

HotSwapPll Technical User Manual



Table of Contents

- HotSwapPII Technical User Manual 1
- 1. Introduction3
- 2. Installation.....3
 - 2.1 Method 1: Using Poetry (Recommended) 3
 - 2.2 Method 2: Using pip 3
- 3. Main Features.....3
- 4. Sidebar Guide4
- 5. Tab-by-Tab Guide.....5
 - 5.1. PII Detection Tab 5
 - 5.2. Model Evaluation Tab..... 8
 - 5.3. Model Dataset Benchmarks Tab.....12
 - 5.4. Model Comparison Tab 14
 - 5.5. Custom Pipeline Tab.....15
- 6. Configuration Guide17
 - 6.1. Model Configuration17
 - 6.2. Detection Settings.....17
 - 6.3. Anonymization Options17
 - 6.4. Entity Mapping.....18

1. Introduction

HotSwapPII is a comprehensive system for detecting and anonymizing personally identifiable information (PII) in text documents. The application provides an interactive interface with multiple detection engines, anonymization methods, and evaluation capabilities.

2. Installation

Recommended to setup up a fresh virtual environment before installation below.

2.1 Method 1: Using Poetry (Recommended)

- i. Install Poetry if not already installed: `pip install poetry`
- ii. Install dependencies: `poetry install`
- iii. Run the application: `poetry run streamlit run app.py`

2.2 Method 2: Using pip

- i. Install dependencies: `pip install -r requirements.txt`
- ii. Run the application: `streamlit run app.py`

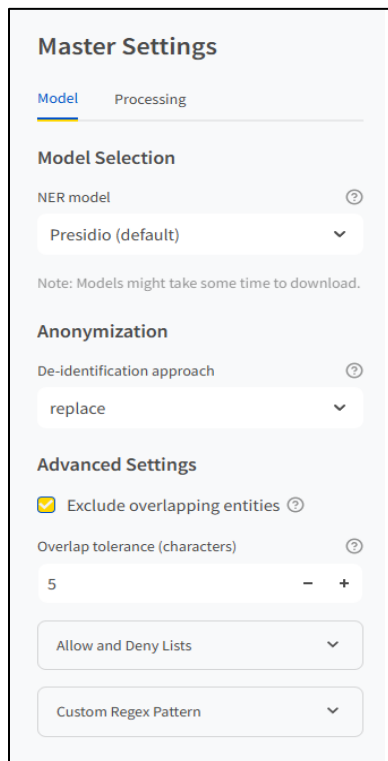
3. Main Features

- Multiple NER engines support (Presidio, SpaCy, HuggingFace Transformers, GLiNER)
- Various anonymization methods
- Synthetic data generation
- Model evaluation capabilities
- Entity-type performance metrics
- Customizable detection settings

4. Sidebar Guide

The sidebar has global settings that can be applied to all the features across the tabs.

This includes at a model level:



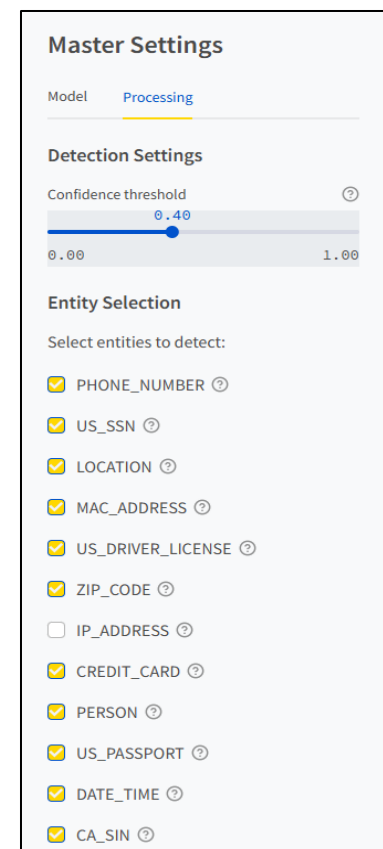
The screenshot shows the 'Master Settings' sidebar with the 'Model' tab selected. It includes sections for 'Model Selection' (NER model dropdown set to 'Presidio (default)'), 'Anonymization' (De-identification approach dropdown set to 'replace'), and 'Advanced Settings' (checkbox for 'Exclude overlapping entities' is checked, 'Overlap tolerance (characters)' is set to 5, and dropdowns for 'Allow and Deny Lists' and 'Custom Regex Pattern').

- a. A dropdown for model selection
- b. Anonymization format
- c. Handling of overlapping entities (if overlapping entities should be excluded or not)
- d. Overlap tolerance (how many characters of overlap should be considered as “overlapping”)
- e. Allow lists and deny lists (this is specifically for Presidio models)
- f. Custom Regex Patterns (this is specifically for Presidio models, only persists during session)

Figure 1: Model Settings

At the entity selection level we have:

- a. Minimum confidence threshold for predictions to qualify
- b. Entity selection (which entities to detect for)
 - i. This list contains the entities defined by the Presidio framework. The config files contain an exhaustive mapping between Presidio entities and other model entities and the expected ground truth entities. This list can be changed to match the ground truth entities instead



The screenshot shows the 'Master Settings' sidebar with the 'Processing' tab selected. It includes 'Detection Settings' (Confidence threshold slider set to 0.40) and 'Entity Selection' (a list of entities to detect, each with a checkbox and a help icon). The entities listed are PHONE_NUMBER, US_SSN, LOCATION, MAC_ADDRESS, US_DRIVER_LICENSE, ZIP_CODE, IP_ADDRESS, CREDIT_CARD, PERSON, US_PASSPORT, DATE_TIME, and CA_SIN. Most are checked, except for IP_ADDRESS.

