

🐶 One - Personal Supercomputing for Everyone

# One agent. Three tiers. One owner.

The agent is the product; the phone, the desk, and the cloud are where it runs. One consent fabric, one transparency log, one owner — the same Puppy scales from your pocket to a private supercomputer without ever changing hands.

## IN HAND · PRIMARY

### The phone is your primary AI supercomputer

iPhone you own (Android on the roadmap)

- The identity anchor — the phone number every institution banks your data against
- Always-on, always-with-you on-device inference; Secure Enclave holds your keys
- The hero node: the agent lives here first and is loyal to exactly one party — you

**Membership from \$69.69/mo · on the device you already own**

▼ same agent, same consent fabric, automatic placement ▼

## AT HOME / OFFICE · OWNED CORE

### The personal supercomputer you own outright

MacBook Pro M5 Max → NVIDIA DGX Spark / Station

- Runs the models too large for the phone — on hardware you own, data never leaving the building
- Up to 128GB unified (Mac) to 748GB coherent / trillion-parameter (DGX Station)
- White-Glove Human-First install; the consent ceremony sets every permission to your choice

**From \$3,669.69 · Ultra by reservation · reserve for \$0.69**

▼ same agent, same consent fabric, automatic placement ▼

## BURST · GOVERNED SURGE

### Always-on private supercomputing you rent and govern

Your own Google Cloud subscription (Azure, AWS, NVIDIA DGX Cloud coming)

- For the workload that exceeds the desk — model-scale train, fine-tune, batch beyond local memory
- A tenant you own: you hold the contract, the keys (CMEK), and the bill — we never mark it up
- Confidential computing + consent receipts to the same transparency log as your device

**BYO subscription + governance fee · usage-based · reserve for \$0.69**

🐶 One - Tier 3 - Bring Your Own Cloud Supercomputing

# Owned core. Governed surge.

When a workload exceeds the device, hushh One Puppy bursts into a cloud tenant you own and govern — your contract, your keys, your bill, our orchestration. We start with Google Cloud, done excellently; the architecture is multi-cloud and earns each provider as it ships.

Provider	Status	Best-in-class resources	Role in the burst
Google Cloud	LAUNCH	Ironwood TPU v7 (4.6 PFLOPS, 192GB HBM3e) · Axion Arm · A4/A4X Blackwell	Model-scale AI; ~44% lower TCO vs GB200; Arm always-on substrate
Microsoft Azure	COMING	ND-series NVIDIA (GB200/H100) · Cobalt Arm · Maia	Enterprise/BFSI estates already on Azure
AWS	COMING	Trainium2/3 + Inferentia · Graviton4 Arm · P5/P6 NVIDIA	First-party training silicon + broadest GPU supply
NVIDIA DGX Cloud	COMING	Managed GB200 / GB300	Cloud mirror of the Tier-2 DGX — desk↔cloud workload parity

## HOW PLACEMENT WORKS

The agent's placement scheduler chooses the cheapest tier that fits the workload:

- phone — always-on, identity, light inference
- desk — heavy models you own outright
- cloud burst — model-scale train / fine-tune / batch beyond local memory, in your governed tenant

Every hop — device to cloud included — is a logged consent event. Nothing moves silently.

**Reserve any tier for \$0.69 · [hushh.ai/one](https://hushh.ai/one)**