GROUND TRUTH · CHECKED AGAINST MAIN ON 19 SEPTEMBER 2026

What Bolt and Aeira actually do

Straight answers to the questions and objections that come up most, with what is shipped today, what is still staged, and where Bolt is honestly not the right tool.

22	41	8	48
SECTIONS	CAPABILITIES SHIPPED	STAGED OR ROADMAP	QUESTIONS ANSWERED

> If a capability is not on this page, do not assume we have it. Every row is checked against the running code, and the ones that are not finished say so.

01 The one thing no one else does

We do not claim to beat every specialist tool at its single specialty. We claim something narrower and true: **no other product unifies all of the following in one governed, self-hostable surface.**

Offline daily driver

A real desktop launcher with local utilities, clipboard, notes, screenshot OCR and visual search, and local file search, useful before any model is called.

Device plus SaaS, unified

First-party context on your machine and third-party enterprise systems in one place an agent can act on.

Governance that survives a swap

Masking, approvals, audit, and identity stay the same when you change the model or the backend.

Extensible by configuration

Open MCP plus a few lines of YAML brings almost any system or format in, without waiting on a vendor roadmap.

Copilot has the Office surface but not sovereignty or model choice. Enterprise search tools have retrieval but not the local daily driver or governed action. Local-file tools have device context but no governance, no enterprise reach, no audit. Sovereign model platforms have the infrastructure but not the end-user workspace. Each owns a slice. Bolt and Aeira own the unification, and the unification is the point.

---

02 Where is my data? Evidence, not a promise

This is the first question a serious buyer asks, so here is the plain answer, and then how you prove it yourself without taking our word for anything.

On your laptop. The free desktop build keeps clipboard, notes, screenshot indexes, and your local files in an encrypted database on your own machine, and Sparcle has no copy. Nothing phones home for licensing: the desktop needs no license, and even the paid team server validates its license offline from a signed token, so there is no runtime entitlement call. The only calls to a Sparcle-controlled address are downloading the app and its updates, a version check against our public release feed that carries none of your content, and that a network-restricted or air-gapped install never makes at all. The local database runtime downloads once on first run, and opt-in AI models come from third parties such as Hugging Face, not from us. Your content is never part of any of it.

For a team. Bolt and Aeira run inside your own perimeter: your VPC, your data center, or an air-gapped network. Sparcle operates no cloud that receives your data and has no hosted ingest path. There is no Sparcle-run environment holding your data to audit, because there is not one.

Where data legitimately moves. When you connect Gmail or Jira, Bolt talks to that system's API as you configured it, which is action on your systems, not exfiltration to us. When a prompt genuinely needs a model, it goes to the model you chose under your own key, and PII is masked at that boundary first, even if you chose a public model. You define where the boundary sits; Bolt enforces it.

You do not have to trust the sentence "we never see your data." You can prove it. We do not give you a promise to file away; we give you a design you can test.

Turn off Wi-Fi. On the desktop build, local utilities, clipboard, notes, and file search keep working. Data that never needed the network never touched it.

Read your own egress logs. In a self-hosted deployment everything runs in your VPC, so your own flow logs and network monitoring are the proof, not our word. Deploy Bolt in a subnet with no route to the internet, and your own logs show it cannot reach Sparcle at all.

Verify the audit chain yourself, offline. Bolt emits a tamper-evident, independently verifiable audit trail, so you and your auditor confirm what actually happened in your environment, over time, instead of waiting on our attestation.

There is nothing at Sparcle to breach. The claim is architectural, not contractual. There is no hosted path, so there is no vendor database of your data that can leak.

> **Why this is a big deal.** Every SaaS AI tool you route data through is one more third party in your data path, and third parties are now one of the fastest-growing breach vectors. Verizon's 2026 Data Breach Investigations Report found the share of breaches involving a third party rose again, from 30 percent to 48 percent in a year, across more than 22,000 confirmed breaches. You cannot patch, monitor, or air-gap a vendor cloud you do not control. The only third-party path you never have to defend is the one that does not exist, and that is the one Bolt removes: Sparcle is not in your data path, so there is one less place your data can be lost.

---

03 For the people who sign off

A security review does not ask how many features you have. It asks where data goes, what happens the moment it has to cross a boundary, and whether you can prove any of it. Here is how Bolt answers each question a CISO, a General Counsel, or a data-protection officer actually asks, with evidence rather than a promise.

WHAT YOUR SECURITY TEAM ASKS	THE BOLT ANSWER
Where does our data live?	Inside your perimeter, always. Self-hosted in your VPC, data center, or air-gapped network; the free build stays on the laptop. Sparcle operates no cloud that receives your data, so there is no vendor store to breach, and you confirm it from your own egress logs, not our word.
What happens when data has to reach a model (AI-DLP)?	PII, secrets, and regulated identifiers are masked at the model boundary before any prompt leaves, then restored in the response, with real check-digit validation so look-alikes are caught, not just regex-matched. The everyday reformat, decode, compute, and redact tasks that tempt people into public chatbots are absorbed locally, so the leak never starts.
Does it cover the regulations we are bound by?	Yes, and you can extend it. 28 signed packs ship in the box: 22 jurisdiction regimes (GDPR, UK GDPR, CCPA and US state acts, DPDP, PIPL, APPI, PIPA, LGPD, PDPA, POPIA, KVKK, and more) plus healthcare, finance, defense, legal, and logistics cores. Enabling a pack masks that regime's identifiers before any LLM call and on every share, honors its data-subject rights and residency, and lights up the safeguards dashboard. Packs install as signed bundles even air-gapped, and you can author your own. They deliver safeguards and evidence, not a compliance certification or a consent manager, so pair them with your DPO or counsel.
Is data encrypted at rest?	Yes. Clipboard, notes, and indexes live in a whole-database-encrypted local store with an on-device key; connector and OAuth tokens sit in a separate AES-256-GCM vault. Key custody is swappable to your own Vault, cloud KMS, or HSM. Nothing local sits in plaintext.
Can we keep our own model and identity?	Yes. Bring any LLM and key, or a local model, and swap anytime with zero token markup; authenticate with your own OIDC or SAML. You are not welded to one vendor's cloud, model, or directory.

WHAT YOUR SECURITY TEAM ASKS	THE BOLT ANSWER
If we change the model, do we re-run compliance?	No. Masking, human approvals, audit, and identity-aware access attach to the boundary, not the model. Move from a frontier API to an air-gapped local model and the same governance applies, so a model change is not a compliance event.
What if we switch identity provider or SaaS suite later?	You are not locked to any of them. Authenticate with any OIDC or SAML provider (Okta, Microsoft Entra, Auth0, Google, Zoho, or any generic OIDC or SAML) and change it later; Bolt sits above Microsoft 365, Google Workspace, Zoho, Slack, Jira, and more through the open Model Context Protocol, so you can run a mix of suites or move from one to another and Bolt does not change. Copilot is welded to Microsoft and Gemini to Google; Bolt is welded to no one. Your masking, approvals, audit, and identity-aware access are defined once and survive every swap, whether it is the identity, the suite, the model, or the deployment, so a vendor change is never a governance rewrite.
Can we prove what actually happened?	Yes. A tamper-evident, independently verifiable audit chain you check yourself, offline, in your own environment, over time, instead of waiting on a vendor's annual attestation window. Two details matter to an auditor and are live: every report you produce is archived as the rendered bytes that were shown, not as inputs to regenerate later, because a regenerated report is a different claim about the same period; and each report states how much of the window each category was actually recorded for, so a category that was switched off reads as unmeasured rather than printing as a clean zero.
Who can take actions, and are they gated?	Destructive actions (send, delete, schedule, pay) require a human approval that survives a restart; commands Bolt injects and agent actions cross a native policy chokepoint where admins allow or deny by rule; RBAC and identity-bound access throughout.
How much new attack surface are we adding?	Less, not more. One governed app absorbs roughly two dozen ungoverned point tools, so there are fewer vendors and fewer places data can leak. Third-party involvement in breaches reached 48 percent in Verizon's 2026 report, up from 30 percent the year before; the path you remove is the one you never have to defend.

