. 开箱即用部署

- 预配置模型环境与容器化部署脚本，2小时内完成从硬件上电到多人可用状态。
- 支持离线API对接，无缝集成企业内部系统（如数据库、办公软件）。

#### 4. 灵活扩展性

- 支持模型替换（如Qwen、ChatGLM等开源模型切换）。
- workflow 扩展，满足更多的需求。
- 支持mcp扩展，实现更多功能。

## 功能介绍

### 智能助手

- 闲聊模式。
- 根据用户输入，结合上传文件、图片作分析，回答问题。包括翻译，总结，润色。
- 论文总结，对论文作总结，分析。
- 调用分析工具。

### 本地检索

- 半自动生成本地数据，加密/水印文档需要手动处理，以及部分数据。
- 通过用输入，精准定位相关的知识，然后作整理反馈给用户。

### 复杂任务规划

### 工具调用

## 典型应用场景

- 量化交易公司：本地解析市场数据、回测策略讨论，杜绝交易逻辑外泄。
- 法律事务所：客户案件分析、合同条款生成，保障隐私案卷安全。
- 研发型企业：技术方案研讨、专利文档撰写，避免核心创意暴露于公网。

## 技术架构亮点

- 硬件平台：NVIDIA Jetson AGX Orin（64GB内存，8核ARM v8.2 CPU，2048核Ampere GPU）
- 软件栈：TensorRT模型加速、Docker容器化部署、RESTful本地API接口
- 安全特性：端到端加密通信、访问权限分级、操作日志审计追踪
- RAG：采用图+模块化结构，大大提高了检索的速度和精度。

