

AI IN SIGHTLINE

Sightline Help Documentation



Table of Contents

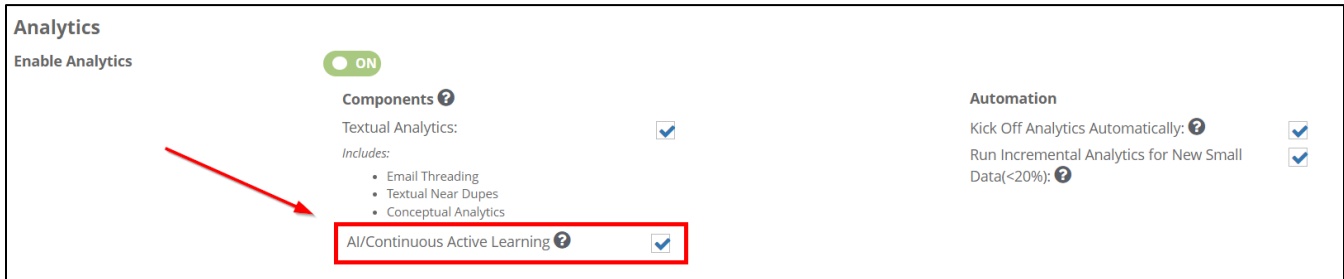
Table of Contents	2
AI in Sightline	2
Enabling AI in a Database	3
Enabling AI in a New Project	3
Enabling AI in an Existing Project	3
Enabling AI on a Tag	4
AI Models	4
Creating a New Model	6
Tag Selection	6
Training Configuration	7
Scoring Configuration	9
Editing an AI Model	10
Tag Selection	10
Training Configuration	10
Scoring Configuration	10
View Scoring History	11
Training Information	11
Scoring Information	11

AI in Sightline

Sightline can add continuous active learning (CAL) to any Tag in the database in a quick, easy, streamlined workflow to enable always on CAL with predictive hints in the DocViewer using AI. AI uses our Consilio Envize Engine and Continuous Active Learning to provide real time, always on, predictive hints to the Reviewer. Once AI is enabled for a Tag, an AI Model is created in the AI home page. This Model can be adjusted to hone the Training and Scoring to get more accurate scores and hints.

Enabling AI in a Database

When creating a new project in Sightline, there is a new option under the Analytics header for AI/Continuous Active Learning. To enable AI in a project, click the checkbox.



Enabling AI in a New Project

If this is a new project, AI/CAL analytics will automatically be run over all new data as it is added to the database. There is no need to manually run the AI indexing and analytics. By default, the AI/Continuous Active Learning option is turned off. If you know that you would like to use AI in your new project, be sure to check the box on the Create Project page.

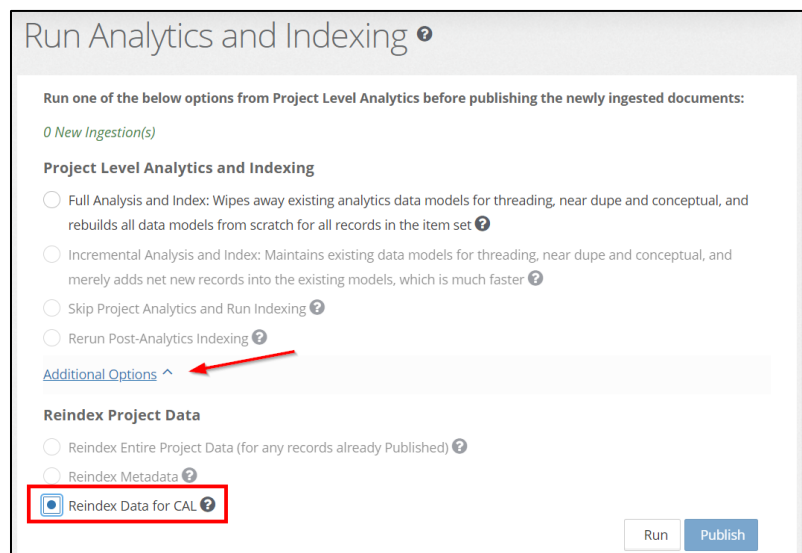
Enabling AI in an Existing Project

If this is an existing project that is being updated to enable AI/CAL, the User will need to first enable the AI/Continuous Active Learning from the Edit Project Page as a Domain Administrator. Once the AI/CAL has been enabled, the User will need to reindex the Data for CAL manually. After the initial, manual, run, CAL analytics will be run automatically on future data loads.