Figure 2: Session settings

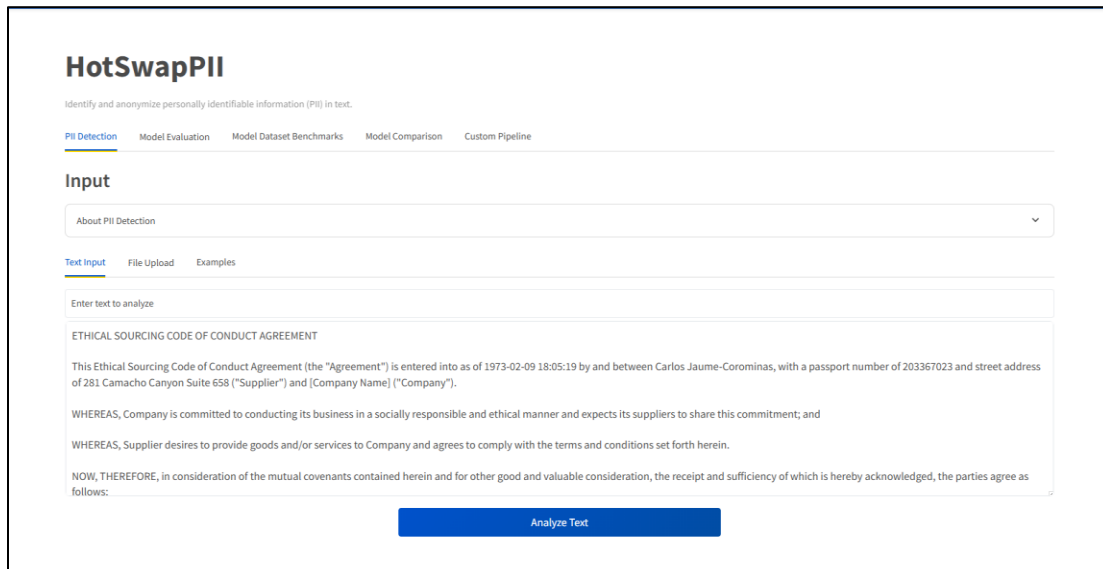
5. Tab-by-Tab Guide

5.1. PII Detection Tab

This is the main tab for processing text and detecting PII.

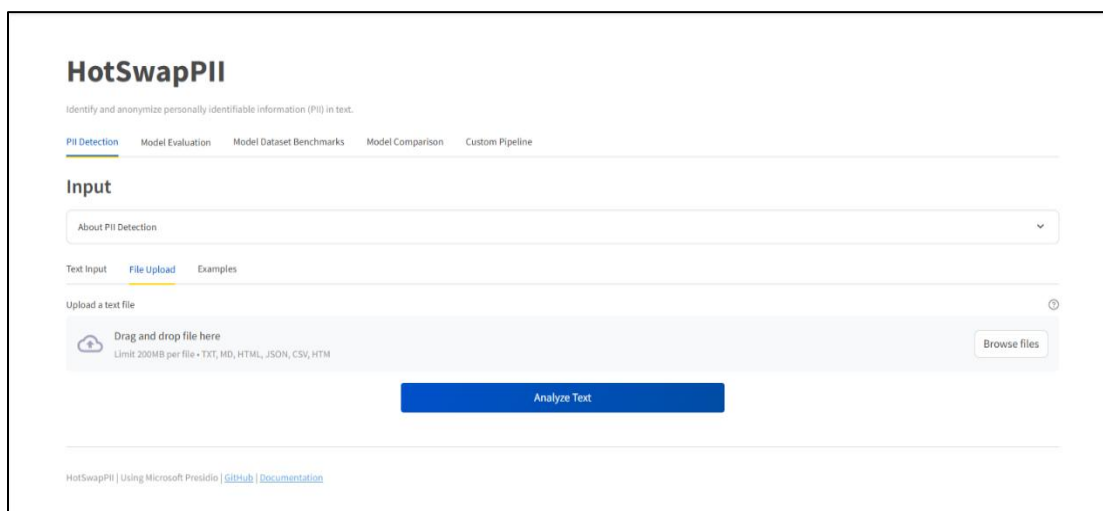
How to Use:

1. Input text either by:
 - a. Typing/pasting directly into the text area
 - b. Uploading a text file



The screenshot shows the 'HotSwapPII' interface with the 'PII Detection' tab selected. Under the 'Input' section, the 'Text Input' sub-tab is active. A text area contains a sample 'ETHICAL SOURCING CODE OF CONDUCT AGREEMENT' text. Below the text area is a blue 'Analyze Text' button.

Figure 3: Typing/pasting directly into the text area



The screenshot shows the 'HotSwapPII' interface with the 'PII Detection' tab selected. Under the 'Input' section, the 'File Upload' sub-tab is active. It features a file upload area with a 'Browse files' button and a blue 'Analyze Text' button. The footer includes links for 'HotSwapPII | Using Microsoft Presidio | GitHub | Documentation'.

