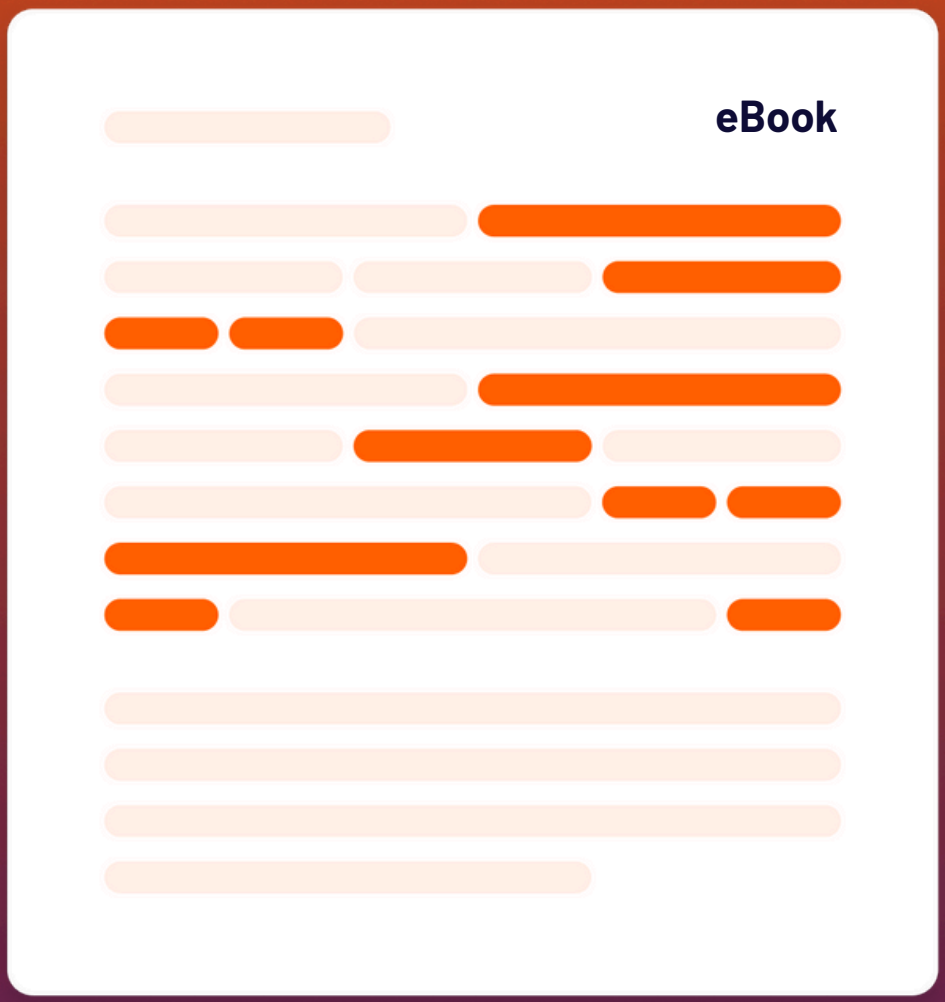
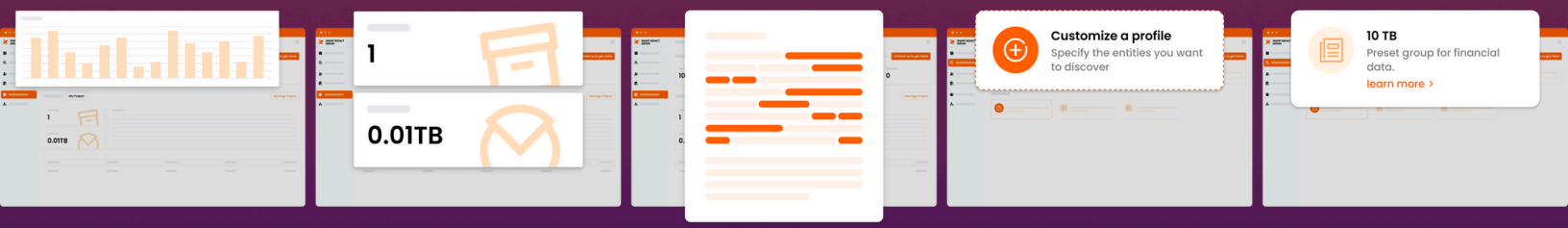
Automated Redaction for Your [redacted] Growing [redacted] Business

THE ULTIMATE GUIDE





The Ultimate Guide: Automated Redaction for Your Growing Business

In a tech-centric, data-driven world, privacy protection is paramount, not only for personal safety but also for legal compliance.