## 为什么选择LANChat?

> 在数据即资产的时代，LANChat将大模型能力“锁进保险箱”——既享受AI智能化带来的效率革命，又牢牢握住数据安全的钥匙。

**适合人群：**CTO、信息安全负责人、AI基础设施团队

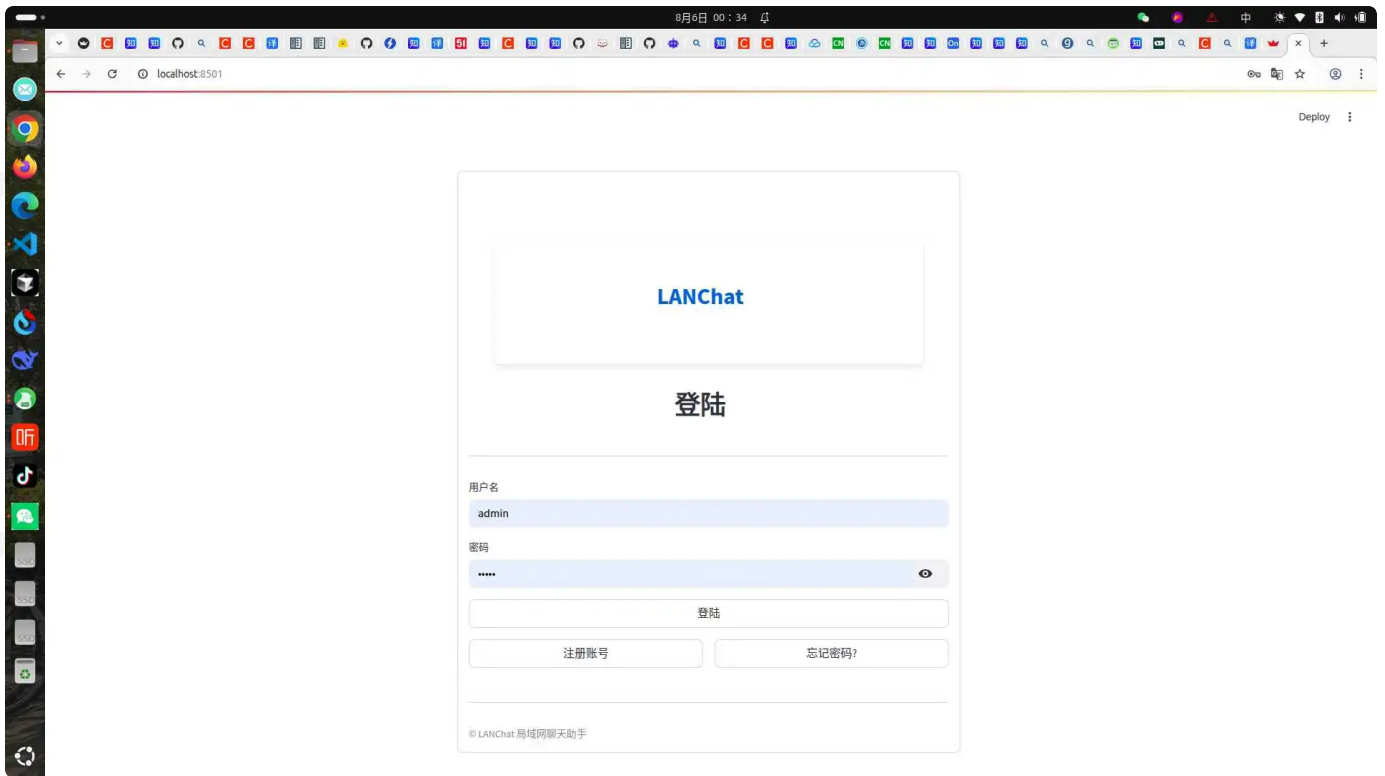