Figure 4: Uploading a text file

2. Settings for processing will be picked from the Sidebar configuration
3. Click "Analyze Text" to process
4. The results have multiple features split over different tabs
 - a. Detection Results tab (View detection results)
 - i. Found entities

Found PII Entities

	Entity Type	Text	Start	End	Confidence	
0	CA_SIN	203367023		217	226	1.00
1	DATE_TIME	1973-02-09	132		142	0.85
2	PERSON	Carlos Jaume-Corominas		167	189	0.85
3	DATE_TIME	hours	999		1,004	0.85
4	DATE_TIME	ten (10) days	2,158		2,171	0.85

[Download Detection Results \(CSV\)](#)

Detection Summary

Total PII Entities
5

Most Common Type
DATE_TIME (3)

Avg. Confidence
0.88

Highlighted Text

ETHICAL SOURCING CODE OF CONDUCT AGREEMENT

This Ethical Sourcing Code of Conduct Agreement (the "Agreement") is entered into as of 1973-02-09 DATE_TIME 18:05:19 by and between Carlos Jaume-Corominas PERSON, with a passport number of 203367023 CA_SIN and street address of 281 Camacho Canyon Suite 658 ("Supplier") and [Company Name] ("Company").

WHEREAS, Company is committed to conducting its business in a socially responsible and ethical manner and expects its suppliers to share this commitment; and

WHEREAS, Supplier desires to provide goods and/or services to Company and agrees to comply with the terms and conditions set forth herein.

Figure 5: Found entities

ii. Highlighted text

Highlighted Text

ETHICAL SOURCING CODE OF CONDUCT AGREEMENT

This Ethical Sourcing Code of Conduct Agreement (the "Agreement") is entered into as of 1973-02-09 DATE_TIME 18:05:19 by and between Carlos Jaume-Corominas PERSON, with a passport number of 203367023 CA_SIN and street address of 281 Camacho Canyon Suite 658 ("Supplier") and [Company Name] ("Company").

WHEREAS, Company is committed to conducting its business in a socially responsible and ethical manner and expects its suppliers to share this commitment; and

WHEREAS, Supplier desires to provide goods and/or services to Company and agrees to comply with the terms and conditions set forth herein.

NOW, THEREFORE, in consideration of the mutual covenants contained herein and for other good and valuable consideration, the receipt and sufficiency of which is hereby acknowledged, the parties agree as follows:

1. Fair Labor Practices. Supplier shall comply with all applicable laws and regulations regarding employment, including but not limited to those related to wages, hours DATE_TIME, overtime, child labor, forced labor, and discrimination.
2. Human Rights Standards. Supplier shall not use or support the use of forced labor, whether in the form of prison labor, indentured labor, bonded labor, or otherwise. Supplier shall ensure that its employees are not subjected to any form of physical, sexual, psychological, or verbal harassment or abuse.
3. Environmental Sustainability. Supplier shall comply with all applicable environmental laws and regulations and shall take reasonable steps to minimize its environmental impact, including but not limited to reducing energy consumption, minimizing waste, and promoting recycling.
4. Monitoring and Reporting. Supplier shall maintain accurate and complete records to demonstrate its compliance with this Agreement. Supplier shall allow Company and its designated representatives to conduct audits and inspections of its facilities and records to ensure compliance with this Agreement.
5. Termination. Company may terminate this Agreement immediately upon written notice if Supplier breaches any material provision of this Agreement or if Supplier fails to cure any such breach within ten (10) days DATE_TIME after receipt of written notice thereof.
6. Governing Law.

HotSwapPII | Using Microsoft Presidio | [Github](#) | [Documentation](#)

Figure 6: Highlighted text

- b. Anonymized Text tab (View the anonymized text according to the global settings for anonymization)

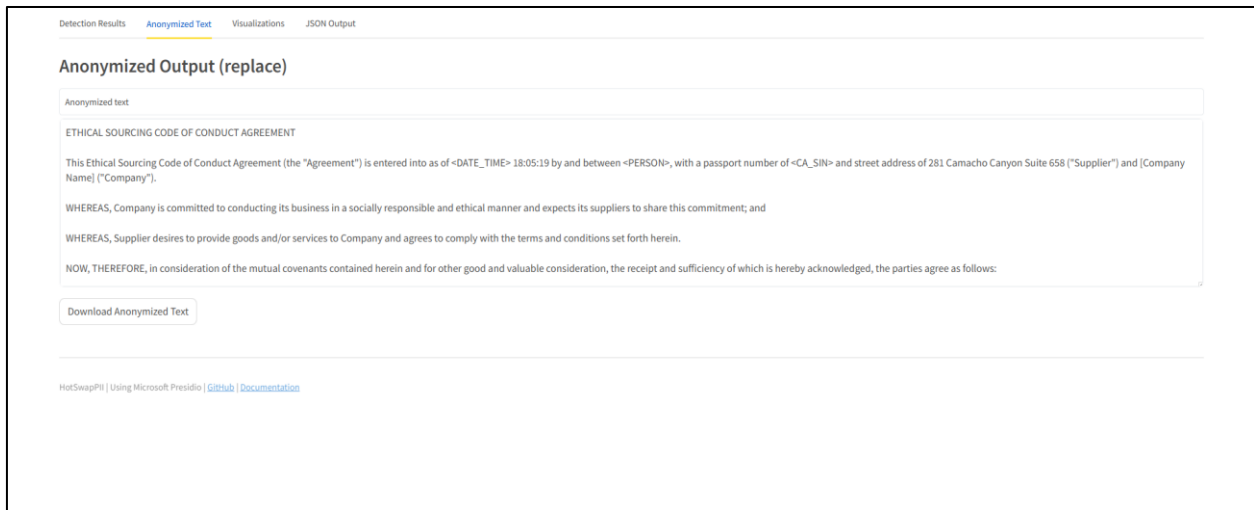


Figure 7: Anonymized Text tab

- c. Visualizations tab (View analytics related to the result)



