



CALL RECORDING

Data Redaction

Admin
Quick Reference

 **MOMENTUM**

Powered by:  **miarec**

1. About Data Redaction

Redaction is used to remove sensitive content from the transcripts and audio recordings. For example, you can use data redaction to eliminate or mask sensitive personal information such as credit card numbers, phone numbers, or dates of birth from call recordings.

MiaRec data redaction engine relies upon the use of redaction rules to locate text in transcripts to redact. Data is redacted from both transcripts and the associated audio files.

⚠ Important

The redaction feature is designed to identify and remove sensitive data. However, due to the predictive nature of machine learning (i.e. transcription) and the variability of human language, MiaRec Data Redaction may not identify and remove all instances of sensitive data in your transcripts and recordings.

Since both false positive and false negative errors can occur, it is important to understand how both types of errors might affect your overall system. False negatives could lead to personal information leakage. False positives could lead to non-sensitive data being recognized as sensitive and thus redacted.

We strongly recommend that you review any redacted output to ensure it meets your needs.

The screenshot displays a 'MEDIA PLAYER' interface with a waveform and a 'TRANSCRIPT' section. The transcript shows the following text:

- Agent [0:20]: Thank you. How can I help you? **Redacted: Redact digits**
- Customer [0:24]: It's a *********.
- Agent [0:27]: Thank you. I see. You're insured with us. **What is the make year and model** of the vehicle you're calling about?
- Customer [0:35]: It's a ********* Toyota Sequoia.
- Agent [0:40]: Thanks. So how can I help you?

1.1 How to Apply Data Redaction: Best Practices

- [Create a set of redaction rules](#) and [test them](#) on a few individual recordings.
- Use advanced search criteria (on the **Recordings > Advanced Search** page), and find recordings in your platform that represent good candidates for redaction, for example, if you target calls with credit card number information, then search

by words "credit card" or even just words "card" or "number". Tag those recordings as a testing dataset with a [custom tag](#), for example, "test-redaction".

- [Run a job](#) in a test mode (highlight only) on the recordings that were selected for testing.
- Review the highlighted results and fine-tune the redaction rules as necessary.
- Once the results are satisfactory, [configure a job](#) to run periodically in production mode.
- Periodically, do a manual spot check if the redaction rules are still applicable to a new vocabulary that agents may use in conversations.

2. Configure Data Redaction

2.1 Add New Rule

To create a new rule, navigate to **Administration > Speech Analytics > Data Redaction** page, click the **Add** button to create a new rule or the **Edit** button to modify the existing rule.

On the Edit Data Redaction Rule page, you can configure:

- Name of the rule
- Color, which helps to visually distinguish different rules in the redaction results, when multiple rules apply to the same recording
- Optional description
- The option to redact the transcript/audio of the matched side or both sides (see **Note 1**)
- One or multiple expressions to match sensitive data

Note 1. Redact both sides vs the matched side only

MiaRec application by default records audio in two-channel format (stereo), where each side of a conversation is stored in a separate audio channel (left and right).

In most cases, we recommend redacting data in both audio channels due to a potential echo effect in the audio signal. Echo is a common flaw in telephony when the spoken words on one side of a conversation are replayed back on another side with a slight delay and distortion.

Let's look at a common case of redacting credit card numbers in recordings. A target for redaction would be a caller speaking consecutive digits after an agent spoke some trigger phrases like "What is your credit card number?" or "What is a three-digit security code?". Due to the echo effect, the spoken digits can be heard in both audio channels, possibly with slight distortion, for example, a clear "five" on one side and a distorted "fine" on another side. If a redaction is applied to one side only, sensitive data could be potentially recovered from the echoed signal.

Another common case is when an agent repeats after the caller each spoken word as some form of confirmation.

A redaction of both sides provides extra safety in removing sensitive data.

For each expression, you can configure the **Replacement** text that will substitute the original text in a transcript, for example, `*****` or `[redacted]`, and left and right padding (see **Note 2**).

Note 2. Left and right padding

The padding setting specifies how much data will be redacted to the left and right side from the targeted data.

