


# Rescued Hardware, Rescued Bees — Building Tech From What Others Throw Away

February 12, 2026 · by Mark Bartlett

[Download this post as PDF](#)

 **More on this topic:** [Why mycoSwarm Was Born](#) · [What Open Source Was Supposed to Be](#) · [Budget Local AI PC](#) · [Best Used GPUs for Local AI](#)

I keep bees and I build distributed AI systems out of used hardware. These feel like unrelated hobbies until you notice the pattern.

---

## The Colony in the Floorboards

---

East Bay Bees does rescue work. Someone's tearing down a shed, renovating a bathroom, demolishing a deck, and they find a colony of wild bees living in the structure. The bees have been there for months, sometimes years. They built comb, stored honey, raised brood, survived winters. A functioning society inside someone's wall.

The construction crew doesn't care. The building comes down on Tuesday. The bees die or they don't. That's where we come in.

Bee rescue isn't heroic. It's slow and hot and you get stung. You cut open walls, find the comb, transfer it into frames, locate the queen, brush thousands of bees into a box, strap the box to your truck, drive home, and hope they stay. Half the time they abscond anyway. You show up again the next day and try something different.

The bees were fine. They didn't need saving. They needed someone to not destroy them. There's a difference, and it matters.

What you're doing is redirecting, not rescuing. The colony has everything it needs to survive. You're just moving it out of the path of a bulldozer and giving it a box where nobody's going to tear the walls down.

---

## The ThinkCentre in the Junk Bin

---

The Lenovo ThinkCentre M710Q on my desk cost \$85 on eBay. Previous owner was probably a company that refreshed their fleet. Hundreds of these machines get dumped every quarter when IT departments decide three-year-old hardware is “end of life.”

The M710Q has a quad-core i5, 16GB of RAM, and a 256GB SSD. It runs Linux. It runs Ollama. It handles embeddings and small model inference without complaint. It draws 35 watts under load. Less than a lightbulb.

The RTX 3090 in my workstation cost \$850 used. Someone upgraded to a 4090 and sold this one on r/hardwareswap. The card still has 24GB of VRAM. It still runs 70B quantized models. It still generates 45+ tokens per second on a 7B model. Nothing about it is worse than the day it was manufactured. The only thing that changed is the number on the box.

The Raspberry Pi that coordinates the swarm cost \$35. It’s been out for years. Nobody’s excited about it anymore. It does its job.

Node	Cost	Source	Role in Swarm
RTX 3090 workstation	\$850	r/hardwareswap	Heavy inference, 32B+ models
ThinkCentre M710Q	\$85	eBay fleet liquidation	Embeddings, light inference, API routing
Raspberry Pi 4	\$35	Sitting in a drawer	Coordinator, health monitoring
<b>Total</b>	<b>\$970</b>		

\$970 for a distributed AI lab. Every piece of it was headed for a landfill, a recycling bin, or a drawer where it would sit until the owner felt guilty enough to throw it away.

There is a parallel here and I didn’t plan it.

---

## Wu Wei and the Path of Least Resistance

---

I practice Tai Chi. Not the park-bench-morning-exercise version, the martial art. Push Hands, application work, the whole system. One of the principles that keeps coming back is Wu Wei.

Wu Wei gets translated as “non-action” or “effortless action,” which makes it sound passive. It’s not. It’s the difference between forcing a result and finding where the energy already wants to go. In Push Hands, you don’t overpower your partner. You feel where they’re already moving and redirect. Less effort, better outcome.

The tech industry does the opposite. It forces. Need more performance? Buy a newer GPU. Need more VRAM? Buy a more expensive card. Need AI capabilities? Subscribe to a cloud API. The answer is always acquisition. Spend more, upgrade faster, replace what you have with what they're selling.

Wu Wei asks a different question: what can the hardware I already own actually do?

Turns out, a lot. That M710Q can run a 7B model on CPU at usable speeds for simple tasks. The 3090 handles anything up to 70B quantized. The Pi coordinates. Together, they're a system. Nobody had to buy a \$1,600 RTX 4090 or a \$2,000 Mac Studio. The path of least resistance was \$85 on eBay and a used card someone else decided they didn't want.

mycoSwarm's design philosophy comes directly from this. The first thing a node does when it joins the swarm is detect its own capabilities. What GPU do I have? How much VRAM? What models are loaded? What can I actually run? It doesn't assume you have top-tier hardware. It works with what you give it.

That's not a limitation. That's the point.

---

## Mycelium Doesn't Have a Data Center

---

The name mycoSwarm comes from mycelium, the underground fungal networks that connect trees in a forest. If you've never read about mycorrhizal networks, the short version is this: the root systems of trees are connected by fungal threads. When one tree has excess sugar, the mycelium moves it to a tree that needs it. When a tree is under attack from insects, chemical warning signals travel through the network. Resources flow where they're needed without anyone directing traffic.

There's no central server. No hub-and-spoke architecture. No CEO of the forest deciding which tree gets nutrients. Individual fungal threads are simple. They move molecules in one direction or another based on concentration gradients. The intelligence, if you want to call it that, is emergent. It comes from connection, not control.