Figure 8: PII Detection Visualization

d. JSON Output tab (View the results as a JSON output)



Figure 9: JSON Output tab

5.2. Model Evaluation Tab

Used for evaluating model performance against labeled datasets.

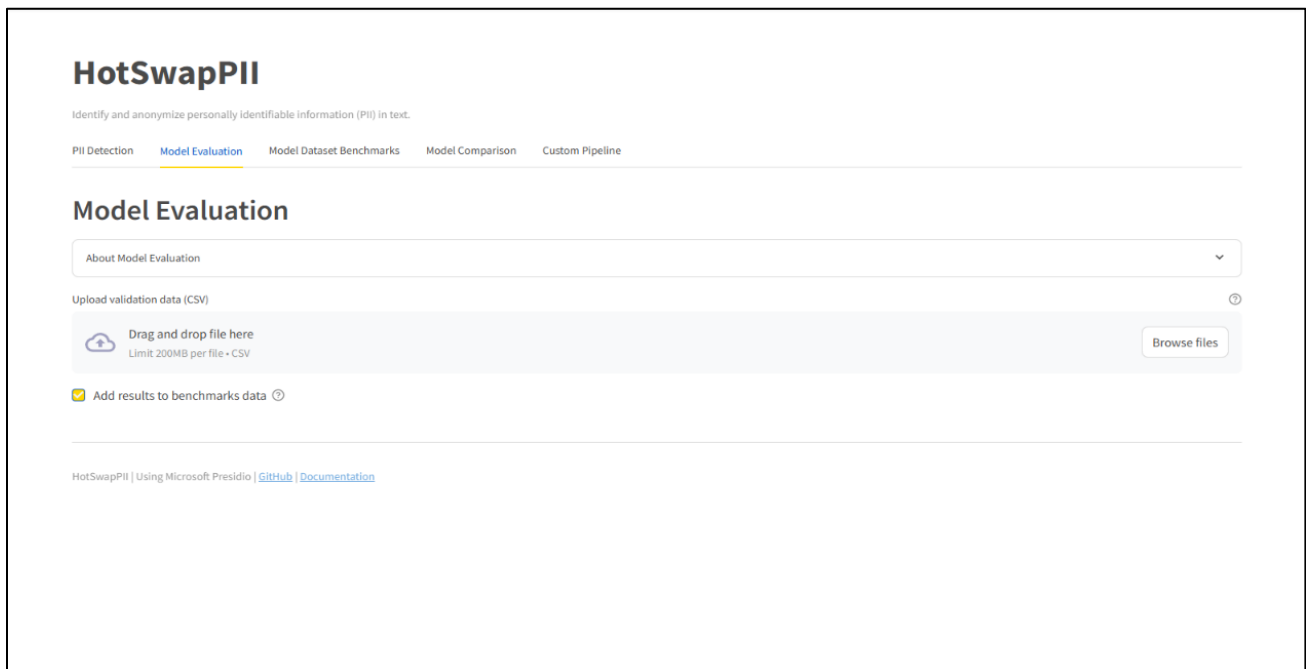
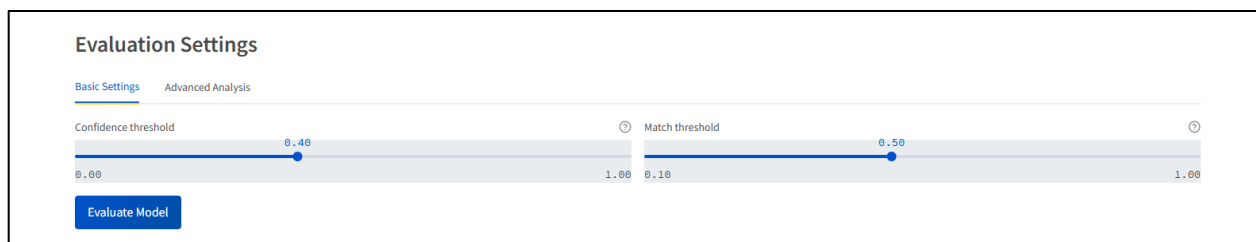


Figure 10: Model Evaluation

How to Use:

1. Upload a CSV file with data in the format as outline in the tooltip. The data should contain:
 - a. “text” column with input text
 - b. “label” column with JSON annotations
2. Choose model (from sidebar)
3. Configure evaluation settings
4. Run evaluation
5. View results (further explained below)



The screenshot displays the 'Evaluation Settings' panel. It features two tabs: 'Basic Settings' (active) and 'Advanced Analysis'. Under 'Basic Settings', there are two sliders. The 'Confidence threshold' slider is set to 0.40, with a range from 0.00 to 1.00. The 'Match threshold' slider is set to 0.50, with a range from 0.10 to 1.00. Both sliders have a blue dot indicating the current value. Below the sliders is a blue button labeled 'Evaluate Model'.