We recommend redacting at least 500ms to the left and right from the detected sensitive data to compensate for the negative effects of echo and potential inaccuracy in transcription timestamps.

Transcription timestamps can fluctuate for 200-300ms on average from the ground truth. Such small fluctuation is hardly noticeable by a human, but can expose a word or two in audio if not redacted properly.

Administration > Speech Analytics > Data Redaction

Edit Data Redaction Rule

Name *

Color #3498d8 ✕
Format is #000000

Description

Redact audio Both sides Matched side

Redact transcript Both sides Matched side

EXPRESSIONS

Expression	Replacement	Left pad (ms)	Right pad (ms)	
1 <input [0-9]{4,}""="" style="width: 90%;" type="text" value="REGEX "/>	<input style="width: 90%;" type="text" value="*****"/>	<input style="width: 50%;" type="text" value="500"/>	<input style="width: 50%;" type="text" value="500"/>	✕

2.1.1 Expression Examples

This chapter demonstrates a few common examples of using the expression syntax to redact sensitive data in call recordings. We will demonstrate the capabilities of the engine by gradually increasing the complexity of redaction rules.

Goal: redact a credit card number from call recordings.

To perform a redaction, we will use the [MQL expression](#) to search for sensitive data in a transcript.

Assuming a credit card number consists of 16 digits, we could specify a simple rule, like:

```
R"[0-9]{16}"
```

Such a rule will search for the digits 0 to 9 that appear in a text exactly 16 times, for example, like in the phrase **"My credit card number is 1234567890123456"**.

In reality, a credit card number could be pronounced by a speaker in many different ways:

1. each digit spoken individually, like in **"My credit card number is 1 2 3 4 5 6 7 ..."**
2. digits are spoken in groups, like in **"My credit card number is 1234 5678 ..."**
3. digits are separated by a dash or other punctuation symbol, like in **"My credit card number is 1234-5678-..."**
4. a speaker uses filler words like in **"My credit card number is 1234 um 5689 ..."**
5. the digits are not a credit card number at all, like in **"A tracking number is 123456789"**

To cover all those variations, we need a more sophisticated rule. Let's improve it.

Our next try will be the following rule:

```
R"[0-9][0-9\-\,\. ]{2,}[0-9]"
```

This rule will search for a single digit 0 to 9 at the beginning of a text (defined by the first `[0-9]` expression), then, it expects to find two or more digits or punctuation symbols (defined by the middle `[0-9\-\,\.]{2,}` expression), and, finally, it expects to find a single digit in the end (defined by the last `[0-9]` expression).

Such an expression would allow us to match all of the above-mentioned variations of the credit card number, but it will match the tracking number phrase as well, which we don't want to redact.

To exclude the tracking number from a redaction, we can improve the expression further and add a trigger phrase to the search expression:

```
R"[0-9][0-9\-\,\. ]{2,}[0-9]" AFTER:10 ("credit card" OR "card number")
```

In this new expression, we use the operator `AFTER:10`, which instructs the redaction engine to search only for the digits that are spoken after triggering phrases "credit card" or "card number". `10` in `AFTER:10` means a maximum allowed distance between a trigger phrase and the searched digits (a distance is specified in words).

With such an expression, we can redact the digits that are related to credit card numbers but ignore moments in conversations when digits are spoken in other contexts.

We recommend testing the expression rules on your real call recordings before enabling the rule in production.

To test the expression, click the **Test Expressions** button and follow the instructions in the [Test Expressions](#) section.

Administration > Speech Analytics > Data Redaction

Data redaction rule «Redact digits»

Edit Test expressions Delete

Name: **Redact digits**

Color: #3498d8

Tenant: **Star Assistance**

Redact the transcript of the opposite side: Yes

Description:

EXPRESSION	REPLACEMENT	LEFT PAD	RIGHT PAD
REGEX "[0-9][0-9\-\,\.]{1,}[0-9]"	*****	500	500

2.2 Test Expression

The Test Expression page allows you to test the expression rule on real call recordings without altering the original transcript and audio file.

