

AI REDACTION TOOLS

Reduce what sensitive information reaches outside AI services while preserving useful work.

Client Education Series

An exclusive publication of Sandbox Technologies, Inc.

Sandbox Technologies

3868 W Carson St, Suite 316
Torrance, CA 90503
www.sandboxtech.com

Tel. 424.207.5130

SANDBOX TECHNOLOGIES

Small Business Solutions
Enterprise GrowthPath®
EGP Secure®
EGP Cloud™
ConstructIT®

AI Redaction Tools

Reduce what sensitive information reaches outside AI services while preserving useful work.

How the process works

In this arrangement, an approved AI redaction service provides a route to an outside AI model. It examines prompts and supported attachments for configured categories of sensitive information, then removes or replaces detected values before sending the request onward. Where supported, the service restores the values it replaced with placeholders in the answer returned to the employee. Masked values remain generalized, blocked content is not sent to the external model, and files generated by the model may retain placeholders. This restoration is sometimes called rehydration.

For example, a tool may replace customer names and account numbers with placeholders while leaving enough context for the model to help draft a response. Which information it detects, which files it supports, and where the original information is processed depend on the selected service.

Choose it for appropriate work

Redaction can be a practical option for research, correspondence, and supported document work where masking information leaves the question useful. It may require less initial investment than dedicated AI hardware. Compare subscription costs, usage charges, implementation, support, and the number of employees who need access. A private server may be more suitable where original information must remain within a controlled environment, the work needs internal integrations, or dedicated capacity makes financial sense. Redaction may be a simpler starting point where masking information still leaves the task useful. Compare both approaches against the same work and information requirements.

The redaction service may itself receive the original information. Review its processing location, access controls, retention, logs, and provider arrangements. Confirm that the approved interface covers the way staff actually work, including attachments and any connected applications.

Test both protection and usefulness

Detection is not foolproof. A name, an image, an unusual identifier, or sensitive information implied by context may be missed. Removing the wrong detail can also change the question. Replacing a specific term such as tax fraud with a broad phrase such as a crime may produce a less useful answer.

Test representative information and file types before routine use. Check what reaches the outside model, what remains in logs, and whether restored answers are accurate. Define what happens when the tool cannot process a file safely. Direct staff to the approved route and review the configuration as information and uses change.

We're here to help and have structured services to address each of these areas individually or as part of a coordinated plan. To discuss what deserves attention first, contact your Sandbox Technologies Account Manager or Consulting CIO.

SANDBOX TECHNOLOGIES

Copyright 2026, Sandbox Technologies, Inc.