> The through-line: every row is answered by architecture you can inspect, not a certificate you have to trust. That is the difference between "we promise we protect your data" and "here is why your data never reaches us, and here is how you verify it."

---

04 What crosses the boundary, and what never does

High-level promises are easy. Here is the itemized version a security architect can check: for every kind of data Bolt touches, where it lives and whether it ever leaves the boundary you defined.

DATA	WHERE IT LIVES	DOES IT LEAVE YOUR BOUNDARY?
Clipboard, notes, snippets, screenshot index, local file index	Encrypted on your device, or your team's own store	**No, never.** These work with the network turned off.
A prompt that genuinely needs a model	The model you chose, under your own key	**Only to the model you selected, and PII is masked first.** Point it at a local or in-perimeter model and it never leaves your boundary at all.
An action on a connected system (Gmail, Jira, and more)	That system's own API, as you configured it	**Only to the system you already use.** That is action on your systems, never a copy to Sparcle.
Audit and governance events	Your deployment, and your SIEM if you forward them	**Stays in your environment.** Forwarded events are PII-free.
The app and its updates	Downloaded from our public release feed	**A version check only, no content,** and a network-restricted or air-gapped install never makes it.
Optional on-device AI models (voice, visual search)	Fetched once, on your opt-in, from third parties such as Hugging Face	**Model weights come in; none of your content goes out.**
Anything reaching Sparcle	There is no hosted ingest path	**Nothing, ever.** There is no path for your content to reach us.

Read the table top to bottom and the pattern is the only claim that matters: your content either stays on your device or goes to a destination you chose and control (your model, your SaaS, your SIEM). The one row that is always empty is your content reaching Sparcle.

05 Runs on your laptop, works offline

This is the part external writeups miss most, because these features live inside the app and were under-described in public copy. They are shipped and generally available today. You can verify the offline claim by turning off Wi-Fi and watching local tasks keep working.

CAPABILITY	STATUS	WHAT IT DOES
Overlay launcher `Cmd/Ctrl+Shift+K`	GA	Non-disruptive overlay on top of your focused app; the daily entry point.
90 instant offline utilities	GA	JSON/YAML/SQL, base64/hex, hashes, JWT, regex, timestamps and time zones, currency and unit conversion, UUIDs, color, and more, with no LLM call and no network.
Clipboard history	GA	Captures copied text, images, and files into a searchable, pinnable history, stored encrypted on your device.
Notes and image annotation	GA	Device-local rich notes (formatting, lists, checklists, tables, links) with full-text search, encrypted at rest. Mark up screenshots and images (arrows, text, redaction) and keep or share the result. See secure sharing below.
Browser extension and live context	GA	A Chrome extension (side panel, omnibox) brings the current tab and selection to Bolt. In the launcher, pull your screen, selection, current tab, clipboard, or an on-device OCR of what is on screen into a prompt as first-class context. The sensitive ones are masked at the boundary before any model sees them.
Timers, reminders, and snippets	GA	Local timers, reminders, notifications, and text snippets with expansion, so the launcher stays the daily driver for small tasks too.
Screenshot capture and OCR	GA	On-device optical character recognition makes the text inside screenshots and images searchable.
Screenshot visual search	GA, OPT-IN	Search screenshots by visual meaning with an on-device image model, downloaded once on your explicit opt-in.

CAPABILITY	STATUS	WHAT IT DOES
Local file search	GA	Indexes and searches your own file locations by name and content, with exclusions you control.
Desktop agent	GA	With your permission, browse local files, run terminal commands, and open applications, sandboxed and permission-controlled.
On-device voice input	OPT-IN	Speak to the launcher: the microphone is captured and transcribed entirely on-device with Whisper, so audio never leaves the machine. It adopts a local Whisper model if present, otherwise offers a one-time download. Currently an opt-in Labs experiment in builds that include it, not on by default.
On-device encrypted storage	GA	Clipboard, notes, and indexes live in an encrypted local database; an embedded database ships in the app, so no external database is required for the desktop build.

Encrypted at rest, everywhere. Clipboard history, notes, and screenshot indexes are stored in a whole-database-encrypted local store, with the encryption key derived on-device and file-permission locked. Connector and OAuth tokens are held in a separate AES-256-GCM vault. Nothing local sits in plaintext.

Precise note on search: screenshot visual search uses an on-device image model. General document retrieval in Bolt is keyword and entity based; Aeira is the dedicated plane for enterprise document search.

06 The browser inside Bolt, and the check on the way out

Bolt ships its own browser, not only the Chrome extension. It is a real browsing surface inside the app, and the thing that makes it worth using is what happens when you type or paste into a page: the content is screened on your machine before it is allowed to leave it. This is the part external writeups miss most, because the extension is the only half visible from outside.

CAPABILITY	STATUS	WHAT IT DOES
Screened before it leaves	GA	Paste a customer list into an AI site and Bolt names what it found, grouped by type with a count and shown without repeating the values back, then holds the page while you decide. Proceeding requires a written justification of at least ten characters and writes a durable incident record with the sealed enforcement entry, not only a counter.
A tab that will not overstate itself	GA	The tab shows "Guarded" only once the in-page guard has actually reported in. Where it cannot screen, it says "Not screened" rather than implying a protection it does not have. A shield that lies is worse than no shield.
Your logins fill in the page	GA	A vault offer per site, unlocked with a PIN on Bolt's own surface, so a password reaches the form without a round trip through another app. Installing a vault is not permission to read it: consent is explicit and per vault.
One session across every tab	GA	Clipboard, notes, and the current page stay reachable from the same bar in any tab, so the browser is part of the launcher rather than a separate app you alt-tab to.
Ads and trackers blocked by default	GA	A native content rule list is applied per tab, on by default, with a single switch in settings.
Sites Bolt will not capture from	GA	Banking, health, and sign-in domains ship on a capture deny-list, so context from those pages is never staged in the first place. You edit the list.
Private tabs that are actually private	GA	A private tab records no history, no session snapshot, and no address in the diagnostic logs either. That last sink is the one a privacy promise usually forgets; we found ours in an audit and closed it in 0.1.162.
Delete browsing data,	GA	The time range applies to every class you tick, not only to history.

CAPABILITY	STATUS	WHAT IT DOES
by range		
An admin off-switch that holds	GA	Turning the embedded browser off removes it, enforced at the route itself and not only on the link that opens it, so it cannot be reached by typing the address or restoring a tab.
Passkeys	STAGED	The sign-in ceremony is wired end to end in both the in-shell browser and the extension, and approval happens on a Bolt-owned window rather than an in-page badge, because a page can draw a convincing badge. Registering a new passkey still refuses, and no ceremony has been run against a live site yet, so this is staged and not GA.
The one browser your org can leave open	STAGED	The operating-system network filter is built to allow AI destinations only from applications that mask at their own boundary, which is Bolt and a browser carrying the Bolt extension. Presence cannot be inferred from a bundle id, so the extension reports it and the server verifies the report. Enforcement is still observe-only; today it names, and does not yet block.

Honest scope: macOS and Windows. On Linux the in-shell browser is off, and that is a decision rather than a porting gap. The screening depends on a page-to-native channel that exists for WKWebView and WebView2; there is no WebKitGTK implementation yet, so on Linux every tab would be unscreened. Shipping a browser branded as governed that is structurally unable to govern is the wrong default, so it stays off until the transport exists. The Chrome extension covers Linux in the meantime.

07 Depth: it does not just recognize, it validates and protects

The local surface is not a handful of toys. It is deep, and much of it is security you get for free on every paste, offline.