For those in heavily regulated industries, redaction plays a critical role in ensuring confidential data doesn't fall into the wrong hands. When done manually, redaction can be labor-intensive, requiring users to pore over documents one-by-one to obscure sensitive data.

Organizations must painstakingly redact sensitive information from thousands of documents before distribution to meet compliance requirements and avoid litigations.

What can companies do to make this process easier and more efficient?

Enter automated redaction!

Rather than printing out each file and scoring over text with a marker pen, you can use an automated redaction software to determine what text relates and redact sensitive data.

This guide will discuss automated redaction in detail, including how it works, the laws mandating redaction, what data needs to be redacted, and how businesses can save money using automated redaction. Let's dive in!

Table of Contents

What Is Automated Redaction?

Redaction means obscuring or censoring a part of a file (text, image, audio, video, etc.) This is done before releasing a document to hide some of its information for security, legal, or compliance purposes.

What about automated redaction?

Automated redaction is the use of technology to automate the process of obscuring sensitive information before its released to the public.

Traditionally, redaction has been done manually, but as mentioned, the manual process is incredibly labor-intensive and time-consuming. Manual processes are also prone to errors. And this is why automated redaction is preferred.

Automated data redaction addresses all the challenges of manual redaction. Using optical character recognition (OCR) and artificial intelligence (AI) tools, data can be scanned and sensitive information redacted for security purposes.

Automated redaction software describes tools that help users identify and blackout Personal Identifiable Information (PII) and Personal Health Information (PHI). These tools can come in the form of a standalone application that users upload documents into or as inbuilt features in a larger automation platform.

Either way, automated redaction increases throughput and processing speed by reducing the amount of labor needed from users.

What Information Need to be Redacted?

Redaction is usually done to comply with the legal requirements.

However, it's also an ethical responsibility to obscure confidential information of individuals and corporations. Here's a list of the type of information that should be redacted.

- Personally identifiable information
- Faces or persons
- Objects
- License plates

Let's briefly go over this information.

1. Personally Identifiable Information (PII)

Personal identifiable information, commonly referred to as PII, is any data that when used alone or with other relevant information can be used to identify a person.

Going by that definition, PII includes but is not limited to the following:

- Social security number (SSN)
- Passport number
- Driver's license
- Taxpayer identification number
- Financial account number
- Patient identification number

- Credit card number
- Personal address information

All this information needs to be redacted for privacy and compliance reasons. Should the information fall into the wrong hands, the consequences can be dire, including identity theft which can cause physical, social, and financial harm to an individual.

Identity theft has become prevalent in the US today, with 15 million people in the US falling victim to identity theft fraud each year. It is no surprise that Americans worry more about identity theft than being murdered.

Redaction can help organizations mitigate the incidences of identity theft and its consequences. Concealing personally identifiable information (PII) is a requirement under the Federal Rules of Civil procedure Rules 5.2. The rule requires any personally identifiable information unrelated to the court to be redacted.

2. Faces or Persons

Generally, a person's face or body doesn't need to be redacted. However, there are circumstances where this might be needed.

The first one is where an individual hasn't consented to be featured in a video. You commonly see this in prank videos where people don't even know they're being captured on video. It's ethical to conceal their faces for privacy reasons.

Secondly, certain compliances accord the "right to erasure" of individual data held by a company. An example of this is seen in the GDPR act article 17. This rule gives individuals the right to request organizations to erase any information they may have on them.

