RESOURCE · ARCHITECTURE

Healthcare security & privacy architecture brief

Also available as a printable PDF (/resources/ai-leak-guard-security-brief.pdf) for compliance review, 3 to 5 pages.

> **IN ONE PARAGRAPH**
>
> AI Leak Guard is a browser extension for healthcare staff. It reads text you paste into supported AI tools, files you attach to them, and (on ChatGPT, Claude, and Gemini) the message at the moment you send it, on your device, and warns you before patient identifiers are sent. It does not modify your files, it does not block, and it does not phone home. The user makes the final call.

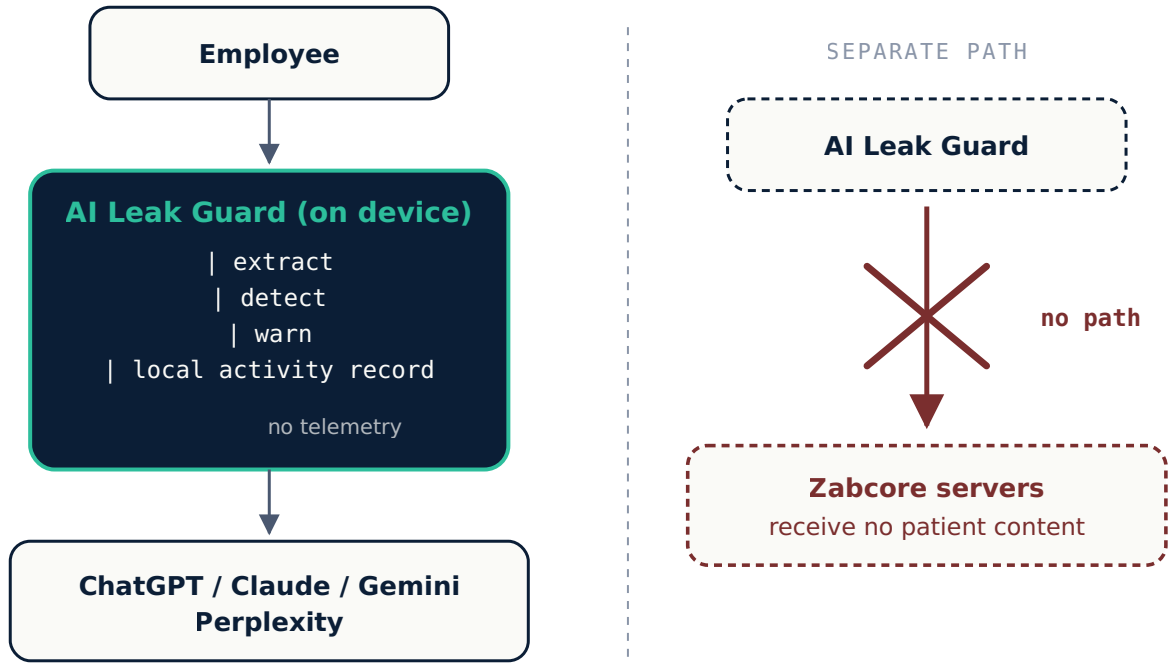
Architecture

- **On-device extraction.** When you attach a supported file type (PDF, Word, Excel, PowerPoint, text), the extension extracts text from the file in the browser, before the upload proceeds. Files it cannot inspect as text (scanned images, encrypted PDFs) are reported as "couldn't inspect", never as safe.
- **Send-time protection (v1.3).** On ChatGPT, Claude, and Gemini, the extension reads the message you are about to send at the moment you send it, so typed or dictated content is caught even if you never pasted it. Your draft is always kept. If the check ever fails, the message still sends: fail-open, never blocks.
- **On-device detection.** Pattern and context based detection against a bundled ruleset, executed in the browser. No detection request leaves the device.
- **Warn, not block.** When identifiers are found, a warning surfaces with the categories and counts. The user chooses to send anyway or cancel. The extension does not silently rewrite input.
- **Local activity log.** Metadata-only, on the device, in browser extension storage.
- **Local export.** CSV or JSON, built and saved in the browser. No new permission required.

