Sovereign AI: Running Useful, Self-Hosted Agents on Modest Hardware

- **Speaker**: Brady Dibble
  * **Room**: 332
  * **Time**: Friday, Oct 23, 9:10 am – 9:50 am
- **Format**: Talk (50 min)
  * **Difficulty**: Moderate
- **Track**: Technical, AI Users
- **Additional Tags**: Local LLM, AI Agents, Open Source, Privacy
- **Experience**: Director of Product at CIQ (Enterprise Linux, security, compliance, and HPC/AI infrastructure) and homelab enthusiast.

Description:

AI assistants like Claude and ChatGPT are powerful, convenient, and a trap. The cost, in addition to the price itself, is that your data and your code are being handed off to a corporation and who knows who else. Breaking your dependency on cloud AI takes effort and sacrifices convenience, but the gap between local LLMs and frontier models is narrowing every day. The hidden challenge is the harness—everything that wraps around the LLM to make it useful.

This talk provides a practical tour of an agentic AI stack you host and own. It begins with an honest look at the tradeoffs regarding reasoning quality, context length, and speed on consumer hardware. From there, it demonstrates how to rebuild capability using open tools and how to create a Beowulf cluster using inexpensive secondhand mini-PCs to turn recycled hardware into useful compute. By implementing a small routing layer to put backends behind one OpenAI-compatible endpoint, you can target your own hardware instead of a vendor. The session closes on the importance of sovereignty: ensuring your data stays on your network, your models remain consistent, and the cloud becomes a choice rather than a dependency.

Project URL: www.bradydibble.com/ai-soverignty

Target Audience:

- Privacy-conscious AI users
  * Homelab enthusiasts and tinkerers
  * Developers looking to reduce cloud AI dependency

From:
https://wiki.livingcartoon.org/ - **Living Cartoon Company**

Permanent link:
https://wiki.livingcartoon.org/events/seagl_2026_sovereign_ai

Last update: **2026/10/01 22:25**