CAPABILITY	STATUS	WHAT IT DOES
Offline knowledge catalog	GA	Tens of thousands of reference entries answer instantly with no network: currency, country and airport codes, time zones, HTTP status codes, ports, MIME types, ICD-10, Unix exit codes, chmod bits, ASCII, and more.
Security scan on every paste	GA	Local scan for hidden and dangerous characters (bidirectional and zero-width, the Trojan-Source class), homograph and confusable look-alike attacks, and leaked secrets (cloud keys, tokens, private keys), before anything is sent anywhere.
Identifier validation	GA	Validates pasted identifiers with real check-digit math, offline: Luhn, IBAN, ISO 7064, Verhoeff (Aadhaar), plus 17+ national validators (NHS, CPF/CNPJ, MyNumber, and more). Catches look-alikes that pass a naive regex.
Structural shape recognition	GA	Recognizes hundreds of value shapes on sight: git SHAs, UUIDs with mint time, semver, MAC and IP/CIDR, crypto addresses, CVE/GHSA/CVSS, DOI/arXiv/PMID, tickers, and roughly 50 regional government IDs.
PII masking at the model boundary	GA	Emails, phones, cards, secrets, and more are masked before any prompt reaches a model, and restored in the response, from an authoritative server-side catalog. Optional pluggable NER (name, address, org) for deeper detection.
Redact before you paste out	GA	A one-motion redact so sensitive values are stripped locally before you copy anything into an outside tool.
Screen-share-safe mode	GA	One toggle widens on-screen masking to an extended tier (email, phone, and more) so nothing identifying is painted while you screen-share or demo; secrets and marked-sensitive values stay hidden regardless. (Automatic meeting detection is staged; the manual toggle is live.)
Cross-surface search	GA	Gmail-grade operators (from:, to:, is:unread, has:attachment, before:/after:, assignee:, due:) parsed locally and applied

CAPABILITY	STATUS	WHAT IT DOES
operators		across email, calendar, tasks, files, people, and clips at once, not just one keyword-matched folder.

08 Private recall: memory that is not a screen recorder

Bolt keeps a private working memory of what you were doing, built from lightweight on-device signals rather than a screen recording. It is distilled into session cards and a local entity graph, encrypted, never uploaded, and anything that reaches a model is masked at the boundary first. This is the counter to screen-recording recall: signal, not pixels.

CAPABILITY	STATUS	WHAT IT DOES
Session cards and activity recall	GA	Distilled cards of what you were working on (the app in focus, files touched, what you searched), so you can pick up where you left off and answer "what was I doing".
On-device signal capture	GA	Memory is assembled from OS signals, not screenshots or a screen recorder. Far lighter to run and far less to leak. Raw signals age out on a short TTL.
Local entity graph	GA	People, projects, and things you touch are linked in a graph that stays on the device.
"Coming up" extraction	GA	Surfaces the likely next action from your recent context, so recall turns into a next step rather than a search.
Consent, pause, and forget	GA	Recall is opt-in and can be paused, scoped, or cleared at any time. Nothing is uploaded.
Encrypted at rest	GA	Memory lives in the same whole-database-encrypted local store as clipboard and notes.
Semantic search over activity	STAGED	Vector search across your activity uses an on-device embedder, wired on when the embedding model is enabled.
Visual, scrubbable timeline	ROADMAP	A visual timeline you can scrub is on the roadmap. Text-based recall is live today.

How this differs from screen-recording recall (Rewind/Limitless, Microsoft Recall, LittleBird): those capture the whole screen into a large personal index, often cloud-assisted. Bolt captures distilled signals on the device, keeps them local and encrypted, masks anything sent to a model, and wires the memory into governed action. Recall becomes the next step, not a movie to scrub.

09 Secure sharing, on your own storage

When you do need to share a note or an annotated screenshot, Bolt turns it into an expiring, revocable link, and the file lives on storage you control, not on Sparcle.

The bytes never touch Sparcle. Bolt encrypts every file on your device before it is stored, then mints an expiring capability link. Storage is your own cloud drive (Google Drive, OneDrive, or Dropbox, connected automatically) or, for teams, your organization's own S3 or MinIO bucket.

You keep the controls. Set an expiry (one hour to never), require a password, limit downloads, restrict to your organization, and revoke access at any time, with a per-open audit trail. The encryption keys and the revoke control stay with you.

Annotate first, locally. Mark up a screenshot or image with arrows, text, and redaction, then share the result.

> Individuals share through their own connected drive automatically; a central team deployment adds server-enforced org-only access, passwords, download limits, per-open audit, and instant revoke on top. Either way, nothing is stored on Sparcle infrastructure.

10 A governed terminal, and the agents you bring

Bolt is also a power surface, and the governance follows it there.

CAPABILITY	STATUS	WHAT IT DOES
Policy-governed terminal	GA	A real terminal and tmux inside the launcher. Every command that Bolt itself injects (a saved recipe, a tmux send, an agent action) crosses a native policy chokepoint where admins allow or deny by rule, and the whole capability can be turned off by policy for regulated orgs. What you type by hand is never governed. (The confirm-before-destructive dialog is staged; deny and catastrophic-command enforcement are live.)
Bring your own coding agent	GA	Run Devin, Claude Code, Codex, or Gemini locally on your workspace through the open Agent Client Protocol, under the same approvals and audit, so you are not locked to one vendor's agent. (Bolt hosts the agent locally; deeper server-side orchestration is staged.)
Password and secret vault	GA	Search across your connected vaults (1Password, Bitwarden, Apple Keychain, Windows Credentials) and copy a password, username, or TOTP in one keystroke, locally, or let it fill the login directly in Bolt's own browser. Results stream in as each vault answers, so one slow provider cannot stall the list. A vault being installed is not permission to read it: consent is explicit and per vault, and locking clears every cache that held it. Revealing a corporate or shared secret is policy-gated and audited; personal vaults stay ungoverned. Generates strong passwords on demand.

11 Agent Gateway: mask secrets before your agents reach the cloud

When your engineers run a third-party coding agent, every prompt can carry keys, tokens, and customer data straight into a vendor cloud. The Agent Gateway masks those values on the device before the request leaves, and restores the real values locally when the agent's

tools run. Unlike the network appliances the rest of the category ships, this sits client-side, on the agent's own wire.

CAPABILITY	STATUS	WHAT IT DOES
Secret and PII masking for agent CLIs	LIVE: 3 CLIS	Masks API keys, tokens, and structured PII in a coding agent's outbound traffic before it reaches the vendor. Three wire formats are handled today: Anthropic Messages (Claude Code), Gemini generateContent (Gemini CLI), and OpenAI Responses (Codex), each proven against a real network path.
Local restore inside tool calls	GA	Real values are restored on your machine when the agent's tool calls run, so the work still works while the vendor model only ever saw placeholders.
Client-side redirect, layered	GA	Traffic is redirected through Bolt via config and environment, so there is nothing new for you to route around by hand. Claude Code and the Gemini CLI take a base-URL environment variable; Codex is redirected by a line in its own config file, because the environment variable that several published guides recommend does not exist in the shipped Codex binary. We checked rather than repeated it.
Streaming responses	GA: CLAUDE, GEMINI	A streaming answer is relayed as it arrives and masked per completed block, rather than buffered to the end. Deltas are held until the block closes, because a secret split across two chunks matches no pattern in either half.
Streaming for Codex	STAGED	Codex streaming is the one gap, and it fails open: a request that asks for a stream is forwarded to OpenAI unmasked and logged as such, rather than being silently half-masked. Until this lands, run Codex through the gateway without streaming if the traffic is sensitive.

CAPABILITY	STATUS	WHAT IT DOES
OS network backstop	STAGED	An operating-system network filter is being brought from observe-only to enforcement, so the governed path becomes non-optional rather than advisory.
Admin policy UI and auto-provisioning	ROADMAP	Central policy for which CLIs may run and automatic provisioning of the redirect on a managed machine. Today the redirect is set per machine.

Honest framing: three CLIs are covered today, with Codex streaming the one named gap. An unmanaged admin on their own machine can defeat any on-device control, so the Gateway is strongest paired with managed policy and the network backstop. It is layers, not a single lock.