I keep seeing people build AI infrastructure the same way corporations build everything else. One powerful machine. One big GPU. One person in charge of one system. The server room model scaled down to a desk.

Mycelium doesn't work like that. Neither do bee colonies, for that matter. A hive has no manager. No bee tells other bees what to do. Individual bees follow simple rules (dance to share location,

fan to cool the hive, feed larvae that are the right age) and the collective behavior emerges from those rules applied in parallel across thousands of individuals.

mycoSwarm is trying to be more like mycelium and less like a server room. Simple nodes. Automatic discovery. Resources flowing to where they're needed. The 3090 doesn't "control" the ThinkCentre. They announce their capabilities and the routing layer decides where each task fits best.

Is this idealistic? Probably. Does it actually work at scale? I don't know yet. But the architecture feels right in the same way that watching a healthy hive feels right. Lots of small things doing their own work, and something larger happening because of it.

---

## The Common Thread

---

I didn't set out to build a philosophy. I rescued some bees because they were going to die. I bought used hardware because it was cheap. I practiced Tai Chi because my back hurt. The connections between these things became obvious later.

The instinct in all three is the same: don't throw away things that still work.

A bee colony in a wall isn't broken. It's in the wrong place. A ThinkCentre from 2019 isn't obsolete. It's been declared obsolete by people who sell new ThinkCentres. A body that's stiff and sore isn't damaged. It's been forced into postures it wasn't designed for. The fix in each case isn't replacement. It's redirection.

The tech industry has a replacement addiction. New GPU every two years. New framework every six months. New subscription every time you need a capability. The old stuff goes in a pile and you try not to think about it.

I'd rather build from the pile.

Not because I'm cheap (though I am). Not because I'm making a political statement (though I suppose I am). Because the pile is full of perfectly good hardware that works. Because the energy it took to mine the lithium, smelt the silicon, assemble the board, and ship the box across an ocean is already spent. The least wasteful thing you can do with a working machine is use it.

---

## What This Looks Like in Practice

---

The romantic version of this philosophy is bees and Tai Chi and mycelium. The practical version is three machines on a shelf running Ollama.

Here's what the \$970 swarm actually does:

The **3090 workstation** handles heavy lifting. When I need a 32B model for code review, document analysis, or complex reasoning, the query routes there. It runs Qwen 2.5 32B at Q4\_K\_M with room to spare. This machine does maybe 20% of the total queries but handles 80% of the hard ones.

The **M710Q** handles everything lightweight. Embeddings for [RAG pipelines](#). Small model inference for simple Q&A, text classification, routine drafts. It runs a 7B model on CPU at about 8 tokens per second. Not fast. Fast enough for tasks where quality matters more than speed. This machine handles the majority of daily queries because most queries don't need a 32B model.

The **Raspberry Pi** coordinates. It monitors which nodes are online, tracks which models are loaded where, and maintains the capability map that the routing layer uses. It doesn't do any inference. It just watches and directs traffic. A \$35 air traffic controller for AI workloads.

Could I do all of this on the 3090 alone? Yes. I did, for months. But the 3090 draws 350 watts at load. The M710Q draws 35. For the 80% of queries that don't need heavy compute, I'm burning ten times the electricity for no benefit. The distributed approach matches resources to requirements. Heavy queries get heavy hardware. Light queries get light hardware.

Wu Wei. Path of least resistance. Don't use a 3090 to answer "what's the capital of France."

---

## Why I'm Telling You This

---

This isn't a sales pitch. There's nothing to buy. mycoSwarm is MIT-licensed on [GitHub](#), there's no company behind it, and there never will be.

I'm telling you this because I think a lot of people in the local AI community are sitting on hardware they've been told is worthless. A 3060 from 2021. A Mac Mini from 2020. A NUC gathering dust since the kids got phones. Old laptops. Retired gaming PCs. Machines that were supposed to be recycled but never quite made it to the bin.

Those machines work. They have CPUs, RAM, sometimes GPUs. They can run Ollama. They can serve embeddings. They can coordinate tasks. They can be part of something.


The bee colony in the floorboards didn't need a new hive. It needed someone to notice it was there and give it somewhere to go. The ThinkCentre in the eBay listing doesn't need a faster processor. It needs someone to plug it in and give it a job.

I don't know if mycoSwarm will become something other people use. The code is early. The discovery layer works. The capability detection works. The [GPU sharing works](#). The full orchestration is still being built. Maybe the coordination overhead turns out to be a dealbreaker. That's fine. The experiment is worth running.

But even if mycoSwarm goes nowhere, the philosophy holds. Build from what you have. Rescue what others discard. Work with the grain instead of against it.

The mycelium has been doing this for 400 million years. The bees figured it out too. I'm just a guy with a ThinkCentre and a bee suit, trying to keep up.

---

 **Related:** [Why mycoSwarm Was Born](#) · [What Open Source Was Supposed to Be](#) · [From 178 Seconds to 19](#) · [Best Used GPUs for Local AI](#)

**mycoSwarm:** [GitHub](#)

---

Source: <https://insiderllm.com/blog/rescued-hardware-rescued-bees/>

Free guides for running AI locally