

Forum:[Hack Forums](#)

Thread: [\[INTEREST CHECK\] - Local Uncensored LLM Auto-Reply Engine w/ Full Local Support](#)

Ratt Pack - April 21, 2026, 9:17 am

TL;DR: Built a platform-agnostic local LLM uncensored chat automation tool, ~80% complete. Looking for feedback on real-world use cases and rough pricing expectations before final push.

Overview

The tool is a local-first automation layer that sits between a GGUF-compatible LLM runtime (Ollama) and any chat interface, undetected. Core value prop: full stealth workflow customization without cloud dependencies or token costs - operating on cron or looping.

Current Feature Set:

- Platform-agnostic hook system — attach to any chat surface or build headless pipelines
- Native GGUF model loading with context/parameter control per workflow node
- Visual/config-driven workflow builder — conditional branching, multi-turn chains, uncensored persona assignment
- Synthetic data generation mode with structured output formatting
- Internal chatbot deployment mode — persistent context, session management

Target Use Cases:

- You tell me. I can add features to support your use case before full build is deployed for testing.

Questions for the community:

1. Does this solve a problem you've actually hit?
2. What platform integrations would be non-negotiable for you?
3. Pricing — one-time license, tiered sub, or open core + enterprise? What's fair / you'd pay?
4. Would you want this as a desktop app, CLI, or self-hosted web UI?

Will drop a demo video if there's real interest here. Appreciate any honest feedback including 'this already exists, see X.'