Figure 11: Evaluation Setting

Optionally, one can also choose to update the benchmark dataset metrics if using one of the benchmark datasets for the evaluation. This is ideal when wanting to test out changes to models and see how the changes affect the results on benchmark datasets. This will only work however if the name of the dataset being evaluated is the same as one of the datasets in the benchmark data folder/config. If you want to add a new dataset to the benchmark collection, the config list must be updated with the relevant dataset's file name. The dataset csv must also be added to the benchmark datasets folder to be able to view the sample data.

```

# File paths for benchmark dataset samples
DATASET_SAMPLE_FILE_PATH: List[str] = [
    "./data/benchmark_datasets/1_original_gretel_ai_conformance_data_500.csv",
    "./data/benchmark_datasets/2_another_gretel_ai_data_500.csv",
    "./data/benchmark_datasets/3_simple_generated_data_500.csv",
    "./data/benchmark_datasets/4_fake_data_with_variance_final.csv",
    "./data/benchmark_datasets/5_enron_data.csv"
]

# File paths for benchmark results
DATASET_BENCHMARK_RESULTS_FILE_PATH: List[str] = [
    "./data/benchmark_results/1_original_gretel_ai_conformance_data_500_results.json",
    "./data/benchmark_results/2_second_gretel_ai_data_500_results.json",
    "./data/benchmark_results/3_simple_generated_data_500_results.json",
    "./data/benchmark_results/4_fake_data_with_variance_final_results.json",
    "./data/benchmark_results/5_enron_data_results.json"
]

```

Figure 12: `config/benchmark_config.py`

The evaluation results include several different metrics that can be quite useful.

Firstly, there are two methods of metric calculations you can view results by:

- a. NERvaluate (<https://github.com/MantisAI/nervalue>)
 - i. This library provides further metric schemas to use when calculating model performance i.e. matching by Type Partial etc. Viewing the results for each schema can be chosen from its specific tab
- b. In House (our own metrics calculation algorithm)

Once you choose your calculation method there are multiple result tabs. For each type of calculation, you can view results at different granularities:

- a. Summary tab: This tab contains overall metrics for the model. Here you can see how it performed on the overall dataset.

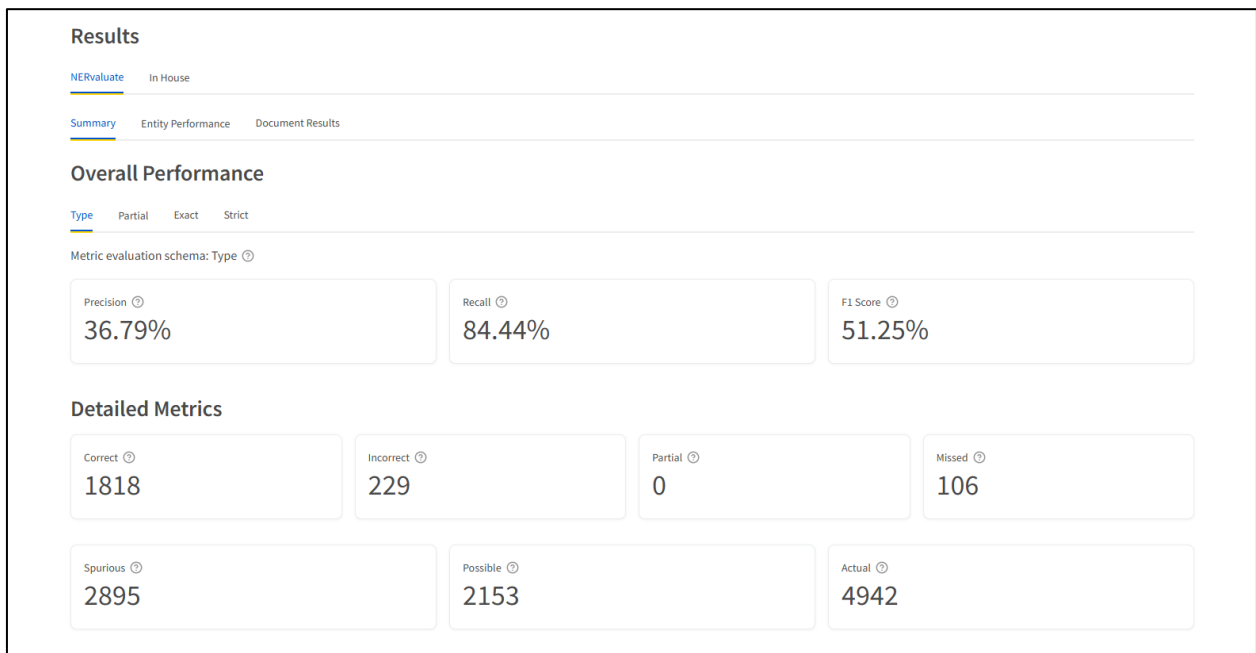


Figure 13: Results - Overall Performance

- b. Entity Performance tab: This tab contains metrics broken down by PII type. Here you can see how the model performed per tag along with raw values for number of predictions.