Under this rule, if an organization holds a video of a person and the person has requested for it to be erased, then the organization has two options:

- Delete the video from its system
- Redact the individual's face

The freedom of Information Act (FOIA) is another body where the redaction of people applies. If a government agency requests to disclose information to the public, the faces of the individuals in question need to be redacted during such disclosures.

Lastly, when submitting video or image files evidence during court proceedings, the faces of anyone appearing in the video who is not relevant to the case need to be redacted. However, there's a precedence to follow for that, as redaction, if not done accordingly, may be seen as hiding the truth.

3. Objects

The laws of redaction aren't restricted to personally identifiable information. They also apply to objects, though the laws regarding object redaction are less stringent.

Generally, buildings, screens, blackboards, etc., are redacted from video to maintain confidentiality. These objects could contain information that bad actors can utilize to cause harm to an organization.

4. License Plates

License plates can also be redacted for security reasons. While not considered personally identifiable identification, license plates can be used to identify the whereabouts of people, hence may need to be redacted.

In 2017, for instance, the California Supreme Court recommended carrying out such redaction to protect the driver's identity.

The Laws Mandating Redaction

Having explored various types of information that need to be redacted, let's now touch on the laws mandating redaction. Although numerous compliances might enforce redaction, we'll cover a few that we believe can impact your business.

1. HIPAA

Each year, thousands of patient records get posted on the dark web. Nowadays, it's possible to access all your medical information with just one click. Technology advancements have yielded many benefits, including easy access to one's medical data, communication with healthcare providers, and more advanced diagnostic methods.

But technology is a double-edged sword. Along with the benefits come many security risks. As much as we want to protect our medical data, it can easily fall into the hands of bad actors and get sold on the dark web.

Unlike other personally identifiable data that can sell as low as $5, stolen healthcare data sells for as much as $1,000 per record, according to Experian.

Sometimes, medical records can also be shared with third parties, including the government and educational institutions, for research purposes. In line with that, redaction of Protected Health Information (PHI) becomes necessary, which is what HIPAA addresses.

The Health Insurance Portability and Accountability Act (HIPAA) was passed in 1996 to provide rules for the privacy and security of PHI.

All entities in the healthcare sector and business associates that collect PHI are bound to follow the HIPAA rules. HIPAA requires redaction of PHI data before being shared publicly or with other professionals.

Failure to comply with the HIPAA rules may result in penalties ranging from $100 to $50,000 per violation with an annual max of $25,000 for a repeat violation." In certain circumstances, however, covered entities can disclose PHI without the patient's authorization. These include:

- When directed by law for public health purposes
- When reporting violence cases and abuse victims
- For clinical research purposes
- When required by a law enforcement agency

Under the HIPAA act, covered entities are held accountable for proper handling and redaction termed as "deidentification" of PHI before disclosure. Therefore, redaction

is required to obscure a person's sensitive data from medical records before sharing it.

2. GDPR

The General Data Protection Regulation act, commonly known as GDPR, is viewed as a tough uphill battle many organizations fail to conquer. Research by DLA Piper has shown that in 2020, GDPR penalties rose by 40%, amounting to a staggering $158 million.

At its core, the GDPR enforces policies to ensure that EU's citizen's privacy is protected and their PII is not exposed to unauthorized parties. And this is where redaction comes in. Rules governing redaction are contained in Article 9, Article 15, and Article 17 of the GDPR.

GDPR Article 9

Article 9 of the GDPR prohibits the use of personally identifiable data without explicit consent of the data subject unless the use falls under special circumstances defined by law. All personally identifiable data needs to be redacted before being shared online.

GDPR Article 15

As per Article 15, individuals have the right to access information organizations have on them and the related information linked to its usage. This may also include any video footage or audio recording data an organization holds linked to that individual.

GDPR Article 17

Article 17 provides individuals or companies the "right to erasure."

Hence, data subjects can request the organizations involved to erase their data. Data controllers can have the information deleted from their system or redacted from text, images, audio, or video files.

3. Freedom of Information Act (FOIA)

The Freedom of Information Act, commonly known as FOIA, allows the public to make a request to disclose information held by public agencies.