Network behaviour

`Zero outbound network requests from the extension.` The extension does not contact any server, send telemetry, or fetch anything at runtime. Detection patterns are bundled inside the extension and change only when the extension itself updates through the Chrome Web Store.

Separately, zabcore.com collects the contact information you choose to submit through a form. The two systems share no identifier, and no detection data ever reaches the website.



Detection and document extraction run entirely on the device. Zabcore never receives patient content.

Permissions

The extension declares `storage` only, plus content-script matches for the supported hostnames. It does not request `tabs`, `history`, `cookies`, `webRequest`, or any host access outside the sites in the coverage matrix. **No new permissions were added in V1.2.** This is verifiable from the Chrome Web Store listing.

Activity log

The log lives in `chrome.storage.local` on the device. Each entry records:

- Timestamp
- Site (hostname)
- Input type (paste, document, or send)
- Action taken (proceed, cancel)

- Categories of identifiers detected
- Count per category

Never recorded: the pasted text, the sent message, the file contents, the filename, or the detected values themselves. The log is not sent anywhere. If the extension is uninstalled, the log is removed by the browser along with the extension's storage.

Export

The user can export the log to CSV or JSON. Export is generated in the browser and written by a standard download prompt. No file leaves the device unless the user chooses to send it. No additional permission is required.

Coverage matrix

COVERAGE MATRIX. THREE PROTECTIONS, ONE ROW PER SITE.

Site	Send-time	Paste	Document upload
ChatGPT	Protected	Protected	Protected
Claude	Protected	Protected	Protected
Gemini	Protected	Protected	Protected
Perplexity	Planned	Protected	Protected
Microsoft Copilot (copilot.microsoft.com)	Planned	Protected	Protected
Microsoft 365 Copilot (m365.cloud.microsoft)	Not supported	Not supported	Not supported

Send-time protection on Perplexity and Microsoft Copilot is planned for a later release. Microsoft 365 Copilot (m365.cloud.microsoft) is not interceptable by a browser extension at all; that environment is governed by the tenant's Microsoft 365 and Purview controls, not by AI Leak Guard.

HIPAA Security Rule mapping

SECURITY RULE MAPPING. THE PRACTICE'S OBLIGATIONS ON THE LEFT, HOW THE TOOL SUPPORTS THEM ON THE RIGHT.

The practice's obligation	How AI Leak Guard supports it
§164.308(a)(1)(ii)(A) Risk analysis	Names workforce use of public AI tools as a risk category most small practices have not documented.
§164.308(a)(1)(ii)(B) Risk management	A technical control at the point of disclosure, reducing the likelihood of impermissible disclosure.
§164.308(a)(5) Security awareness and training	The warning is a teaching moment at the moment of risk, not an annual slide deck.
§164.308(a)(1)(ii)(C) Sanction policy	The AUP template gives a written policy staff can be held to.

This is not legal advice, and the practice's risk analysis remains theirs to conduct.

Limitations

WHAT AI LEAK GUARD DOES NOT DO

- Detection is pattern and context based. It will not catch every possible identifier, and it will sometimes flag text that is not patient information.
- **No OCR in V1.2.** Scanned or image-only PDFs, and encrypted PDFs, are reported as "couldn't inspect", never as safe.
- **Browser only.** Does not cover desktop applications, email clients, Teams, Outlook, or file transfers outside the browser.
- **Warn, not block.** The user makes the final decision. If they choose to send anyway, the extension will not stop them.
- **Supported sites only**, per the coverage matrix above. There is no universal AI coverage claim. Send-time protection is on ChatGPT, Claude, and Gemini today; Perplexity and Microsoft Copilot are planned for a later release.
- Does not make any organization HIPAA compliant. There is no HIPAA certification for software.

BOTTOM LINE

AI Leak Guard is a technical control at the point of disclosure: on-device inspection of what your staff paste and attach to consumer AI, with an honest warning and a user decision. It reduces risk. It is not a compliance product.

For the ungated CSV/JSON export as user-held evidence that the control is operating, see the HIPAA page (/resources/hipaa). For the exact data flow, see what data we receive (/resources/what-we-receive).

Describes AI Leak Guard v1.3.0. Page last updated 7 September 2026.