

Domestic AI

Domestic AI

Essay 3 — Understanding AI Course

What if you could have a powerful AI assistant that runs entirely on your own computer — no internet required, no data going to any company, no subscriptions, no privacy worries? That's what **local AI** (also called **domestic AI** or **private AI**) is all about. In this essay, we'll look at what it takes to run AI at home, what your options are, and whether it's the right choice for you.

So... AI On Your Own Computer?

Yes — it's genuinely possible. You don't need to be a tech expert, though some technical comfort helps.

Running AI locally means installing software on your own machine that lets you interact with a language model without sending your data anywhere. Everything stays on your device.

This is attractive for several reasons:

- **Privacy:** Nothing leaves your computer. Your conversations, your files, your questions — completely private.
- **No internet required:** Once installed, local AI works offline. Useful in areas with poor connectivity.
- **No ongoing subscription costs:** You pay for the hardware once, and that's it (though you may need to pay for some software).
- **Customisation:** You can tweak and configure things more freely than with a corporate AI service.
- **Learning experience:** Running your own AI gives you a much deeper understanding of how it works.

But there are trade-offs too, which we'll come to.

What Hardware Do You Need?

This is where it gets practical. Running AI locally requires a computer with a reasonably powerful processor, and — crucially — a decent **graphics processing unit (GPU)** or **Apple Silicon** chip.

Why does the GPU matter?

AI models do a lot of math, and they can do it much faster when they have hundreds or thousands of small processors working in parallel. A GPU, originally designed for rendering video games, happens to be very good at this. Think of it like having a team of mathematicians working simultaneously rather than one person working alone.

Minimum requirements (rough guide)

Specification	Minimum	Good	Ideal
RAM	8 GB	16 GB	32 GB or more
GPU VRAM	6 GB	8–12 GB	16–24 GB
Processor	Modern quad-core	Modern 6–8 core	Modern 8+ core
Storage	50 GB free	100 GB+ free	200 GB+ free (SSD recommended)
GPU type	NVIDIA GTX 10-series or better	RTX 3060, 4060, AMD equivalent	RTX 4080, 4090, or Apple M-series chip

Important: Not all computers can upgrade their GPU. Most laptops can't. Many pre-built desktops can. Apple Macs with M1, M2, M3, or M4 chips are excellent for local AI — the Neural Engine in Apple Silicon is designed specifically for AI workloads.

Cost in GBP: You can potentially use an existing computer or laptop if it has a compatible GPU or Apple Silicon chip. If buying new: - A decent second-hand PC with a suitable GPU: **£400–£800** - A new desktop PC: **£800–£1,500+** - A new Mac with Apple Silicon: **£1,000–£2,500+**

Your Software Options

Once you have the hardware, you need the software. Here are the main options:

1. Ollama (easiest for beginners)

Ollama is a free, open-source tool that makes running local AI remarkably simple. You download it, run a single command in the terminal, and the AI is running. It supports many popular models including Llama, Mistral, Gemma, and others. Available for Mac, Linux, and Windows.

Pros: Very easy to set up, good range of models, active community.

Cons: Command-line interface (no fancy app window), limited graphical interface.

2. LM Studio

A more user-friendly application with a graphical interface. You can download models, adjust settings, and chat with them in a clean app window. Free version available; paid version adds extra features.

Pros: Attractive interface, easy model management. **Cons:** More resource-intensive than Ollama in some cases.

3. GPT4All

Another open-source option designed to be as accessible as possible. Includes a downloadable app and a curated list of models.

Pros: Very easy to use, good documentation. **Cons:** May be slower on older hardware.

4. Jan (formerly Fireworks AI)

An open-source alternative to ChatGPT-style interfaces, designed to run locally.

Pros: Clean interface, active development. **Cons:** Relatively newer, less community support than Ollama.

5. Open WebUI

A self-hosted, open-source WebUI that works with Ollama. Gives you a ChatGPT-like experience in your browser, running entirely on your machine.

Pros: Excellent interface, feature-rich. **Cons:** Requires a bit more setup than LM Studio.

6. Apple Intelligence (built into Mac, iPhone, iPad)

Apple has integrated local AI features into its latest operating systems. On supported Macs (M1 and later), certain AI tasks run entirely on-device using Apple's own foundation models.

Pros: Already built in, very smooth integration. **Cons:** Limited to Apple hardware, Apple controls the models, less customisable.

Which Model Should You Choose?