However, public agencies aren't obliged to disclose any information that falls under the nine exemptions of FOIA, personally identifiable information included. If any information falling under any of the nine exemptions needs to be disclosed, then agencies need to redact such information before sharing it publicly.

4. California Rules of Court

The California Rules of Court, Rule 1.201 section (a) requires all personal information presented as evidence in court to be redacted.

Rule 8.83 requires all records to be made available for public access.

However, personally identifiable information needs to be redacted before public disclosure. Other courts in different jurisdictions within the US also have similar requirements for public filings and disclosure.

How Automated Redaction Works

The automated redaction process is straightforward as it doesn't require a lot of manual input. The document redaction process only takes three simple steps.

- The document is scanned and converted into a digital format with OCR
- The software identifies PII in searchable digital files
- Sensitive information is obscured and the redacted file stored

Automated redacted allows for fast processing of large volumes of data without needing to view and process each file manually.

Using OCR for Redaction

Optical character recognition, commonly known as OCR technology, allows rules-based search engines to identify and mark sensitive data within documents for redaction. The OCR redaction process involves a few steps.

- Automated software scans the document for personally identifiable information. Note that paper and microfilm must be scanned first.
- The service cleans the file, redacting marked information
- The file is reproduced in its new, redacted format and restored with other files in cloud storage or document management system (DMS)

Manual vs. Automated Redaction: What's the Difference?

A typical manual redaction workflow involves a team of lawyers and support staff poring over large document files for hours. The team uses bounding boxes to redact elements present within by blurring, pixelating, or covering them with an opaque color.

The process is incredibly time-consuming and is not the best solution for enormous amounts of video content. Automated redaction, on the other hand, makes the process easier and more efficient through artificial intelligence.

The AI detects PII information, faces, objects, and other elements in the video, which are then redacted easily. While the process may sound efficient and straightforward, there may be obstacles along the way.

For instance, AI is not very accurate when working with low-bitrate video and often mislabels objects or assumes them to be duplicated. In some cases, it may consider some regions of a frame as a face, where there may not be any traces of facial elements.

In real-world scenarios, most of the videos obtained from sources such as CCTV, dashboard cams, and mobile phones are not in 4K or Full-HD 1080p. Hence, AI may not provide the best results working with such videos. Still, automated redaction is preferable to manual redaction for the reasons discussed in the next chapter.

Here's a quick comparison of manual vs. automated redaction.

Manual Redaction	Automated Redaction
Slow – Each document must be scanned manually, which can be time-	**Fast** – The software scans the document for PII and automatically

consuming when you have voluminous files	redacts it. Documents can be quickly redacted in bulk, saving time, and is ideal for voluminous work.
Manual redaction has a significantly higher level of accuracy for simpler or single files	More accurate in most cases but not the best option for low-resolution content.
Quality loss- physically marking the documents may degrade the quality of the documents. Manual redaction may also dirtify the document.	**No quality loss** – Automated redaction doesn't degrade the document's quality.
Not secure – Using analog methods compromises the security of the documents since many people view them. Analog methods of redaction also lead to resource wastage as you may end up using a lot of ink, paper, and people	**Highly secure** – Digitally redacted files don't need to be printed, scanned, or viewed by multiple people, so there's increased security. There's also no wastage of resources as you don't need marker ink or people to print and view the papers.
Not flexible – Analogue redaction methods aren't easy to adjust, especially if there are compliance changes.	**Flexible** – You can easily adjust redaction procedures to keep pace with changing compliance requirements

The Problem with Manual Redaction

Organizations relying on analog redaction methods run the risk of exposing sensitive personal information to bad actors.

Moreover, they are opening themselves up to potential compliance failures, which can have devastating consequences. Take the GDPR, for example. Organizations that violate these rules may face penalties, fines, court trials, reputational damages, and loss of customers.

It's no surprise that upholding compliance with regard to privacy has taken precedence in many organizations worldwide.

There's also the time and cost aspect. Scouring thousands of documents to physically redact with a marker pen is a painstakingly time-consuming task that's also subject to human error.