2.2.1 Procedure

To test the expression rule:

1. On the **Data Redaction** page, navigate to the existing data redaction rule and click **Test Expressions**.

Administration > Speech Analytics > Data Redaction

Data redaction rule «Redact digits»

[Edit](#)
[Test expressions](#)
[Delete](#)

Name: **Redact digits**
 Color: **#3498d8**
 Tenant: **Star Assistance**
 Redact the transcript of the opposite side: **Yes**
 Description:

EXPRESSION	REPLACEMENT	LEFT PAD	RIGHT PAD
REGEX "[0-9][0-9\-\,\.\]{1,}[0-9]"	*****	500	500

2. On the **Step 1** tab, select a call, on which you want to test your expression. You can apply filtering criteria to quickly find the target call recording.

Wide view

Test redaction rule

Redact digits

[Manage rule](#)

Step 1. Select a call | Step 2. Test expressions

Call - Transcript | Search query: card

[+ Add criteria](#) | [Search](#)

0-20 of 35

	DATE	TIME	DURATION	CALLER PARTY	CALLED PARTY
Select 2/2	Oct 8, 2022	5:23 PM	1:47	623441916396	542243175 (Tracy Butler)
store real quick. I guess I don't understand why that would take money off my gift card . OK. Let me see. Can I return something wouldn't give me return shipping? I mean					
Select 3/3	Oct 8, 2022	5:22 PM	0:15	557165206658	432548072 (Mallory Molina)
just Yes. came out that your card is showing twenty five dollars are available. OK, thank you so much. You bet have a great day and thank					

3. On the **Step 2** tab, apply the desired changes to the expression rules, if necessary and click **Test rule** to run a simulated data redaction process.

Digits

[Manage rule](#)

Step 1. Select a call Step 2. Test expressions

EXPRESSIONS

Expression	Replacement	Left pad (ms)	Right pad (ms)
1 REGEX "[0-9]{4,}"	*****	500	500

[+ Add Expression](#)

Redact audio Both sides Matched side Highlight


Redact transcript Both sides Matched side Highlight

[Test rule](#) [Save](#) [Save and close](#)

4. Review the results of the redaction for any mistakes and adjust the rule as necessary. Hover over the redacted text to see what rule was applied.

MEDIA PLAYER

Wide view



0 30 60 1:30 2:00 2:30 3:00 3:30 4:00 4:30 5:00 5:30 6:00 6:30 7:00 7:30

[Play](#) [x1](#) [x1.2](#) [x1.5](#) [x1.7](#) [x2](#) [Save audio file](#)

REDACTIONS

[Redact digits \(2\)](#) [Digits \(4\)](#)

TRANSCRIPT

Agent [0:20]: Thank you. Are you have your date of birth?
Redacted: Redact digits

Customer [0:24]: It's a *****.

Agent [0:27]: Thank you. I see. You're insured with us. What is the make year and model of the vehicle you're calling about?

Customer [0:35]: It's a ***** Toyota Sequoia.

Agent [0:40]: Thanks. So how can I help you?

When you are satisfied with the results, click the **Save** or **Save and close** button to save the redaction rule, if it was edited.

We recommend testing rules on multiple call recordings to confirm that rule covers a variability of spoken language.

2.3 Run the Job

Prerequisites:

- Data redaction works on transcribed recordings only.

2.3.1 Create the job

Navigate to **Administration > Speech Analytics**, switch to the **JOBS** tab and click **Create** to create a new job.