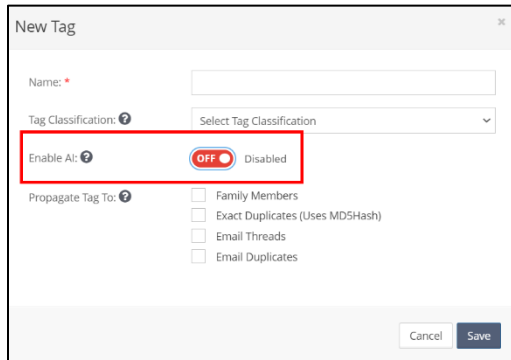
To reindex the Data for CAL, log into the database as a Project Administrator, and go to **Ingestions -> Analytics & Indexing**.

Click on the **Additional Options** link and then select **Reindex Data for CAL**.

Once selecting, click **Run** to execute the reindex.



Enabling AI on a Tag

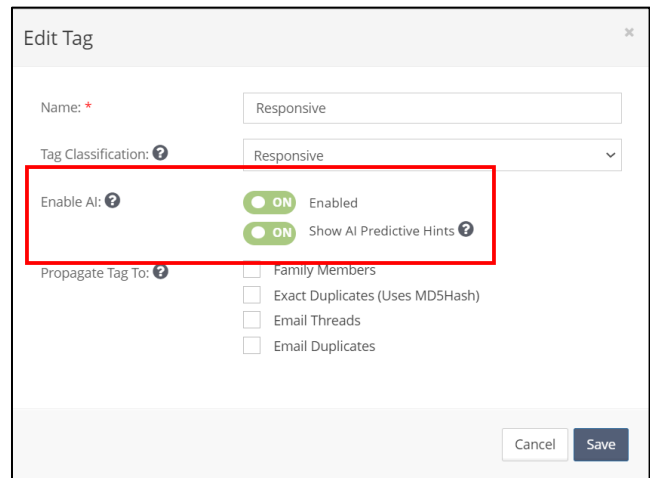


Continuous Active Learning in Sightline is based on the Tag. AI can be enabled on any Tag in the database by turning on the toggle for Enable AI. This can be done when creating a new Tag in the database or editing an existing Tag.

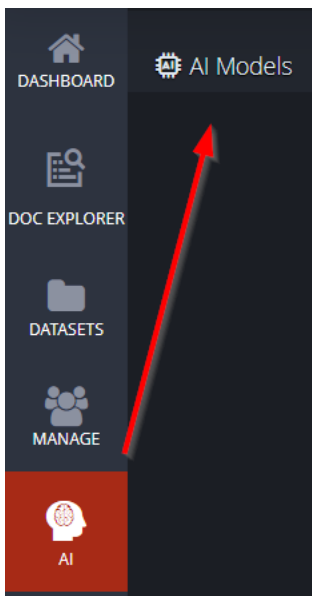
Additionally, AI Predictive Hints can also be enabled. AI Predictive Hints will provide a visible cue to the Reviewer as to how likely the document is to be responsive to the Tag.

By default, AI is enabled for the Responsive Tag in every new database. This can be turned off by visiting the Manage Tags and Folders Page, clicking on the Responsive Tag, and choosing Edit from the Action dropdown.

Once in the Edit window, turn the toggle Off for Enable AI. This will turn off AI for the Responsive Tag, and delete the AI Model. Optionally, the Predictive Hints can be turned off, while still allowing AI to run on the Tag.

