

LAB NOTEBOOK · dated



THE LAB NOTEBOOK

A private AI that lived on your own machine

In April 2025 we built MyBrain: a browser plug-in that captured your own browsing, fine-tuned a small local model with LoRA, and let you ask questions of your own history — nothing leaving the device. It worked. The honest finding: on consumer hardware, retrieval usually beat fine-tuning for personal data, and we say so.

The dated R&D archive



Research — measured, partial

A private AI that lived on your own machine

In April 2025 we built MyBrain: a browser plug-in that captured your own browsing, fine-tuned a small local model with LoRA, and let you ask questions of your own history — nothing leaving the device. It worked. The honest finding: on consumer hardware, retrieval usually beat fine-tuning for personal data, and we say so.



01 The entry, dated

02 What it was

03 What we built

04 What we learned — including the honest negative

05 Where it went / status

06 Sources, glossary & the honesty ledger

Every claim carries its tier · nothing faked

PREPARED

15 April 2025

VERIFIED BY

Jon Ross — Founder, Vocabotics —
15 years building safety-critical
systems

HOW TO READ IT

Drafted with AI, verified by a
named human. Sources at the back.

01 The entry, dated



The entry, dated

Apr 2025

entry date

Models

category

~1 year ahead

lead over the world



Public

access

What the world had at the time

Spring 2025: 'personal AI' still mostly meant sending your data to someone else's cloud. On-device fine-tuning and private RAG over your own life were fringe experiments, not products — the big 'AI that knows you, privately' wave came later.

We fine-tuned a model on one person's own browsing, entirely on their own machine, and asked it questions — a private AI, a year before that became the pitch everyone gives.

The costly signal

The measured result

Every number carries its own honesty tier — the tiering is the credibility.

LoRA

local fine-tune on consumer hardware



0 bytes

personal data leaving the device



RAG > FT

the honest result for personal data



The pitch that every AI company now gives — *an assistant that knows you, and keeps your data private* — was, in early 2025, still an experiment nobody had shipped well. We built one to find out what was real about it and what was marketing.

02 What it was



MyBrain was a privacy-first personal knowledge system. A browser plug-in captured your own browsing; a small language model was fine-tuned on it locally; and you could then ask natural-language questions of your own history. The constraint that made it interesting was absolute: **nothing left the device**. No cloud, no upload, no third party holding your life.

03 What we built



A working end-to-end system with every part present:

- A **Chrome extension** that captured browsing data on-device.

- A local **LoRA fine-tuning** pipeline (Python, transformers, PEFT) that adapted a small model to that personal corpus.

- A **retrieval (RAG)** path for asking questions against the same data.

- A Node/Express + React app tying it together, with SQLite for local storage.



RESEARCH — MEASURED, PARTIAL

It ran. You could fine-tune on your own browsing and query it, with the model and the data both sitting on your own hardware. That was the whole thesis — *ahead of the curve on personal, private AI* — and it held.

04 What we learned — including the honest negative



The most valuable finding was a negative one, and it changed how we build.

Retrieval usually beat fine-tuning — and we kept building around that

For personal data, **RAG generally won over fine-tuning**. Retrieval was fresher (your data changes daily; a fine-tune is frozen the moment it finishes), more honest (it can point at the source instead of half-remembering it), and far cheaper on consumer hardware. LoRA fine-tuning on a laptop is slow, fiddly, and prone to overfitting a small personal corpus. We had expected the fine-tune to be the magic; the retrieval turned out to be the workhorse.

That lesson — *prefer a system that can cite what it knows over one that has memorised it* — is a straight line to the calculated-mind work, where the whole design goal is a model that reasons over knowledge and cites it, rather than bluffing from baked-in weights.

05 Where it went / status

Archived as a **research prototype**, its lessons absorbed. The private, on-device, "your data never leaves your machine" principle became a founding commitment rather than a feature — it is the reason the north-star product puts a model that runs *on your own hardware* at the centre. We proved the private personal AI was buildable in 2025; we also proved which half of the popular recipe actually carries the weight.

What is still open — kept visible

The honest edges of this lane, shown next to the wins. None of these is a footnote; each is a claim we have not yet earned — and marking it is what earns the right to be believed on the rest.

- ☐ LoRA fine-tuning on consumer hardware has real limits — it is slow, fiddly, and easy to overfit on a small personal corpus.
- ☐ For most personal-data questions, retrieval (RAG) beat fine-tuning on both freshness and honesty — the fine-tune's value was narrower than we hoped.
- ☐ Capturing browsing is a serious privacy surface; keeping it local is the whole point, and it constrains how much you can do.

06 Sources, glossary & the ledger

Where every claim comes from — and what the plain words mean.

Sources

- 01 **Vocabotics project audit — MyBrain (local fine-tune + RAG), Apr 2025** — Vocabotics internal project history (as of April 2025)
- 02 **LoRA: Low-Rank Adaptation of Large Language Models** — the public fine-tuning technique the experiment used (as of 2021)

THE HONESTY LEDGER

- Shipped — measured, gate-clean.
- 🔬 Research — measured, partial; the gap is shown.
- 🔭 Vision — designed, honest; not yet built.

The bigger claim goes up a tier, never down. We show the eval beside the demo — the tiering is the credibility.

See where 'owned, private AI' is heading

A Friend that runs on your own machine and is cryptographically yours — the direction this early private-AI experiment pointed at.

[/research/novaterra](#)

Nothing faked.

A real, dated lab report from our own R&D. Measured, not asserted; the honest negatives are kept on the page next to the wins — because a lab that hides its failures forfeits the right to be believed on its results.