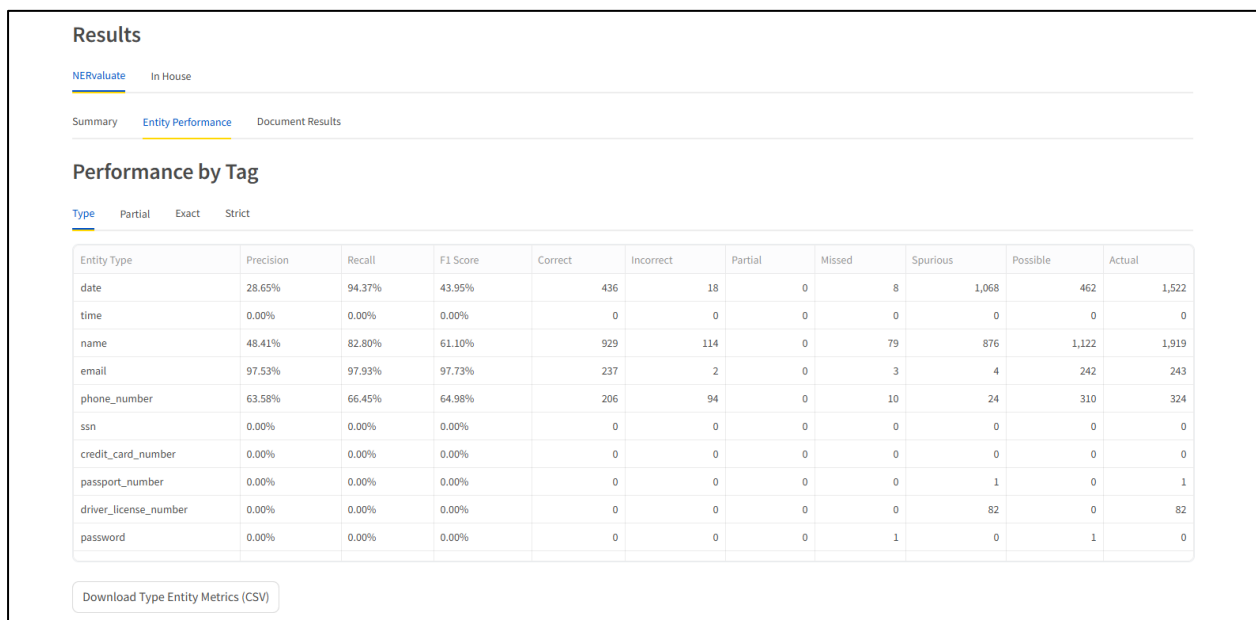


Figure 14: Results - Performance by Tag

c. Document Results tab:

- i. This is a document wise breakdown of metrics and raw counts of predictions

Results

[NERvaluate](#)
[In House](#)

[Summary](#)
[Entity Performance](#)
[Document Results](#)

Results by Document

Text	Precision	Recall	F1 Score	TP	FP	FN
A follow up to your questions on the accounting for the above: HPL - when Gosset began marking t	16.67%	50.00%	25.00%		1	5
John, EGS are asking whether Orlando is yours for PRC and bonus purposes? Presumably no doub	40.00%	100.00%	57.14%		2	3
As you know, Enron Net Works (ENW) and Enron Global Strategic Sourcing (GSS) recently executed	36.84%	58.33%	45.16%		7	12
I hope we didn't pay them on the unwind!! -----Original Message----- From: Lavorato, John Sen..	50.00%	61.54%	55.17%		8	8
Please make a note of the following: Discussion: Contractual Relations EES-NPW Date: Thursday,	40.00%	100.00%	57.14%		2	3
yes, i went over it with Errol. The positive value I gave you was change in existing deals versus L..	28.57%	66.67%	40.00%		4	10
I already had you down for 2000, did you do an incremental 2? -----Original Message----- From: La.	19.05%	57.14%	28.57%		4	17
You want to see perf reports now? -----Original Message----- From: Lavorato, John Sent: Thursda.	20.00%	57.14%	29.63%		4	16
If so, what time?	0.00%	0.00%	0.00%		0	0
spinnaker, one of fred's customer's , unwound a hedge today because they dont want exposure to	0.00%	0.00%	0.00%		0	1

Download Document Results (CSV)

Figure 15: Result by Document

5.3. Model Dataset Benchmarks Tab

Compare model performance across different datasets. The datasets along with their file paths are located in the benchmark_config.py file as mentioned above.

How to Use:

1. Select benchmark dataset from the dropdown
2. Choose model to benchmark (from the sidebar)
3. View dataset results
4. Choose which metric schema to view results by (Type, Partial etc)