**部署要求：**单设备支持6-8人，仅需电源与局域网连接

**联系我们：**获取定制化部署方案与压力测试报告，让AI在您的防火墙内安全赋能！ [redstromth@live.com](mailto:redstromth@live.com)

**\*LANChat：安全与智能，无需妥协。\***

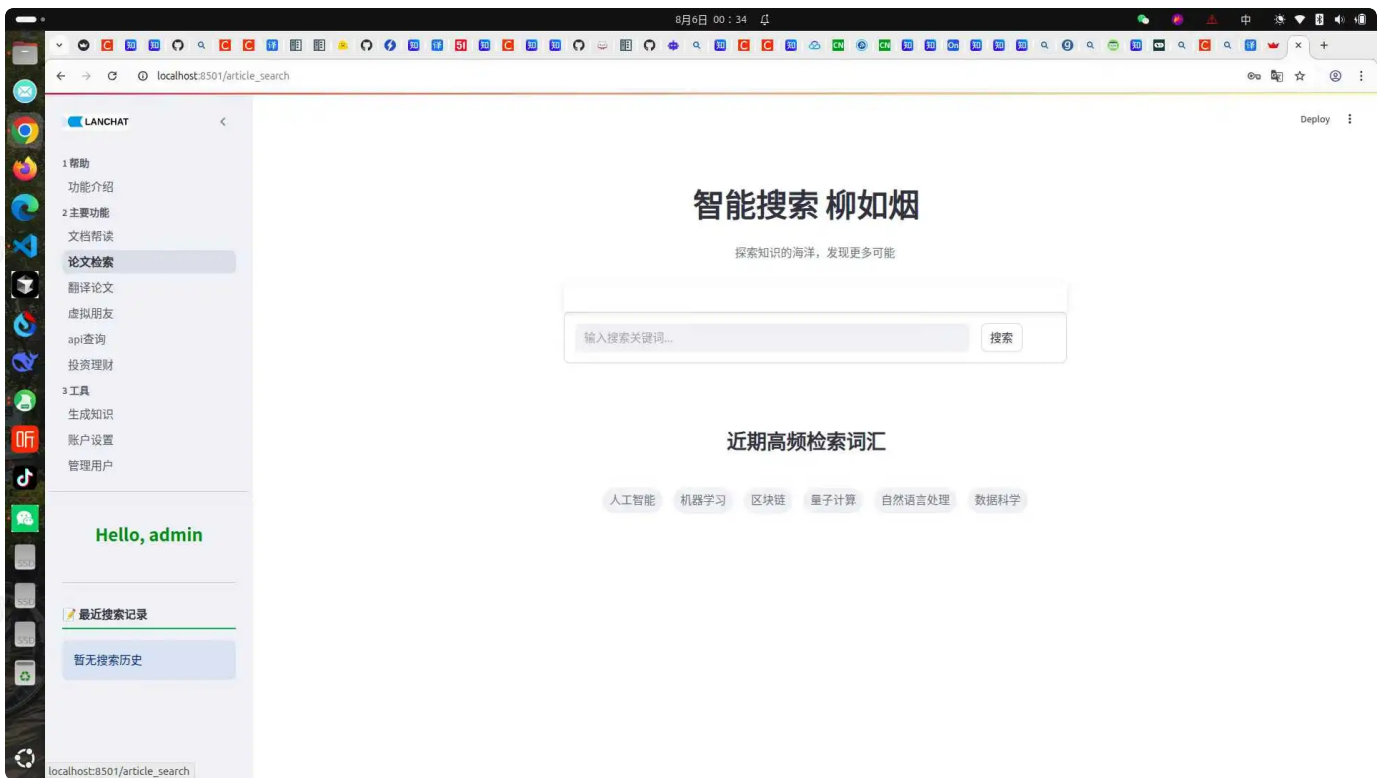
## 界面展示

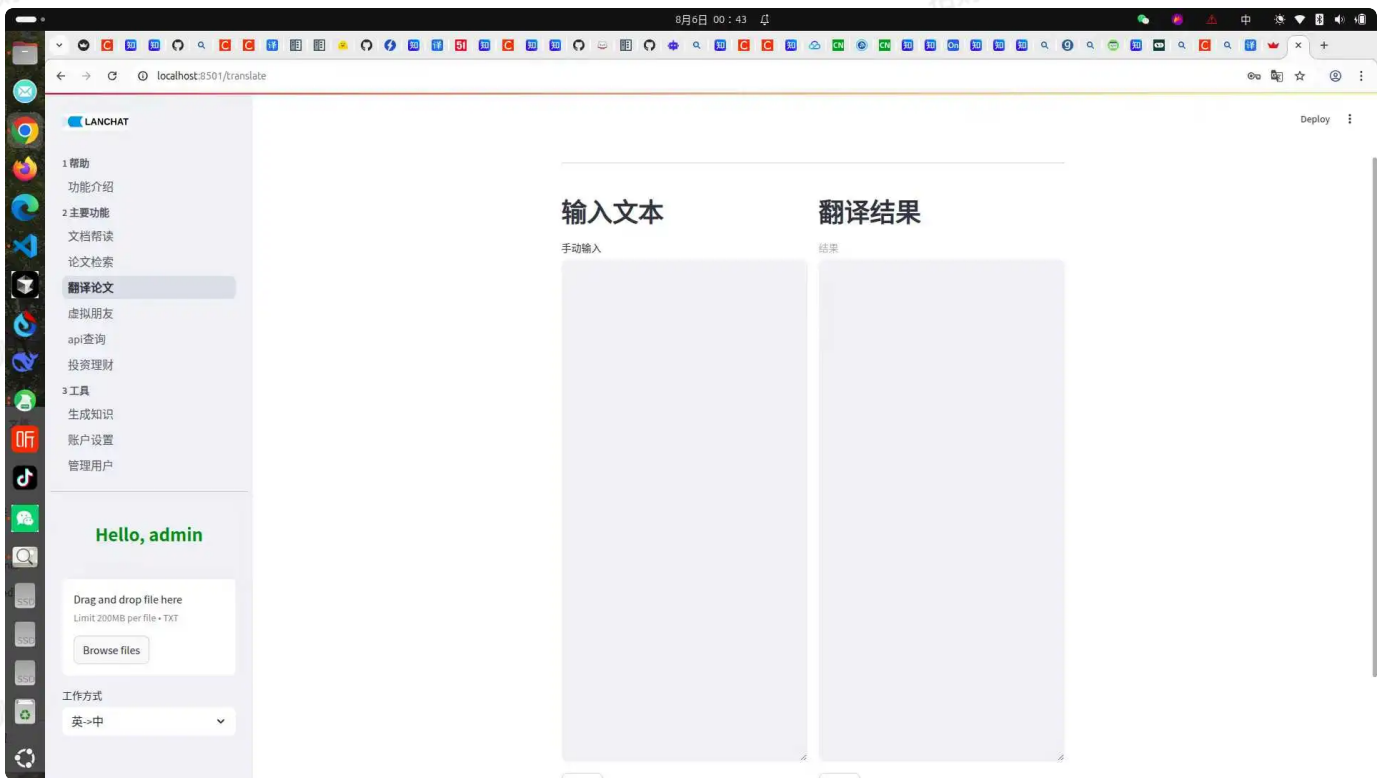
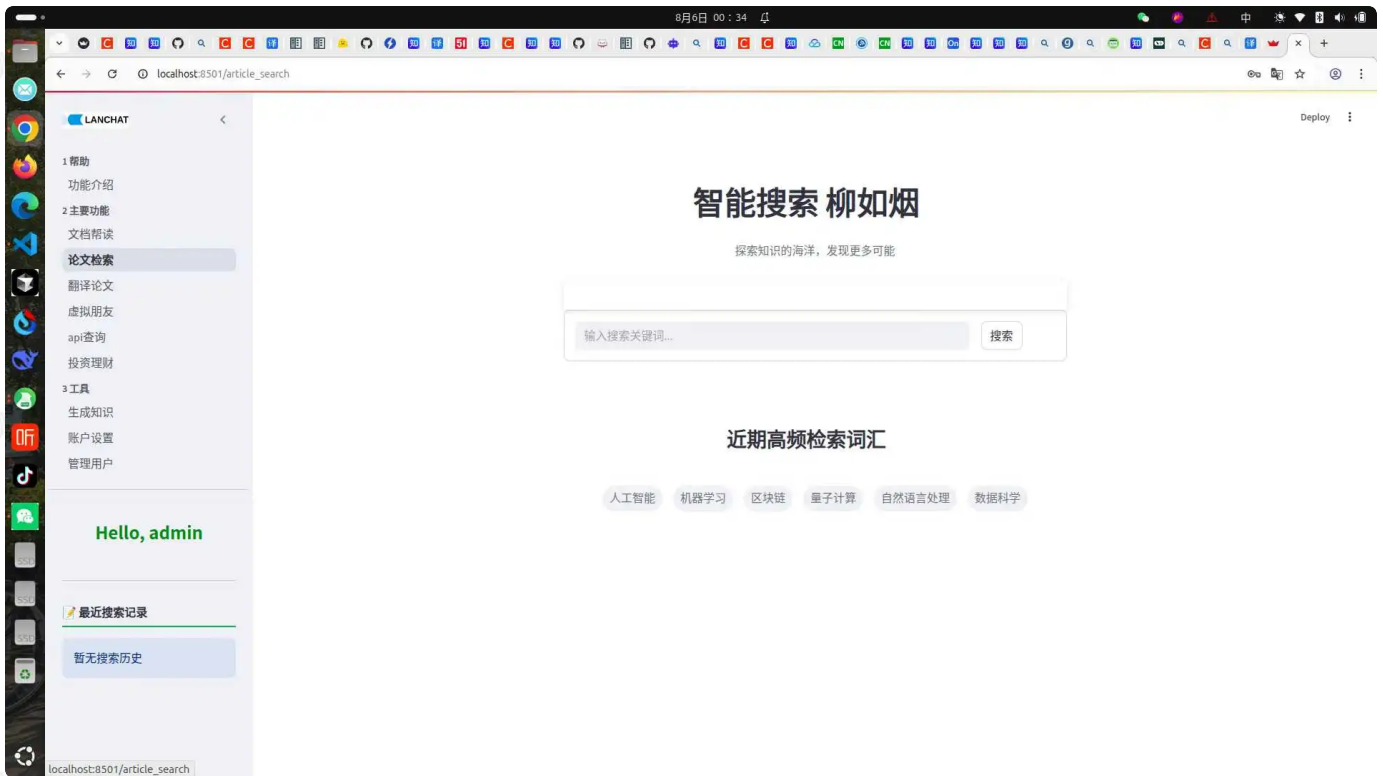
admin账户演示界面