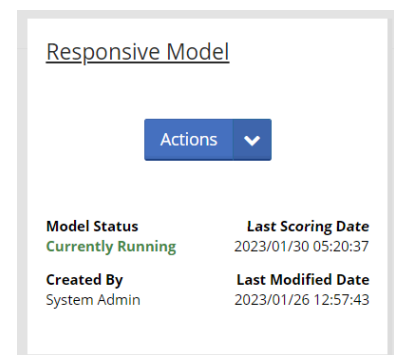


AI Models

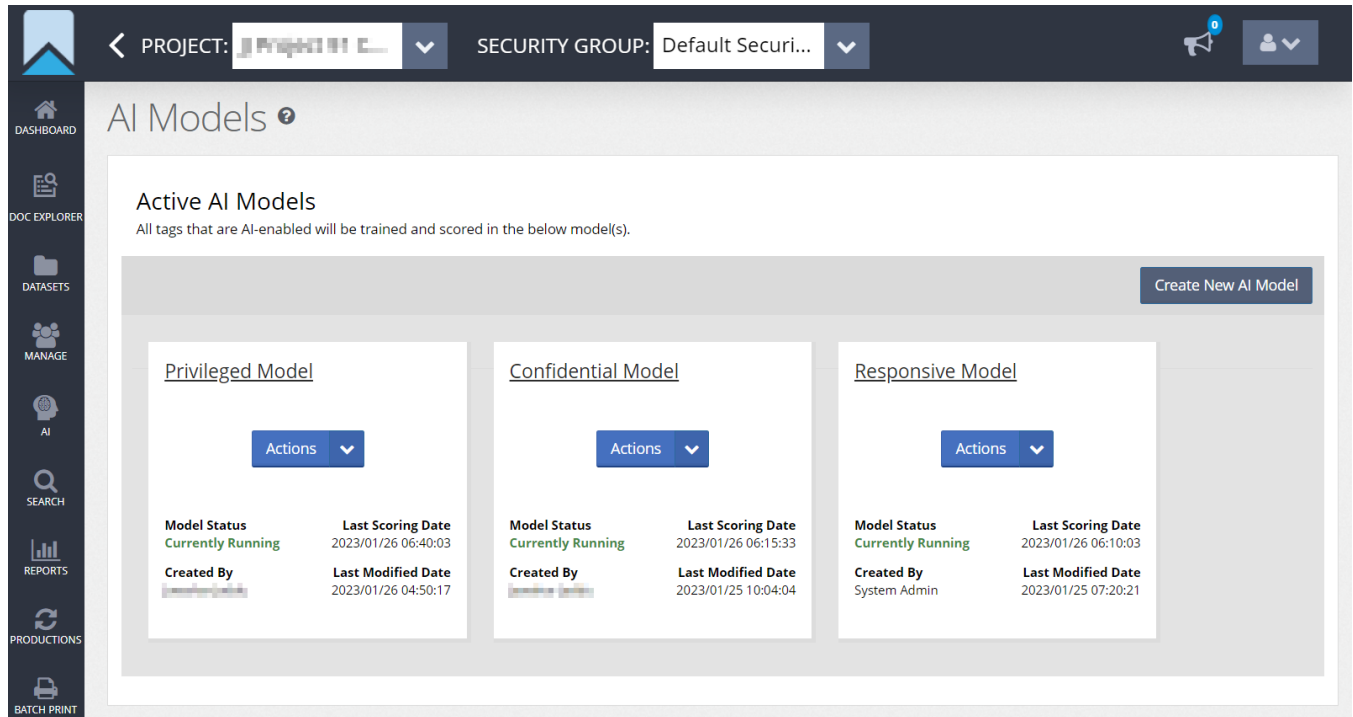


Once a Tag has had AI enabled, an AI Model is created. The AI Model will allow full control over how the AI is configured for that Tag. To view the AI Models, click the AI menu on the left-hand side and select AI Models. This will open the AI Models Page.

The AI Models Page contains all the models within the database. Each model is represented by a tile that shows the Model Status, the Last Scoring Date, the name of the User who created it, and the Last Modified Date. The Action dropdown allows for editing the Model, viewing the Scoring History and deleting the Model.