5. Download as csv button below the results

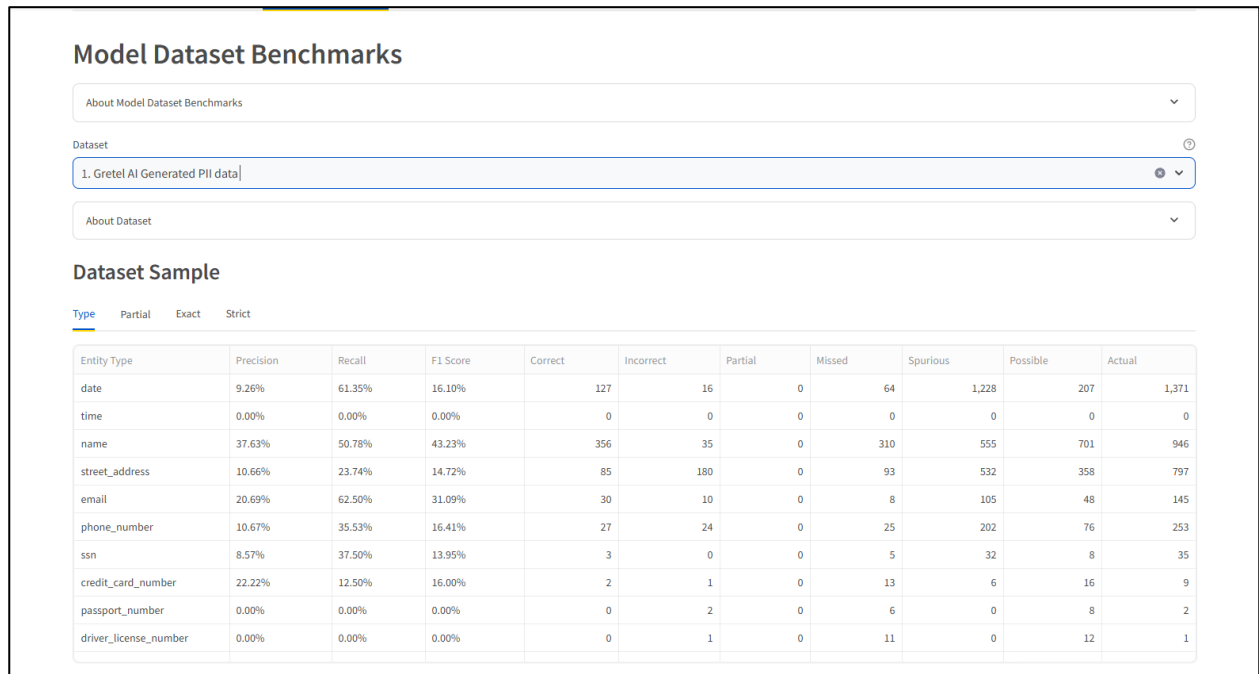


Figure 16: Model Dataset Benchmarks

5.4. Model Comparison Tab

Direct comparison between different models. The comparison happens off datasets in the benchmarks tab. Since it uses the results from constructed benchmarks and compares model results for models that have results for that dataset, the model selection in the sidebar is not applied here.

How to Use:

1. Select benchmark dataset from the dropdown
2. View side-by-side comparison results
3. Choose which metric schema to view results by (Type, Partial etc)
4. Choose which metric to compare by (F1, Recall, Precision)
5. Download as csv button below the results

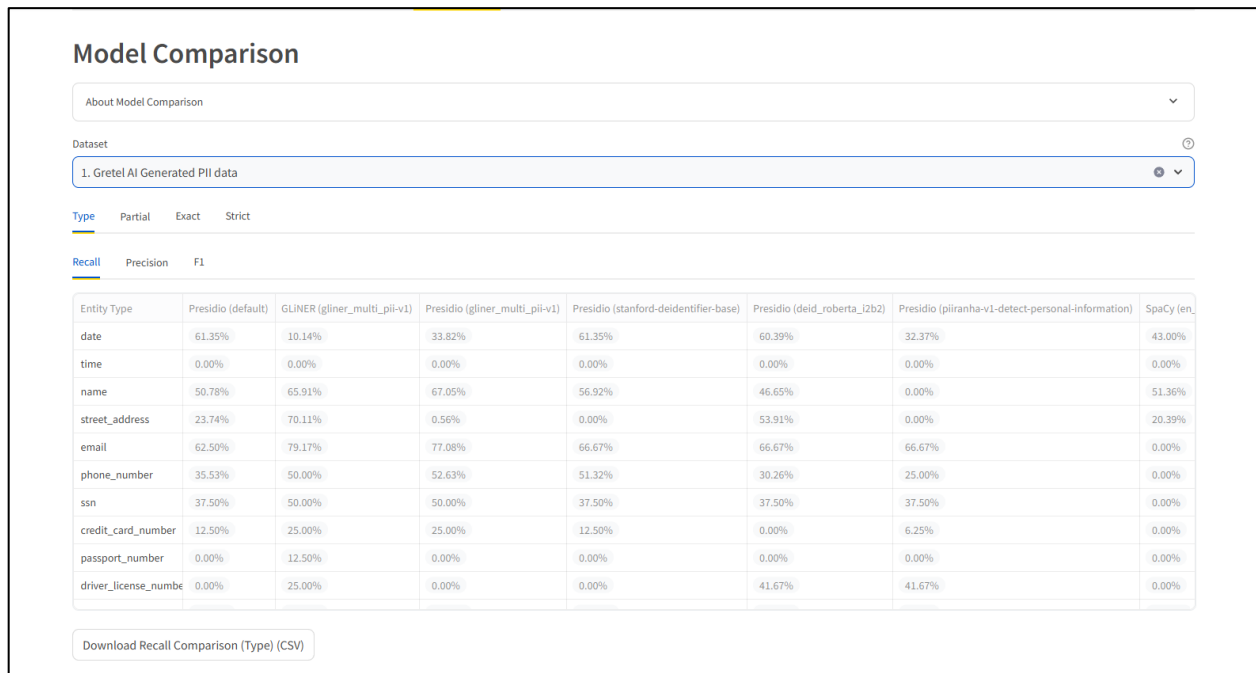


Figure 17: Model Metrics Comparison

5.5. Custom Pipeline Tab

Create, explore and manage custom detection pipelines. You can mix and match a custom pipeline here, giving different models precedence of predictions when it comes to specific entities based on previous results or if you want to test out different configurations.