## v0.4

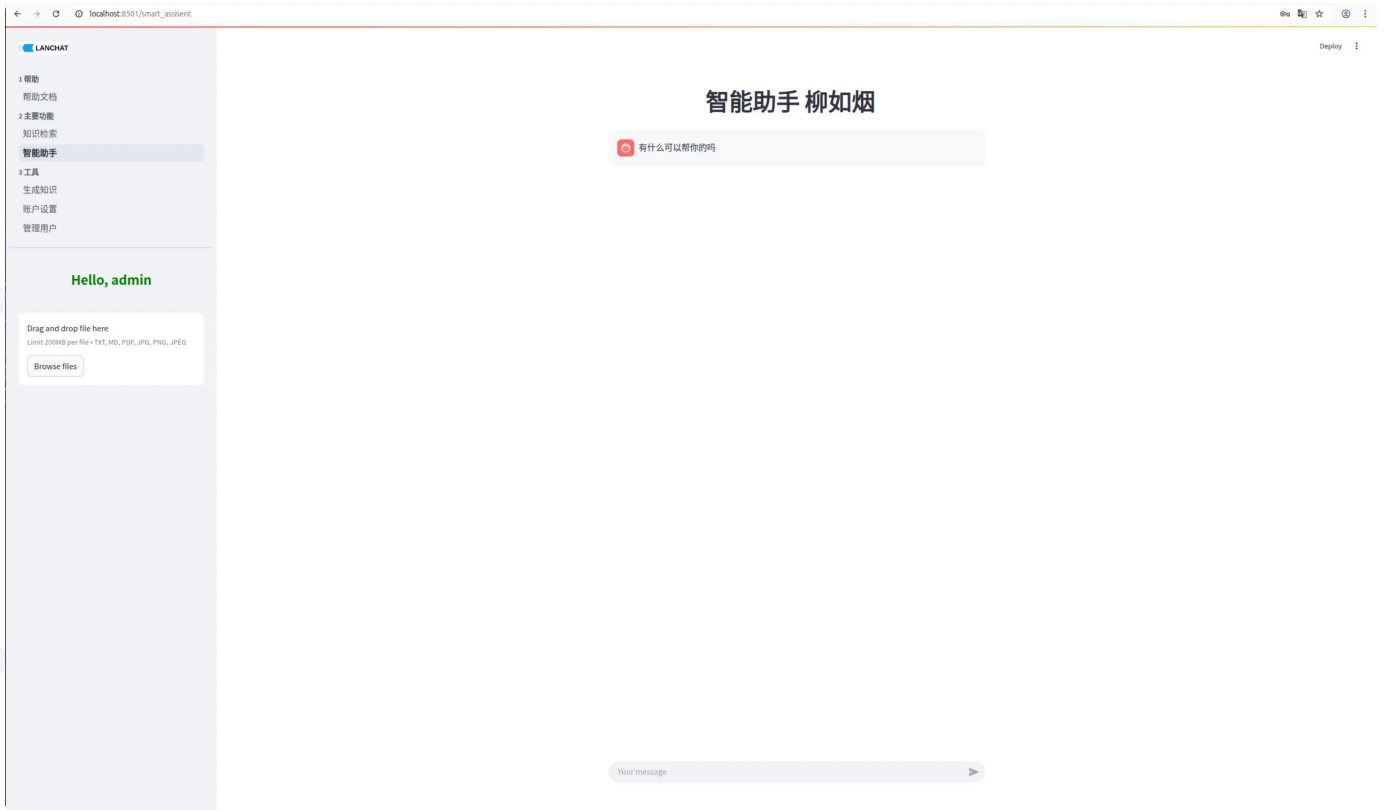
每个agent都是独立的一个界面，没有意图感知这块，速度更快。





v0.5

对多个agents合并，由于不会前端开发，最后一个没有作合并。加入了意图识别，消耗资源相对更多。



v0.6

on the way

# 其他资源

[LANchat.wps](#)

[LANchat.pdf](#)