Administration > Speech Analytics > Data Redaction

Data Redaction

RULES **JOBS**

+ Create x Delete 0-1 of 1 < >

<input type="checkbox"/>	JOB NAME	STATUS	SCHEDULE	LATEST RUN	
<input type="checkbox"/>	Clear Sensitive Keywords in Calls	Finished	Not scheduled	Today, 8:50 AM	View Start Edit

20 per page 0-1 of 1 < >

Fill out the required configuration parameters:

Administration > Speech Analytics > Data Redaction

Add Job «Clear Sensitive Keywords in Calls»

Name *

Access scope *

- Unrestricted - All tenants, including System
- Tenants only - All tenants, excluding System
- One tenant

Test only * This is a test-drive. Write to a log file, but do not modify data

Parallel execution * workers

Mode *

- Full
- Incremental

- **Name** - give the job a distinctive name.
- **Access Scope** - specify for which tenants this job applies. This setting is visible only for a multi-tenant environment.
- **Test Only** - enable this option to run a job in a testing mode
- **Mode** - full or incremental mode.
 - **Full** - will process all the recordings every time the job is run.
 - **Incremental** - remembers which records have been already processed and do not process them on the next job run.

2.3.2 Rules

Under the Rules section, choose the [previously configured](#) rules, that should be applied to the recordings.

RULES

Data redaction rules x v x

[+ Add Rule](#)

2.3.3 Filtering Criteria

The optional filtering criteria allow you to limit what call recordings will be processed. For example, you can process the calls for a specified date interval.

FILTERING CRITERIA

x

[+ Add criteria](#)

2.3.4 Action After Successful Processing

Optionally, you can clear/assign a tag once recordings are processed. With this capability, you can create a chain of post-processing, and mark it with relevant flags. For example, you can process the recordings and mark them with a relevant tag. Then, you can email the recordings with this tag. And once the recordings are sent via email, you tell the system to clear the tag from the recordings.

ACTION AFTER SUCCESSFUL PROCESSING (OPTIONAL)

Clear a tag ✕ ▼

Assign a tag ✕ ▼

For more information about other settings that can be applied to the job, see [Advanced Settings](#).

Click **Save** to save your changes.

2.3.5 Start the job

To start the job manually, navigate to the **JOBS** tab and click the **Start** button next to the target job.

Administration > Speech Analytics > Data Redaction

Data Redaction

RULES JOBS

+ Create
✕ Delete

0-2 of 2 < >

	JOB NAME	STATUS	SCHEDULE	LATEST RUN	
<input type="checkbox"/>	Clear Sensitive Keywords in Calls	Not started	Not scheduled	Today, 8:39 AM	View Start ✎ Edit

20 per page ▼

0-2 of 2 < >

The message will appear informing you that the job has started. The **Status** flag will also inform you whether the run has been finished, aborted or is in progress.

Click **View** to see the results of the job run.

Administration > Speech Analytics > Data Redaction

Job is started successfully. ✕

Data Redaction

RULES **JOBS**

+ Create **✕ Delete** 0-1 of 1 < >

<input type="checkbox"/>	JOB NAME	STATUS	SCHEDULE	LATEST RUN	
<input type="checkbox"/>	Clear Sensitive Keywords in Calls	Finished	Not scheduled	Today, 8:51 AM	View Start Edit

20 per page ▼ 0-1 of 1 < >

The page with job results will display the number of processed records and the number of replaced keywords, if any.

Administration > Speech Analytics > Data Redaction

Run #6 of Job «Clear Sensitive Keywords in Calls» Delete

Job Name: **Clear Sensitive Keywords in Calls**

Status: **Finished**

Start Time: **Oct 13, 2022, 8:50 AM**

Total Execution Time: **3 seconds**

Stage: **Finished [Finished]**

Total replacements for sensitive keywords: **4**

Skipped (files not transcribed): **1**

Total records to process: **202**

Processed: **201**

Skipped: **1**

Remaining: **1**

2.4 Review Redaction Results

The following screenshot demonstrates the results of data redaction, where sensitive data is removed from both the audio file and transcription.

The screenshot displays a 'MEDIA PLAYER' interface with a waveform and a transcript. The waveform shows three redaction markers: a multi-colored bar at 0:30, a yellow bar at 4:00, and a multi-colored bar at 6:00. The transcript shows the following text:

Agent [0:20]: Thank you. Are you Can I have your date of birth?
Customer [0:24]: It's a ~~Redacted- Redact digits~~ *****.
Agent [0:27]: Thank you. I see. You're insured with us. What is the make year and model of the vehicle you're calling about?
Customer [0:35]: It's a ***** Toyota Sequoia.
Agent [0:40]: Thanks. So how can I help you?

