

Governance & Security

Enterprise control enforced at the endpoint, by architecture — not by policy. Built for banking, healthcare, and high-security teams.



ARCHITECTURE

Security by inheritance

- Runs as an installed app under the user's own roles and governance
- It can't do anything the user couldn't already do — "pre-governed"
- No separate agent-permission layer to build and reconcile with optional org level settings for further restrictions
- Govern the laptop and you already govern Cephable



JAILBREAK-RESISTANT

Hard rules, not prompts

- IT sets allow / deny rules that fire before Cephable acts
- Enforced at the tool-call level — it checks the app and page first
- "Ignore previous instructions..." can't defeat them; not prompt-based
- Block banking, payroll, any app or site org-wide



IT ADMIN PORTAL

Governance dashboard

- Allow / block apps, sites, connectors, skills — org-wide or per member
- Per-user allow-lists override an org-wide block; rules addable via API
- Restrict public and imported skills; keep regulated fleets closed
- Control which model sizes, Skills, and MCP servers are permitted



ON-DEVICE

Data never leaves for inference

- AI inference, camera / mic, documents and prompts stay on the device
- MCP servers are called from the client; internal ones stay behind your firewall
- Only login, updates, and optional profile sync touch the network
- We don't train on your data



HUMAN-IN-THE-LOOP

Approval gates

- Confirm before remote tool calls and transactions
- Confirm before destructive actions: delete / move files, terminal commands
- See exactly what data goes up and down on every remote call
- Start locked-down, loosen as trust builds



DATA PROTECTION

Encryption & scoping

- History and model assets encrypted to both the Cephable and OS user
- Can't be decrypted outside Cephable running for that user
- Restrict Cephable to a folder + subfolders, or grant whole-machine
- Run history exports to JSONL for third-party audit