12 Productive without burning tokens

The "AI is too expensive" objection is answered by an engine, not a dashboard. Before the agent loop runs, a lightweight classifier decides whether the query even needs a model: knowledge, greeting, and meta questions skip the model and tool-loading entirely, and most everyday queries resolve on the instant local engines with no model call at all. When a model is genuinely needed, Bolt caches the stable prompt prefix for a large discount and enforces a token budget, and because you bring your own key there is zero token markup. The PII masking still runs on the classifier path, so cost discipline never costs you the privacy guarantee.

13 Healthcare, finance, and defense, without a code release

Regulation and vertical rules are not hard-coded. They ship as swappable, signed packs, so a new jurisdiction or industry is a data change, not a software update. This is how one product serves regulated verticals honestly.

28 signed packs ship in the box, including healthcare, finance, defense, legal, and logistics cores, plus 22 jurisdiction privacy packs: GDPR, UK GDPR, CCPA and US state acts (Indiana, Kentucky, Rhode Island), LGPD, DPDP (India), PIPL (China), APPI (Japan), PIPA (Korea), PDPA (Singapore and Thailand), PDP (Indonesia), Australia Privacy, PIPEDA and Quebec Law 25, POPIA, NDPA, KVKK, PDPL, revFADP, and more.

An enabled pack masks its identifiers before any LLM call and on every share, validates national IDs with real check-digit math so look-alikes are caught rather than just regex-matched, honors the regime's data-subject rights and residency, and lights up the safeguards dashboard.

Packs install offline as signed bundles. New regulation reaches an air-gapped deployment without a software update.

Cryptographic erasure and evidence export support GDPR and HIPAA right-to-be-forgotten obligations, with an offline-verifiable audit proof.

You can author your own pack in a short YAML manifest. Multi-jurisdiction is stacking packs.

> Honest scope: the packs deliver defensible safeguards, masking, and evidence. They are not a consent manager and do not by themselves issue a compliance certification; pair them with your DPO or counsel. Automated DSR request-intake is staged; the erasure cascade and evidence export are live today.

14 Your device and your SaaS, one surface

The same launcher that does the local work reaches outward. Your first-party context (local files, clipboard, notes, terminal, other machines you own, reachable by phone-to-desktop bridge and multi-machine routing) and your third-party systems (email, calendar, Slack, Jira, Salesforce, ServiceNow, SAP, and 100+ more via the open Model Context Protocol) sit together, so an agent can act across both under one set of rules. This is what removes the "two AI worlds" problem, where people keep a cloud assistant for work systems and separate local tools for their own machine.

15 No lock-in, and governance that outlives your model

Lock-in in enterprise AI usually happens at several layers. Bolt refuses every one.

Model layer. Bring your own LLM at every tier: OpenAI, Anthropic, Amazon Bedrock, Google Vertex, Ollama, NVIDIA NIM, LM Studio, or any OpenAI-compatible endpoint. Hot-swappable, zero token markup, you pay your provider directly.

Ecosystem layer. You are not tied to one vendor's suite. Bolt sits above Microsoft 365, Google Workspace, Zoho, Slack, Jira, Salesforce, ServiceNow, SAP, and more, and works with any OIDC identity. Copilot is welded to Microsoft and Gemini to Google; Bolt is welded to no one, so mixing or changing suites does not change Bolt.

Protocol layer. Integrations ride the open MCP standard, not a proprietary plugin marketplace. Bring your own MCP servers.

Deployment layer. Self-host anywhere, including air-gapped. No dependency on a Sparcle-operated cloud.

Agent layer. Bring any ACP coding agent (Devin, Claude Code, Codex, Gemini) and run it locally under the same approvals and audit. You are not locked to one vendor's agent any more than one vendor's model.

The differentiator most competitors cannot match: **your governance is defined once and survives the swap.** PII masking, human approvals, audit, and identity-aware access attach to the boundary, not to a particular model or vendor. Change the model, change the backend, even move from a frontier API to a local model in an air-gapped room, and the same masking, approvals, and audit chain apply. You do not re-run compliance because you changed a model. A cloud AI assistant cannot offer this, because its governance is its cloud.

---

16 Bring anything into Bolt

The honest boundaries below are narrow because Bolt is built to be extended, not replaced, when you need something it does not ship. The practical answer to most "does Bolt do X?" questions is: if X has an API or an MCP server, yes, by configuration, usually in minutes, and under your governance.

If it speaks MCP, it plugs in. Point Bolt at any MCP-compliant server and its tools become governed actions the agent can take.

If it has an API, you can wrap it. A short YAML utility manifest turns an external API or internal system into a first-class Bolt action.

If it is a data format, Bolt probably already knows it, and if not, you teach it in a few lines of the same manifest.

The shape of a user utility: a `recognize` block matches the input, a `transform` pipeline shapes and routes it, locally or to any MCP tool.

```
utility: acme-ticket
recognize:
  - pattern: "ACME-\\d+"          # matches ACME-1234 anywhere you type or paste
    read_as: "Acme ticket id"
transform:
  pipe:
    - mcp: acme.get_ticket         # call your own MCP server
    - pick: [id, title, status, owner]
    - table                        # render a clean result in the launcher
```

That is the whole idea: you are not waiting on our roadmap to reach your systems.

---

17 One app replaces a stack, and shrinks the attack surface

Because the daily surface is broad and local, one governed app absorbs what is usually a drawer of single-purpose tools:

Utilities and dev tools: clipboard manager, snippet expander, JSON/YAML/SQL formatter, regex tester, base64/hash/JWT tools, timezone and currency converters, UUID and password generators, QR maker, color tools.

Local knowledge and files: screenshot-OCR search, local file search, a notes and annotation app, and a reference catalog (ports, HTTP codes, ICD-10, and more).

Secure workflow: password-vault access, secure file sharing with expiry and revoke, image annotation, and a terminal.

Work systems and AI: enterprise search, a cross-system agent, and an AI chat client, all under one identity and one audit trail.

That is easily two dozen point tools folded into one surface you deploy and govern once, with fewer licenses, fewer vendors, and fewer places your data can leak.

On risk, we quantify honestly. Every point tool Bolt retires and every task it keeps local is attack surface removed and one less place data can leak. The size of that problem is real and third-party documented (see the evidence below), but we will not quote a made-up risk-reduction percentage. The honest, concrete number is the consolidation itself: one governed surface instead of two dozen ungoverned ones.

18 The problem we remove, and why it is expensive

The offline surface matters because the best documented AI risk in the enterprise is people pasting sensitive data into public chatbots because it is fast. Every task Bolt absorbs locally is a leak that never happens. The pain is documented by independent researchers (figures below are attributed to their publishers, not our own measurement):

FINDING	SOURCE
Employees routinely paste into public GenAI tools, much of it through unmanaged personal accounts, many times a day, and a meaningful share of those pastes contain PII or regulated data.	LayerX Enterprise AI/SaaS Security Report; Verizon DBIR coverage
A significant share of files uploaded to GenAI and of text pasted into it contains PII or payment data.	Cloud Security Alliance; LayerX
Shadow AI added roughly $670,000 to the average breach, and most breached organizations had no AI governance policy.	IBM Cost of a Data Breach 2025
One in four malicious breaches is now AI-enabled, a 56 percent rise in a year, at about $6 million against a $4.99 million global average. More than one organization in five reported a breach that targeted an AI model or application, most often through a compromised API, application or plug-in, or a misconfigured cloud AI workload.	IBM Cost of a Data Breach 2026
Shadow AI is now the third most common non-malicious insider action, a fourfold rise in a year, measured across 858,440 data-loss events aimed at generative AI tools. Source code is the leading kind of data sent to an external model.	Verizon 2026 DBIR, shadow AI section, on LayerX browser telemetry
45 percent of enterprise employees now use AI on a corporate device, up from 15 percent a year earlier, and 67 percent of that use runs through personal accounts an enterprise control never sees.	LayerX, cited in the Verizon 2026 DBIR
One enterprise user in five runs at least one AI-enabled browser extension, and 73 percent of those extensions hold high or critical permission scope.	LayerX Browser Extension Security Report 2026
74 percent of enterprise leaders believe they could pass an AI compliance audit; 27 percent have a fully mature AI governance programme, 44 percent have documented AI incident response, and 29 percent feel ready for the EU AI Act. Only 36 percent of boards regularly discuss third-party AI risk.	Schellman, State of AI Governance 2026 (500+ US enterprise leaders)