The 'REDUCTIONS' section shows 'Redact digits (2)' and 'Digits (4)'. The transcript shows redaction markers (lock icons) next to the customer's responses.

When browsing through the recordings, you can quickly track which calls contain redacted data by spotting the lock icon next to the file.

<input type="checkbox"/>	TENANT	DATE	TIME	DURATION	CALLER PARTY	CALLED PARTY	TAGS	
<input type="checkbox"/>	PeriSolutions Ltd.	Jun 14, 2022	11:05 PM	2:25	252815881	583232816		<input type="checkbox"/>
<input type="checkbox"/>	PeriSolutions Ltd.	Jun 14, 2022	7:10 AM	7:59	168780123	928732197		<input type="checkbox"/>
<input type="checkbox"/>	PeriSolutions Ltd.	Dec 11, 2021	10:07 PM	0:24	467255031	307449420	encrypted	<input type="checkbox"/>
<input type="checkbox"/>	PeriSolutions Ltd.	Nov 22, 2021	9:11 PM	1:48	256814801076	271327389 (Kent Clark)	imported	<input type="checkbox"/>
<input type="checkbox"/>	PeriSolutions Ltd.	Nov 20, 2021	9:11 PM	1:48	256814801076	271327389 (Kent Clark)	imported	<input type="checkbox"/>
<input type="checkbox"/>	PeriSolutions Ltd.	Nov 19, 2021	9:11 PM	1:48	256814801076	271327389 (Kent Clark)	imported	<input type="checkbox"/>
<input type="checkbox"/>	PeriSolutions Ltd.	Nov 20, 2019	12:48 AM	1:42	+17029034115 (BTP)	+14084148202	imported2	<input type="checkbox"/>
<input type="checkbox"/>	PeriSolutions Ltd.	Nov 20, 2019	12:47 AM	1:01	+17029034115 (BTP)	+14084148202	imported2	<input type="checkbox"/>

Also, you can use the advanced search and apply **Call - Redacted** and **Call - Redacted By Rule** filtering criteria to search call recordings with redacted data.

Recordings

wide view

ALL CALLS
ACTIVE CALLS
MY CALLS
BY USER
BY CLIENT
UNASSIGNED CALLS
BY TAG
ADVANCED SEARCH

[Manage Saved Searches](#)

Call - Redacted

Is true (enabled)

+ Add criteria

Search
Save Search

wide view

Recordings

ALL CALLS
ACTIVE CALLS
MY CALLS
BY USER
BY CLIENT
UNASSIGNED CALLS
BY TAG
ADVANCED SEARCH

[Manage Saved Searches](#)

Call - Redacted By Rule

Is

Digits

+ Add criteria

Search
Save Search

Contact Us

A team of expert Momentum product customer support professionals are here to assist with technical issues, questions related to billing, feature usage, and service upgrades, as well as any other general inquiries you may have. Simply contact us and a friendly representative will help you with your request.

In order to provide optimal support, we recommend that the account owner (or a fully Authorized Contact) contact us via our toll-free support number regarding any critical or timely issues that may require troubleshooting or impact billing, and submit support ticket for questions or requests. A phone call is always the fastest way to get expert assistance. And always keep your account number handy to help us better assist you right when you need us.

Support by Phone

1.888.538.3960

Live phone support is available 24/7 for fast emergency assistance.

Online Support

www.GoMomentum.com/support

Bookmark this page to quickly and easily access the support phone numbers, ticket submission tools, and customer care information. For fast resolution, include the issue, details of your efforts to resolve (if any), and your organization's Momentum account number.

Online Documentation

Momentum University

Bookmark this page for an online knowledge base of training content that will equip you with all you need to learn about Momentum's best-in-class cloud services and features. This library includes a routinely updated and ever-expanding set of product user guides, quick reference tools, videos, and more. These resources are available 24/7.