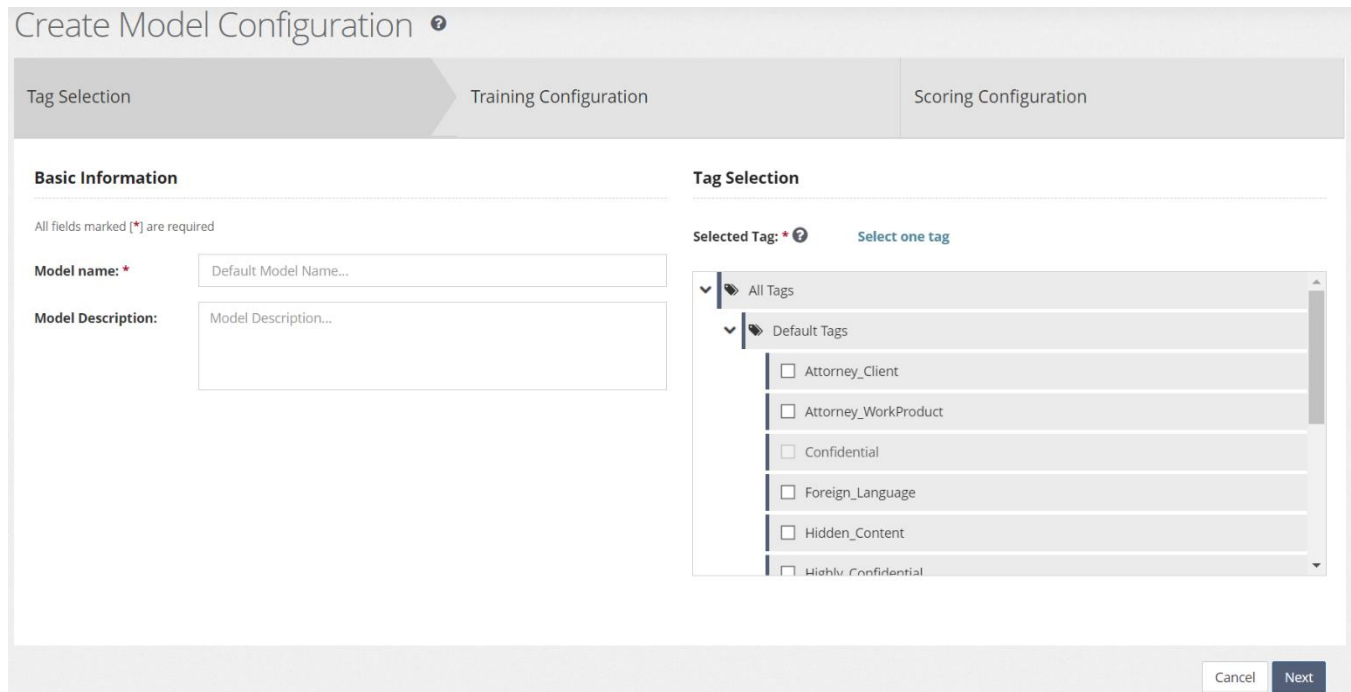
Once a Model is deleted, it cannot be restored. All scoring history and predictive hints are deleted with the Model. To re-enable AI for the Tag, a new Model will need to be created, and new scores generated from newly reviewed and tagged documents.



In the above screenshot, there are 3 Tags that have AI enabled. For each Tag that has AI enabled, there is a corresponding Model. The Models are represented as tiles that show basic information, and an Action dropdown that can be used for the additional actions.

Creating a New Model

In addition to being able to create an AI Model from the Manage Tags and Folders Page by enabling AI on the Tag itself, you can also create a Model from the AI Models Page. To do so, click the Create New AI Model button on the upper right of the Page to open the Create Model Configuration Page. From here you will see 3 tabs: Tag Selection, Training Configuration, and Scoring Configuration.



Tag Selection

From the first tab, on the left-hand side, provide a Name and optional Description for the Model. If the Model is created from the Tag, the Name will be the Tag name followed by the word, “Model.” If you enabled AI on the Confidential tag, the Model Name would be Confidential Model. From this page, name the Model manually. Consilio recommends following the same naming convention.

On the right-hand side of the Tab is the Tag Selection. You can only select one Tag per Model, and Tag cannot be used in multiple Models at the same time. If a Tag is already associated with a Model, it will be greyed out.

Make the appropriate selections and click Next to move to the next tab.

Training Configuration

The second tab presents the current configuration and allows for editing the configuration or reverting to the default settings. The left section is the Training Parameters, and the right section defines what is included in the training.

- **Training Type** will be Automated Training for all instances of AI in Sightline 9.1. Additional options may be made available in future releases.
- The **Automation Status** is a toggle that indicates if the Model is running as per the settings immediately below this option. The toggle may be turned on or off depending on the needs of the user.
- The next option sets the minimum number of documents to be Tagged before the training should start. The default option is 50, but that can be adjusted by typing into the box.
- The last section of the Training Parameters is the **Training Frequency**. This sets how often, and under what parameters the system updates the training of the Model.