Unlike machines, humans suffer from fatigue, which inevitably leads to productivity loss. It's easy to miss some files, especially when you have a heap of documents to redact. Worst of all, OCR technology can often identify words that have been redacted manually.

Aside from manual redaction being time-consuming and expensive, they are also prone to errors, which can be costly. It is no surprise that redaction failures have hit the headlines in recent years and keep making the news.

An example of manual redaction failures can be seen in the case involving Trump presidential campaign chair Paul Manafort. In this case, lawyers failed to properly redact pleadings for Paul. As a result, journalists were able to read through the text

and publicize the redacted passages in the documents, revealing shocking information about the case, including details of Paul's involvement with a suspected Russian spy.

When to Redact Content

Organizations have different requirements for information retention and workflows that work along with redaction requirements. That said, here are five scenarios that might warrant you to redact content.

1. Upon Acquisition

When you receive company reports, you might want to redact certain information.

Perhaps the information received is not pertinent to the job function of the employee viewing the report. Or the reports contain sensitive information that shouldn't be accessed by staff who don't have clearance for such information.

For example, a sales rep might not need to know the diagnostic codes outlined in some insurance documents. In such cases, you could use page redaction to obscure the sensitive data.

The process should be run before the information, or reports, become available to everyone in the organization. It's imperative to have someone with proper clearance double-check the reports to ensure they are properly redacted.

2. Prior to Distribution

Your organizational reports may be distributed to parties inside or outside the organization. If the data included in the reports is necessary for specific workflows, it won't make sense to redact the documents upon acquisition.

When it's time to distribute the reports, however, it may be the time to redact the sensitive information. Again, you could use page redaction to obscure the page content or pattern matching redaction to remove specific texts or phrases from the document.

3. Once the Work is Done

Most organizations shred sensitive data once it's not needed anymore. Instead of shredding the documents you don't need, you could redact the pages that contain sensitive data.

This way, you get to keep the bulk for record-keeping without the heightened risk of exposing PII data to bad actors. Automated redaction makes the most sense for these scenarios as it's cost-effective and easier to deploy.

4. Prior to Archive

When a report or any other organizational document has reached its end use, redacting the sensitive information lowers the risk of exposing it to bad actors. Again, redaction ensures you get to keep the document for record-keeping without risking anything.

Your organization's archiving procedures should include using a redaction tool to obscure or strip out sensitive data.

5. Prior to Disposal

When it comes to disposing of documents and reports containing sensitive information, redacting as much information as possible leaves you with the highest level of security. This way, if a bad actor recovers a discarded report, it would be extremely difficult for them to recover any of the information contained in the documents.

How Businesses Can Save Money Using Automated Redaction

Text redaction is an unavoidable business activity, but the amount of time, energy, and resources it takes detract your team from focusing on revenue-generating activities.

For this reason, forward-thinking businesses resort to automated reduction, owing to its efficiency and cost-saving aspects. Here's how automated redaction can save your business money and improve operational efficiency.

Saves Time and Money

Automation has increased business efficiency in virtually all areas, and text redaction continues that legacy.

Leveraging artificial intelligence and optical character recognition (OCR) capabilities, users can dramatically cut redaction processing time and focus on revenue-generating activities. Here's how automated redaction tools help fulfill this task.

Increased Redaction Accuracy

As mentioned, manual processes are prone to errors, especially when redacting thousands of documents. Users can miss or omit important documents and fail to censor, leading to all sorts of problems—from compliance issues to compromised privacy to reputational damages.

There's also the risk of partly censoring texts, giving bad actors enough leverage to uncover the rest of the texts.

Users can leverage automated redaction software to automatically identify every sensitive information and block it out completely. AI-powered tools can even find variations of information, such as misspelled words or acronyms. This increase in accuracy ensures there are no errors that can cost your business money or drain your employees' time.

Quickly Discover Sensitive Information

When it comes to speed, humans can't compare with computers.

Even the most experienced user will go at a snail's pace compared to the processing speed of computers. Take OCR, for example.