FINDING	SOURCE
The share of breaches that involved a third party went from 30 percent to 48 percent in a single year, across more than 22,000 confirmed breaches. Every vendor in your data path is one you cannot patch, monitor, or air-gap yourself.	Verizon 2026 Data Breach Investigations Report
The EU AI Act has been enforceable since 2 August 2026 for transparency obligations, general-purpose-model supervision, and the penalty regime, which reaches 35 million euros or 7 percent of global turnover for prohibited practices and 15 million euros or 3 percent for most other breaches. The high-risk obligations themselves were deferred by the Digital Omnibus: stand-alone Annex III systems to 2 December 2027, AI embedded in regulated products to 2 August 2028. A deferral is not a reprieve, because the evidence a high-risk filing needs is accumulated over the intervening period, not assembled at the deadline.	EU Regulation 2024/1689 as amended by the 2026 Digital Omnibus; European Commission timetable

Bolt is the structural antidote: the fast, sanctioned, local alternative for the everyday reformat, decode, compute, validate, and redact tasks, so the reason to paste into an outside tool goes away, while masking, approvals, and audit govern the times a model genuinely is needed. That is data locality delivered as daily productivity, not as a policy nobody follows.

The developer path is the one nobody has measured. The citations above describe the browser: pasting, uploading, personal accounts. That is the path researchers can observe. The path that is harder to observe, and increasingly the larger one, is an AI coding assistant or a terminal agent running with an engineer's own credentials, sending source, configuration, tokens and customer records straight to a model vendor from the laptop. Because that traffic does not pass through a gateway, it rarely appears in anyone's statistics at all. We are not going to invent a number for it. It is what Bolt's governed terminal and agent gateway exist for, and it is measurable on one machine in an afternoon.

Every figure above is somebody else's average. The useful number is yours. The free desktop build runs on one laptop, with no server, no cluster, no Docker, no Helm, and nothing that needs an NDA. In a day it will show what actually left your boundary through the AI tools already in use, and leave a record of it. **If the answer turns out to be nothing, that is a useful finding too, and we will say so.** A wider pilot is the second step, and it is far easier to justify once the first one has produced a number.

19 What you need to run it, and what it costs

Setting expectations plainly, because a governed tool you cannot budget for or deploy is no use.

Two ways to run it

On your laptop, free. The full desktop build runs on your own machine with your own API key. No server to stand up, no external database (it ships an embedded one), and no data reaches Sparcle. This is the individual and evaluation path.

Self-hosted, for teams. Deploy inside your own perimeter: Docker Compose for staging, Kubernetes for production high availability, on-prem, or fully air-gapped. It runs in your VPC or data center, not ours.

What you provide

Your infrastructure (self-host only). A container host or Kubernetes cluster sized to your seats: in practice, standard mid-size VMs plus storage, which fits inside an existing cloud or on-prem budget. The free desktop tier needs none of this.

Your LLM. Bring your own key (OpenAI, Anthropic, Bedrock, Vertex, and more) or your own served or local model. Bolt adds zero token markup, so your model spend is exactly your provider's price, and a local model costs nothing per token.

Do I need GPUs, a cluster, or a vector database?

No, on all three, and this is a common misread. **Bolt does not run models,** it brings yours, so it needs no GPU. The free desktop build needs no server at all. A team deployment is standard VMs or a small Kubernetes namespace sized to your seats, inside the cloud or data center you already run, and Bolt reaches your existing SaaS through their own APIs. You are not standing up a new data plane; you are governing the one where your apps already live.

Aeira, the enterprise search plane, also needs no GPU to start. Its baseline is keyword search on a plain Postgres, CPU-only and air-gappable, with no vector database. Semantic ranking is an optional add-on: a vector store (pgvector or Qdrant) plus an embedding backend you point it at, which can itself run on CPU. A GPU only raises that backend's throughput; it is never required. The keyword engine, vector store, and embedder each sit behind a swappable interface, so Aeira degrades gracefully: with no embedder you get keyword results, and if an embedder is configured but unavailable it falls back to keyword mid-query rather than failing. You add semantic ranking, and the hardware for it, only when the value is worth it.

What it costs

Three visible parts: the per-seat license, your own infrastructure, and your own token spend. The desktop build is free (bring your own key). Paid self-hosted tiers start at $30 per seat per month with a 10-seat minimum, and founding customers get 25 percent off for 24 months. Because you bring the model there is no vendor token margin, which is why a Bolt deployment typically undercuts stacking a cloud assistant plus an enterprise search tool plus an automation tool. See pricing for the current tiers, the TCO one-pager (PDF) for a modeled 1,000-seat scenario, or ask for a tenant-specific model in an architecture review.

Who it is for

Anyone can run the free build on a laptop to evaluate with real data. The paid tiers suit teams that want central deployment and governance, and Bolt has the lowest seat minimum in its category (10 seats). The sharpest fit is regulated and security-sensitive organizations, but that is a fit, not a gate. If your need is one of the honest non-fits below, we say so.

---

20 "We already built something internally"

Usually true, and usually about a different problem. This is the most common reason an evaluation stops, so here is the honest version of it.

Internal AI programmes tend to produce three things, and they work: a gateway or proxy that routes model calls, meters spend and logs that a call happened; a retrieval assistant over a wiki, a ticket system or a chat archive; and managed Copilot or Cursor licences behind SSO. **Bolt is not a better version of any of those, and it does not ask you to replace them.**

What an internal build almost never produces, because nobody builds it until an auditor asks for it, is **evidence**: a record of what content actually reached a model, bound to who acted and under what authority, which can be verified by someone who does not trust the system that produced it.

So the question worth asking internally is not whose tool is better. It is **what did you build it to produce, function or evidence?** If the honest answer is function, the gap is narrow, and closing it displaces nothing that already works.

There is a second gap worth checking, and it is a layer difference rather than a quality one. An internal gateway governs the calls that pass through it. It does not see an AI coding assistant calling a model vendor straight from a laptop, a browser tab signed into a consumer AI account, a local MCP server making its own outbound calls, or a terminal agent running on a developer's own credentials. None of that traffic traverses the gateway. Two questions settle whether that gap is real in your environment:

> Does your gateway see what goes to an AI coding assistant from a terminal, or only what goes through the gateway?
>
> When something does pass through it, do you keep the content that reached the model, or the fact that a call happened?

---

21 When Bolt is not the right tool

Credibility depends on saying where we stop. Bolt is not, and does not pretend to be:

NOT THIS	WHAT TO USE INSTEAD
Inline network security (secure web gateway, forward proxy, CASB, firewall)	Keep your proxy. Bolt does not sit inline on your network and does not replace it. The reason both are needed is a layer difference rather than a quality one: **a proxy governs the path it owns, and much of the AI exposure sits on paths it never sees.** An AI coding assistant calling the model vendor straight from a laptop. A browser tab on a consumer AI account. A local MCP server making its own outbound calls. A terminal agent running on a developer's own credentials. None of that traverses the gateway. Bolt governs at the endpoint, where the content is assembled and sent.
Endpoint security, antivirus, or EDR	Run your EDR alongside Bolt. Bolt is not a threat-detection or malware-defense agent.
GPU or MLOps infrastructure (scheduling, fine-tuning, model serving)	Pair Bolt with your serving stack (for example NVIDIA NIM or vLLM). Bolt brings your model; it does not run the model farm.
In-document Office co-authoring	If your entire AI value is inside Word or Excel and cloud AI is acceptable, a native Office assistant fits that specific need. Bolt is an overlay plus connectors, not an AI in the ribbon.
A system of record	Bolt acts across your systems; it does not replace your database, CRM, or ticketing system.

