Privacy Policy

Introduction

This extension only intentionally and inevitably captures a small amount of personal information- the personal API key(s) used to power the web agent with a frontier AI model from one of the big AI labs.
Your personal API keys (for the AI cloud providers) are only stored locally in your browser.

If all tasks and batches of unsafe-action-annotations avoid pages with sensitive information, then the extension will not capture any further personal information.

Reasons why sensitive information may be collected

However, to avoid 'pages with sensitive information', the user would need to refrain from starting tasks or annotation-batches on websites which the user is signed in to, as well as websites which have forms (e.g. login, newsletter-subscription, or payment pages) where the browser might auto-fill credentials, payment information, addresses, phone numbers, etc. Additionally, sometimes a task that starts on one website will lead to navigating to a different website, and that second site could also have such sensitive pages.

In practice, then, this extension can potentially capture a great deal of personal information, depending on the tasks that the user asks the agent to perform and the batches of unsafe-action annotations that the user collects.
This is because screenshots (and under-the-hood webpage information called 'HTML') are collected both in the web agent use case (for the agent's decision-making and for later review by the user) and in the unsafe-actions annotator use case.
These screenshots and HTML data may contain personal information (e.g. from the user previously filling some things in, a password manager automatically filling things in, or the user already being signed in to a given website).

How customer information is used

This data is only ever sent to a server of the user's choice (see "AI Model Provider" configuration option). Even then, it is only sent so that it can be fed to an AI model to make decisions on the user's behalf.
The user should consult the privacy policy of their chosen AI Model Provider for how it handles customer data that is sent to its API's. Many of them commit to not training their models on data sent to their API's (a guarantee they frequently do not make for use of their chatbot web pages).

As of 2024 November 11th:
OpenAI commits to not training on data sent to its API's unless you opt into that. Likewise, Anthropic commits to not training on data sent to its API's unless you opt into that. Please be aware that the Google Gemini API's Free Tier will train on user data, but the Google Gemini API's Paid Tier will not.

The data is stored *locally* on the user's machine to allow for review of the agent's decisions and of the behavior of the extension. However, a given screenshot or other piece of stored information is automatically deleted after 14-28 days. In

rare cases, data may persist for several weeks longer than this if the browser interferes with the extension's delayed-action triggers.
Finally, the collection of potentially-sensitive data only occurs during a user-initiated task or batch of annotations, and the user can terminate the task or annotation batch at any time. Technically, after the end of an annotation-batch and before page reload or navigation, the extension still passively monitors mouse movements and page state changes to allow for the user to start another annotation batch with less delay. This monitoring would only cause the storage of log messages containing snippets of information about HTML 'iframes' and it would be unlikely for those snippets to contain sensitive data.

Full list of information collected

- Website URL's visited during tasks or explored during batches of annotations
- Screenshots of websites visited during tasks or explored during batches of annotations
- HTML data of websites visited during tasks or explored during batches of annotations (including any personal information that may be present on the page)
- Configuration choices, including API keys for AI cloud providers
- Logs of the agent's actions and decisions during tasks
- Logs of the user's choices when annotating unsafe actions
- Logs of miscellaneous internal activity in the extension's programming
- Mouse position within the browser's viewport and/or side-panel
    - This is only collected in some contexts, to implement features like the extra-scalable details pop-up for the status updates section of the side panel and the annotator mode's target element selection.