The Training Frequency determines how, and when, the Model gets training. Sightline’s AI uses a continuous learning process, whereby the model is trained as the review is ongoing. The User can set the frequency of the learning to provide Model results that fit their needs.

Tag Selection
Training Configuration
Scoring Configuration

Training Parameters

This section contains the parameters to control when the automated training runs.

All fields marked [*] are required

Training type: [* ?] Automated Training

Automation status: [* ?] ON Currently Running

Minimum docs tagged to start training: [* ?] 50

Training frequency: [* ?]

Train and score model every 0 hrs 30 mins

AND at least 30 new docs tagged.

OR

Train and score when 500 new docs tagged.

Include in Training

This section allows you to choose the file types and tagging methods to be included in training.

**Note: When "Images" file type is selected, only images that have OCR extracted text will be included in training.

File types to be included in training: [* ?]

- Adobe Portable Document Format (.pdf)
- Archive Files(.zip, .rar, etc.)
- Database Files(.mdb, .mdf, etc.)
- Email(.msg, Notes, .eml, etc.)
- Images(.jpg, .gif, .tiff, .cad, etc.)
- Multimedia(.wav, .mp3, .mp4, etc.)
- Other(i.e.Uncategorized, unknown, etc.)
- Presentation Files(.PPT, .PUB, Visio, etc.)
- Spreadsheets(.xls, .xlsx, .csv, etc.)
- System Files(.exe, .dll, .lnk, .class, .java, etc.)
- Word / Text Files(.doc, .txt, .rtf, etc.)

Training methods to be included in training: [* ?]

- Bulk Tagging
- Tag Propagation

Cancel
Reset to Default Settings
Back
Next

By default, the Model will be trained every 30 minutes if at least 30 new documents have been tagged, or if at any point within those 30 minutes, 500 new documents were tagged. If 35 documents were tagged during review with 30 minutes, the Model will be trained with the newly tagged documents. If, at the end of the next 30 minutes, only 15 documents have been tagged, the model will not train.

If, at any time within those 30 minutes, 500 documents have been tagged, then the Model will train regardless of time. If the review rate is very high, and a lot of docs are being quickly coded, then you might not want to wait for 30 minutes to send the coded docs for training. This condition allows you to initiate a training and scoring round when 500 documents have been coded, even though it has not yet been 30 minutes since the previous training round. This will also consider bulk tagged documents, if that is enabled as a tagging method to be included in training.

As part of the Training Configuration Tab, the file types and tagging methods that should be included in the training can be adjusted. The analytics engine uses the text from these documents to train and mature the Model. By default, the below file types are included in training:

- Adobe PDF
- Email
- Images (only those images with text)
- Presentation
- Spreadsheets
- Word

The following filetypes will not be included by default: Archives, Databases, Multimedia, System Files and Other uncategorized or unknown files. To change any filetype either check or uncheck the associated box.

In addition to explicit tagging of documents, the system can also train the model using two other tagging methods: Bulk Tagging and Tag Propagation. By default, any documents that meet the filetype criteria and that have been bulk tagged are eligible for training the Model. Documents that are tagged through Tag Propagation – either by family or thread – are not eligible to train the model. Documents tagged using any of the CodeSync options are considered Explicit Tagging for Training purposes.

At any time during setup, you can revert to the Default settings by clicking the Reset to Default Settings at the bottom of the page.

Scoring Configuration

Create Model Configuration ?

Tag Selection
Training Configuration
Scoring Configuration

Scoring Population
This section allows you to configure the scoring Population.

All fields marked [*] are required

Population in score* ?

All Documents in the Security Group

Select a Subset of Documents

Include in Scoring
This section allows you to choose the file types to be included in scoring.

****Note:** When "Images" file type is selected, only images that have OCR extracted text will be included in scoring.

File types to be included in scoring: ?

- Adobe Portable Document Format (.pdf)
- Archive Files(.zip, .rar, etc.)
- Database Files(.mdb, .mdf, etc.)
- Email(.msg, Notes, .eml, etc.)
- Images(.jpg, .gif, .tiff, .cad, etc.)
- Multimedia(.wav, .mp3, .mp4, etc.)
- Other(i.e.Uncategorized, unknown, etc.)
- Presentation Files(.PPT, .PUB, Visio, etc.)
- Spreadsheets(.xls, .xlsx, .csv, etc.)
- System Files(.exe, .dll, .lnk, .class, .java, etc.)
- Word / Text Files(.doc, .txt, .rtf, etc.)