Where we are early rather than unsuitable, we say so: public reference logos at Fortune-500 scale and multi-year connector-maintenance SLAs are things a pre-scale company earns over time. That is real, and it is a reason some conservative buyers wait. It is not a capability gap. Note the shape of the list above: the genuine non-fits are inline network security, endpoint security, ML infrastructure, in-document Office authoring, and being a system of record. Almost everything else people ask for is reachable by configuration.

22 Full FAQ

Does Bolt work offline, and is it really useful before the AI fires?

Yes. Bolt opens as a desktop launcher and does 90 instant utilities (JSON, base64, hashes, JWT, regex, time zones, currency, and more), clipboard history, notes, screenshot OCR, and local file search with no model call and no network. The launcher earns its keep before any AI cost starts. You can verify by turning off Wi-Fi and watching local tasks keep working.

Does Bolt have clipboard history?

Yes. Bolt captures copied text, images, and files into a searchable, pinnable clipboard history, stored encrypted on your device.

Can Bolt search my screenshots and local files?

Yes. On-device OCR makes the text inside screenshots and images searchable, and an opt-in on-device image model adds visual search. Bolt also indexes and searches your own file locations (for example Home, Desktop, Documents, Downloads) by name and content, with exclusions you control.

Does Bolt keep notes, and can I annotate images?

Yes to both. Bolt has device-local rich notes (formatting, lists, checklists, tables, links) with full-text search, encrypted at rest. You can also mark up screenshots and images with arrows, text, and redaction, then keep or securely share the result.

Can I share files and notes securely, and where are they stored?

Yes. Bolt turns a note or an annotated screenshot into an expiring, revocable link. Every file is encrypted on your device before it is stored, and the storage is yours: your own cloud drive (Google Drive, OneDrive, Dropbox) or, for teams, your organization's own S3 or MinIO bucket. The bytes never touch Sparcle. You control expiry (one hour to never), password, download limits, org-only access, and instant revoke, with a per-open audit trail.

Does Bolt have a terminal, and is it governed?

Yes. There is a real terminal and tmux in the launcher. Commands you type by hand are never governed, but any command Bolt itself injects (a saved recipe, a tmux send, an agent action) crosses a native policy chokepoint where admins allow or deny by rule, and the whole capability can be disabled by policy for regulated orgs. Deny and catastrophic-command enforcement are live today; a confirm-before-destructive dialog is staged.

Can I run coding agents like Devin or Claude Code inside Bolt?

Yes. Through the open Agent Client Protocol you can run Devin, Claude Code, Codex, or Gemini locally on your workspace, under the same approvals and audit, so you are not locked to one vendor's agent. Bolt hosts the agent locally; deeper server-side orchestration is a staged follow-up.

Is running AI in Bolt expensive?

It is designed not to be. A lightweight classifier runs before the agent loop and skips the model and tool-loading entirely for knowledge, greeting, and meta questions, and most everyday queries resolve on instant local engines with no model call at all. When a model is needed, Bolt caches the stable prompt prefix for a large discount and enforces a token budget, and because you bring your own key there is zero token markup. PII masking still runs on that fast path.

Can Bolt use my password vault?

Yes. Bolt searches across your connected vaults (1Password, Bitwarden, Apple Keychain, Windows Credentials) and copies a password, username, or TOTP in one keystroke, locally. Revealing a corporate or shared secret is policy-gated and audited; personal vaults stay ungoverned. Bolt can also generate strong passwords on demand.

Can Bolt use my screen, selection, or current tab as context?

Yes. A Chrome extension brings the current tab and selection to Bolt, and the launcher can pull your screen, selection, open tab, clipboard, or an on-device OCR of what is on screen into a prompt as first-class context. The sensitive ones are masked at the boundary before any model sees them.

How many tools can Bolt replace?

Easily two dozen categories: clipboard manager, snippet expander, JSON/regex/hash/JWT tools, timezone and currency converters, UUID and password generators, QR and color tools, screenshot-OCR search, local file search, a notes and annotation app, a password-vault front end, a secure file-sharing tool, a terminal, enterprise search, a cross-system agent, and an AI chat client, all under one identity and audit. Fewer licenses, fewer vendors, and fewer places your data can leak.

Is there a screen-share-safe mode for demos and meetings?

Yes. One toggle widens on-screen masking to an extended tier (email, phone, and more) so nothing identifying is painted while you screen-share or present; secrets and marked-sensitive values stay hidden regardless. Automatic meeting detection is a staged follow-up; the manual toggle is live.

Does Bolt have voice search or voice input?

Yes, and it is fully on-device. You can speak to the launcher; the microphone is captured and transcribed locally with Whisper, so audio never leaves the machine, unlike cloud dictation. Bolt adopts a local Whisper model if one is present, otherwise offers a one-time download. Voice input is currently an opt-in Labs experiment in builds that include it, not on by default.

Is my local data encrypted at rest?

Yes. Clipboard history, notes, and screenshot indexes are stored in a whole-database-encrypted local store, with the encryption key derived on-device and file-permission locked. Connector and OAuth tokens are held in a separate AES-256-GCM vault. Nothing local sits in plaintext.

How deep is the offline surface, really?

Deep. Beyond the utilities, Bolt carries a large offline knowledge catalog (currency, country and airport codes, time zones, HTTP status, ports, MIME types, ICD-10, exit codes, chmod, ASCII, and more), recognizes hundreds of value shapes on sight, and validates identifiers with real check-digit math offline (Luhn, IBAN, ISO 7064, Verhoeff for Aadhaar, plus 17+ national validators). Every paste is also scanned locally for hidden and dangerous characters, homograph attacks, and leaked secrets. None of this needs a network.

Does Bolt protect against Trojan-Source and homograph attacks?

Yes, offline, on every paste. Bolt flags hidden bidirectional and zero-width characters (the Trojan-Source class), homograph and confusable look-alikes, and leaked secrets such as cloud keys and private keys, before the content goes anywhere.

How does Bolt serve regulated industries like healthcare, finance, and defense?

Through swappable, signed compliance and vertical packs. 28 ship in the box, including healthcare, finance, defense, legal, and logistics cores plus 22 jurisdiction privacy packs (GDPR, UK GDPR, CCPA and US state acts, LGPD, DPDP, PIPL, APPI, PIPA, PDPA, POPIA, and more). An enabled pack masks its identifiers before any LLM call and on every share, validates national IDs with real check-digit math, and honors the regime's rights and residency. Packs install offline as signed bundles, so new regulation reaches an air-gapped deployment without a software update, and you can author your own pack in a short YAML manifest.

Can I add a new country or regulation myself, air-gapped?

Yes. A regulation pack is a short signed YAML manifest (jurisdiction, residency, retention, rights, safeguards, recognizers). You, a partner, or your counsel can author and sign one, and it installs offline with a review-and-consent step. Multi-jurisdiction is stacking packs.

Is Bolt a SaaS? Should I evaluate it like one?

No, and this trips up a lot of security reviews. Bolt is self-hosted software that runs inside your own perimeter; Sparcle operates no cloud that receives your data and has no hosted ingest path. So you do not evaluate a Sparcle-run production environment the way you would a SaaS vendor, because there is not one handling your data. What you evaluate is the software and the architecture, running under your own controls. That is a different, and for regulated buyers usually stronger, security posture.

Do you have SOC 2 Type II?

Short answer: your data never touches our servers, so the core of what a SaaS SOC 2 certifies, how a vendor guards your data inside its cloud, does not exist here. Bolt is self-hosted; there is no Sparcle-run environment holding your data to audit.

Better still, the real substance of SOC 2 Type II is evidence that controls operate effectively over a period of time. Because you run Bolt yourself and it emits a tamper-evident, independently verifiable audit trail, you and your auditor can verify the deployed controls and their operation continuously, in your own environment, over time, instead of waiting on a vendor's annual attestation window.