How to Use:

1. (Optional) Select benchmark dataset from the dropdown:
 - a. You can select a benchmark dataset that you think performed according to your expectations or in alignment with a particular format of data. The custom pipeline tab will then load the configuration that is best for each entity type given that benchmark dataset.
 - b. The metric used for comparison is Recall using the Partial schema

Entity Type	Best Model	Recall Score
date	Presidio (default)	34.54%
time	Presidio (default)	0.00%
name	Presidio (gliner_multi_pii-v1)	37.59%
street_address	GLINER (gliner_multi_pii-v1)	43.58%
email	Presidio (gliner_multi_pii-v1)	45.83%
phone_number	Presidio (gliner_multi_pii-v1)	36.18%
ssn	GLINER (gliner_multi_pii-v1)	31.25%
credit_card_number	GLINER (gliner_multi_pii-v1)	12.50%
passport_number	Presidio (deid_roberta_l2b2)	25.00%
driver_license_number	Presidio (deid_roberta_l2b2)	29.17%

Figure 18: Custom Pipeline Builder

2. Adjust pipeline configuration accordingly

Deploy

Custom Pipeline Configuration

Select the model to use for each entity type:

date	time	name	street_address	email
<div>Presidio (default)</div>	<div>Presidio (default)</div>	<div>Presidio (gliner_multi_pii-v1)</div>	<div>GLINER (gliner_multi_pii-v1)</div>	<div>Presidio (gliner_multi_pii-v1)</div>
phone_number	ssn	credit_card_number	passport_number	driver_license_number
<div>Presidio (gliner_multi_pii-v1)</div>	<div>GLINER (gliner_multi_pii-v1)</div>	<div>GLINER (gliner_multi_pii-v1)</div>	<div>Presidio (deid_roberta_l2b2)</div>	<div>Presidio (deid_roberta_l2b2)</div>

Apply Custom Pipeline

Current Custom Pipeline Configuration

Entity Type	Selected Model
date	Presidio (default)
time	Presidio (default)
name	Presidio (gliner_multi_pii-v1)
street_address	GLINER (gliner_multi_pii-v1)
email	Presidio (gliner_multi_pii-v1)
phone_number	Presidio (gliner_multi_pii-v1)
ssn	GLINER (gliner_multi_pii-v1)
credit_card_number	GLINER (gliner_multi_pii-v1)
passport_number	Presidio (deid_roberta_l2b2)
driver_license_number	Presidio (deid_roberta_l2b2)

Download Custom Pipeline Configuration (CSV)

Figure 19: Custom Pipeline Configuration

3. Apply custom pipeline (Activate)
4. Use in the PII detection tab or Evaluation tab as needed (an alert will be present if the custom pipeline is active)
5. Download configuration as csv if needed

6. Configuration Guide

6.1. Model Configuration

Models can be added permanently by adding them in the `models_config.py`. Currently only two types of base models and three types of model families are available to set up as follows:

1. Base Model Selection:
 - a. Presidio: Microsoft's PII detection engine
 - b. Independent: Standalone models
2. Model Family Options:
 - a. SpaCy
 - b. HuggingFace (any Huggingface Transformer model)
 - c. GLiNER

More types of base models and model families can be added but there needs to be a robust set up in accordance with the implementation of the rest of the models to have it work correctly.

6.2. Detection Settings

A short summary of the detection settings is as follows:

1. Threshold: Configure confidence threshold (0.0 - 1.0)
2. Entity Types: Select which PII types to detect
3. Overlap Handling: Configure how overlapping entities are managed
4. Custom Lists:
 - a. Allow List: Always detect these terms
 - b. Deny List: Never detect these terms

6.3. Anonymization Options

A short summary of the redaction settings is as follows:

1. Redaction: Complete removal of PII
2. Replacement: Entity type placeholders
3. Masking: Character masking
4. Highlighting: Visual marking
5. Synthesis: AI-generated replacements

6.4. Entity Mapping

The entity list in the UI is a list of Presidio default entity types and any entity types loaded in the recognizers.yaml. The entity_config.py file has the settings for the default selected entities on startup (DEFAULT_ENTITY_SELECTION), the entities available for selection in the UI (CORE_ENTITIES), entity mapping for the GLiNER recognizer used with Presidio (GLiNER_ENTITY_MAPPING), the list of entities GLiNER searches for (GLiNER_LABELS) and a general mapping of ALL model entity types to the ground truth entity types i.e. the entity types expected in the ground truth labelling (MODEL_ENTITIES_TO_STANDARDIZED_ENTITY_MAPPING).

This setup can be changed as required e.g. use Presidio as the entity type for the unified mapping as well as the expected tags in the ground truth labelling.

This manual provides a comprehensive overview of the HotSwapPII system. For specific technical details or advanced configurations, please refer to the source code documentation or reach out to the development team.