There are many AI models you can run locally, each with different strengths:

Model	Best For	Size (approx.)
Llama 3 (Meta)	General conversation, writing, coding	8B–70B parameters
Mistral	Good all-rounder, efficient	7B parameters
Gemma (Google)	General use, relatively efficient	2B–7B parameters
Phi-3 (Microsoft)	Smaller, efficient, still capable	3.8B parameters
Qwen (Alibaba)	Multi-lingual, good for non-English	Various
DeepSeek R1	Reasoning and problem-solving	Various

The number refers to how many parameters (tiny adjustable settings) the model has. Larger models tend to be smarter but need more powerful hardware.

My advice for beginners: Start with **Ollama + Llama 3 or Mistral** on a model with around 7–8 billion parameters. This gives a good balance of capability and accessibility.

The Problems — Let's Be Honest

Local AI isn't all upside. Here are the genuine challenges:

Capability gaps

The most powerful models (GPT-4 class, Claude 3 Opus) are far beyond what most local hardware can run. Local AI is improving rapidly, but if you need the absolute best performance, corporate AI still leads — for now.

Technical know-how

While tools like Ollama have made this much easier, there's still a learning curve. Expect to troubleshoot, read documentation, and maybe use a command line.

Hardware costs

Getting a machine capable of running the better models can be expensive. It's an investment.

Maintenance

Software updates, model updates, handling errors — it all takes time. You're the IT department.

No internet-aware models by default

Many local models don't have access to the web (though this is changing with tools like Open WebUI with search integration).

The Benefits — Why Bother?

And yet, many people find it genuinely worthwhile:

- **Total privacy:** Your thoughts, your data, your files — nothing leaves your machine.
 - **No censorship or content restrictions:** Local models can be configured more freely.
 - **Always available:** Works even without internet.
 - **Satisfying to run yourself:** There's something empowering about understanding and controlling the technology yourself.
 - **Great for learning:** Running and experimenting with AI locally gives you a much deeper understanding than just using a web interface.
-

How Can This Help Me and My Family?

Here are some genuinely useful things you can do with a home AI:

- **Homework helper:** Explain concepts, help with writing, solve problems — without sharing your child's work with a third party.
- **Drafting and writing:** Write letters, CVs, emails, stories — all offline and private.
- **Research and learning:** Ask questions, get explanations, explore topics without tracking.
- **Coding help:** Learn to program, debug your code, build small tools.
- **Creative projects:** Brainstorm ideas, write stories, compose music (with certain tools).
- **Accessibility:** For some users, a private AI assistant that works completely offline is far more useful than a cloud-based one.

For families, local AI can be a way to explore and learn together — without the privacy concerns of sending everything to a corporation.

What Have We Learned?

- Local AI runs entirely on your own hardware, keeping your data private
 - You need a reasonably powerful computer, ideally with a dedicated GPU or Apple Silicon chip
 - Tools like Ollama and LM Studio make setup much easier than it used to be
 - There are real trade-offs: less power than corporate AI, more technical effort required
 - But for privacy, learning, and control, it can be deeply worthwhile
-

Glossary of Terms

Term	Definition
Local AI	AI software that runs entirely on your own device, with no internet connection or data sharing required. Also called private or domestic AI.

Term	Definition
GPU (Graphics Processing Unit)	A computer component originally designed for video games, but very good at the kind of math AI uses. More GPUs = faster AI.
VRAM	Video RAM — the memory on your GPU. Crucial for running AI models; more VRAM means larger models you can run.
Apple Silicon	Apple's own chips (M1, M2, M3, M4 etc.) used in Macs and iPads. Include a Neural Engine designed for AI tasks and use unified memory (RAM and VRAM are the same pool).
Parameters	Tiny adjustable settings in an AI model. More parameters generally means a smarter model, but also requires more computing power to run.
Ollama	A free, open-source tool that makes running local AI very simple. Available at ollama.com
LM Studio	A free desktop application for running local AI models with a graphical interface.
Open-source	Software whose code is publicly available, meaning anyone can inspect, modify, and use it freely.
Unified Memory	In Apple Silicon, RAM and VRAM are combined into one pool, meaning you can run larger models than a PC with the same total RAM but separate VRAM.
Terminal	A text-based interface for giving commands to your computer, used in Mac, Linux, and Windows.

Further Thinking

- What would you use a private, offline AI for in your home or daily life?
 - Is the privacy benefit of local AI worth the extra technical effort for you?
 - What size model do you think would be "good enough" for most everyday tasks?
-

Essay 3 of 8 — Understanding AI Course Authors: Bea Groves-McDaniel and SAL-9000 Licensed under Creative Commons (NC, ND) — Share with attribution, no commercial use.