What a company-level SOC 2 still speaks to is our own engineering and security practices. On that: we hold neither SOC 2 Type I nor Type II today and we are not going to put a date on it that we have not bought; a readiness package and gap analysis are available now; and our Trust Center maps SOC 2, HIPAA, GDPR, and ISO 27001 criteria to concrete evidence. If your procurement hard-requires a completed vendor SOC 2 Type II report today, that is a fair reason to wait.

What exactly does Bolt's DLP cover, and where does it stop?

It covers the request path into the model. Two layers scan for sensitive values before the LLM adapter is invoked: an HTTP middleware over JSON request bodies on protected routes, and the agent loop itself, which masks every prompt, memory and tool result it assembles. The agent loop is the load-bearing one, because it is where the text that actually reaches a model is built, and it is active when a detector is configured; with no detector wired there is nothing to mask with. Detected values are replaced with opaque tokens and the mapping is rehydrated on the way out, so the model sees tokens and your user sees real values. It sits above a swappable provider layer by design, so changing model or backend does not switch the policy off, and tool results are re-masked before they go back into the agent loop.

Where it stops today: remembered values now route through the same chokepoint on their way into the system prompt, so the model sees a token, but the underlying text is still persisted unmasked at rest and the retrieval hints stored beside it are plaintext by design. At-rest protection there is your storage and key posture, not the masking layer. Separately, name detection in free text is a bring-your-own sidecar with no endpoint configured by default, so recall on names depends on you wiring it; pattern and check-digit detection for structured identifiers does not. If your policy treats stored-at-rest identifiers as in scope, encrypt that store and hold the keys, or keep user memory off until you have.

Is the audit trail complete? Can I prove nothing is missing?

Complete for what it covers, and it is worth being precise about what that is. Aeira search seals an event for every query it serves, on all three exit paths (denied, router-allowed, keyword-allowed), and it cannot fail open: the same Postgres that wires the audit sink is required to run a search at all, so the failure mode is an error, not a silent unrecorded answer. Tool calls, LLM calls, agent-run state transitions, human approvals, masking and authorization decisions, admin policy changes, the security-relevant identity transitions (sign-in success and failure, sign-out, session delegation, delegated-session revocation, and session attach), and the privileged key and identity operations (KMS rotation, crypto-shred, PII reveal, SCIM token issue and revoke) are in the chain too, sealed into an Ed25519-signed Merkle structure that a standalone verifier checks offline with nothing from Sparcle in the loop. Two of those refuse rather than proceed unevidenced: a crypto-shred is declined if the disposal receipt could not be sealed, and a PII reveal is declined if auditing has been switched off, so neither can happen into a void.

What is not in the chain today: the audit-seal operation itself still writes a log line rather than a sealed record. Two identity events are left out on purpose and pinned that way by tests, OAuth access-token refresh and the read-only SCIM token list, because neither grants anyone new authority and sealing them would bury the transitions that do. Aeira's catalog browse routes and Bolt's own connector-side search endpoints return results without emitting an event. Routing both through the same emit chokepoint is in flight. On proof: the verifier confirms that what was recorded has not been altered. It cannot confirm that everything that happened was recorded, and no chain construction can. Detecting a missing event means comparing the chain against a complementary signal, typically your SIEM or your own access logs. We say the same thing on the Trust Center rather than only when a reviewer asks.

Can I push a policy change to every installed Bolt from one console?

Not as a push, not today, and I would rather say so than let you plan a rollout around it. Policy itself lives server-side: the desktop client fetches its capability set from the server, so a central policy change does change what a user can do the next time their client asks. That part is real, and it is how terminal execution and the vertical packs are gated now.

What now exists is device policy, read from the tools you already run. On macOS the desktop app reads the configuration profile your MDM pushes (Jamf and equivalents), and the browser extension reads managed policy on Chrome, Edge, Brave, Firefox and Safari, nineteen keys covering the DLP guards, the risky-site list, page capture and the deployment URL. Device policy, your server policy and the built-in floor are reconciled by a single merge rule with a documented precedence, so an MDM setting and a server setting cannot quietly disagree. Windows Group Policy for the desktop app is not built yet; the seam it plugs into is.

What still does not exist is device fleet management. There is no enrollment, no device registry, no push channel, no targeting by group or platform, and no delivery confirmation telling you which machines picked a change up. Distributing the app itself goes through whatever you already run (Intune, Jamf, your own packaging). If your requirement is to change a policy and see confirmed application across thousands of endpoints inside an hour, Bolt does not do that yet and you should score it accordingly.

What is the access-control model, and how much of it is on by default?

Layered, and most of it is on out of the box. Session middleware, admin-only middleware, and OAuth scope checks gate the HTTP surface. The agent loop runs a per-tool capability check before it invokes anything. Aeira applies the per-tenant ACL as a SQL pre-filter on search, with no bypass flag, so rows a user cannot see never enter the candidate set. Those are default-on wherever there is a database. The capability list itself is compiled into the server rather than declared in configuration, which is deliberate: no manifest or YAML edit can invent a capability or widen an existing one.

The attributes that access decisions are made on now come from your identity provider rather than being inferred: country, office, role band, and employment type are read from real claims per provider (Entra, Okta, Google Workspace, Workday, or generic OIDC), and a session cannot name its own access dimensions in a query string. One honest scope note: on the search path the ACL is compiled into the query, so rows you cannot see never enter the candidate set, while the entity catalog still applies its ACL in the application layer after the rows leave the database. Both are enforced; only the search path is enforced by the database itself.

What is off by default is the optional policy decision point, the rule-based layer for finer-grained capability policy. It ships in the box, no shipped configuration file enables it, and turning it on is a deliberate deployment choice we walk through during a pilot. So RBAC enforced at multiple layers is accurate; every governance feature running the moment you install is not.

How much work is it to get Bolt's audit events into our SIEM?

Between an afternoon and a sprint, depending on the route. Route one is the one most teams already have: set LOG_FORMAT=json and Bolt emits structured events to stdout, where Vector, Fluent Bit, an OpenTelemetry Collector, or a Datadog Agent picks them up like any other workload. If you already ship container logs, this is configuration rather than integration. Route two is a forwarder built into bolt-api that authenticates directly to Splunk HEC, Microsoft Sentinel through the HMAC-signed collector API, Microsoft Sentinel through a Data Collection Rule with AAD OAuth2 and automatic token refresh, or syslog over TCP.

Two honest caveats on route two: it is off by default and no shipped values file enables it, so treat it as a deployment step you plan for; and it sends best-effort with no retry buffer, so a transient SIEM outage drops events. If you need guaranteed delivery, use route one with a collector in front, or run both. Either way there is no Sparcle-side ingest to configure, because there is no Sparcle cloud in the path.

Is my data sent to Sparcle?

No. Sparcle operates no cloud that receives your data and has no hosted ingest path. Local device data (clipboard, notes, screenshot indexes) is stored encrypted on your machine. When you connect a third-party system such as Gmail or Jira, Bolt talks to that system's API as you configured it; that is action on your systems, not exfiltration to a vendor. PII is masked before any prompt reaches a model.

You do not have to take this as a promise. On the desktop build, turn off Wi-Fi and local tasks keep working; in a self-hosted deployment everything runs in your VPC, so your own egress logs are the proof, and you can deploy Bolt in a subnet with no internet route so your logs show it cannot reach us at all. This matters because every vendor in your data path is one more place data can be lost, and third-party involvement in breaches doubled to 30 percent in Verizon's 2025 report; the one path you never have to defend is the one that does not exist. See Where is my data for the full picture.

Can I use my own LLM, and can I switch models later?

Yes at every tier. Point Bolt at OpenAI, Anthropic, Bedrock, Vertex, Ollama, NVIDIA NIM, LM Studio, or any OpenAI-compatible endpoint. Models are hot-swappable with zero token markup; you pay your provider directly. Bolt also auto-detects and adopts a local OpenAI-compatible runtime if one is already running on your machine.

If I switch models or backends, do I lose governance or have to re-run compliance?