These tools can scan documents at a fast speed and convert their contents into usable text in a blink of an eye. Users can then deploy AI to automatically redact sensitive data, resulting in a fast and efficient process.

Use Fewer Resources

Leveraging automated redaction tools saves your business money in many ways.

First, you won't need to outsource redaction work to third-party vendors or hire additional staff to perform the tasks. Outsourcing can be legally challenging to start and manage, especially if you've never outsourced before.

Plus, outsourcing may not be a good idea, especially when you have sensitive data about renowned figures that shouldn't be leaked to outsiders.

Secondly, automated redaction utilizes fewer resources. You only need a few employees to manage the process, and the task is not labor-intensive, so your employees can manage the redaction while performing other duties. This increases their productivity and saves you the money you would have otherwise spent hiring additional staff.

How Businesses can Limit Liability and Strengthen Credibility with Automated Redaction

According to 2021 IBM's Cost of Data Breach Report, a whopping 44% of data breaches included personally identifiable information (PII).

With an average cost of $161 million in penalties and fines for non-compliance (any unredacted record), PII and PHI non-compliance risk is expensive. In today's data-driven world, protecting sensitive information is no longer an option.

Failure to protect PII data will from your customer interactions will put you at risk of data breaches, lawsuits, expensive penalties, and loss of customer trust. Sure, manual redaction techniques can get the job done.

But legacy redaction techniques are prone to errors and compliance misses and lack the flexibility to redact the specific information you want to protect from the public.

Leveraging analog redaction methods to protect sensitive PII/PCI data from millions of customer conversations can be highly laborious, inaccurate, and unscalable. But with an automated redaction solution that leverages AI, personally identifiable information (PII) can be redacted automatically from documents.

Automated redaction also eliminates individual oversight or human error, preventing you from running into compliance issues or customer complaints. The liability of leaked client data resulting from failed redactions is also significantly reduced, adding more credibility to your business.

Common Data Redaction Challenges and How to Overcome Them

Data controllers encounter many challenges when redacting data.

A major challenge faced by many people is ensuring that all personal data is redacted or removed from a document collection.

If even a single document is missed out or the data redacted partly, it could potentially compromise the privacy of the individual and put the company at risk of violating compliance requirements. This makes accuracy critical to any data reduction method.

Another challenge is ensuring the redacted data cannot easily be restructured, as seen in Paul Manafort's case. Data can also be easily restructured if the names are removed but pertinent information, such as ages and addresses, is left out.

In fact, researchers from two European universities have designed a method that can correctly re-identify 99% of individuals in anonymized data sets with just 15 variables. The challenge here is redacting consumer data to make it impossible for anyone to read through or restructure the data.

Below are a few tips for overcoming the above challenges.

Leverage Automated Redaction Techniques

Unlike legacy redaction techniques, automated redaction is more accurate and reliable. These technologies scan the entire document (s) to identify PII and any other information that may need to be removed or obscured. This leaves no room for errors that could lead to data leaks, compliance risks, and other consequences.

Make sure to use automated redaction software from a trusted source designed specifically for data redaction.

Work with a Professional

If you are not sure of how to redact data, seek the help of a professional.

Data redaction is a complicated process, and working with an expert in the field can acquaint you with the process and help to ensure your redacted data isn't easily restructured. They can also help you choose a reliable data redaction platform that will increase the redaction efficiency and help you save.

Implement Access Control Policies

Even with robust data reduction software, data leaks can still happen if your access control policies are lacking.

Unauthorized personnel could still access data before distribution and leak it to bad actors. Research has shown that insiders are responsible for 22% of security incidents. To enhance the security and efficiency of your data redaction process, implement access control policies to ensure only personnel with the right clearance can access data.

Wrapping Up

Automated redaction is the use of technology to automate the process of obscuring sensitive information before its released to the public. Unlike manual redaction, automated redaction is more accurate and can lead to significant cost savings for your business.

Automated redaction can also help limit liability and strengthen your business's credibility. Since it's more accurate and efficient than manual redaction, automated redaction can help ensure compliance by reducing the risks that result from failed redactions.