Cancel
Reset to Default Settings
Back
Save

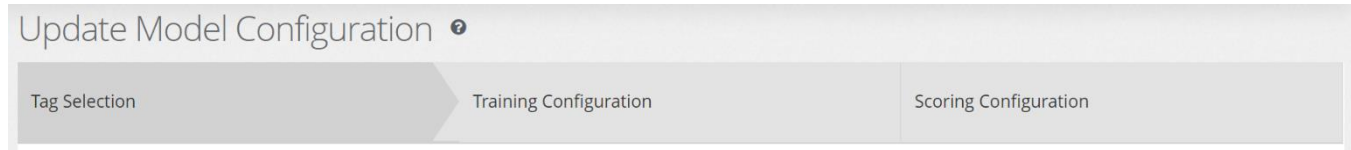
The final tab sets the configuration for the Scoring of documents. The Scoring Configuration is broken into two sections: The Scoring Population and Include in Scoring. In the first section set the population of records that will be scored. This can be either all the documents in the Security Group or a subset of documents based on a Saved Search or a Folder. By default, the Models will score all the documents in the Security Group. If this option is changed at a later point for a particular Model, any documents that are not in the selection will have their scored deleted for that Model.

If a subset of documents is selected, please note that the training population is always a subset of the scoring population, e.g., if you select a Folder of documents for your scoring model, only documents from that folder will be considered for the training model.

The second section of the Scoring Configuration is the Include in Scoring settings. Just like the training documents, certain filetypes can be selected for scoring. The system defaults are the same as the Training defaults and only Image files with text will be considered. At a minimum, the same file types for training must be used for scoring. However, you can include additional file types for scoring (even if you didn't select to send those files for training). The scores will help filter out irrelevant documents.

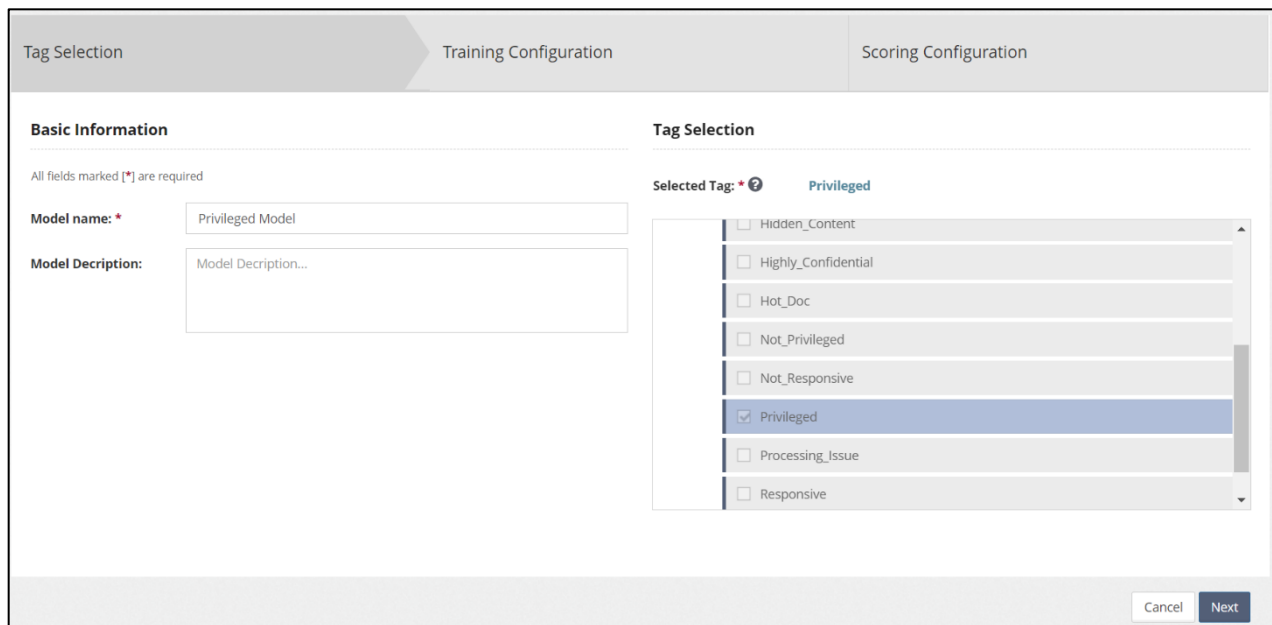
Editing an AI Model

To Edit a Model, either click the name of the Model, or use the Action dropdown to Edit the Model and access the Update Model Configuration Page.



