

On-Premises Personal LLMs for Each User

[Assem Hijazi – Experience, Knowledge & Ideas](#)

Yes, as the title says, **on-premises personal LLMs for each user**—this is going to happen, whether we like it or not. The future isn't waiting for us to be ready; it's charging ahead, and personal AI is leading the way. The idea might sound exciting, empowering, even revolutionary, but it's also a little terrifying. What happens when every individual has the power to own, train, and deploy their own LLM on their laptop, desktop, or even their smartphone?

Think about it. The same technology that today sits behind corporate firewalls or massive cloud infrastructures will soon be running locally, tailored for individual needs. No gatekeepers. No cloud. Just you and your AI. It's a dream for creativity and independence—and a potential nightmare for regulation and ethics. Personal LLMs will replace many of the apps and tools we rely on today, giving us power and control like never before. But at what cost?

This shift will change everything: the way we create, the way we work, and the way we interact with technology. It will redefine industries and challenge the dominance of big players like Microsoft, OpenAI, and Apple. The question isn't if this will happen—it's how soon. And when it does, are we really ready for the implications? Let's dive in.

You don't believe me? Think twice. What does it mean to already have **personal AI** embedded in your Apple phone? Doesn't it give you a clear hint of what's coming? The foundation is already here, quietly sitting in your pocket, running tasks locally, and learning about you. If that's not a glimpse into the concept of personal LLMs, what is?

And here's the kicker: what seems today like an expensive, resource-heavy feat—requiring massive hardware, deep learning systems, and endless training data—will soon become as simple as installing an app on your laptop. Even quantum computing, once reserved for science fiction, will shrink to personal devices, enabling anyone to run, train, and customize their own LLM without breaking a sweat. The future is closer than you think, and it's moving fast.

The Complex Matrix of LLM Intersections

Today, the main players—OpenAI, Microsoft, Google, and others—are running the show, providing **generic LLMs** to serve the masses. But the stage is getting crowded. Countries are stepping in, launching their own localized LLMs, claiming better services tailored to their citizens. They're using open-source models, building expensive infrastructures, and signing partnerships with major players. Meanwhile, enterprise companies are following suit, building **enterprise-focused LLMs or even smaller, specialized models (SLMs)** to meet their unique needs.

And now, another road has joined this already crowded intersection: **Apple's personal AI**. This is, without a doubt, a huge innovation and a bold step forward. But imagine, as we look 3 to 5 years ahead, how many more roads will converge at this intersection.

Why is it becoming more complicated? Because all the **expensive requirements** for building and running LLMs—massive infrastructure, deep learning, training data, and specialized hardware—will soon become significantly easier and cheaper to manage. Decentralization is no longer a distant dream. It's becoming increasingly feasible to build **on-premises LLMs or SLMs**, or whatever you choose to call them. Hardware vendors, from laptops to IoT devices, will begin integrating personal AI capabilities into their products. Giants like Apple are already leading this wave, and others will follow, adding even more complexity.

Picture this: a future where every laptop, phone, or IoT device offers **personal AI**. LLMs won't just exist in the cloud or on massive data centers. They'll be embedded everywhere, tailored for individual users. It will make the entire ecosystem—spanning governments, corporations, and global players—**unbelievably fragmented**. The competition will be chaotic. Who will you trust? Your local government's LLM? A global corporation's LLM? Your enterprise's SLM? Or, simply, your **own personal LLM**, running securely and privately on your own device?

The **only human-centric solution** is this: every individual needs to have their **own LLM**, running on their premises. When that happens, people will care less about global players, local governments, or even enterprise models. Why rely on someone else's AI when your own can do the job just as well, if not better?

Sound scary? Yes, it is. But this is exactly where we're heading.

LLM Is Not a Technology Only

Some might try to fit LLMs into neat frameworks, dismissing them as a bubble or a phase, much like the dot-com boom. Others view them through the lens of Gartner's Hype Cycle, predicting a rise, crash, and eventual stabilization. But **LLMs are not just a technology**.

They represent something far more profound—a **cloning of human thinking**. They simulate reasoning, creativity, decision-making, and problem-solving. These aren't mere tools or temporary trends; they're the foundation of a future where machines don't just process—they understand. They think. They adapt.

This is not about hype or cycles. It's a fundamental redefinition of intelligence and interaction. To dismiss LLMs as "just another technology" is to miss the point entirely. They're reshaping our world, and we're only beginning to see the impact.

Decentralized Is the Future

I've just scratched the surface in this article. In this series on the **future of LLMs and Generative AI**, we'll dive deeper into these concepts. But one thing is clear: **decentralization is the future—100%**.

This is not just speculation; it's something I started working on four years ago when I built the **Digital Human System**, a framework based on decentralized accounts for every human. This is the direction we're heading, whether we like it or not.

We cannot continue in a world where every human needs to repeatedly customize and re-enter their profile for every system. The current model is broken. What will happen soon is inevitable: **all systems must become compatible with a single profile for every human**, enabling seamless interaction and providing **personal AI for everyone**.

This is the future. It's not just a possibility—it's a necessity. And it's happening faster than we think. Let's get ready.