No. PII masking, human approvals, audit, and identity-aware access attach to the boundary, above a swappable provider layer. Change the model, change the backend, or move to an air-gapped local model, and the same governance applies. You do not re-run compliance because you changed a model.

Do you support [our system]?

Almost certainly, by configuration. If it speaks MCP, point Bolt at the server and its tools become governed actions. If it has an API, a short YAML utility manifest turns it into a first-class Bolt action. Bolt ships 100+ integrations today (105 in-house connectors, roughly 350 more via Airbyte, plus any MCP-compliant server), and you can add your own without waiting on our roadmap.

How hard is it to add a new integration or format?

A few lines of YAML. A utility manifest has a recognize block that matches the input and a transform pipeline that shapes and routes it, locally or to any MCP tool. No fork, no vendor ticket.

Is there vendor lock-in?

No, at several layers. Model: bring and swap any LLM. Ecosystem: not tied to one vendor's suite, Bolt works across Microsoft 365, Google Workspace, Zoho, Slack, Jira, Salesforce, ServiceNow, SAP, and more via open MCP, with any OIDC identity. Protocol: open MCP, not a proprietary plugin marketplace. Deployment: self-host anywhere including air-gapped, with no dependency on a Sparcle cloud. Agent: bring any ACP coding agent.

Am I locked into Microsoft or Google if I use Bolt?

No. That is the point. Copilot is welded to Microsoft 365 and Gemini to Google Workspace; Bolt sits above whatever suite you run and connects to Microsoft 365, Google Workspace, Zoho, Slack, Jira, Salesforce, ServiceNow, SAP, and more through the open Model Context Protocol, with any OIDC identity. If you run a mix of suites, or change one later, Bolt does not change. You get the cross-ecosystem reach a single-vendor assistant cannot offer.

Is Bolt just enterprise search? Do I have to pay for parts I will not use?

No. Bolt is a daily-driver workspace that people open for local utilities, clipboard, notes, and search from day one, then use for governed action across connected systems. Aeira is the separate ACL-aware document-retrieval plane; you can run Bolt without it, or add it when you need governed enterprise search.

Is Bolt a network proxy, firewall, CASB, or antivirus?

No. Bolt is not an inline network device and not endpoint security. It governs what your workforce does with AI. Keep your proxy and your EDR; Bolt complements them.

Does Bolt run, fine-tune, or serve models on GPUs?

No. Bolt is the governed workspace, not ML infrastructure. You bring your model, and you or your serving stack (for example NVIDIA NIM or vLLM) run it. If your primary need is a model farm, pair Bolt with that infrastructure.

Aeira, the search plane, is the same story: it needs no GPU to start (keyword search on a plain Postgres, CPU-only and air-gappable), and semantic ranking is an optional add-on served by an embedding backend you supply, which can also run on CPU. See the dedicated Aeira answer below.

How does Aeira search work, and does it need a GPU or a vector database?

Aeira does hybrid retrieval: keyword search (BM25 and full-text) is the baseline, an optional vector lane adds semantic ranking, and the two are fused with reciprocal rank fusion, all behind an ACL pre-filter so every result is identity-bound. The baseline runs keyword-only on a plain Postgres, CPU-only and air-gappable, with no GPU and no vector database. Semantic ranking is an optional add-on: a vector store (pgvector or Qdrant) plus an embedding backend you supply, which can run on CPU, so a GPU only raises throughput and is never required. The keyword engine, vector store, and embedder each sit behind a swappable interface, so if no embedder is configured you get keyword results, and if one is configured but unavailable Aeira falls back to keyword mid-query rather than failing.

Is local model inference fully on-device today?

Partly. Bolt auto-detects and adopts a local OpenAI-compatible runtime and masks PII unconditionally at the boundary; that is generally available. The fully on-device generation loop is a committed next slice, and we do not claim it is finished.

Does Bolt work inside Word and Excel like an Office assistant?

No. Bolt is an overlay plus connectors, not an AI embedded in the Office ribbon. If your entire AI value is in-document and cloud AI is acceptable, a native Office assistant fits that specific need. Bolt fits when you need sovereignty, model choice, cross-system action, and a governed local surface.

Can I deploy air-gapped?

Yes. Docker Compose for staging, Kubernetes for production HA, on-prem, or fully air-gapped for ITAR and classified environments. License validation uses an offline-signed token, with no outbound calls.

What does it cost, and is there a free tier?

The full desktop build is free on your own machine with your own API key. Paid self-hosted tiers start at $30/seat/month (10-seat minimum). Founding customers get 25% off any tier, locked for 24 months, through the end of 2026 or until 50 customers sign. See the pricing page for current tiers.

Do I need my own servers to run Bolt?

Not for the free desktop build, which runs entirely on your laptop with an embedded database and no external services. For a team deployment you self-host inside your own perimeter (Docker Compose for staging, Kubernetes for production HA, on-prem, or air-gapped), so you provide a container host or cluster sized to your seats. In practice that is standard mid-size VMs plus storage, inside your existing cloud or on-prem budget.

Do I need to bring my own LLM, and does that cost extra?

Yes, you bring the model, and that is the point: Bolt adds zero token markup. Use your own key (OpenAI, Anthropic, Bedrock, Vertex, and more) or your own served or local model. Your model spend is exactly your provider's price, and a local model costs nothing per token. There is no vendor inference margin on top.

What is the total cost of ownership?

Three visible parts: the per-seat license, your own infrastructure, and your own token spend. The desktop build is free; paid tiers start at $30 per seat per month. Because there is no token markup and one governed app replaces a stack of point tools, a Bolt deployment typically undercuts stacking a cloud assistant plus an enterprise search tool plus an automation tool. Ask for a modeled scenario in an architecture review, or see the pricing page.

Who is Bolt for? Is it only for large regulated companies?

Anyone can run the free build on a laptop to evaluate with real data. The paid tiers suit teams that want central deployment and governance, and Bolt has the lowest seat minimum in its category (10 seats). Regulated and security-sensitive organizations are the sharpest fit, but that is a fit, not a gate.

Do you have Fortune-500 reference logos today?

Public reference logos at that scale are something a pre-scale company earns over time, and many of our target buyers are in environments where references are contractually restricted. If your procurement requires named public references before any purchase, that is a fair reason to wait. It is a stage limitation, not a capability gap.

Does Bolt have its own browser, and what does it actually check?

Yes, on macOS and Windows. It is a browsing surface inside the app, and the point of it is what happens on the way out: what you type or paste into a page is screened on your machine first. When something is found, Bolt names what it found grouped by type with a count, without repeating the values back, holds the page while you decide, and requires a written justification of at least ten characters to proceed, which lands in the audit record as a durable incident rather than a bare counter. The tab only claims "Guarded" once the in-page guard has reported in, and says "Not screened" where it cannot screen. Ads and trackers are blocked by default, banking, health and sign-in domains are on a capture deny-list, private tabs record nothing including in the diagnostic logs, and an admin can remove the feature entirely. On Linux it is off by design: the page-to-native screening channel exists for WKWebView and WebView2 only, and a governed browser that cannot govern is worse than none. Use the Chrome extension there.

Does Bolt support passkeys?

Staged, and we will not call it shipped yet. The sign-in ceremony is wired end to end in both Bolt's own browser and the Chrome extension, and approval happens on a Bolt-owned window rather than an in-page badge, because a page can draw a convincing badge and a real click on a fake one is still a real click. What is not done: registering a new passkey refuses by name, and no ceremony has yet been run against a live site. Every failure path falls back to your platform authenticator, so a broken Bolt costs you a Bolt sign-in and never the ability to log in.

How do I verify the data-locality claims myself?

Install the free desktop build, turn off Wi-Fi, and confirm local utilities, clipboard, notes, and file search keep working. For the enterprise deployment, everything runs inside your perimeter, so your own network monitoring is the proof. Request a pilot or an architecture review to run this end to end.

Every claim on this page reflects shipped behavior as of 19 September 2026, with staged items marked. For verification, request a pilot or an architecture review at sparcle.app. Contact: bolt@sparcle.app.