Tag Selection

In the first tab, you can update the Model Name and Model Description. If you are editing an existing Model, you cannot change the Tag. If you wish to change the Tag you will need to delete the current Model and create a new Model.



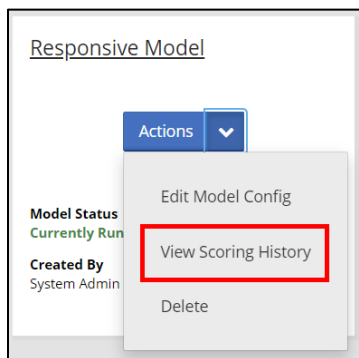
Training Configuration

The second tab presents the current configuration and allows for editing the configuration or reverting to the default settings. The left section is the Training Parameters, and the right section defines what is included in the Training. This is the same page as is presented when creating a new Model. See above for more information on this tab.

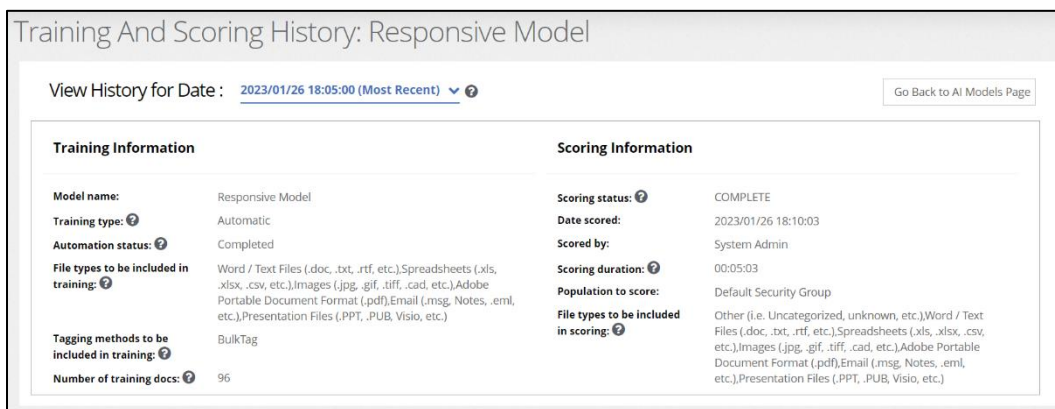
Scoring Configuration

The third tab presents the current configuration and allows for editing the configuration or reverting to the default settings. The left section is the Scoring Population, and the right section defines what is included in the Scoring. This is the same page as is presented when creating a new Model. See above for more information on this tab.

View Scoring History



Once the Model has been configured and run, the Training and Scoring History can be viewed using the Action dropdown on the Model’s tile. Choosing this option will open the Training and Scoring History Page. This Page will present the Training Information on the left-hand side, and the Scoring information on the right. The dropdown at top displays the results for the most recent Scoring but can be clicked to select a previous Date and Time for Scoring. The Date and Time presented are in UTC.



Training Information

The training information presents the following information:

- Model name – The name given when the Model was created
- Training type – Will be Automatic until future options are integrated
- Automation status – The status of the training (Completed, Currently Running, Not Running)
- File Types included in training – The options selected for this Model
- Tagging methods included in training – The options selected for this Model
- Number of training docs – The number of documents selected for training

Scoring Information

The scoring information presents the following information:

- Scoring status – The status of the scoring (If running, will present the percent complete)
- Date scored – The timestamp of when the specific scoring completed
- Scored by – This will be name of the User who created the Model for Automated, or the Name of the person who manually ran the Scoring
- Scoring duration – The amount of time it took for the training and scoring to complete
- Population to score – The scoring population against which the scoring was run (All docs in the Security Group vs a subset)
- File types included in scoring – The options selected for this Model